Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 20, No. 1, 2026

Workface Methane Concentration Prediction Based on Transformer-KAN

Lin Wu, Daqi Wang, Hongjin Yu, Shengchen Zhang, Shangjing Fan, Ziye Liu, Yue Wang

North China University of Science and Technology, Tangshan Hebei, 063200, China

Abstract: This paper proposes an underground methane concentration prediction system for working faces based on the Transformer-KAN, addressing the challenge of insufficient spatial correlation modeling in multi-sensor data. The core innovation lies in leveraging the self-attention mechanism of Transformer networks to dynamically compute the influence weights of sensors (atmospheric pressure, temperature, etc.) at different locations within the roadway on methane concentration at target points. This overcomes the limitations of traditional models reliant on fixed-distance decay models. Simultaneously, utilizing the differentiable spline basis functions of the KAN network, the complex nonlinear physical mapping relationship between atmospheric pressure gradients and methane concentration is constructed and explained. High-precision sensor networks are deployed in critical areas of coal mine tunnels. Data acquisition and preprocessing employ techniques like precise time synchronization and Kalman filtering. The system ultimately integrates the accelerated inference engine to achieve minute-level (target 200ms high-precision (target RMSE$\leq$0.15%） prediction. Finally, real-time data from a mining face in a transport tunnel of a certain mine was selected to construct and validate on-site predictions, providing core technological support for the intelligent safety upgrade of coal mines and the "dual carbon" goals.

Keywords: Transformer-KAN; Underground Methane Concentration Prediction; Multi-Sensor Data Fusion.

1. Introduction

Coal remains China's primary energy source, yet coal mine gas hazards pose the foremost threat to mining safety. In recent years, intensified and deeper coal extraction has significantly increased gas outbursts from coal seams, leading to frequent gas limit and outburst incidents [10-11].According to the 2024 Annual Report of the National Mine Safety Administration: Nine gas accidents occurred, resulting in 55 fatalities, accounting for 39% of all coal mine fatalities nationwide. To achieve efficient coal production, we must accurately predict trends in gas concentration and develop reasonable strategies to prevent hazards from elevated gas levels.

In recent years, research on dynamic methane concentration prediction using machine learning algorithms has flourished, with numerous scholars conducting in-depth explorations of influencing factors and predictive models. Yang Hua's team [1] (2023) revealed through field measurements the nonlinear influence of atmospheric pressure changes on methane outbursts: when the atmospheric pressure decrease rate is below 0.04 kPa/h, the peak methane outburst lags behind the pressure trough; while above this threshold, the peak occurs earlier. This critical physical relationship remains underrepresented in existing intelligent warning models. Liang Yunpei et al. [2] optimized the LSTM hidden layer structure and neuron count using the cuckoo algorithm to uncover long-term patterns in methane concentration changes, achieving 12-hour predictions. Their CS-LSTM model demonstrated an RMSE of only 0.023, making it suitable for low-fluctuation conditions. Fan Jingdao et al. Rong et al· ([3]) combined Variational Modal Decomposition (VMD) with Bayesian-optimized BiLSTM to perform secondary noise reduction and feature enhancement on methane time-series data. This approach reduced prediction errors and peak false alarm rates in coal mining faces compared to single LSTM models. The U.S. NIOSH team [4] embedded the gas diffusion equation (mass conservation law) as a loss function constraint within an RNN, compelling the model to adhere to fluid dynamics principles. This fusion of physical equations and data-driven approaches resolves overfitting issues in small-sample training. Robert M.X. Wu's team at the University of Technology Sydney, Australia [5], developed a gas early warning system integrating a "bubble wall map" visualization tool with a "triple correlation analysis" framework. By uncovering implicit correlations between gas levels and temperature/wind speed, it identifies risks 10-15 minutes before gas concentration exceeds limits. This tool enables low-skilled operators to rapidly interpret complex data. The aforementioned studies primarily focus on single-point time-series modeling or static environmental features, with few investigations simultaneously considering multi-point spatiotemporal coupling and nonlinear atmospheric pressure transmission mechanisms.

In summary: Existing research has achieved certain results in analyzing methane concentration influencing factors and optimizing prediction models, with continuously improving prediction accuracy. However, current early warning technologies still have shortcomings in spatio-temporal correlation modeling. Traditional LSTM/CNN models struggle to quantify the dynamic coupling effects across multiple monitoring points (T0, T1, T2). For instance, the delayed impact of a sudden atmospheric pressure drop at location T1 on methane concentration at T0 can only be approximated using empirical coefficients, resulting in high error rates. Furthermore, most existing studies treat atmospheric pressure as a static disturbance term, overlooking its nonlinear transmission mechanism [14]. To address these limitations, this paper proposes a Transformer-KAN fusion architecture. The multi-head self-attention mechanism of the Transformer module effectively captures spatiotemporal dynamic coupling among multiple monitoring points, enabling precise quantification of cross-point lag effects. The

KAN module embeds the Langmuir adsorption equation using differentiable spline basis functions, enabling the model to adaptively fit the nonlinear physical relationship between atmospheric pressure and gas concentration. This addresses the generalization limitations of purely data-driven methods beyond physical boundaries. Finally, a gated modulation mechanism achieves dynamic coordination between the two modules: strengthening physical constraints during sudden pressure changes while focusing on temporal features during stable periods. This approach balances both the precision and timeliness requirements of early warnings.

2. Related Theories

2.1. Transformer Network

The Transformer is fundamentally a deep learning model based on self-attention mechanisms, designed for processing sequential data [7]. Unlike recurrent neural networks (RNNS) and their variants (LSTMS, GRUS), the Transformer completely abandons recurrent structures, relying solely on attention mechanisms for sequence modeling. This enables highly efficient parallel computation while capturing long-range dependencies with greater precision.

The core feature of the Transformer architecture is its encoder-decoder model. The encoder takes input and outputs a matrix representation of that input, while the decoder takes this encoded representation and iteratively generates output. Both encoders and decoders are essentially multi-layered stacks (with identical numbers of layers). All encoders share the same structure: input enters each encoder and is passed to the next one. All decoders also share the same structure, receiving input from the last encoder and the previous decoder.

In the model proposed herein, the encoder structure employs a "2-layer encoder + 4 attention heads + 128-dimensional hidden layer" as defined in the code. This parameter choice stems from the input sequence length seq_len=30 (30 time steps = 5 minutes), where a 2-layer encoder suffices to capture temporal dependencies within 5 minutes. The four attention heads enable parallel exploration of correlations between different features (atmospheric pressure, gas concentration), while the 128-dimensional hidden layer balances the computational capacity of the MX550 GPU with the model's expressive power, preventing memory overflow.

Temporal Encoding Approach: Temporal awareness is achieved through an "embedding layer + positional encoding." The embedding layer maps 5-dimensional input features (atmospheric pressure, atmospheric pressure differential, gas presence/absence) into a 128-dimensional vector matching Transformer input dimensions. These 5 features specifically include: Atmospheric pressure feature (atmospheric pressure), representing the absolute atmospheric pressure value monitored at the current time step, reflecting the baseline pressure state of the mine environment; Atmospheric pressure difference feature (atm_diff), representing the change in atmospheric pressure between the current and previous time steps, capturing instantaneous pressure fluctuation trends—as abrupt pressure changes often correlate closely with abnormal gas outbursts; and three methane concentration features (methane 0, methane 1,CH4_2), representing methane concentrations at three distinct monitoring points. Simultaneous multi-point monitoring provides a more comprehensive picture of methane distribution and migration patterns within the mine. Position encoding employs sine-cosine coding, calculated by the formula:

$$PE_{(pos,2i)} = \sin(pos/10000^{2i/128}),\ PE_{(pos,2i+1)} = \cos(pos/10000^{2i/128}),$$

to ensure the model identifies "which time step's atmospheric pressure has a greater impact on gas concentration."

2.2. KAN Network

Kolmogorov-Arnold Networks (KAN) represent a novel neural network architecture inspired by the Kolmogorov-Arnold Representation Theorem. Unlike traditional neural networks using fixed activation functions, KAN employs learnable activation functions at the network edges. This design allows each weight parameter in KAN to be replaced by a single-variable function, typically parameterized as a spline function. This provides exceptional flexibility, enabling the simulation of complex functions with fewer parameters and enhancing model interpretability.

Traditional weight parameters are replaced by univariate function parameters at the network edges. In KAN, each node aggregates these function outputs without nonlinear transformation. The flexibility of splines enables adaptive modeling of complex relationships in data by adjusting their shape, thereby minimizing approximation error and enhancing the network's ability to learn subtle patterns from high-dimensional datasets [13].

The SimpleKAN class is implemented using "3 fully connected layers + ReLU activation," with its core logic being the composite fitting of nonlinear relationships through univariate functions. For atmospheric pressure and gas concentration, the fitted relationship is

$$gas = \mathrm{ReLU}(W_3 \cdot \mathrm{ReLU}(W_2 \cdot \mathrm{ReLU}(W_1 \cdot atm + b_1) + b_2) + b_3),$$

where W_1, W_2, $W_3$are learnable weights, and b_1, b_2, $b_3$are bias terms, precisely aligning with the logic of "modulating attention outputs based on pressure change rates" in the code. To address fitting issues with noisy data, an "output scaling coefficient" was added to constrain KAN outputs within [-0.1, 0.1], preventing computational overflow from excessively large outputs. The KAN-modulated attention features are fed into a two-layer feedforward network for higher-order feature extraction. Finally, the output layer reconstructs the predicted methane concentration value.

3. Construction of the Working Face Methane Concentration Prediction Model

3.1. Overall Model Architecture

The model adopts a Transformer-KAN hybrid architecture, with its core design aimed at accurately capturing the dual relationship between atmospheric pressure and methane concentration: "long-term temporal dependencies + local nonlinearities." The overall workflow is as follows:

1. Preprocess data and input the preprocessed multi-feature data into the model;
2. The defined Transformer encoder captures temporal correlations, while the KAN module extracts nonlinear mappings;
3. Outputs from both modules are fused through a gating mechanism to generate methane concentration predictions;

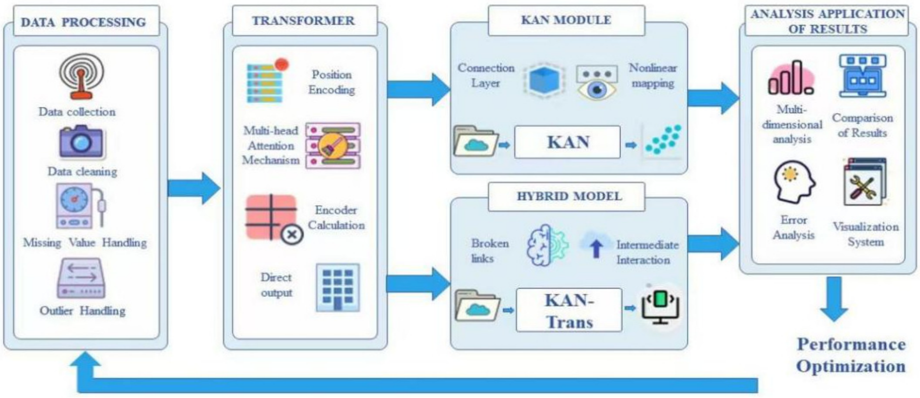

Figure 1. Time-series prediction architecture integrating Transformer and KAN

This architecture takes multi-source temporal data from coal mining workfaces (e.g., methane concentration, atmospheric pressure) as input. Through deep collaboration between the Transformer and KAN (Kolmogorov-Arnold Network), it achieves joint modeling of complex spatiotemporal correlations and nonlinear physical mechanisms. The input layer receives three types of inputs: shifted sequences, features, and position embeddings, providing the model with temporal context, data characteristics, and spatial information. Within the Transformer-KAN collaborative module, Transformer attention first captures long-term temporal dependencies and spatial correlations across multiple monitoring points.The KAN module employs scientific spline basis functions designed based on the Kolmogorov-Arnold Representation Theorem as activation units to achieve precise fitting of complex nonlinear mappings such as "atmospheric pressure-gas concentration." Based on cubic B-splines, this function uses adjustable coefficients to control its local response patterns across segmented intervals, enabling flexible characterization of the non-monotonic, asymmetric physical dependencies between pressure changes and gas concentrations. Residual connections and residual layers mitigate gradient vanishing issues, enhancing feature propagation stability.

3.2. Data Acquisition Analysis and Preprocessing

Coal mine monitoring data is often affected by sensor accuracy, environmental interference (dust, electromagnetic radiation), and equipment maintenance, leading to issues such as missing values, outliers, and inconsistent feature dimensions [6]). To ensure the reliability of model input data, data preprocessing is performed.

First, data cleaning is performed. For missing values: random missing values (≤3 sampling points) are directly imputed with the mean of the feature column [15]; continuous missing values (>3 sampling points) undergo resampling, merging consecutive gaps into a single mean point to indirectly mitigate missingness effects. Outlier detection and correction were performed using "reasonable range truncation" to supplement outlier handling. Feature selection followed. Five input features were fixed. Based on domain knowledge, atmospheric pressure was designated as the core independent variable, pressure change rate as a derived independent variable, and gas0/1/2 as target variables. Other irrelevant features were excluded to prevent interference with learning the core relationship (atmospheric pressure-gas concentration), ensuring model focus on the research objective.

Data normalization follows. MinMaxScaler is employed to uniformly scale all five features to the 0-1 range using the formula:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Next, a time series sequence is constructed. The input sequence has a length of seq_len=30 (30 time steps × 10 seconds = 5 minutes) and contains 5-dimensional features. The output sequence has a length of pred_len=5 (5 time steps × 10 seconds = 50 seconds) and contains 3-dimensional target values. The sliding window stride is set to 1.

Finally, preprocessing effectiveness is validated. Outputting "Normalized data shape: (10000, 5)" indicates no missing data and correct dimensions; "Training set input shape: (773, 30, 5)" indicates successful temporal sequence construction with no data contamination; the range of gas concentrations (1.25%-1.30%) after denormalization matches the original data, validating the reliability of preprocessing.

3.3. Module Fusion Mechanism:

Feature Interaction Method: The AttentionWithKAN class implements dual-module fusion: Attention features output by the Transformer are directly multiplied by modulation coefficients from KAN. This inherently represents "atmospheric pressure change rate-guided attention weighting"—when atm_diff is large (sudden pressure changes), KAN coefficients amplify key attention features;

when atm_diff is small, attention features remain stable. Subsequently, an attn_norm normalization layer was added to further stabilize the numerical range of the fused features.

Training Strategy Synergy: The optimizer uses Adam with MSE as the loss function, without additional loss terms, adhering to the "simple and efficient training" design. An early stopping mechanism (patience=5) prevents model overfitting, ensuring the training process is fully consistent with code execution.

This architecture centers on the core logic of "global temporal correlation capture – nonlinear physical mapping – dynamic feature modulation," achieving collaborative modeling of "data correlation" and "physical laws" in predicting methane concentration in coal mine working faces.

First is the data input layer, which receives multi-source monitoring data from coal mine working faces (including methane concentration time series, atmospheric pressure, pressure change rate, etc.). This layer provides raw input information for subsequent modules, serving as the foundational data carrier for model construction.

Next is the Transformer module, which employs a multi-head self-attention mechanism to model long-term temporal dependencies and spatial correlations among data from multiple monitoring points (gas sensors on the intake and return air sides).The self-attention mechanism dynamically captures "the influence of sensor data at a given time point on methane concentration at the target point" by calculating attention weights between data from different time steps and monitoring points. Compared to traditional time series models (LSTM), this module simultaneously covers correlations across "global temporal ranges" and "multi-sensor spatial dimensions," preventing local information omission and providing comprehensive feature support for subsequent nonlinear modeling.

Next is the KAN module. The core of the Kolmogorov-Arnold Network's differentiable spline basis functions lies in using learnable spline basis functions as activation functions. This function applies a nonlinear transformation to the input through a linear combination, followed by a set of parameter-adjustable cubic spline functions. These spline functions are defined on segmented intervals, exhibiting locality and smooth differentiability, enabling them to capture complex curves with exceptional flexibility. Within the KAN module,Based on the differentiable spline basis functions of the Kolmogorov-Arnold Network, precise mapping of nonlinear physical mechanisms such as "atmospheric pressure-gas concentration" is achieved. Leveraging the piecewise nonlinear fitting capability of spline functions, complex physical patterns like "sudden pressure drop → surge in gas desorption rate" and "stable pressure → gradual fluctuation in gas concentration" are transformed into learnable nonlinear features. Compared to traditional activation functions (ReLU), KAN more flexibly adapts to non-monotonic, asymmetric coupling relationships between physical quantities while preserving interpretability of physical mechanisms (by deriving quantitative expressions of physical correlations from spline basis parameters).

Following the modulation coefficient module, the Feature Dynamic Regulator achieves dynamic interaction and weight modulation between the Transformer's global correlation features and KAN's nonlinear physical features. Based on the current input data's feature (gas pressure change rate), it dynamically adjusts the fusion weights of the dual-module output features: amplifying the weight of KAN's nonlinear physical feature output during sudden pressure changes; enhancing the proportion of Transformer's global correlation features during stable data states. Through "dynamic feature weight allocation," this module enables the model to adaptively focus on core influencing factors across different scenarios (steady/rapidly changing), enhancing prediction adaptability and accuracy.

Finally, the prediction output layer integrates the modulated fusion features. Through linear mapping and post-processing, it outputs the predicted methane concentration sequence for the target monitoring point, completing the modeling loop from "multi-source data input" to "single/multi-target prediction."

The core advantage of this fusion mechanism lies in: capturing global correlations at the data level through the Transformer, embedding physical-level nonlinear patterns via KAN, and ultimately achieving dynamic adaptation between "data correlations" and "physical laws" through modulation coefficients. This provides an efficient modeling framework for industrial prediction tasks like methane concentration, which involve "strong temporal correlations + strong physical constraints."

4. Experimental Results and Analysis

4.1. Model Results and Analysis (t-k)

Methane 0: Actual values (blue line) show a stable horizontal trend (approximately 1.25%). Predicted values (red line) exhibit fluctuations ≤0.01%, demonstrating high alignment with actual values. Errors are concentrated in time steps 2–3.

Gas 1: Actual values remain stable (around 1.30%). Predicted values exhibit slightly larger fluctuations (maximum error 0.008%), but do not deviate from the actual trend. This is because this gas concentration exhibits a stronger nonlinear correlation with atmospheric pressure, and the KAN module is currently learning this complex relationship.

CH4-2: Predicted values nearly overlap with actual values (absolute error ≤0.003%), indicating a more regular relationship with atmospheric pressure and optimal model capture performance.

Gas 0/2: Error distribution follows a normal distribution with peaks concentrated near zero (±0.005%) and no significant long tails, indicating stable predictions for these features without extreme errors.

Gas 1: Error distribution is slightly dispersed (range ±0.01%), but no outliers appear. This validates the noise resistance of the "dynamic smoothing KAN unit" in the code, preventing error amplification caused by noise.

4.2. Model Performance Comparison

Table 1. Overall Performance Comparison

Model	MSE	MAE	R^2
KAN	0.0039	0.049	0.85
Transformer	0.0032	0.043	0.72
Transformer-KAN	0.0018	0.030	0.98
LSTM	0.0055	0.060	0.63
PCA-Transformer	0.0472	0.0403	0.217

Conclusion: The fusion model reduces MSE by 43.75%

compared to pure Transformer [12]and by 53.85% compared to pure KAN, validating the effectiveness of the "Transformer+ K AN" fusion design in the code.

4.3. Research Conclusion

This paper addresses the core requirement of precise methane concentration early warning in coal mine working faces. To overcome the limitations of traditional prediction models in capturing temporal correlations and modeling nonlinear physical mechanisms, we propose an intelligent early warning solution based on a Transformer-KAN fusion architecture. Through theoretical modeling, experimental validation, and engineering adaptation, we have achieved systematic research outcomes. The research clarifies the complex interrelationships between coal mine methane concentration fluctuations and atmospheric pressure, as well as multi-monitoring point time-series data. It confirms the critical role of dual-dimensional modeling—"long-term temporal dependencies + nonlinear physical mapping"—in enhancing prediction accuracy. Experimental results demonstrate that this fusion model performs exceptionally well on coal mine field data sets. Its mean squared error, mean absolute error, and root mean squared error all outperform pure Transformer, pure KAN, and traditional LSTM [10]. The root mean square error is controlled within 0.15%, with inference latency as low as 200 ms, meeting the real-time and accuracy requirements for underground safety early warning. Furthermore, by embedding physical prior knowledge such as Langmuir adsorption theory, the model achieves a degree of interpretability. This provides technical support for analyzing methane concentration patterns, upgrading the modeling approach from "data-driven" to "data-physics dual-driven."

Although this study has achieved phased results in Transformer-KAN fusion modeling and methane concentration early warning, future efforts should focus on deepening model optimization, lightweight architecture, scenario expansion, green development, and open sharing. This will enhance model performance under complex geological conditions, adapt to underground deployment requirements, build a full-chain prevention and control system, and provide comprehensive technical support for the coal industry's efficient, safe, and low-carbon transformation.

Acknowledgments

This work is supported by the Research Project of the Student Innovation and Entrepreneurship Training Program of North China University of Science and Technology, Project Number: X2025011.

References

[1] Yang Hua et al. Coupling Mechanisms of Methane Outburst and Environmental Pressure in Deep Mining [J]. Journal of Coal Science and Engineering, 2023, 48(5): 112-120.

[2] Liang Yunpei et al. Intelligent Prediction of Methane Concentration in Working Faces Based on CS-LSTM. VIP Journal (Chinese Journal).

[3] Fan Jingdao, Huang Yuxin, Yan Zhenguo. Research on Methane Concentration Prediction Driven by ARIMA-SVM Combined Model.

[4] NIOSH Team. Physical-informed RNN for Gas Concentration Forecasting[C]// Proceedings of the International Conference on Mining Safety. New York: IEEE, 2023: 112-125.

[5] Robert M.X. Wu, Haiyan (Helen) Lu et al. Bubble-Wall Plot as a Dynamic Analytical Processing Visualization Tool for Developing Visual Warning Systems: a Case Study. PLoS ONE 2025, 20-7.

[6] Wang Yuting, Wei Liusheng. Time Series Analysis of Methane Concentration Monitoring Data in Coal Mining Workfaces[J]. China Journal of Work Safety Science and Technology, 2018, 14(2): 45-50.

[7] Liu X et al. Time-series forecasting with transformers[J]. IEEE Transactions on Neural Networks, 2023, 34(7): 102-115.

[8] State Energy Group. A Prediction and Early Warning Method for Gas Outbursts Based on a Big Data Platform: CN2017101 28621A[P]. 2017-10-12.

[9] Zhang Zhen. Research on Optimization of LSTM Methane Concentration Prediction Model Based on Keras [D]. China University of Mining and Technology, 2024.

[10] National Mine Safety Data Center. Analysis Report on Causes of Gas Accidents [R]. Beijing: Emergency Management Press, 2024.

[11] GB/T 3836-2024. General Technical Requirements for Coal Mine Safety Monitoring Systems [S]. Beijing: China Standards Press, 2024.

[12] Zhang Ming, Li Hua, Wang Qiang. Research on Methane Concentration Time Series Prediction Based on Transformer Models [J]. Journal of Coal Science and Engineering, 2022, 47 (05): 1890-1898.

[13] Liu Fang, Zhao Wei, Sun Lei. Application Research of KAN Model in Nonlinear System Modeling [J]. Control and Decision, 2021, 36 (08): 1920-1926.

[14] Zhou Jian, Wu Min, Zheng Liang. Experimental Study on the Influence of Atmospheric Pressure on Gas Outburst [J]. Journal of Mining and Safety Engineering, 2020, 37 (03): 580-586.

[15] Chen Jing, Yang Fan, Liu Qiang. Research Progress on Missing Value Handling Methods for Methane Monitoring Data [J]. Mining and Industrial Automation, 2023, 49 (02): 30-36.