



OPEN Time series prediction model using LSTM-Transformer neural network for mine water inflow

Junwei Shi✉, Shiqi Wang, Pengfei Qu & Jianli Shao

Mine flooding accidents have occurred frequently in recent years, and the predicting of mine water inflow is one of the most crucial flood warning indicators. Further, the mine water inflow is characterized by non-linearity and instability, making it difficult to predict. Accordingly, we propose a time series prediction model based on the fusion of the Transformer algorithm, which relies on self-attention, and the LSTM algorithm, which captures long-term dependencies. In this paper, Baotailong mine water inflow in Heilongjiang Province is used as sample data, and the sample data is divided into different ratios of the training set and test set in order to obtain optimal prediction results. In this study, we demonstrate that the LSTM-Transformer model exhibits the highest training accuracy when the ratio is 7:3. To improve the efficiency of search, the combination of random search and Bayesian optimization is used to determine the network model parameters and regularization parameters. Finally, in order to verify the accuracy of the LSTM-Transformer model, the LSTM-Transformer model is compared with LSTM, CNN, Transformer and CNN-LSTM models. The results prove that LSTM-Transformer has the highest prediction accuracy, and all the indicators of its model are well improved.

Keywords Mine water inflow, LSTM-Transformer model, Time series prediction, Long short-term memory, Self-attention mechanism

Background

As one of the “five major disasters” in coal mines, mine water hazards have become the second largest “killer” after gas accidents in terms of threatening coal mine safety and workers’ lives¹. Mine water inflow refers to the phenomenon of groundwater or underground water sources entering the mine or underground works due to underground pressure or other factors. Typically, mine water continuously and slowly flows into the mine and is pumped to the surface through underground drainage equipment, without affecting the normal construction and production of the mine. However, in some cases, this water can suddenly flood into the underground workspace in a short period, leading to localized production interruptions by damaging equipment, or even causing casualties and mine flooding accidents. From 2000 to 2022, a total of 1206 water hazard accidents occurred in coal mines nationwide, resulting in 5018 deaths, including 103 major water hazard accidents that caused 2039 deaths². These statistics indicate that water hazard accidents pose a severe threat to coal mine safety and miners’ lives. One of the important causes of mine water hazard accidents is the sudden water inrush at burst points, which leads to a rapid increase in water inflow. The occurrence of burst points is often unpredictable, and once it happens, the water inflow can increase sharply in a short time, seriously threatening miners’ safety and causing damage to mine equipment. Therefore, predicting the water inflow at burst points is crucial for preventing mine water hazard accidents.

Currently, establishing accurate water inflow prediction models based on monitoring data has become an important task. The goal of the water inflow prediction model is to achieve real-time prediction of water inflow at burst points, providing scientific evidence for mine water hazard early warning and emergency response. However, water inflow prediction faces multiple challenges due to the significant variability in the water-bearing system caused by mining disturbances, such as overlying strata lithology and structures. This results in a high degree of uncertainty in the distribution of water-conducting fracture structures, including data complexity, uncertainty, and diversity, as well as the sudden and uncontrollable nature of water inrush events. Therefore, achieving real-time prediction of water inflow during burst events to ensure the timeliness and accuracy of the early warning system is key to solving this problem.

College of Management Science and Engineering, Shandong Technology and Business University, Yantai 264000, China. ✉email: shijunwei302@sdtbu.edu.cn

At the same time, mine water inflow data is often influenced by various factors such as geological conditions, meteorological changes, groundwater flow, and mining activities. These factors interact, causing the water inflow data to exhibit high nonlinearity and uncertainty. Additionally, the suddenness and high frequency of mine water inrush events increase the difficulty of water inflow prediction. In this complex context, traditional prediction methods often fail to meet practical needs, requiring advanced technical means to improve prediction accuracy and real-time performance.

Therefore, facing the severe challenge of mine water hazards, improving the prediction accuracy of water inflow at mining faces is crucial. To address key issues such as the small sample size of water inflow data, unreliable selection of hydrogeological parameters, and high nonlinearity and uncertainty of water inflow data in the construction of prediction models, this paper introduces an integrated deep learning model, LSTM-Transformer, and its optimization techniques. This approach can more comprehensively consider the relevant factors of water inflow at the working face, effectively improving the real-time performance and accuracy of water inflow prediction, and providing strong technical support for mine water hazard early warning and emergency response.

Research status

Mine inflow prediction is a crucial part of ensuring safe production in mines. At present, in the research field of mine inflow prediction in China, the following three methods are mainly used: statistical analysis methods, theoretical research methods, and prediction methods based on deep learning. Statistical analysis methods typically involve probability calculations and comprehensive analysis. These methods are applied to issues involving complex systems and multivariate data, such as the properties of amorphous solids³, mechanisms of material aging^{4,5}, defect studies in metallic glasses⁶, and the nonequilibrium evolution of inherent structures in disordered materials⁷. Regarding the issue of predicting gas emission quantity, researchers have proposed analytical formulae for predicting mine water inflow in the past, such as the ones proposed by Singh et al.⁸ in 1985 laid the groundwork for this field of research. These formulas have an explicit definition of the constraints and applicability of the prediction conditions for accurate mine inflow prediction, which provides significant theoretical support. According to research, there is a degree of regularity to the seemingly irregular, disorderly and complex phenomenon of mine water inflow fluctuations⁹. Through the use of Hurst index trend analysis, we successfully capture the regularity in the nonlinear time series and convert it from qualitative to quantitative by taking into account the length and correlation of the series¹⁰. However, the Hurst index is a pure mathematical statistic, and the results usually lack a mechanistic explanation of the mine water inflow hypothesis, thus making it difficult for the predictions to cope with the complexity of the actual mine water inflow synthesis hypothesis. For the mine water inflow uncertainly, the cloud-weighted Markov prediction model combines the stochasticity and ambiguity of the cloud model with the state transfer feature of the Markov chain. The innovation of the model is that it is able to “soft-divide” the mine inflow state to deal with its uncertainty in the prediction, providing a new methodology for mine inflow prediction¹¹. At the same time the method has some drawbacks due to the fact that the assumption of Markov forecasting is in some cases oversimplified, thus failing to capture the long-term dependencies in the time series.

Theoretical research methods tend to utilize nonlinear mathematical models to study various aspects from microscopic mechanisms^{12,13} to macroscopic behavior^{14,15}. For example, these methods are used to investigate the mechanisms of defect formation¹⁶, in conjunction with molecular dynamics to quantify key parameters¹⁷, as well as to capture and understand disordered structures¹⁸. Numerical simulation is a common mathematical model used to predict the inflow of water, the core of which is to take the watershed as the basic unit and different lithological strata as the basic structural layer, based on the regional groundwater overflow, groundwater level data, etc., model fitting and simulation prediction¹⁹, relative Comparison²⁰ and Analysis methods²¹, numerical methods provide a more realistic portrayal of the mine inflow prediction entity, and thus numerical methods are more accurate in predicting mine inflows²². As the general mine water inflow prediction is for a certain production stage or time period, the predicted value belongs to the instantaneous amount, rather than the average amount². Therefore, on the basis of numerical simulation model identification, utilizing a quantitative method combined with a pumping test inversion of the model for correction, access to the simulation area based on mining hydrogeological parameters for correction. By developing numerical models to represent the groundwater flow system, a miner can intuitively predict the single face in the process of mining back to the water inflow rule, but they can also predict the adjacent multiple mined faces in the new face under the influence of mining back to the water inflow changes in the characteristics of the workings of the workface, emphasizing the dynamic changes in the water inflow in conjunction with the mining process back to the law²³. Part of the reason for the large error in the water inflow prediction compared to the measured water inflow from the mine is that changes in the permeability characteristics of the overburden fissures caused by mining have not been taken into account. Based on the development characteristics of water-conducting fissure zone and mine compaction data, the dynamic change law of overburden rock mining permeability was analyzed, and the analytical method was used to predict the mine water inflow in the working face according to the trend of the development of water-conducting fissure zone, which provided a new way of thinking for the prediction of mine water inflow²⁴. However, nonlinear mathematical models are usually complex and difficult to explain the parameters and internal mechanisms embedded in them, which may limit the in-depth understanding of the inflow mechanism and reduce the interpretability of the model in the mining industry.

With the advancement of technology, deep learning methods have shown great potential in prediction tasks such as anomaly detection²⁵, predicting changes in the microstructure of alloys²⁶, predicting the overall energy of metallic glasses²⁷, and forecasting alloy durability²⁸. Therefore, applying deep learning methods to the problem of predicting water inflow quantities holds significant promise. Deep learning models such as, ARIMA²⁹, XG-Boost³⁰, GRU³¹ show some advantages in mine water inflow prediction, but a single deep learning model

has the disadvantages of not being able to capture complex data relationships and being more sensitive to noise and outliers³². Recurrent neural network (RNN) has shown strong capabilities in natural language processing, image recognition, time series prediction^{33,34}, but RNN has the problem of gradient explosion and gradient vanishing during the training process³⁵. As a modified version of LSTM, it can effectively avoid these problems and has significant advantages with regard to long-term dependencies³⁶. With the development of self-attention mechanisms, self-attention has shown great potential in time series problems such as wind speed prediction and electricity forecasting. Multi-head self-attention mechanisms excel at capturing complex dependencies, making them widely applied and recognized in many fields. Based on this, we chose to introduce the Transformer model, which uses the self-attention mechanism as its core module, into water inflow prediction to enhance the performance of the prediction model. The Transformer model, due to its ability to process data in parallel, can effectively handle long-distance dependencies, making it perform excellently when dealing with sequence data. However, although the Transformer has advantages in capturing global dependencies, its performance may be lacking when processing local time series information. To address this, we use the LSTM model as the Decoder part of the Transformer, leveraging LSTM strong temporal modeling capabilities to enhance the model's ability to capture local temporal dependencies, thereby improving prediction accuracy.

Research gap

Collecting data on water inflow within mines is challenging, and the number of samples available for time series prediction is often limited. Small sample datasets tend to be imbalanced, making it difficult for traditional deep learning prediction models to fully understand and capture the variations in water inflow at inflow points, thereby affecting the accuracy and stability of the prediction models. Therefore, effectively utilizing limited data resources and adjusting and optimizing prediction models according to specific geological and hydrological conditions within the mine becomes one of the key challenges in improving water inflow prediction capabilities.

The variation of water inflow in mines is often nonlinear, influenced by various nonlinear factors such as geological structures, precipitation, and groundwater levels. Traditional statistical methods often perform poorly in handling these nonlinear relationships, failing to capture the complex patterns of water inflow variations. Theoretical analysis methods, primarily based on numerical simulation models, are usually established for specific mines and conditions, requiring the satisfaction of applicable conditions and the provision of reliable parameters before use. However, it is difficult to meet these requirements with field geological and hydrogeological data, making these methods hard to generalize and limiting their applicability. The promotion and application of models across different mines or geological conditions require readjustment and recalibration, resulting in weak generalization capabilities.

Therefore, this paper aims to improve the accuracy of water inflow prediction from three key perspectives: small sample prediction of water inflow at burst points, nonlinear variation characteristics of water inflow data, and how to enhance the generalization ability of the prediction model for water inflow at water gushing points:

Conduct smoothness tests on water inflow data at burst points and divide the dataset into different proportions to meet the prediction needs of small sample datasets. In water inflow prediction, smoothness tests can help identify whether the data is stationary, allowing the model to be trained and predicted on a more stable data basis, thereby improving prediction accuracy. Therefore, the Augmented Dickey–Fuller (ADF) test is chosen to verify the stationarity of the water inflow data. Different dataset division ratios can help identify and address data imbalance issues, improving the robustness and generalization ability of the model. For small sample datasets, analyzing different division ratios can better balance the quantity of training and testing data, ensuring the model achieves the best learning effect and evaluation results under limited data conditions. Therefore, this paper deeply analyzes the impact of dataset division on model performance by comparing evaluation metric results of different dataset division ratios.

Construct a model for predicting nonlinear long-term series of water inflow at burst points. The data on water inflow at mine burst points is influenced by various factors such as geological structures and hydrological characteristics. Its variations are complex and difficult to predict, exhibiting significant nonlinearity and temporal correlation characteristics. To better capture the complex nonlinear temporal characteristics of water inflow, this paper integrates the Transformer model, which has a self-attention mechanism as its core, with the long short-term memory (LSTM) model. The LSTM-Transformer model's parallel structure is used to process the time series prediction of mine water inflow.

Enhance the model's generalization ability by combining random search and Bayesian optimization methods. When dealing with complex time series data, the LSTM-Transformer model requires tuning of hyperparameters to capture patterns and regularities in the data. Initial exploration can be done using random search, followed by fine-tuning with Bayesian optimization. Random search and Bayesian optimization can flexibly adjust model parameters in the face of dynamic changes and uncertainties in the data, enhancing the model's robustness and stability, and further improving model performance.

Theory and modeling

Prediction model based on single neural network

Recurrent neural network (RNN) is not only fully connected between adjacent network layers, but also neurons are interconnected between internal³⁷. RNN has an important structural feature that allows information to be passed from one time step to the next to capture temporal dependencies in the data³⁸. The network structure is shown in Fig. 1. But there are some problems with traditional recurrent neural networks, such as gradient explosion and gradient vanishing, this limits its performance when dealing with long sequences. In order to solve

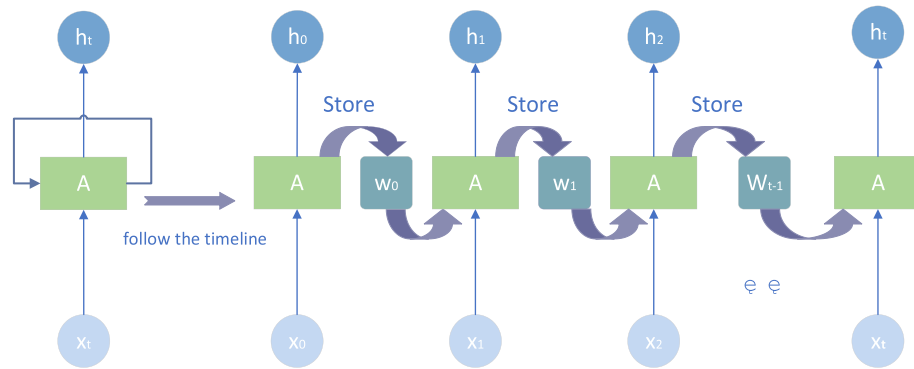


Figure 1. RNN network structure.

this series of problems, the architecture of RNN has been improved and some more efficient neural network structures have been put forward, such as long and short-term memory networks and the Transformer model.

LSTM is a variant of RNN designed to process data and capture and memorize long-term dependencies. Each structure of LSTM has a clear division of labor, memory cells are responsible for storing and transmitting information, the input gate determines what information can be added to the memory cell, the forget gate determines which information is to be deleted or forgotten from the memory unit, the output gate determines what information will be output at the current time. Gated structure avoids the problem of vanishing gradient and gradient explosion after multistage backpropagation³⁹.

The Transformer is completely based on the attention mechanism, with no need for recursion and convolution. Allows the model to efficiently capture correlations between different locations in the input sequence, thus better handling long distance dependencies⁴⁰. The self-attention mechanism is one of the most important components of the Transformer. The self-attention mechanism controls the model to assign different weights to each position in the sequence in order to focus on the information at different positions when processing the inputs. Transformer is usually made up of encoder-decoder, the encoder is used to process the input sequence, the decoder is used to generate the output sequence. Moreover, compared with conventional convolutional neural networks and recurrent neural networks, Transformer uses a new positional encoding mechanism to capture time series information between input data.

In the beginning, the Convolutional neural network (CNN) was originally designed for image processing, but is now extensively used for natural language processing, such as text generation, text categorization, time series classification. Compared to traditional RNN, CNN has fewer parameters because the convolutional kernel shares weights over the entire time series. This reduces the risk of overfitting, making the CNN more robust to small data samples. CNN structure generally consists of three parts⁴¹. The first part is the convolutional layer, which is one of the core components of the CNN, and the convolutional operation involves sliding a convolutional kernel over the input data in order to extract local features from the sample data. The second part is the pooling layer, which is used to reduce the size of the feature mapping and reduce computational complexity. The pooling operation also makes the model more robust to small changes in the inputs and improves the generalization ability of the model. The third part is the fully connected layer, the fully connected layer transforms the output of the convolutional and pooling layers into the final prediction or feature representation.

Prediction model based on multi-model

The CNN-LSTM model is a combination of convolutional neural network and long and short-term memory model. In this paper, the historical data of mine water inflow is used as the input for the prediction of water inflow at the next moment, and the structure of the constructed CNN-LSTM mine water inflow prediction model is shown in Fig. 2. The CNN-LSTM model is generally divided into five parts: input layer, data preprocessing layer, CNN layer, LSTM layer and output layer.

Input layer: This paper takes the mine water inflow as the object of study, and before inputting the mine water inflow data, it is necessary to define the inflow sequence data $x = (x_1, x_2, \dots, x_n)$, divide the dataset into training set proportionally $x_{\text{train}} = (x_1, x_2, \dots, x_m)$ ($m \in (0, n)$) and test set $x_{\text{test}} = (x_{m+1}, x_{m+2}, \dots, x_n)$.

Data preprocessing layer: In order to ensure the reliability of data prediction and to improve the accuracy of the prediction results, a series of preprocessing should be carried out on the data. In order to avoid the excessive influence of some features on the model training, it is necessary to normalize the data by adjusting the range of values of different features to similar scales. A smoothness test and a white noise test are also required to determine whether the time series is smooth and white noise free.

CNN layer: The CNN model has significant advantages in feature extraction, firstly, feature extraction is performed in the convolutional layer of the CNN model, and after the convolutional operation, a non-activated linear function is usually applied in order to introduce non-linear properties, so as to facilitate the model to learn complex time-series features. The pooling layer can also be used in time-series data processing to reduce the size of the time dimension and reduce the computational complexity.

LSTM layer: The LSTM model has the ability to capture long-term dependencies. The pooled information is inputted into the LSTM layer. The memory units in the LSTM can transfer and update the information between

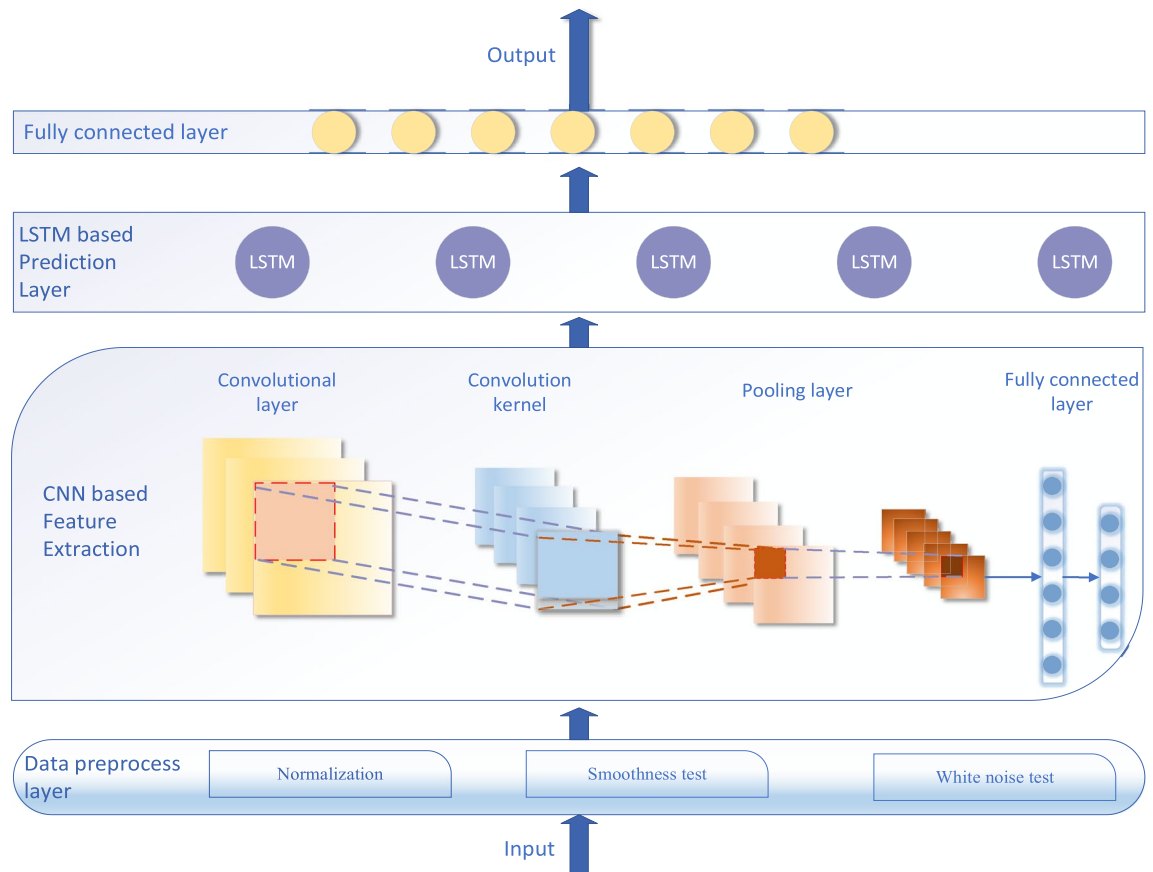


Figure 2. The CNN–LSTM model structure.

different time steps, which makes the neural network more suitable for time-series analysis and modeling. Layer-by-layer adaptive feature extraction gives the network a stronger learning ability and more effective access to the information in the time series.

Output layer: The output layer is to optimize and reduce the loss of the predicted value and the real value, and provide the result of m moments to the prediction module for the prediction of $m + 1$ moments, and finally get the prediction result of mine water inflow.

Both Transformer and LSTM are neural network models for processing time series data, and Transformer is a network architecture based on a self-attention mechanism. Due to the design of the self-attention mechanism, Transformer can focus on different positions of the input sequence at the same time to better capture the long-term dependencies of the time series⁴², and it is not constrained by the processing order of the time steps as LSTM⁴³. LSTM can selectively choose to update or forget memories through the design of gating units to capture important content in a time series⁴⁴, and LSTM can effectively control the problem of gradient descent or gradient explosion⁴⁵. So, in this paper, Transformer and LSTM are combined to utilize the advantages of both of them for time series forecasting, and the working principle of the hybrid model is shown in Fig. 3. In addition to having the standard model's input layer and data preprocessing mechanism, the structure of LSTM-Transformer consists of the following three main core parts:

LSTM layer: Firstly, the starting stage of the model is the LSTM layer, which processes the input sequences sequentially in time steps. At each time step, the LSTM updates its hidden state, which can effectively capture and retain long-term dependencies in the sequence and provide a basis for subsequent processing.

Transformer layer: Second, the output of the LSTM is used as an input to the Transformer, which captures the intrinsic relationships of the individual time steps in the sequence through a self-attentive mechanism. This step not only preserves the sequential information of the time series, but also enhances the model's ability to deal with complex patterns, such as periodic or seasonal features.

Output layer: Finally, the output of the Transformer can be passed to one or more fully connected layers to generate the final forecast output. In time series forecasting, this is usually the value at some point in the future.

Performance evaluation

In order to further compare the advantages and disadvantages between different models, this paper establishes an evaluation index system for model fitting effectiveness, which includes RMSE, MAE and R^2 .

RMSE is the root mean square error, which represents the sample standard deviation of the difference between predicted and observed values. When doing nonlinear fitting, the smaller the value of RMSE, the smaller the difference between the predicted and actual values of the model, and the more accurate the prediction of the model.

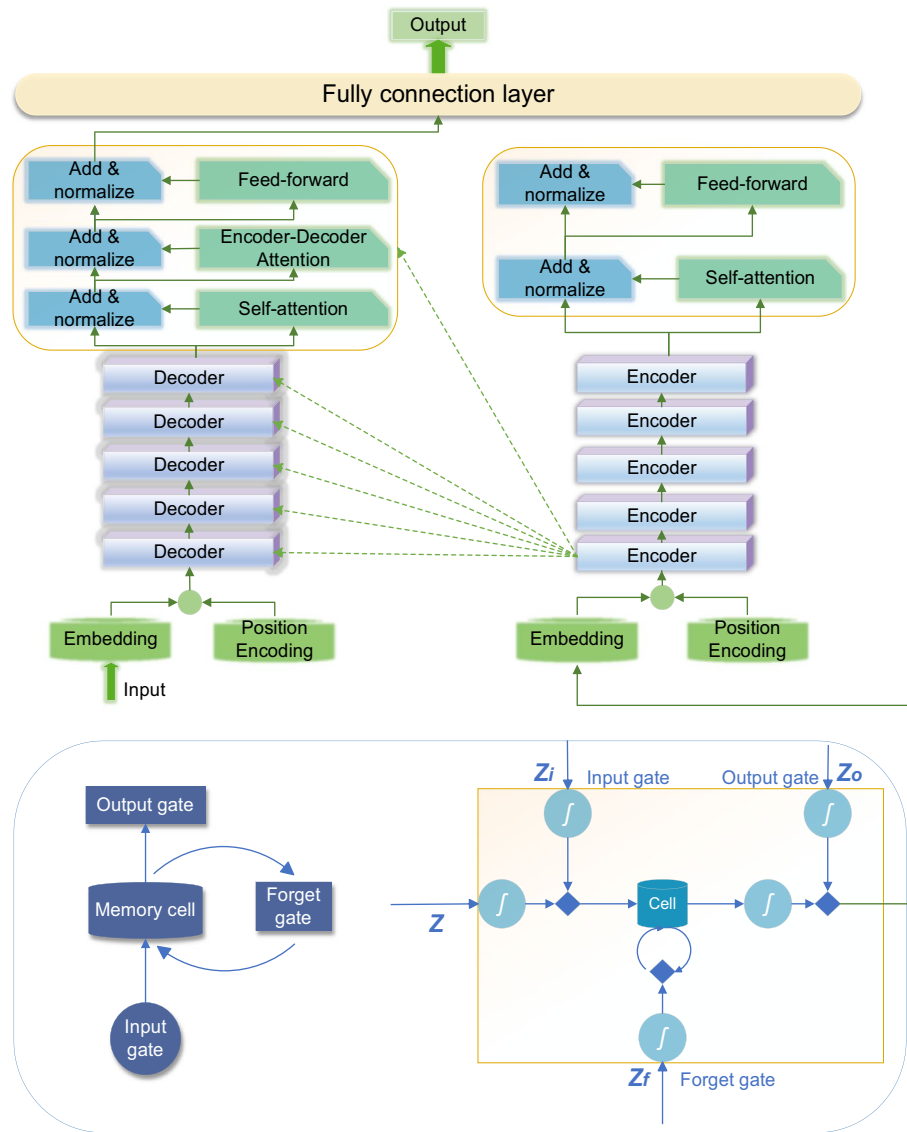


Figure 3. The LSTM-Transformer model structure.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \tag{1}$$

MAE is the Mean Absolute Error, which represents the average of the absolute errors between the predicted and real values. MAE assigns the same weight to all errors and does not increase significantly due to individually large errors.

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \tag{2}$$

Goodness of fit refers to the degree of fit of the regression curve to the predicted value, the coefficient of determination is a measure of the goodness of fit statistics, in general, the coefficient of determination is closer to 1 the better the fit.

$$R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{3}$$

n indicates the sample size, y_i indicates the sample value, \bar{y} is average value, \hat{y}_i is the predicted value.

Experiment validation

In order to verify the effectiveness of the above mentioned the LSTM-Transformer deep learning neural network for mine inflow prediction, it is compared with other different time series prediction algorithms (CNN, LSTM, Transformer, CNN-LSTM) as follows. In this paper, the water inflow monitoring data of Baotailong mine in Qixhetai, Heilongjiang Province is used as experimental data, and the five algorithms are verified to compare the advantages and disadvantages of these five algorithms, and the experimental results are compared and analyzed.

Data collection

The inflow data used in this paper are derived from field monitoring. In this paper, the water bin level method⁴⁶ was chosen to measure the surge water of Baotailong mine in Qitaihe, Heilongjiang. The water level method requires calculating the area of the free water surface of the temporary water silo in the downstream channel of the first mining face, recording the water level at the time of stopping the pump, the time of stopping the pump, and the water level after stopping the pump for a period of time, and the principle of the calculation is shown below.

$$Q = (H_2 - H_1) \times F \times 3600/t \quad (4)$$

Q-inflow water volume, unit m³/h. H₁-Water level of the water bin at the time of stopping the pump, unit m. H₂-Pump stopping time t silo rising water level, unit m. F-Area of free water surface in the silo, unit m². t-The time required for the water level of the silo to rise from H₂ to H₁, unit s.

The basic information such as springs points, surge volume and mine temperature need to be recorded after the measurement is finished, so as to facilitate the subsequent data integration and processing. The gushing water observation ledger is shown in Table 1, with the gushing water observation on January 17, 2022 as an example.

The supply water source of Heilongjiang Baotailong mine is mainly atmospheric precipitation, and the mine is a bedrock fissure water-filled deposit, with atmospheric precipitation directly recharging the groundwater, and the hydrogeological type is II type 1, meteorological water-filled deposit. In order to fully understand the water-rich nature of aquifers, old kiln waterlogged area, water-rich nature of trap columns and other mine hydrogeological structures in the Baotailong mine area, the transient electromagnetic detection method is used to detect the 25# layer of the left piece of the working face of the seven wells in the return airway and the roadway, and the results are shown in Fig. 4.

Figure 4 shows the proposed cross-section of the detected apparent resistivity of the roof of the roadway in the left working face of the 25# layer of the NO.7 well detected with the corresponding color scale. Different colors in the graph indicate different resistivity values, which are regularly distributed from cool to warm. Therefore, the cross section filled with different colors can be more intuitively judged the distribution of the resistivity of the roof plate in the detection area.

Normally the resistivity of a rock does not change under the same conditions. But if the rock has fissures or is filled with water, then the resistivity of the rock will change greatly. As shown in Fig. 4, we can find the difference of apparent resistivity, which indicates that there is a water anomaly in the area of about 90 m depth of the top plate of the left working face of layer 25#, and there is a risk of water inflow abnormality.

There were 245 monitoring data collected from 2012.1.10 to 2022.1.17 on the water anomaly in the 25# layer of the NO.7 well at the Heilongjiang Baotailong Mine. In the machine learning model, the data set is generally divided into 2 categories: one is the training set, and the other is the test set; the model is trained using the data of the training set in order to obtain the best parameters, and then the model is validated on the test set. The generalization ability of the model is verified on the test set. The division ratio of the training and test sets is crucial for model performance optimization. In order to accurately evaluate the model performance, different data division ratios are used in this study for cross-validation. The performance of the model under different ratios is described below and shown in detail in Fig. 5. Through this approach, we are able to understand more comprehensively the specific impact of the data division ratio on the model results, and guide us to achieve the optimal model performance.

Different models do not perform the same on different ratios of training sets, and different model structures have different adaptations to different features. Some models are better suited to capture specific types of patterns and associations, so these may be more apparent in specific data. After testing on various training sets with different ratios, the results show that the LSTM-Transformer model performs best at all ratios. Meanwhile, the Transformer, LSTM and CNN-LSTM models achieve optimal performance with a 6:4 data division ratio. The

Discharge time		2022.1.7		Outflow point location	NO.7 mine 25#layer left one working face return airway roof plate	
Outflow layer		Roof				
Observation method	Water bin level method	Water yield (m ³ /h)	Start	195.6 m ³ /h	Type of water quality	Fissure karst water
			Maximum	205.9 m ³ /h		
			Stable	200.9 m ³ /h		
Temperature	Temperatures		-21.6 °C			
	Roadway temperature		15 °C			
	water temperature		Normal water temperature			

Table 1. Mine water inflow observation ledger.

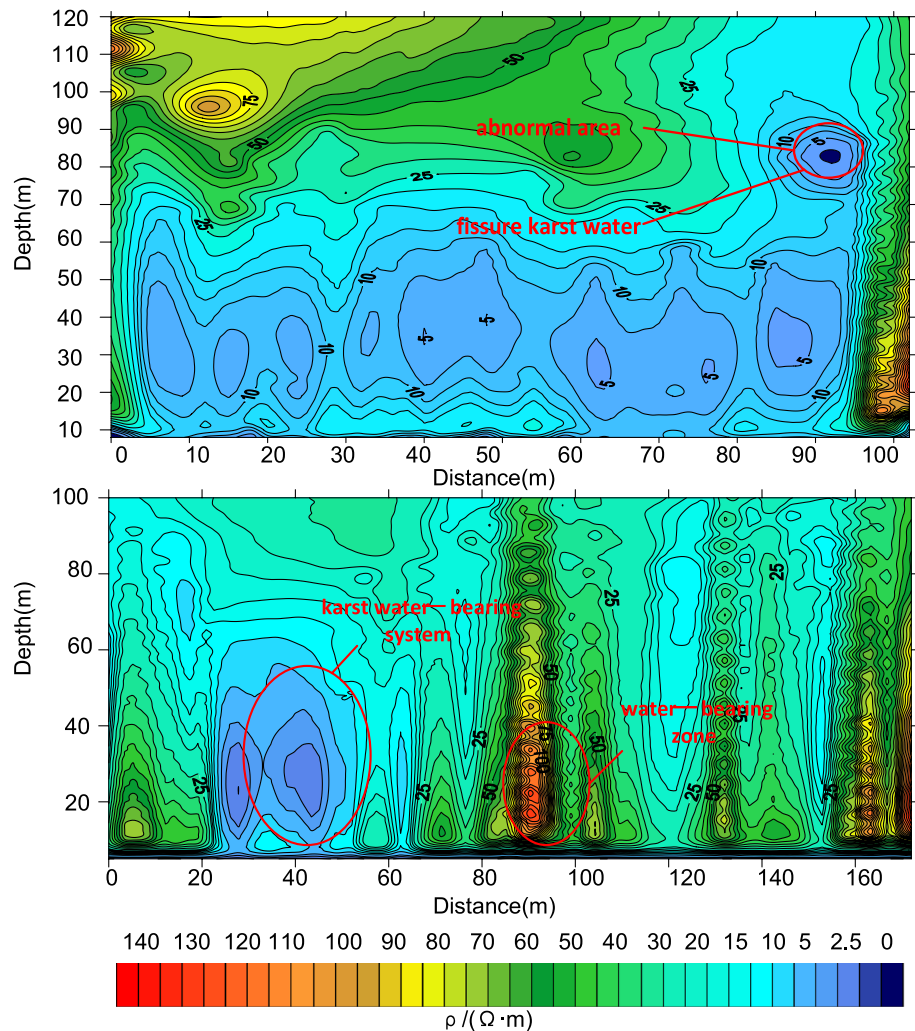


Figure 4. Proposed cross-section of resistivity of the roof in the return airway and the roadway of the left piece of working face of the 25# layer of the No.7 well.

CNN model, on the other hand, performs best with a 7:3 ratio. In this paper, we choose to show the prediction accuracy of the hybrid models LSTM-Transformer, CNN-LSTM and Transformer, which has the best performance ability among the single models, for different data division ratios are shown in Fig. 6.

To ensure the effectiveness of model training and to improve the performance of the final model, pre-processing the raw data is an indispensable step. The preprocessing process involves a range of techniques and methods aimed at improving the data quality. These technical tools reduce the noise in the data and enhance the feature representation of the data so that the model is able to learn from cleaner and more standardized data, which in turn improves the accuracy, robustness, and generalization of the model. A well-designed and implemented data preprocessing step before the in-depth training of the model lays a solid foundation for the construction of an efficient and reliable prediction model.

In order to scale the data with different features to a similar size or range, to avoid the influence of different features on the training of machine learning models, leading to a decrease in convergence speed and affecting the accuracy of model prediction. The maximum–minimum normalization method is usually chosen when predicting the time-series of a single feature because it maintains the relative order and time-series of the data, as shown in Eq. (5).

$$x_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (5)$$

Smoothness test is an important step in time series analysis, its purpose is to assess whether the time series data has smoothness, usually time series with smoothness is easier to analyze and predict. Augmented Dickey–Fuller test (ADF test) is a commonly used unit root test for determining whether a time series data has a unit root. If the p value of the ADF test is less than or equal to the level of significance (0.05) then the data is smooth, and vice versa, the data is non-smooth and has a unit root. The smoothness of the moving average line can help to determine the volatility of the data, to visualize the smoothness test, the EMA line that does not have

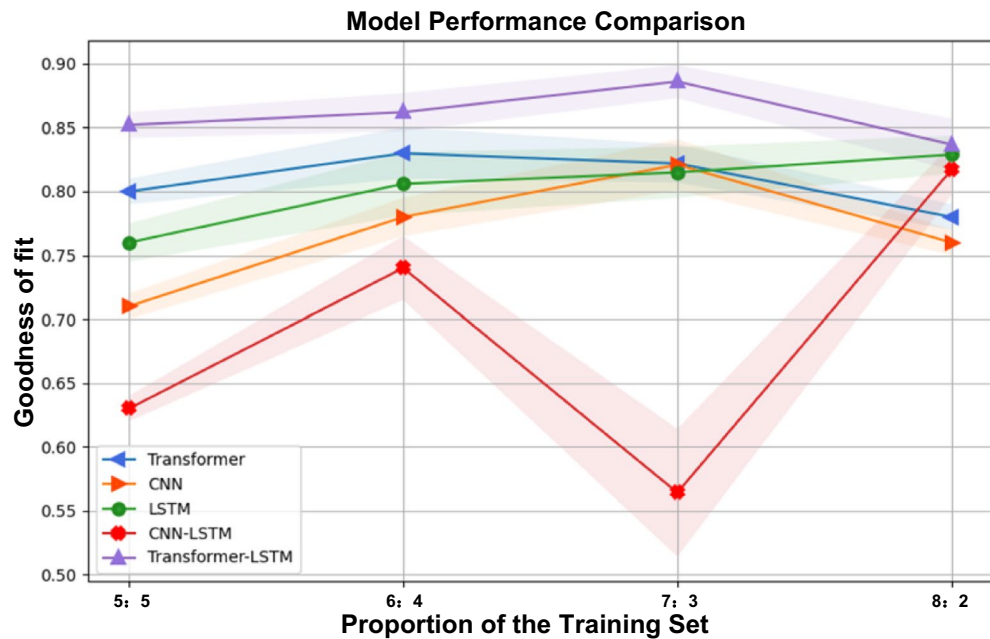


Figure 5. Model training results at different ratios.

significant spikes or deep valleys, but rather changes relatively gently over time, this indicates that the data is somewhat smooth after being treated by the EMA, as shown in Fig. 7.

The purpose of the white noise test is to test the randomness of time series data to determine whether there is a trend and whether there is a cyclical factor. Ljung-box is usually used in the processing of time-series data for white noise test. In Ljung-Box test, the judgment is based on the comparison of the test statistic Q with the level of significance, the Q statistic is less than the level of significance (0.05), then the data may exhibit the characteristics of white noise, there is no autocorrelation within the lagged order. The calculation of the Q statistic is shown in Eq. (6).

$$Q = n(n+2) \sum_{k=1}^q \frac{\hat{\rho}_k^2}{n-k} \quad (6)$$

n is the sample size. $\hat{\rho}_k^2$ is the correlation coefficient of the k th order lag of the sample.

Plotting the autocorrelation function (ACF) is a common visualization method used to make a check on whether the data has white noise or not. In ACF plot, we can observe the autocorrelation coefficients at different time lags and determine the nature of the data by their significant properties. The ACF plot is shown below in Fig. 8.

The results of the ADF test and Ljung-box on the mine inflow data show that the test statistic is infinitely close to 0, much less than the significance level of 0.5. In addition to the results of the EMA trend analysis and ACF visualization, this suggests that the mine inflow time series data is smooth, non-noisy, and easily interpretable.

In the case of too many training parameters and small training samples, the trained model is prone to overfitting phenomenon, which is manifested in the following: the model has a small loss function and high prediction accuracy on the training data, but has a larger loss function and lower prediction accuracy on the test data⁴⁷. In order to alleviate the overfitting problem and enhance the generalization ability of the model, regularization techniques are widely used. Among these, L2 regularization, also known as ridge regression or weight decay, is an effective form of regularization that penalizes model complexity by adding a regularization term to the loss function that is proportional to the square of the paradigm of the weight vector L2. The principle of the L2 regularization formula is shown in Eq. (7).

$$L_{new} = L_{original} + \frac{\lambda}{2} \sum_i w_i^2 \quad (7)$$

L_{new} is the new function after adding regularization. $L_{original}$ is the original loss function. λ is the regularization factor. w_i is the model weight. $\sum_i w_i^2$ indicates the sum of squares of all weights.

Modeling parameter adjustment

Hyperparameter tuning is a key step in machine learning and deep learning for improving the performance and accuracy of a model. Commonly used methods for tuning parameters in machine learning include grid search, stochastic search, and Bayesian optimization. Grid Search is a technique to optimize the hyperparameters of a model by traversing a given grid of parameters⁴⁸. Grid search searches for parameters, i.e., within a specified

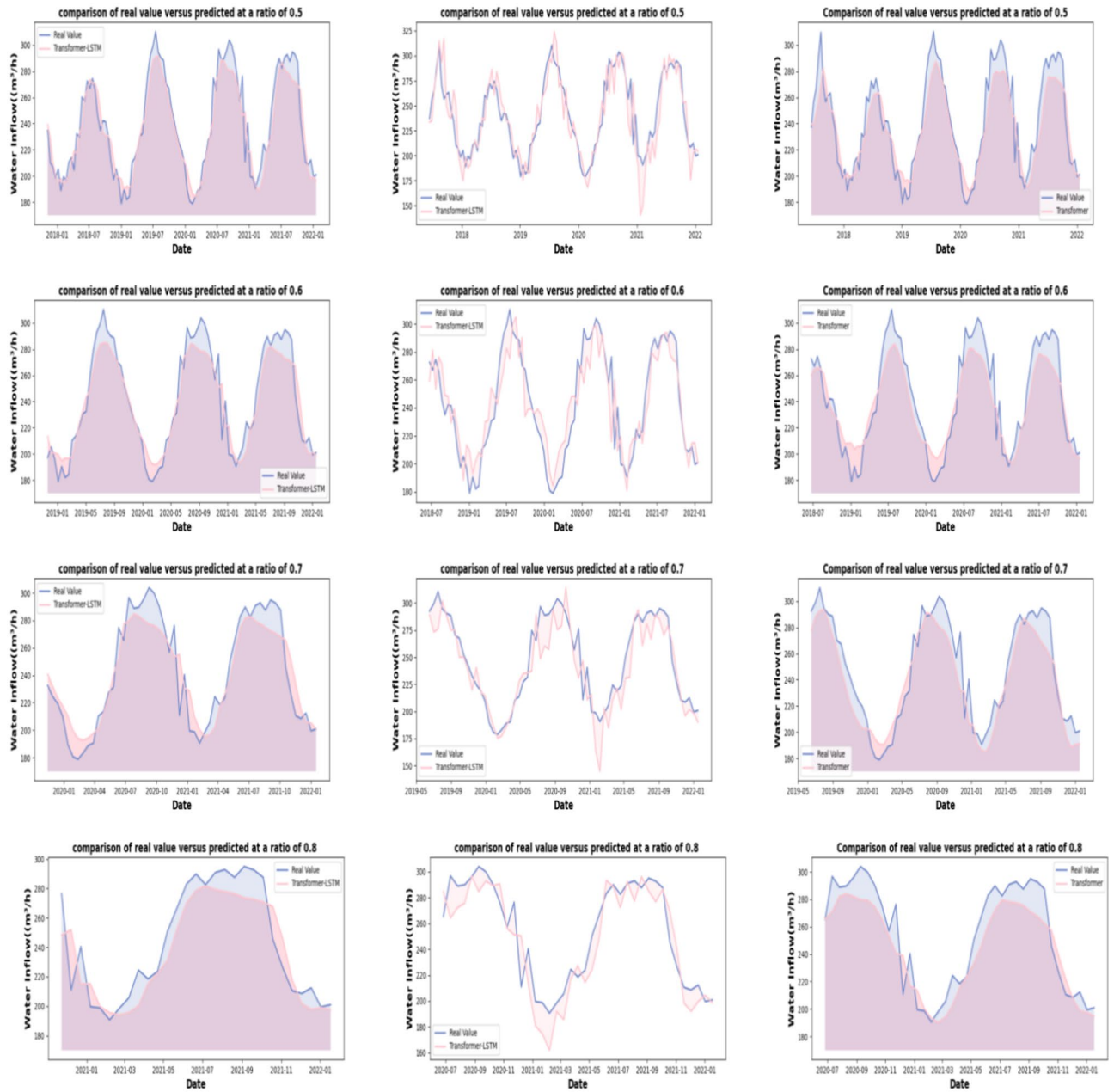


Figure 6. LSTM-Transformer (left) CNN-LSTM (center) Transformer (right) at different ratios.

range of parameters, adjusts the parameters sequentially by step size, uses the adjusted parameters to train the learner, and finds the parameter with the highest accuracy on the validation set from all the parameters, which is actually a training and comparison process. However, grid search uses a fixed step size to explore the hyperparameters, which may result in missing the optimal value, especially when the optimal value is located between grid points. Random search is a parameter optimization technique which, in contrast to grid search, uses a strategy of randomly selecting parameter values for the optimization of hyperparameters. Instead of trying every possible combination of parameters, random search tries to randomly sample a fixed number of parameter combinations from the distribution of parameters. Afterwards, the performance of each set of parameters is recorded and the parameter combination with the best evaluation results is selected.

Bayesian optimization uses probabilistic models to predict the performance of different hyperparameter combinations. This approach is more efficient than traditional grid or random searches because it can use previous evaluation results to guide the search process and locate optimal or near-optimal hyperparameter combinations more quickly. Therefore, in this work, a combination of random search and Bayesian optimization is chosen to tune the hyperparameters of the model, initially using random search for fast exploration in the hyperparameter space, which helps to identify better hyperparameter regions, and then using the data collected during the random search phase as a starting point for Bayesian optimization. The combination of random search and Bayesian optimization takes advantage of the fast exploration of random search over a large area and the ability of Bayesian

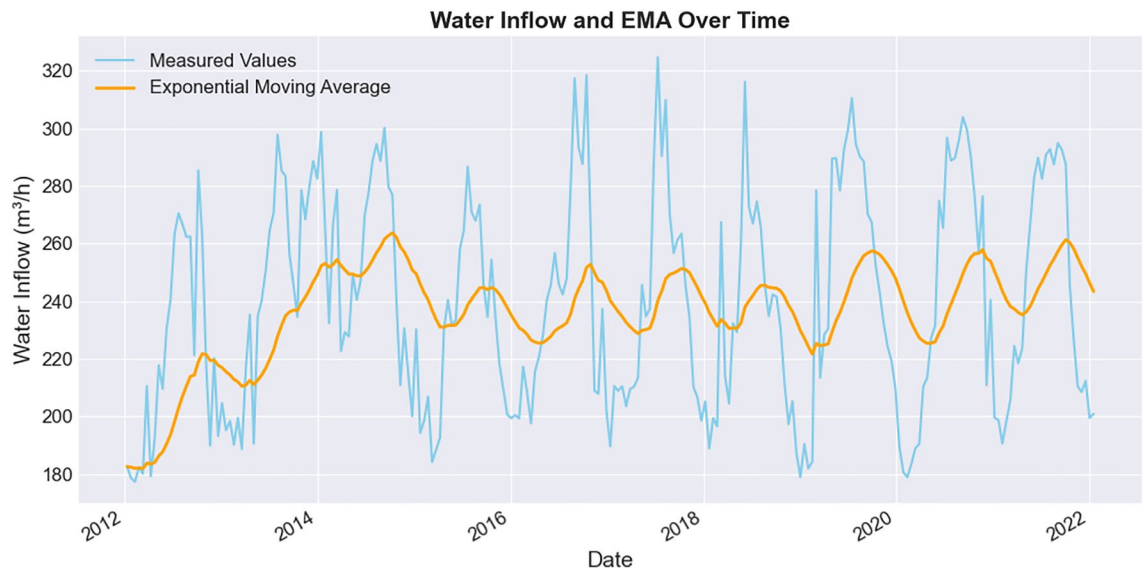


Figure 7. Water inflow and EMA over time.

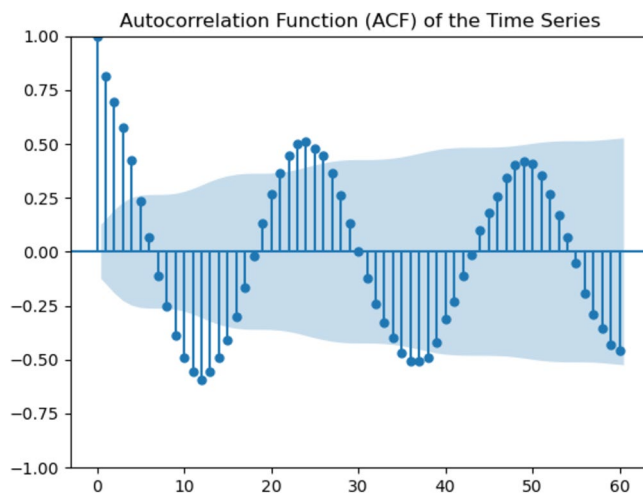


Figure 8. Autocorrelation function of the time series.

optimization to perform fine tuning over a local area. The hyperparameter settings for LSTM-Transformer are shown in Table 2.

Prediction performance

After determining the best network structure for each model best in the current assignment, the current model was trained using the training set until the model converged, a test set was used for evaluation and the predicted results and real values of different models are shown in Fig. 9. From the figure, it can be observed that the results of the LSTM-Transformer model on the prediction of mine water inflow and the real trend of water inflow are in good agreement, most of the predicted values do not differ much from the real values, and there is a certain deviation only in the prediction at the turning point and the prediction of some peaks, but the deviation is not significant. The fitted curves basically fluctuate up and down in a small range around the real data curves, and the trend of the predicted values is also basically consistent with the trend of the real influx. The goodness of fit between the real data and the predicted data indicated that the LSTM-Transformer model could capture 88.6% of the explained variance.

This result shows that the LSTM-Transformer model performs well in the task of mine inflow prediction, accurately reflecting the trend of inflow. The predictive ability of the model remains relatively robust even when key turning points and peaks in the data occur.

Comprehensive evaluation of the model

In order to better compare the performance of different models, RMSE, MAE and R^2 are selected as the evaluation indexes in this paper. Meanwhile, in order to validate the accuracy of the LSTM-Transformer model for the

Parameter	Setup parameter	Meaning
hidden_layers	2	LSTM number of hidden layers
LSTM Units	64	The number of neurons in both LSTM and Transformer hidden layers is 64
n_head	4	Number of attention heads in the transformer self-attention mechanism
d_model	2	Number of transformer encoder layers
dropout	0.1	Dropout reduces overfitting by randomly zeroing out the outputs of some neurons during training
batch_size	8	Volume of input data for a single batch
loss function	MSE	The loss function is used to calculate the difference between the real value and the predicted value to evaluate the predictive performance of the model
optimizer	Adam	Adam is an extended algorithm for stochastic gradient descent
weight_decay	0.1	Weight_decay is to reduce overfitting by adding a penalty term to the loss function
epochs	100	Update the number of iterations
learning_rate	0.001	The learning rate is the key hyperparameter used to adjust the rate of gradient descent

Table 2. The LSTM-Transformer model hyperparameters.

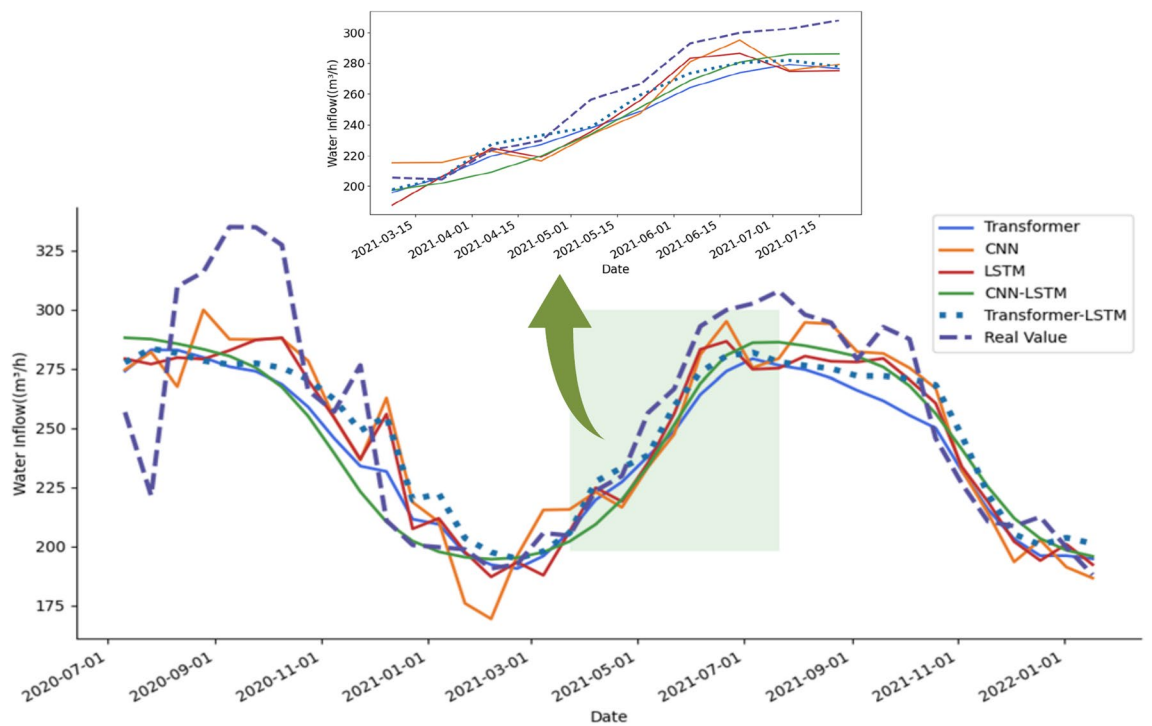


Figure 9. Comparison between predicted results and real values of different models.

prediction of mine inflow, a single deep learning model CNN, LSTM, Transformer, and a hybrid deep learning model CNN–LSTM are introduced as comparison models. It was found that LSTM-Transformer achieved better performance than other models for mine inflow prediction on time series datasets, and the average absolute error, root mean square error, and goodness of fit of the LSTM-Transformer model on the test set were improved compared to the single deep learning model CNN, LSTM, and Transformer hybrid deep learning model CNN–LSTM. Specifically, the average absolute error and root mean square error of the LSTM-Transformer model are reduced by 18.2% and 21.3%, and the goodness of fit is improved by 9.1% compared with CNN; the average absolute error and root mean square error are reduced by 7.4% and 13.3%, and the goodness of fit is improved by 6.4% compared with LSTM; the average absolute error and root mean square error are reduced by 17.1% and 13.3%, and the goodness of fit is improved by 5.9% compared with Transformer, the average absolute error and root-mean-square error are reduced by 17.1% and 13.3% respectively, and the goodness-of-fit is improved by 5.9%; and compared with the hybrid model CNN–LSTM, the average absolute error and root-mean-square error are reduced by 10.4% and 12.7% respectively, and the goodness-of-fit is improved by 8.3%. Comparison of prediction performance of different models is shown in Fig. 10.

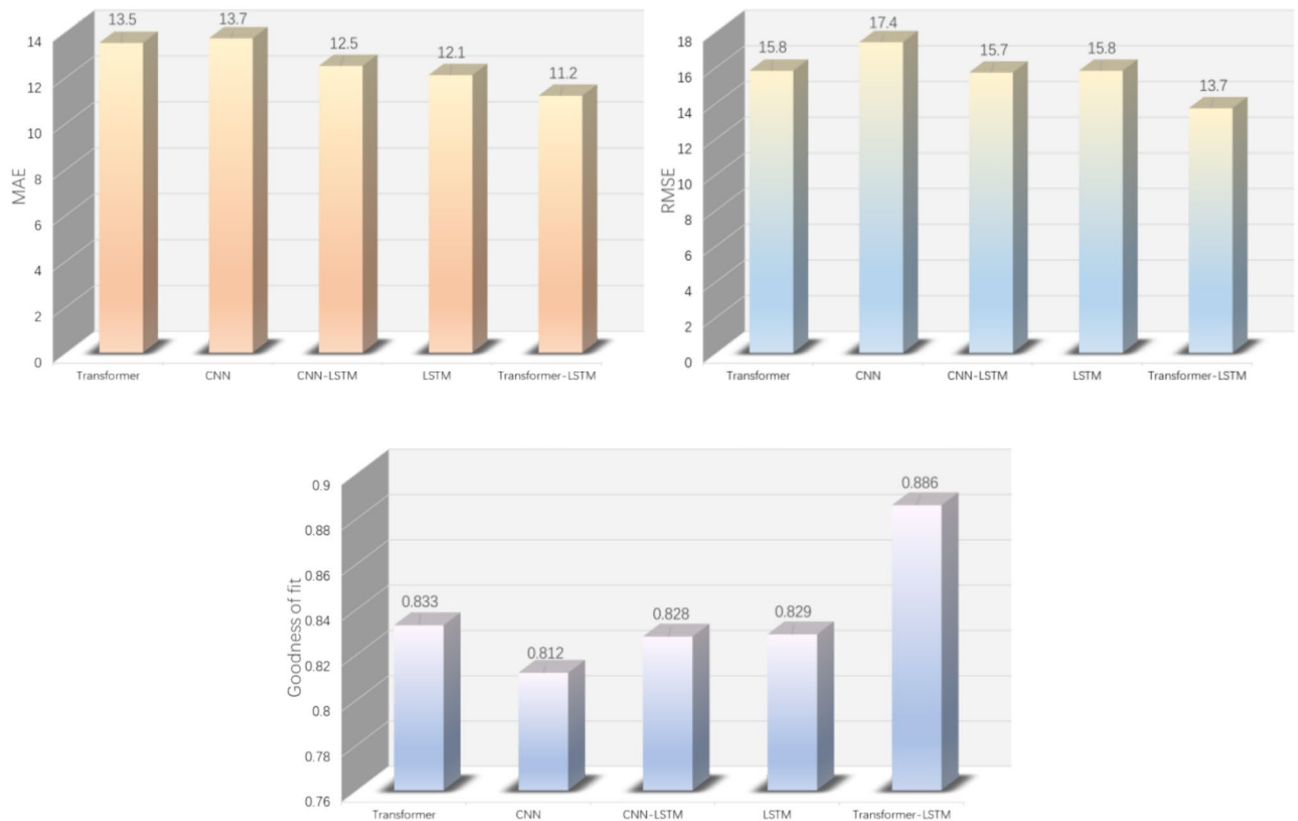


Figure 10. Comparison of prediction performance of different models.

Discussion

Firstly, in the data preprocessing stage, the smoothness test is a key step. By conducting a smoothness test on the raw data, noise and outliers in the data can be effectively removed, thereby simplifying the complexity of the model and improving the accuracy of the predictive model. Mine water inflow data is often influenced by various external factors, such as climate change and geological conditions, resulting in abrupt changes or outliers in the data that could adversely affect the performance of the predictive model. Through smoothness testing, we can effectively filter out these anomalies while retaining the main trends in the data, making the model more stable and reliable.

Secondly, regarding data set partitioning, this study adopts a strategy that comprehensively considers the rational division of training and test sets. A reasonable data set partitioning helps make full use of limited data resources and effectively prevents the problem of model overfitting. Overfitting is a common challenge in small sample datasets; if the dataset is not properly divided, the model may perform well on the training set but poorly on the test set. By adopting a 7:3 training to test set ratio, we can dynamically adjust the model parameters during training and effectively validate the model's generalization ability on unseen data.

At the same time, to enhance the model's performance and more effectively capture the trends and patterns in time series data, this study employs a series of rigorous preprocessing steps. Through continuous empirical testing, we ultimately chose to apply L2 regularization to the loss function to optimize the model's generalization performance and robustness. Additionally, we introduced the Adam optimizer to accelerate the model's convergence process and automatically adjust the learning rate during training, which helps address gradient differences among various features. To comprehensively evaluate the model's performance, a series of validation experiments were conducted to explore the impact of different training set ratios on the prediction results. In this process, it was found that when the training to test set ratio was 7:3, the Transformer-LSTM model achieved a good fit of 0.886 on the training set. This indicates that the model achieved a high degree of fit on the training set while maintaining good generalization performance.

In the instance validation section, by comparing the performance of different models in predicting mine water inflow, it was found that the Transformer-LSTM model performed the best on the training set. The seasonal atmospheric precipitation influences mine water inflow, and such seasonal time series usually contain complex long-term dependencies. Therefore, we chose to combine the LSTM and Transformer models to fully utilize their advantages in time modeling and feature extraction. LSTM helps capture potential trends and periodicity in seasonal water inflow data by retaining long-term memory in the data. The output of LSTM is used as the input to the Transformer, allowing the Transformer to better capture complex relationships between different time points in the time series and learn important features in the sequence.

For a single deep learning model, compared to CNN and LSTM models, the optimal performance of the Transformer model is based on two aspects: firstly, the self-attention mechanism in the Transformer allows the model to consider information from different positions in the sequence simultaneously, without the need to process the sequence sequentially like traditional RNNs or LSTMs. Secondly, the scalability of the Transformer model enables it to easily adapt to sequences of different lengths, making it suitable for various time prediction tasks.

The collection of mine water inflow data is constrained by mine conditions and economic status, making data collection challenging. Simultaneously, using the CNN–LSTM model faces multiple hyperparameters that need adjustment, including convolution kernel size and the number of LSTM units. With a smaller sample size, the model's fit to the data is lower, making it more difficult to find suitable hyperparameter combinations, which may be one reason for the CNN–LSTM model's poor performance in this water inflow prediction task.

Finally, to enhance model performance, we chose to combine random search and Bayesian optimization for hyperparameter tuning. This combined method leverages the advantages of both optimization techniques, further enhancing the model's generalization ability. In practice, random search is first conducted in a larger parameter space to initially filter out a batch of well-performing parameter combinations. Subsequently, Bayesian optimization is used to perform a more refined search around these initially filtered parameter combinations, further optimizing the model's hyperparameters. This approach significantly improves the model's generalization ability, effectively enhancing its performance on both training and test sets. This combined method not only improves the model's prediction accuracy but also enhances the model's adaptability to unseen data, validating its reliability and stability in practical applications. In the task of predicting water inflow at burst points, using the combined method of random search and Bayesian optimization for hyperparameter tuning successfully improved the performance of the Transformer–LSTM model, enabling it to more accurately capture trends and patterns in time series data.

Conclusions

This study systematically explores how to achieve higher accuracy in water inflow prediction from three key perspectives: small sample prediction of water inflow at burst points, the nonlinear variation characteristics of water inflow data at burst points, and how to enhance the generalization ability of water inflow prediction models. Specifically, the study proposes a mine water inflow prediction model based on the Transformer–LSTM structure, which demonstrated significant advantages in experiments. Compared to single neural network models such as CNN, LSTM, Transformer, and hybrid neural network models like CNN–LSTM, the Transformer–LSTM model can not only effectively capture long-term dependencies in time series but also fully leverage the advantages of the Transformer in global dependency relationships, achieving more accurate and robust prediction results.

In the data preprocessing stage, the combined use of data smoothness testing and reasonable data set partitioning can significantly improve the performance of time series prediction models. Smoothness testing removes noise and outliers from the data, simplifying the model structure and improving prediction accuracy. Reasonable data set partitioning effectively utilizes the data and prevents overfitting, showing better potential in small sample prediction. The combined effect of these two methods makes the model exhibit higher robustness and reliability in practical applications.

The water inflow at mine burst points shows dynamic changes and some trend changes. The LSTM model is better at capturing such dynamic patterns and trends, so time series data is processed by the LSTM model before being input into the Transformer model. The self-attention mechanism in the Transformer model allows the input data to be computed in parallel, improving training efficiency and prediction accuracy. The Transformer–LSTM model provides a good approach for predicting dynamic time series.

By combining random search and Bayesian optimization, the model parameters can be precisely adjusted under limited data conditions, effectively dealing with complex time series data structures and patterns. This optimization not only improves the model's fit on the training set but also enhances the model's generalization ability on new data. Therefore, the improved Transformer–LSTM model performs excellently in experimental environments and demonstrates strong adaptability and prediction effectiveness in practical applications, providing an efficient and reliable prediction tool for mine safety management.

The water inflow prediction model proposed in this study improves prediction accuracy, providing a solid technical support and foundation for mine flood warning and post-disaster rescue work. However, this study still has some limitations. First, the scale and diversity of the dataset are limited, which may affect the model's generalization ability. Therefore, future work should focus on expanding the scale and diversity of the dataset to further improve the model's generalization performance and applicability. These improvements and challenges will help further perfect the water inflow prediction model, providing more reliable technical support and foundation for mine safety production and post-disaster rescue work.

Data availability

The data that support the findings of this study are available on request from the corresponding author, contact information of the author: shijunwei302@sdtbu.edu.cn upon reasonable request.

Received: 18 May 2024; Accepted: 5 August 2024

Published online: 07 August 2024

References

- Hu, W. Y. & Yan, L. Analysis and consideration on prediction problems of mine water inflow volume. *Coal Sci. Technol.* **44**(1), 13–18+38. <https://doi.org/10.13199/j.cnki.cst.2016.01.003> (2016).
- He, X. L., Pu, Z. G. & Ding, X. Improved methods for prediction of mine water inflow and determination of accuracy of results. *Coal Sci. Technol.* **48**(08), 229–236. <https://doi.org/10.13199/j.cnki.cst.2020.08.029> (2020).
- Liu, C. & Fan, Y. Emergent fractal energy landscape as the origin of stress-accelerated dynamics in amorphous solids. *Phys. Rev. Lett.* **127**, 215502. <https://doi.org/10.1103/PhysRevLett.127.215502> (2021).
- Wang, Y. & Fan, Y. Incident velocity induced nonmonotonic aging of vapor-deposited polymer glasses. *J. Phys. Chem. B* **124**, 5740–5745. <https://doi.org/10.1021/acs.jpcc.0c02335> (2020).
- Liu, C. *et al.* Unraveling the non-monotonic ageing of metallic glasses in the metastability–temperature space. *Comput. Mater. Sci.* **172**, 109347. <https://doi.org/10.1016/j.commatsci.2019.109347> (2020).
- Liu, C., Guan, P. & Fan, Y. Correlating defects density in metallic glasses with the distribution of inherent structures in potential energy landscape. *Acta Mater.* **161**, 295–301. <https://doi.org/10.1016/j.actamat.2018.09.021> (2018).
- Fan, Y., Iwashita, T. & Egami, T. Energy landscape-driven non-equilibrium evolution of inherent structure in disordered material. *Nat. Commun.* **8**, 15417. <https://doi.org/10.1038/ncomms15417> (2017).
- Singh, R. N. & Atkins, A. S. Application of idealised analytical techniques for prediction of mine water inflow. *Min. Sci. Technol.* **2**(2), 131–138. [https://doi.org/10.1016/S0167-9031\(85\)90346-9](https://doi.org/10.1016/S0167-9031(85)90346-9) (1985).
- Tian, G. H., Cui, Z. L. & Hu, T. C. Time series analysis and quantitative prediction of water mine inflow based on R/S analysis. *J. Ind. Miner. Process.* **50**(10), 1–5+9. <https://doi.org/10.16283/j.cnki.hgkwyjg.2021.10.001> (2021).
- Chen, J. F., Ying, C. & Ma, Y. Quantitative prediction of mine inflow based on Hurst exponent. *J. Saf. Coal Mines* **47**(12), 199–202. <https://doi.org/10.13347/j.cnki.mkaq.2016.12.054> (2019).
- Xie, D. W. & Shi, S. L. Mine water inrush prediction based on cloud model theory and Markov model. *J. Central South Univ. (Sci. Technol.)* **43**(06), 2308–2315 (2012).
- Wu, B., Bai, Z., Misra, A. & Fan, Y. Atomistic mechanism and probability determination of the cutting of Guinier–Preston zones by edge dislocations in dilute Al–Cu alloys. *Phys. Rev. Mater.* **4**, 020601. <https://doi.org/10.1103/physrevmaterials.4.020601> (2020).
- Fan, Y., Kushima, A., Yip, S. & Yildiz, B. Mechanism of void nucleation and growth in bcc Fe: Atomistic simulations at experimental time scales. *Phys. Rev. Lett.* **106**, 125501. <https://doi.org/10.1103/PhysRevLett.106.125501> (2011).
- Jiang, L. *et al.* Deformation mechanisms in crystalline–amorphous high-entropy composite multilayers. *Mater. Sci. Eng. A* **848**, 143144. <https://doi.org/10.1016/j.msea.2022.143144> (2022).
- He, M., Yang, Y., Gao, F. & Fan, Y. Stress sensitivity origin of extended defects production under coupled irradiation and mechanical loading. *Acta Mater.* **248**, 118758. <https://doi.org/10.1016/j.actamat.2023.118758> (2023).
- Fan, Y. *et al.* Mapping strain rate dependence of dislocation–defect interactions by atomistic simulations. *Proc. Natl. Acad. Sci.* **110**, 17756–17761. <https://doi.org/10.1073/pnas.1310036110> (2013).
- Fan, Y., Kushima, A. & Yildiz, B. Unfaulting mechanism of trapped self-interstitial atom clusters in bcc Fe: A kinetic study based on the potential energy landscape. *Phys. Rev. B* **81**, 104102. <https://doi.org/10.1103/PhysRevB.81.104102> (2010).
- Fan, Y. *et al.* Editorial: Modeling of structural and chemical disorders: From metallic glasses to high entropy alloys. *Front. Mater.* <https://doi.org/10.3389/fmats.2022.1006726> (2022).
- Luo, Q. B., Bin, B. F. & Mao, X. G. Prediction and analysis of mine water inflow based on numerical simulation method. *J. Northwest Univ. (Nat. Sci. Edition)* **52**(06), 1100–1110. <https://doi.org/10.16152/j.cnki.xdxzbz.2022-06-014> (2022).
- Wen, W. F. & Cao, L. J. Comparative analysis for analogue method and analytical method in prediction of water inflow in some coal mine. *China Coal* **37**(07), 38–40. <https://doi.org/10.19880/j.cnki.ccm.2011.07.011> (2011).
- Duan, J. J., Xu, H. J. & Wang, Z. H. Correlational analysis method applied to prediction of mine water inflow quantity. *Coal Sci. Technol.* **41**(06), 114–116+76. <https://doi.org/10.13199/j.cnki.cst.2013.06.82.duanjj.032> (2013).
- Liu, J., Wang, Q. M. & Yang, J. Mine inflow simulation and dynamic prediction based on visual Modflow. *Saf. Coal Mines* **49**(03), 190–193. <https://doi.org/10.13347/j.cnki.mkaq.2018.03.051> (2018).
- Hou, E. K., Xi, H. Q. & Wen, Q. Prediction of water inflow volume in the coal mining workforce below the concealed fire area based on GMS. *J. Saf. Environ.* **22**(05), 2482–2492. <https://doi.org/10.13637/j.issn.1009-6094.2021.0852> (2022).
- Cheng, X. G., Qiao, W. & Li, L. Model of mining-induced fracture stress–seepage coupling in coal seam over-burden and prediction of mine inflow. *J. China Coal Soc.* **45**(08), 2890–2900. <https://doi.org/10.13225/j.cnki.jccs.2019.0651> (2020).
- Tian, L. *et al.* Identifying flow defects in amorphous alloys using machine learning outlier detection methods. *Scr. Mater.* **186**, 185–189. <https://doi.org/10.1016/j.scriptamat.2020.05.038> (2020).
- Wang, Y. *et al.* Predicting the energetics and kinetics of Cr atoms in Fe–Ni–Cr alloys via physics-based machine learning. *Scr. Mater.* **205**, 114177. <https://doi.org/10.1016/j.scriptamat.2021.114177> (2021).
- Liu, C. *et al.* Concurrent prediction of metallic glasses’ global energy and internal structural heterogeneity by interpretable machine learning. *Acta Mater.* <https://doi.org/10.1016/j.actamat.2023.119281> (2023).
- Wang, Y. *et al.* Nonmonotonic effect of chemical heterogeneity on interfacial crack growth at high-angle grain boundaries in Fe–Ni–Cr alloys. *Phys. Rev. Mater.* **7**, 073606. <https://doi.org/10.1103/physrevmaterials.7.073606> (2023).
- Liu, X. D. & Pan, G. Y. Prediction of mine water inflow based on three time series models. *Min. Saf. Environ. Prot.* **49**(02), 91–95+101. <https://doi.org/10.19835/j.issn.1008-4495.2022.02.016> (2022).
- Dong, D. L., Zhang, L. Q. & Zhang, E. Y. A rapid identification model of mine water inrush based on PSO–XGBoost. *Coal Sci. Technol.* **51**(07), 72–82. <https://doi.org/10.13199/j.cnki.cst.2023-0446> (2023).
- Li, Z. L., Xing, J. S. & Jin, H. M. Prediction of mine water inflow based on CEEMD–GRU model. *J. Beijing Univ. Technol.* **47**(08), 904–911. <https://doi.org/10.11936/bjtxb2020120022> (2021).
- Han, J. *et al.* Online estimation of the heat flux during turning using long short-term memory based encoder–decoder. *Case Stud. Therm. Eng.* **26**, 101002. <https://doi.org/10.1016/j.csite.2021.101002> (2021).
- Rius, A., Ruisanchez, I., Callao, M. P. & Rius, F. X. Reliability of analytical systems: Use of control charts, time series models and recurrent neural networks (RNN). *Chemom. Intell. Lab. Syst.* **40**(1), 1–18. [https://doi.org/10.1016/s0169-7439\(97\)00085-3](https://doi.org/10.1016/s0169-7439(97)00085-3) (1998).
- Yin, Q. X., Zhang, R. X. & Shao, X. CNN and RNN mixed model for image classification. In *MATEC web of conferences* 02001, Vol. 277 (2019). <https://doi.org/10.1051/mateconf/201927702001>.
- Rehmer, A. & Kroll, A. On the vanishing and exploding gradient problem in gated recurrent units. *IFAC-Papers Online* **53**(2), 1243–1248. <https://doi.org/10.1016/j.ifacol.2020.12.1342> (2020).
- Landi, F., Baraldi, L., Cornia, M. & Cucchiara, R. Working memory connections for LSTM. *Neural Netw.* **144**, 334–341. <https://doi.org/10.1016/j.neunet.2021.08.030> (2021).
- ArunKumar, K. E., Kalaga, D. V., Kumar, C. M. S., Kawaji, M. & Brenza, T. M. Forecasting of COVID-19 using deep layer recurrent neural networks (RNNs) with gated recurrent units (GRUs) and long short-term memory (LSTM) cells. *Chaos Solitons Fractals* **146**, 110861. <https://doi.org/10.1016/j.chaos.2021.110861> (2021).
- Zhang, J. & Li, S. Air quality index forecast in Beijing based on CNN–LSTM multi-model. *Chemosphere* **308**, 136180. <https://doi.org/10.1016/j.chemosphere.2022.136180> (2022).
- Wang, D., Zhao, L., Hao, T. & Du, Y. Multiple sequence long and short memory network model for corner gas concentration prediction on coal mine workings. *ACS Omega* **7**(42), 37980–37987. <https://doi.org/10.1021/acsomega.2c05188> (2022).

40. Vaswani, A. *et al.* Attention is all you need. *Adv. Neural Inf. Process. Syst.* <https://doi.org/10.48550/arXiv.1706.03762> (2017).
41. Zhao, B. D., Lu, H. Z. & Chen, S. F. Convolutional neural networks for time series classification. *J. Syst. Eng. Electron.* **28**(1), 162–169. <https://doi.org/10.21629/jsee.2017.01.18> (2017).
42. Khalil, M. M. Y. *et al.* Cross-modality representation learning from transformer for hashtag prediction. *J. Big Data* **10**, 148. <https://doi.org/10.1186/s40537-023-00824-2> (2023).
43. Elsheikh, A. H. *et al.* Productivity forecasting of solar distiller integrated with evacuated tubes and external condenser using artificial intelligence model and moth-flame optimizer. *Case Stud. Therm. Eng.* **28**, 101671. <https://doi.org/10.1016/j.csite.2021.101671> (2021).
44. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**(8), 1735–1780. [https://doi.org/10.1162/neco.9\(8\).1735](https://doi.org/10.1162/neco.9(8).1735) (1997).
45. Mu, Y., Wang, M., Zheng, X. & Gao, H. An improved LSTM-Seq2Seq-based forecasting method for electricity load. *Front. Energy Res.* **10**, 1093667. <https://doi.org/10.3389/fenrg.2022.1093667> (2023).
46. Yang, X. H. Multi-level automation drainage control system for mine. *Saf. Coal Mines* **45**(08), 122–125. <https://doi.org/10.13347/j.cnki.mkaq.2014.08.036> (2014).
47. Xu, D. G., Wang, L. & Li, F. Review of typical object detection algorithms for deep learning [J]. *Comput. Eng. App.* **57**(08), 10–25. <https://doi.org/10.3778/j.issn.1002-8331.2012-0449> (2021).
48. Ogunsanya, M. *et al.* Grid search hyperparameter tuning in additive manufacturing processes. *Manuf. Lett.* <https://doi.org/10.1016/j.mfglet.2023.08.056> (2023).

Author contributions

S.J. designed the experimental methodology and wrote the first draft of the paper; W.S. and S.J. prepared all figures; W.S. and Q.P. applied statistical, mathematical, computational, or other formal techniques to analyze and synthesize the study data. All authors reviewed the results and approved the final version of the manuscript.

Funding

This research was supported by the Ministry of Education's Humanities and Social Sciences Research Youth Fund Project (21YJCZH135): Research on the evaluation mechanism and control system of major disaster risk in coal mine based on complex system theory.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.S.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024