Journal of Network Intelligence
Taiwan Ubiquitous Information

Volume **8**, Number **3**, August **2023**

Coal Miner Height Trend Prediction Based on Dropout_LSTM_LEC Deep Learning Model

Chang-Zhou Lv

School of Computer Science and Technology
Henan Polytechnic University
Jiaozuo 454000, China
212109010030@home.hpu.edu.cn

Wei-Peng An*

School of Software
Henan Polytechnic University
Jiaozuo 454000, China
2597511146@qq.com

Wei-Hao Qiao

School of Computer Science and Technology
Henan Polytechnic University
Jiaozuo 454000, China
376842083@qq.com

Xue-Han Zhao

School of Computer Science and Technology
Henan Polytechnic University
Jiaozuo 454000, China
228914636@qq.com

*Corresponding author: Wei-Peng An

Received November 26 2022, revised January 19 2023, accepted March 20 2023.

Abstract. *The problem of low accuracy of the current traditional coal mining machine memory cut-off technology is due to the influence of the non-linear and the local abrupt change of cut-off trajectory data. To solve this problem, this paper proposes a coal mining height pre-model (Dropout_LSTM_LEC) based on the Dropout optimization algorithm and deep Long Short-Term Memory (LSTM) neural network coupled with Local Error-Correct (LEC) to improve the prediction accuracy of coal mining height trend, which is extremely necessary to realize the automated production of coal mining machines. The proposed model is based on LSTM and builds a multi-level LSTM prediction model to improve the prediction accuracy; meanwhile, the Dropout optimization algorithm is used to complete the training of the model to alleviate the model overfitting problem; finally, the local mutation correction method is combined to correct the prediction results to reduce the prediction error caused by the local mutation of the truncated trajectory. Through practical validation, the proposed model shows better performance in terms of mean absolute error, mean absolute percentage error, and root mean square error compared with the gradient boosting regression tree and support vector regression algorithms.*
Keywords: Coal mining machine, Highly predictable, LSTM, Dropout algorithm, Deep Learning, Local Error-Correction.

---

1. **Introduction.** China is currently the world's largest coal consumer and producer, and in the longer term, the status of coal as the main energy source in China will not change, due to the low level of automation and reliability of coal mining equipment, a large number of equipment relying on manual operation of underground operators, resulting in coal production is still a high-risk industry. Therefore, there is an urgent need to achieve the goal of less humanized or even unmanned underground workings [1] and to improve the level of automation of mining equipment in order to improve the safety of coal production.

The coal mining machine is one of the core equipment to complete coal mining operations, and the realization of automatic cutting of the coal mining machine is the key to influencing the automation level of comprehensive mining working face [2]. Influenced by the complex working environment under the mine, it is difficult for the coal miner to adjust the drum cutting height adaptively. With the in-depth exploration and research on the principle of the coal-rock interface by domestic and foreign scholars. Currently, the methods to adjust the height of the drum are roughly divided into the following two categories. The first one is the adjustment of cut-off drum height by direct recognition of the coal-rock interface through sensors. These include the ray detection method [3], infrared detection method [4], image recognition method [5], and other technologies. However, a large number of industrial tests show that most of the above methods have the limitation of coal rock identification due to the influence of the complex environment of the comprehensive mining working face. For example, the ray detection method does not apply to the working face with too much gangue in the coal seam, and the infrared detection method applies to the coal seam with a hard roof. Therefore, the use of the direct detection method does not have good universal applicability [6]. The second one is to determine the coal-rock partition indirectly through a coal mining machine working the information and parameters, which mainly includes the memory cut-off method etc. The principle of the memory cut-off method is to predict the cut-off trajectory of a limited number of coal mining cycles afterward by manually showing the stored coal mining working parameters, thus avoiding the difficulty of directly identifying the coal-rock interface [7]. However, when the remembered cut-off path is inconsistent with the actual coal seam conditions, manual operation is required, leading to frequent downtime and reducing productivity and equipment life, so it is not ideal for coal mining applications.

With the rapid development of deep learning technology, it has become relevant to our lives, and it is found in the fields of medicine [8, 9], finance [10], urban traffic congestion [11], and so on. To improve the fit between the cut-off path and the actual coal seam. Wang et al. [12] proposed an iterative learning control algorithm to track and control the cut-off drum trajectory, and the target trajectory is accurately tracked by an adaptive iterative algorithm. Zhang et al. [13, 14] used a particle swarm algorithm and genetic algorithm for optimizing the memory cutoff path to achieve the path optimization of coal mining machines quickly and effectively. Wang et al. [15] used an adaptive weight particle swarm algorithm to solve the affiliation function of the optimized cutoff signal to achieve accurate identification of the coal-rock interface. Li et al. [16] obtained the predicted value of coal miner height by establishing a gray prediction model and coupling a Markov chain stochastic process to improve the prediction accuracy. Data-driven type of methods are applied to coal miner height prediction in the above-related studies, and these methods have the advantage of more accurate prediction only when the training samples are sufficient because they do not require complex degradation models. However, the mechanism of action and influencing factors of roller height present nonlinear, uncertain, and mechanical characteristics, and the model suffers from poor generalization ability and insufficient prediction ability when the training samples are insufficient. At present, deep learning is widely used in many fields because of its good stability and high prediction accuracy. It is not difficult to find that coal miner height prediction is a sequential prediction problem, which is suitable for deep learning-related techniques. Chen et al. [17] proposed an MSLSTM neural network by improving the neural network structure, which improved the model's multi-step prediction capability for the cut-off height but did not consider the problem that the neural network is prone to overfitting with the increase of the number of layers. It is well known that model prediction reflects the relationship between input and output values, but the raw data information of error series is often easily ignored. Yu et al. [18] constructed a prediction model by adding error factors to time series prediction; Xiao et al. [19] proposed a method based on the local climbing error correction of wind speed to improve the lag of predicted wind speed, both of which can further improve the accuracy of prediction.

In this paper, we first use a long and short-term memory-based neural network model for trajectory prediction, while optimizing the hyperparameters of the neural network to improve the prediction accuracy. We also introduce the Dropout optimization algorithm to deactivate some neurons and thus avoid the overfitting problem during network training. In addition, a local error correction scheme is used to solve the prediction lag and large prediction error caused by the unsteady nonlinear raw truncated data to achieve accurate prediction.

2. **Related work.** Recurrent Neural Network (RNN) [20] introduces a temporal recursive design, but suffers from the problem of gradient explosion and gradient disappearance, resulting in RNNs that do not have long-term memory function. To solve this problem, Hochreiter et al. [21] proposed the LSTM model, which, compared with the traditional recurrent neural network, adds memory units to each neuron and controls the degree of forgetting and memory capacity of time-series memory information through gating units, achieving better convergence and avoiding the problem of falling into local minima easily due to too fast gradient descent. The peephole LSTM [22], which is the popular variant structure of the LSTM, is used as the neural network unit in this paper, with the addition of the peephole connection compared to the LSTM, which means that the control gate is made to accept the input of the united state. The underlying structure is shown in Figure 1, where h and g are the input and output activation functions, respectively, usually, the *tanh* function; σ is the gate activation function, usually, the *sigmoid* function; and C is the unit memory state.

The LSTM accepts the cell input x_t at the current moment t and the cell output h_{t-1} at the previous moment $t-1$ through input gates, output gates, and forgetting gates. After updating the cell memory state using the input and forgetting gates, the LSTM cell output is controlled by the output gates and the nonlinear activation function. The LSTM cell update is as follows.

- The input gate i_t is:

$$i_t = \sigma(W_f[h_{t-1}, x_t] + P_f \circ c_{t-1} + b_f) \tag{1}$$

- The oblivion gate f_t is:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + P_f \circ c_{t-1} + b_f) \tag{2}$$

- The output gate o_t is:

$$o_t = \sigma(W_o[h_{t-1}, x_t] + P_o \circ c_{t-1} + b_o) \tag{3}$$

- The unitary memory state $\hat{c}_t$ is:

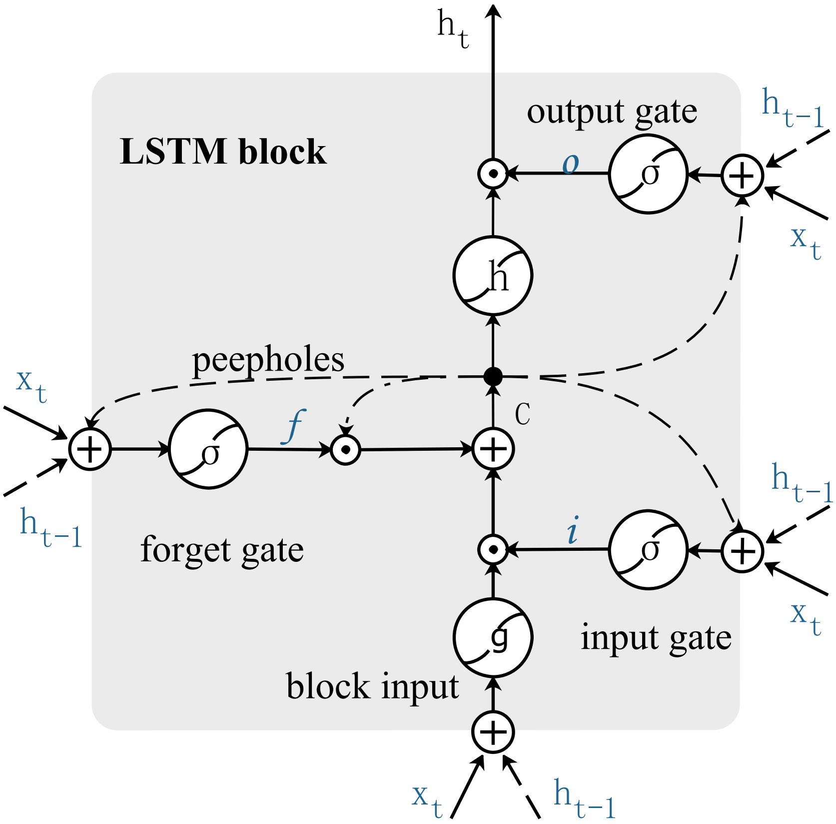

Figure 1. LSTM unit

$$\hat{c}_t = tanh(W_c[h_{t-1}, x_t] + P_c \circ c_{t-1} + b_c) \tag{4}$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \hat{c}_t \tag{5}$$

- The cell output h_t is:

$$h_t = o_t \circ tanh(c_t) \tag{6}$$

Where c_{t-1} is the unit memory state at the previous moment, $\circ$ is the Hadamard product, W_i, W_f, W_o, W_c, P_i, P_f, P_o and P_c, are the input gate, forgetting gate, output gate, and the current unit state corresponding to the weight moments b_i, b_f, b_o and b_c are the corresponding bias matrices that need to be trained for parameter search.

Deep neural networks can capture better data features through learning, and not only can automatically extract rules between data but also have a powerful nonlinear fitting ability. Therefore, to maximize the mining of nonlinear relationships between data and improve the sequence prediction accuracy, the deep LSTM neural network is chosen to predict the truncated trajectories. Its simple network structure is shown in Figure 2. The deep network model has better generalization capability compared to the single-layer LSTM network, as each layer of the former learns the input data, and its output value is also used as the input value of the next layer. The choice of a deep LSTM neural network facilitates better learning, feature extraction, and classification capability of the model.

Each LSTM cell in the model can be viewed as a sequence of different truncated trajectories. Before the model training, the generation of random numbers between 0 and 1 in the initial state is completed as the weight matrix with bias term assignment. Finally, the model is constructed and then the training process of the network is completed by the

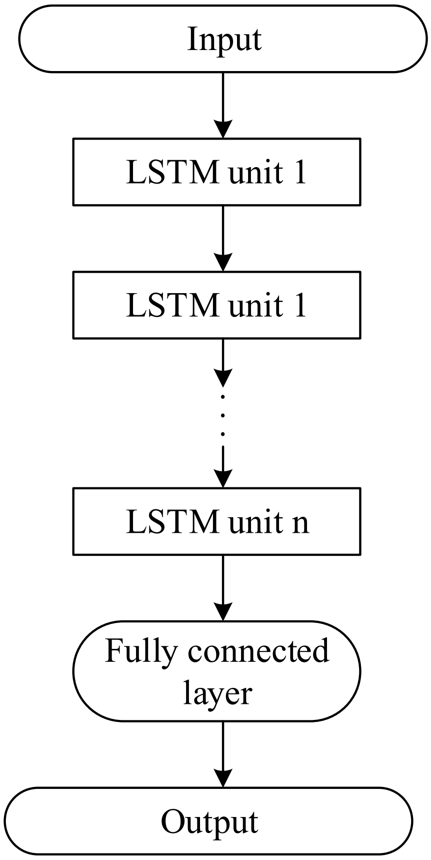

Figure 2. Deep LSTM network structure

adaptive moment estimation algorithm [23]. In the subsequent subsections, the improved Dropout algorithm and local error correction are used to optimize the training process of the network model.

3. Dropout_LSTM_LEC height prediction model construction.

3.1. **Dropout optimization algorithm.** The LSTM unit poses the following problems due to the presence of three gating structures. Firstly, the network structure can be very complex as the network depth increases. Secondly, for larger-scale deep learning networks, the amount of truncated trajectory height data is small. In summary, for the complex network structure and the small number of samples, the overfitting phenomenon is easy to occur, whereby the trained prediction model is more accurate in the training set and less accurate in the test set. In this regard, the Dropout algorithm is introduced.

The dropout algorithm is a common method to prevent the overfitting of neural networks [24], which is based on the principle of letting some neuron activation values be set to 0 with a certain probability during the training process. Figure 3 below shows the comparison before and after applying the Dropout algorithm to neural networks, where the dashed part indicates the neurons after deactivation, which can make the prediction value not too dependent on local features and weaken the inter-neuron. The joint adaptation is used to improve the generalization ability of the model.

The Dropout algorithm does not need to be used on a test set, as the test results are not expected to be random, otherwise, the prediction results would be compromised. However, the training process requires the generation of each neuron deactivation probability through Bernoulli distribution, which leads to the growth of the Dropout algorithm model

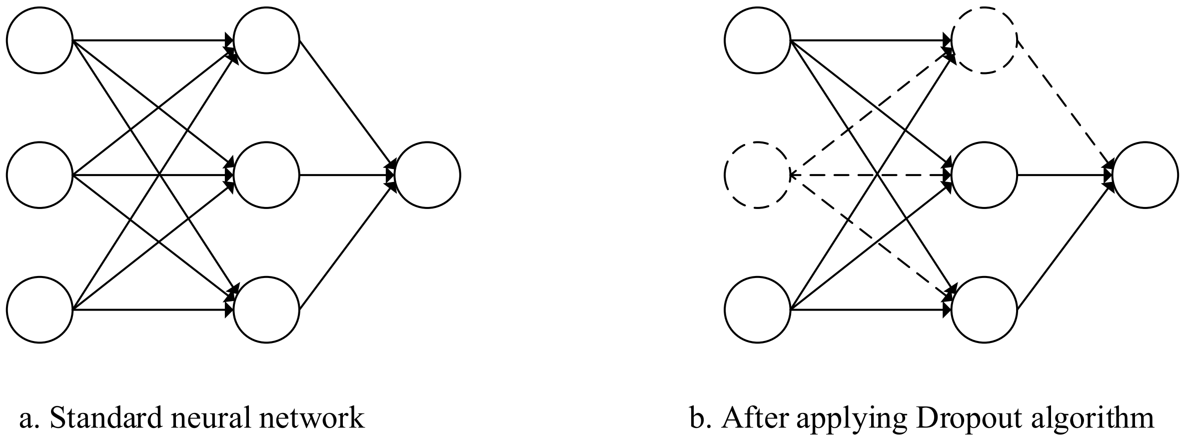

FIGURE 3. Before and after comparison of neural networks using Dropout algorithm

training time [25], and the effectiveness of the Dropout algorithm is reduced when the number of samples is small [26]. Therefore, the proposed Dropout optimization algorithm, which randomly discards neurons more than the Dropout algorithm, the optimization algorithm will preferentially deactivate the neurons with higher correlations to improve the model generalization ability. To calculate the inter-neuron correlation, the variance of the output value of each neuron needs to be calculated, after which the vector acting on the neurons in its lower layer can be obtained by calculating the output of a node in a certain layer. The variance of the output values of each layer neuron in a deep LSTM neural network is calculated as follows:

$$s_i^2 = \sum_x (e_{it} - \bar{e}_i)^2 \tag{7}$$

Where, e_{it} denotes the output of the i^{th} neuron at acquisition point x, and $\bar{e}_i$ denotes the expected value of all the outputs of the first neuron. The correlation between the i^{th} neuron and the j^{th} neuron is calculated as follows:

$$R_{ij} = \frac{1}{\sqrt{s_i^2}} \frac{1}{\sqrt{s_j^2}} \sum_t (e_{it}e_{jt} - te_ie_j) \tag{8}$$

The correlation between each neuron is calculated according to Equation (8), while a threshold value Θ, to distinguish strong and weak correlations, is introduced and set by network complexity and experience, and is generally a value between 0 and 1. The activation and deactivation states of neurons are denoted by 1 and 0, respectively, and the update state of neurons during network training is calculated as follows:

$$U_i(t+1) = \begin{cases} \sum_j R_{ij}U_j(t) \geq \theta, & 0 \\ else, & 1 \end{cases} \tag{9}$$

Where $U_j(t)$ is the activation state of j neuron at moment t and θ is a constant.

3.2. **Local error correction.** When there is a local abrupt change state of coal mining height and an upward and downward trend in a short period, the prediction results will have obvious errors compared with the real values. In addition, due to the randomness of the coal mining height trajectory, the prediction results are prone to extreme values. To reduce the errors caused by local extremes and abrupt change states and further improve the prediction effect, a local error correction is proposed. Assuming that two adjacent collection points are x and $x+1$, the absolute value α of the corresponding coal mining height difference is calculated as follows:

$$\alpha = |t(x+1) - t(x)| \tag{10}$$

Where $t(x)$ and $t(x+1)$ denote the height of the coal miner at the collection point x and $x+1$, respectively. When satisfying $\alpha \geq \lambda$, x is defined to be a local mutation point. If x is a local mutation point or an extreme value, the correction scheme is calculated as follows:

$$\alpha \geq \lambda \tag{11}$$

$$c(x) = pred(x) + err(x) \tag{12}$$

Where, λ is the threshold constant, which will be obtained in the experimental part by finding the best. $c(x)$ is the corrected prediction value at x, $pred(x)$ is the height value predicted by the network model at x, and $err(x)$ is the error prediction value at x. The basic framework of the local error correction method proposed in this section is shown in Figure 4, where the error values are first obtained from the sequence of predicted and true values, then the error values are predicted, and then the predicted values are corrected.

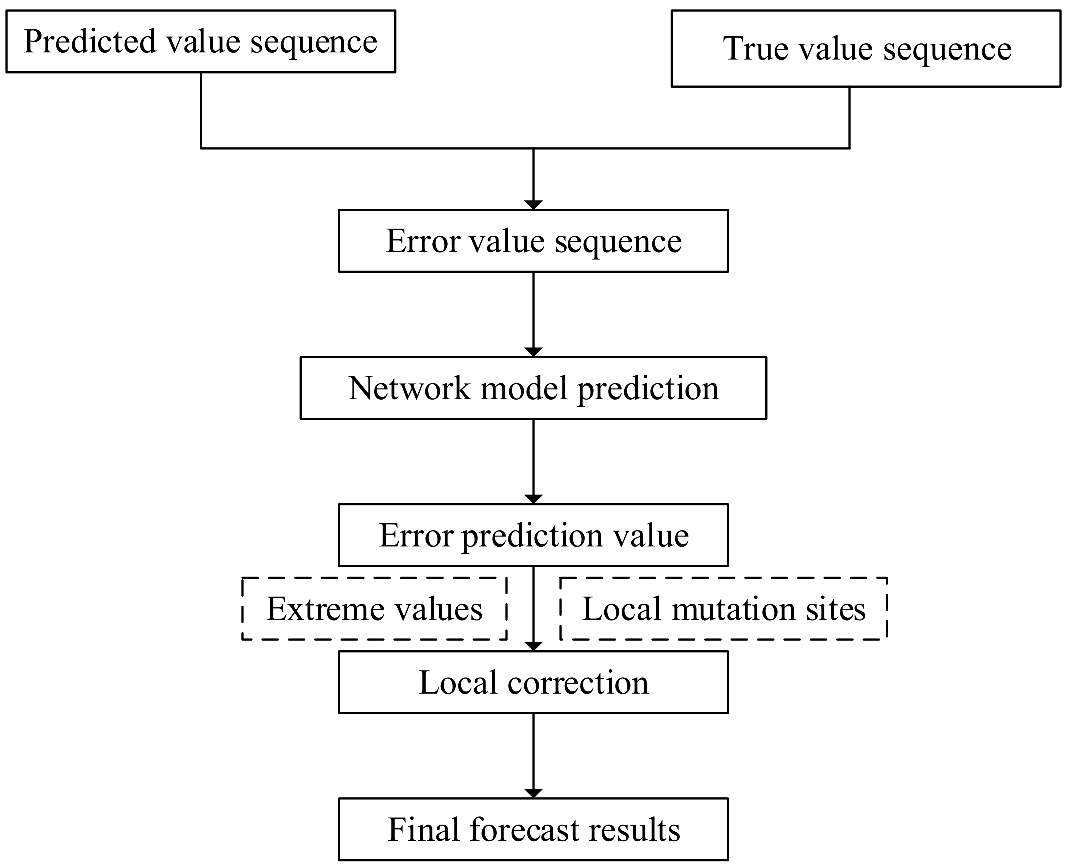

FIGURE 4. LEC Framework

3.3. **Dropout_LSTM_LEC model.** From the above analysis, it can be seen that the smaller number of height samples is not easy to fully explore the hidden information of the input sequence, which not only easily causes overfitting problems but also leads to the accuracy of the model prediction results. The Dropout_LSTM_LEC model proposed in this paper solves this problem better. The model builds a multilayer LSTM neural network based on the LSTM neural network to improve prediction accuracy. The improved Dropout algorithm is also introduced to improve the generalization ability of the model and avoid the model overfitting problem. The overall framework of the Dropout_LSTM_LEC model can be illustrated in Figure 5.

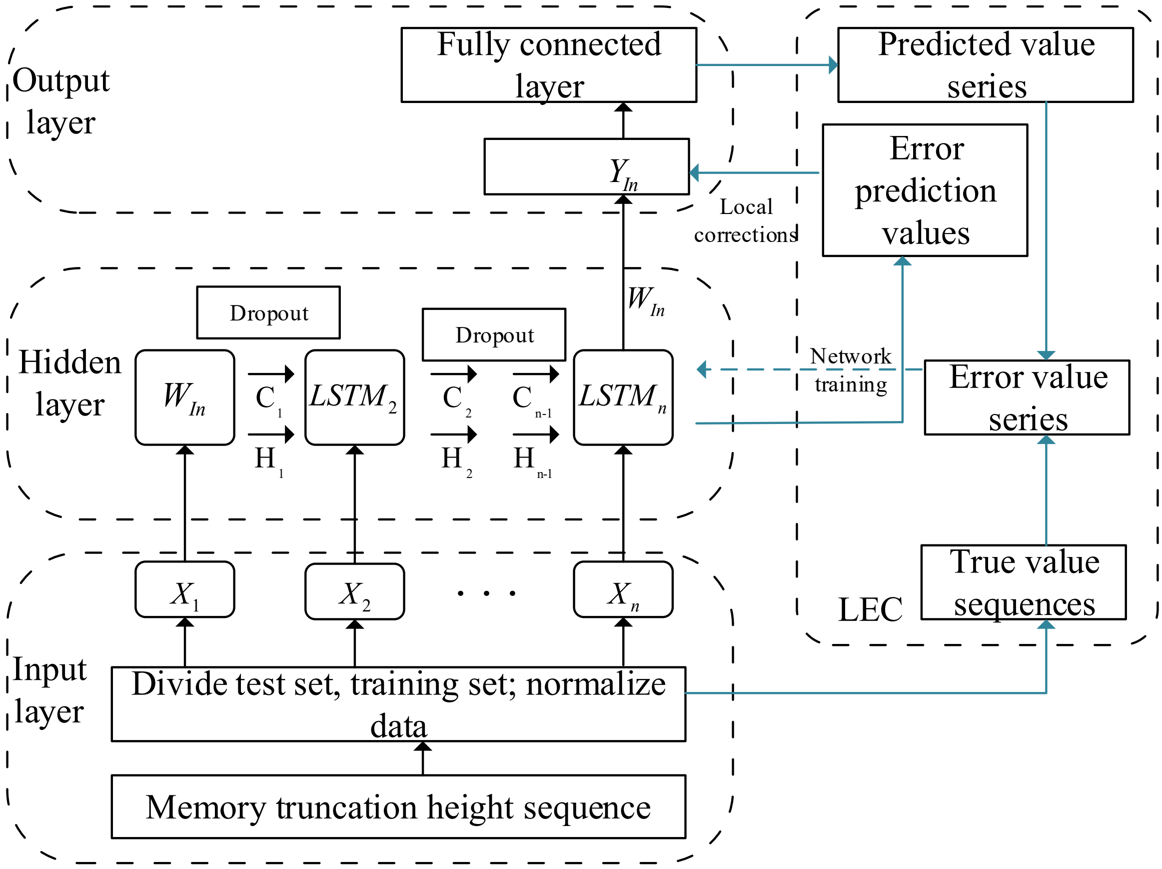

FIGURE 5. Dropout_LSTM_LEC model prediction framework

In the figure, X is the height of the roller in front of the coal miner at the acquisition point, C and H are the memory states and outputs of the LSTM unit, and Y is the output calculated by weighting. The basic training process of the Dropout_LSTM_LEC model is as follows:

Step 1: Obtain historical truncated data, preprocess the input data with normalization, and partition the data into a training set and a test set.

Step 2: Construct a deep LSTM network, and initialize the weight matrix and bias matrix in the network. Use Adam's algorithm to iteratively train the training set, and determine whether the error requirement is satisfied after each iteration, if not, continue to iterate and make the number of iterations plus one until the upper limit of iterations is reached and then reset the neural network, if satisfied, end the cycle. The training process is shown in Figure 6.

Step 3: Calculate the inter-neuron correlation according to Equation (7) and update the neuron activation state according to Equation (8) (the threshold value of the Dropout algorithm stage in this paper is taken as 0.3). At the same time, the appropriate hyper-parameters are selected by seeking the best.

Step 4: Determine if the truncated height data are not involved in the training. If yes, the new sample sequence is performed as in step 2. If no, the network parameter iteration is stopped and the results of the stage prediction are output.

Step 5: Calculate the network model to get the difference between the predicted and true value series, then use the network model to predict the error prediction value by the difference, and finally get the final prediction result after the local error correction of the mutation point and extreme value point according to the correction rule.

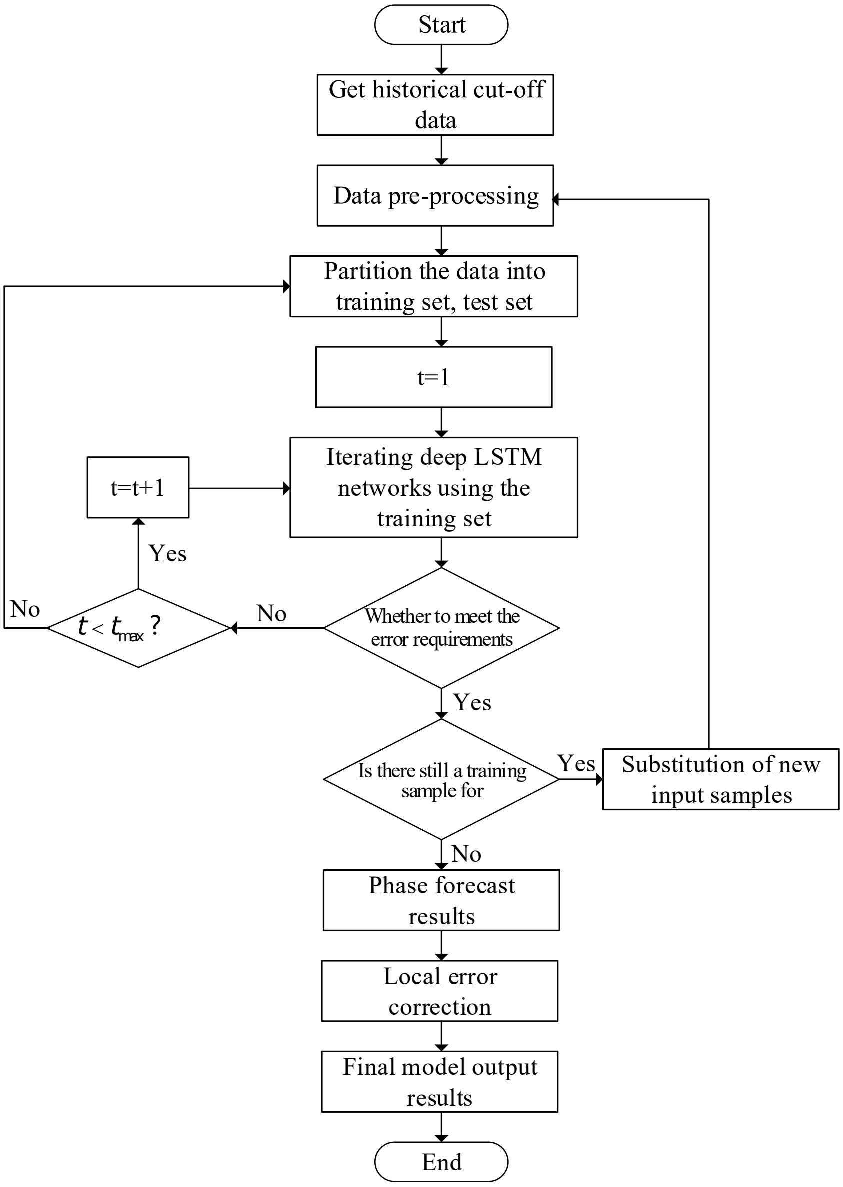

FIGURE 6. Dropout_LSTM_LEC model training process

4. **Analysis of experimental results.**

4.1. **Experimental data.** In the experiment, the data of coal miner operation in the actual production working face of a coal mine were selected. In the coal wall operation, to describe the coal rock distribution, 40 collection points were set for each cut, and one collection point was set every $3m$ to record the height of the center of the front drum of the coal miner. Table 1 data are the height data of the coal miner cutting at the collection point $x_1 - x_{40}$ cycles, and a total of 25 cycles were recorded. To shorten the model training time and improve the function convergence speed. Then, the training and test data are normalized, and the relevant calculation equations are as follows:

$$x'_{ab} = \frac{x_{ab} - x_{min}}{x_{max} - x_{min}} \tag{13}$$

Where x_{max} and x_{min} are the extreme and minimal values in the sample data, respectively, and x_{ab} is the roller center height value at the b^{th} collection point of the a^{th} cut.

TABLE 1. Coal rock distribution boundary collection point data ($/m$)

Collection points	First	Second	Third	Fourth	$\cdots$	Twenty-second	Twenty-third	Twenty-fourth	Twenty-fifth
x_1	3.03	3.11	3.04	3.23	$\cdots$	3.39	3.45	3.42	3.51
x_2	3.09	3.19	2.96	3.35	$\cdots$	3.34	3.41	3.50	3.44
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
x_{39}	3.64	3.74	3.69	3.78	$\cdots$	3.82	3.78	3.88	3.79
x_{40}	3.67	3.72	2.65	3.74	$\cdots$	3.85	3.90	3.81	3.84

4.2. **Experimental evaluation index.** In the experiments, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) were selected as evaluation indexes and defined in Equations (14-16). Where N is the number of test samples, $\hat{y}_i$ and y_i are the predicted and actual values of the i^{th} prediction point, respectively.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \tag{14}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i| \tag{15}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|\hat{y}_i - y_i|}{y_i} \times 100\% \tag{16}$$

4.3. **Model parameter setting.** To make full use of the latest data, we used the first two sets of drum height data to predict the drum height of the latter set and constructed a neural network prediction model with a two-dimensional input and one-dimensional output in a rolling prediction method. In addition, the number of network layers l, the number of neural units q in the hidden layer, and the mutation threshold λ in the local error correction in the LSTM model, have a great influence on the training effect, are used for parameter optimization.

TABLE 2. The prediction effect of taking different values of l and q

l values	q values	$RMSE/(mm)$	$MAE/(mm)$	$MAPE/(\%)$
2	8	121.60	111.41	3.654
2	16	112.41	106.91	2.968
2	32	113.23	107.16	3.068
3	8	102.75	91.32	2.862
3	16	93.82	84.08	2.668
3	32	99.62	93.82	2.960

4.3.1. *Selection of the number of network layers l and the number of hidden layer cells q.* Different l and q have different effects on the prediction effect. As l and q increase, the prediction effect tends to be better and then worse, indicating that underfitting occurs when the model parameters are too low and overfitting occurs when they are too high. Details are shown in Table 2.

From Table 2, it can be seen that the evaluation index is best when $l = 3$ and $q = 16$, and the best prediction is achieved in this case. Therefore, the number of network layers and the number of hidden layer cells in this paper are chosen to be 3 and 16.

4.3.2. *Selection of mutation threshold λ.* From the definition of the mutation threshold λ in the local error correction, the value of λ is extremely important. If the threshold λ is appropriately small it will lead to more local mutations in the global and ultimately better correction, but if λ is too small, it will ignore its error with the problem of over-correction, which will lead to poor correction instead. Therefore, we choose $\lambda = 0, 10, 20, 30, 40, 50$ for the experiment, and the evaluation results of the prediction when λ takes different values are shown in Table 3. The best prediction effect is obtained when $\lambda = 10$, so the mutation threshold is 10 in this paper.

TABLE 3. The prediction effect of taking different values of λ

λ values $/(cm)$	$RMSE/(mm)$	$MAE/(mm)$	$MAPE/(\%)$
0	95.60	84.14	2.241
10	**82.41**	**76.91**	**1.965**
20	91.63	78.84	2.036
30	89.84	79.64	1.974
40	96.82	89.08	2.191
50	102.62	91.82	2.387

4.4. **Experimental results and analysis.**

4.4.1. *Model prediction performance analysis.* In this paper, we use the stage prediction model Dropout-LSTM to continuously predict the 21st cut data and get the stage prediction results as shown in Figure 7. Meanwhile, the actual data of roller height are added to Figures 7 to 12 to compare the prediction effect.

From Figure 7, the predicted and realized values of the drum center height are generally close, but the predictions are poor at the abrupt change points in the truncated trajectory and the extreme value points, so the next stage of local error correction is performed. The

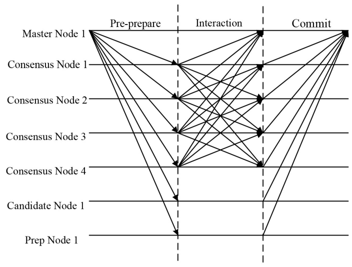

FIGURE 7. Dropout_LSTM model prediction effect

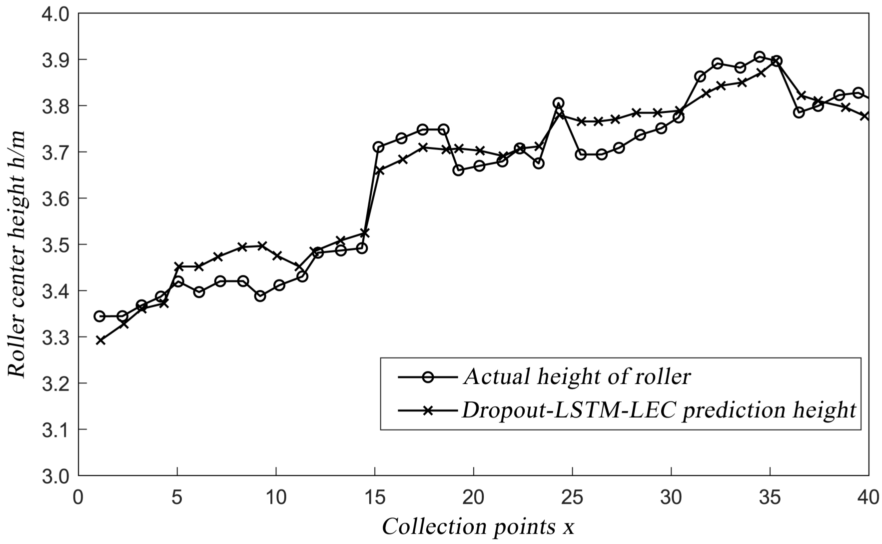

FIGURE 8. Prediction effect after local error correction

result of the model evaluation in this model is 93.82% for $RMSE$, 84.08% for MAE, and 2.668% for $MAPE$.The prediction effect with local error correction is shown in Figure 8.

The evaluation results of the model after local error correction are: $RMSE$ is 82.41, MAE is 76.91, and $MAPE$ is 1.965%. From Figure 8, the prediction performance of the model is better compared to the Dropout_LSTM model, which shows that the $RMSE$ improves by 13.85%, MAE by 8.5%, and $MAPE$ by 26.3%, indicating the effectiveness of the local error correction method in predicting the roller height in the model.

4.4.2. *Model Comparison.* To accurately evaluate the prediction performance of each algorithm, four other height prediction models: LSTM, deep LSTM, Gradient Boosting Regression Tree (GBRT), and Support Vector Regression (SVR) were selected and compared under the same data set, and the prediction results of each prediction model are shown in Figure 9-12, and their corresponding evaluation results are shown in Table 4. As can be seen from Figures 7 to 12, the Dropout_LSTM_LEC model predicts the data closest to the actual data, mainly because the model adds the optimized Dropout technique to the deep LSTM to alleviate the problem of overfitting of the network model, as well as adding the local error correction method to make the prediction model fully exploit the internal features of the error sequence, thus improving the model accuracy. The deep LSTM and LSTM models fit second best, because the LSTM neural network has unit state memory, which is more advantageous in the prediction of long series data, and the appropriate number of network layers can further improve the fitting effect. The GBRT and SVR models have the worst prediction results, especially in the collection point $x = [5, 15]$ with more obvious deviations, due to the more simplistic structure of the GBRT and SVR networks and the poor nonlinear fitting effect.

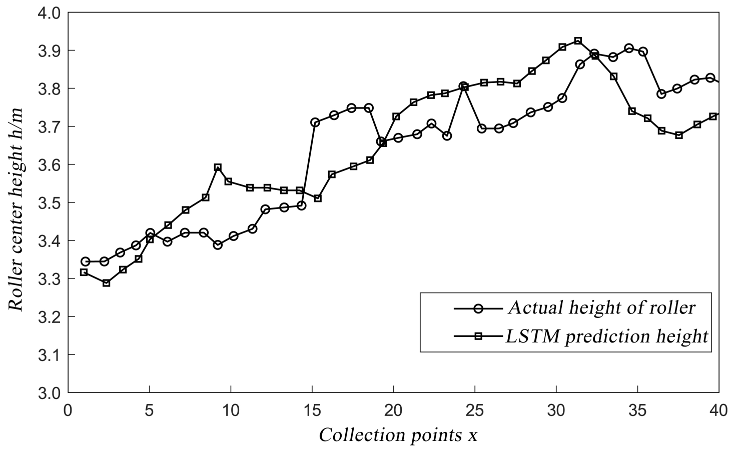

FIGURE 9. LSTM model prediction results

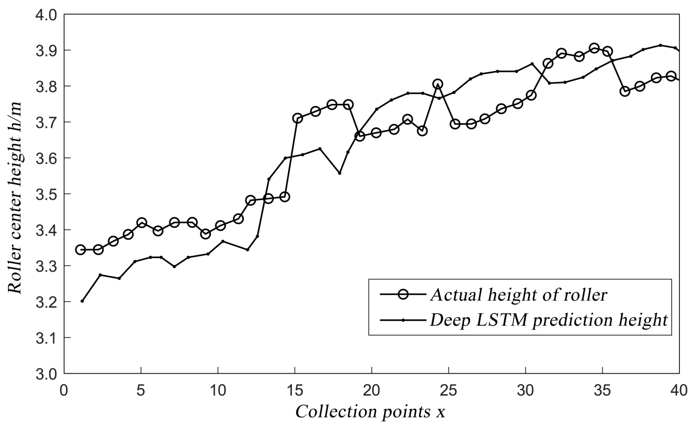

FIGURE 10. Deep LSTM model prediction results

TABLE 4. Comparison of prediction performance by models

Models	$RMSE/(mm)$	$MAE/(mm)$	$MAPE/(\%)$
GBRT	168.21	149.11	3.906
SVR	134.29	115.74	3.165
Single layer LSTM	129.02	110.84	2.836
Deep LSTM	108.84	99.64	2.274
Dropout_LSTM	96.82	88.08	2.191
Dropout_LSTM_LEC	**82.41**	**76.91**	**1.965**

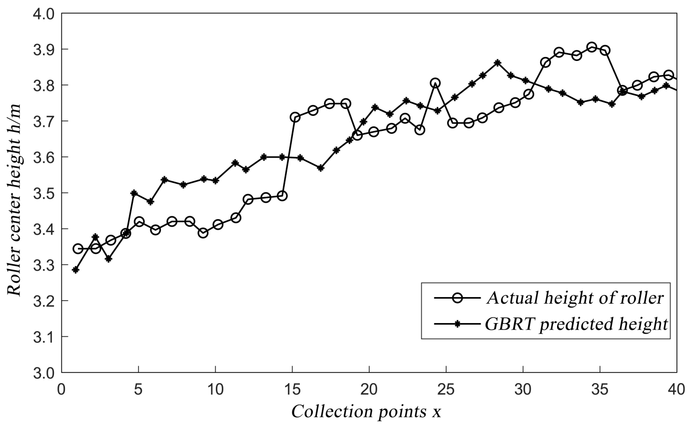

FIGURE 11. GBRT model prediction results

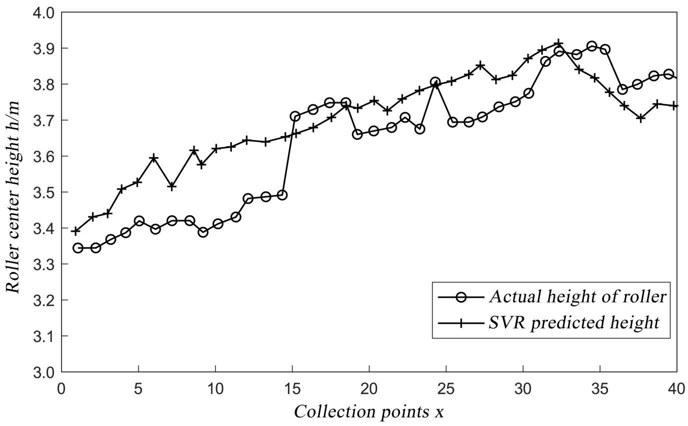

FIGURE 12. SVR model prediction results

As can be seen from Table 4, each evaluation result of the Dropout_LSTM_LEC model is better than the other five models, so the model can better extract the features in the nonlinear height data. Comparing the evaluation results of the last four prediction models in Table 4, we can see that extending the single-layer LSTM to a multi-layer network, introducing the Dropout technique, and adding local error correction can effectively improve the prediction accuracy of the model. the Dropout_LSTM_LEC model has improved the prediction ability compared with the first two prediction models in Table 4, in which the $RMSE$ is reduced by 51.0% and 38.6%, MAE decreased by 48.4% and 33.5%, and $MAPE$ decreased by 49.7% and 37.9%. In summary, the Dropout_LSTM_LEC model is more advantageous than other prediction models for the prediction of nonlinear data with truncated height.

5. **Conclusion.** In this paper, we propose a Dropout_LSTM_LEC model based on the LSTM model to solve the problem of low accuracy of coal miner memory cutoff. Firstly, we extend the original LSTM to multiple layers to improve the feature extraction and learning ability of the model; secondly, we optimize the network by using the improved Dropout technique to prioritize the nodes with strong deactivation correlation, which

solves the problem of easy overfitting due to the small amount of data at the center of the roller and the more complex structure of the multi-layer network, and reduces the dependence of the model prediction results on local nodes; finally, the We take into account the characteristic of local mutation in the truncated trajectory, and propose a local mutation correction method to correct and compensate the model prediction value. The experimental results show that the proposed architecture and method are feasible and effective, and the specific conclusions are as follows.

(1) The model has better generalization ability. Compared with the traditional neural network model, the Dropout_LSTM_LEC neural network model has higher prediction accuracy and effectively alleviates the problem of overfitting.

(2) The model has better prediction results. Due to the existence of unit memory structure and the proposed local error correction of LSTM, it has advantages in continuous nonlinear coal miner height prediction with $RMSE$ of 82.41, MAE of 76.91, and $MAPE$ of 1.965%, which are lower than GBRT, SVR, and LSTM.

However, this paper is only a preliminary study on the automation technology of coal mining machines based on neural networks, and we will further study the problems of coal rock cut-off state recognition and automation controller in the future to achieve the purpose of accurate memory cut-off of coal mining machines.

Acknowledgment. This work is partially supported by the National Natural Science Foundation of China (No.61872126).

REFERENCES

[1] F. Zhang, and Z.-C. Xu, "A mining image reconstruction method based on residual neural network," *Journal of Coal*, vol. 44, no. 11, pp. 3614–3624, 2019.

[2] S.-R. Ge, Z.-B. Wang, and S.-B. Wang, "Study on key technology of internet plus intelligent coal shearer," *Coal Science And Technology*, vol. 44, no. 7, pp. 1–9, 2016.

[3] S.-L Bessinger, and M.-G. Nelson, "Remnant roof coal thickness measurement with passive gamma-ray instruments in coal mines," *IEEE Transactions on Industry Applications*, vol. 29, no. 3, pp. 562–569, 1993.

[4] J.-R Markham, P.-R Solomon, and P.-E Best, "An FT-IR based instrument for measuring spectral emittance of material at high temperature," *Review of Scientific Instruments*, vol. 61, no. 12, pp. 3700–3708, 1990.

[5] J.-P Sun, and J. She, "Wacelet-based coal-rock image feature extraction and recognition," *Journal of China Coal Society*, vol. 38, no. 10, pp. 1900–1904, 2013.

[6] E. Yang, S.-B Wang, and S.-R. Ge, "Study on the principle of hyperspectral recognition of coal-rock interface," *Journal of China Coal Society*, vol. 43, no. S2, PP. 646–653, 2018.

[7] C.-S. Liu, Q. Yang, and C.-H. Liu, "Simulation of shearer drum cutting with memory program controlling by fuzzy control," *Journal of Coal Science*, vol. 166, no. 7, pp. 822–825, 2008.

[8] Y.-R. Ma, Y.-J. Peng, and T.-Y. Wu, "Transfer learning model for the false positive reduction in lymph node detection via sparse coding and deep learning," *Engineering Village*, vol. 43, no. 2, pp. 2121–2133, 2022.

[9] K.-K. Tseng, Z. Zhang, and C.-M. Chen, "DNetUnet: a semi-supervised CNN of medical image segmentation for super-computing AI service," *The Journal of Supercomputing*, vol. 77, no. 4, pp. 3594–3615, 2021.

[10] M.-E. Wu, J.-H. Syu, and C.-M. Chen, "Kelly-Based Options Trading Strategies on Settlement Date via Supervised Learning Algorithms," *Computational Economics*, vol. 59, no. 4, pp. 1627–1644, 2022.

[11] F.-Q Zhang, T.-Y. Wu, Y.-O. Wang, R. Xiong, M. Peng, ang L.-L. Liu, "Application of Quantum Genetic Optimization of LVQ Neural Network in Smart City Traffic Network Prediction," *IEEE Access*, vol. 8, 2999608, 2020.

[12] F.-Z. Wang, and Y. Gao, "Tracking Control of Shearer Cutting-drum Using Iterative Learning," *Control Engineering of China*, vol. 25, no. 6, pp. 966–971, 2018.

[13] L.-L. Zhang, C. Tan, and Z.-B. Wang, "Optimization of Mnemonic Cutting Path for Coal Shearer Based on Corpuscular Group Algorithm," *Coal Science And Technology*, vol. 38, no. 4, pp. 69–71, 2010.

[14] L.-L. Zhang, C. Tan, and Z.-B. Wang, "Optimization on memory cutting route of coal shearer based on genetic algorithm," *Coal Engineering*, vol. 388, no. 2, pp. 111–113, 2011.

[15] H.-J. Wang, M.-D. Huang, and X.-Y. Gao, "Coal-rock interface recognition based on multi-sensor information fusion considering pick wear," *Journal of China Coal Society*, vol. 46, no. 6, pp. 1995–2008, 2021.

[16] W. Li, "Memory cutting of adjacent coal seams based on a hidden Markov model," *Arabian Journal of Geosciences*, vol. 7, no. 12, pp. 5051–5060, 2014.

[17] W.-H. Chen, P.-F. Nan, and X.-H. Yan, "Prediction and model optimization of shearer memory cutting trajectory based on deep learning," *Journal of China Coal Society*, vol. 45, no. 12, pp. 4209–4215, 2020.

[18] Z.-J. Yu, S.-L. Yang, and X.-J. Wang, "Research on error correction method based on data transformational GM(1,1)," *Systems Engineering-Theory* & *Practice*, vol. 35, no. 09, pp. 2339–2347, 2015.

[19] Y. Xiao, C.-H. Li, and R.-P. Li, "Optimal wind power prediction based on local ramp error correction of wind speed," *Electric Power Automation Equipment*, vol. 39, no. 03, pp. 182–188, 2019.

[20] G.-E. Hinton, S. Osindero, and Y.-W. Teh, "A Fast Learning Algorithm for Deep Belief Nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2014.

[21] S. Hochreiter, and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[22] F.-A Gers, and J. Schmidhuber, "Recurrent nets that time and count," in *IEEE-INNS-ENNS International Joint Conference on Neural Networks*. IEEE, 2000.

[23] D. Kingma, and J. Ba, "Adam: A Method for Stochastic Optimization," *Computational Research Repository*, vol. 2014, 1412980, 2014.

[24] X.-X. Ren, "ST Segment Waveform Classification Algorithm Based on Dropout Deep Convolutional Neural Network," *Journal of Transduction Technology*, vol. 31, no. 8, pp. 1217–1222, 2018.

[25] N. Srivastava, G. Hinton, and A. Krizhevsky, "Dropout: A simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[26] W.-W. Shen, Y. Li, and Z.-H. Yang, "Attribute reduction with avoiding overfitting," *Application Research of Computers*, vol. 37, no. 9, pp. 2665–2668, 2020.