IEEE *Access*

Multidisciplinary | Rapid Review | Open Access Journal

A Gated Recurrent Unit Network Model for Predicting Open Channel Flow in Coal Mines Based on Attention Mechanisms

ZHANLI LI[1], TIANYU GAO [ID][2], CHENG GUO[3], AND HONG-AN LI [ID][1]

[1]College of Computer Science and Technology, Xi'an University of Science and Technology, Xi'an 710054, China
[2]College of Safety Science and Engineering, Xi'an University of Science and Technology, Xi'an 710054, China
[3]Xi'an Institute of Applied Optics, Xi'an 710065, China

Corresponding authors: Tianyu Gao (gvastsky@gmail.com) and Hong-An Li (honganli@xust.edu.cn)

ABSTRACT The prediction of water inflow during coal mining is an important issue. There are many factors that can affect the water inflow in mines. The intercoupling of these factors makes it difficult for current water inflow forecasting methods to meet the needs of real-time forecasting. Open channels are the main devices used for mine drainage, and their flow rate reflects the water inrush of a mine to some extent. This paper uses a hybrid neural network model combining attention mechanisms and a gated recurrent unit network to make real-time predictions of open channel flow. First, attention mechanisms are used to learn the interdependence between multisource hydrosensor data, and then, a gated recurrent unit network is employed to capture the dependencies on different time scales to improve the prediction accuracy of the neural network model. Finally, we design a series of comparative experiments to verify and analyse the performance of the hybrid neural network model. The experimental verification shows that the proposed model can learn the dependency relationships among multisource sensors, and the modelling of these dependencies can greatly improve the prediction accuracy of real-time flow in open channels.

INDEX TERMS Mine water damage, flow prediction, neural network, attention mechanism.

I. INTRODUCTION

As the cornerstone of China's economic development, coal will continue to be the country's primary source of energy for the foreseeable future. The issue of safe coal production has always been a focus of attention because of the complex hydrological and geological conditions in China's mining areas. As China's demand for coal continues to grow, the scale of coal mining gradually widens, the depth of mining gradually deepens, and the probability of flooding and other water damage accidents increases [1], [2]. Mine water inflow is the primary reason for rationally setting up mine drainage systems and formulating mine water prevention and control measures. The dynamic prediction of mine water inflow has become an extensively studied problem.

The existing methods of predicting mine water inflow can be roughly divided into two types: analytical methods and numerical simulations. Analytical methods use

The associate editor coordinating the review of this manuscript and approving it for publication was Jenny Mahoney.

hydrogeological parameters to establish a conceptual model of hydrogeology in a mining area and calculate and predict an analytical solution to water inflows during mining operations [3], [4], [5]. Among these solutions, the big well method has become a commonly used prediction approach since it allows the use of specific assumptions and simple geological and boundary conditions [6]. Li *et al.* proposed a generalized large well method for dynamically predicting and evaluating groundwater levels during mining operations, taking the Yimin open-pit mine as an example to verify the effectiveness of the analysis method [4]. Wu *et al.* used the large well method and a numerical simulation to calculate the comprehensive mine inflow volume in three goafs and compared and analysed the water levels in different periods to provide a reference for decision makers to improve the level of safety in mine production [7].

Numerical simulation is currently the most widely used method; it also employs hydrogeological parameters to establish a groundwater flow model in the mining area and uses the model to simulate the change of groundwater in the

mining area during the mining process to numerically predict mine water inflows [8]–[11]. Zhang *et al.* established a three-dimensional hydrogeological conceptual model of the Baku phosphate deposit and predicted the groundwater flow and groundwater flow field evolution at different development stages of the Baku phosphate deposit [12]. Xue *et al.* simplified the groundwater level in the central and western regions of China to a single-layer groundwater system and optimized the geological parameters of each unit area to obtain a reasonable steady-state groundwater distribution model [13]. Zhang *et al.* characterized the flow of groundwater through physical model tests, seepage tests and numerical simulations [14]. Chen *et al.* used a basic Darcy model and Biot's classic theory to establish a full-porosity elastic model to estimate the seepage flow during longwall mining in coal mines [15]. Mu *et al.* established a conceptual and mathematical model of a karst flow system and carried out a numerical simulation of the karst flow system using a finite difference method [16].

Both analytical methods and numerical simulations can achieve satisfactory prediction accuracy when hydrogeological parameters are detailed and the established model is sufficiently accurate. However, analytical methods are based on certain assumptions and specific boundary conditions, which limits their applicability to different mining situations. For example, most analytical solutions can neither directly explain the gushing water from the floor of a mine [17] nor simulate the head and saturated and unsaturated flow conditions of a confined aquifer [18]. Numerical simulation methods require many hydrogeological parameters such as aquifer permeability coefficients, aquifer transmittances, and rainfall data. Obtaining and determining these parameters are very difficult, expensive, and time consuming. In addition, these parameters will increase the cumulative error of the model, thereby increasing the uncertainty of the final result.

In recent years, there has been a great success using machine learning algorithms in areas such as target recognition [19], natural language processing [20], and sequence prediction [21]–[23]. In particular, neural networks such as recurrent neural networks (RNNs) [24], [25], deep belief networks [26], and radial basis function networks [27] have demonstrated excellent performance in sequence prediction problems. Therefore, neural networks are widely used in the field of mine water damage prediction, including the prediction model of flow height in a fractured zone [28] and the framework of probability assessment of mine water inrush accidents [29]. Bahrami *et al.* [30] designed two hybrid methods coupling artificial neural networks with genetic algorithms and simulated annealing methods to predict the head of an open-pit mine. The results show that these two hybrid methods have the ability to compete with numerical models to some extent. Ardejani *et al.* [31] used a neural network model to predict the rebound process of groundwater after a dewatering stop at a restored open-pit coal yard in East Midlands, UK. The predicted value was very close to the

monitoring result, indicating that the model prediction result was satisfactory.

Although the models proposed by these studies have achieved good prediction accuracy, their prediction time steps are measured in days, months and even years. Forecasting with such a long time step can only be used to analyse the general trends of mine water inrushes, which is less helpful for real-time warnings of mine water damage. Open channels are the main devices used for coal mine water discharges. The effective prediction of open channel flows can not only help predict the amount of water in a mine but also acquire the precursory information of water damage problems, such as water permeability, in a timely manner by comparing and analysing historical data.

Attention mechanisms are first proposed in the field of visual images. In the 2014 Google DeepMind team's research [32], attention mechanisms are used in a recurrent neural network model to classify images, making them a research topic of interest. Subsequently, in [33], Bahdanau *et al.* used an attention-like mechanism to perform translation and alignment simultaneously on machine translation tasks. Their work was the first to apply the attention mechanism to the field of natural language processing. Since then, attention mechanisms have been widely used in various natural language processing tasks based on neural network models such as recurrent neural networks or convolutional neural networks. In 2017, a study published by the Google Machine Translation team [34] made extensive use of self-attention to learn text representation. Self-attention has also become a recent research area of interest and has been applied to various natural language processing tasks. Attention mechanisms have shown success in the field of visual images and natural language processing since they can model dependencies without considering their distance in the input or output sequence.

Gated recurrent unit (GRU) networks are proposed by Cho *et al.* [36] as a variant of long short-term memory (LSTM) networks. In addition to solving the long-term dependence problem in traditional RNNs [37], GRU networks also simplify the calculation of the gating function in LSTM networks and improve the calculation efficiency. This paper proposes a hybrid neural network model of attention mechanisms and GRU networks to predict the flow of open channels and increase the performance compared to conventional methods. The proposed methodology can provide an important basis for the design of coal mine drainage capacity and prevention measures. The main contributions are outlined as follows.

1) The proposed method addresses the problem of an excessively large step size in current mine water inrush prediction research. Reducing the prediction step size can improve the real-time prediction of mine water inrushes.

2) The attention mechanism combined with the GRU network is used to strengthen the GRU network's ability to mine dependencies. This enables the network to more

comprehensively obtain the association relationship of the input data and improve the network prediction performance.

3) The elevation, water temperature, burial depth, open channel flow, water temperature, and dibhole liquid level data are chosen as the inputs to obtain more comprehensive information. The results validate the performance improvement.

This paper describes the algorithm of the attention mechanisms, the GRU and the overall architecture of the hybrid neural network model in Section II. In Section III, the experimental data are explained, and the optimization method of the model is introduced. Section IV presents the experimental design, results and analysis, and Section V offers the conclusions.

II. ARCHITECTURES AND ALGORITHMS
A. ATTENTION MECHANISMS
The input of an attention algorithm consists of sequence queries, keys and values. An attention algorithm can be described as a mapping from a query to a series of keys and values. The two most commonly used attention functions are dot-product (multiplicative) attention and additive attention [33]. To improve the operation efficiency, we pack the sequences *queries*, *keys*, and *values* into the matrices $Q(query_1 \ldots, query_T)$, $K(key_1 \ldots, key_T)$, and $V(value_1 \ldots, value_T)$, where $query_i$ is the queries vector stacked into the matrix $Q \in \mathbf{R}^{T \times L}$, key_i is the keys vector stacked into the matrix $K \in \mathbf{R}^{T \times L}$, $value_i$ is the values vector stacked into the matrix $V \in \mathbf{R}^{T \times L}$, v_{ij} is an element of matrix V, $\mathbf{R}$ is the set of real numbers, T is the number of packed vectors in the matrices, and L is the dimension of the input vectors, $i = 1, 2 \ldots, T$, $j = 1, 2 \ldots, L$. The computation of the elements in the attention matrix can be expressed abstractly as [35]:

$$a_{ij} = \sum_{l=1}^{L_x} \alpha_{il} v_{jl} \tag{1}$$

$$\alpha_{ij} = \text{soft max}(s_{ij}) = \frac{\exp(s_{ij})}{\sum_{l=1}^{L_x} \exp(s_{il})} \tag{2}$$

$$s_{ij} = Similarity(query_i, key_j) \tag{3}$$

where $Similarity(query_i, key_i)$ is used to calculate the correlation or similarity between *queries* and *keys*. Different attention functions use different calculation methods. In our work, considering different prediction targets, the target parameters and other input parameters have different dependencies. We use a feed-forward neural network that is jointly trained with other components of the predictive network to compute similarity, i.e.,

$$s_{ij} = Similarity(query_i, key_j) = \sum_{l=1}^{L_x} w_{il} q_l k_l \tag{4}$$

where q_l and k_l are elements in vector $query_i$ and key_j, respectively, and w_{il} is the weight to be trained in the feed-forward neural network. Then, the SoftMax function, also

called the normalized exponential function, is used to normalize the calculated correlation. In this way, on the one hand, the correlation can be organized into a probability distribution with a sum of 1. On the other hand, the weights of important elements can be made more prominent through the internal mechanism of the SoftMax function. Finally, the weighted summation can obtain the value of attention.

Self-attention is also called intra-attention and is calculated in the same way as attention. Different from the attention mechanism, the queries, keys, and values are the same in the self-attention input sequences, that is, $query_i = key_i = value_i$. The purpose of the same input sequences is to learn the dependencies between the sequences and link the different positions of a single sequence to calculate the sequence representation.

In practice, the input X exists in the form of a matrix. Suppose that x_{ij} is the element in matrix $X \in \mathbf{R}^{m \times n}$, n represents the time step of interception, and m is the number of sensor parameters. The attention mechanism used in this work can be expressed as:

$$a_{ij} = \sum_{l=1}^{m} \alpha_{il} x_{jl} \tag{5}$$

$$\alpha_{ij} = \text{soft max}(s_{ij}) = \frac{\exp(s_{ij})}{\sum_{k=1}^{m} \exp(s_{ik})} \tag{6}$$

$$s_{ij} = \sum_{l=1}^{m} w_{il} x_{il} x_{jl} \tag{7}$$

For example, we take the monitoring values of 5 sensors at t times to calculate the attention matrix and analyse the correlation between the sensors. The input X can be represented as

$$X = \begin{bmatrix} sensor1_1 & sensor1_2 & \cdots & sensor1_t \\ sensor2_1 & sensor2_2 & \cdots & sensor2_t \\ \vdots & \vdots & \ddots & \vdots \\ sensor5_1 & sensor5_2 & \cdots & sensor5_t \end{bmatrix} \tag{8}$$

where $sensor1i$, $sensor2i \ldots$, $sensor5i$ are the monitoring values of 5 sensors at times i, where $i = 1, 2 \ldots, t$. Then, using Eq. (6) and Eq. (7), the similarity matrix shown in Fig. 1 can be calculated.

As shown, the similarity matrix is composed of elements with a value of 0-1. The greater the value of the element is, the higher the correlation between the corresponding sensors. Therefore, when using the elements in the similarity matrix as the weight α_{ij} and calculating the attention matrix through (5) with the original input X, the value with a high degree of correlation can be enlarged according to the weight to achieve the purpose of focusing attention.

B. GATED RECURRENT UNIT
The LSTM introduces three gate functions on the basis of the traditional recurrent neural network structure—the input gate, forget gate and output gate—to control the input value, memory value and output value, respectively, to save long-term memory and solve the long-reliance problem in the traditional

	sensor1	sensor2	sensor3	sensor4	sensor5
sensor1	0.96	0.25	0.70	0.69	0.11
sensor2	0.27	0.98	0.08	0.19	0.02
sensor3	0.71	0.05	0.98	0.71	0.25
sensor4	0.69	0.17	0.69	0.95	0.79
sensor5	0.13	0.03	0.24	0.78	0.99

FIGURE 1. Illustration of the similarity matrix.

recurrent neural network. The GRU combines the forget gate and input gate in the LSTM into a single update gate z:

$$z = \sigma(W_z x + U_z h_{t-1}) \tag{9}$$

where x is the input, ht-1 is the previous hidden state, W_z and U_z are the learned weight matrices, and σ is the logistic sigmoid function. Similarly, the GRU replaces the output gate in the LSTM with a reset gate r:

$$r = \sigma(W_r x + U_r h_{t-1}) \tag{10}$$

where W_r and U_r are the learned weight matrices. In the end, the output of the hidden unit is:

$$\tilde{h}_t = \tanh(Wx + U(r * h_{t-1})) \tag{11}$$

$$h_t = z h_{t-1} + (1 - z)\tilde{h}_t \tag{12}$$

where $\tilde{h}$ is a new hidden state, W and U are the learned weight matrices, and h_t is the hidden state at the current moment. An illustration of the hidden activation function is shown in Fig. 2.
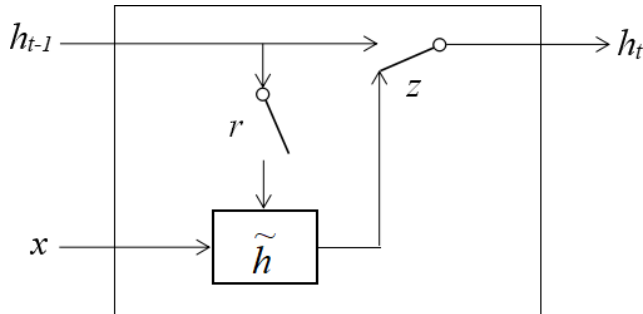

FIGURE 2. Illustration of the hidden activation function.

The update gate z determines whether the hidden state is to be updated with a new hidden state $\tilde{h}$. The reset gate r decides whether the previous hidden state is ignored.

C. MODEL ARCHITECTURE

Because of its network architecture, a recurrent neural network exhibits excellent performance in processing sequence data [38]. The neural network model proposed in this paper adds attention mechanisms to a GRU network architecture. The overall architecture is shown in Fig. 3.

First, the input of the neural network uses the ability of the attention mechanism to learn dependencies of the coal mine hydrological monitoring parameters. Then, a GRU network is
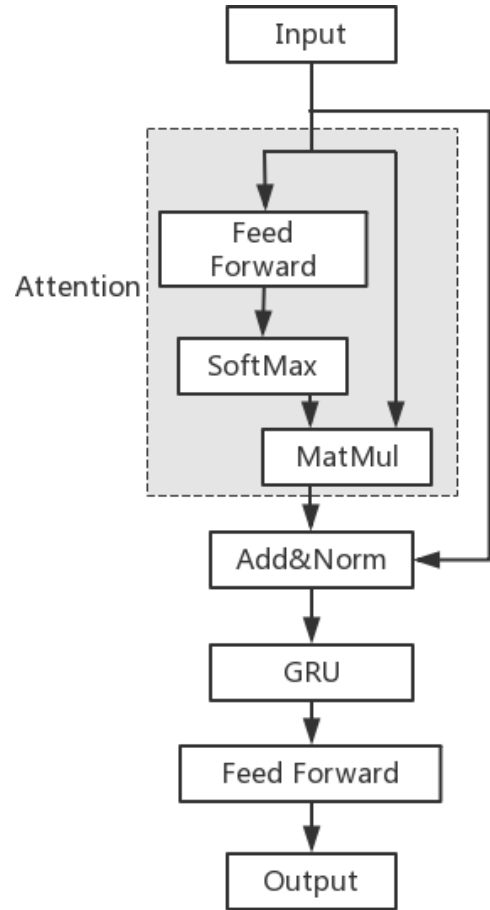

FIGURE 3. The model architecture.

utilized to capture the dependencies in different time dimensions. We employ a residual connection [39] between the GRU network and the attention mechanism, followed by layer normalization. Finally, a simple fully connected feed-forward network is used for output.

III. LEARNING ALGORITHM FOR MODEL
A. DATA DESCRIPTION

The data used for training and testing in this paper come from the data collected by hydrological sensors in the same mine. We design two sets of data to verify the ability of the hybrid neural network to learn data dependencies. One set is single-sensor data. We input the flow data monitored at time $t - n$ to t in an open channel and predict the flow at time $t + 1$ in this open channel. The other set is multi-sensor data. We input the elevation, water temperature, and burial depth of the entire underground, including aquifer sensors, open channel flow and water temperature, dibhole liquid level data from a total of 16 hydrological sensors $t - n$ to t, and predict the flow at the same open channel as the single-sensor data. To better reflect the changes of the hydrological monitoring data under the mine and highlight the correlation between the various parameters when they are changed or affected by other parameters, single-sensor data and multi-sensor data use the change values of current time and previous time.

Algorithm 1 Adam

> **begin**
>> α: learning rate;
>> $\beta_1, \beta_2 \in [0,1]$: exponential decay rates for the moment estimates;
>> $f(\theta)$: stochastic objective function with parameters θ;
>> Initialize the initial parameter vector θ_0;
>> Initialize the first moment vector $m_0 \leftarrow [0\ldots, 0]$;
>> Initialize the second moment vector $v_0 \leftarrow [0\ldots, 0]$;
>> Initialize timestep $t \leftarrow 0$;
>> **while** *Termination criterion* **do**
>>> $t \leftarrow t+1$;
>>> Calculate the gradient of stochastic objective function at timestep t: $g_t \leftarrow \nabla_\theta f_t(\theta_{t-1})$;
>>> Update moment estimates:
>>>> $m_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$;
>>>> $v_t \leftarrow \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2$;
>>> Compute bias-corrected moment estimates:
>>>> $\hat{m}_t \leftarrow m_t / (1 - \beta_1^t)$;
>>>> $\hat{v}_t \leftarrow v_t / (1 - \beta_2^t)$;
>>> Update parameters
>>>> $\theta_t \leftarrow \theta_{t-1} - \alpha \cdot \hat{m}_t / (\sqrt{\hat{v}_t} + \varepsilon)$;
>> **end while**

B. ADAM OPTIMIZATION ALGORITHM

Adam is an efficient stochastic optimization method that only requires first-order gradients [40]. Unlike a traditional stochastic gradient descent that keeps a single learning rate to update all the weights, Adam calculates the first-order moment estimation and second-order raw moment estimation of the gradient to design independent adaptive learning rates for different parameters. Adam combines the advantages of adaptive subgradient methods (AdaGrad) [41] and root mean square propagation (RMSProp) optimization algorithms. It not only works well with sparse gradients but also works well in on-line and non-stationary settings.

Let $f(\theta)$ be the objective function that needs to be optimized, and θ be the parameter that needs to be solved in the objective function; then, the gradient g_t at timestep t can be expressed as

$$g_t = \nabla_\theta f_t(\theta_{t-1}) \tag{13}$$

The first moment estimates m_t and second moment estimates v_t of the gradient g_t represent the expected estimates of g_t and g_t^2, respectively. m_t and v_t can be given by

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t \tag{14}$$
$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2 \tag{15}$$

where the hyperparameters $\beta_1, \beta_2 \in [0,1]$ control the exponential decay rates of m_t and v_t. g_t^2 indicates the elementwise square of g_t. Considering that moment estimates are biased towards zero, m_t and v_t are initialized to 0 vectors, especially when the initial time step and the decay rates are small.

Therefore, to offset the initialization bias, the bias needs to be corrected. Taking m_t as an example, Eq. (14) can be given by

$$m_t = (1 - \beta_1) \sum_{i=1}^{t} \beta_1^{t-i} \cdot g_t \tag{16}$$

By taking the expectations on both sides of Eq. (16) at the same time, it can be concluded that

$$E[m_t] = E[g_t] \cdot (1 - \beta_1^t) + \zeta \tag{17}$$

where ζ is 0 or a very small number; then, the remaining $(1 - \beta_2^t)$ is the initialization bias that we need to correct. The bias-corrected estimates $\hat{m}_t$ and $\hat{v}_t$ can be given by

$$\hat{m}_t = m_t / (1 - \beta_1^t) \tag{18}$$
$$\hat{v}_t = v_t / (1 - \beta_2^t) \tag{19}$$

The final parameter update calculation is as follows

$$\theta_t = \theta_{t-1} - \alpha \cdot \hat{m}_t / (\sqrt{\hat{v}_t} + \varepsilon) \tag{20}$$

where α is the preset learning rate and ε is the preset blur factor. Adam's pseudocode is shown as Algorithm 1.

IV. EXPERIMENT AND ANALYSIS

A. EVALUATION METHOD AND EXPERIMENT SETUP

To obtain a reliable and stable model, verification of the model is indispensable. In this experiment, because the data have a strong timing dependence, nested cross-validation is used to evaluate the model.

The nested cross-validation process can provide a nearly unbiased estimate of the true error. To generate a better estimate of the model prediction error, we perform multiple training and test data segmentation and then calculate the average value of the error on these segmentations. A schematic diagram of the nested cross-validation is shown in Fig. 4.
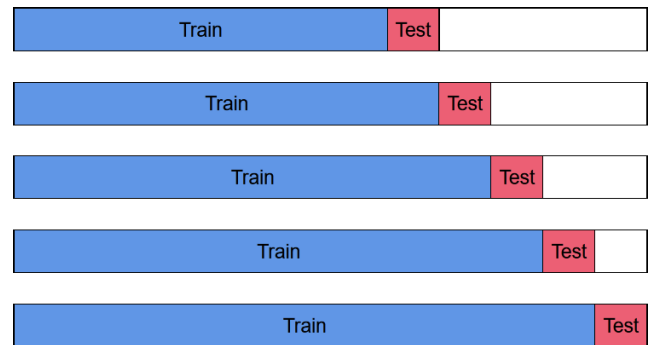

FIGURE 4. Schematic diagram of the nested cross-validation.

We split the training data and test data in a 7:3 ratio. To facilitate further segmentation during the nested cross-validation, the test data are rounded up to 10 validations. The number of training data and test data is shown in Table 1.

The parameters of the hybrid neural network model are set as in Table 2. The four parameters α, β_1, β_2 and ε of the

TABLE 1. Number of training data and test data.

Data Type	Number of training data	Number of test data
Multi-sensor data	641	200

TABLE 2. The parameters of the hybrid neural network model.

Parameters	Value
Input time steps	10
Batch Size	30
GRU Units	50
Epoch	15
α	0.001
β_1	0.9
β_2	0.999
ε	1×10^{-8}

Adam optimization algorithm use the default values in the original paper. Other parameters obtain better performance through multiple experiments. The number of epochs is set as 15 because the accuracy is stable according to repeated experiments. The results of the nested cross-validation are shown in Fig. 5.
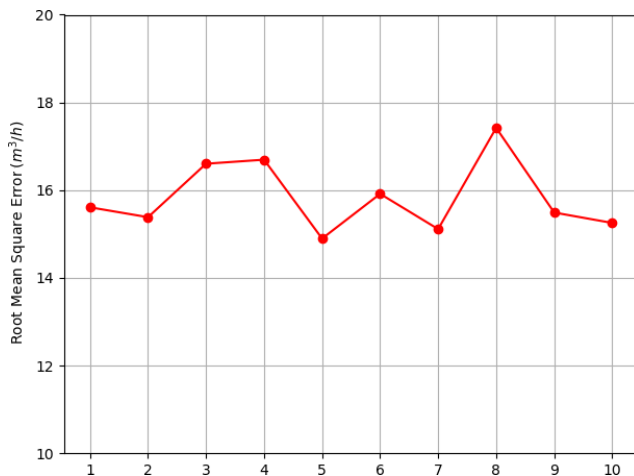

FIGURE 5. The results of the nested cross-validation.

B. EXPERIMENTAL RESULTS AND ANALYSIS

This paper uses a GRU network model that introduces the attention mechanisms for the real-time prediction of open channel flow in mines and compares this model with three recurrent neural network models: the LSTM, the GRU, and the LSTM model that also introduces the attention mechanism. The average training cost and prediction error of repeated experiments are shown in Table 3, and one of the experimental prediction results is shown in Fig. 6 and Fig. 7.

1) MODEL PERFORMANCE ANALYSIS

Fig. 6 shows the prediction results of single-sensor data and multi-sensor data using the LSTM or GRU model alone. Comparing the performance of the LSTM and GRU under the same set of data, the prediction curves of the two models under the same set of data are similar. However, the prediction results of the GRU are slightly stronger than those of the LSTM at some times.

As shown, in terms of the prediction error, the RMSE and MAE of the GRU models are slightly better than those of the LSTM models under the two sets of data, but the difference is not obvious. In terms of the training cost, the GRU model is also lower than the LSTM model. Especially under multi-sensor data, the difference between the two models is more than one second.

It can be seen that when the flow prediction problem of open channels is investigated under the data selected in this paper, although the GRU model combines and simplifies the gate functions in the LSTM model, they are still valid to pass the information obtained from the previously hidden state to the currently hidden state, thereby helping the recurrent neural network to remember long-term information. At the same time, due to the simplification of the gate function, the hidden state is more compactly represented, which reduces the overall calculation amount of the model. In particular, multi-sensor data require a large amount of data calculation, and the reduction in GRU model training costs can be seen more clearly.

2) DEPENDENCY MODELLING ANALYSIS

The LSTM and GRU have separate gate functions for each hidden unit that have the ability to capture dependencies on different time scales. Taking the GRU as an example, when the reset gates in the hidden unit are active, the short-term dependencies are captured, and when the update gates are active, the long-term dependencies are captured. As shown in Fig. 6, the prediction results of the same model using single-sensor data and multi-sensor data are compared: under multi-sensor data, because the input data contain more information, the model captures more timing dependencies. Therefore, the prediction result is not as close to the average line as in single-sensor data, and the prediction accuracy has improved.

The prediction results of the LSTM and GRU models after introducing attention mechanisms are shown in Fig. 7. Comparing the results of models without attention mechanisms in Fig. 6, the prediction accuracy of the former is much higher than that of the latter. For example, in the case of multi-sensor data, the RMSE and MAE of the prediction results of the pure GRU model are 19.689 m³/h and 16.927 m³/h, respectively, and the RMSE and MAE of the prediction results are reduced to 15.265 m³/h and 12.656 m³/h, respectively, after the attention mechanism is added. After adding attention mechanisms, the prediction curves of the models are closer to the original data curve. This demonstrates the ability of the attention mechanism to learn inter-data dependencies and verifies the importance of dependency modelling to improve prediction accuracy in such prediction problems.

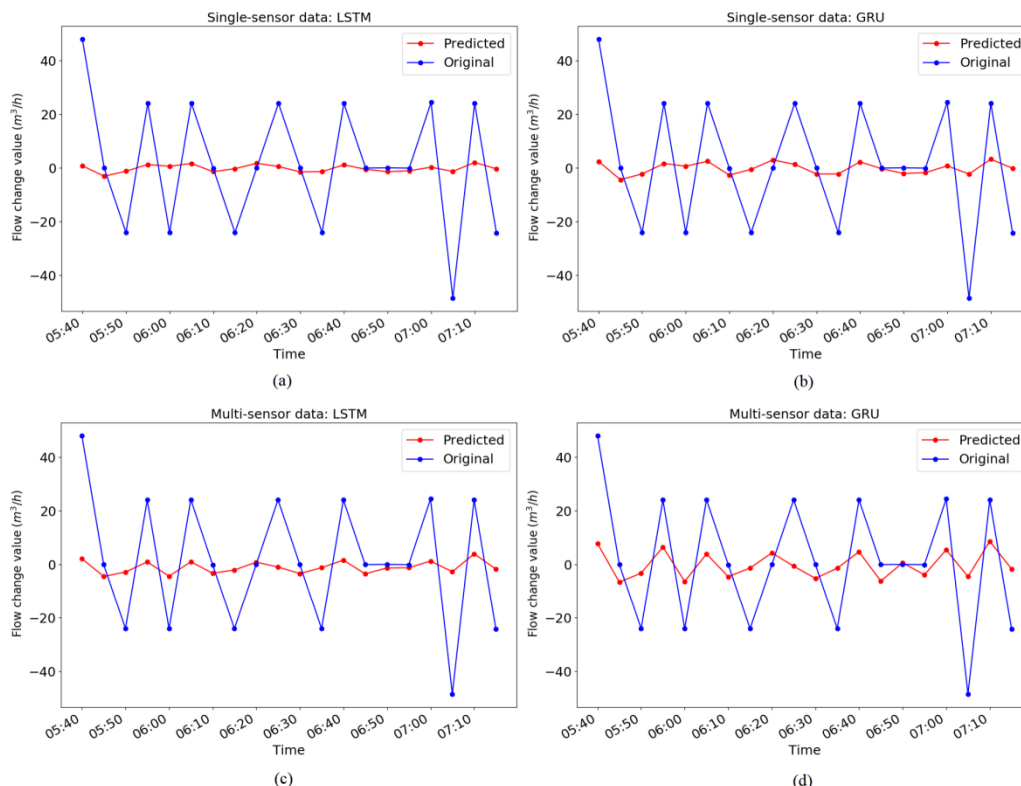The GRU model with an attention mechanism is used to perform another open channel flow prediction experiment.

FIGURE 6. Comparison of prediction results of the LSTM and GRU models.

TABLE 3. Comparison of model training time and prediction error.

Training Data	Model	Training Cost(s)	RMSE(m³/h)	MAE(m³/h)
Single sensor data	LSTM	2.81	22.452	17.782
	GRU	2.58	21.972	17.895
	Attention+LSTM	4.09	21.578	17.563
	Attention+GRU	3.66	20.714	16.758
Multi-sensor data	LSTM	6.39	22.475	17.970
	GRU	5.02	19.993	16.908
	Attention+LSTM	7.54	17.927	14.744
	Attention+GRU	6.39	**16.425**	**13.472**

TABLE 4. Other open channel flow prediction errors.

Training Data	Model	RMSE(m³/h)	MAE(m³/h)
Multi-sensor data	Attention+GRU	8.223	6.306

Table 4 shows the prediction error of open channel flow, with an RMSE of 8.223 m³/h and MAE of 6.306 m³/h; Fig. 8 compares the curves of the actual flow and predicted flow. From Table 4 and Fig. 8, it is shown that the hybrid neural network model proposed in this paper has also achieved good results in other open channel flow prediction experiments. This further validates the predictive ability of the model proposed in this paper.

C. COMPARISON RESULTS

To further verify the performance of our proposed model in the real-time prediction of open channel flow, we compared

our model with a series of models in the same research area and those in other research areas with similar algorithm structures. The models for comparison include back propagation neural networks (BPNNs), the hybrid method coupling artificial neural networks with genetic algorithm methods (ANN-GA) for predicting groundwater inflow in mines [30], the random forest regression method (RFR) for predicting the height of fractured water-conducting zones in coal roof strata [42], the hybrid method coupling LSTM with support vector machines (LSTM+SVM) for fault prediction [43], and a deep neural network-based traffic flow prediction model (DNN-BTF) that also uses attention mechanisms and GRU networks [44]. The comparative experiment was conducted using the same samples in Table 1. Due to the differences in data dimensions, some parameters of the above models are adjusted in the experiment; for example, the batch size is unified as 30, the output unit is changed to 1, the LSTM/GRU
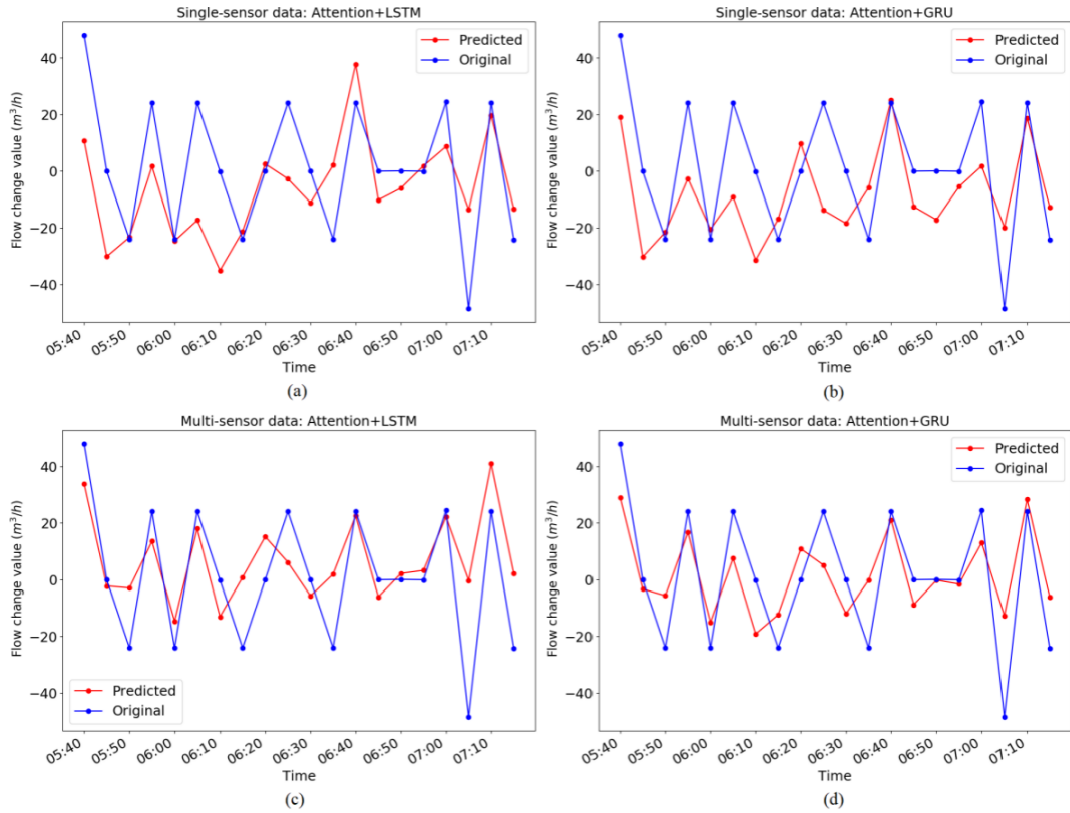
FIGURE 7. Comparison of model prediction results after adding attention mechanisms.
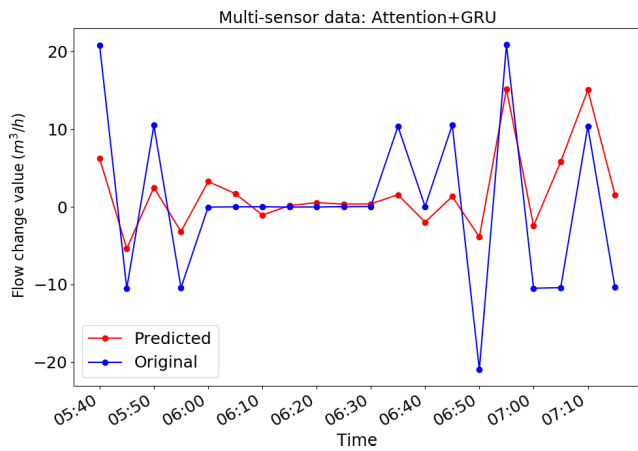

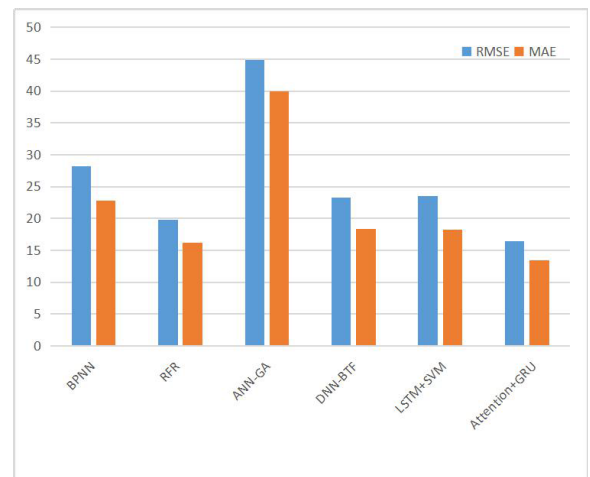**FIGURE 8.** Another open channel flow forecast result.

FIGURE 9. Comparison results for the proposed method and similar methods, which include BPNNs, RFR, ANN-GA, DNN-BTF, LSTM+SVM.

units are changed to 50, the epochs is set to 20, and the remaining parameters were unchanged. The average error of prediction after repeated experiments is shown in Fig. 9 and Table 5.

The results show that the proposed method exhibits better performance than BPNNs, RFR, ANN-GA, DNN-BTF and LSTM+SVM. GRU networks can extract long short-term dependence from data according to hidden layer units, but BPNNs, RFR, and ANN-GA cannot extract this information. Although LSTM+SVM uses LSTM networks to

obtain time dependence, it lacks the attention mechanism to learn the dependence relationship between multi-sensor data. DNN-BTF also uses the attention mechanism and GRU networks, but on the one hand, the method we proposed uses the self-attention mechanism, and a feed-forward network is used to replace the traditional similarity calculation function so that the method can better learn the dependency relationship for the target data. On the other hand,

TABLE 5. Comparison results for the proposed method and similar methods.

Methods	BPNN	RFR	ANN-GA	DNN-BTF	LSTM+SVM	Attention+GRU
RMSE(m³/h)	28.153	19.761	44.917	23.288	23.555	16.425
MAE(m³/h)	22.849	16.236	39.954	18.384	18.206	13.472

convolutional networks are used in DNN-BTF, and convolutional networks will weaken the input boundary features when extracting features, which is extremely disadvantageous for the prediction of the time sequence.

In general, the real-time prediction method of open channel flow based on GRU networks and attention mechanisms proposed in this paper is a noteworthy method for early-warning research on coal mine water disasters.

V. CONCLUSION
In this work, we propose a hybrid neural network model combining the attention mechanism and a GRU structure for real-time prediction of open channel flow in a mine. After performing and analysing a series of comparative experiments, the model demonstrated a higher prediction accuracy than the traditional LSTM and GRU algorithms. In addition, these experiments also expose the dependence relationship between the sensor data. The exploration of the dependence relationship helps us predict the amount of water in the mine. Compared with some traditional recurrent neural network models, the GRU network model based on attention mechanisms can better learn the dependence relationship between input data.

In future work, we will use more detailed monitoring data for training on the basis of this hybrid neural network model and at the same time improve the model such that it can more fully and reasonably establish the dependency model between data. This work will further improve the accuracy of the prediction results, provide technical support for the prevention of water inflows and guarantee coal mine safety. We can also apply the hybrid neural network to automated manufacturing systems [45] and social networks [46].

REFERENCES
[1] Q. Wu, "Progressproblems and prospects of prevention and control technology of mine water and reutilization in China," *J. China Coal Soc.*, vol. 39, no. 5, pp. 795–805, May 2014.
[2] J. Wu, F.-P. Cui, S.-Q. Zhao, S.-J. Liu, Y.-F. Ceng, and Y.-W. Gu, "Type classification and main characteristics of mine water disasters," *J. China Coal Soc.*, vol. 38, no. 4, pp. 561–565, 2013.
[3] Y. Yihdego and A. Paffard, "Predicting open pit mine inflow and recovery depth in the durvuljin soum, zavkhan province, mongolia," *Mine Water Environ.*, vol. 36, no. 1, pp. 114–123, Mar. 2017.
[4] T. Li, H. Song, G. Huang, Y. Bi, and X. Li, "Assessment of groundwater changing trends through the generalized large well method with confined–unconfined flow model in open-pit mine area," *Environ. Earth Sci.*, vol. 72, no. 11, pp. 4599–4606, Dec. 2014.
[5] F. Marinelli and W. L. Niccoli, "Simple analytical equations for estimating ground water inflow to a mine pit," *Ground Water*, vol. 38, no. 2, pp. 311–314, Mar. 2000.
[6] K. Zhang, B. Cao, G. Lin, and M. Zhao, "Using multiple methods to predict mine water inflow in the Pingdingshan no. 10 coal mine, China," *Mine Water Environ.*, vol. 36, no. 1, pp. 154–160, Mar. 2017.
[7] C. Wu, X. Wu, G. Zhu, and C. Qian, "Predicting mine water inflow and groundwater levels for coal mining operations in the pangpangta coalfield, China," *Environ. Earth Sci.*, vol. 78, no. 5, Mar. 2019.
[8] M. Shabanimashcool and C. C. Li, "Numerical modelling of longwall mining and stability analysis of the gates in a coal mine," *Int. J. Rock Mech. Mining Sci.*, vol. 51, pp. 24–34, Apr. 2012.
[9] B. Miladinović, V. R. Vakanjac, D. Bukumirović, V. Dragisić, and B. Vakanjac, "Simulation of mine water inflow: Case study of the Štavalj coal mine (Southwestern Serbia)," *Arch. Mining Sci.*, vol. 60, no. 4, pp. 955–969, Dec. 2015.
[10] H. Mengistu, A. Tessema, T. Abiye, M. Demlie, and H. Lin, "Numerical modeling and environmental isotope methods in integrated mine-water management: A case study from the witwatersrand basin, south africa," *Hydrogeol. J.*, vol. 23, no. 3, pp. 533–550, May 2015.
[11] W. Li, Z. Liu, H. Guo, N. Li, and W. Kang, "Simulation of a groundwater fall caused by geological discontinuities," *Hydrogeol. J.*, vol. 19, no. 6, pp. 1121–1133, Sep. 2011.
[12] H. Zhang, Y. Wang, R. Yang, and R. Ye, "Modeling the effects of phosphate mining on groundwater at different stages of mine development," *Mine Water Environ.*, vol. 37, no. 3, pp. 604–616, Sep. 2018.
[13] S. Xue, Y. Liu, S. Liu, W. Li, Y. Wu, and Y. Pei, "Numerical simulation for groundwater distribution after mining in zhuanlongwan mining area based on visual MODFLOW," *Environ. Earth Sci.*, vol. 77, no. 11, Jun. 2018.
[14] Y. Zhang, Y. Xu, K. Wang, P. Chen, X. Wang, Q. Zheng, and Z. Zhang, "The fracturing characteristics of rock mass of coal mining and its effect on overlying unconsolidated aquifer in Shanxi, China," *Arabian J. Geosci.*, vol. 11, no. 21, Nov. 2018.
[15] Y. Chen, A. P. S. Selvadurai, and W. Liang, "Computational modelling of groundwater inflow during a longwall coal mining advance: A case study from the Shanxi Province, China," *Rock Mech. Rock Eng.*, vol. 52, no. 3, pp. 917–934, Mar. 2019.
[16] W. Mu, Q. Wu, Y. Xing, C. Qian, Y. Wang, and Y. Du, "Using numerical simulation for the prediction of mine dewatering from a karst water system underlying the coal seam in the yuxian basin, northern China," *Environ. Earth Sci.*, vol. 77, no. 5, Mar. 2018.
[17] T. Hanna, E. Azrag, and L. Atkinson, "Use of an analytical solution for preliminary estimates of ground water inflow to a pit," *Mining Eng.*, vol. 46, no. 2, pp. 149–152, 1994.
[18] F. Doulati Ardejani, R. N. Singh, E. Baafi, and I. Porter, "A finite element model to: 1. Predict groundwater inflow to surface mining excavations," *Mine Water Environ.*, vol. 22, no. 1, pp. 31–38, Apr. 2003.
[19] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, no. 2, Jan. 2012.
[20] J. Gehring, M. Auli, D. Grangier, D. Yarats, and Y. N. Dauphin, "Convolutional sequence to sequence learning," 2017, *arXiv:1705.03122*. [Online]. Available: http://arxiv.org/abs/1705.03122
[21] R. Li, Y. Huang, and J. Wang, "Long-term traffic volume prediction based on K-means Gaussian interval type-2 fuzzy sets," *IEEE/CAA J. Automatica Sinica*, vol. 6, no. 6, pp. 1344–1351, Nov. 2019.
[22] S. Gao, M. Zhou, Y. Wang, J. Cheng, H. Yachi, and J. Wang, "Dendritic neuron model with effective learning algorithms for classification, approximation, and prediction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 2, pp. 601–614, Feb. 2019.
[23] J. Bi, H. Yuan, M. Zhou, and Q. Liu, "Time-dependent cloud workload forecasting via multi-task learning," *IEEE Robot. Autom. Lett.*, vol. 4, no. 3, pp. 2401–2406, Jul. 2019.
[24] E. Principi, D. Rossetti, S. Squartini, and F. Piazza, "Unsupervised electric motor fault detection by using deep autoencoders," *IEEE/CAA J. Automatica Sinica*, vol. 6, no. 2, pp. 441–451, Mar. 2019.
[25] Y. Gu, W. Lu, X. Xu, L. Qin, Z. Shao, and H. Zhang, "An improved Bayesian combination model for short-term traffic prediction with deep learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 3, pp. 1332–1342, Mar. 2020.

[26] G. Wang, J. Qiao, J. Bi, W. Li, and M. Zhou, "TL-GDBN: Growing deep belief network with transfer learning," *IEEE Trans. Autom. Sci. Eng.*, vol. 16, no. 2, pp. 874–885, Apr. 2019.

[27] H. Li, Y. Wang, X. Xu, L. Qin, and H. Zhang, "Short-term passenger flow prediction under passenger flow control using a dynamic radial basis function network," *Appl. Soft Comput.*, vol. 83, Oct. 2019, Art. no. 105620.

[28] H. Guo, M. Ji, K. Chen, Z. Zhang, Y. Zhang, and M. Zhang, "The feasibility of mining under a water body based on a fuzzy neural network," *Mine Water Environ.*, vol. 37, no. 4, pp. 703–712, Dec. 2018.

[29] J. Wu, S. Xu, R. Zhou, and Y. Qin, "Scenario analysis of mine water inrush hazard using Bayesian networks," *Saf. Sci.*, vol. 89, pp. 231–239, Nov. 2016.

[30] S. Bahrami, F. Doulati Ardejani, and E. Baafi, "Application of artificial neural network coupled with genetic algorithm and simulated annealing to solve groundwater inflow problem to an advancing open pit mine," *J. Hydrol.*, vol. 536, pp. 471–484, May 2016.

[31] F. D. Ardejani, M. Sadeghiamirshahidi, R. N. Singh, T. E. Kish, and S. M. Reed, "Prediction of the groundwater rebound process in a backfilled open cut mine using an artificial neural network," *Mine Water Environ.*, vol. 32, no. 4, pp. 251–257, Dec. 2013.

[32] V. Mnih, N. Heess, A. Graves, and K. Kavukcuoglu, "Recurrent models of visual attention," 2014, *arXiv:1406.6247*. [Online]. Available: http://arxiv.org/abs/1406.6247

[33] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*. [Online]. Available: http://arxiv.org/abs/1409.0473

[34] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2017, *arXiv:1706.03762*. [Online]. Available: http://arxiv.org/abs/1706.03762

[35] D. Britz, A. Goldie, M.-T. Luong, and Q. Le, "Massive exploration of neural machine translation architectures," 2017, *arXiv:1703.03906*. [Online]. Available: http://arxiv.org/abs/1703.03906

[36] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, *arXiv:1406.1078*. [Online]. Available: http://arxiv.org/abs/1406.1078

[37] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[38] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.

[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[40] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[41] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *J. Mach. Learn. Res.*, vol. 12, no. 7, pp. 257–269, 2011.

[42] D. Zhao and Q. Wu, "An approach to predict the height of fractured water-conducting zone of coal roof strata using random forest regression," *Sci. Rep.*, vol. 8, no. 1, pp. 1–12, Jul. 2018.

[43] S. Zhang, Y. Wang, M. Liu, and Z. Bao, "Data-based line trip fault prediction in power systems using LSTM networks and SVM," *IEEE Access*, vol. 6, pp. 7675–7686, 2018.

[44] Y. Wu, H. Tan, L. Qin, B. Ran, and Z. Jiang, "A hybrid deep learning based traffic flow prediction method and its understanding," *Transp. Res. C, Emerg. Technol.*, vol. 90, pp. 166–180, May 2018.

[45] X. Zan, Z. Wu, C. Guo, and Z. Yu, "A Pareto-based genetic algorithm for multi-objective scheduling of automated manufacturing systems," *Adv. Mech. Eng.*, vol. 12, no. 1, pp. 1–15, 2020.

[46] L. Yang, Z. Yu, M. A. El-Meligy, A. M. El-Sherbeeny, and N. Wu, "On multiplexity-aware influence spread in social networks," *IEEE Access*, vol. 8, pp. 106705–106713, 2020.

ZHANLI LI received the B.S. degree from Xianyang Normal University, Xi'an, China, in 1984, the M.S. degree from the Xi'an University of Science and Technology, Xi'an, in 1989, and the Ph.D. degree from Xi'an Jiaotong University, Xi'an, in 1997. He is currently a Professor and the Dean of the School of Computer, Xi'an University of Science and Technology. He is also a Professor of engineering with Iwate University, Japan. He has authored more than 50 technical papers for conferences and journals. He holds more than ten invention patents. His main research interests include artificial intelligence, danger perception, and image identification.

TIANYU GAO received the B.S. and M.S. degrees from the Xi'an University of Science and Technology, Xi'an, China, in 2013 and 2016, respectively, where he is currently pursuing the Ph.D. degree in safety science and engineering. His main research interests include artificial intelligence and danger perception.

CHENG GUO received the B.S. degree from the Beijing Institute of Technology, Beijing, China, in 1998, the M.S. degree from Xi'an Technological University, Xi'an, China, in 2007, and the Ph.D. degree from the Xi'an Institute of Applied Optics, Xi'an, in 2013. He is currently a Senior Engineer with the Xi'an Institute of Applied Optics. He has authored more than ten technical papers for conferences and journals. He holds two invention patents. His main research interests include artificial intelligence and photoelectric information processing.

HONG-AN LI received the M.S. and Ph.D. degrees in computer science and technology from Northwest University, Shaanxi, China, in 2009 and 2014, respectively. From 2014 to 2019, he was an Associate Professor with the College of Computer Science and Technology, Xi'an University of Science and Technology. His research interests include computer graphics, computer-aided geometric design, computer animation, virtual reality, and image processing.

● ● ●