sustainability

MDPI

Article

Construction and Application of an Intelligent Prediction Model for the Coal Pillar Width of a Fully Mechanized Caving Face Based on the Fusion of Multiple Physical Parameters

Zhenguo Yan [1], Huachuan Wang [2,3,*], Huicong Xu [2,4,*], Jingdao Fan [1,4] and Weixi Ding [4]

[1] College of Safety Engineering, Xi'an University of Science and Technology, Xi'an 710054, China; yanzg@xust.edu.cn (Z.Y.); fanjd@126.com (J.F.)
[2] State Key Laboratory of Green Low Carbon Development of Oil-Rich Coal in Western China, Xi'an 710054, China
[3] Department of Civil and Environmental Engineering, University of Strathclyde, Glasgow G1 1XJ, UK
[4] College of Energy Resources, Xi'an University of Science and Technology, Xi'an 710054, China; dwx2755050680@163.com
* Correspondence: huachuan.wang@strath.ac.uk (H.W.); xhcxust@163.com (H.X.)

Abstract: The scientific and reasonable width of coal pillars is of great significance to ensure safe and sustainable mining in the western mining area of China. To achieve a precise analysis of the reasonable width of coal pillars in fully mechanized caving face sections of gently inclined coal seams in western China, this paper analyzes and studies various factors that affect the retention of coal pillars in the section, and calculates the correlation coefficients between these influencing factors. We selected parameters with good universality and established a data set of gently inclined coal seams based on 106 collected engineering cases. We used the LSTM algorithm loaded with a simulated annealing algorithm for training, and constructed a coal pillar width prediction model. Compared with other prediction algorithms such as the original LSTM algorithm, the residual sum of squares and root mean square error were reduced by 27.2% and 24.2%, respectively, and the correlation coefficient was increased by 12.6%. An engineering case analysis was conducted using the W1123 working face of the Kuangou Coal Mine. The engineering verification showed that the SA-CNN-LSTM coal pillar width prediction model established in this paper has good stability and accuracy for multi-parameter nonlinear coupling prediction results. We have established an effective solution for achieving the accurate reservation of coal pillar widths in the fully mechanized caving faces of gently inclined coal seams.

Keywords: intelligent coal mining; gently inclined; coal pillar; physical indicators; prediction methodology; LSTM

Citation: Yan, Z.; Wang, H.; Xu, H.; Fan, J.; Ding, W. Construction and Application of an Intelligent Prediction Model for the Coal Pillar Width of a Fully Mechanized Caving Face Based on the Fusion of Multiple Physical Parameters. *Sustainability* **2024**, *16*, 986. https://doi.org/10.3390/su16030986

Academic Editor: Cun Zhang

Received: 29 December 2023
Revised: 15 January 2024
Accepted: 22 January 2024
Published: 23 January 2024

1. Introduction

The width of a coal pillar is a vital factor affecting the stationarity of the coal pillar and the difficulty of stope maintenance. Coal mines in western China are under intense exploitation. The scientific and rational width of coal pillars is of great significance to ensure safety and sustainable mining in the western mining area. Therefore, a machine learning method to determine the width of coal pillars is our current research focus [1]. The gently inclined coal seam reserves in western China are huge, and are an important factor in the current national energy supply strategy. However, the geomechanical environment in the western region is complex. Affected by indicators such as its time–space relationship, overlying strata, and geo-framework, the superposition of multi-field stress poses problems for on-site safety, prevention, and control, which seriously restricts safe and efficient operation on-site [2–5]. If a coal pillar is overly wide, it will lead to a huge

waste of resources, which seriously violates the "Scientific-Green-Precise" mining concept advocated by China. It is of vital scientific importance to quantitatively determine a suitable coal pillar width to realize precise and safe mining and environmental protection in western China.

At present, research on the appropriate width of coal pillars mostly adopts empirical analysis, similarity simulation experiments, numerical simulation, and real-time monitoring methods [6–8]. Suo [9] used UDEC4.0 numerical simulation software to study the width of upper and lower coal pillars during extremely close-range coal seam mining. They judged the appropriate width of the middle coal pillar in an adjacent horizontal subsection of a coal seam. Taking the mining situation of the 15,208 working faces of Xinjing Coal Mine as their engineering background, Wang [10] used theoretical methods such as limit equilibrium to determine the effect of mine pressure on the width of the plastic zone. The width of the effective waterproof elastic core was determined by the semi-inverse solution method, and the width of the water pressure failure zone was determined by taking full account of the softening due to water seepage. The theoretical width of a waterproof coal pillar was obtained. Zhang [11] aimed at combating the wasted resources caused by excessively wide protective coal pillars in the working face 30103 of Yayaomao Coal Mine. Through the research methods of field exploration, mechanical analysis, numerical simulation, and real-time monitoring, a coal pillar width calculation model was established. The surrounding stress, displacement distribution, and plastic zone characteristics under different coal pillar width conditions were revealed, and the suitable width of the coal pillar was determined. In recent years, experts, scholars, and research teams have introduced methods such as machine learning and deep learning into mining engineering research, using artificial intelligence to solve traditional mining problems, and trying to solve the problems of the three-zone distribution of mines and rockburst prediction by establishing machine learning or deep learning models. Some research results have been obtained [12–15]; some research teams have also introduced deep learning methods into their research on the suitable width of coal pillars [16].

The combination of deep learning solutions and traditional mining engineering problems provides a scientifically effective method and means for the design of coal pillar widths [17,18]. The LSTM network model has a good predictive performance on non-time series data, with special gating units and decent non-linear mapping capabilities [19–21]. It could overcome the long-term dependence on data characteristics in forecasting research. The parameters required for model establishment, that is, the physical indicators, are easy to measure and quantitative descriptions of qualitative characteristics are easy to determine. This provides a credible new method for the accurate calculation and quantitative design of the safe coal pillar width. At present, most studies on the suitable width of coal pillars are based on engineering experience, and numerical simulation methods combined with field measurements are used to analyze and obtain reasonable values. Existing studies using deep learning methods to predict the width of remaining coal pillars lack the correlation verification of their parameter selection, and the prediction effect of their models still needs to be improved.

2. Physical Indexes Affecting Coal Pillar Width

2.1. Selection of Physical Indicators

Usually, the mechanical factors of coal rock mass and the mining method used in the coal seam chiefly affect coal pillar width. For the sake of studying the suitable width of the reserved coal pillar, the selected physical indicators should have good measurability, facilitate data acquisition, and comprehensively consider the correlation of the influence of various factors on the width. This study selected eight physical indicators, including coal seam burial depth, coal seam dip angle, coal seam thickness, tensile strength, elastic modulus, Poisson's ratio, cohesion, and internal friction angle, as parameters for subsequent model training samples. In this paper, a total of 106 pieces of gently inclined coal

pillar width multi-parameter data were selected for training [22–24]. Due to space limitations, 20 sets of sample examples were selected, as shown in Table 1.

Table 1. Raw data for coal pillar width prediction.

No.	Buried Depth/m	Coal Seam Dip Angle/(°)	Coal Seam Thickness/m	Tensile Strength/MPa	Elastic Modulus/GPa	Poisson's Ratio	Cohesion/MPa	Angle of Internal Friction/(°)	Coal Pillar Width/m
1	600	8	3.20	0.50	2.00	0.34	2.50	32.00	20
2	400	20	1.00	5.83	4.91	0.30	0.08	40.07	10
3	350	10	2.40	0.12	0.80	0.40	0.10	25.00	17
4	700	25	4.40	0.74	0.20	0.27	1.83	36.86	30
5	700	22	4.20	0.66	0.17	0.36	2.87	20.69	28
6	400	8	5.87	0.28	2.36	0.24	0.28	59.17	30
7	220	23	5.25	0.12	0.64	0.28	1.15	24.00	15
8	300	9	4.67	2.10	4.00	0.44	1.00	30.00	20
9	300	10	4.87	2.10	4.00	0.44	1.60	19.00	15
10	600	9	4.75	0.53	2.95	0.22	5.60	22.00	25
11	200	8	6.50	0.49	2.00	0.33	2.70	28.10	6
12	450	25	7.88	1.08	0.60	0.26	1.13	26.00	12
13	300	19	2.20	0.57	4.20	0.12	1.50	42.00	12
14	500	8	5.50	0.55	2.97	0.32	2.05	35.00	40
15	400	9	5.25	0.35	2.43	0.31	0.50	24.00	7
16	266	22	3.50	0.20	0.65	0.28	0.80	17.50	32
17	300	9	4.00	1.45	5.96	0.34	1.70	36.70	6
18	440	10	5.86	0.32	2.36	0.26	0.28	30.00	30
19	400	17	2.47	0.51	1.56	0.12	1.80	38.00	22
20	200	8	5.00	1.83	2.50	0.28	1.46	28.00	20

2.2. Establishment of a Physical Index Fusion Data Set

Multi-parameter deep learning prediction usually involves many types of parameters with different dimensions, or that are dimensionless, and the parameters of each dimension having a wide range of distribution. In the process of learning such samples, the model is easily overwhelmed by data with a large or small parameter distribution range. This influence, resulting in a reduced model training effect, ultimately leads to a poor prediction effect. After some deep learning models are stretched unevenly in each dimension their optimal solution will not change, but, usually, deep learning models will contain iterative algorithms, and if the distribution range of parameters in each dimension is too different, this may lead to iterative divergence. The convergence speed of the algorithm slows down or directly causes the iterative algorithm not to converge. Normalizing the source data can ensure that parameters with different dimensions can eliminate the influence of their dimension, and, at the same time, this can speed up the data processing speed of the model.

The LSTM model selected in this paper is less affected by multi-parameter dimension differences, but an iterative algorithm is still used. Therefore, when selecting physical indicators, the original data are normalized, and the Min–Max normalization method is applied. The formula is:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Apply Formula (1) to normalize the 8 physical indicators selected to obtain the coal pillar width predictive learning and training sample data sets. The sample data sets after calculation and processing are shown in Table 2.

Table 2. Coal pillar width prediction data set.

No.	Buried Depth/m	Coal Seam Dip Angle/(°)	Coal Seam Thickness/m	Tensile Strength/MPa	Elastic Modulus/GPa	Poisson's Ratio	Cohesion/MPa	Angle of Internal Friction/(°)	Coal Pillar Width/m
1	0.80000	0.00000	0.31977	0.06655	0.31606	0.68750	0.43841	0.34797	0.41176
2	0.40000	0.70588	0.00000	1.00000	0.81865	0.56250	0.00000	0.54164	0.11765
3	0.30000	0.11765	0.20349	0.00000	0.10881	0.87500	0.00362	0.17999	0.32353
4	1.00000	1.00000	0.49419	0.10858	0.00518	0.46875	0.31703	0.46460	0.70588
5	1.00000	0.82353	0.46512	0.09457	0.00000	0.75000	0.50543	0.07655	0.64706
6	0.40000	0.00000	0.70785	0.02802	0.37824	0.37500	0.03623	1.00000	0.70588
7	0.04000	0.88235	0.61773	0.00000	0.08117	0.50000	0.19384	0.15599	0.26471
8	0.20000	0.05882	0.53343	0.34676	0.66149	1.00000	0.16667	0.29998	0.41176
9	0.20000	0.11765	0.56250	0.34676	0.66149	1.00000	0.27536	0.03600	0.26471
10	0.80000	0.05882	0.54506	0.07180	0.48014	0.31250	1.00000	0.10799	0.55882
11	0.00000	0.00000	0.79942	0.06480	0.31606	0.65625	0.47464	0.25438	0.00000
12	0.50000	1.00000	1.00000	0.16813	0.07427	0.43750	0.19022	0.20398	0.17647
13	0.20000	0.64706	0.17442	0.07881	0.69603	0.00000	0.25725	0.58795	0.17647
14	1.40000	0.94118	0.94477	0.11734	0.54231	1.37500	0.38587	1.25990	1.35294
15	1.20000	1.00000	0.90843	0.08231	0.44905	1.34375	0.10507	0.99592	0.38235
16	0.93200	1.76471	0.65407	0.05604	0.14162	1.25000	0.15942	0.83993	1.11765
17	1.00000	1.00000	0.72674	0.27496	1.05872	1.43750	0.32246	1.30070	0.35294
18	1.28000	1.05882	0.99709	0.07706	0.43696	1.18750	0.06522	1.13991	1.05882
19	1.20000	1.47059	0.50436	0.11033	0.29879	0.75000	0.34058	1.33189	0.82353
20	0.80000	0.94118	0.87209	0.34151	0.46114	1.25000	0.27899	1.09191	0.76471

2.3. Correlation Calculation of the Physical Indicators

In the prediction of multi-parameter deep learning models without time series, the correlation between various parameters will affect the model's recognition of its features and the actual prediction performance of the final model. In this paper, a total of eight physical indicators are selected as the parameters of the data set. To ensure the effectiveness of the model training, it is necessary that a correlation calculation is carried out for the selected physical indicators, and that the correlation between the physical indicators and the target predicted value is tested.

Thirty samples from the data set were used to carry out the correlation calculation in a fitting software to verify the accuracy of the multivariable prediction based on eight physical indexes. The accuracy of the multivariate prediction of the eight physical indexes, from burial depth to internal friction angle, was assessed using Spearman's rank correlation. Its value was irrelated to the specific value of two correlation variables, and only related to the relationship between their numerical range. The Spearman rank correlation is calculated based on the variance between the ranks of each pair of ranks in two columns, so it is also called the rank variance method. Usually represented by the Greek letter ρ, its calculation formula is

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \quad (2)$$

In this formula, x and y represent two variables, but in practical applications, Formula (2) is usually simplified as

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (3)$$

In this formula, d represents the rank difference of the corresponding variable, that is, the position of the paired variable after the two variables are sorted respective to each other, and n is the quantity of variables observed. The correlation index was calculated using Formula (3). The calculation results are shown in Figure 1.

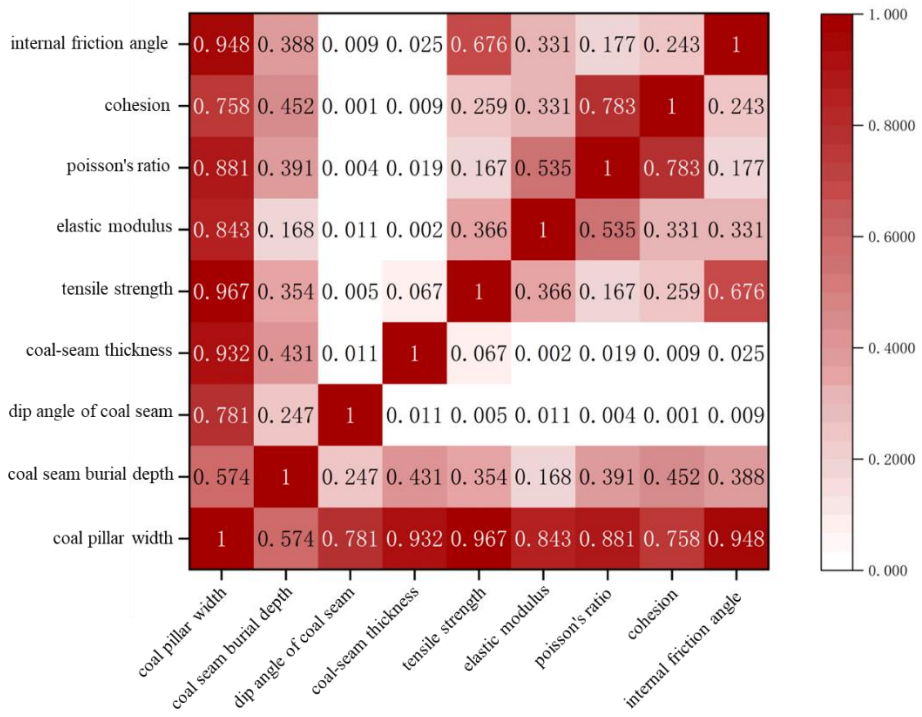

Figure 1. Correlation matrix.

The correlation coefficient in the correlation matrix is expressed as an absolute value, and the color order is related to the strength of correlation: the closer the value is to 1, the deeper the color of the block is. The closer the value is to 0, the closer the color of the block is to white, while the correlation between a variable and itself is 1. The closer the correlation coefficient is to 1, the higher the correlation strength is. Under normal circumstances, a correlation coefficient of about 0.7 or above can be considered a strong correlation. When the correlation coefficient tends to zero, the color block tends to white, and it can be considered that the correlation between the two physical indicators is weak. The correlation coefficients of the eight physical indicators are 0.737, 0.683, 0.915, 0.967, 0.718, 0.842, 0.726, and 0.931. It is calculated that there is a strong correlation between the eight physical indicators and coal pillar width [25]. Among them, the correlation coefficient between the coal seam dip angle and coal pillar width is the lowest, and their correlation coefficient is 0.683. It can be seen intuitively from Figure 1 that the eight physical indicators all have a strong correlation with the width and can be used as parameters for the multi-parameter fusion prediction of the width of reserved coal pillars.

3. Establishment of the LSTM Prediction Model

3.1. Long Short-Term Memory Network

Recurrent neural networks can describe the relationship between the output information and the state before the output information. This network has very time-sensitive characteristics when processing sequence–structure information and can apply the information from the previous moment to the output of subsequent nodes. In the time dimension, it has the characteristics of accumulation from front to back. However, in an RNN there exists the defect of gradient explosion and dispersion. To solve the shortcomings of the RNN structure, the LSTM network adds 3 gate units based on RNN, including the forget gate, input gate, and output gate, thus controlling the state iteration of the LSTM unit, its memory, input acquisition, and output transmission, thus improving the efficiency and stability of the network [26]. Multiple "GATE" structure neurons positioned in each layer can memorize any time state of the hidden layer. No matter how long the gradient propagation path is, it will never disappear or drop to zero. The long short-term

memory neural network (LSTM) is a time-recursive neural network, which is improved from the RNN (Recurrent Neural Network), solving the problems of gradient dispersion and gradient explosion in the RNN model, and is suitable for processing and predicting data sets with multi-parameter features.

LSTM consists of seven components including input X_t, unit state C_t, temporary unit state $\tilde{C}_t$, hidden state h_t, forget gate f_t, memory gate i_t, and output gate o_t, and the subscript t represents a certain moment. The iterative process of LSTM can be simply described as follows: Firstly, information is forgotten, and new information is remembered; only effective data for subsequent cell data processing are transmitted, and the hidden layer state h_t is output at each time step. Among these data, the forgetting parameter, memory parameter, and output parameter are commanded by the corresponding module, calculated by the hidden layer state h_{t-1} of the previous moment and the present input. Its calculation formula is as follows:

In the forgetting gate,

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}, X_t\right] + b_f\right) \tag{4}$$

where f_x is the forget gate at time t; σ(x) is the sigmoid activation function; W_f is the weight vector of the forget gate; h_{t-1} is the output at time t−1; X_i is the input at this time; and b_f is the offset of the forget gate. After the function output, the value of f_t is (0, 1), and f_t will be multiplied by C_{t-1} bit by bit. When the value of a certain bit of f_t is 0, then so is the corresponding bit of C_{t-1}: the value will be forgotten. When a certain bit of f_t is 1, the corresponding value of C_{t-1} will be retained.

There are two parts in the updated gate layer, the formulas of the first part are as follows:

$$i_f = \sigma\left(W_i \cdot \left[h_{t-1}, X_t\right] + b_i\right) \tag{5}$$

$$\tilde{C}_t = tanh\left(W_h \cdot \left[h_{t-1}, X_t\right] + b_C\right) \tag{6}$$

The second part of the formula is

$$C_t = f \times C_{t-1} + i \times \tilde{C}_t \tag{7}$$

In this formula, i_t is the input gate at time t; b_i and b_c are the offsets of the input gate and unit information update, respectively; $\tilde{C}_t$ is the new information at time t; C_t and C_{t-1} are the units at the corresponding time information states; $\tan h$ is the activation function; and W_i and W_c are the weight vectors for the input gate and unit information update, respectively. The i_t in the first part is like the forget gate, it is the part of the information that is retained; $\tilde{C}_t$ is the information introduced by the new input, and its value is normalized to −1 or to 1 after $\tan h$. C_t, in the second part of the formula, is the cell information state after adding the new cell information and the updated cell information at this moment.

In the output gate layer,

$$o_t = \sigma\left(W_o \cdot \left[h_{t-1}, X_t\right] + b_o\right) \tag{8}$$

$$h_t = o_t \times tanh(C_t) \tag{9}$$

In this formula, o_t is the output gate; W_t is the weight vector of the output gate; b_o is the offset of the output gate; and h_t is the final output result.

The LSTM network's internal unit structure is shown in Figure 2.

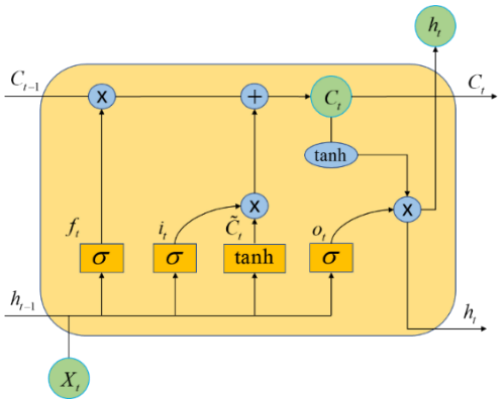

Figure 2. LSTM internal unit structure diagram.

The input layer of the LSTM algorithm mainly consists of two parts, namely, the initial hidden state and the initial unit state, represented by parameters C_{t-1} and h_{t-1}, respectively, as shown in Figure 2. The output layer format is consistent with the input layer, and it is also a hidden state and a unit state. But the labeling of the two parameters is different from the input layer.

In practical applications, the LSTM model does not exist in the form of one unit, but is connected to multiple units; each unit is encapsulated, and the processed data are passed to the next unit, as shown in Figure 3.

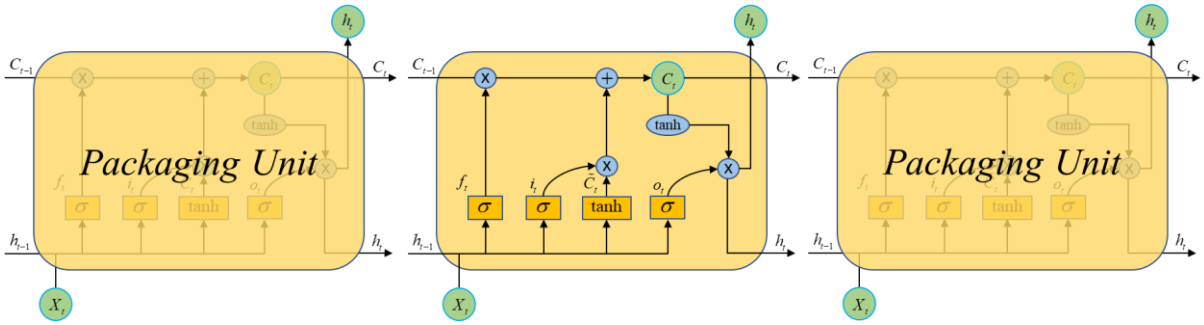

Figure 3. Basic structure of LSTM model.

3.2. The Coal Pillar Width Physical Index Fusion Prediction Model Based on Simulated Annealing Algorithm Optimization

The simulated annealing algorithm (SA) was first introduced by N. Metropolis et al. in 1953 [27], and it is a method for approximately solving the optimization problem, designed using Monte Carlo thinking. The search for the optimal solution can be regarded as the search for the minimum energy value of a complex system. Therefore, when the system temperature drops, the energy will decline step by step, and, correspondingly, the solution to the problem also "falls" to the minimum, which shows the optimal solution. The Metropolis discriminant criterion is as follows:

$$p = \begin{cases} 1, E(x_{new}) < E(x_{old}) \\ exp\left(\dfrac{E(x_{new}) - E(x_{old})}{T}\right), E(x_{new}) < E(x_{old}) \end{cases} \quad (10)$$

In this formula, p is the poor solution acceptance probability, also known as the state acceptance function; E represents the energy state; x_{new} represents the new state; and x_{old} represents the current state. The algorithm steps are as follows (Figure 4):

Step 1. Define the incipient temperature T_0, the incipient solution state X_1, and the rounds of iterations L of each T value (this is the Metropolis sampling stability criterion).

Step 2. Use the state transition function to convert a new solution X_2.

Step 3. Calculate the increment $\Delta E = E(X_2) - E(X_1)$. $E(X)$ is the evaluation function.

Step 4. If $E(X)$ meets the requirements, then accept X_2 as the new current solution, otherwise accept X_2 with a probability (this is the Metropolis sampling stability criterion).

Step 5. When the termination condition is not met, gradually decrease T and jump to Step 2 to start iteration.

Step 6. If the control parameter T satisfies the termination condition, the algorithm will end the recursion and choose the last solution as the optimal solution.

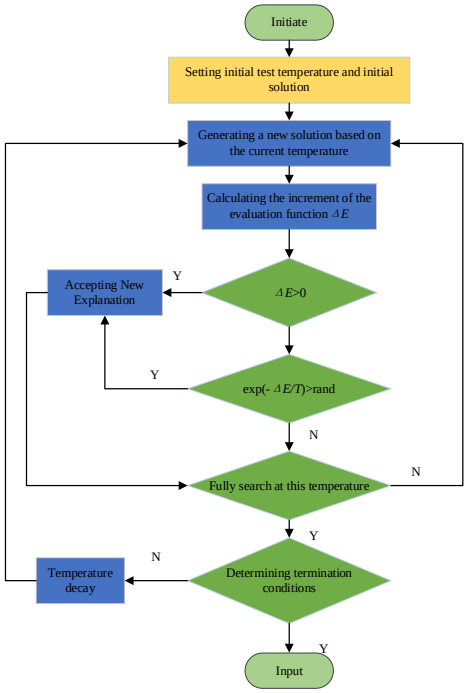

Figure 4. Flow chart of simulated annealing algorithm.

The normalized data are imported into the model. This simulated annealing algorithm was applied to optimize the number of layers(lstm_layer, ly), the number of neurons in the hidden layer (lstm_nets, ln), the number of layers in the connection layer (dense_layer, dy), and the number of neurons in the connection layer (dense_nets, dn) to fit the optimal parameter combination, and then when combined with the CNN-LSTM model a coal pillar width prediction model is formed. The model architecture is shown in Figure 5.

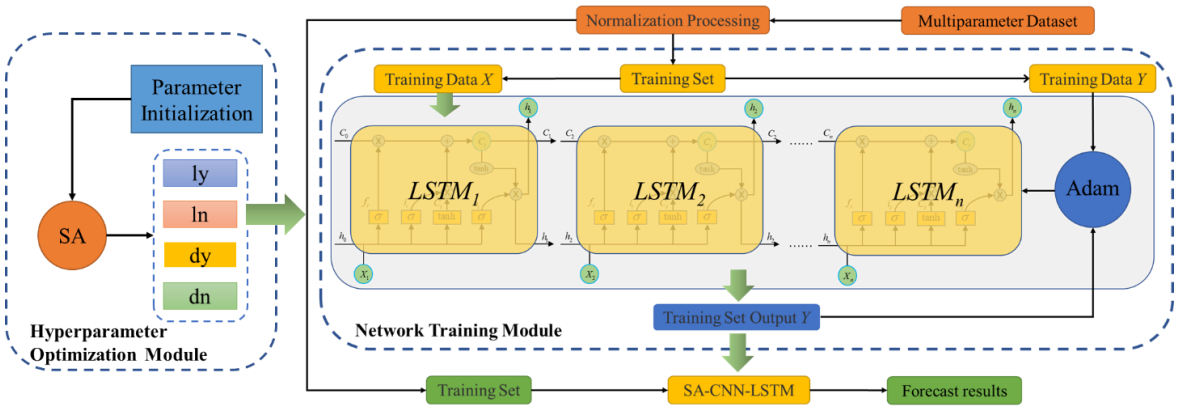

Figure 5. SA-CNN-LSTM architecture diagram of coal pillar width prediction model.

4. Analysis and Verification of the Sample Prediction Effect

4.1. Model Training

The normalized coal pillar width prediction data set was divided, and 89 pieces of information were selected as the model training set to be imported into the SA-CNN-LSTM algorithm for training, and the prediction model, optimized using the simulated annealing algorithm, was generated. In the actual training, the parameter T of the simulated annealing algorithm is initialized. If the value of T is too large, the algorithm may stop iterating when it reaches the local optimal value; if the value of T is too small, the optimization speed will be slowed down, and the overall performance of the algorithm will be reduced. Therefore, this paper adopts an exponential descending formula to control the parameter T, and adopts a larger T value at the beginning of the iteration, which gradually decreases with the algorithm's iteration. The parameter T control formula is

$$T(n) = q^{T(n)}, (n = 1,2,3\ldots) \quad (11)$$

In this formula, q is the initial value. After analysis, this was set to q = 0.91 to trigger the optimum computational power of the algorithm. During the iterative process of model training, the simulated annealing algorithm continuously optimizes and adjusts the network parameters of the model. The number of hidden layers and the number of neurons, the number of connection layers and the quantity of neurons in the LSTM are optimized by a genetic algorithm to determine the best parameter combination. The simulated annealing algorithm optimizes the LSTM parameter range, as shown in Table 3.

Table 3. Neural network structure's parameter settings.

Parameter Name	Parameter Value
Hidden Layers	[1, 5]
Hidden Layer Neurons	[1, 200]
Connection Layers	[1, 5]
Connected Layer Neurons	[1, 100]

For the sake of improving the convergence effect of the model, we choose Adam as the internal optimizer of LSTM. MSE was selected as the model loss function to calculate the loss value. To compare the optimization effects of different schemes on the model architecture, we input the optimal hyperparameters found by each optimization algorithm into the LSTM architecture. The quantity of iterations was set to 100. The loss function of each prediction model is shown in Figure 6.

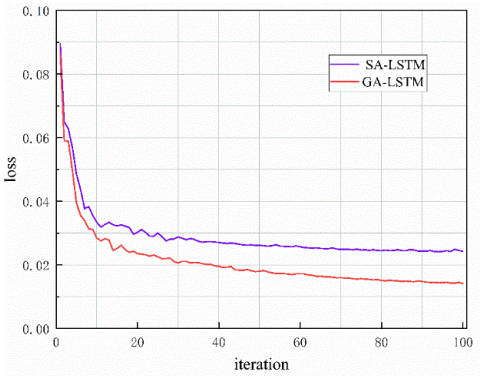

Figure 6. Prediction model's loss function curve.

To compare the optimization performance of the SA algorithm, two models were selected for the experiment. It can be understood from Figure 6 that both prediction models

converge within the set number of iterations, and the loss function eventually tends toward a stable value. Among them, the SA-LSTM model established in this paper has a minimum loss value of 0.0141 after stabilization, and its convergence speed is faster than the GA-LSTM prediction model's. It is seen intuitively from Figure 6 that the SA-LSTM training model established in this paper has the best performance, which verifies the effect of the simulated annealing algorithm on the hyperparameter optimization of the LSTM model, and finally determines to use the simulated annealing algorithm to optimize the coal pillar width prediction model.

4.2. Prediction Result Evaluation Index

When evaluating the prediction performance, this article considers three commonly used performance indicators: the residual sum of squares (*RSS*), root mean square error (*RMSE*), and correlation coefficient (R^2).

The residual sum of squares' formula is

$$RSS = \sum_{i=1}^{N} (\hat{I}_i - I_i)^2 \tag{12}$$

The root mean square error's formula is

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (\hat{I}_i - I_i)^2 \tag{13}$$

The correlation coefficient's formula is

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (\hat{I}_i - I)^2}{\sum_{i=1}^{N} (I_i - \bar{I})^2} \tag{14}$$

To verify the training effect of the selected verification set, the model evaluation index is calculated and compared with other models to verify the prediction effect of the SA-CNN-LSTM model.

4.3. Analysis of Forecast Results

To verify the training effect of the model, 15 values were selected as verification sets. The data in the verification set have the characteristics of a typical gently inclined coal seam, which can be used to verify the training effect. The trained model was used to predict the width of coal pillars, and the predicted results were reverse normalized to facilitate intuitive analysis, and compared with the actual values, as shown in Figure 7.

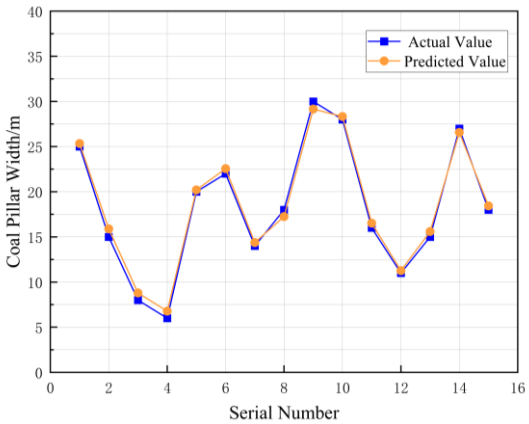

Figure 7. Comparison chart of predicted values and real values.

By calculating the error between the obtained estimated value and the real value, the parameter values of the performance evaluation index are calculated, as shown in Table 4.

Table 4. Parameter values of the model performance evaluation index.

Parameter	Value
RSS	0.00073
RMSE	0.00837
R^2	0.96733

When the same data are selected and the same calculation method is used for the parameter estimation and predictive analysis of the classic LSTM algorithm, R^2 = 0.85924, which is 12.6% higher than the model used in this paper, *RSS* = 0.00574, and *RMSE* = 0.01095; compared with the model used in this paper, they are reduced by 27.2%, 41.7%, and 24.2%, respectively.

5. Practical Application Based on the Engineering Background

5.1. Coal Pillar Setting Width in Gently Inclined Mine Sections

Based on the comprehensive prediction comparison chart and model prediction effect evaluation index, it can be concluded that the SA-CNN-LSTM model trained using the coal pillar width data set in gently inclined mine sections is useful for prediction. Its prediction results have small dispersion and high prediction accuracy. It can be used to guide engineering practice, provide more accurate reference data for coal pillar reservation in gently inclined mine sections, and provide safe and reliable data support for mining in gently inclined mines.

5.2. Engineering Example Verification

The engineering verification section used real data from the W1123 working face of Kuangou Coal Mine. The inclined width of the W1123 working face is about 192 m and the strike length is about 1468 m. The working face features fully mechanized caving mining, with a mining height of 3.2 m, a coal caving height of 6.3 m, and a mining-to-caving ratio of 1:1.97. The width of the coal pillar between the W1121 and W1123 working faces is 15 m, with an upper groove of 4.2 m and a lower groove of 4.5 m on the W1123 working face. The trained prediction model was applied to the design of the coal pillar width of a certain coal seam in a coal mine and eight physical index parameters were obtained and input into the prediction model, and the predicted value of the coal pillar width was obtained. The absolute error and relative error value calculated for the single predicted value are shown in Table 5.

Table 5. Prediction of coal pillar width in the W1123 coal seam in Kuangou Coal Mine.

Name	Predictive Effect		Difference	
	Actual Value/m	Predictive Value/m	Absolute Error/m	Relative Error/%
SA-CNN-LSTM	15	15.19	0.19	1.25%
GA-LSTM	15	15.76	0.76	4.82%
PSO-LSTM	15	16.44	1.44	8.76%
LSTM	15	17.81	2.81	15.78%

It can be seen from Table 5 that the absolute error of SA-CNN-LSTM in predicting the coal pillar width in this Kuangou Coal Mine section is 0.19, and the relative error is 1.25%, which can satisfy the requirements for the engineering example's application. The verification of engineering examples shows that the SA-CNN-LSTM model based on small

samples has incomparable superiority to other methods (such as GA-LSTM, PSO-LSTM, LSTM, etc.), and can save a lot of manpower in practical engineering applications.

6. Conclusions

(1) A data set of physical indicators related to coal pillar width, based on the measured data, was established, and the data set of coal pillar widths suitable for the final model training was obtained by normalizing the obtained data set. The correlation coefficient between the physical indicators and the width of the reserved coal pillar was obtained by applying correlation calculations. The calculation results have verified that there is a strong correlation between the eight physical indicators and the width of the reserved coal pillar. It has been verified that they can be used as basic data for predicting the width of reserved coal pillars in gently inclined fully mechanized mining faces with on multi-parameter fusion.

(2) A prediction model of coal pillar width based on a simulated annealing algorithm was established, and the simulated annealing algorithm was introduced to optimize the hyperparameters of the model. The effect of the optimization algorithm was compared and analyzed, and it was confirmed that the simulated annealing algorithm has the best optimization effect for the prediction effect of the whole algorithm. The prediction effect of the model was verified via four performance evaluation indexes, including the residual sum of squares (*RSS*), mean square error (*MSE*), root mean square error (*RMSE*), and correlation coefficient (also known as the certainty coefficient or determinability coefficient, R^2). Compared to the basic LSTM algorithm, the correlation coefficient obtained by our model was increased by 12.6%, and the residual sum of squares, mean square error, and root mean square error were reduced by 27.2%, 41.7%, and 24.2%, respectively.

(3) The prediction model was applied to predict and analyze the width of the coal pillar in the W1123 working face of Kuangou Coal Mine, verifying the feasibility and reliability of the model for predicting actual coal pillar widths. Due to the differences in geological conditions and mining processes across different mines, we will expand the sources of our data samples in the future and attempt to obtain data sets based on a wider range of physical indicators in mines. This study verifies the universal influence of the characteristics of eight physical indicators, including coal pillar width, Poisson's ratio, cohesion, and the internal friction angle, on the reserved coal pillar width, through deep learning.

Author Contributions: Conceptualization, Z.Y.; methodology, H.W.; preparation of deep learning models, H.X.; validation, W.D.; formal analysis, Z.Y.; data preparation and process, H.W.; writing—original draft preparation, H.W.; writing—review and editing, H.X.; visualization, W.D.; supervision, J.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported by the National Natural Science Foundation of China (No. 51904227), the basic research program of the Natural Science in Shaanxi Province (No. S2019-JC-LH-QY-SM-0102), the Key R&D Project of Yulin National High-tech Industrial Development Zone (No. 2D-2021-01), the Outstanding Ph.D Dissertation Cultivating Program of XUST (No. PY22001), and the National Foundation for studying abroad (No. [2022]87). Support from these agencies is gratefully acknowledged.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Wang, R.; Bai, J.B.; Yan, S.; Pan, G.Q.; Zhang, D.; Zhu, Q.C.; Dou, Z. Structure partition and reasonable width determination of waterproof coal pillar in strip mining. *Lithosphere* **2021**, *2021*, 3339797. http://doi.org/10.2113/2021/3339797.
2. Xu, H.; Lai, X.; Shan, P.; Yang, Y.; Zhang, S.; Yan, B.; Zhang, Y.; Zhang, N. Energy dissimilation characteristics and shock mechanism of coal-rock mass induced in steeply-inclined mining: Comparison based on physical simulation and numerical calculation. *Acta Geotech.* **2023**, *18*, 843–864. https://doi.org/10.1007/s11440-022-01617-2.
3. Xin, J.; Jiang, Q.; Liu, Q.; Zheng, H.; Li, S. A shear constitutive model and experimental demonstration considering dual void portion and solid skeleton portion of rock. *Eng. Fract. Mech.* **2023**, *281*, 109066. https://doi.org/10.1016/j.engfracmech.2023.109066.
4. Xin, J.; Jiang, Q.; Li, S.; Chen, P.; Zhao, H. Fracturing and Energy Evolution of Rock Around Prefabricated Rectangular and Circular Tunnels Under Shearing Load: A Comparative Analysis. *Rock Mech. Rock Eng.* **2022**, *56*, 9057–9084. https://doi.org/10.1007/s00603-023-03532-8.
5. Zhou, L.; Zhu, Z.; Xie, X.; Hu, Y. Coupled thermal–hydraulic–mechanical model for an enhanced geothermal system and numerical analysis of its heat mining performance. *Renew. Energy* **2022**, *181*, 1440–1458. https://doi.org/10.1016/j.renene.2021.10.014.
6. Wang, P.; Qi, Z.; Ma, C.; Cai, M. Anisotropic behavior of the seepage-stress coupling mechanical model of coal pillars of underground reservoirs. *Géoméch. Geophys. Geo-Energy Geo-Resour.* **2023**, *9*, 3. https://doi.org/10.1007/s40948-023-00549-9.
7. Zhang, C.; Zhao, Y.X.; Huang, H.F.; Bai, Q.S. Coal pillar failure analysis and instability evaluation methods: A short review and prospect. *Eng. Fail. Anal.* **2022**, *138*, 106344. https://doi.org/10.1016/j. engfailanal.2022.106344.
8. Bai, J.B.; Hou, C.J.; Huang, H.F. Numerical simulation study of narrow coal pillars in roadway driving along goaf. *Chin. J. Rock Mech. Eng.* **2004**, *23*, 3475–3479.
9. Suo, Y.L.; Liu, J.D.; Zhou, L.S.; Mao, X.W.; Shang, T.L. Study on the Reasonable Width of Coal Pillars in the Mining Section of Very Close Range Coal Seams mining. *Coal Eng.* **2014**, *40*, 8–14.
10. Wang, Q.; Gao, H.; Jiang, B.; Li, S.; He, M.; Wang, D.; Lu, W.; Qin, Q.; Gao, S.; Yu, H. Research on reasonable coal pillar width of roadway driven along goaf in deep mine. *Arab. J. Geosci.* **2017**, *10*, 466. https://doi.org/10.1007/ s12517-017-3252-1.
11. Zhang, J.G.; Cheng, Z.H.; Cheng, H.Y.; Yin, S.F.; Chen, L.; Xue, A. Analysis and optimization of remaining width of coal pillars in the section of yayaomao coal mine. *Coal Sci. Technol.* **2022**, *50*, 60–67.
12. Dai, L.; Pan, Y.; Li, Z.; Wang, A.; Xiao, Y.; Liu, F.; Shi, T.; Zheng, W. Quantitative mechanism of roadway rockbursts in deep extra-thick coal seams: Theory and case histories. *Tunn. Undergr. Space Technol.* **2021**, *111*, 103861. http://doi.org/10.1016/j.tust.2021.103861.
13. Dong, L.J.; Hu, Q.C.; Tong, X.J.; Liu, Y.F. Velocity-free MS/AE source location method for three-dimensional hole- con-taining structures. *Engineering* **2020**, *6*, 827–834.
14. Du, J.; Chen, J.; Pu, Y.; Jiang, D.; Chen, L.; Zhang, Y. Risk assessment of dynamic disasters in deep coal mines based on mul-ti-source, multi-parameter indexes, and engineering application. *Process Saf. Environ. Prot.* **2021**, *144*, 575–586.
15. Huang, L.Q.; Li, J.; Huang, H.; Li, X.B. Micro-seismic event detection and location in underground mines by using Con-volutional Neural Networks (CNN) and deep learning. *Tunn. Undergr. Space Technol.* **2018**, *81*, 265–276.
16. Wang, Z.Y.; Lai, X.P.; Liu, X.M.; Cui, F. Construction and application of the GRNN model of coal section pillar width prediction in fully mechanized face. *J. Xi'an Univ. Sci. Technol.* **2019**, *39*, 209–216.
17. Jia, Y.P.; Lv, Q.; Shang, Y.Q. Rockburst prediction using partice swarm optimization algorithm and grneral regression neural network. *Chin. J. Rock Mech. Eng.* **2013**, *32*, 343–348.
18. Zhou, L.M.; Zhu, S.; Zhu, Z.D. Cosserat ordinary state-based peridynamic model and numerical simulation of rock fracture. *Comput. Geotech.* **2023**, *155*, 105240. https://doi.org/10. 1016/j.compgeo.2022.105240.
19. Gers, F.A.; Schmidhuber, J.; Cummins, F. Learning to Forget: Continual Prediction with LSTM. *Neural Comput.* **2000**, *12*, 2451–2471. https://doi.org/10.1162/089976600300015015.
20. Wu, S.H.; Huang, Q.H.; Zhao, L. A deep learning-based network for the simulation of airborne electromagnetic re-sponses. *Geophys. J. Int.* **2023**, *233*, 253–263.
21. Azizi, N.; Yaghoubirad, M.; Farajollahi, M.; Ahmadi, A. Deep learning based long-term global solar irradiance and tem-perature forecasting using time series with multi-step multivariate output. *Renew. Energy* **2023**, *206*, 135–147.
22. Wu, X.; Lai, X.P.; Guo, J.B.; Cui, F.; Wang, Z.Y. PSO-SVM prediction model of coal pillar width in fully mechanized mining face. *J. Xi'an Univ. Sci. Technol.* **2020**, *40*, 64–70.
23. Li, X.; Zhang, X.; Shen, W.; Zeng, Q.; Chen, P.; Qin, Q.; Li, Z. Research on the mechanism and control technology of coal wall sloughing in the ultra-large mining height working face. *Int. J. Environ. Res. Public Health* **2023**, *20*, 868. https://doi.org/10.3390/ijerph20010227.
24. Zhang, S.; Cao, J.; Liu, Y.; Liu, J.; Hong, C.; Lei, B. Statistical characterization of damage of different surface p-wave velocity sets under dynamic load and study on overall radon detection consistency. *Lithosphere* **2021**, *2021*, 2015665. https://doi.org/10.2113/2021/2015665.
25. Chen, Y.; Bradley, B.A.; Baker, J.W. Nonstationary spatial correlation in New Zealand strong ground-motion data. *Earthq. Eng. Struct. Dyn.* **2021**, *50*, 3421–3440. https://doi.org/10.1002/eqe.3516.

26. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780.
27. Steinbrunn, M.; Moerkotte, G.; Kemper, A. Heuristic and randomized optimization for the join ordering problem. *VLDB J.* **1997**, *6*, 191–208. https://doi.org/10.1007/s007780050040.