Journal of Environmental Health Science and Engineering (2021) 19:401–414
https://doi.org/10.1007/s40201-021-00613-0

RESEARCH ARTICLE

Monitoring and prediction of dust concentration in an open-pit mine using a deep-learning algorithm

Lin Li[1] · Ruixin Zhang[1,2] · Jiandong Sun[2] · Qian He[1] · Lingzhen Kong[1] · Xin Liu[2]

Received: 2 August 2020 / Accepted: 4 January 2021 / Published online: 3 February 2021

Abstract

Purpose Dust pollution is currently one of the most serious environmental problems faced by open-pit mines. Compared with underground mining, open-pit mining has many dust sources, and a wide area of influence and complicated changes in meteorological conditions can result in great variations in dust concentration. Therefore, the prediction of dust concentrations in open-pit mines requires research and is of great significance for reducing environmental pollution and personal health hazards.

Methods This study is based on monitoring of the concentration of total suspended particulate (TSP) in the Anjialing open-pit coal mine in Pingshuo. This paper proposes a hybrid model based on a long short-term memory (LSTM) network and the attention mechanism (LSTM-Attention) and applies it to the prediction of TSP concentration. The LSTM model reflects the historical process of an input time series, and the attention mechanism extracts the inherent characteristics of the input parameters to assign weights based on the importance of the influencing factors. The autoregressive integrated moving average (ARIMA) and LSTM models are also used to predict the TSP concentration. Finally, several statistical measures of error are used to evaluate the accuracy of the model and perform a sensitivity analysis.

Results It was found that, in general, the TSP concentration was highest in the period 08:00–09:00 and lowest in the period 15:00–16:00. In addition to the influence of meteorological parameters and normal operations, the reason for this trend is the presence of an inversion layer above the open-pit mine. The results show that, compared with the ARIMA and LSTM models, the LSTM-Attention model is more stable and has a prediction accuracy that is 5.6% and 3.0% greater, respectively.

Conclusion This model can be applied to the prediction of dust concentrations in open-pit mines and provide guidance on when to carry out dust-suppression work. It has expansibility and is potentially valuable for application in a wide range of areas.

Keywords Deep learning · Dust monitoring · Inversion layer · LSTM-attention · Open-pit mine · TSP prediction

Introduction

With the development of society, people have higher expectations for their quality of life and for the natural environment. However, exposure to dust pollution is currently a significant public health problem [1–4]. Open-pit mining is an important method for extracting mineral resources, and it involves operations including perforation, blasting, loading, transport, and soil disposal [5]. In the process of these operations, the crushing of a large amount of coal and rocks produces a large concentration of dust. This is called composite dust, and it includes coal dust, rock dust, and small amounts of other substances [6]. The composite dust floats in the air for a long time and is harmful to workers and the environment. It can reduce visibility in the workspace, cause personal injury, present a risk of combustion or explosion, and increase mechanical losses [7, 8]. Thus, accurate prediction of dust concentration is essential for mitigating these problems and ensuring effective control of atmospheric pollution around open-pit mines.

A large number of scholars have proposed many analysis and prediction methods related to the problem of dust concentration around mines. Some have proposed mathematical models to estimate dust concentration [9, 10]. However, these models have some limitations and only apply to certain special

✉ Lin Li
lilin@student.cumtb.edu.cn

[1] School of Energy and Mining Engineering, China University of Mining & Technology (Beijing), Beijing 100083, China

[2] School of Safety Engineering, North China Institute of Science and Technology, Sanhe 065201, Hebei, China

cases, meaning they cannot be applied universally. Due to the influence of open-pit mining on airflow, the problem of ventilation in these mines becomes complicated, and this also increases the difficulty of analysis using mathematical modeling [11]. Because of the problems with mathematical modeling in open-pit mining situations, making predictions through data analysis has been considered.

Neural networks have been widely used in prediction of air pollution [12–15]. Kukkonen et al. [16] used neural-network, linear statistical, and deterministic model systems to predict urban concentrations of NO_2 and particulate matter with an aerodynamic diameter less than or equal to 10 μm (PM10), and their results showed that the neural-network model had the best performance. Wang et al. [17] proposed an improved neural network and optimized it using a genetic algorithm, and their system showed good performance in PM2.5 (particulates with aerodynamic diameter < 2.5 μm) concentration prediction. Choubin et al. [18] used a random forest algorithm, bagged classification and regression trees (BCRT), and mixture discriminate analysis (MDA) to predict the danger of PM10 in the Barcelona Province of Spain [19, 20], finding that the BCRT model was superior to the MDA model. Zhang et al. [21] proposed a wavelet-based autoregressive moving average (ARMA)/autoregressive integrated moving average (ARIMA) model for short-term serial prediction of PM10 concentration in Taiyuan City, Shanxi Province, China. It was found that the wavelet-ARMA/ARIMA method can effectively reduce the prediction error and improve the prediction accuracy. Bui et al. [22] studied the prediction and control of PM10 concentrations caused by open-pit drilling operations and provided a new artificial-intelligence system. Li et al. [23] proposed a PM2.5 concentration-prediction algorithm based on time series and an interactive multi-model (IMM). They found that the IMM algorithm had higher prediction accuracy for PM2.5 [24]. Ahn et al. [25] designed a microchip made from sensors that can record measurement data regularly and proposed a model to estimate atmospheric changes using deep learning. Neural-network models are effective for predicting and controlling dust concentrations [12]. However, the emission and diffusion of dust are not only related to the nature of the dust itself, but also to meteorological conditions [11]; the prediction of dust concentrations needs to consider multiple influences. In this case, the difficulty of using traditional shallow models is increased. Therefore, to better predict dust concentrations, it is more appropriate to adopt a model that is good at dealing with multiple variables.

Long short-term memory (LSTM) networks are good at dealing with multiple variables. Their memory feature makes them better at solving complex time-series forecasting problems [26–28]. Pak et al. [29] proposed a combined model involving a spatiotemporal convolutional neural network (CNN) and an LSTM network and applied it to predict the PM2.5 concentration of the next day in Beijing. Lin et al. [30] studied the training and prediction performance of neural networks using three models, a fast Fourier transform (FFT)–deep neural network, an FFT–LSTM network, and a one-dimensional CNN to predict the surface roughness in the milling process by combining vibration signals with a deep-learning prediction model. Kim et al. [31] developed a deep recursive neural-network system based on an LSTM model and applied it to the daily prediction of PM10 and PM2.5 in Korea. Their results showed that the LSTM model had better performance than a 3D chemistry-transport model. Zhang et al. [32] proposed an LSTM-network prediction method based on actual monitoring data from coal mine production. Their results showed that the LSTM model predicts more accurately than the traditional model. Kim et al. [33] proposed using an LSTM model to predict the hourly fine dust concentration of a target location in Seoul. As these sources show, LSTM models have been widely used in dust-concentration prediction. However, these studies mainly consider the influence of time on changes in dust concentration and cannot consider the influence of different weightings of input variables. The weighting relationships between the input variables are crucial for improving the accuracy of the prediction results.

The attention mechanism can be used to apply adaptive attention to input data or features in a neural network and improve its feature-extraction ability [34–36]. Li et al. [37] proposed a new spatiotemporal prediction model based on densely connected convolutional networks and attention LSTM. Wang et al. [38] proposed a new attention model based on an LSTM encoder structure to predict long-term time series. Their experimental results showed that the accuracy of the model was at least 2% better than the latest benchmark. Li et al. [39] proposed a new LSTM code model based on hierarchical time attention for individual position sequence prediction. By incorporating the calendar period of individual travel schedules into location predictions, frequent and periodic movement patterns were revealed. Zou et al. [40] predicted the water quality of Beilun Estuary based on a multi-time-scale two-way LSTM network. Their experimental results showed that the model was superior to single LSTM or bidirectional LSTM in predicting water quality. When attention-related inputs are different, the attention mechanism can focus on different parts of the input features to better extract the features, which can improve the accuracy of the prediction model.

The above research shows that although there has been significant study of the prediction of dust concentrations, the techniques are still imperfect. Furthermore, our literature review found that no specific studies examining the regional dust concentrations around open-pit mines have been carried out. Due to the different meteorological conditions in various regions, the changes in dust concentrations around different mines will also vary. It is therefore necessary to conduct

targeted research into the dust-concentration problem. Shanxi province is a large coal-producing province in China, and the Pingshuo Anjialing open-pit coal mine is representative of mines in the area. Since this mine is close to residential areas, air pollution from it is bound to affect the health of residents. This study first TSP concentration changes in the Anjialing open-pit coal mine. Then the attention mechanism was used to tap the advantages of the hidden information in the input parameters. LSTM-Attention was used with a view to improving the accuracy and stability of predictions of TSP concentration. Finally, the results of the proposed LSTM-Attention model and other models (ARIMA and LSTM) were compared and evaluated. The results of this study have reference value for the prediction of dust concentrations in the field of environmental research and mining engineering and the adoption of dust-reduction measures.

Materials and methods

Area description

The Pingshuo mining area is located in the Pinglu District of Shuozhou City, Shanxi Province, China. It is one of the five largest open-pit mining areas in China, and is in a warm, temperate, semi-arid climate zone. The annual average temperature is 6.5 °C, the extreme minimum temperature is −32.4 °C, and the extreme maximum temperature is 37.9 °C. The highest, lowest, and average annual precipitation are 757.4 mm, 195.6 mm, and 428.00 mm, respectively, and this mostly falls in July, August, and September, months which account for 75% of the annual precipitation. The dominant wind direction in the region is northwest, with an average annual wind speed of 1.8 m/s and a maximum of 10 m/s.

The Anjialing open-pit coal mine is the largest open-pit mine in the Pingshuo mining area. Its annual output is 2000 Mt/year, and the average amount of rock stripped is about 9000 Mm^3/year. The distances between the three nearby villages and the Anjialing open-pit coal mine are 1.6 km, 2.0 km, and 2.8 km. These villages are located downwind from the mine, as shown in Fig. 1, and environmental pollution from the mine is therefore bound to affect them.A large amount of dust is generated during coal production at the mine, and because dust pollution lasts for a long time, the pollution situation is complicated and has a wide scope. The main areas producing dust pollution are mining operations, transportation operations, and disposal operations. An example of the dust produced during these operations is shown in Fig. 1. The main hazards of dust pollution at the mine itself are workers developing pneumoconiosis, a reduction of visibility potentially causing more workplace accidents, and the dust increasing mechanical losses. The proximity of the mine to local villages will also inevitably have an impact on their air quality, harming the health of villagers. Therefore, it is necessary to predict dust concentrations to facilitate the timely implementation of reasonable control measures.

Data collection

We installed seven online dust monitoring stations at the Anjialing open-pit coal mine to monitor dust concentration. Because the prevailing wind direction at the mine is northwest, we deployed three of these in the upwind direction and four downwind. These locations and a photograph of one of the monitoring stations are shown in Fig. 2.

The recorded monitoring data included TSP, PM10, PM2.5, temperature (*T*), relative humidity (RH), wind direction (WD), wind speed (*V*), air pressure (*P*). The equipment parameters used in this monitoring process are shown in Table 1.

We conducted continuous monitoring at the Anjialing open-pit mine from 10 September 2019 to 9 September

Fig. 1 Location map of the study area (left) and a photograph showing the generation of dust pollution during mining operations (right)

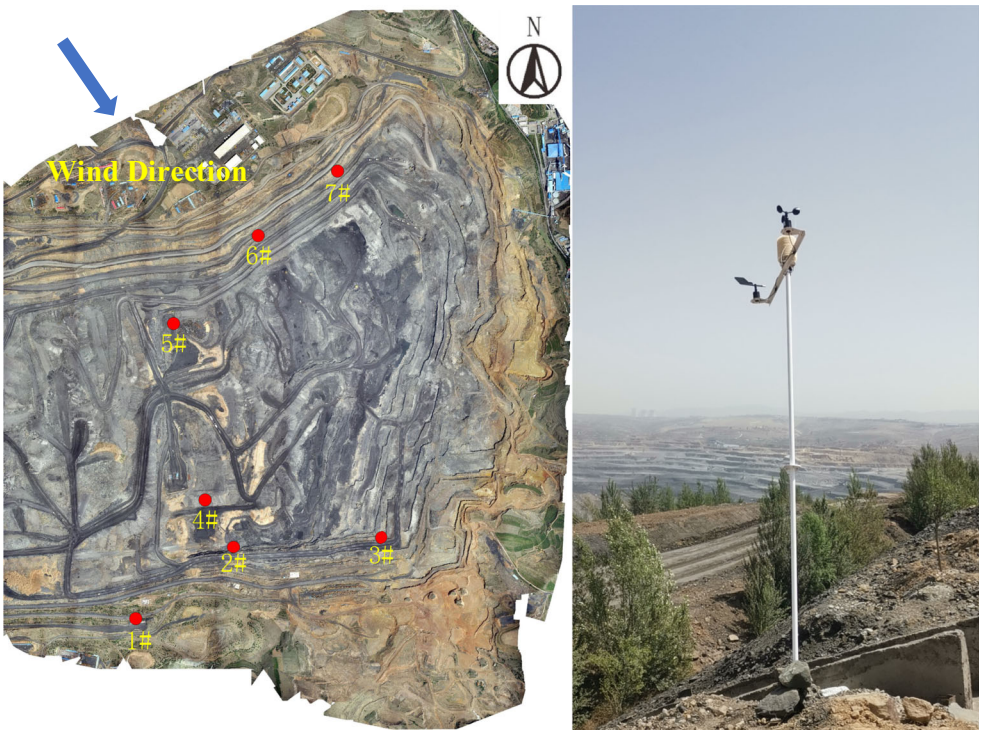

Fig. 2 Locations of the monitoring stations (left) and an example photograph of the monitoring equipment

2020. The collected data were preprocessed so as to improve their quality. There will be gaps in the data collected from air-quality monitoring equipment due to equipment failure and other uncontrollable factors. When there were several missing values in a data record, we directly removed them; when there was only one missing value, linear interpolation was used to fill this gap [32]. Four-hour average concentration values were used for the data sets. After data processing, we obtained 2179 sets of valid data, as shown in Table 2. The average, median, and mode are statistics that reflect the central tendency of the data. The standard deviation is used to evaluate the degree of variation in the data distribution; skewness and kurtosis are used to evaluate whether the sampling distribution is normal; the minimum and maximum values indicate the magnitude of changes in the data.

It can be seen from Table 2 that the average concentration of PM2.5 was 60.89 μg/m^3 and the average concentration of PM10 was 114.25 μg/m^3. This indicates that the air quality was good. However, the highest concentration of PM2.5 reached 225.62 μg/m^3 and the highest concentration of PM10 reached 405.64 μg/m^3. This indicates that the air was moderately polluted. Therefore, prediction of dust concentrations is crucial so that a convenient dust-reduction plan can be prepared.It can also be seen from Table 1 that the high standard deviation of the TSP concentration indicates that the data is rather dispersed, which increases the difficulty of predicting this value.

ARIMA method

The time-series analysis method is mainly used to summarize and reveal the features of a time series using autocorrelation [41]. Depending on whether the original time series is stable and which part is contained in the regression, it can be divided into an autoregressive process AR(p), a moving-average process MA(q), an autoregressive moving-average process

Table 1 Monitoring equipment parameters

Parameter	Value
Working environment	−20 °C to 60 °C
Built-in sensors	T, RH, P, TSP, PM10, PM2.5, V, WD
Sensor accuracy	T: ±0.01 °C, RH: ±2%, P: 0.2 Pa, TSP: measurement range: 0.01–10,000 μg/m^3; measurement accuracy: ±10%
PM2.5+PM10	Effective range: 0–1000 μg/m^3
Wind speed	Measurement range: 0–30 m/s, measurement accuracy: ±1 m/s
Wind direction	Range of wind direction: 0–360°, measurement accuracy: ±3°

Table 2 Characteristics of the data used

Statistic	TSP (μg/m^3)	PM10 (μg/m^3)	PM2.5 (μg/m^3)	T (°C)	RH (%)	V (m/s)	P (kPa)
Mean	250.30	114.25	60.89	9.07	60.85	1.44	86.53
Median	231.98	105.55	56.06	11.77	60.10	1.00	86.60
Mode	227.42	65.73	31.96	−2.00	48.30	0.00	86.80
Standard deviation	121.08	55.37	31.59	9.84	20.53	1.50	0.57
Skewness	1.20	1.25	1.21	−0.65	0.02	1.11	0.13
Kurtosis	2.55	2.64	2.47	−0.27	−0.94	0.86	−0.60
Min.	17.18	11.66	8.02	−23.00	14.47	0.00	85.30
Max.	896.36	405.64	225.62	34.17	98.97	8.80	88.10

ARMA(p,q), or a summation autoregressive moving average process ARIMA(p,d,q). For a stationary series, after the series difference is processed, we can use the ARIMA model to perform fitting prediction. This model is described mathematically as:

$$\Delta^{\mathrm{d}} y_t = \theta_0 + \sum_{i=1}^{p} \phi_i \Delta^d y_{t-1} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-1} \quad (1)$$

where: y_t is the original time series; $\Delta^d y_t$ represents the stationary series of y_t after d difference; ε_t represents a white-noise random-error sequence with zero mean; ϕ_i ($i = 1, 2, \ldots, p$) and θ_j ($j = 1, 2, \ldots, q$) are parameters to be estimated for each model; and p and q are the orders of the models.

LSTM method

In a standard recurrent neural network (RNN), the state of the hidden layer at each moment is determined by the current input and the previous state of the hidden layer. Because the memory capacity is limited, any gradient disappears rapidly [32 42]. An LSTM network is an improved version of an RNN. Its main features are a forget gate, an input gate, and an output gate. The LSTM model architecture is shown in Fig. 3, in which σ represents the sigmoid activation function, the output range is 0–1, and tanh is used to adjust the value of the function. The output range is between −1 and 1.

The forget gate is used to control how the state of the previous moment is retained in the current unit state to realize screening of the memory content. The output of the forget gate f_t is:

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

where: W_{xf} is the weight matrix between the input and the forget gate; W_{hf} is the weight matrix between the historical output and the forget gate; b_f is the forget gate bias term; x_t is the current input state; and h_{t-1} is the output state at the previous moment.

The input gate is used to update the unit state, input the previous state and the current input information into the sigmoid activation function, and obtain an output value between 0 and 1 to determine the updating of information. In this system 0 indicates *not important* and 1 indicates *important*. The previous state and input information are

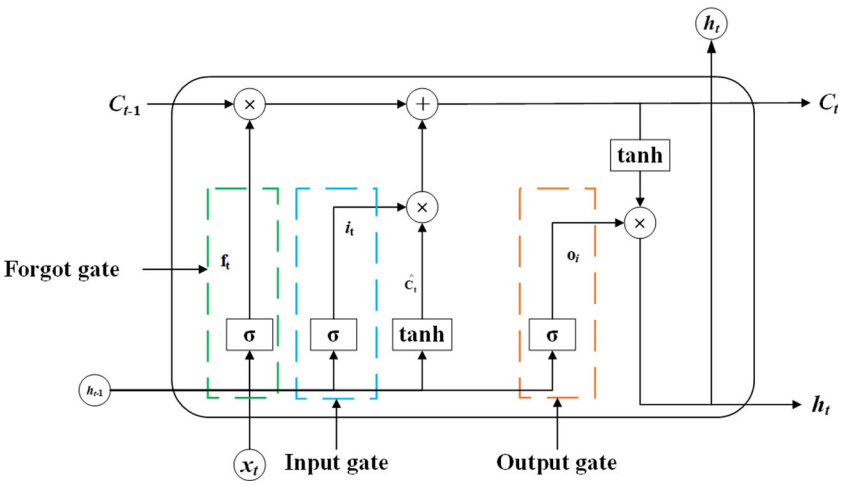

Fig. 3 Diagram of LSTM model architecture

fed into the tanh function to obtain a candidate unit state between −1 and 1. The input gate output i_t and the candidate unit status $\widehat{c}_t$ are:

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

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

where: W_{xi} is the weight matrix between the input and the input gate; W_{hi} is the weight matrix between historical output and input gate; W_{xc} is the weight matrix between the input and the unit state; W_{hc} is the weight matrix between the historical output and unit state; b_i is the input gate bias term; and b_c is the unit state bias term.

After obtaining the outputs of the forget gate and the input gate, the current unit state c_t is obtained by adding two things. One is the information retention determined by the multiplication of the output of the forget gate and the unit state at the previous moment, and the other is the information addition determined by the multiplication of the input gate and the current candidate state. The unit status is:

$$C_t = f_t C_{t-1} + i_t \widehat{c}_t. \tag{5}$$

The output gate is used to control the final output of the unit state. The unit output is obtained by multiplying the output gate output and the current unit state information through tanh activation function. The output gate o_t and the unit output h_t are:

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

$$h_t = o_t \tanh(C_t), \tag{7}$$

where: W_{xo} is the weight matrix between the input and output gates; W_{ho} is the weight matrix between historical output and the output gate; and b_o is the output gate bias term.

Attention method

In this study, an attention-mechanism layer was added to better capture the effective information in the data. This attention-mechanism layer overcomes the problem of a standard LSTM model not being able to fully learn detailed information about the sequence encoding when the same state vector is used for each prediction. The attention mechanism of the neural network can be described as follows [34]:

$$M = \tanh(W_h H), \tag{8}$$

$$\alpha = soft\max\left(\omega^T M\right), \tag{9}$$

$$r = H\alpha^T, \tag{10}$$

where: $W \in R^{d \times N}$, $\alpha \in R^N$, $r \in R^d$, $W_h \in R^{d \times d}$, and $w^T \in R^d$ is the parameter matrix that the subsequent model needs to train. $H \in R^{d \times N}$ represents the matrix composed of the LSTM network output vector $[h_1, h_2, \ldots, h_N]$. The attention mechanism will eventually generate attention weight vector a and r.

LSTM-attention method

For a given sequence, the *importance* of each element will be different; that is, the weight of each element will be different. This study used an attention mechanism to retain the intermediate output results of the input sequence of the LSTM encoder. A model was then trained to selectively learn these inputs and associate the output sequence with it. The LSTM-Attention model architecture as shown in Fig. 4.

In Fig. 4, $[h_1, h_2, \ldots, h_t]$ is equivalent to H in Eq. (8); $[\alpha_1, \alpha_2, \ldots, \alpha_t]$ is equivalent to α in Eq. (9), which is the attention weight of feature H, and r is calculated using Eq. (10). The dust-concentration prediction output is located at the end of

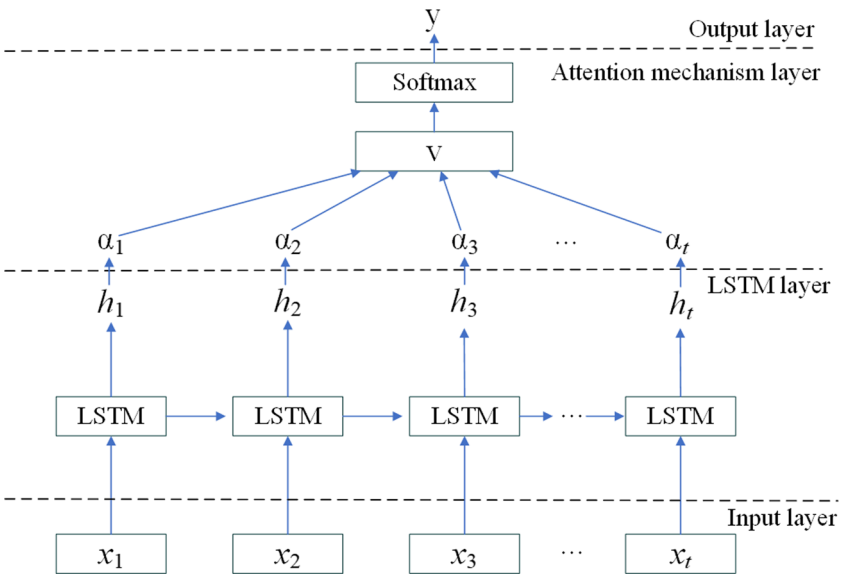

Fig. 4 Diagram of the LSTM-Attention model architecture

the network architecture diagram. It is a single-layer, fully-connected neural network. The prediction formula is as follows:

$$h_s = tanh\left(W_p r + W_x h_t\right) \tag{11}$$

$$y = W_s h_s + b_s \tag{12}$$

Accuracy evaluation

Four evaluation indicators, the root mean square error (RMSE), mean absolute percentage error (MAPE), mean absolute error (MAE), and correlation coefficient (R^2), were used to evaluate the accuracy of the TSP dust-concentration prediction model. These can be calculated using:

$$RMSE = \sqrt{\frac{1}{l}\sum_{i=1}^{l}(O_i - P_i)^2}, \tag{13}$$

$$MAPE = \frac{1}{l}\sum_{i=1}^{l}\frac{O_i - P_i}{O_i}, \tag{14}$$

$$MAE = \frac{1}{l}\sum_{i=1}^{l} O_i - P_i, \tag{15}$$

$$R^2 = 1 - \frac{\sum_i (O_i - P_i)^2}{\sum_i \left(O_i - \overline{O}\right)^2} \tag{16}$$

In Eqs. (13)–(16), O_i, P_i, and $\overline{O}$, represent the observed, predicted, and average values of the output, respectively, and l is the number of evaluation samples.

The RMSE and the MAE are used to evaluate the absolute error; the smaller the values of these parameters, the better the performance of the model. The MAPE is a measure of relative error; the smaller its value, the closer the predicted result is to the actual value. The value of R^2 represents the ratio of the explained sum of squares and is between 0 and 1. This can be used as a measure of the goodness of fit of the regression equation; the larger the value of R^2, the better the fitting effect.

Hybrid prediction algorithm

Here, we propose a hybrid algorithm using an attention mechanism with an LSTM to improve the accuracy of dust-concentration predictions in open-pit mines. To consider the influences of different factors on the concentration of TSP as much as possible, we used time and date, PM10, T, RH, WD, V, and P as independent variables for TSP concentration prediction. The inputs to the model are PM10, T, RH, WD, V, and P their corresponding time series every 4 h on average; the output based on the weather conditions and air-quality data from the previous 4 h, is a prediction of the average TSP concentration in the next 4 h. The accuracy of the model is then evaluated. The steps we took were as follows (Fig. 5):

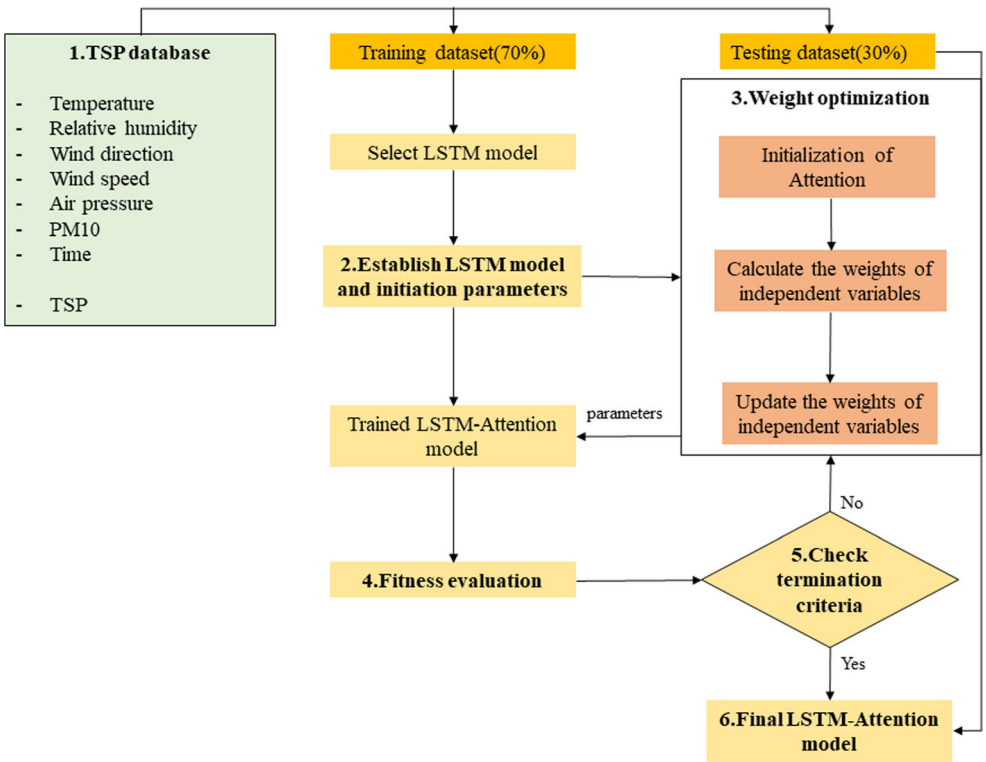

Fig. 5 The proposed LSTM-Attention model for estimating TSP concentration

Step 1. Standardize raw data (PM10, *T*, RH, WD, *V*, and *P* and the corresponding time series every 4 h on average) and perform data preprocessing;
Step 2. Observe parameters and analyze changes in them;
Step 3. Turn the dataset into a supervised learning problem;
Step 4. Construct traditional prediction models (ARIMA and LSTM) and a combined LSTM-Attention prediction model; sequentially carry out TSP concentration prediction and related analysis;
Step 5. Evaluate the prediction accuracies of the traditional and hybrid models.

Results

Influence of air-quality data on TSP concentration prediction

To evaluate the impact of air-quality data on the accuracy of TSP forecasts, we separately input meteorological data alone and meteorological data along with air-quality data (PM2.5) into the ARIMA model. The results of this analysis are shown in Table 3.

From the results in Table 3, we can see that the air-quality data improve the TSP-prediction performance. Based on the difference in the R^2 value, the accuracy of the model including air-quality data is 22% higher than that of the model only using meteorological data. The inclusion of air-quality data therefore notably improves the accuracy of the model.

Variation of dust concentration

Figure 6a shows a graph of the concentrations of PM2.5, PM10, and TSP over the observation period. To examine diurnal variations in the TSP concentration, a TSP concentration map based on data from 00:00 on 16 September 2019 to 00:00 on 17 September 2019 was drawn, as shown in Fig. 6b.

It can be seen from Fig. 6a that the trends in the changes in the PM2.5, PM10, and TSP concentrations are similar and there is a certain correlation between them. Additionally, Fig. 6a indicates that the dust concentrations exhibit a high degree of daily fluctuation. This can be seen more clearly in Fig. 6b, which shows that the dust concentration is high in the morning and low in the afternoon. The dust concentration reaches a maximum value at around 00:00–01:00 every day, before falling to a minimum value at around 05:00–06:00 and then starting to rise again. After reaching another maximum around 08:00–09:00, it begins to decline, and reaches a minimum around 16:00–17:00 and starts to rise again before reaching the maximum around 00:00–01:00 once more.

In addition to normal operations and the influence of meteorological parameters, the main reason for the above trends in dust concentration is the inversion layer above the open-pit mine. With the increase in the mining depth, the open-pit mine stope has a concave structure. When the wind is calm or there is a slight breeze, under the heat mainly caused by solar radiation, the particular topographic structure of the stope makes it easy to form an inversion layer. Due to the existence of the temperature-inversion layer, the upper part of the pit dissipates heat quickly and the cold air circulates from the pit edge to sink to the bottom of the pit. The warmer air at the bottom of the pit is pushed up by the cold air, and the temperature rises after the convection are completed, resulting in the inversion of high and low temperatures. As a result, the dust concentration of the open-pit mine continuously decreases during the day and rises at night. A temperature-inversion layer is more likely to appear in the period 00:00–08:00, and the probability of inversion decreases in the period 09:00–10:00. It will gradually weaken or even disappear in the period 12:00–16:00. This has an impact on the dust concentration.

Since the PM2.5, PM10, and TSP concentrations are correlated, only PM2.5 or PM10 needs to be selected as one of the independent variables. We applied both PM2.5 and PM10 as independent variables and used the LSTM network to predict the concentration of TSP so that it could be established which of the two will give greater prediction accuracy. The results of this analysis are shown in Table 4.

It can be seen from Table 4 that the errors in the TSP concentration prediction using PM10 as an independent variable are smaller than those using PM2.5. Therefore, PM10 was used as an independent variable to predict the TSP concentration.

Correlations between influencing factors

Figure 7 shows a matrix of graphs of the relationships between the input variables. The matrix is symmetric, and the sub-

Table 3 Analysis of the impact including air-quality data on TSP predictions

Influencing factors	Learning paradigm	RMSE (μg/m^3)	MAE (μg/m^3)	MAPE (%)	R^2
Meteorological data	ARIMA	61.853	40.303	18.639	0.739
Meteorological data and air quality		31.393	21.106	11.842	0.902

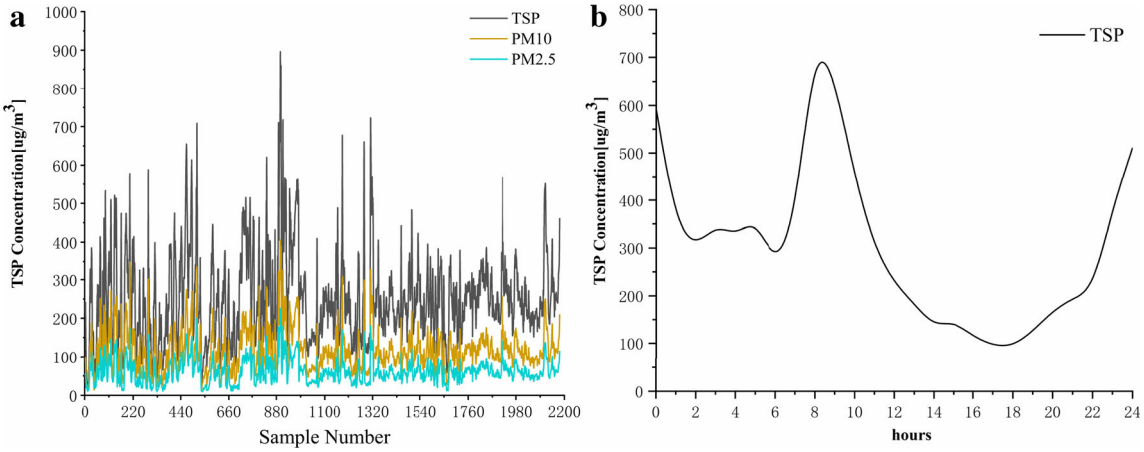

Fig. 6 Changes in dust concentrations; **a** variations of PM2.5, PM10, and TSP concentration over the study period; **b** changes in TSP concentration over 16 September 2019

graphs on the diagonals represent the internal relationships of the variables while the sub-graphs on the non-diagonal elements represent the correlations between variables.

It can be seen from Fig. 7 that most of the input variables are not highly correlated, indicating that the parameters are reasonable as variables. The graphs on the diagonal line show the interval distributions of each variable. Taking PM10 as an example, it can be seen from that the concentration of PM10 corresponding to the abscissa of 65 μg/m^3 has the largest proportion, and the proportion above 300 μg/m^3 is the smallest. This shows that the PM10 concentrations are mainly clustered around 65 μg/m^3. The other variables here show similar tendencies.

Forecasting of TSP with ARIMA and LSTM

The ARIMA [21, 41] and LSTM [26, 28, 32] methods were used to establish models for predicting the TSP concentration, and the performances of the two methods were compared. According to the principle of using 70% of the data for model training and 30% of the data for model testing, in this study, the first 1525 sets of data among the 2179 sets of data were used as training samples for model training, and the remaining 654 were used as test samples to verify the model. The prediction results from the ARIMA and LSTM models are shown in Fig. 8.

It can be seen from Fig. 8a that both the ARIMA model and the LSTM model can be used to predict the TSP concentration, and both generally give good results. However, at the positions of peaks, their predictions are comparatively poor, and the predicted values are too low. It can be seen from Fig. 8b that the correlation coefficients between the predicted and the actual values of the two models are greater than 0.9 (specifically 0.922 and 0.946), indicating that the prediction abilities of the two models are similar and generally good. But when the dust concentration is less than 180 μg/m^3 or greater than 300 μg/m^3, the prediction errors of the two algorithms begin to increase.

Forecasting of TSP with ARIMA and LSTM-attention

The ARIMA [21, 41] and LSTM-Attention methods were used to establish models for predicting the TSP concentration, and the performances of the two methods were compared. Again, the first 1525 sets of data were used as training samples and the remaining 654 were used as test samples to verify the model. The results are shown in Fig. 9.

Figure 9a shows that that both the ARIMA and LSTM-Attention model can be used to predict the TSP concentration, and both give reasonable results. Again, the under-prediction of the ARIMA model at peaks can be seen. The LSTM-Attention model, however, predicts the peaks significantly better. Figure 9b shows that the correlation coefficient of the values predicted by the LSTM-Attention model is 0.986, very much higher than that of the ARIMA model. When the dust concentration is lower than 180 μg/m^3, the prediction ability of the ARIMA model is only average, and the error in the prediction is higher., The LSTM-Attention model, however, has a small overall error and only gives small deviations from the actual values.

Table 4 Analysis of prediction errors with PM2.5 and PM10 as the independent variable

Independent variable	Learning paradigm	RMSE (μg/m^3)	MAE (μg/m^3)	MAPE (%)
PM2.5	LSTM	29.517	21.215	11.573
PM10		23.204	19.435	8.537

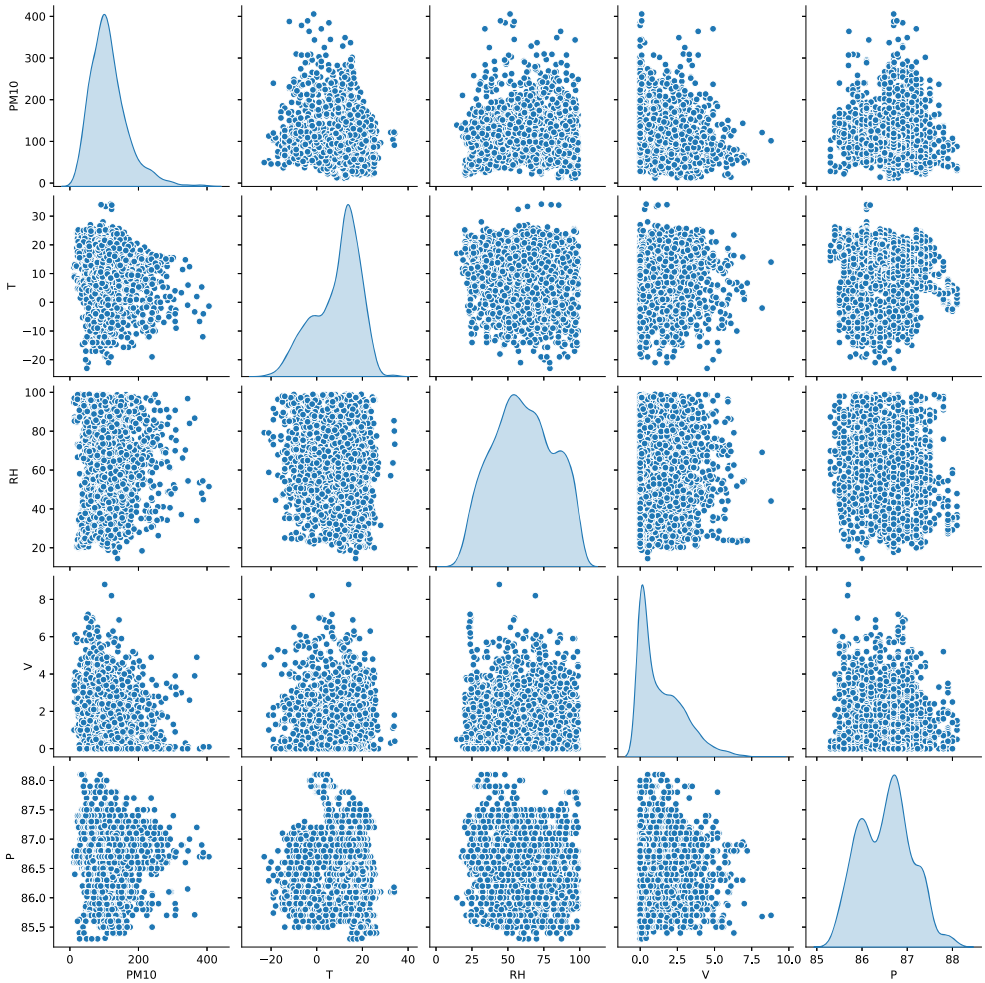

Fig. 7 Diagram of the relationship between variables

Forecasting of TSP with LSTM and LSTM-attention

The LSTM [26, 28, 32] and LSTM-Attention methods were used to establish models for predicting the TSP concentration, and the performances of the two methods were compared. Once more, the first 1525 sets of data were used as training samples and the remaining 654 were used as test samples to verify the model. The results are shown in Fig. 10.

Similar to the analysis in the previous section, it can be seen from Fig. 10a that both the LSTM model and the LSTM-Attention model can be used to predict the TSP concentration, and both give good results. However, once more, in contrast to the LSTM-Attention model, at the positions of peaks, the overall predictions of the LSTM model are too large. This is again reflected in the correlation coefficients. When the dust concentration is lower than 300 μg/m^3, the errors of the two prediction algorithms are small. When the dust concentration is higher than 300 μg/m^3, the errors of the two prediction algorithms increase, but the prediction error of the LSTM model is greater than that of the LSTM-Attention model, which further shows that the LSTM-Attention model is better than the LSTM model.

Comparative analysis of forecast results

In this study, the RMSE, MAE, MAPE, and R^2 values were selected as indicators to evaluate the accuracy of the prediction model. A comparative analysis of the prediction results of the three models is shown in Table 5.

It can be seen from Table 5 that all three prediction models show good prediction ability. On the whole, the predictions of the LSTM-Attention model are better than those of the ARIMA model. This is related to the advantages of the LSTM model. The LSTM model can handle a large amount of air-quality and meteorological data, and even long-term sequence data without gradual changes. In other words, the LSTM model can fully reflect the historical process data in the input time series, which is the advantage of the LSTM model over the ARIMA model. Furthermore, the predictions of the LSTM-Attention model are better than those of the LSTM model. This shows that when the attention mechanism is added to the LSTM model, the inherent characteristics of the input data can be effectively extracted, and the prediction performance can be improved. Based on our proposed hybrid algorithm, the values of RMSE, MAE, and MAPE are

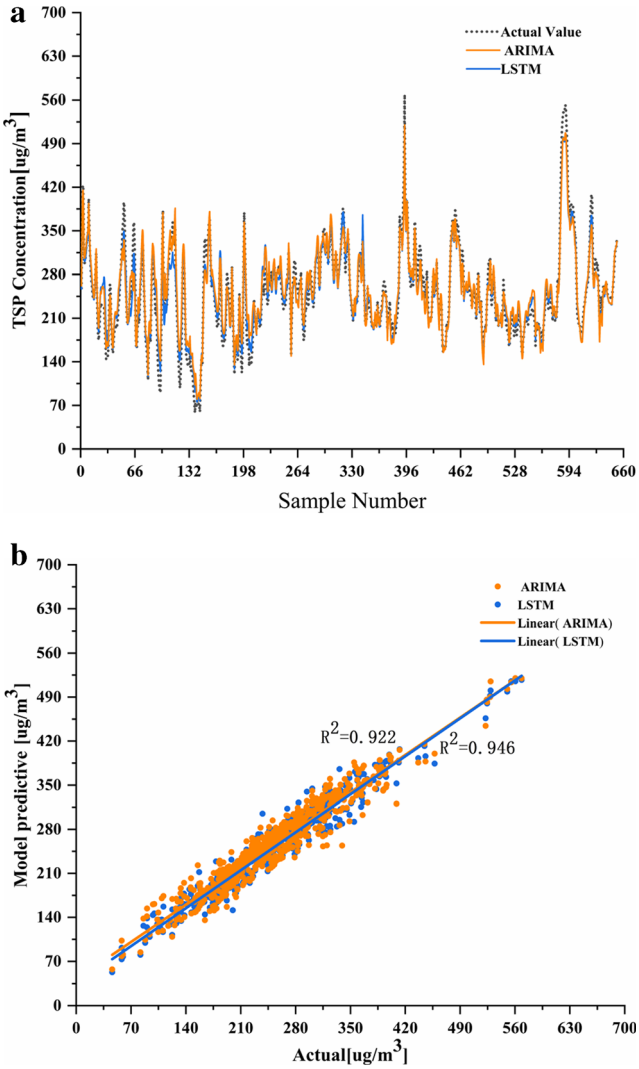

Fig. 8 Forecasting results from the ARIMA and LSTM models. **a** Graph of predicted changes in TSP concentration. **b** Graph of correlation coefficient analysis

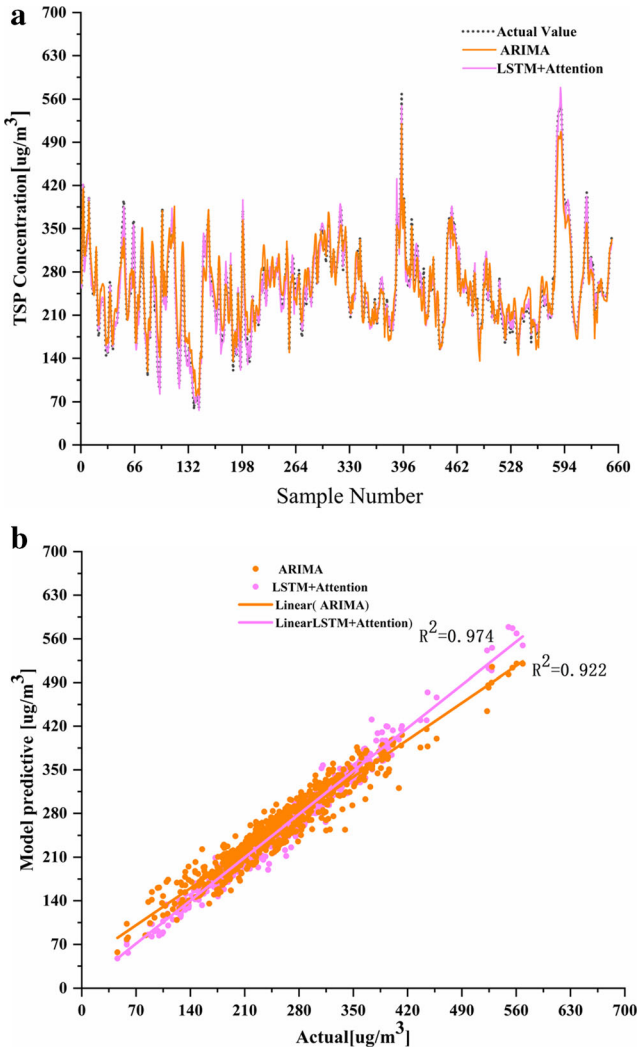

Fig. 9 Forecasting results from the ARIMA and LSTM-Attention models. **a** Graph of predicted changes in TSP concentration. **b** Graph of correlation coefficient

significantly reduced compared to the benchmark model, and R^2 has also been significantly improved.

Discussion

Sensitivity analysis

To verify the impact of small data changes on the prediction results, a sensitivity analysis of the three methods was conducted. The concentration of PM10 can have a large influence on dust-concentration prediction. Since the PM10 concentration interval 70 μg/m^3–300 μg/m^3 accounts for 2/3 of the overall data volume, this range was chosen for conducting a sensitivity analysis of this parameter. The PM10 value was varied in this range with a step size of 1, and different dust concentration values were obtained. The results are shown in Fig. 11.

It can be seen from the sensitivity analysis in Fig. 11 that in 231 tests, the value of dust concentration predicted using the LSTM-Attention method had relatively little change, with a maximum change rate of 0.18. In contrast, the maximum change rate of the value of dust concentration predicted using the LSTM method was 0.26, and the value of dust concentration predicted using the ARIMA method had relatively large changes, with a maximum rate of change of 0.29. This shows that LSTM-Attention is the least sensitive to changes in the input value, but it also has a certain sensitivity, with small changes and a certain degree of stability. This model can effectively avoid the problem of inaccuracy in the predicted concentration value caused by algorithm defects.

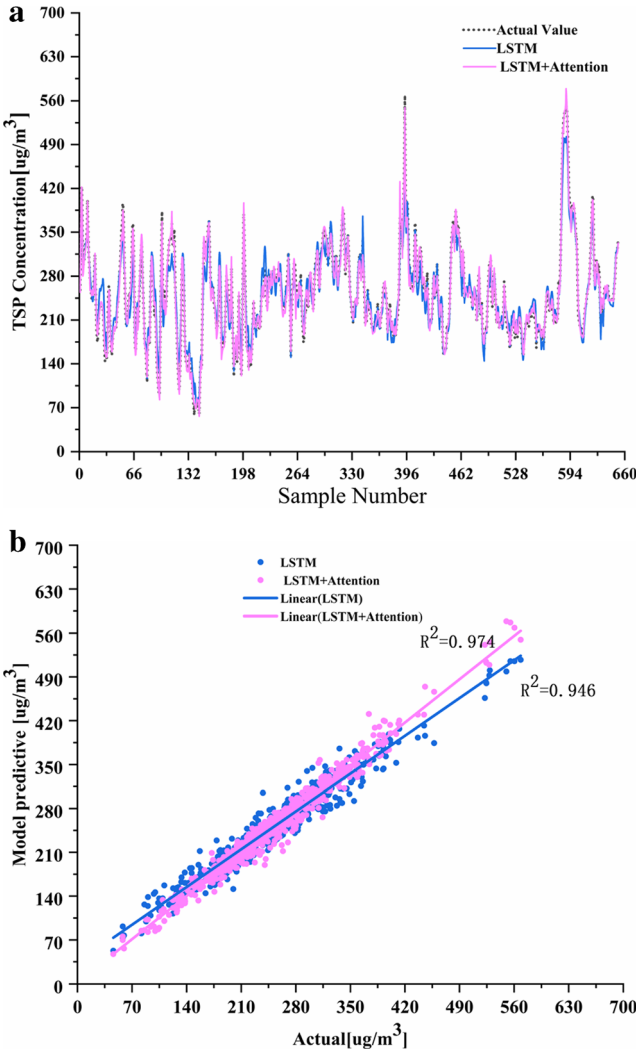

Fig. 10 Forecasting results from the LSTM and LSTM-Attention models. **a** Graph of predicted changes in TSP concentration. **b** Graph of correlation coefficient analysis

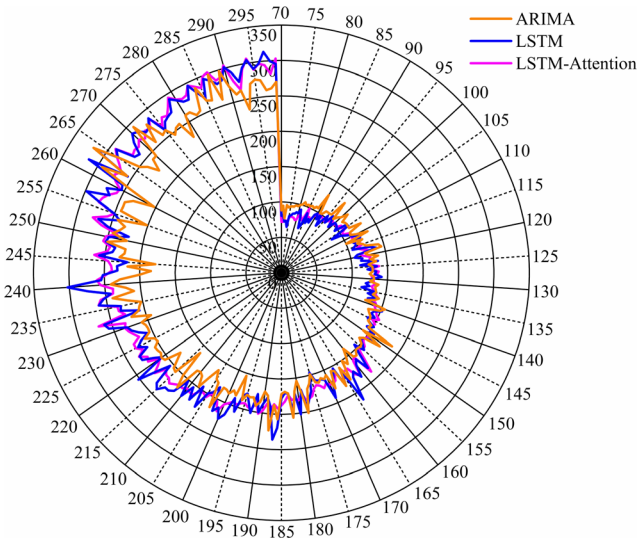

Fig. 11 Sensitivity analysis of the effects of changing PM10 concentration on the three algorithms

Model comparison

To verify the feasibility and effectiveness of the proposed LSTM-Attention model, we compared it with three other models: simple RNN [16, 17], ARIMA [21, 22], and LSTM [26–28]. After training all models with the same training set, according to ambient air quality standards (GB3095–2012), we divided 654 sets of validation data into two groups: those with TSP concentration less than 300 μg/m^3 and those with TSP concentration greater than 300 μg/m^3. A performance evaluation was then performed on each model, and the results are shown in Tables 6 and 7.

It can be seen from Tables 6 and 7 that with higher concentrations, the predictive abilities of all models become worse. This conclusion is consistent with the previous analysis. The three algorithms based on time series (ARIMA, LSTM, and LSTM-Attention) are better than the algorithm based on non-time series (Simple RNN). However, when the attention mechanism is added to the LSTM model, the overall difference in prediction accuracy for different concentrations is minimal. This result is consistent with the results of the sensitivity analysis and also further indicates the superiority of the LSTM-Attention model.

Conclusion

In this study, the temporal characteristics of dust concentration in an open-pit mine and its variation within the 24-h cycle were revealed. The TSP concentration was predicted using meteorological data, time data, and a deep-

Table 5 Comparative analysis of forecast results

Learning paradigm	RMSE (μg/m^3)	MAE (μg/m^3)	MAPE (%)	R^2
ARIMA	24.797	18.905	10.604	0.922
LSTM	20.298	16.743	7.002	0.946
LSTM-Attention	16.298	13.028	6.108	0.974

Table 6 Comparative analysis of forecast results for data with TSP concentration < 300 μg/m^3

Learning paradigm	RMSE (μg/m^3)	MAE (μg/m^3)	MAPE (%)
Simple RNN	30.718	22.623	14.059
ARIMA	23.652	19.523	10.102
LSTM	22.392	18.960	8.349
LSTM-Attention	21.831	18.061	7.539

Table 7 Comparative analysis of forecast results for data with TSP concentration > 300 μg/m^3

Learning paradigm	RMSE (μg/m^3)	MAE (μg/m^3)	MAPE (%)
Simple RNN	35.752	25.236	15.267
ARIMA	26.952	20.423	12.153
LSTM	24.897	20.777	9.513
LSTM-Attention	22.204	18.498	7.863

learning method. The training and prediction performances of the ARIMA, LSTM, and LSTM-Attention models were evaluated. All three models had good training performance under the same conditions, with $R^2 >$ 0.90 and MAPE <12%. Comparing the prediction results, the prediction ability of ARIMA was the worst, and LSTM was second. However, the attention mechanism was able to extract complex features to provide better prediction ability and improve prediction accuracy, and the LSTM-Attention model had the best prediction ability and the highest stability.

The results of this study indicate the following. (1) The average concentrations of PM2.5, PM10, and TSP in the Anjialing open-pit coal mine over the study period were 60.89 μg/m^3, 114.25 μg/m^3, and 250.30 μg/m^3, respectively. The dust concentration changes were the highest in the period 08:00–09:00 and the lowest in the period 15:00–16:00. In addition to the influence of meteorological parameters and normal operations, the reason for this trend is the inversion layer above the mine. (2) The performance evaluation results show that the prediction accuracy is significantly improved, by 22%, when both air-quality and meteorological data are used, as compared to the case of using meteorological data alone. (3) The proposed hybrid algorithm combines an LSTM model with the attention mechanism. When compared with the traditional ARIMA and LSTM models, this hybrid algorithm improves the overall prediction accuracy by 5.6% and 3.0%, respectively. Therefore, it is suggested that the LSTM-Attention model should be used to predict TSP concentration in open-pit mines.

However, this study still has the following deficiencies. (1) Because the inversion layer distribution is related to the dust concentration of the open-pit mine,, we recommend further study of the factors influencing the formation and dissipation of the inversion layer and their relationships with open-pit mine dust concentrations [43]. (2) On the basis of the current research, we will conduct further study into the relationship between field operations and the distribution of dust concentration. (3) With reasonable design modifications, the hybrid algorithm proposed in this paper is a good alternative for the prediction of other air pollutants in other areas.

Authors' contributions Author Contributions: Conceptualization, L.L., RX.Z. and JD.S.; methodology, L.L. and RX.Z.; software, L.L. and Q.H.; validation, L.L. and LZ.K.; formal analysis, L.L. and LZ.K.; investigation, L.L. and X.L. resources, L.L. and RX.Z.; data curation, L.L. and LZ.K.; original draft preparation, L.L., LZ.K. and Q.H.; manuscript review and editing, L.L., RX.Z., LZ.K., and Q.H.; visualization, L.L. and LZ.K.; supervision, RX.Z.; project administration, RX.Z.; funding acquisition, RX.Z. All authors have read and agreed to the published version of the manuscript.

Funding This research was funded by the National Key R&D Program, grant number 2018YFC0808306.

Data availability The data used to support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

1. Liu D, Li L. Application study of comprehensive forecasting model based on entropy weighting method on trend of PM2. 5 concentration in Guangzhou, China[J]. Int J Environ Res Public Health. 2015;12(6):7085–99. https://doi.org/10.3390/ijerph120607085.
2. Csavina J, Field J, Félix O, Corral-Avitia AY, Sáez AE, Betterton EA. Effect of wind speed and relative humidity on atmospheric dust concentrations in semi-arid climates[J]. Sci Total Environ. 2014;487:82–90. https://doi.org/10.1016/j.scitotenv.2014.03.138.
3. Liang X, Zou T, Guo B, et al. Assessing Beijing's PM2. 5 pollution: severity, weather impact, APEC and winter heating[J]. Proc Royal Soc A Math Phys Eng Sci. 2015;471(2182):20150257. https://doi.org/10.1098/rspa.2015.0257.
4. Afrad M S I, Monir M B, Haque M E, et al. Impact of industrial effluent on water, soil and Rice production in Bangladesh: a case of Turag River Bank[J]. J Environ Health Sci Eng. 2020;18(2):825–834. https://doi.org/10.1007/s40201-020-00506-8.
5. Paithankar A, Chatterjee S, Goodfellow R, Asad MWA. Simultaneous stochastic optimization of production sequence and dynamic cut-off grades in an open pit mining operation[J]. Res Policy. 2020;66:101634.
6. Gonzalo Morera de la Vall Gonzalez. Dust production in mining. Suppression measures in quarry blasting[D]. College of mining and energy engineering: Department of Geology and mining engineering. 2018.
7. Zhang Y, Zhang Y, Liu B, Meng X. Prediction of the length of service at the onset of coal workers' pneumoconiosis based on neural network[J]. Arch Environ Occup Health. 2020;75(4):242–50. https://doi.org/10.1080/19338244.2019.1644278.
8. Tripathy D P, Dash T R, Badu A, et al. Assessment And Modelling Of Dust Concentration In An Opencast Coal Mine In India[J]. Global Nest Journal. 2015;17(4):825–834.
9. Bray CD, Battye W, Aneja VP, Tong D, Lee P, Tang Y, et al. Evaluating ammonia (NH3) predictions in the NOAA National air Quality Forecast Capability (NAQFC) using in-situ aircraft and satellite measurements from the CalNex2010 campaign[J]. Atmos Environ. 2017;163:65–76. https://doi.org/10.1016/j.atmosenv.2017.05.032.
10. Zhou G, Xu J, Xie Y, Chang L, Gao W, Gu Y, et al. Numerical air quality forecasting over eastern China: an operational application of

WRF-Chem[J]. Atmos Environ. 2017;153:94–108. https://doi.org/10.1016/j.atmosenv.2017.01.020.
11. Aljerf L. Reduction of gas emission resulting from thermal ceramic manufacturing processes through development of industrial conditions[J]. Sci J King Faisal Univ. 2016;17(1):1–10.
12. de Gennaro G, Trizio L, Di Gilio A, et al. Neural network model for the prediction of PM10 daily concentrations in two sites in the Western Mediterranean[J]. Sci Total Environ. 2013;463:875–83. https://doi.org/10.1016/j.scitotenv.2013.06.093.
13. Li X, Zhang C, Zhang B, Liu K. A comparative time series analysis and modeling of aerosols in the contiguous United States and China[J]. Sci Total Environ. 2019;690:799–811. https://doi.org/10.1016/j.scitotenv.2019.07.072.
14. Yi L, Mengfan T, Kun Y, Yu Z, Xiaolu Z, Miao Z, et al. Research on PM2. 5 estimation and prediction method and changing characteristics analysis under long temporal and large spatial scale-a case study in China typical regions[J]. Sci Total Environ. 2019;696:133983. https://doi.org/10.1016/j.scitotenv.2019.133983.
15. Yuan W, Wang K, Bo X, Tang L, Wu J. A novel multi-factor & multi-scale method for PM2. 5 concentration forecasting[J]. Environ Pollut. 2019;255:113187. https://doi.org/10.1016/j.envpol.2019.113187.
16. Kukkonen J, Partanen L, Karppinen A, et al. Extensive evaluation of neural network models for the prediction of NO2 and PM10 concentrations, compared with a deterministic modelling system and measurements in Central Helsinki[J]. Atmos Environ. 2003;37(32):4539–50. https://doi.org/10.1016/S1352-2310(03)00583-1.
17. Wang X, Wang B. Research on prediction of environmental aerosol and PM2. 5 based on artificial neural network[J]. Neural Comput & Applic. 2019;31(12):8217–27. https://doi.org/10.1007/s00521-018-3861-y.
18. Choubin B, Abdolshahnejad M, Moradi E, Querol X, Mosavi A, Shamshirband S, et al. Spatial hazard assessment of the PM10 using machine learning models in Barcelona, Spain[J]. Sci Total Environ. 2020;701:134474. https://doi.org/10.1016/j.scitotenv.2019.134474.
19. Bagnall A, Flynn M, Large J, et al. Is rotation forest the best classifier for problems with continuous features?[J]. arXiv preprint arXiv:1809.06705, 2018.
20. Sutton CD. Classification and regression trees, bagging, and boosting[J]. Handbook Stat. 2005;24:303–29. https://doi.org/10.1016/S0169-7161(04)24011-1.
21. Zhang H, Zhang S, Wang P, Qin Y, Wang H. Forecasting of particulate matter time series using wavelet analysis and wavelet-ARMA/ARIMA model in Taiyuan China. J Air Waste Manage Assoc. 2017;67(7):776–88. https://doi.org/10.1080/10962247.2017.1292968.
22. Bui X, Lee C, Nguyen H, et al. Estimating PM10 Concentration from Drilling Operations in Open-Pit Mines Using an Assembly of SVR and PSO[J]. Appl Sci. 2019;9(14). https://doi.org/10.3390/app9142806.
23. Li J, Li X, Wang K, et al. Atmospheric PM2.5 Concentration Prediction Based on Time Series and Interactive Multiple Model Approach[J]. Adv Meteorol. 2019: 1–11. https://doi.org/10.1155/2019/1279565.
24. Ali Shah SA, Aziz W, Ahmed Nadeem MS, Almaraashi M, Shim SO, Habeebullah TM. A novel phase space reconstruction-(PSR-) based predictive algorithm to forecast atmospheric particulate matter concentration[J]. Sci Program. 2019;2019:1–12.
25. Ahn J, Shin D, Kim K, et al. Indoor air quality analysis using deep learning with sensor data[J]. Sensors. 2017;17(11):2476. https://doi.org/10.3390/s1711247626.
26. Yu Y, Hu C, Si X, Zheng J, Zhang J. Averaged bi-LSTM networks for RUL prognostics with non-life-cycle labeled dataset[J]. Neurocomputing. 2020;402:134–47. https://doi.org/10.1016/j.neucom.2020.03.041.
27. Li M, Lu F, Zhang H, Chen J. Predicting future locations of moving objects with deep fuzzy-LSTM networks[J]. Transportmetrica. 2018;16(1):119–36. https://doi.org/10.1080/23249935.2018.1552334.
28. Wu D, Jiang Z, Xie X, Wei X, Yu W, Li R. LSTM learning with Bayesian and Gaussian processing for anomaly detection in industrial IoT[J]. IEEE Trans Industr Inform. 2019;16(8):5244–53. https://doi.org/10.1109/TII.2019.2952917.
29. Pak U, Ma J, Ryu U, Ryom K, Juhyok U, Pak K, et al. Deep learning-based PM2. 5 prediction considering the spatiotemporal correlations: a case study of Beijing, China[J]. Sci Total Environ. 2020;699:133561. https://doi.org/10.1016/j.scitotenv.2019.07.367.
30. Lin W, Lo S, Young H, et al. Evaluation of deep learning neural networks for surface roughness prediction using vibration signal analysis[J]. Appl Sci. 2019;9(7). https://doi.org/10.3390/app9071462.
31. Kim HS, Park I, Song CH, Lee K, Yun JW, Kim HK, et al. Development of a daily PM10 and PM2. 5 prediction system using a deep long short-term memory neural network model[J]. Atmos Chem Phys. 2019;19:12935–51. https://doi.org/10.5194/acp-19-12935-2019.
32. Zhang T, Song S, Li S, et al. Research on gas concentration prediction models based on LSTM multidimensional time series[J]. Energies. 2019;12(1):161. https://doi.org/10.3390/en12010161.
33. Kim S, Lee J M, Lee J, et al. Deep-dust: Predicting concentrations of fine dust in Seoul using LSTM[J]. arXiv preprint arXiv:1901.10106. 2019.
34. Ding Y, Zhu Y, Feng J, Zhang P, Cheng Z. Interpretable Spatio-temporal attention LSTM model for flood forecasting[J]. Neurocomputing. 2020;403:348–59. https://doi.org/10.1016/j.neucom.2020.04.110.
35. Wang Q, Hao Y. ALSTM: an attention-based long short-term memory framework for Knowledge Base reasoning[J]. Neurocomputing. 2020;399:342–51. https://doi.org/10.1016/j.neucom.2020.02.065.
36. Pang X, Zhou Y, Li P, Lin W, Wu W, Wang JZ. A novel syntax-aware automatic graphics code generation with attention-based deep neural network[J]. J Netw Comput Appl. 2020:102636. https://doi.org/10.1016/j.jnca.2020.102636.
37. Li W, Tao W, Qiu J, Liu X, Zhou X, Pan Z. Densely connected convolutional networks with attention LSTM for crowd flows prediction[J]. IEEE Access. 2019;7:140488–98. https://doi.org/10.1109/ACCESS.2019.2943890.
38. Wang Z, Zhang L, Ding Z. Hybrid time-aligned and context attention for time series prediction[J]. Knowl-Based Syst. 2020: 105937. https://doi.org/10.1016/j.knosys.2020.105937.
39. Li F, Gui Z, Zhang Z, Peng D, Tian S, Yuan K, et al. A hierarchical temporal attention-based LSTM encoder-decoder model for individual mobility prediction[J]. Neurocomputing. 2020;403:153–66. https://doi.org/10.1016/j.neucom.2020.03.080.
40. Zou Q, Xiong Q, Li Q, Yi H., Yu Y., Wu C. A water quality prediction method based on the multi-time scale bidirectional long short-term memory network[J]. Environ Sci Pollut Res. 2020:1–12. https://doi.org/10.1007/s11356-020-08087-7.
41. Maleki A, Nasseri S, Aminabad MS, Hadi M. Comparison of ARIMA and NNAR models for forecasting water treatment plant's influent characteristics[J]. KSCE J Civ Eng. 2018;22(9):3233–45. https://doi.org/10.1007/s12205-018-1195-z.
42. Dai S, Li L, Li Z. Modeling vehicle interactions via modified LSTM models for trajectory prediction[J]. IEEE Access. 2019;7:38287–96. https://doi.org/10.1109/ACCESS.2019.2907000.
43. Tukkaraja P, Keerthipati M, French A. Simulating temperature inversions in surface mines using computational fluid dynamics[C]. Proc South Dakota Acad Sci. 2016;95:119.