



A Novel ABRM Model for Predicting Coal Moisture Content

Fan Zhang^{1,2,3} · Hao Li¹ · ZhiChao Xu^{1,3} · Wei Chen^{1,2}

Received: 29 September 2021 / Accepted: 2 December 2021 / Published online: 3 February 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract

Coal moisture content monitoring plays an important role in carbon reduction and clean energy decisions of coal transportation-storage aspects. Traditional coal moisture content detection mechanisms rely heavily on detection equipment, which can be expensive or difficult to deploy under field conditions. To achieve fast prediction of coal moisture content, a novel neural network model based on attention mechanism and bidirectional ResNet-LSTM structure (ABRM) is proposed in this paper. The prediction of coal moisture content is achieved by training the model to learn the relationship between changes of coal moisture content and meteorological conditions. The experimental results show that the proposed method has superior performance in terms of moisture content prediction accuracy compared with other state-of-the-art methods, and that ABRM model approaches appear to have the greatest potential for predicting coal moisture content shifts in the face of meteorological elements.

Keywords Coal moisture content · Meteorological elements · Deep learning · CNN · LSTM

1 Introduction

In September 2020, China set the path and timeline for achieving peaking carbon dioxide emissions and net zero carbon emissions, namely “carbon neutrality”. Currently, the energy industry is actively implementing an energy transition to gradually move towards the goal of achieving carbon peaking. At the same time, the coal industry is also promoting a new process of using fossil energy in a more environmentally friendly way, and achieving cleaner use

of coal and lower carbon emissions through technological innovation and intelligent technologies in coal mining, storage, and transportation [1]. And the dynamic monitoring of coal moisture in coal storage and transportation will provide effective technical means for coal storage and transportation, dust suppression in coal yards and clean utilization. Therefore, it is important to study coal moisture content prediction and monitoring technology to improve coal transportation and storage safety, reduce environmental pollution in the yard, and reduce carbon emission index of coal storage and transportation [2].

The online measurement methods of coal moisture can be categorized into two types, namely, direct hardware sensing and indirect soft sensing. Cutmore et al. [3] proposed a contactless microwave gauge for the on-belt determination of total moisture content in coal preparation plants, thus achieving the online detection of moisture content in coal. Deliang Zeng et al. [4] presented a soft sensing model for coal moisture based on the energy and mass balance of material in the inlet and outlet of a positive pressure, direct combustion, MPS-type mill, which reduces the delay time for moisture measurement. Yuman Wang [5] modified the microwave transmission coal moisture content detection method using least squares support vector machine algorithm to further improve the detection accuracy. Mao et al. [6] proposed a rapid detection of the total

✉ Fan Zhang
zf@cumtb.edu.cn

✉ Hao Li
lh@student.cumtb.edu.cn

ZhiChao Xu
xzc@cumtb.edu.cn

Wei Chen
chenwdavior@163.com

¹ School of Mechanical Electronic & Information Engineering, China University of Mining and Technology (Beijing), Beijing, China

² Key Laboratory of Intelligent Mining and Robotics, Ministry of Emergency Management, Beijing, China

³ Institute of Intelligent Mining and Robotics, China University of Mining and Technology (Beijing), Beijing, China

moisture content of coal using low-field nuclear magnetic resonance (NMR), thereby providing a new and effective tool for online detection of moisture content in coal from coal processing plants.

The above coal moisture online detection methods have been applied to coal storage yards and played a role in improving the decision of coal storage and transportation control, but there are still some shortcomings in detection speed, measurement accuracy and stability: (1) They require long moisture content analysis time with poor real-time detection performance of the target, thus have difficulty in providing real-time effective decision support. For example, the commonly used vacuum high temperature drying method requires at least 2 h to determine the moisture content. (2) During coal storage and transportation, water mist and dust is common interference in the environment. The above-mentioned methods have lower detection accuracy. (3) Dynamic monitoring models cannot be constructed, and the whole system may be susceptible to external perturbations such as meteorological changes and operational activities.

In recent years, with the extensive research on machine learning technology, deep learning has been widely used in many fields [7–11]. Currently, in regards to moisture dynamics prediction, neural network models are often applied to soil, crops, wood, etc [12–17]. There is a complex interaction between moisture changes and meteorological factors in these objects, and only by fully extracting the features of this complex relationship can reliable prediction results be obtained [18, 19]. According to relevant studies, the moisture content is a dynamic and continuous process in the time dimension, i.e., there is an interaction between the moisture content of adjacent moments [20]. The mechanism of change of moisture content in coal is similar to that of soil and other substances, which is susceptible to the influence of meteorological factors such as temperature and humidity, and shows dynamic continuity in the trend of change [21]. Thus, by obtaining a large amount of time series data on moisture and meteorology, a data-driven prediction model for coal moisture content can be built. This process does not require complex statistical or physical equations for a single variable, but is based entirely on the learning of a set of predictors for the variables of interest. Therefore, the technical idea of the research on the prediction of moisture content of coal stored and transported in open yards is considered in conjunction with the characteristics of coal moisture variability as well as in-depth extraction of the correlation feature analysis of moisture and meteorological factors [22]. In view of this, this paper proposes a machine learning-based online monitoring method for coal moisture content, which effectively solves the above problems by deep learning.

In this paper, based on deep neural network theory and its prediction method, a coal moisture prediction model based on the residual nets (ResNets) and bidirectional long short-term

memory (Bi-LSTM) is proposed. A one-dimensional convolution can ensure correct understanding of sequence data and better extracting features. A simplified ResNet can prevent a deep CNN from overfitting on small-scale datasets, and at the same time avoid the negative impact of redundant convolutional layers [23]. Bi-LSTM can learn the effect of both antecedent and postcedent sample features on the current time step [24]. The attention mechanism can further strengthen the key information in the time series by redistributing the weight of each time step's output [25]. The model combines one-dimensional ResNet, Bi-LSTM and attention mechanism, incorporating the feature extraction capability of CNN and the time-series feature memory capability of Bi-LSTM, to achieve the trend prediction of water content of coal in open dumps by training the model using multi-source data such as meteorological conditions in a specific region.

The rest of this paper is organized as follows. The architecture of our proposed methodology is presented in Section 2. Section 3 describes the experimental setup and evaluation metrics, as well as the analysis and discussion of the experimental results. Finally, Section 4 summarizes our conclusions.

2 ABRM: An Attention-based Bidirectional ResNet-LSTM Model

In order to enable the network model to correctly understand the serial data and extract its key features, this paper constructs a time series data consisting of two parts: meteorological features and coal moisture content features. Based on the data feature extraction capability of the CNN and the contextual association capability of long short-term memory (LSTM), we propose an attention-based bidirectional ResNet-LSTM model (ABRM). The network structure of the ABRM is shown in Fig. 1. The model consists of input layer, one-dimensional convolution layers, Bi-LSTM layer, attention layer, and output layer. The residual convolution layer is composed of one-dimensional-convolution-based ResNet to accomplish feature extraction; this is described in detail in Section 2.1. The recurrent layers and attention layers together perform the sequence prediction task, which is described in detail in Sections 2.2 and 2.3.

2.1 Convolutional Neural Network

The CNN is one of the most important core technologies in the field of computer vision (CV). It is a neural network structure composed of multiple convolutional layers, which have the characteristics of sparse interaction, parameter sharing, and equivalent representation. When performing image processing tasks, the input data of a CNN are a matrix of image pixels so that its pixel points in each direction have essentially the same

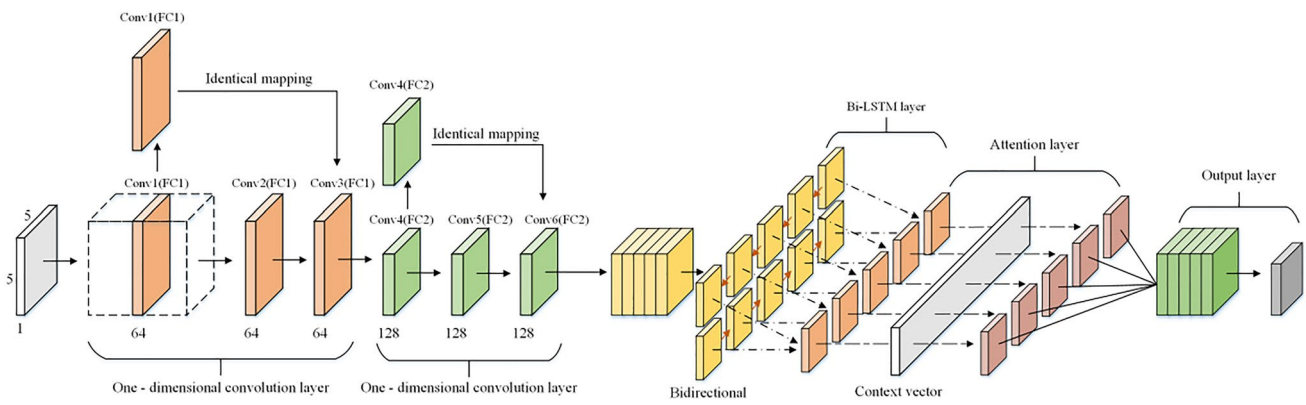


Fig. 1 Network structure of the ABRM

correlation with each other. While using the time series data as the input of the CNN, the sequence step length is usually taken as the width, the number of features as the height, and each feature value as the pixel point value. Assuming that the sequence step size is n and the number of features is d , the input to the CNN forms an $n \times d$ matrix. It is obvious that the input sequences are not equally correlated in different directions. In this study, a 1D CNN is used to ensure that the model understands the data correctly while making the width of the convolution kernel equal to the number of features of the sample points. With the top-down sliding convolution of the convolution kernel, the features and interrelationships of each time step can be extracted efficiently.

When the CNN is deep and the size of the dataset is not huge, the overfitting problem is easily triggered. ResNet can avoid this and has shown excellent performance [23]. Moreover, ResNet adds identity mapping between different convolutional layers so that the information in the network can flow across layers, ensuring that the model does not suffer from increasing errors as the depth of the network increases. In other words, the deep network has the possibility to transform autonomously into a shallow network, thereby avoiding the negative effects of redundant convolutional layers. Figure 2 shows the structure of the basic residual unit in ResNet. The calculation of the residual unit is as follows:

$$x_{l+1} = x_l + F(x_l, \{W_l\}). \tag{1}$$

where x_l is the identity mapping part; $F(x_l, \{W_l\})$ is the residual mapping part, which is composed of two layers of convolution (see Fig. 2). x_{l+1} is the output of the residual unit.

2.2 Bidirectional Long Short-term Memory (Bi-LSTM)

LSTM is an important technique in Natural Language Processing (NLP) [26] and solves the gradient vanishing or

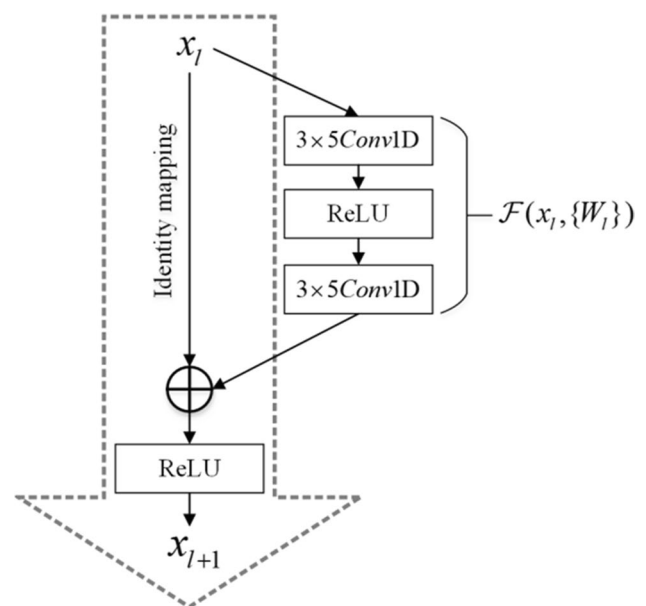


Fig. 2 Structure of the residual unit

exploding problems that may occur in traditional Recurrent Neural Networks (RNNs) when processing longer sequences. The LSTM was designed with three gate structures: the forget gate, update gate (input gate), and output gate. Through the weighting of the input vector by these three gates, LSTM can selectively extract the antecedent information, and better understand and preserve the global information to accomplish more accurate prediction. Therefore, LSTM has been widely used in tasks such as speech recognition, sentiment analysis, and text analysis. Bidirectional LSTM can read the sequence data in the reverse direction for re-learning. The final output of Bi-LSTM is determined by forward and reverse learning. This structure further ensures the backward and forward correlation of the sequence prediction results. Therefore,

in some tasks, Bi-LSTM will perform better than LSTM. Taking the input time step t as an example, the specific computational flow of the Bi-LSTM is as follows.

- (1) The output h_{t-1} of the previous time step and the input x_t of the current time step are combined, and sent to the forget gate for selective forgetting:

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f), \tag{2}$$

where f_t is the output of the forget gate, σ represents the Sigmoid activation function, W_{xf} and W_{hf} are the weights assigned to x_t and h_{t-1} by the forget gate, respectively, and b_f is the bias of the forget gate.

- (2) The stitching vectors of x_t and h_{t-1} are selectively memorized through the input gate as shown in Eq. 3. Meanwhile, x_t and h_{t-1} are scaled by the activation function \tanh to produce the candidate memory cell \tilde{c}_t that stores the information of the current time step, as shown in Eq. 4.

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i), \tag{3}$$

$$\tilde{c}_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c). \tag{4}$$

Here, i_t is the output of the input gate, W_{xi} and W_{hi} are the weights assigned to x_t and h_{t-1} by the input gate, respectively, and b_i is the bias of the input gate.

- (3) The contents of the antecedent memory cell c_{t-1} are updated by deciding which new information in the current candidate memory cell \tilde{c}_t will be stored; the latest memory cell c_t is obtained as follows:

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t. \tag{5}$$

- (4) The output gate determines what information should be included in the output state o_t of the current stage:

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o), \tag{6}$$

where o_t is the output of the output gate, W_{xo} and W_{ho} are the weights assigned to x_t and h_{t-1} by the output gate, respectively, and b_o is the bias of the output gate.

- (5) The output value h_t of LSTM is obtained by calculating the output o_t of the output gate and the state c_t of the current cell, as follows:

$$h_t = o_t * \tanh(c_t). \tag{7}$$

- (6) Bi-LSTM adds an inverse learning cell. It will combine the forward and backward outputs of the LSTM cell as the final output based on the above operations:

$$h_t = [\vec{h}_t \oplus \overleftarrow{h}_t], \tag{8}$$

where \oplus is the summation operation, \vec{h}_t is the output of the forward LSTM cell, \overleftarrow{h}_t is the output of the corresponding backward cell, and h_t is the final output.

The LSTM cell structure is shown in Fig. 3, and the Bi-LSTM model structure is shown in the orange part of the recurrent layer in Fig. 1.

2.3 Attention-based Model

The attention mechanism was proposed in 2014 and has achieved great success in machine translation [25]. In recent years, attention mechanisms have been widely used in various different types of deep learning tasks such as NLP [27] and image recognition [28] and have become one of the most interesting and insightful approaches in the DL field. In a representative study in 2017, the Google Brain team abandoned the classic RNN/CNN structure and proposed a transformer model composed of only the attention mechanism [29].

The attention mechanism is able to redistribute the input weights, thus enabling the selection of information that is more critical to the current task goal from among the many pieces of information. In the model of the present paper, the attention mechanism assigns different weights to the output $[h_1, h_2, \dots, h_t, \dots, h_T]$ of LSTM units, enabling the model to give different levels of attention to different time steps in the sequence. Taking the output time step t as an example, the specific calculation process is as follows.

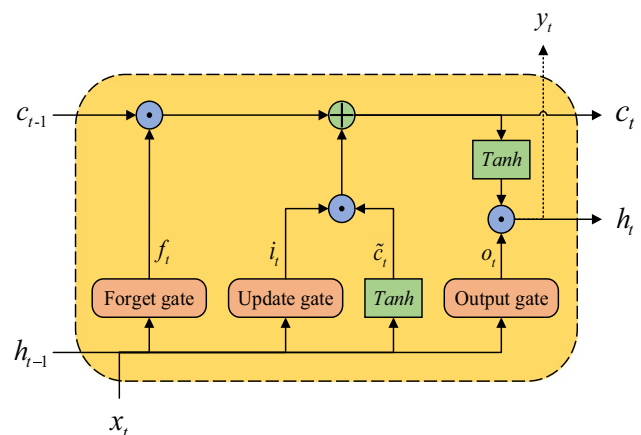


Fig. 3 LSTM cell

- (1) When the Bi-LSTM model produces a hidden output h_t at time step t , it is given a weight W_ω to obtain u_t as follows:

$$u_t = \tanh(W_\omega h_t + b_\omega). \quad (9)$$

- (2) The importance of the current time step u_t is then calculated using the similarity between u_t and u_ω , and is normalized to obtain α_t :

$$\alpha_t = \frac{\exp(u_t^T u_\omega)}{\sum_i \exp(u_i^T u_\omega)}, \quad (10)$$

where u_ω denotes the context vector, which is randomly initialized and jointly learned in the training phase.

- (3) Finally, by weighted summation of these weights α_t , the comment vector S is summarized. This vector summarizes the information of all time steps in the sequence:

$$S = \sum_t \alpha_t h_t. \quad (11)$$

The attention structure is shown in the attention layer in Fig. 1.

3 Experimental Process and Analysis Results

In this section, we first introduce the experimental datasets, and discuss the parameter settings and evaluation metrics. Finally, we compare the proposed ABRM model with several other advanced algorithms. All of the experimental results originate from the proposed algorithm which is performed 50 times independently and the average value is obtained.

3.1 Description of Datasets

In this experiment, the data of moisture content of the coal stack surface as well as the meteorological statistics from on-site weather stations were sampled hourly during June to August 2020, October to December 2020, and January to April 2021 at Huanghua Port coal stockpile in Hebei Province, China (As shown in Figs. 4 and 117°48'46 "E, 38°18'52 "N). The coal moisture content was obtained by the drying method. The drying was done by maintaining a constant temperature of 80 °C in a vacuum for 120 min. Real-time meteorological data were obtained from the nearest weather monitoring stations distributed within the coal yard to the stacks. The meteorological data contained four characteristics: temperature, air humidity, air pressure, and wind speed. We integrated the moisture content data and meteorological data by time consistency, and eliminated the anomalous data, and finally retained 4350 valid data records.



Fig. 4 Experimental scene of sprinkling operation in coal storage yard of Huanghua Port

The valid data were divided into sequences using the sliding window method, and a total of 2004 sequences of length 5 were generated. The sample data contained five characteristics: temperature, air humidity, air pressure, wind speed, and moisture content at the previous moment. The current moment moisture content was used as the label. By comparing other deep learning application cases, the amount of data obtained in this experiment can provide reliable support for the training of prediction models for a single coal species [22]. Finally, the ratio of number of training set samples to that of test set samples was set to 5:1. Partial sample data are shown in Table 1. Table 2 shows the minimum (Min), maximum (Max), mean (Mean), and standard deviation (SD) of each variable of the dataset, including training set and test set. Because the data was collected in distinct seasons, the variation is large in meteorological data, especially for temperature and humidity. The observed values of coal moisture content and the fluctuation values are small, in the range of [0.90%, 27.00%].

3.2 Model Parameters and Experimental Environment

The parameter settings of the ABRM for this experiment are shown in Table 3.

According to the parameter setting of ABRM, we can know the operation process of the model as follows: the input training dataset is a three-dimensional data vector (batchsize, 5, 5), in which the parameters in parentheses represent batch size, time step and feature size, respectively. First, the data enter a 1D convolutional layer with a residual structure to extract features and obtain a 3D output vector (batchsize, 5, 128), in which the number of output channels is 128. Next, the vector enters the

Table 1 Partial sample data

Temperature [°C]	Humidity [%]	Atmospheric pressure [kPa]	Wind velocity [m/s]	Coal moisture content in the previous hour [%]	Coal moisture content of current [%]
23.79	101.51	100.9	2.27	14.80	12.70
23.93	102.33	100.89	1.51	12.70	12.50
24.54	100.23	100.8	1.92	12.50	12.30
24.77	91.01	100.8	1.9	12.30	11.20
25.22	81.72	100.8	2.09	11.20	10.70
25.59	79.07	100.75	1.95	10.70	9.10
25.52	83.58	100.71	1.89	9.10	7.15
25.4	87.47	100.75	1.48	7.15	5.20
25.25	87.66	100.7	1.12	5.20	4.42
25.18	92.63	100.7	0.94	4.42	3.76
24.92	94.3	100.71	0.93	3.76	3.19
24.34	100.46	100.8	1.61	2.71	2.31

Table 2 Statistical analysis results of each variable

Variable	Unit	Min	Max	Mean	SD
Temperature	°C	-15.75	34.91	12.18	12.91
Humidity	%	12.17	119.28	93.97	24.62
Atmospheric pressure	kPa	99.80	104.10	102.14	1.31
Wind velocity	m/s	0.41	7.56	2.48	1.13
Coal moisture content	%	0.90	27.00	6.84	4.57

Table 3 Parameter setting of ABRM

Parameter	Value
Number of convolution layer filters	15
Convolution layer kernel size	3
Convolution layer activation function	ReLU
Convolution layer padding	1
Pooling layer pool size	1
Pooling layer padding	None
Pooling layer activation function	ReLU
Number of hidden units in LSTM layer	16
Number of LSTM layers	2
LSTM layer activation function	tanh
Dropout	0.3
Time step	5
Batch size	2
Learning rate	0.01(*0.1 every 2500 epochs)
Optimizer	Adam
Loss function	MSE Loss
Number of epochs	10,000

max-pooling layer, which does not change the vector size, and a 3D output vector (batchsize, 5, 128) is obtained. Then, the output vector enters the Bi-LSTM layer and the attention layer for training, and the output vector (batchsize, 32) is obtained, where 32 is the hidden size of the Bi-LSTM. Finally, the output vector of Bi-LSTM is interpreted by the fully connected layer to obtain the output value. Dropout was added to the LSTM layer to prevent overfitting [30]. Training was completed for 10,000 epochs, and the learning rate decayed to 1/10 of the previous rate for every 2,500 epochs that passed.

The training environment of the model was a graphics workstation, which was configured with CPU: Intel (R) Xeon (R) CPU 4210R*2@3.20 GHz, GPU: Nvidia Geforce RTX3080 and RAM: 64GB. The Anaconda platform was adopted as the basic platform for deep learning training, pytorch (version 1.6.0) was the deep learning framework, CUDA technology was used to realize parallel computing. The Python version was 3.7.

3.3 Evaluation Metrics

In order to validate and evaluate the performance of the model, five evaluation metrics are used in this paper: the mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), correlation coefficient (R-squared, R^2), and mean absolute percentage error (MAPE). These evaluation metrics have been widely used in prediction tasks. The formulae of these metrics are expressed as follows:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_{Pred} - y)^2, > \quad (12)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (|y_{Pred}| - |y|)^2}, \quad (13)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_{Pred}| - |y|, \quad (14)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_{Pred} - \bar{y})^2}{\sum_{i=1}^N (\bar{y} - y)^2}, \quad (15)$$

$$MAPE = \frac{100\%}{N} \sum_{i=1}^N \left| \frac{y_{Pred} - y}{y} \right|. \quad (16)$$

Here, N is the sample size of the test set. y_{Pred} is the predicted value, and y is the true value. \bar{y} is the average of the true values of all samples in the test set.

Among the above indicators, the closer the value of R^2 is to 1, the better discriminating the model is. The closer the values of MAE, RMSE, MAE, and MAPE are to 0, the higher the prediction accuracy.

3.4 Comparing Performance of Different Models

In this section, six models, SVM, RNN, LSTM, CNN, A-LSTM (attention-based LSTM), and simple CNN-LSTM (without residual unit and attention mechanism), were selected for comparison and analysis with the proposed ABRM. The selected models include traditional ML models, common recurrent/convolutional neural network models, and combined models of neural networks. These have been applied to various sequence data prediction tasks and have achieved satisfactory results. We trained each model with the same training dataset and parameters. After the training, the models were tested using the same test dataset. The prediction accuracy of each model for coal moisture content was evaluated by comparing the predicted values with the observed values for the next 1 h.

Table 4 shows the results of prediction accuracy comparison between the ABRM and other models. In this experiment, the R^2 and MAE scores are particularly important to indicate the performance of the model. R^2 is used to detect the degree of fit of each prediction model to the observed moisture content of the samples; MAE reflects the deviation of the observed moisture content of all individual samples from the arithmetic mean in the experiment. MAE intuitively reflects the magnitude of the actual prediction error. The results in Table 4 show that the R^2 and MAE scores of the ABRM model proposed in this paper are 0.9971 and 0.0812 (boldfaced in Table 4), respectively, which are significantly better

Table 4 Results of prediction accuracy comparison between the ABRM and other models

Models	MSE	RMSE	MAE	R^2	MAPE
SVM	7.2537	2.6933	1.2535	0.6024	34.0171
RNN	0.9048	0.9512	0.6103	0.9504	17.6118
LSTM	0.6806	0.825	0.4783	0.9627	14.1196
CNN	0.4200	0.6481	0.3447	0.977	10.4491
A-LSTM	0.3404	0.5834	0.321	0.9813	9.6539
CNN-LSTM	0.1375	0.3708	0.1643	0.9925	3.6014
ABRM (proposed model)	0.0528	0.2299	0.0812	0.9971	1.6117

than the performance metrics of other models. In addition, the performance of MSE, RMSE and MAPE metrics of ABRM also outperformed other models, reaching 0.0528, 0.2299 and 1.6117, respectively.

Figure 5 demonstrates the predicted and observed values of coal moisture content for model. For better comparison, we label the R^2 scores of each model in red text. It is apparent from this figure that the fitting ability of the model improves to different degrees as the complexity of the model increases. The R^2 score of both the simple CNN-LSTM model and ABRM in this paper exceeds 0.99. The ABRM proposed in this paper performs relatively better, with an R^2 score of 0.9971 and achieving a satisfactory fitting effect (Fig. 5(g)). It can be seen that the feature extraction capability of the CNN and the sequence prediction capability of LSTM are fully utilized in the coal moisture content prediction task. It is worth noting that in Fig. 5(g), there are 2 obvious anomalies, which appear at 4% and 14%, respectively, also in (a) to (f). By querying the dataset, it was found that these 2 anomalies were caused by sudden changes in meteorological parameters due to the occurrence of short-term strong convective weather at the experimental site. This indicates that the model proposed in this paper is less capable of handling such changes. However, the model is able to capture the smoother meteorological changes very well. In other words, in most cases, the predicted values of moisture content output by the model can be trusted and guide industrial production.

In addition, we conducted a comparative analysis of the absolute errors of the predicted values of each model and plotted the error distribution in Fig. 6. The mean absolute error (MAE) scores of each model are labeled in each sub-figure, and the points inside the horizontal black dashed line are the prediction points with absolute errors in the range of $[-2, 2]$. Although most of the models are basically able to control the absolute error within the range of $[-2, 2]$.

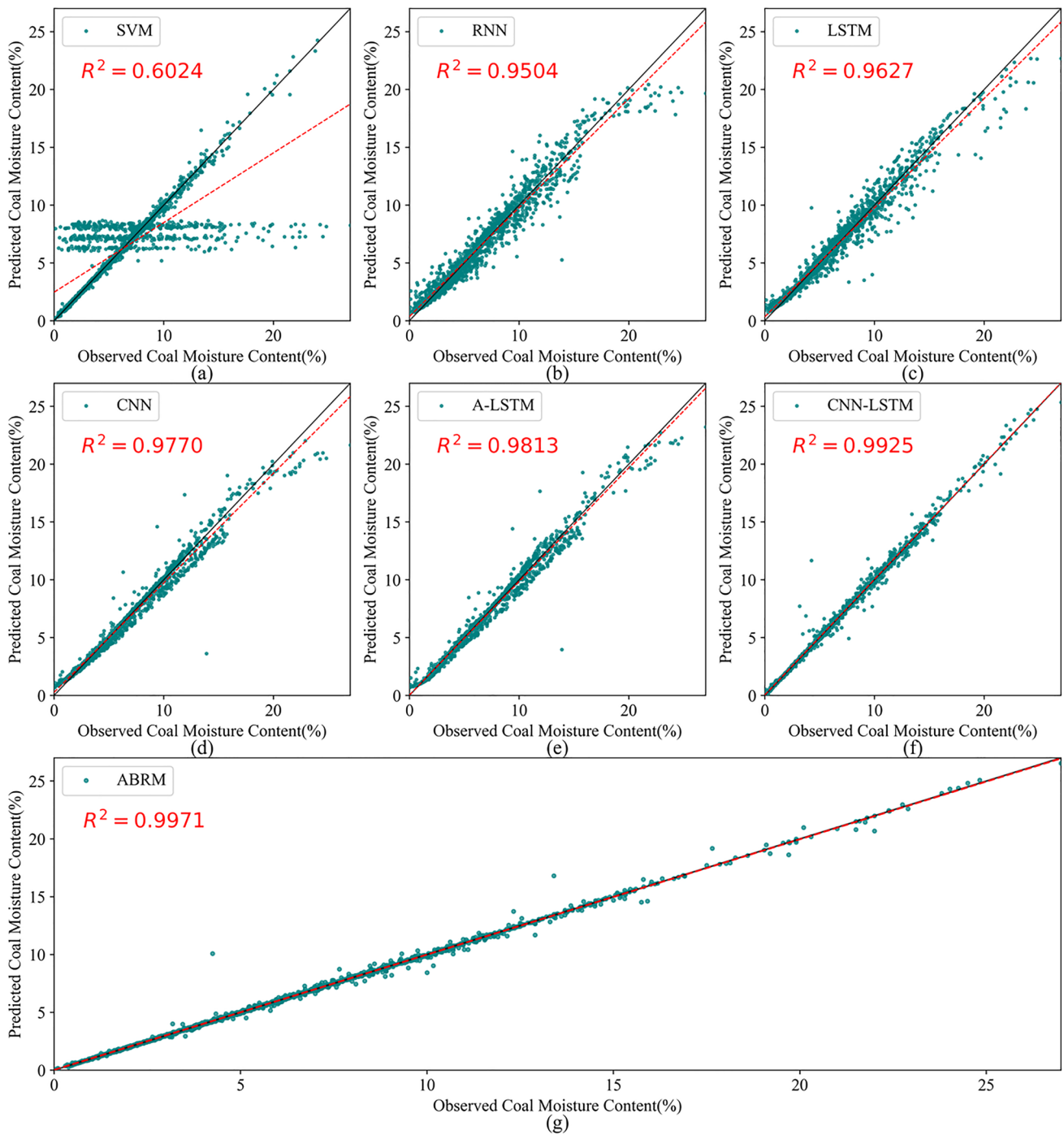


Fig. 5 Comparison of the predicted and observed values of the SVM, RNN, LSTM, CNN, A-LSTM, CNN-LSTM, and ABRM models

Table 4 summarizes the comparison results of prediction accuracy between the ABRM and other models. The range of moisture content is only [0.9,27] and the SD is only 4.57, that is, the absolute error range of [-2,2] does not meet the requirements of long-term guiding production.

Comparatively, ABRM can control most of the errors within the range of [-0.5,0.5], and the prediction performance is more superior.

Figure 7 further illustrates the predicted values of the ABRM on the test set compared to actual values. Different

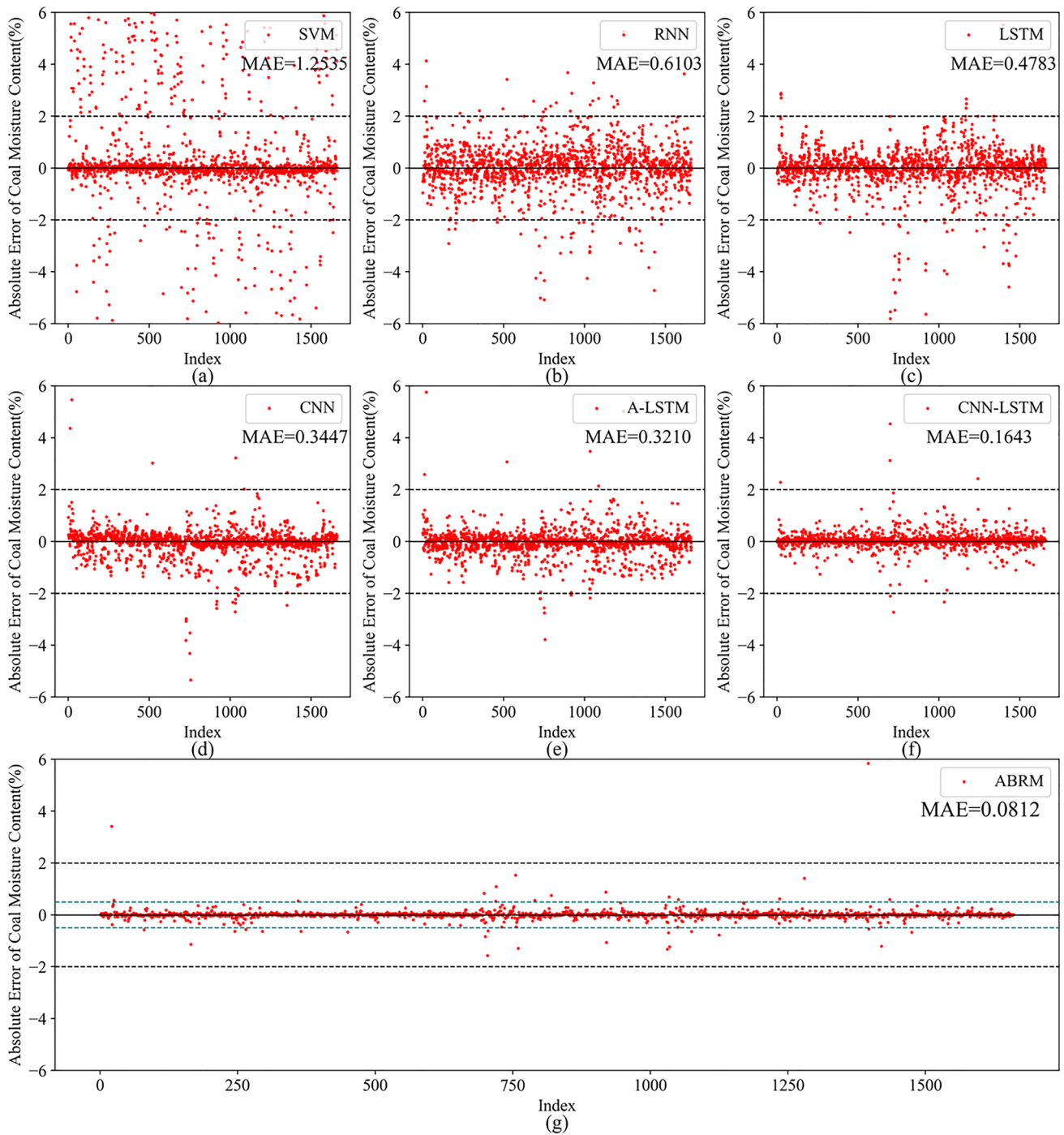


Fig. 6 Absolute error analysis of the SVM, RNN, LSTM, CNN, A-LSTM, CNN-LSTM, and ABRM models

sequences of testing data and their corresponding prediction curves are put in the same plot. In Fig. 7, the majority of ABRM prediction curves marked in green highly overlap with real observation marked in red, showing good generalization capability of the model.

In summary, we find that the ABRM constructed in this paper has relatively optimal prediction performance. The model structure ensures that the neural network correctly understands the sequence data and extracts key features, while structurally suppressing the overfitting issue to ensure the network’s robustness against perturbations.

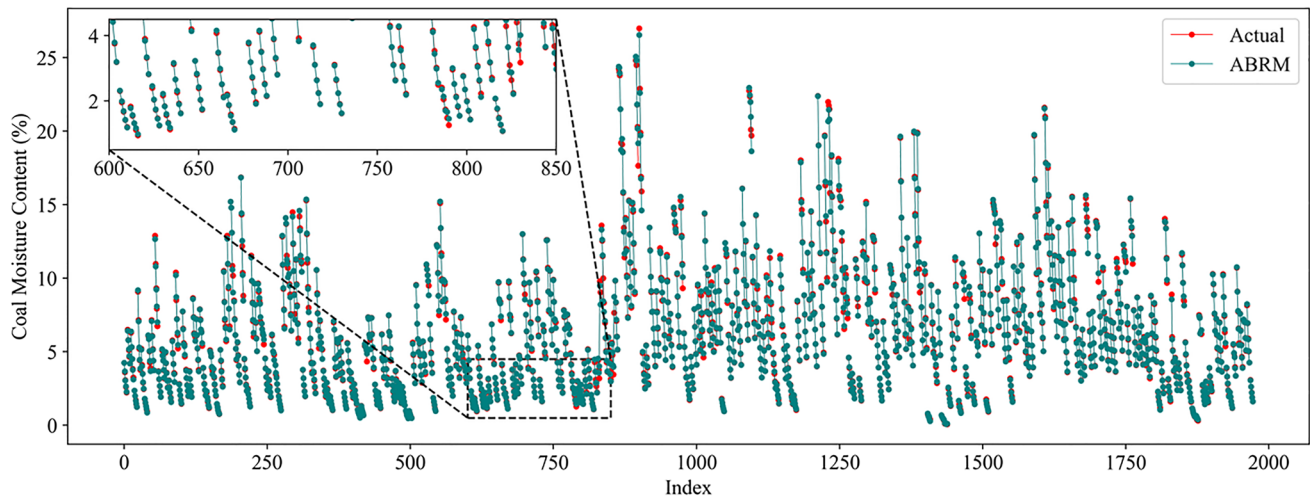


Fig. 7 Comparison between the predicted values and the actual values of the ABRM

4 Conclusions

In this paper, an ABRM model structure for predicting the moisture content of coal in the surface layer of coal stacks is proposed. The model combines ResNet based on one-dimensional convolution and Bi-LSTM based on attention mechanism. It fuses the ResNet feature extraction ability and the time series feature memory ability of Bi-LSTM. The sample features extracted by ResNet are used as the input sequence of LSTM for learning to obtain the final prediction results. The experimental results show that the ABRM model proposed in this paper outperforms common regression prediction models in terms of prediction accuracy and convergence rate. The proposed model scores better than other models in all evaluation metrics. Excellent performance is achieved in predicting the moisture content of coal in the surface layer of coal stacks. The high precision coal moisture content prediction value provides an important reference for the coal storage and transportation management system of coal storage base, which eases the management work for dust-suppression sprinkling and spontaneous combustion prevention.

Future work can incorporate light intensity, coal type, etc. into the data feature set to improve the prediction stability of the model. It is meaningful to develop a multi-task learning algorithm to further predict the change in coal moisture content for several hours by predicting the change in meteorological parameters for several hours in the future. In addition, this method can be applied to most natural minerals, such as sand and dirt whose moisture content is influenced by the environment. It is hoped that interested scholars will verify it in future experiments.

Author Contributions Professor Fan Zhang made primary contributions to the conceptualization or design of the work, review and editing. Master Hao Li completed original draft, data curation and algorithm development, and Dr. Zhichao Xu completed the validation and visualization. Professor Wei Chen made optimization of the software architecture, review.

Funding This research was supported by the National Natural Science Foundation of China (grant number 51874300), Fundamental Research Funds for the Central Universities (grant number 2021YJSJD02), the Graduate Program of ideological and political construction, China University of Mining and Technology of Beijing (grant number YKC-SZ202100404S), and the Open Research Fund of Key Laboratory of Intelligent Mining and Robotics, China University of Mining and Technology of Beijing (grant number U03462).

Data Availability The data used to support the finding of this study are available from the corresponding author upon request.

Declarations

Ethics Approval Approval was obtained from the ethics committee of the China University of Mining and Technology of Beijing.

Consent to Participate Informed consent was obtained from all individual participants included in the study.

Consent for Publication The authors has consented to the submission of the research manuscript to the journal.

Conflict of Interest The authors declare no conflict of interest.

References

1. Fu, C.H.E.N., Haochen, Y.U., Zhengfu, B.I.A.N., et al.: How to handles the crisis of coal industry in China under the vision of carbon neutrality[J]. *J. China Coal Soc.* **46**(06), 1808–1820 (2021)

2. Wang, G., Xu, Y., Ren, H.: Intelligent and ecological coal mining as well as clean utilization technology in China: Review and prospects. *Int. J. Min. Sci. Technol.* **29**, 161–169 (2019). <https://doi.org/10.1016/j.ijmst.2018.06.005>
3. Cutmore, N., Abernethy, D., Evans, T.: Microwave Technique for the On-Line Determination of Moisture in Coal. *J. Microw. Power Electromagn. Energy.* **24**, 79–90 (1989). <https://doi.org/10.1080/08327823.1989.11688079>
4. Zeng, D., Hu, Y., Liu, J., Zhao, Z., Gao, S.: Soft sensing of coal moisture. *Measurement.* **60**, 231–239 (2015). <https://doi.org/10.1016/j.measurement.2014.09.080>
5. Yuman, W.A.N.G.: Mechanism and methods of Coal Moisture Measurement Based on Microwave Transmission Method[D]. North China Electric Power University (2016)
6. Mao, Y., Xia, W., Xie, G., Peng, Y.: Rapid detection of the total moisture content of coal fine by low-field nuclear magnetic resonance. *Measurement.* **155**, 107564 (2020). <https://doi.org/10.1016/j.measurement.2020.107564>
7. Tai, Y., Qian, K., Huang, X., Zhang, J., Jan, M.A., Yu, Z.: Intelligent Intraoperative Haptic-AR Navigation for COVID-19 Lung Biopsy Using Deep Hybrid Model. *IEEE Trans. Ind. Inf.* **17**, 6519–6527 (2021). <https://doi.org/10.1109/TII.2021.3052788>
8. Du, S., Li, T., Yang, Y., Horng, S.-J.: Deep air quality forecasting using hybrid deep learning framework. *IEEE Trans. Knowl. Data Eng.* **33**, 2412–2424 (2021). <https://doi.org/10.1109/TKDE.2019.2954510>
9. Han, X.-F., Laga, H., Bennamoun, M.: Image-based 3D object reconstruction: state-of-the-art and trends in the deep learning era. *IEEE Trans. Pattern Anal. Mach. Intell.* **43**, 1578–1604 (2021). <https://doi.org/10.1109/TPAMI.2019.2954885>
10. Wang, L., Wu, T., Fu, H., Xiao, L., Wang, Z., Dai, B.: Multiple contextual cues integrated trajectory prediction for autonomous driving. *IEEE Robot. Autom. Lett.* **6**, 6844–6851 (2021). <https://doi.org/10.1109/LRA.2021.3094564>
11. Ma, D., Song, X., Li, P.: Daily traffic flow forecasting through a contextual convolutional recurrent neural network modeling inter- and intra-day traffic patterns. *IEEE Trans. Intell. Transport. Syst.* **22**, 2627–2636 (2021). <https://doi.org/10.1109/TITS.2020.2973279>
12. Jia, Y., Jin, S., Savi, P., Yan, Q., Li, W.: Modeling and theoretical analysis of GNSS-R soil moisture retrieval based on the random forest and support vector machine learning approach. *Remote Sens.* **2020**(12), 3679 (2020). <https://doi.org/10.3390/rs12223679>
13. Sanuade, O.A., Hassan, A.M., Akanji, A.O., Oloajo, A.A., Olanjoye, M.A., Abdulraheem, A.: New empirical equation to estimate the soil moisture content based on thermal properties using machine learning techniques. *Arab. J. Geosci.* **13**, 377 (2020). <https://doi.org/10.1007/s12517-020-05375-x>
14. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature.* **521**, 436–444 (2015). <https://doi.org/10.1038/nature14539>
15. Chai, H., Chen, X., Cai, Y., Zhao, J.: Artificial neural network modeling for predicting wood moisture content in high frequency vacuum drying process. *Forests* **2018**(10), 16 (2018). <https://doi.org/10.3390/f10010016>
16. Sun, Q., Zhang, M., Yang, P.: Combination of LF-NMR and BP-ANN to monitor water states of typical fruits and vegetables during microwave vacuum drying. *LWT.* **116**, 108548 (2019). <https://doi.org/10.1016/j.lwt.2019.108548>
17. Chatterjee, S., Dey, N., Sen, S.: Soil moisture quantity prediction using optimized neural supported model for sustainable agricultural applications. *Sustain. Comput. Infor. Syst.* **28**, 100279 (2020). <https://doi.org/10.1016/j.suscom.2018.09.002>
18. Fang, K., Pan, M., Shen, C.: The value of SMAP for long-term soil moisture estimation with the help of deep learning. *IEEE Trans. Geosci. Remote Sensing.* **57**, 2221–2233 (2019). <https://doi.org/10.1109/TGRS.2018.2872131>
19. ElSaadani, M., Habib, E., Abdelhameed, A.M., Bayoumi, M.: Assessment of a spatiotemporal deep learning approach for soil moisture prediction and filling the gaps in between soil moisture observations. *Front. Artif. Intell.* **4**, 636234 (2021). <https://doi.org/10.3389/frai.2021.636234>
20. Zhang, B., Zou, G., Qin, D., Lu, Y., Jin, Y., Wang, H.: A novel Encoder-Decoder model based on read-first LSTM for air pollutant prediction. *Sci. Total Environ.* **765**, 144507 (2021). <https://doi.org/10.1016/j.scitotenv.2020.144507>
21. Teng, T., Gao, F., Ju, Y., Xue, Y.: How moisture loss affects coal porosity and permeability during gas recovery in wet reservoirs? *Int. J. Min. Sci. Technol.* **27**, 899–906 (2017). <https://doi.org/10.1016/j.ijmst.2017.06.016>
22. Yu, J., Zhang, X., Xu, L., Dong, J., Zhangzhong, L.: A hybrid CNN-GRU model for predicting soil moisture in maize root zone. *Agric. Water Manag.* **245**, 106649 (2021). <https://doi.org/10.1016/j.agwat.2020.106649>
23. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Recognition, P. (CVPR), IEEE, Las Vegas, NV, USA, pp. 770–778 (2016). <https://doi.org/10.1109/CVPR.2016.90>
24. Xie, W., Wang, J., Xing, C., Guo, S., Guo, M., Zhu, L.: Variational autoencoder bidirectional long and short-term memory neural network soft-sensor model based on batch training strategy. *IEEE Trans. Ind. Inf.* 1–1 (2020). <https://doi.org/10.1109/TII.2020.3025204>
25. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. *ArXiv:1409.0473 [Cs, Stat]*. <http://arxiv.org/abs/1409.0473> (2016). Accessed 17 Apr 2021
26. Jelodar, H., Wang, Y., Orji, R., Huang, S.: Deep sentiment classification and topic discovery on Novel Coronavirus or COVID-19 online discussions: NLP Using LSTM recurrent neural network approach. *IEEE J. Biomed. Health Inform.* **24**, 2733–2742 (2020). <https://doi.org/10.1109/JBHI.2020.3001216>
27. Basiri, M.E., Nemati, S., Abdar, M., Cambria, E., Acharya, U.R.: ABCDM: An attention-based bidirectional CNN-RNN Deep Model for sentiment analysis. *Future Gener. Comput. Syst.* **115**, 279–294 (2021). <https://doi.org/10.1016/j.future.2020.08.005>
28. Li, X., Jiang, Y., Li, M., Yin, S.: Lightweight attention convolutional neural network for retinal vessel image segmentation. *IEEE Trans. Ind. Inf.* **17**, 1958–1967 (2021). <https://doi.org/10.1109/TII.2020.2993842>
29. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. *Adv. Neural Inf. Process. Syst.* **30**(NIPS 2017), 30 (2017)
30. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Fan Zhang received the Ph.D. degree in communications and information systems from China University of Mining and Technology, Beijing, China, in 2012. He joined the School of Mechanical Electronic & Information Engineering, China University of Mining and Technology, Beijing, where he is a professor. He has published more than 50 research papers and obtained more than 20 authorized invention patents. His research interests include image reconstruction, computer vision, deep learning and digital twins.

Hao Li is a master's student of China University of Mining and Technology, Beijing. His research interests concern intelligent monitoring and time series predictive analytics.

ZhiChao Xu is a faculty member of China University of Mining and Technology, Beijing. He received his Master's degree in Communication and Information Systems in 2020. His research interests deal with digital image processing, super-resolution imaging and cyber-security problems.

Wei Chen received the B.Eng. degree in medical imaging and the M.S. degree in paleontology and stratigraphy from China University of Mining and Technology, Xuzhou, China, in 2001 and 2005, respectively, and the Ph.D. degree in communications and information systems from China University of Mining and Technology, Beijing, China, in 2008.

In 2008, he joined the School of Computer Science and Technology, China University of Mining and Technology, where he is currently a professor. He is a member of IEEE and EAI. He has published more than 80 research papers. He is an associate editor of international journal Array. His research interests include machine learning, image processing, and computer networks.