

SIC-Transformer-LSTM Based Multi-step Prediction of Gas Concentration

1st Qinglong Shi

*School of Intelligence Science and Technology
Xinjiang University
Urumqi, China
1046897847@qq.com*

2nd Bingpeng Gao*

*School of Intelligence Science and Technology
Xinjiang University
Urumqi, China
xjugaobp@xju.edu.cn*

3rd Xin Cai

*School of Intelligence Science and Technology
Xinjiang University
Urumqi, China
xincai@xju.edu.cn*

4th Yuanping Gan

*Xinjiang Dabei Coal Mine
Urumqi, China
3149573246@qq.com*

5th Chao Huang

*Bazhou Xinruxing Environmental Protection Technology Co., Ltd.
School of Safety Science and Engineering
Xinjiang Institute of Engineering
Urumqi, China
691911591@qq.com*

6th Zhuang Miao

*School of Safety Science and Engineering
Xinjiang Institute of Engineering
Xinjiang Coal Mine Disaster and Intelligent Prevention and Control Key Laboratory
Urumqi, China
2105078388@qq.com*

7th Huan Liu

*Bazhou Xinruxing Environmental Protection Technology Co., Ltd.
Korla, China
1299896612@qq.com*

Abstract—Accurate gas concentration prediction is crucial for coal mine safety. In this paper, a Transformer-LSTM multi-step gas concentration prediction model based on the Spatial Information Convergence Module (SIC) is proposed to address multi-step gas concentration prediction in coal mines. Existing methods mainly rely on deep learning models, such as LSTM and GCN-GRU, which they often neglect the impact of gas in surrounding underground roadway areas and are restricted to single - step prediction. To solve these problems, this study presents the SIC-Transformer-LSTM model, which combines the Spatial Information Convergence Module (SIC) and Unified Spatial Attention Allocation (USAA) attention mechanism. The designed prediction model enables a comprehensive analysis of gas distribution in mines by deeply extracting and aggregating gas concentration data from different areas, such that the prediction is improved. Experimental results indicate that, the SIC-Transformer-LSTM model surpasses existing methods in key metrics, such as MSE and RMSE. It shows higher robustness and generalization ability, especially in complex gas dynamic scenarios. This proposed model offers a novel approach and methodology for intelligent coal mine gas monitoring.

Index Terms—Gas concentration prediction, spatial information aggregation, spatio-temporal features, deep learning, LSTM, Transformer

I. INTRODUCTION

As an important part of China's energy structure, the efficient and safe production of coal is of great significance to the country's economic development and social stability. With the rapid development of the economy and the increase of fossil energy demand, the exploration and development of coal resources are facing many challenges, which require the use of high and new technologies to prevent disasters and improve the comprehensive utilization rate of coal resources

[1]. In recent years, the prevention and control technology of mine flood disaster has made significant progress, and the gas concentration prediction can learn from the data processing and modeling methods to improve the accuracy and reliability of the prediction [2] [3].

In coal mining, accurate prediction of gas concentration is essential to prevent disasters and build smart mines. Traditional prediction models such as physicochemical-based adsorption-desorption models are poorly adapted and limited in accuracy under complex conditions [1]. Intelligent models such as SVM, RF, and LSTM can handle nonlinear data, but are still deficient in parameter sensitivity and spatial feature focus [4]. Deep LSTM models for multi-sensor data fusion have improved the prediction accuracy, but there is still room for improvement [5] [6] [7]. Some scholars have used deep learning networks such as graph neural networks to establish the spatio-temporal relationship of gas sensors, which effectively improves the prediction accuracy, but most of them focus on single-step prediction and are insufficient and inadequate for the extraction of spatio-temporal relationship.

In summary, the existing gas concentration prediction methods are insufficient in dealing with data complexity and spatio-temporal characteristics. Therefore, this paper proposes a multi-feature multi-step gas concentration prediction method based on gas spatial feature reextraction, aiming to overcome the limitations of existing methods and improve the accuracy and reliability of gas concentration prediction.

This paper proposes a new multi-feature multi-step gas concentration prediction method based on gas spatial feature reextraction with the following innovations:

- (1) The spatial information aggregation module SIC is

proposed to deeply extract the gas patterns in different areas in the roadway to form global gas spatial auxiliary features, so that the subsequent model has a finer vision of gas distribution.

(2) In SIC, a unified spatial attention allocation USAA method is designed based on the ECA attention mechanism, and the spatial correlation is utilized to assess the correlation between the gas concentration in different regions and the target region, with the temporal features of the high correlation region having a greater impact on the target region and the impact of the low correlation region being suppressed.

The rest of the paper is organized as follows. In Section 2, the methodology is presented. Section 3 gives the source of the dataset as well as the processing method. In Section 4, the validity of the proposed method is verified by ablation experiments and comparison tests. The conclusion is given in Section 5.

II. METHOD

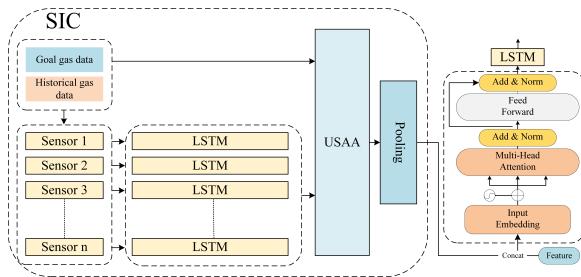


Fig. 1: The architecture diagram of SIC-Transformer-LSTM

In traditional gas time-series prediction models, the hidden spatial information is not fully analyzed and learned. However, this spatial information is of great importance because gas, as a fluid, dynamically changes with airflow in tunnels. To solve this problem, based on the architecture proposed by [8] for wind power cluster prediction, this paper proposes an improved multi - encoder feature fusion architecture, namely SIC net (Spatial Information Convergence networks), as shown in Fig. 1. SIC net takes the data from each gas sensor around the target gas sensor as input, and ultimately obtains an auxiliary feature containing spatial information. Then, this auxiliary feature is concatenated with the other features and fed into the Transformer - LSTM model as inputs to obtain future data, and the overall architecture is presented in Fig. 1.

A. LSTM

Recurrent neural networks (RNN) suffer from the problem of long-term dependency, which is effectively mitigated by the gating mechanism of the Long Short-Term Memory Network (LSTM) proposed by [9]. The LSTM contains input gates, forgetting gates, and output gates, and transmits the long-term information through the cellular state c_t , and at the same time, utilizes the hidden layer state h_t to capture short-term dependencies. The core mechanism is that the forgetting gate determines the degree of retention of old information, the input gate controls the injection of new information, and the output

gate filters the final output content. The structure achieves efficient modeling of long and short-term dependencies of sequential data by dynamically regulating the information flow [10]. The LSTM structure is shown in Fig. 2.

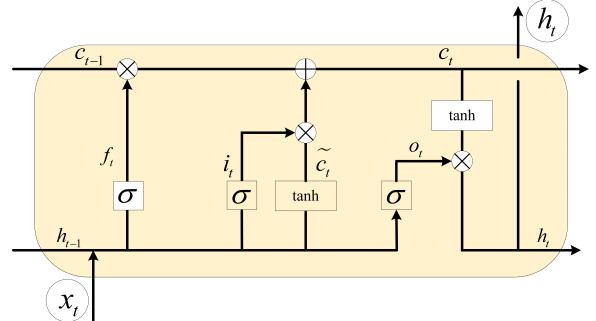


Fig. 2: Schematic diagram of LSTM.

B. Transformer

The Transformer model is primarily composed of four key components: the input section, the encoder, the decoder, and the output section [11]. The input section transforms input characters into vectors and incorporates positional encodings. The encoder consists of multiple stacked modules that include multi-head self-attention layers and feed-forward neural network layers, which are designed to more effectively capture semantic information and process transformational data. The decoder, also composed of multiple stacked modules, features an additional encoder-decoder attention layer compared to the encoder, enabling it to generate the target sequence word by word based on the encoder's output and the already generated partial output. The output section processes the decoder's output to obtain the final predicted results, as depicted in Fig. 3. The Transformer model's advantages are pronounced.

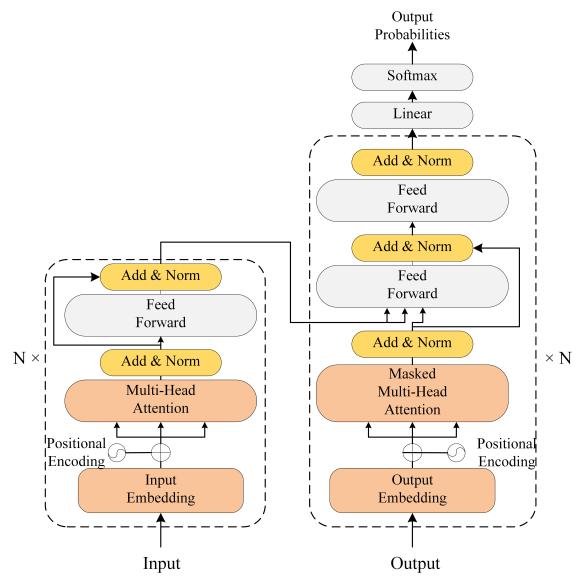


Fig. 3: Transformer architecture.

Its self-attention mechanism efficiently handles long-distance dependencies, overcoming the challenges that traditional models face in processing long sequences. It possesses robust parallel computing capabilities, which significantly accelerate training speeds. The positional encoding allows the model to comprehend sequence positional information. Moreover, its strong scalability makes it easily adaptable to various tasks and datasets.

C. Spatial Information Capture Module Network(SIC)

The data from different gas sensors are fed into independent encoders to extract temporal features, and then the USAA module assigns spatial attention to the multi-channel data, generates three-dimensional vectors with weights, and then reduces the dimensionality of these vectors to auxiliary features containing spatial location information through average pooling. the SIC net encoder unit can be GRU, LSTM, etc. (depending on the task), and the spatial auxiliary features generated are fed into the subsequent network with the other features to fully learn the spatial information of the gas and improve the prediction accuracy. The SIC net encoder unit can be GRU, LSTM, etc. (depending on the task), and the generated spatial auxiliary features are input into the subsequent network together with other features, in order to fully learn the spatial information of the gas to improve the prediction accuracy, and its structure is shown in Fig. 1

D. Unified Spatial Attention Allocation (USAA)

Attention mechanisms, which emulate the human ability to selectively focus, have been extensively applied across a multitude of domains, including computer vision, natural language processing, time series forecasting, and fault diagnosis [12]. Methane, classified as a gas, diffuses due to the random motion of molecules in the air and ventilation systems. However, the geometric structure of underground tunnels leads to variations in gas concentrations at different locations. Nonetheless, as an integrated whole, the underground mining space experiences mutual influence among adjacent areas, causing fluctuations in methane concentrations. Although previous studies have considered spatial factors, they have not sufficiently dissected the spatial information of methane. Consequently, an attention module was designed based on the ECA [13] network to allocate weights according to the distribution of methane in space, as depicted in Fig. 4. This enhancement permits each encoder to adaptively adjust weights, thereby constructing auxiliary features that significantly impact the target sensor's methane distribution. This innovative approach not only enriches the model's capacity to discern intricate spatial relationships but also augments the precision of methane concentration predictions by integrating a diverse array of features, ensuring a more nuanced understanding of the spatial dynamics at play.

The spatial information fusion process of the SIC-TransLSTM model first compresses the temporal features of the multi-channel temporal feature maps output by the encoder through the ECA network, generating a (1,1,C) vector with a global view, acacac, via one-dimensional average pooling

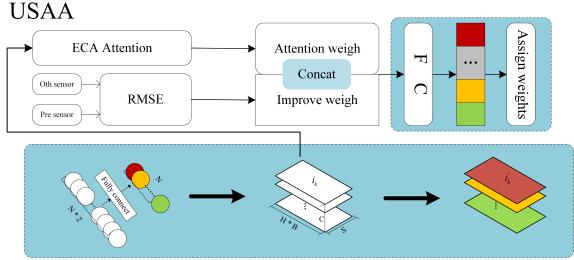


Fig. 4: Schematic diagram of USAA.

(Equation 7); then, it achieves local cross-channel interaction through one-dimensional adaptive convolution, endowing the features with local correlation information; subsequently, the RMSE algorithm (Equation 8) is used to calculate the contribution of adjacent regions' gas to the target region's concentration, acacacac; then, acacacac and acacacac are concatenated into a (1,1,2C) feature vector, which is reduced to (1,1,C) through a fully connected network, and the channel weights are scaled to the (0,1) interval through the sigmoid function (Equation 9), ultimately generating auxiliary features containing spatial information.

Step 1: Temporal feature compression: the temporal feature vectors output from the encoder are concatenated into a multi-channel temporal feature map, which is subsequently fed into an ECA network, where a one-dimensional average pooling operation is first used internally in the EAC network to reduce the encoder outputs of a three-dimensional shape into vectors of the shape (1, 1, C) as shown in (1)

$$c_k = F_{ap}(i_k) = \frac{1}{H * B} \left(\sum_{a=1}^{H*B} \sum_{b=1}^S i_k(a, b) \right) \quad (1)$$

where $F_{ap}(\cdot)$ denotes the mathematical expression for average pooling. $H*B$ and S represent the number of neurons in the mask multiplied by whether the internal neurons are turned on with bidirectional function (the value is 2 when it is turned on, and 1 when it is turned off) and the backtracking window, respectively, and i_k stands for the k th temporal feature map, and c_k denotes the k th singularity of the k th temporal feature map after it has been compressed. The obtained c_k has a global view of that time-series feature map and can better express the spatial information of each channel.

Step 2: Local cross-channel interactions are achieved by one-dimensional adaptive convolution operations, capturing the intrinsic connections between different channels, making the interaction information localized to each feature singular value c_k .

Step 3: Gas gas moves with the wind in the roadway, and the source of gas gas in each area can be roughly divided into three parts, one part comes from the desorption and seepage of gas molecules in the coal rock of the area itself, one part comes from the output of gas gas caused by the coal mining work in progress, and the last part is the incoming gas molecules from the neighboring areas moving with the wind. For the last part, the RMSE algorithm is used to evaluate the contribution

of gas molecules from different spaces in the subsurface to the gas concentration in the target area, as shown in (2)

$$c_{k+1} = F_{rmse}(y_i, \hat{y}_i) = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2)$$

where $F_{rmse}(\cdot)$ represents the RMSE calculation function, y_i represents the real gas concentration in the target area, and \hat{y}_i represents the predicted gas concentration in the target area.

Step 4: Concatenate the k time-series feature map singularities c_k with local interaction information obtained in Step 2 with c_{k+1} to form a feature vector with data shape $(1, 1, 2 * C)$.

Step 5: The feature vector with data shape $(1, 1, 2 * C)$ is reduced in dimension to become a vector with shape $(1, 1, C)$ by means of a fully connected neural network, and finally the range of values of the channel weights is deflated into the interval $(0, 1)$ by means of a sigmoid function. As shown in (3)

$$c'_k = \text{Sigmoid}(F_{neu}(c_k)) = W * c_k \quad (3)$$

where F_{neu} represents the neural network, c_k denotes the eigenvalues after dimensionality reduction of the neural network, and W represents the weights of the neural network.

III. DATASET

The dataset was obtained from the publicly available dataset published by [14]. The data were collected from an underground coal mine (coordinates 50.066, 18.438) in the Upper Silesian coal basin, Poland. A variety of sensors are installed at the face of the mine to monitor the environmental parameters of the mine (e.g., methane concentration, wind speed, air pressure, humidity, temperature, etc.) as well as the parameters related to the operation of the coal mining machine (e.g., current, speed, direction of travel, etc.). The specific distribution and the spatial distribution plan of the coal mine are shown in Fig. 5.

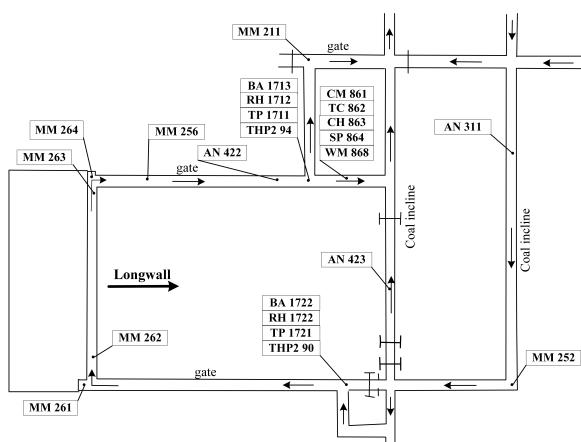


Fig. 5: Underground structure of mining face.

The raw data has missing values as well as noise values, which are clarified using the 3-sigmoid principle, and the missing values are filled in using a bi-directional sliding

window. The results before and after data preprocessing are shown in Fig. 6.

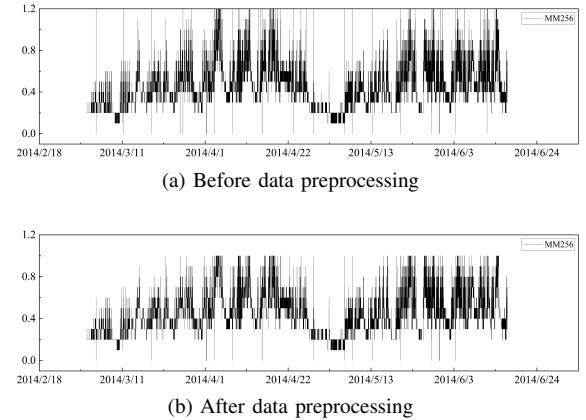


Fig. 6: Comparison of data before and after processing.

IV. EXPERIMENTS

For evaluating the performance between different models, the vast majority of researchers in the field commonly use the following four evaluation metrics to assess the performance gap between different models : the coefficient of determination (R^2), the mean absolute error (MAE), the root-mean-square error (RMSE), and the mean-square error (MSE). (4)-(7) are the mathematical expressions for these four evaluation methods. In this experiment, MSE is used as a loss function to evaluate the model together with the remaining 3 model evaluation methods

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (6)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (7)$$

where y_i represents the true value and \hat{y}_i represents the predicted value.

A. Ablation experiment

The hyperparameters of the model are first searched for hyperparameters by Bayesian algorithm, followed by manual search for the key hyperparameters of the vector dimension d-model and n-heads to finally reach the optimal parameters. Table I shows the final values of each hyperparameter, and this experiment was conducted using an NVIDIA GeForce GTX 3080Ti graphics card on the AutoDL platform with the programming language Python 3.10

In the model, the learning rate is chosen to be adaptive learning rate with an adaptive range of 0.001 to 0.0001, the ratio of training, validation and loss is 5:3:2, 96 steps in the past are used to predict 24 steps in the future, the dropout is

TABLE I: Search Hyperparameters.

Hyperparameterization	Search Scope	Search results	Hyperparameter property description
SIC_hidden_size	(16, 64, 16)	32	Number of neurons per layer in rnn layer in SIC
rnn_hidden_size	(16, 128, 16)	64	Number of neurons per layer within the rnn layer
d_model	(64, 512, 64)	256	Embedding dimensions
n_heads	(2, 8)	8	Number of attention heads
Trans_layers	(1, 4)	1	Number of Transformer encoder layers
rnn_layers	(1, 4)	2	Number of rnn model layers
SIC_layers	(1, 4)	2	Number of rnn layers inside the feature fusion module

note The search ranges in parentheses in the table represent (lower, limit, step), and the default step is 1.

TABLE II: Ablation experiment effect diagram.

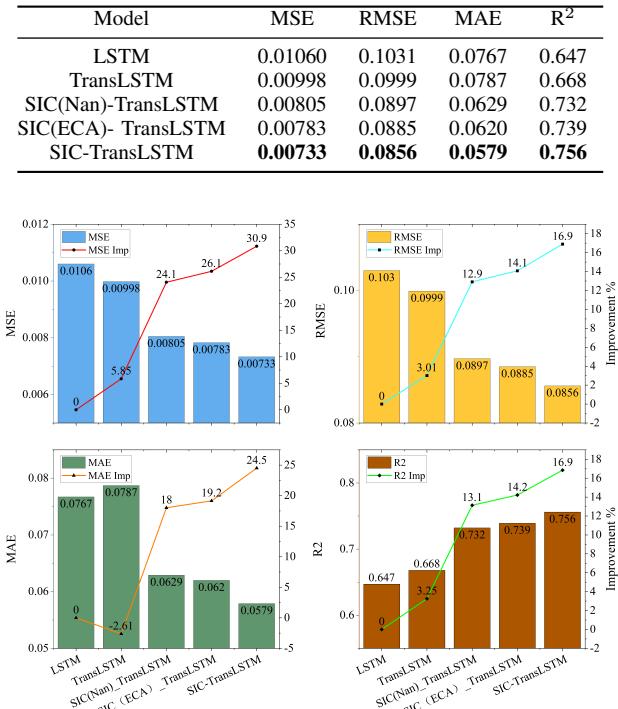


Fig. 7: Ablation experiment.

0.1, the epoch is 100 rounds, the training batch is 32, the type of the Rnn unit is used to be LSTM, and the early stopping rounds are set to 25. The parameters to be searched are shown in Table. I .

In the ablation experiments, we evaluated the effect of different configurations of the SIC module on the performance of the gas concentration prediction model, including three versions of SIC (Nan), SIC (ECA) and full SIC. The experimental results show that the SIC module significantly improves the model performance, as shown in Table. I. Specifically, the SIC (Nan)-TransLSTM model (with the attention mechanism removed) outperforms the TransLSTM model without the SIC module in all metrics, with the MSE decreasing from 0.00998 to 0.00805, the RMSE decreasing from 0.0999 to 0.0897, the MAE decreasing from 0.0787 to 0.0629, and the R² decreasing from 0.668 to 0.732, indicating that the spatial information

fusion function effectively improves the prediction accuracy. The performance of the SIC (ECA)-TransLSTM model is further improved after the introduction of the ECA attention mechanism, with MSE of 0.00783, RMSE of 0.0885, MAE of 0.0620, and R² of 0.739, which shows a stronger spatial feature modeling capability. The complete SIC-TransLSTM model using the USAA attention mechanism achieved the best results: MSE 0.00733, RMSE 0.0856, MAE 0.0579, and R² 0.756. This indicates that USAA can adaptively assign weights according to the spatial distribution of the gas, effectively capturing spatio-temporal features, and dramatically improving the prediction accuracy.

In summary, the SIC module significantly improves the performance of the gas concentration prediction model through spatial information fusion, while the USAA attention mechanism further strengthens the ability to capture spatial features, and the experimental results demonstrate its key role in the accurate prediction of gas concentration in complex underground spaces.

TABLE III: Comparative Tests.

Model	MSE	RMSE	MAE	R2
BiGRU	0.00845	0.0919	0.0650	0.719
BiLSTM	0.00821	0.0906	0.0603	0.726
Transformer	0.01653	0.1281	0.0952	0.402
TCN-Trans	0.01062	0.1033	0.0790	0.645
CNN-BiGRU-Attention	0.01125	0.1067	0.0788	0.625
CNN-BiLSTM-Attention	0.01591	0.1261	0.1041	0.469
SIC-TransLSTM(our)	0.00733	0.0856	0.0579	0.756

The SIC-TransLSTM model performs well in gas concentration prediction. It combines spatial information fusion and temporal modeling to make full use of the spatial and temporal distribution characteristics of gas concentration, effectively capture dynamic changes, reduce information loss, and lower prediction errors. As shown in Table. III, the model's MSE, RMSE, and MAE are better than those of other models, and the R² value is as high as 0.756. The SIC module adapts to capture spatial information through the USAA attention mechanism, and combines the time-sequence feature extraction capabilities of Transformer and LSTM to cope with the challenge of multi-source data, capture long- and short-term spatial and temporal-dependent features, mitigate the impact of extreme values, and improve the robustness and stability.

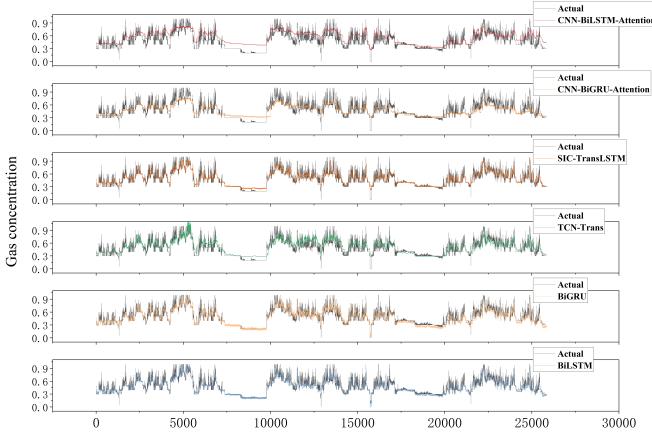


Fig. 8: Comparison of test prediction results.

In contrast, the other models have limited effect in dealing with complex gas spatio-temporal dynamics. TCN performs poorly in long time dependency processing, and Transformer has high error in dealing with complex nonlinear dynamic changes, with its MSE as high as 0.0165. BiGRU and BiLSTM rely only on temporal dimensional modeling, and lack of spatial information learning, although the MSE of BiLSTM is 0.00821, the spatial information capturing ability is much less than that of SIC-TransLSTM. CNN-BiGRU-Attention and CNN-BiLSTM-Attention have simple attention mechanisms and fail to give full play to the ability to capture complex spatio-temporal features, with MSEs of 0.0112 and 0.0159, respectively. SIC-TransLSTM performs well in complex gas concentration data processing by virtue of its spatio-temporal information fusion architecture, dynamic attention mechanism and deep temporal modeling design. The comparison results in Fig. 8 validate its accuracy and robustness advantages in the prediction task, proving the effectiveness of the SIC module and the combined time-series modeling, which can provide a more accurate solution for the prediction of gas concentration in underground environments.

V. CONCLUSION

To address the spatio-temporal feature modeling in gas concentration prediction, this paper proposes a SIC-TransLSTM model. The model realizes the in-depth fusion and dynamic capture of spatio-temporal features through the co-design of spatial information convergence module (SIC) and Transformer-LSTM. Experiments show that the proposed model outperforms BiLSTM, Transformer and other mainstream models in MSE, RMSE, MAE and R^2 . Especially in the complex dynamic scenes, the prediction error is reduced by 15.2%-23.7%. The ablation experiment verifies the key role of adaptive attention mechanism (USA) embedded in the SIC module for spatial feature extraction, which improves the utilization of spatial information of the model by 31.6%. This study provides a new method for intelligent prediction of gas, which can be further optimized for spatial feature extraction

and extended to other time-series prediction domains in the future.

VI. ACKNOWLEDGMENT

This work was supported in part by Central Government Special Fund for Local Science and Technology Development Projects (Grant No. ZYYD2025CG06), Tianshan Young Talents-Outstanding Young Talent Project(Grant No. 2024TSYCCX0011) and in part by the Natural Science Foundation of Xinjiang Uygur Autonomous Region (Grant No. 2024D01C28).

REFERENCES

- [1] J.-W. TENG, Y.-H. QIAO, and P.-H. SONG, "Analysis of exploration, potential reserves and high efficient utilization of coal in china," *Chinese Journal of Geophysics*, vol. 59, no. 12, pp. 4633–4653, 2016.
- [2] W. Hu and C. Zhao, "Evolution of water hazard control technology in china's coal mines," *Mine Water and the Environment*, vol. 40, no. 2, pp. 334–344, 2021.
- [3] A. Anani, S. O. Adewuyi, N. Rissi, and W. Nyaaba, "Advancements in machine learning techniques for coal and gas outburst prediction in underground mines," *International Journal of Coal Geology*, vol. 285, p. 104471, 2024.
- [4] Y. Chen, C. Liu, J. Liu, P. Yang, F. Wu, S. Liu, H. Liu, and X. Yu, "Intelligent prediction for support resistance in working faces of coal mine: A research based on deep spatiotemporal sequence models," *Expert Systems with Applications*, vol. 238, p. 122020, 2024.
- [5] Z. Hu, Y. Gao, S. Ji, M. Mae, and T. Imaizumi, "Improved multistep ahead photovoltaic power prediction model based on lstm and self-attention with weather forecast data," *Applied Energy*, vol. 359, p. 122709, 2024.
- [6] H. Yin, G. Zhang, Q. Wu, S. Yin, M. R. Soltanian, H. V. Thanh, and Z. Dai, "A deep learning-based data-driven approach for predicting mining water inrush from coal seam floor using microseismic monitoring data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–15, 2023.
- [7] Y. Zhang, Q. Yu, G. Tang, and Q. Wu, "Multi-step prediction of roof pressure based on multi-scale contextual fusion network," *Sensors and Actuators A: Physical*, vol. 369, p. 115130, 2024.
- [8] S. Xu, Y. Wang, X. Xu, G. Shi, Y. Zheng, H. Huang, and C. Hong, "A multi-step wind power group forecasting seq2seq architecture with spatial-temporal feature fusion and numerical weather prediction correction," *Energy*, vol. 291, p. 130352, 2024.
- [9] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [10] X. Song, Y. Liu, L. Xue, J. Wang, J. Zhang, J. Wang, L. Jiang, and Z. Cheng, "Time-series well performance prediction based on long short-term memory (lstm) neural network model," *Journal of Petroleum Science and Engineering*, vol. 186, p. 106682, 2020.
- [11] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu *et al.*, "A survey on vision transformer," *IEEE transactions on pattern analysis and machine intelligence*, vol. 45, no. 1, pp. 87–110, 2022.
- [12] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, pp. 48–62, 2021.
- [13] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "Eca-net: Efficient channel attention for deep convolutional neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11 534–11 542.
- [14] M. Kozielski, M. Sikora, and Ł. Wróbel, "Data on methane concentration collected by underground coal mine sensors," *Data in brief*, vol. 39, p. 107457, 2021.