

ORIGINAL ARTICLE

Prediction of longwall mining-induced stress in roof rock using LSTM neural network and transfer learning method

Changkun Qin^{1,2} | Wusheng Zhao^{1,2}  | Kun Zhong^{1,2} | Weizhong Chen¹

¹State Key Laboratory of Geomechanics and Geotechnical Engineering, Institute of Rock and Soil Mechanics, Chinese Academy of Sciences, Wuhan, China

²University of Chinese Academy of Sciences, Beijing, China

Correspondence

Wusheng Zhao, State Key Laboratory of Geomechanics and Geotechnical Engineering, Institute of Rock and Soil Mechanics, Chinese Academy of Sciences, Wuhan 430071, China.
Email: wszhao@whrsm.ac.cn

Funding information

National Natural Science Foundation of China, Grant/Award Number: 51991393 and 52079134

Abstract

Real-time monitoring of three-dimensional stress in the field is an effective approach to detect evolving stress in roof rock and to evaluate rock bursts risk. However, the sensors or data transmission cables may be damaged due to the volatile environment found in coal mines, which can lead to the loss of relevant monitoring data, and some critical information for rock burst prediction may be missed entirely. A number of methods that use historical data to predict missing data or future structural states have been proposed. However, the performance of these methods is poor when the training data are insufficient owing to lack of data. To address this issue, a methodology framework is proposed to predict the mining-induced stress when some monitoring data are missing. The framework uses a long short-term memory neural network integrated with the transfer learning method. The proposed method can transfer the knowledge learned from complete monitored data of adjacent sensor to target sensor to boost forecasting. A case study has been conducted to evaluate the method. The results show that the developed model can significantly improve the prediction performance for the target domain, which can be improved further by increasing the size of the target domain training data available.

KEYWORDS

data missing, LSTM, mining-induced stress, monitoring, stress prediction, transfer learning

1 | INTRODUCTION

Rock bursts are a common dynamic disaster in coal mines, and they crucially affect the safety, economics, and efficiency of mining operations. The mitigation and control of rock bursts is challenging owing to their violent, unpredictable characteristics.^{1,2} Rock bursts are characterized by the sudden release of elastic strain energy in rock and coal during mining or roadway excavation. The mining-induced redistributed high-stress regions around

surrounding rocks are crucial for the evaluation of rock bursts risk, particularly when using the longwall mining method. The excavation is surrounded by thick layers of hard, intact rock capable of storing high levels of strain energy.³ Therefore, the determination of mining-induced stresses is essential for the evaluation of rock bursts risk in coal mines. Several approaches have been employed to assess and calculate the stress distribution around the extraction field, such as empirical and analytical,^{4–9} numerical simulation,^{10–15} and field monitoring^{16–20} approaches.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Energy Science & Engineering* published by Society of Chemical Industry and John Wiley & Sons Ltd.

Three-dimensional (3D) stress monitoring is a crucial and effective approach for revealing the stress state, predicting rock burst events, and adopting countermeasures in coal mines.²¹ Generally, long-term stress monitoring is helpful for analyzing and predicting the evolution of stresses in surrounding rock. Therefore, the accuracy and reliability of early disaster warnings rely on the quantity and quality of the in situ monitoring data.²² However, the sensors and data transmission cables are often damaged in the volatile environment in underground mines, which leads to the loss of monitoring data.^{19,23,24} The loss of data greatly affects the early warning of surrounding rock instability. Therefore, it is essential to reconstruct the missing data and predict the stress state in the following few days to enable evaluations of safety, reliability analyses, and real-time early warnings of disasters. The reconstruction and completion of missing data in field monitoring can be converted to a time series prediction task, and the process has been studied widely. Numerous methods of model-based and data-driven approaches have been proposed in the structure health monitoring (SHM) field. These approaches include time series state space modeling, autoregressive modeling, Gaussian process modeling, Bayesian multi-task learning methodologies, support vector regression methodologies, and long short-term memory (LSTM) neural networks.^{25–27} However, the performance of these methods is highly dependent on the quantity of training data available. When monitoring data are lost due to the failure of sensors or optical fiber cables, sufficient training data may not be available to train the models. Conventional methods do not perform well with insufficient data. Therefore, in this situation, how to improve the prediction accuracy is a critical issue.

To this end, an LSTM neural network based on transfer learning for stress prediction is proposed. The LSTM neural network is a special and advanced recurrent neural network (RNN) that is capable of learning the long and short time series patterns from historical data. This type of networks has been proved to exhibit excellent performance for addressing time series problems and has the best performance compared with the other algorithms.^{34,41} Recently, LSTM neural networks have been widely used and achieved great success in civil engineering, for example, stress-strain behavior of soils,⁴² tunneling-induced ground settlement,⁴³ performance of EPB shield tunneling,^{44,45} and seismic bearing capacity of foundations.⁴⁶ However, few studies have been performed to predict the mining-induced stress using LSTM neural networks. More importantly, insufficient training data limit the effective performance of such a network. Therefore, in this study, we integrate the LSTM neural network and the transfer learning method to predict the mining-induced stress

and improve the prediction accuracy. Transfer learning, on the other hand, is an important development in machine learning that aims to improve the prediction performance by transferring knowledge learned from source domains to target domains. The source domain refers to the domain with knowledge and more labeled data and is the object that is to be transferred. The target domain refers to a new but related domain with a small amount of labeled data and is the object to which the knowledge is transferred.^{28,29} To test and verify the proposed method, we selected the vertical stress data of monitored results in a coal mine as a case study. The results show that the developed model can significantly improve the prediction performance for the target domain, which can be improved further by increasing the size of the target domain training data available.

This paper is organized as follows. Section 2 presents a detailed analysis of a field case study of a stress monitoring scheme and the results of the study. The methodology framework of the proposed model, base theory of the LSTM network, and transfer learning are described in Section 3. In Section 4, a case study of a model application and experiments over two datasets of adjacent monitoring sections are described, and corresponding experimental results and analysis are presented. Section 5 concludes the paper.

2 | BACKGROUND: IN SITU REAL-TIME STRESS MONITORING IN COAL MINE

2.1 | Geological and mining conditions at Dongtan coal mine

The Dongtan coal mine is located in Ji-ning, Shandong, China (Figure 1). The longwall panel in this study was the 6303 working face. One side of the 6304 panel had been previously extracted. The protective coal pillar between the 6303 panel and the gob was 5-m-wide. The panel overburden depth was approximately 660 m, and the panel was 245-m-wide and 1400-m-long. The immediate roof of the panel comprised mudstone and had an average thickness of 0.8 m. The main roof was fine-grained sandstone with an average thickness of 12.9 m. The immediate floor of the panel was composed of siltstone and was approximately 1.54-m-thick. The main floor comprised fine-grained siltstone and was approximately 7.7-m-thick. The detailed information geological conditions of panel 6303 are shown in Figure 1. As shown in Figure 1, the geological conditions of the monitoring sections from S-1 to S-6 are relatively simple, and they are not located in densely faulted areas. To avoid the influence of geological conditions on

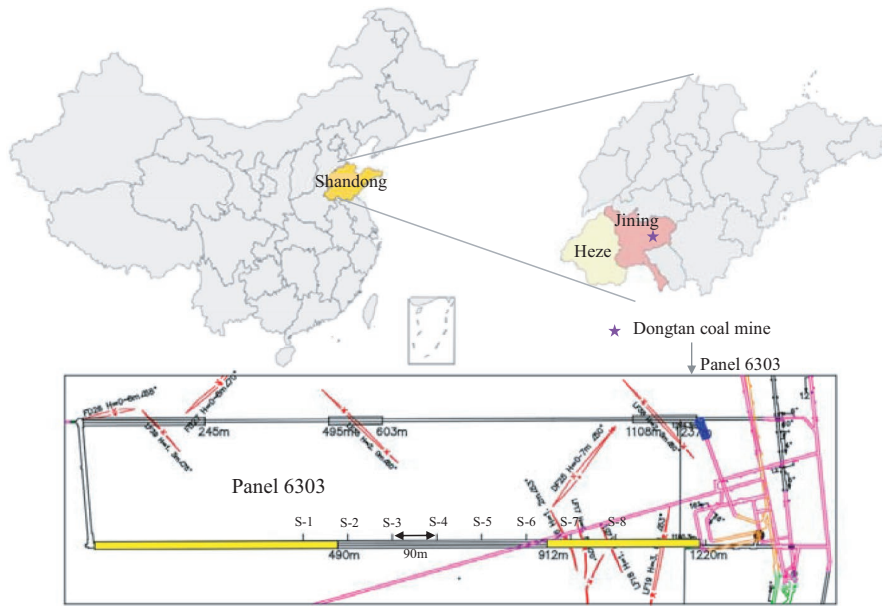


FIGURE 1 The location of Dongtan coal mine and the plan of panel 6303

this study, we selected the monitoring section of S-3 and S-4 as the research subject.

2.2 | Monitoring scheme

The working face of the 6303 panel experienced frequent microseismic events during extraction, which were closely related to the high-stress distribution around the mining area. A Fiber Bragg Grating (FBG) borehole deformation sensor for stress measurement in coal mine roof rock is adopted. The details of the measurement technique can be found in the literature.³⁹ To ensure accurate monitoring of the long-term stress changes during the mining process, sensors were installed in the intact homogenous main roof. This installation avoided the difficulties associated with drilling and installing sensors in the soft coal.

According to the theory of rock mechanics, stress in rock involves both in situ stresses σ_{ij}^0 and excavation-induced stresses $\Delta\sigma_{ij}$. The in situ stresses in a rock mass depend largely on the geological structure, such as discontinuities, faults, folds, and dikes. Excavation-induced stresses are due to the mining process or nearby activities such as excavation, blasting, or pumping. Therefore, the real stress state

of the rock mass is the sum of in situ stresses and induced stresses, which can be denoted as follows.

$$\sigma_{ij} = \sigma_{ij}^0 + \Delta\sigma_{ij} \quad (1)$$

Therefore, the monitoring of the dynamic evolution of mining-induced stress should include two steps, the investigation of the in situ stresses and real-time investigation of the induced stresses.

First, we measured the in situ stresses using the overcoring stress measurement method for the 6303 working face before long-term monitoring. The in situ stresses measured by overcoring tests are shown in Table 1. As shown in Figures 2 and 3, eight FBG borehole stress sensors were installed in the main roof of the 6303 panel at different section positions ahead of the working face. The sensor installation inclination was approximately 35° , the depth was approximately 15–20 m, and the distance between adjacent sensors was approximately 90 m. The detailed information of these stress monitoring boreholes is shown in Table 2. Figure 4 shows the core with sensor by overcoring test and on-site installation work. The dynamic 3D stress was obtained by the sensors with the variations of strains that were calculated from the variation of

TABLE 1 The in situ stresses

Geodetic coordinate system				Working sector coordinate system			
Principal stress	Value (MPa)	Azimuth angle ($^\circ$)	Oblique angle ($^\circ$)	Stress components	Value (MPa)	Stress components	Value (MPa)
σ_{11}	33.2	167.20	4.36	σ_{xx}	16.96	τ_{xy}	-0.52
σ_{22}	24.2	68.51	63.23	σ_{yy}	33.09	τ_{yz}	-0.64
σ_{33}	15.2	259.37	26.36	σ_{zz}	22.43	τ_{zx}	-3.61

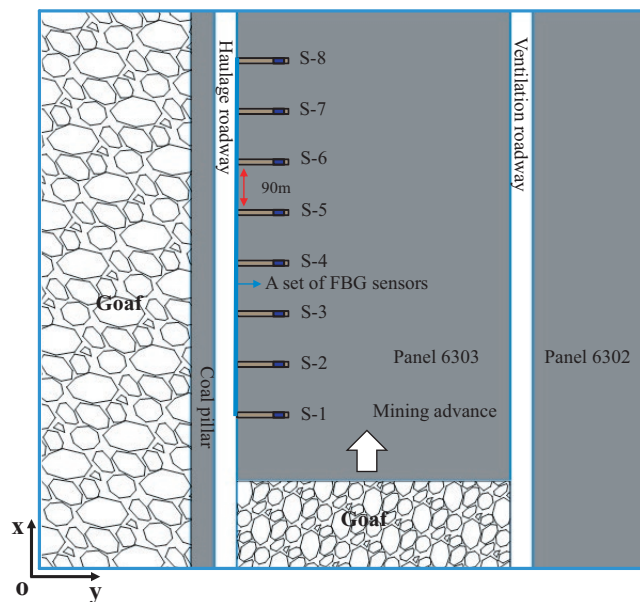


FIGURE 2 Sketch map of FBG sensor installation (Plan)

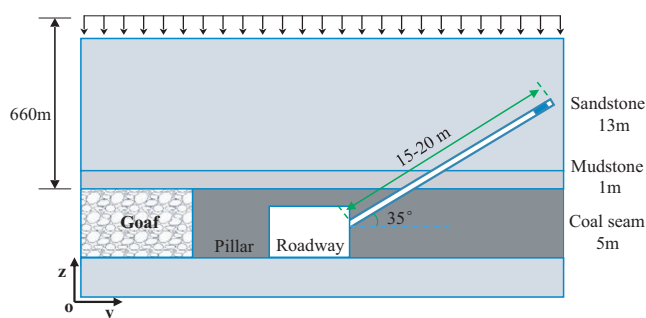


FIGURE 3 Sketch map of FBG sensor installation (Profile)

TABLE 2 The information of stress monitoring boreholes

Monitoring sections	Angel (°)	Depth (m)	Location (distance to the cut-out hole)
S-1	35	15.97	Haulage roadway 420 m
S-2	35	17.13	Haulage roadway 510 m
S-3	35	16.87	Haulage roadway 600 m
S-4	35	14.94	Haulage roadway 690 m
S-5	35	17.14	Haulage roadway 780 m
S-6	35	17.89	Haulage roadway 870 m
S-7	35	19.07	Haulage roadway 960 m
S-8	35	16.91	Haulage roadway 1050 m

wave lengths and the results of three normal stresses and three principal stresses within the rock mass. To illustrate the stress evolution law and data characteristics, normal stress measurements were taken from only two sections (S-2 and S-3) for a detailed analysis.



FIGURE 4 The core with sensor by overcoring test (A) and on-site installation work (B)

2.3 | Variation of normal stresses

The magnitude and distribution of existing in situ stresses around a coal seam are disturbed by a goaf formed due to underground mine excavation. Figure 5 shows typical monitoring results of the changes in the three normal stresses (σ_{xx} , σ_{yy} , σ_{zz}) in the monitoring sections S-2 and S-3. The three normal stresses vary, as shown in Figures 5(a) and 3(b). The figures show that the stress states in the monitoring sections are not affected by mining disturbance and are similar to the in situ stresses at distances >80 m from the working face. As the working face advances the three normal stresses increase; the vertical stress increases rapidly compared to the other two horizontal stresses. At a distance of about 20 m ahead of the working face, all the three normal stresses reach a peak and then decrease sharply because the integrity of the roof and coal seam is violated.

To quantitatively evaluate the degree of disturbance to the roof rock mass caused by mining activities, we assume that k_1 , k_2 , and k_3 are the corresponding stress concentration coefficients of the ratio of the three normal stresses to in situ stresses. The values of the coefficients when the working face is at different distances from the monitoring section can be calculated using the following equation:

$$\begin{cases} \sigma_{xx} = k_1 \sigma_{xx}^0 \\ \sigma_{yy} = k_2 \sigma_{yy}^0 \\ \sigma_{zz} = k_3 \sigma_{zz}^0 \end{cases} \quad (2)$$

Figure 6 shows the variations of the concentration coefficients of the three normal stresses during the coal mining

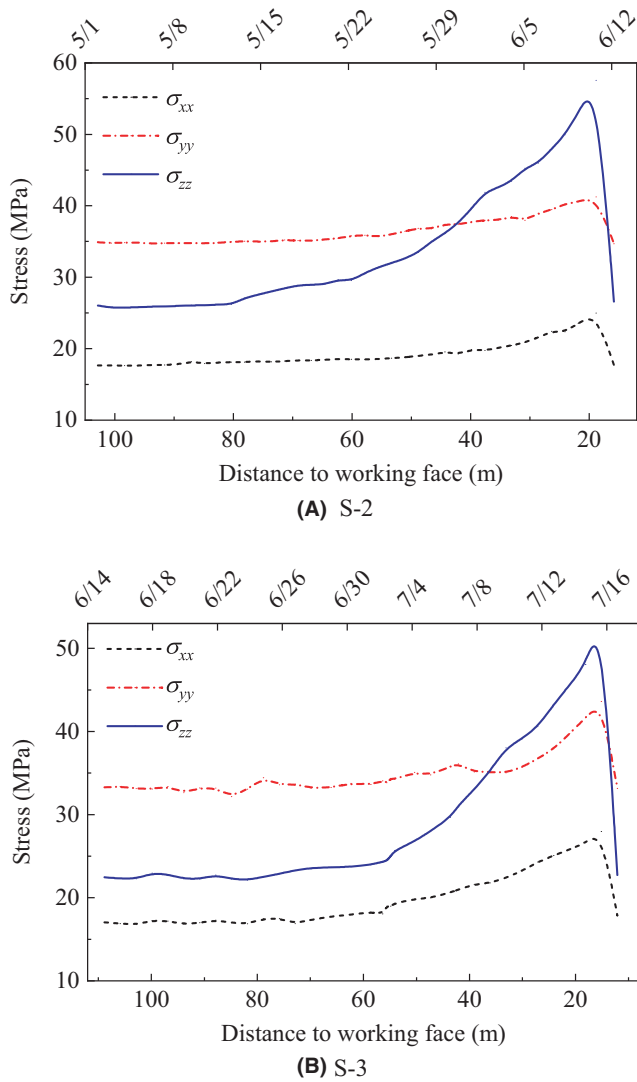


FIGURE 5 Monitoring results of normal stresses of the monitoring sections of S-2 (A) and S-3 (B)

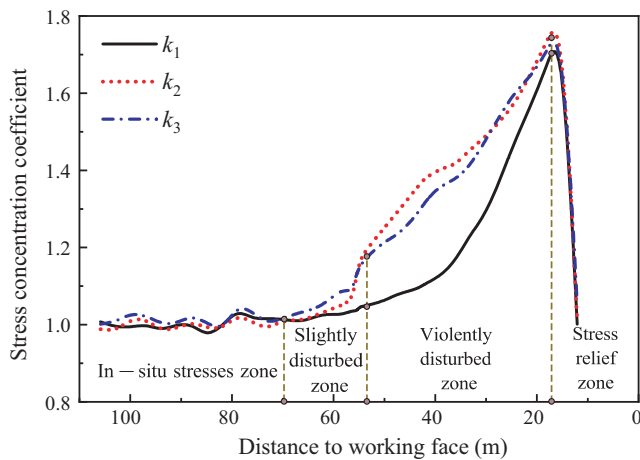


FIGURE 6 Variations of the concentration coefficients of the three normal stresses

process. Therefore, the area ahead of the working face can be divided into four zones, namely, the in situ stresses zone,

slightly disturbed zone, violently disturbed zone, and stress relief zone according to the values of the coefficients.

From above, field measurement can more directly reflect the distribution and variation law of mining-induced stress, and the monitoring data are essential for rock burst warning and prevention. However, it is ubiquitous and inevitable that monitoring data may be lost due to sensor malfunction, optical fiber cable damage caused by coal mining activities, and the volatile environment.³⁰ Because of lost data, information that is critical for safety evaluation may not be available. Thus, data that can be recovered when the sensor fails or short-term stress data predictions can have important implications for the diagnosis and prognosis of disasters in coal mines. For example, relevant destress measures can be adopted to mitigate high-stress concentration, and an early warning can be raised timeously. Therefore, in the next section, a methodology framework of an LSTM neural network that integrates transfer learning is proposed for stress prediction.

3 | METHODOLOGY FRAMEWORK

Figure 7 shows a flow chart of the methodology proposed in this study, and the algorithm of ensemble of the LSTM and transfer learning method is shown in Table 3. To summarize, the ensemble of LSTM and transfer learning method can be generalized into the following steps. First, the time series datasets $D_S = \{(x_1^S, t_1^S), (x_2^S, t_2^S), \dots, (x_n^S, t_n^S)\}$ and $D_T = \{(x_1^T, t_1^T), (x_2^T, t_2^T), \dots, (x_n^T, t_n^T)\}$ are collected to serve as source domain and target domain, respectively. Second, time series dataset D_S is preprocessed by rolling window method to obtain time series samples. The LSTM neural network is used to predict the stress; thus, the next step is to construct a base LSTM model using the source domain data. The grid search method is used to optimize the hyperparameters for the base model. Subsequently, the parameter transfer approach is used to complete knowledge transfer. The weights of hidden layers and the hyperparameters of the above pre-trained base LSTM model act as initialization parameters of the target LSTM model over the target domain data. The model parameters are then fine-tuned according to the test results. The proposed model is used to improve the prediction accuracy to overcome the missing data problem resulting from sensor or optical fiber cable damage caused by mining activities.

3.1 | LSTM neural network

Long short-term memory is a special RNN that is applicable to time series problems, compared to RNNs or

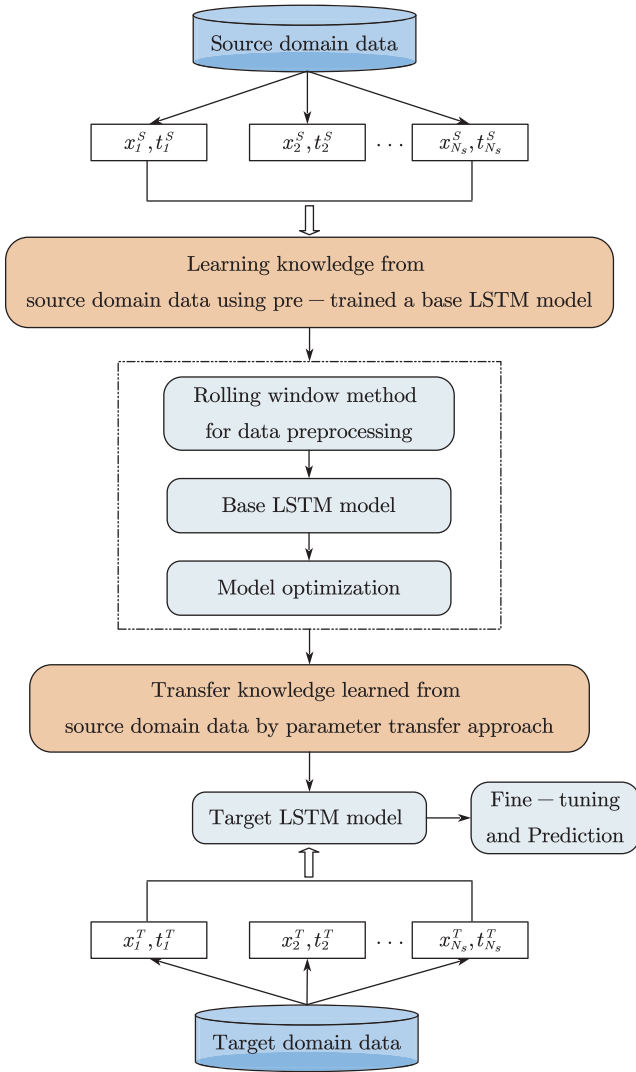


FIGURE 7 Flow chart of the stress prediction model using LSTM neural network and transfer learning method

traditional neural networks, which is best able to solve vanishing gradient and exploding gradient problems of long time series. A single LSTM cell comprises an input gate, a forget gate, an output gate, and the cell state memory (Figure 8). Gates are used to optionally perform information saving, adding, or deleting using the activation functions, thereby updating the cell state to achieve long-term storage of information and resolve the dependence of the time series on time. More specifically, the input gate controls the flow of input activation into the internal cell state. The forget gate controls the LSTM cell to forget or reset the cell's memory adaptively. The output gate controls the flow of output activation into the LSTM cell output.³¹ The activation functions are *sigmoid* and *tanh*.

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

$$\text{tanh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (4)$$

TABLE 3 The algorithm of ensemble of the LSTM and transfer learning

Algorithm 1 The Algorithm of Ensemble of the LSTM and Transfer Learning

Input: Source domain time series dataset

$$D_S = \{(x_1^S, t_1^S), (x_2^S, t_2^S), \dots, (x_N^S, t_N^S)\}$$

Target domain time series dataset

$$D_T = \{(x_1^T, t_1^T), (x_2^T, t_2^T), \dots, (x_N^T, t_N^T)\}$$

Rolling window size Δt

Learning rate η

Hidden layer L

Hidden units U

Epoch times I

Forecast horizon h

Output: The final model generated by the ensemble LSTM and transfer learning.

Pretreatment: Time series dataset D_S is preprocessed by rolling window method to obtain time series samples.

For $D_S = \{(x_1^S, t_1^S), (x_2^S, t_2^S), \dots, (x_N^S, t_N^S)\}$

Calculate the autocorrelation coefficients $\rho_{\Delta t}$ according to Eq. (16).

Determine the Δt_1 according to the values of $\rho_{\Delta t}$.

The D_S was divided into $K = N - (\Delta t + h) + 1$ time series samples.

1: For $D_S = \{(x_1^S, t_1^S), (x_2^S, t_2^S), \dots, (x_N^S, t_N^S)\}$

1.1: Build a base LSTM model, set $h = 1$, and set a group appropriate values for each parameters of η, L, U , and I .

1.2: Traverse the parameter combination to optimize the LSTM model by grid search method.

1.3: Calculate the RMSE according to Eq. (13).

1.4: Get an optimal pre-trained LSTM model.

2: For $D_T = \{(x_1^T, t_1^T), (x_2^T, t_2^T), \dots, (x_N^T, t_N^T)\}$

2.1: Build a new LSTM model, set $h=1, \Delta t_2=\Delta t_1$.

2.2: Transfer the weights of hidden layers and hyperparameters of $h = 1, \Delta t, \eta, L, U$, and I from the above optimal pre-trained LSTM model act as initialization parameters of the new LSTM model over the D_T .

2.3: Fine-tune the new model according to Eq. (13–15).

2.4: Update weights and hyperparameters, obtain the final model.

End For

At the time phase $t (t = 1, \dots, n)$ and inside the l th LSTM network layer, the input state of the LSTM cell is $x_t^{(l)}$; the forget gate is $f_t^{(l)}$; the input gate is $i_t^{(l)}$; the output gate is $o_t^{(l)}$; the hidden state output is $h_t^{(l)}$; and the memory cell state is $c_t^{(l)}$. At the previous time $t - 1$, the cell state memory is $c_{t-1}^{(l)}$ and the hidden state output is $h_{t-1}^{(l)}$. The following equations describe the relationship between these variables.^{31,32}

$$f_t^{(l)} = \sigma \left(\mathbf{W}_f^{(l)} [h_{t-1}, x_t] + \mathbf{b}_f^{(l)} \right) \quad (5)$$

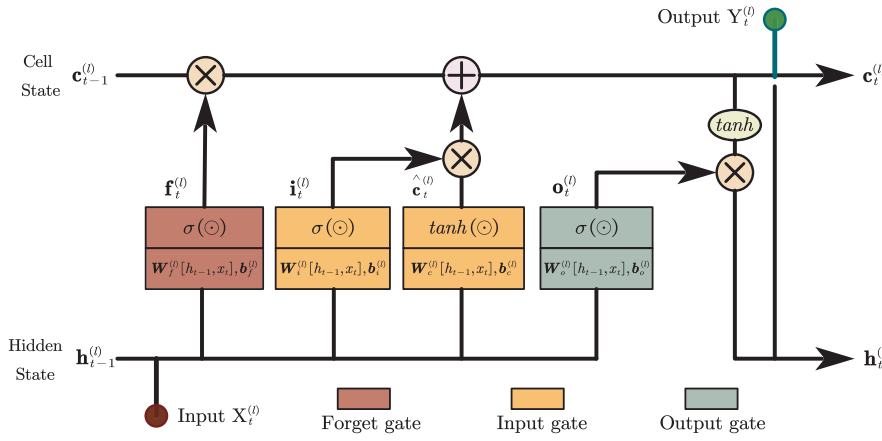


FIGURE 8 Single LSTM cell structure

$$i_t^{(l)} = \sigma \left(\mathbf{W}_i^{(l)} [h_{t-1}, x_t] + \mathbf{b}_i^{(l)} \right) \quad (6)$$

$$\hat{c}_t^{(l)} = \tanh \left(\mathbf{W}_c^{(l)} [h_{t-1}, x_t] + \mathbf{b}_c^{(l)} \right) \quad (7)$$

$$o_t^{(l)} = \sigma \left(\mathbf{W}_o^{(l)} [h_{t-1}, x_t] + \mathbf{b}_o^{(l)} \right) \quad (8)$$

$$c_t^{(l)} = f_t^{(l)} \cdot c_{t-1}^{(l)} + i_t^{(l)} \cdot \hat{c}_{t-1}^{(l)} \quad (9)$$

$$h_t^{(l)} = o_t^{(l)} \cdot \tanh \left(c_t^{(l)} \right) \quad (10)$$

Also shown in Figure 8 for schematic illustration, where $\mathbf{W}_\alpha^{(l)} [h_{t-1}, x_t]$ (with $\alpha = \{f, i, c, o\}$) are the weight matrices corresponding to different input vectors x_t or h_{t-1} , respectively, within different gates. $\hat{c}_{t-1}^{(l)}$ is a vector of candidate memory cells created by the \tanh function. σ and \tanh are the sigmoid and \tanh activation function, respectively.

3.2 | Rolling window method for data preprocessing

The raw data cannot be fed into the proposed model directly because the LSTM neural network expects input or output sequences. The rolling window method is used to transform the raw data into X (input) and Y (output) sequences. In this method, Δt is defined as the rolling window size, and a series of small samples with the same number can be obtained by rolling through the whole sample with a fixed window size. Therefore, to predict the time series value at time $t + 1$, the rolling window feeds not only the value at time t but also those at times $t - 1$, $t - 2$, ..., $t - \Delta t$ to the model. The predicted value at time $t + 1$ is appended to the sequence at time $t + 1$, t , $t - 1$,

TABLE 4 The rolling window algorithm

Algorithm 1 The Rolling Window Algorithm

Input: A time series dataset $D_T = \{T_1, T_2, T_3, \dots, T_n\}$

Rolling window size Δt

Min_periods t

Forecast horizon h

Output: The time series samples generated by rolling window algorithm

1: For $D_T = \{T_1, T_2, T_3, \dots, T_n\}$

1.1: For i in range $(1, n)$

For $\lambda_\alpha = (1, \alpha)$, $\eta_\alpha = (1, \alpha)$

1.1.1: Starting with i , Record data from T_i to $T_{i+\Delta t}$.

1.1.2: If $i + \Delta t + h < n$

1.1.3: Save the array $\{T_i, \dots, T_{i+\Delta t}\}$ as the first input sequence λ_1 , and the array $\{T_{i+\Delta t+1}, \dots, T_{i+\Delta t+h}\}$ as the first output sequence η_1 .

1.1.4: For $\lambda_\alpha, \eta_\alpha$ in range $(1, \alpha)$

1.1.5: Rolling the window to the right, Record data from $T_{i+\alpha t}$ to $T_{i+\alpha t+\Delta t}$.

1.1.6: If $i + \alpha t + \Delta t + h < n$

1.1.7: Save the array $\{T_{i+t}, \dots, T_{i+t+\Delta t}\}$ as the α input sequence λ_α , and the array $\{T_{i+\alpha t+\Delta t+1}, \dots, T_{i+\alpha t+\Delta t+h}\}$ as the α output sequence η_α .

1.1.8: End If

1.2: End For

2:End For

Return: The time series samples generated by rolling window algorithm

$t - 2, \dots, t - \Delta t - 1$, and so on until the last value has been predicted.³³ This can be expressed as follows.

$$f(x_t, x_{t-1}, x_{t-2}, \dots, x_{t-\Delta t}) = x_{t+1} \quad (11)$$

$$f(x_{t+1}, x_t, x_{t-1}, x_{t-2}, \dots, x_{t-\Delta t+1}) = x_{t+2} \quad (12)$$

The algorithm of rolling window is shown in Table 4. Specifically, we assume 10 samples in the dataset, including T_1, T_2, \dots, T_{10} , and set $\Delta t = 6$. An example of the transformation of the raw data to time series samples is shown in Figure 9.³⁴ The appropriate window size for the study will be identified later in Section 4.

3.3 | Transfer learning

Traditional machine learning (supervised learning) relies on the availability of a large amount of labeled data and the identical distribution of the training and test data.³⁵ However, the difference in data distribution and the lack of sufficient labeled data are challenges when tackling practical problems. Transfer learning, as opposed to traditional machine learning, uses the ability of a system to recognize and apply knowledge and skills learned in the source domains or tasks to new but related target domains or tasks. This is an important ability that enables solving the problems described above. Figure 10 shows the learning processes comparing traditional machine learning and transfer learning.

Given a source domain D_S and a target domain D_T , transfer learning aims to help improve the target prediction performance using the transfer of knowledge in the source domain data when dealing with the issue of time series prediction but with few fresh training samples. Based on the definition of transfer learning and learning patterns, transfer learning can be divided into instance transfer, feature-representation transfer, relational-knowledge transfer, and parameter transfer. In this study, the parameter transfer approach is adopted, in which the parameters of hidden layers and the hyperparameters of the above pre-trained LSTM model act as initialization parameters of another LSTM model over the target domain data.

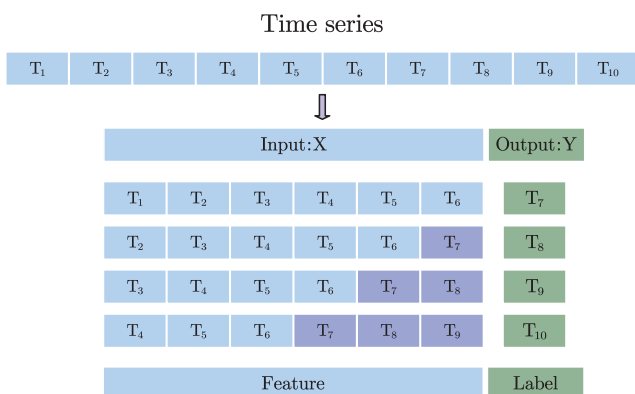


FIGURE 9 A preprocessing example to transform the raw data to time series samples

Some studies in engineering research have recently explored the applicability of deep learning techniques and transfer learning strategies. Li et al.⁴⁰ proposed a model to predict dam displacement data based on transfer learning and deep learning. Transfer learning is used to transfer the knowledge learned from similar sensors to improve prediction accuracy in the target sensor. Ma et al.^{33,34} proposed a method that integrates transfer learning and advanced deep learning to transfer knowledge from existing air quality stations to new stations to predict air quality. For monitoring mining-induced stress, a set of stress sensors is usually installed in different sections. Along the mining direction, the longwall face passes through each monitoring section in turn. When the longwall face is going through a monitoring section, the stress sensor can record complete monitoring data successfully if the sensor does not early fail work. Therefore, according to transfer learning theory, the complete monitoring data of the adjacent stress sensor can be considered the source domain, and the monitoring data of the following stress sensor can be regarded as the target domain.

4 | CASE STUDY

4.1 | Data collection and preprocessing

A detailed analysis of a stress monitoring scheme and the results in a coal mine are presented in Section 2. To test the proposed model, the monitored vertical stress data from the field monitoring case of the Dongtan coal mine were used as an application study. The monitored vertical stress data of S-2 (Figure 5(A)) (Data acquisition time: 8 May 2019–15 June 2019) and S-3 (Figure 5(B)) (Data acquisition time: 14 June 2019–18 July 2019) are used as the source and target domain data, respectively. To avoid the influence of noise data on the model, we artificially eliminate some extremely outlier and inconsistent data and select the average of the data every 6 h as a data sample during the data collection process. However, the model cannot learn the full process of the stress variation because the data must be split into training and test datasets. To learn enough patterns and knowledge from the source domain time series, especially when the peak of stress is reached, the vertical stress data of S-2 are copied to augment the size of the dataset in source domain. This is reasonable because transfer learning only focuses on the performance of the target domain. Then, the rolling window method is used to transform the source domain data to time series samples and split the data into the training and test datasets in proportions of 90%–10%, respectively, to pre-train the base LSTM model.

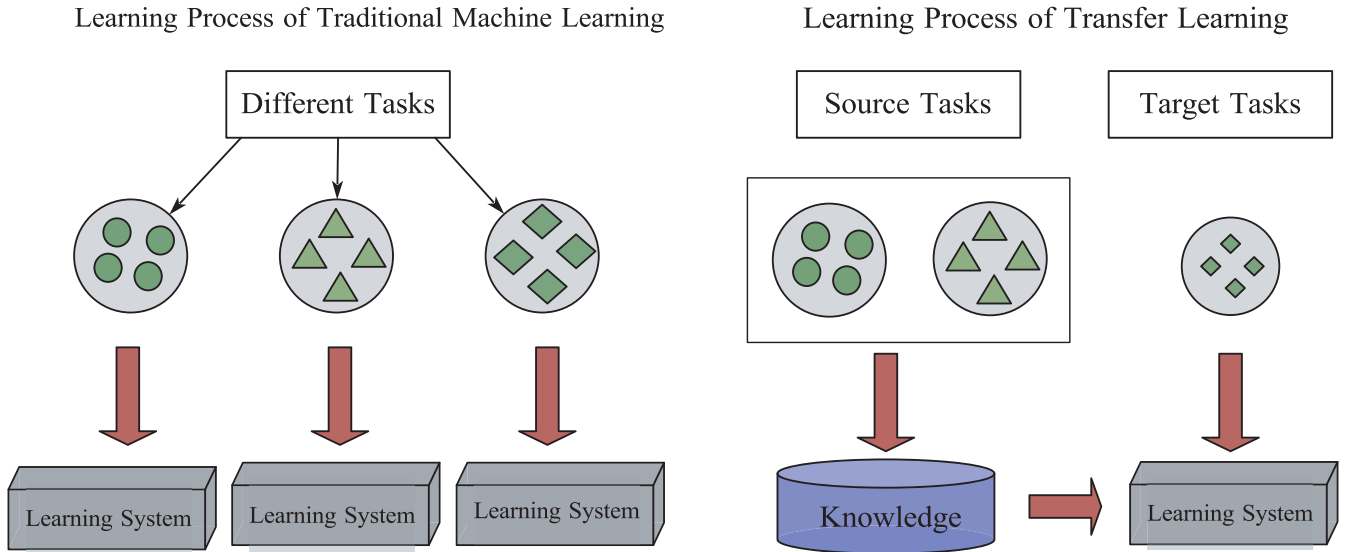


FIGURE 10 Learning process differences between traditional machine learning and transfer learning²⁹

4.2 | Performance evaluate indicators

To evaluate the performance of the proposed model, three widely used evaluation indicators, namely, root mean square error (RMSE), mean absolute percentage error (MAPE), and mean absolute error (MAE), are adopted to measure the prediction error of the model. The equations are described as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (13)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n |1 - \hat{y}_i/y_i| \quad (14)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (15)$$

where n is the number of the prediction, and y_i and \hat{y}_i are the i th actual monitoring stress data and predicted value, respectively. $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$. Low values of the RMSE, MAPE, and MAE indicate high accuracy of the predictions.

4.3 | Determination of hyperparameters and network structure for base LSTM model

4.3.1 | Rolling window sizes

The size of the rolling window Δt influences the prediction performance. This is because data for previous instances

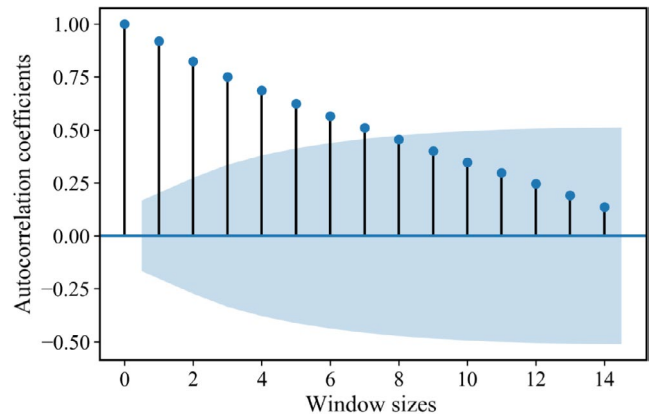


FIGURE 11 The autocorrelation coefficient as a function of window size

might have a strong or weak lagged effect on the data at the next instance. A small window size cannot guarantee that enough information and sample features will be processed for the LSTM neural network inputs, while a large window size might increase unrelated information and, thus, the computation complexity.^{34,36} To determine an appropriate window size, the autocorrelation function³⁷ is used to measure the temporal correlations among stress time series. Higher autocorrelation coefficients indicate stronger time correlations. For a window size Δt , the autocorrelation functions can be calculated as follows:

$$\rho_{\Delta t} = \frac{Cov(y(t), y(t + \Delta t))}{\sqrt{\sigma_{y(t)} \sigma_{y(t + \Delta t)}}} \quad (16)$$

where $y(t)$ and $y(t + \Delta t)$ denote the stress value at time t and $t + \Delta t$, respectively, $Cov(\cdot)$ represents the covariance, and $\sigma(\cdot)$ is the standard deviation.

TABLE 5 Influence of the learning rate on the performance

Network structure	Learning rate	Number of iterations	RMSE	Network structure	Learning rate	Number of iterations	RMSE
2 layers [50,50] cells	0.0001	400	1.318	3 layers [25,25,25] cells	0.0001	1000	0.247
	0.0005	400	2.178		0.0005	700	1.496
	0.001	500	0.515		0.001	300	3.195
	0.01	1000	14.740		0.01	500	4.753
3 layers [100,50,25] cells	0.0001	400	2.487	3 layers [20,10,5] cells	0.0001	400	0.283
	0.0005	700	0.746		0.0005	400	0.718
	0.001	700	3.794		0.001	700	0.818
	0.01	500	17.850		0.01	300	3.105

For the source domain data, Figure 11 shows the autocorrelation coefficients for different window sizes. The autocorrelation coefficients clearly decrease with increasing window size. This confirms that the earlier events have a weaker impact on the current status. In addition, when the window size is <7 , the autocorrelation coefficient is >0.5 . This study follows the range of the window size used in previous studies^{33,38} when the autocorrelation coefficients are >0.5 , which indicate a high temporal correlation. However, for our stress data, the coefficients are >0.5 for a window size ≤ 7 . To ensure enough sample features for model inputs, the window size Δt was therefore set as 7.

4.3.2 | Learning rate and LSTM structure

In addition to the above rolling window sizes, the LSTM neural network structure and other hyperparameters influence the forecasting performance. To obtain an optimal prediction performance on the base LSTM model, the grid search method is used to optimize the hyperparameters. Given the premise of window size $\Delta t = 7$, the ranges of the learning rate (0.0001, 0.01), hidden layers (2, 3), hidden units (5, 100), and epoch times (200, 1000) were used to determine the optimal hyperparameters for the model. The evaluation criteria of the RMSE were used to measure the prediction performance for each parameter combination. In this base LSTM model, the Adam optimizer algorithm was used, which can replace the classic stochastic gradient descent method to update the network weights more effectively.

Table 5 shows the influence of the learning rate on performance. For the models using the four network structures, the prediction error increases significantly when the learning rate is 0.01 and decreases when the learning rate is <0.001 . Therefore, the recommended learning rate for this model is <0.001 . Figure 12 shows the prediction accuracy as a function of the different LSTM network structures and number of iterations (for learning rate = 0.0001). The different network structures provide considerable improvements to the prediction error of the LSTM models. The minimum

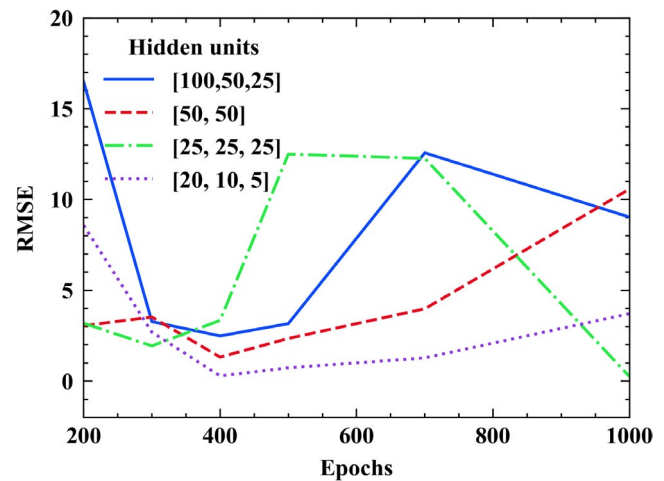


FIGURE 12 RMSE as a function of the network structure (learning rate = 0.0001)

prediction error is at around the 400th iteration. However, the fluctuations are caused by the inherent stochasticity in training or poor combination of parameters, which are usually observed in such cases, and cannot be completely eliminated. Therefore, for these hyperparameters (such as the learning rate of 0.0005 and 0.0001, and network structure as [20,10,5]), the amplitude and frequency of fluctuations can be reduced. Considering the above problems and learning efficiency, the learning rate, network structure, and epochs were set as 0.0001, [20,10,5], and 400 for the base LSTM model, respectively.

4.4 | Analytical results and discussion

4.4.1 | Performance comparison between without transfer learning and with transfer learning

To verify the effect of the proposed model, the complete monitoring data of vertical stress of S-3 were considered to be the target domain data. The hyperparameters and

weights of the hidden layer of the pre-trained LSTM base model are transferred from the source domain data (vertical stress of S-2) to serve as initialization for the target LSTM model to be trained on the target domain data (vertical stress of S-3). It is assumed that the vertical stress data of S-3 are missing some data when the distance to the working face is approximately 33 m. Therefore, the monitored data closer than 33 m can be set as the training set, and all remaining data can be set as the test set. In addition, to compare the performance of the proposed model using the transfer learning method, another LSTM model without using transfer learning, and two time series prediction commonly used models, including Autoregressive Integrated Moving Average (ARIMA) and Recurrent Neural Network (RNN), are also used to generate predictions.

The predicted results without transfer learning (LSTM), with transfer learning, ARIMA, and RNN, are illustrated in Figures 13, 14, 15, and 16, respectively. The 95% confidence intervals calculated using multiple predictions are also shown in these figures. Given the prediction results of both cases, it can be seen from Figures 13, 15, and 16 that the variation of the predicted vertical stress tends to increase indefinitely when the transfer learning method is not used, and the peak of stress cannot be predicted. Figure 14 shows that it can be predicted accurately by the proposed method using transfer learning. This has important implications for the diagnosis and prognosis of rock bursts in coal mines. A comparison of the prediction error by three evaluation criteria is shown in Table 6, which shows that the predicted error without using transfer learning is 2–3 times greater than that obtained by the proposed method. This confirms that the use of the

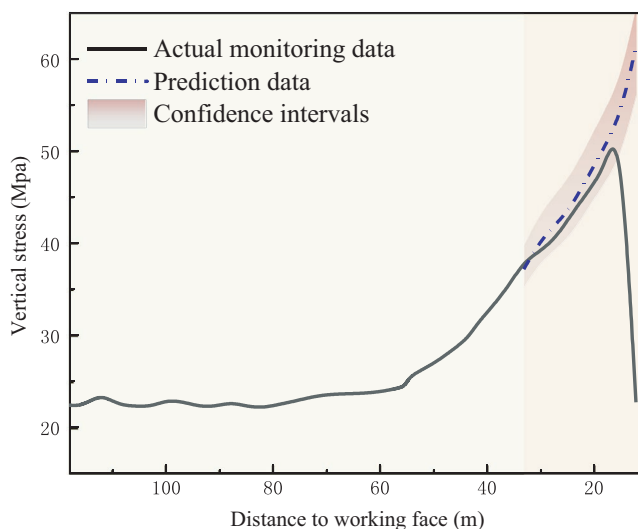


FIGURE 13 Predicted results without transfer learning (LSTM)

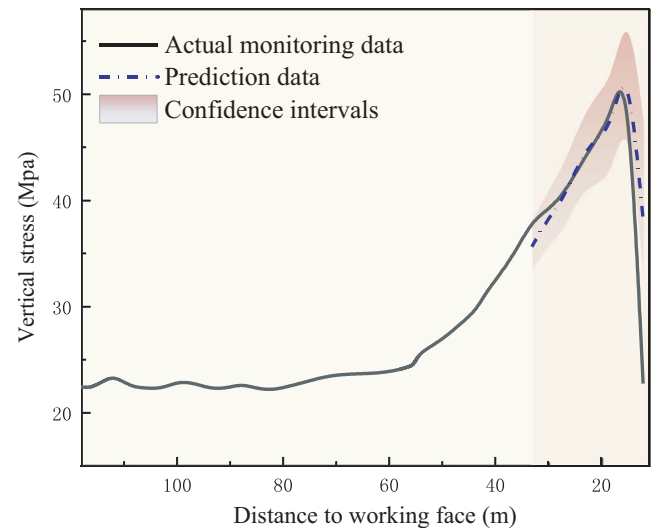


FIGURE 14 Predicted results with transfer learning (TL-LSTM)

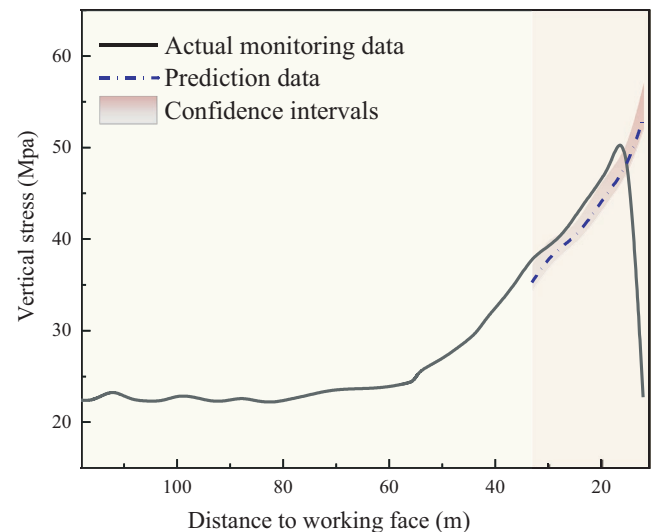


FIGURE 15 Predicted results by ARIMA

proposed method is more efficient than the use of the LSTM, ARIMA, and RNN model alone.

4.4.2 | Performance on different size of the target domain training data availability

In the actual monitoring application, the monitoring data may be lost at any time due to sensor failure, abnormal data transmission, or human factors. Therefore, the availability and amount of target domain data depends strongly on the practical situation and has a significant influence on the prediction results. To test the performance on different target domain training data sizes, six versions of the target domain training data were established. The target

TABLE 6 A comparison of predicted error by four evaluation criteria

Baselines	RMSE (MPa)	MAPE (%)	MAE (MPa)
Without transfer learning (LSTM)	13.71	23.0	5.86
With transfer learning (TL-LSTM)	5.55	10.2	2.67
ARIMA	11.09	21.8	6.11
RNN	13.50	23.8	6.12

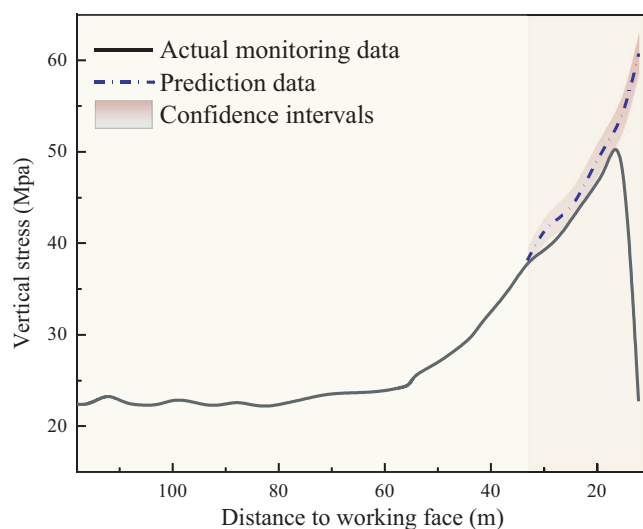


FIGURE 16 Predicted results by RNN

domain training data were set at 50%, 55%, 60%, 65%, 70%, and 80% of the total target domain data, and the test set was assigned to all the remaining data. For each case, the same pre-trained LSTM model was used for transfer learning.

In addition, two metrics, which are RMSE and coefficient of determination (R^2), are adopt to describe the prediction performance on different size of the target domain training data availability of developed model. The RMSE has been defined in Section 3.4, and the R^2 can be calculated using the following equation:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (17)$$

The results showed that, for each case, the stress peak and drop point can be predicted accurately using the transfer learning strategy. The stress drop points varied considerably; however, this was not important as the focus was on predictions of the buildup to the stress

peak. The results of the RMSE and R -squared parameters as a function of size of the training data are shown in Figure 17.

Figure 17 shows that when the training data size was set to 50%, the prediction performance was poor. While the prediction performance improves significantly when the availability of the training data is 55%, the RMSE decreased from 7.21 to 3.48 MPa and the R^2 increased from 0.32 to 0.83. The fundamental reason for this trend is that the stress state was still in the virgin mode with no excavation disturbance at 66 m from the working face (training data availability 50%). The stress state is in an excavation disturbed zone at 54 m from the working face (training data availability 55%). As the training data size increases further, the prediction performance in terms of RMSE and R^2 values gradually improves. When the target domain training data availability is >65%, the RMSE value is <2 MPa and the R^2 is >0.9. This experiment demonstrates that the proposed model in this study is effective and can be used to predict the stress in future and recover the missing stress data when the target domain training data availability is >55%. In other words, the proposed model can significantly improve the prediction performance when the target domain training data availability has reached the stress disturbed zone.

5 | CONCLUSION

In this study, a framework that integrates an LSTM neural network and the transfer learning method to improve the prediction performance for missing stress data is proposed. To test the proposed model, vertical stress data of two monitoring sections obtained from previous field measurement for stresses in the Dongtan coal mine were selected as a case study for stress prediction. The main conclusions from this study are as follows.

For the base LSTM model, excellent prediction performance is obtained when the network structure is set as three layers comprising [20,15,5] cells and when the window size is 7 and the learning rate is 0.0001.

Model application and experimental results showed that the proposed model using transfer learning is more efficient than a model without transfer learning. Moreover, the peak stress can be predicted accurately by the proposed model.

With the increase in the size of the target domain training data availability, the prediction performance measured by the RMSE and R -squared values improves. The proposed model can be used to predict the stress in future and recover missing stress data. The prediction

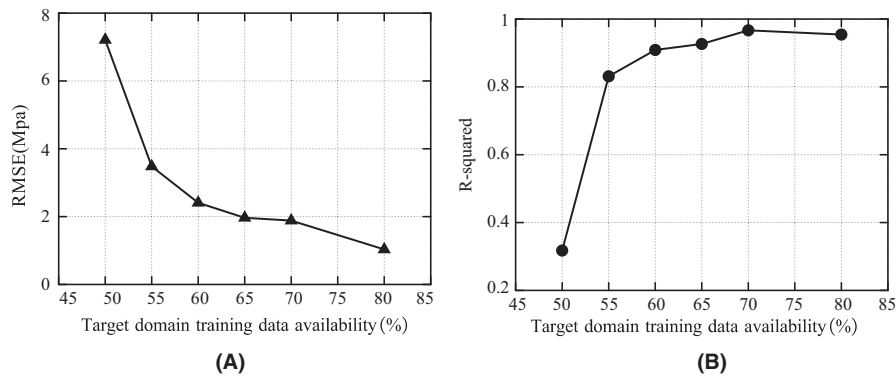


FIGURE 17 Performance as a function of target domain training size: (A) RMSE and (B) R-squared

performance improves significantly when the target domain training data availability is $>55\%$, indicating that the data have reached the stress disturbed zone.

The main contributions of the proposed framework are that the LSTM neural network and transfer learning method are integrated to improve the mining-induced stress prediction performance when data are missing. This can be applied to reconstruct the missing data or predict the stress state in the next few days, which is crucial for the diagnosis and prognosis of the stability of surrounding rock, and provides an important reference for similar projects. However, in the mining field, many factors such as blasting, tremor generated by wave propagation, and energy release of the hard roof fracturing can cause sudden stress changes. The influence of these external factors on stress cannot be accounted for by the proposed model. This challenging and significant work will be carried out in future studies.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Grant numbers [51991393] and [52079134]).

CONFLICT OF INTEREST

We declare that we have no conflict, interest, and personal relationships with other people or organizations that can inappropriately influence our work.

ORCID

Wusheng Zhao  <https://orcid.org/0000-0001-5204-0205>

REFERENCES

- Keneti A, Sainsbury BA. Review of published rockburst events and their contributing factors. *Eng Geol.* 2018;246:361-373. doi:10.1016/j.enggeo.2018.10.005
- Zi LI, He XQ, Dou LM, Song DZ, Wang GF. Numerical investigation of load shedding and rockburst reduction effects of top-coal caving mining in thick coal seams. *Int J Rock Mech Min.* 2018;110:266-278. doi:10.1016/j.ijrmms.2018.08.005
- Konicek P, Schreiber J. Heavy rockbursts due to longwall mining near protective pillars: a case study. *Int J Min Sci Technol.* 2018;28(5):799-805. doi:10.1016/j.ijmst.2018.08.010
- Xue D, Wang J, Zhao Y, Zhou H. Quantitative determination of mining-induced discontinuous stress drop in coal. *Int J Rock Mech Min.* 2018;111:1-11. doi:10.1016/j.ijrmms.2018.08.003
- Yavuz H. An estimation method for cover pressure re-establishment distance and pressure distribution in the goaf of longwall coal mines. *Int J Rock Mech Min.* 2004;41:193-205. doi:10.1016/S1365-1609(03)00082-0
- Kodama J, Miyamoto T, Kawasaki S, Fujii Y, Kaneko K, Hagan P. Estimation of regional stress state and Young's modulus by back analysis of mining-induced deformation. *Int J Rock Mech Min.* 2013;63:1-11. doi:10.1016/j.ijrmms.2013.05.007
- Zhang H, Cao J, Tu M. Floor stress evolution laws and its effect on stability of floor roadway. *Int J Min Sci Technol.* 2013;23(5):631-636. doi:10.1016/j.ijmst.2013.08.002
- Zhu SY, Jian ZQ, Hl H, Xiao WG, Yao P. Analytical model and application of stress distribution on mining coal floor. *J China Univ Min Technol.* 2008;18(1):13-17.
- Salamon MDG. Elastic analysis of displacements and stresses induced by the mining of seam or reef deposits, Part I. *J S Afr I Min Metall.* 1963;64(4):128-149.
- Basarir H, Ferid Oge I, Aydin O. Prediction of the stresses around main and tail gates during top coal caving by 3D numerical analysis. *Int J Rock Mech Min.* 2015;76:88-97. doi:10.1016/j.ijrmms.2015.03.001
- Jiang Y, Wang H, Xue S, Zhao Y, Zhu J, Pang X. Assessment and mitigation of coal bump risk during extraction of an island longwall panel. *Int J Coal Geol.* 2012;95:20-33. doi:10.1016/j.coal.2012.02.003
- Wang H, Jiang Y, Zhao Y, Zhu J, Liu S. Numerical investigation of the dynamic mechanical state of a coal pillar during longwall mining panel extraction. *Rock Mech Rock Eng.* 2012;46(5):1211-1221. doi:10.1007/s00603-012-0337-8
- Yang W, Lin BQ, Qu YA, et al. Stress evolution with time and space during mining of a coal seam. *Int J Rock Mech Min.* 2011;48(7):1145-1152. doi:10.1016/j.ijrmms.2011.07.006
- Thin IGT, Pine RJ, Trueman R. Numerical modelling as an aid to the determination of the stress distribution in the goaf due to longwall coal mining. *Int J Rock Mech Min.* 1993;30:1403-1409.
- Das AJ, Mandal PK, Bhattacharjee R, Tiwari S, Kushwaha A, Roy LB. Evaluation of stability of underground workings for exploitation of an inclined coal seam by the ubiquitous joint model. *Int J Rock Mech Min.* 2017;93:101-114. doi:10.1016/j.ijrmms.2017.01.012

16. Konicek P, Waclawik P. Stress changes and seismicity monitoring of hard coal longwall mining in high rockburst risk areas. *Tunn Undergr Sp Technol*. 2018;81:237-251. doi:10.1016/j.tust.2018.07.019
17. Stas L, Soucek K, Knejzlik L. First results of conical borehole strain gauge probes applied to induced rock mass stress changes measurement. *Acta Geodyn Geomater*. 2006;4(4):77-82.
18. Stas L, Knejzlik J, Palla L, Soucek K, Waclawik P. Measurement of stress changes using a compact conical-ended borehole monitoring. *Geotech Test J*. 2011;34(6):685-693.
19. Shen B, Duan Y, Luo X, et al. Monitoring and modelling stress state near major geological structures in an underground coal mine for coal burst assessment. *Int J Rock Mech Min*. 2020;129:104294. doi:10.1016/j.ijrmms.2020.104294
20. Zhang M, Jiang F. Rock burst criteria and control based on an abutment-stress-transfer model in deep coal roadways. *Energy Sci Eng*. 2020;8:2966-2975. doi:10.1002/ese3.715
21. Wang Y, Zheng G, Wang X. Development and application of a goaf-safety monitoring system using multi-sensor information fusion. *Tunn Undergr Sp Technol*. 2019;94:103112. doi:10.1016/j.tust.2019.103112
22. Wan HP, Ni YQ. Bayesian multi-task learning methodology for reconstruction of structural health monitoring data. *Struct Health Monit*. 2018;18(4):1282-1309. doi:10.1177/1475921718794953
23. Chen Y, Ma S, Yu Y. Stability control of underground roadways subjected to stresses caused by extraction of a 10-m-thick coal seam: a case study. *Rock Mech Rock Eng*. 2017;50(9):2511-2520. doi:10.1007/s00603-017-1217-z
24. Yang JP, Chen WZ, Li M, Tan XJ, Yu JX. Structural health monitoring and analysis of an underwater TBM tunnel. *Tunn Undergr Sp Technol*. 2018;82:235-247. doi:10.1016/j.tust.2018.08.053
25. Wan Z, Li S, Huang Q, Wang T. Structural response reconstruction based on the modal superposition method in the presence of closely spaced modes. *Mech Syst Signal Process*. 2014;42(1-2):14-30. doi:10.1016/j.ymsp.2013.07.007
26. Liu Y, Cheng H, Kong X, Wang Q, Cui H. Intelligent wind turbine blade icing detection using supervisory control and data acquisition data and ensemble deep learning. *Energy Sci Eng*. 2019;7:2633-2645. doi:10.1002/ese3.449
27. Yao R, Pakzad SN. Autoregressive statistical pattern recognition algorithms for damage detection in civil structures. *Mech Syst Signal Process*. 2012;31:355-368. doi:10.1016/j.ymsp.2012.02.014
28. Ye R, Dai Q. A novel transfer learning framework for time series forecasting. *Knowl-Based Syst*. 2018;156:74-99. doi:10.1016/j.knsys.2018.05.021
29. Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng*. 2010;22(10):1345-1359.
30. Ren P, Chen X, Sun L, Sun H. Incremental Bayesian matrix/tensor learning for structural monitoring data imputation and response forecasting. *Mech Syst Signal Process*. 2021;158:107734. doi:10.1016/j.ymsp.2021.107734
31. Zhang R, Liu Y, Sun H. Physics-informed multi-LSTM networks for metamodelling of nonlinear structures. *Comput Method Appl Mech Eng*. 2020;369:113226. doi:10.1016/j.cma.2020.113226
32. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;9(8):1735-1780.
33. Ma J, Cheng JCP, Lin C, Tan Y, Zhang J. Improving air quality prediction accuracy at larger temporal resolutions using deep learning and transfer learning techniques. *Atmos Environ*. 2019;214:116885. doi:10.1016/j.atmosenv.2019.116885
34. Ma J, Li Z, Cheng JCP, Ding Y, Lin C, Xu Z. Air quality prediction at new stations using spatially transferred bi-directional long short-term memory network. *Sci Total Environ*. 2020;705:135771. doi:10.1016/j.scitotenv.2019.135771
35. Lu J, Behbood V, Hao P, Zuo H, Xue S, Zhang G. Transfer learning using computational intelligence: a survey. *Knowl-Based Syst*. 2015;80:14-23. doi:10.1016/j.knsys.2015.01.010
36. Ma J, Ding Y, Gan VJL, Lin C, Wan Z. Spatiotemporal prediction of PM_{2.5} concentrations at different time granularities using IDW-BLSTM. *IEEE Access*. 2019;7:107897-107907. doi:10.1109/access.2019.2932445
37. Box GEP, Jenkins GM. Time series analysis: forecasting and control. *JTime*. 2010;31(3):592. doi:10.1111/j.1467-9892.2009.00643.x
38. Li X, Peng L, Yao X, et al. Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. *Environ Pollut*. 2017;231(P1):997-1004. doi:10.1016/j.envpol.2017.08.114
39. Zhao WS, Zhong K, Chen WZ. A fiber Bragg grating borehole deformation sensor for stress measurement in coal mine rock. *Sensors*. 2020;20(11):3267. doi:10.3390/s20113267
40. Li YT, Bao TF, Chen H, et al. A large-scale sensor missing data imputation framework for dams using deep learning and transfer learning strategy. *Measurement*. 2021;178:109377. doi:10.1016/j.measurement.2021.109377
41. Shen S-L, Atangana Njock PG, Zhou A, Lyu H-M. Dynamic prediction of jet grouted column diameter in soft soil using Bi-LSTM deep learning. *Acta Geotech*. 2021;16(4):303-315. doi:10.1007/s11440-020-01005-8
42. Zhang N, Shen SL, Zhou A, Jin YF. Application of LSTM approach for modelling stress-strain behaviour of soil. *Appl Soft Comput*. 2021;100:106959. doi:10.1016/j.asoc.2020.106959
43. Zhang N, Zhou A, Pan YT, Shen SL. Measurement and prediction of tunnelling-induced ground settlement in karst region by using expanding deep learning method. *Measurement*. 2021;183:109700. doi:10.1016/j.measurement.2021.109700
44. Lin SS, Shen SL, Zhang N, Zhou A. Modelling the performance of EPB shield tunnelling using machine and deep learning algorithms. *Geosci Front*. 2021;12(5):101177. doi:10.1016/j.gsf.2021.101177
45. Wang RH, Li DQ, Chen EJ, et al. Dynamic prediction of mechanized shield tunneling performance. *Automat Constr*. 2021;132:103958. doi:10.1016/j.autcon.2021.103958
46. Hamrouni A, Sbartaï B, Dias D, et al. Probabilistic analysis of ultimate seismic bearing capacity of strip foundations. *J Rock Mech Geotech*. 2018;10(4):8. doi:10.1016/j.jrmge.2021.05.009

How to cite this article: Qin C, Zhao W, Zhong K, Chen W. Prediction of longwall mining-induced stress in roof rock using LSTM neural network and transfer learning method. *Energy Sci Eng*. 2022;10:458-471. doi:[10.1002/ese3.1037](https://doi.org/10.1002/ese3.1037)