INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
Online ISSN 1841-9844, ISSN-L 1841-9836, Volume: 20, Issue: 4, Month: August, Year: 2025
Article Number: 6813, https://doi.org/10.15837/ijccc.2025.4.6813

CCC Publications

# A Time Series-based Data Modelling Approach for Natural Gas Pipeline Intelligent Management System

C. Mei, S. Lu, Z. Song, H. Li, Z. Feng, J. Liu

**Chunxiao Mei***
Hebei Gas Co., Ltd.
Shijiazhuang 050011, China
*Corresponding author: chunxiaomei0725@126.com

**Shengxin Lu**
Hebei Gas Co., Ltd.
Shijiazhuang 050011, China
lc_lc223@163.com

**Zhiyong Song**
Hebei Gas Co., Ltd.
Shijiazhuang 050011, China
luso_0987@163.com

**Hao Li**
Hebei Gas Co., Ltd.
Shijiazhuang 050011, China
yynanj_123@163.com

**Zhihao Feng**
Hebei Gas Co., Ltd.
Shijiazhuang 050011, China
luo_sm1@163.com

**Jiankun Liu**
Hebei Gas Co., Ltd.
Shijiazhuang 050011, China
liuhu_lh123@163.com

**Abstract**

The study proposes an integrated method for intelligent management and risk prediction of natural gas pipelines using time series analysis. The method combines k-nearest neighbor imputation for data preprocessing, long short-term memory networks, and an attention mechanism for temporal data prediction to introduce a new intelligent data management model. The research method is tested by testing real pipeline data from eight scenarios in the gas pipeline operation monitoring dataset and the gas pipeline failure dataset. The experimental results demonstrated a notable enhancement in performance relative to existing methods, with data coverage reaching up to 92%, a classification error rate of 5%, an accuracy in risk prediction of 95%, and a processing time of 12.7 seconds. The framework provides a comprehensive solution for improving the safety, efficiency, and reliability of natural gas pipeline operations through advanced data analytics.

**Keywords:** Time series analysis; Gas pipeline; Long short-term memory network; Attention mechanism; Data management.

## 1 Introduction

In light of the ongoing urbanization and the continued expansion of energy consumption, the secure and effective operation of gas pipelines, as a vital component of urban infrastructure, is of paramount importance. The conventional approach to gas pipeline management entails periodic inspections and manual monitoring, which enhance management efficiency to a degree. However, these methods remain constrained in their ability to process large-scale data, make nonlinear predictions, and address anomalies [1, 2]. Although existing research has made some progress in intelligent gas pipeline management, especially based on the application of Internet of Things and deep learning technologies, there are still significant research gaps. Existing methods are often computationally inefficient when dealing with large-scale and multidimensional gas pipeline data, and it is difficult to cope with the real-time processing demands of massive data [3, 4]. In addition, anomalous and missing data are inevitable in gas pipeline monitoring, and the effectiveness of existing methods in data preprocessing and filling is weak, resulting in insufficient reliability of the overall management system. These problems indicate that there is still a certain research gap in intelligent management of gas pipelines. To fill this gap, the study proposes a gas pipeline intelligent management and prediction model based on Time Series Analysis (TSA), Long Short-Term Memory (LSTM), and Attention Mechanism (AM). The model performs data preprocessing through the K-Nearest Neighbors Imputation (KNNI) algorithm and then combines LSTM and AM to process complex time-series data, which improves the accuracy of risk prediction and management efficiency. The research aims to provide a new intelligent solution in gas pipeline management, which helps to improve the safety of pipeline operation and management efficiency. The research contribution is the combination of KNNI, LSTM, and AM, which provides an efficient solution for natural gas pipeline data preprocessing, pattern recognition, and trend prediction. This study also promotes further development in prediction accuracy and resource utilization efficiency. The study is divided into four main parts. The first part reviews the related literature and background studies. The second part introduces the proposed Intelligent Data Management (IDM) and risk prediction model. The third part demonstrates the experimental design and result analysis. The fourth part discusses the conclusions of the study and future research directions.

## 2 Literature review

Once a gas pipeline leaks, ruptures, or malfunctions, it not only poses a threat to public safety but also causes significant economic losses and environmental pollution. Traditional gas pipeline management methods often rely on regular inspections and manual monitoring, which is not only inefficient but also carries the risk of overlooking hidden dangers. With the development of Internet of Things, big data, and Artificial Intelligence (AI) technologies, gas pipeline management systems based on intelligent means have gradually become a research hotspot. Wang S L et al. proposed an efficient coordinated supply chain management model based on a synergistic model of system dynamics and time series data analysis of supply chains. The experimental results showed that the model demonstrated excellent processing capabilities for both static and dynamic supply chains [5]. However, the cost and efficiency problems in large-scale applications are still not fundamentally solved. To improve the level of digital management of gas pipelines, Kim B C et al. proposed a new data analysis model after combining PID image recognition and deep neural network. Experimental results showed that the model's ability to digitally reshape gas pipeline data reached 96.64% and the average accuracy of topology reconstruction was 96.40% [6]. However, the method relied too much on specific gas pipeline topology reconstruction and lacked sufficient generalization. Odili P O et al., after combining Internet technology and AI algorithm, constructed an intelligent management and prediction method for gas pipeline that can reduce the risk of human operation. The sensitivity of this method to corrosion, temperature, and pressure data of pipelines was the highest at 97.3% [7]. To further strengthen the management of gas pipeline safety, Khan U et al. proposed a new pipeline data sensing method after combining dynamic pressure sensing and deep learning techniques. Experimental results showed that the method perceived gas pipeline data with up to 89.6% accuracy and had superior data processing capability [8]. Ma T et al. concluded that the existing gas pipeline bending strain lacks an intelligent and efficient method for recognizing pipeline features. For this reason, the researchers proposed a natural gas pipeline feature recognition and management method based on shape and hybrid fusion models. Experimental results showed that the method was trained from bending strain data obtained from real natural gas pipelines, and its recognition accuracy was 97.17%. Although it performed well in specific tasks, its ability to adapt to different pipeline features and environments was still limited [9]. To improve the planning and intelligent management of urban gas pipeline networks, after combining with a 3D Geographic Information System (GIS), Huang Y et al. proposed a new type of underground pipe network management model. The experimental results showed that the model could effectively assess the compliance, feasibility, and scientificity of the planning scheme of underground pipeline construction projects. However, it could not adequately cope with the dynamically changing pipeline status [10].

TSA is a statistical technique used to analyze data points arranged in chronological order to identify patterns, trends, periodicity, or outlier. Kumar R et al. found that the existing TSA still has certain drawbacks in financial forecasting analysis of the stock market. Therefore, they constructed a new type of stock market time series data prediction by combining differential evolution method, which significantly improved the selection effectiveness of relevant data [11]. However, it also showed that the existing TSA models still had efficiency problems when dealing with complex and variable data. Hosseini S A et al. used digital aerial images combined with multilayer perceptrons and artificial neural network algorithms to analyze time-series data in order to understand the nitrogen status of crops. The results showed that the average error in estimating nitrogen levels using this method was only 0.145, which is much lower than other methods [12]. Sadeghi J et al. developed an assessment model incorporating fuzzy algorithms to identify and prioritize significant seismic risks involved. Experimental results showed that this method demonstrated good predictive performance and robustness in multiple address risk identification tests [13]. This new method improved the prediction accuracy of some parameters, but still had bias in dealing with the interaction of multiple factors. However, its applicability to long-term and complex pipeline systems was doubtful. Ahmed S et al. optimized the Transformer-based architecture language processing, introduced TSA for a series of extensions, and finally proposed a novel time series Transformer architecture. Experimental results showed that this new architecture could effectively overcome the challenges of training Transformers for TSA [14]. To explore the relationship between tourism and economic growth in India in depth, Singh D et al. proposed a novel feature relationship mining model after combining sector-specific data on macroeconomic variables and TSA techniques. The experimental results showed that the model was able to deeply explore the relationship between tourism and economic growth in India and the data showed non-cyclical and expansionary growth [15].

In summary, existing research has made significant progress in intelligent management of gas pipelines, but there are still difficulties in dealing with complex nonlinear problems. In addition, although the TSA method performs well in predicting the future status of gas pipelines, there are still challenges in dealing with abnormal data and long-term series predictions. Therefore, this study innovatively proposes a new management prediction model that combines KNNI algorithm, LSTM, and AM. A new technical support is provided for improving the accuracy of Gas Pipeline Data Management (GPDM) through real-time collection of pipeline data, data preprocessing, pattern recognition, and trend analysis.

## 3 Research methodology

To construct a complete gas pipeline IDM and prediction model, this study first collects, stores, and transmits real-time gas data. Normalization is performed using the KNNI algorithm in TSA and a novel IDM model for gas pipelines is proposed. Secondly, the risk of GPDM is divided and LSTM is introduced for improvement. After combining with the new gas pipeline IDM model, a new GPDM risk prediction model is proposed.

### 3.1 Time series data management model for gas pipeline

Gas pipelines are an important component of urban infrastructure, responsible for transporting gas from production or storage sites to users' homes. Real time collection, analysis, and monitoring of its operational data are key to achieving precise control and early warning in the later stage. General operational data include temperature, humidity, material, length, pressure, flow rate, and leakage of pipelines [16, 17, 18]. Therefore, this study takes the pipeline construction project of a certain gas group in a certain city as an example, and selects multiple gas pipelines for the collection of irregular operation data. According to the classification of dynamic and static data, the running data is shown in Figure 1.

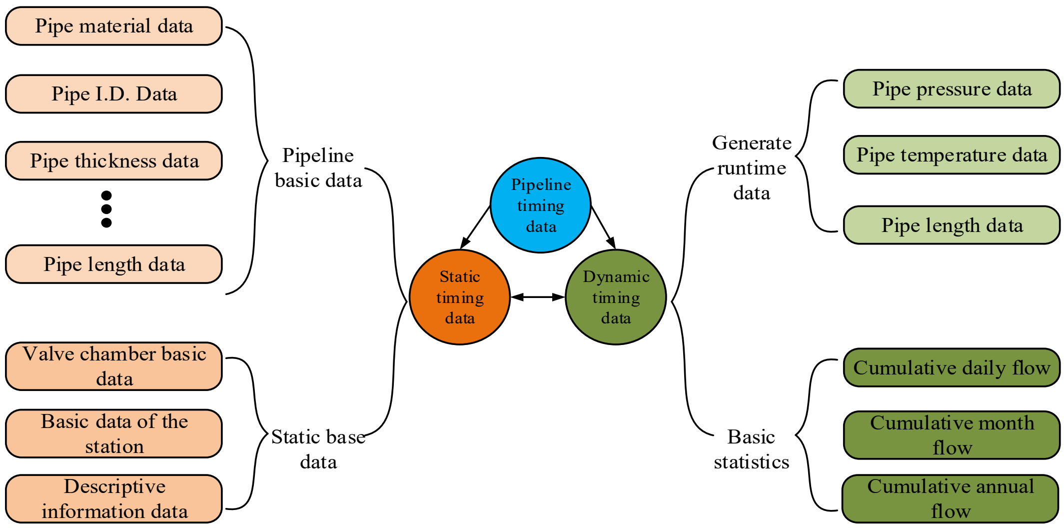


Figure 1: Gas pipeline operation data classification

In Figure 1, the static data architecture consists of two main data entities, namely pipeline basic data and static basic data. The former includes basic information about pipelines, such as pipeline laying, inner diameter, length information, etc. [19, 20, 21]. The latter covers relevant network node data, station basic data, and basic information, such as station basic data. The pipeline operation data are all time-series data, that is, real-time data collected by sensors [22, 23, 24]. Therefore, this study adopts a time-series database for data storage and introduces KNNI for filling in abnormal missing data. The filling process is divided into defining distance metrics and calculating interpolation. The reason for choosing KNNI is that it can effectively fill missing values with information from neighboring data points while maintaining the overall trend and local features of the data. Compared with other interpolation methods, such as linear interpolation or spline interpolation, KNNI performs more consistently and flexibly in dealing with high-dimensional and multivariate gas pipeline time series data [25, 26, 27]. In terms of implementation details, KNNI first determines the K-nearest neighboring observations by calculating the Euclidean distance between each missing data point and other data points. Then, the missing values are replaced based on the weighted average of these neighboring data points [28, 29]. The calculation formula for defining the distance metric is shown in equation (1).

$$d(x_i, x_j) = \sqrt{\sum\nolimits_{k=1}^{m} (x_{i,k} - x_{j,k})^2} \tag{1}$$

In equation (1), $x_i$ and $x_j$ are the data at the $i$ and $j$ time points in the gas pipeline dataset. $d(x_i, x_j)$ is the Euclidean distance between two data points. $k$ is the variable index. For each data point $x_i$ containing missing values, after calculating its Euclidean distance from all data points, the nearest $k$ components are selected to form the set. Interpolation is performed using the values of $K$ neighboring variables, as shown in equation (2).

$$x_{i,p} = \frac{1}{K} \sum\nolimits_{j \in N_i} x_{j,p} \tag{2}$$

In equation (2), $K$ is the set of adjacent data points. $p$ is the variable index of missing value time-series data. $N_i$ is the set of $K$ data points closest to data point $x_i$. The calculation after interpolation and averaging is shown in equation (3).

$$\begin{cases} x_{i,p} = \dfrac{\sum_{j \in N_i} \omega_{ij} \cdot x_{j,p}}{\sum_{j \in N_i} \omega_{ij}} \\ \omega_{ij} = \frac{1}{d(x_i, x_j)} \end{cases} \tag{3}$$

In equation (3), $\omega_{ij}$ is the interpolation weight. By using variable $x_{i.p}$ with missing values, the average or weighted average of its $K$ neighbors is calculated to fill in the missing values. In addition, abnormal pipeline operation timing data are filtered, and abnormal data appear due to the influence of the private network status, as shown in Table 1.

Table 1: Gas pipeline operation anomaly data display

| Time | Pipe temperature/°C | Pipe flow/m$^3$/s | Pipe pressure/Mpa |
|---|---|---|---|
| 2022/7/4 5:13 | 23.11 | 3.44 | 2.12 |
| 2022/7/6 6:44 | 20.12 | 3.69 | 2.13 |
| 2022/7/11 8:08 | 20.53 | 3.58 | 3.54 |
| 2022/7/13 11:28 | 23.54 | 3.17 | 2.16 |
| 2022/7/17 13:46 | 22.77 | 3.22 | 2.22 |
| 2022/7/17 19:01 | 28.18 | 3.19 | 2.14 |
| 2022/7/23 10:55 | 22.33 | 3.27 | 2.16 |
| 2022/7/25 12:22 | 20.46 | 3.52 | 2.20 |
| 2022/7/28 16:34 | 21.29 | 3.26 | 2.18 |
| 2022/7/30 20:17 | 22.56 | 3.18 | 2.19 |

In Table 1, the pipeline temperature data at 19:01 on July 17, 2022 are much higher than the general monitoring values. The pipeline pressure value at 8:08 on July 11, 2022 is much higher than the general detection value. This indicates that due to the influence of the operating network and environment, there are still inevitable data anomalies and losses in the management of gas pipeline timing data. Therefore, this study normalizes the time-series data [30, 31, 32]. The normalization calculation formula is shown in equation (4).

$$X_n = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{4}$$

In equation (4), $X$ is the time-series data sample of the gas pipeline. $X_{\max}$ and $X_{\min}$ are the maximum and minimum values of the gas pipeline operation timing data. $X_n$ is the normalized data target matrix. The normalized data is subjected to inverse normalization processing, and the calculation at this time is shown in equation (5).

$$Y = X \times (X_{\max} - X_{\min}) + X_{\min} \tag{5}$$

In equation (5), $Y$ is the time-series data of pipeline operation after inverse normalization. Based on the above data classification, data filling, data anomaly handling, and data normalization, this study proposes an IDM model for gas pipelines. The process of this model is shown in Figure 2.

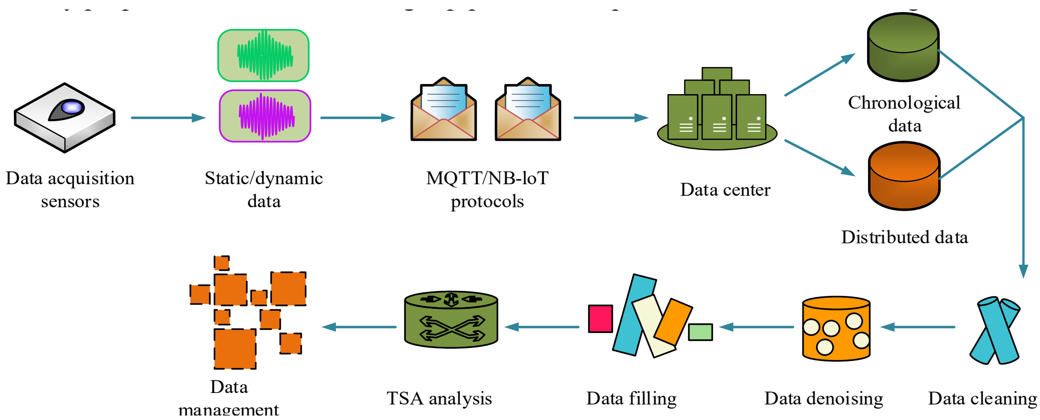


Figure 2: IDM model for gas pipelines

In Figure 2, firstly, various sensors collect real-time static and dynamic data of gas pipelines, and transmit the data to the data center through MQTT protocol and NB-IoT wireless communication network. Subsequently, the data are stored in both time-series and distributed databases. Next, the data are cleaned, denoised, and missing-value filled, and KNNI is used for data preprocessing. Then, TSA technology is utilized to perform pattern, trend, periodicity, and outlier analysis on the data, and based on the analysis results, a predictive model is constructed to warn potential problems in advance.

### 3.2 Intelligent management risk prediction model for gas pipeline

After completing the construction of the IDM model for gas pipelines, this study focuses on risk prediction in their management process. This study analyzes and screens specific risk factors in multiple dimensions to eliminate redundant and low correlation indicators. The risk indicator system at this time is shown in Figure 3.

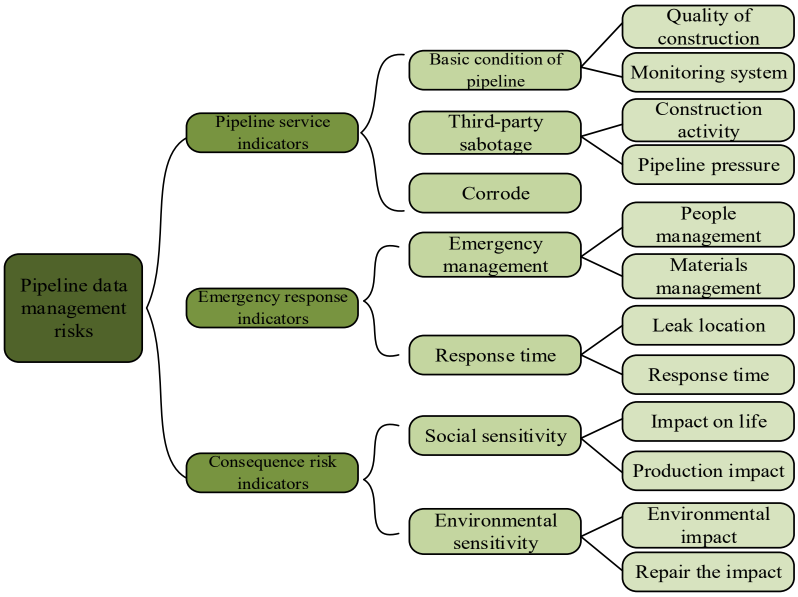


Figure 3: Schematic risk indicators for intelligent management of gas pipelines

In Figure 3, the pipeline service indicators include basic pipeline conditions, third-party damage, and corrosion. The emergency response indicators include emergency management and emergency time. The consequence risk indicators include social sensitivity and environmental sensitivity. To improve the accuracy and real-time performance of multi-dimensional risk prediction, this study introduces LSTM. The reason for introducing LSTM is that there are complex temporal dependencies in natural gas pipeline operation data, such as the periodic impact of temperature changes on pipeline pressure. LSTM can capture this complex temporal dependency [33, 34, 35]. Compared with other traditional TSA models, LSTM has a stronger ability to learn nonlinear features [36, 37]. In terms of implementation details, the study employs a two-layer LSTM structure consisting of 128 units per layer. ReLU is used for the activation function, Adam is chosen for the optimizer, and the learning rate is 0.001. Figure 4 shows the structure of LSTM.

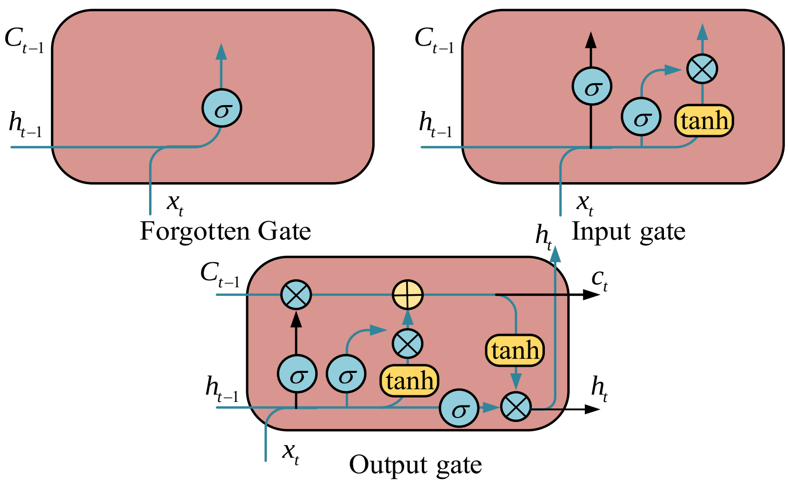


Figure 4: LSTM structure

In Figure 4, the entire LSTM consists of three gate control structures, namely the forget gate, input gate, and output gate. The forget gate determines which information in the memory state of the previous time step needs to be forgotten. The input gate determines how the input of the current time step should be integrated into the memory state. The output gate determines what the output of the current time step is. The core working mechanism of LSTM oblivious gate is shown in equation (6).

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \tag{6}$$

In equation (6), $\sigma$ is the activation function. $f_t$ is the activation value of the forget gate at time $t$. $x_t$ represents the input value for $t$. $h_{t-1}$ is the output when $t-1$. $b_f$ is the bias term. $W_f$ and $U_f$ are the weight values of the parameters. Oblivion gates selectively forget certain information by determining what historical information is no longer important. After determining the update information, the tanh layer creates a candidate value $C_t$ at $t$, which is the memory unit. At this point, the input gate controls how the information of the current time step is updated to the memory cell. Its calculation is divided into two parts, as shown in equation (7).

$$C_t = f_t * C_{t-1} + i_t * \tanh(W_c x_t + U_c h_{t-1} + b_c) \tag{7}$$

In equation (7), $i_t$ is the activation value calculated by the input gate sigmoid layer. $C_{t-1}$ is the memory unit of the previous moment. $W_c$ and $b_c$ are the weights and biases of the forget gate. $U_c$ is the parameter of the forget gate. After calculating the outputs of the forgetting gate and the input gate, the LSTM updates the state of the memory cell with the update formula shown in equation (8).

$$h_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) * \tanh(C_t) \tag{8}$$

In equation (8), $W_i$ and $b_i$ are the weights and biases of the output gate. $U_i$ is the parameter of the output gate [28]. To address the problem of slow training speed and overfitting caused by multiple parameters in LSTM for long sequence data, this study introduces AM. The justification for choosing AM is that it can highlight key features in long time series data and make up for the deficiency of LSTM in long sequence prediction. With the self-attention module, the model can automatically adjust the attention to different time points, which improves the model's ability to handle critical data [38]. Figure 5 shows the results of AM-LSTM.

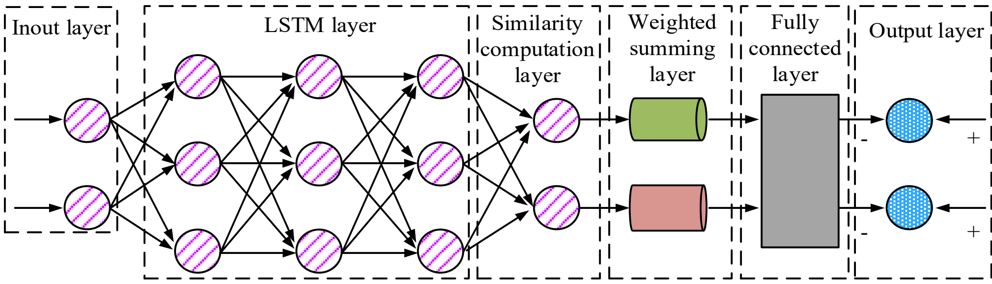


Figure 5: Schematic diagram of AM-LSTM model structure

In Figure 5, the input data are processed by the LSTM layer to generate a hidden state sequence, and then the attention score matrix is obtained through similarity calculation. Subsequently, the self-attention layer normalizes the score matrix, generates attention weights, multiplies them with hidden states, and weights them to obtain a weighted context similarity vector. Finally, the final result is output through the fully connected layer. The expression of self-attention weight is shown in equation (9).

$$\alpha_t = \frac{\exp(e_t)}{\sum_{t'} \exp(e_{t'})} \tag{9}$$

In equation (9), $e_t$ is the attention score at $t$. The contextual similarity is shown in equation (10).

$$c_t = \sum\nolimits_{t'} \alpha_{t'} h_{t'} \tag{10}$$

In equation (10), $h_{t'}$ is the hidden state at time $t'$. $c_t$ is the contextual similarity at $t$. At this point, by assigning different weights to input data at different time steps, important features can be effectively extracted and utilized. The output formula of AM-LSTM is shown in equation (11).

$$h_t^* = o_t \cdot \tanh(C_t) + c_t \tag{11}$$

In equation (11), $o_t$ is the activation value of the output gate. $C_t$ is the state of the memory unit. At this point, the risk prediction of the gas pipeline is shown in equation (12).

$$\hat{y}_t = \sigma(W_i \cdot h_t + b_i) \tag{12}$$

In equation (12), $\hat{y}_t$ is the predicted value of $t$. The GPDM and risk prediction process of gas pipeline IDM model combined with AM-LSTM can be roughly divided into five stages, as shown in Figure 6.

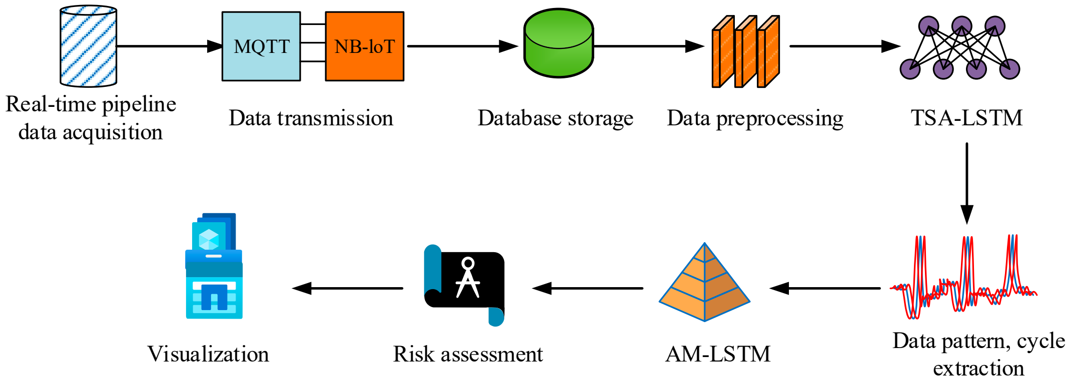


Figure 6: Gas pipeline data management and risk prediction process

From Figure 6, Step 1 is to collect real-time data on pipeline pressure, temperature, flow rate, etc. through sensors, and transmit it to the data center through MQTT protocol and NB-IoT network. Step 2, the data are stored in both time-series and distributed databases, and undergo cleaning, denoising, and missing value imputation. Step 3 is to use TSA and LSTM models to extract patterns, trends, and periodicity of data, and introduce AM to further enhance the model's memory ability and selectivity. Step 4 is to use the AM-LSTM model for risk prediction, generate a risk assessment report, and provide early warning of potential risks. Step 5 is to display the analysis results through a visualization platform, provide decision support, and improve the efficiency and safety of pipeline management. The final model consists of KNNI, LSTM, and AM. KNNI is used to fill in the missing values in the gas pipeline data, LSTM is responsible for dealing with the long-term dependencies in the time-series data, while AM further enhances the model's ability to extract key features in long series prediction. The choice of hyperparameters includes a two-layer structure of LSTM with 128 units per layer using Adam optimizer with a learning rate of 0.001 and ReLU activation function. The model is trained using Mean Square Error (MSE) as the loss function and an early stopping strategy to avoid overfitting. Subsequently, the performance is evaluated by several metrics, including accuracy, recall, and F1 value. In addition, grid search and cross-validation are employed for K values in the range of [2, 10]. Meanwhile, the effects of two-layer, four-layer, six-layer, and eight-layer LSTM structures on the model performance are explored over cross-experiments.

# 4 Results and discussion

To verify the performance of GPDM and prediction models, a suitable experimental environment is set up in this study. The optimal $K$ value and the number of layers in AM-LSTM are verified through testing, and the effectiveness of the final model is validated through ablation testing. A similar model is introduced for comparison. In addition, in real pipeline scenarios, this study compares several advanced methods through multiple indicators to verify the effectiveness of the research methods.

### 4.1 Performance testing of IDM prediction model for gas pipeline

The CPU adopts Intel Core i7, the GPU is NVIDIA GeForce RTX 2070, the memory is 32GB, and the development environment is Python. The activation function is set to ReLU, optimizer to Adam, and learning rate to 0.001. The Gas Pipeline Operation Monitoring Dataset (GPOMD) and Gas Pipeline Fault Dataset (GPFD) are used as data sources. GPOMD includes key parameters such as pipeline pressure, temperature, flow rate, and humidity. GPFD includes information on fault types such as leakage, rupture, corrosion, location and time of occurrence, severity of faults, and corresponding maintenance measures. The key features of the dataset are shown in Table 2.

Table 2: Key features of the dataset

| Dataset name | Dataset type | Data features | Sample size | Time span | Key parameters | Anomaly sample ratio | Sampling frequency | Data source | Missing data ratio |
|---|---|---|---|---|---|---|---|---|---|
| GPOMD (Gas Pipeline Operation Monitoring Dataset) | Real-time monitoring data | Time-series data including pressure, temperature, flow rate | 50,000 | 12 months | Pipeline pressure, temperature, humidity, flow rate | 5% | Hourly | Multiple pipeline sensor networks | 2% |
| GPFD (Gas Pipeline Fault Dataset) | Historical fault data | Fault type, fault occurrence time, location | 8,000 | 5 years | Fault type (leakage, rupture, etc.), severity | 3% | Fault occurrence | Gas company maintenance records | 1% |

Firstly, this study attempts to determine the $K$ value of the KNNI algorithm and the number of layers in AM-LSTM, with the coefficient of determination $R^2$ as the metric. $R^2$ is the proportion of explanatory variables in the model, and the closer the value is to 1, the better the fitting effect of the model. The test results are shown in Figure 7.

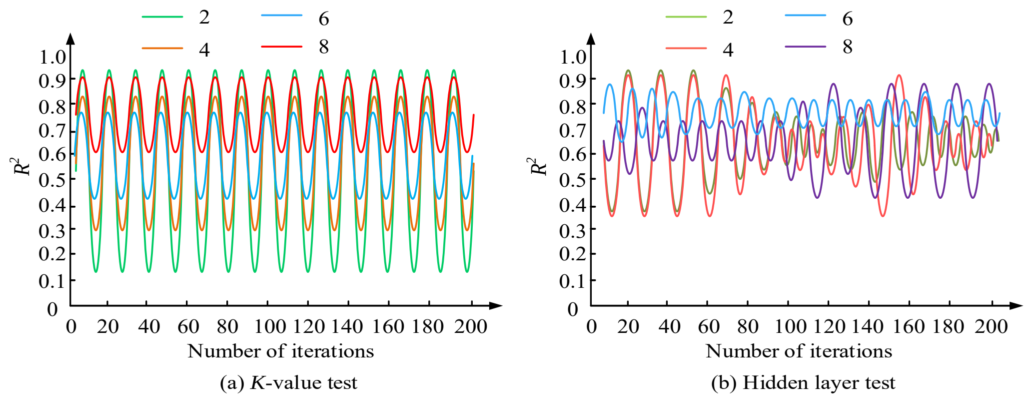


Figure 7: Test results for different $K$ values and number of hidden layers

Figure 7 (a) shows the performance of the GPDM model under different $K$ values and hidden layers. Figure 7 (b) shows the performance effect of the GPDM model with different number of hidden layers. With the increase of K value, the performance of the model shows a tendency of increasing and then decreasing. The coefficient of determination of the model reaches the best when the value of $K$ is 8, at which time $R^2 \approx 0.80$. When the value of $K$ is less than 8, the lack of neighboring data points leads to lower interpolation accuracy. If the $K$ value is too high, it will introduce more irrelevant data, leading to an increase in interpolation error. Therefore, a $K$ value of 8 is the optimal choice to fill in the missing data, which can achieve a balance between data integrity and interpolation accuracy. In addition, when the number of hidden layers of AM-LSTM is 6, the mean value of $R^2$ at time is 0.8, and the fitting effect is the most stable. 2-layer LSTM is difficult to capture complex temporal dependencies when dealing with long sequential data, which leads to a decrease in the performance of the model. Although 8-layer LSTM has more feature extraction capabilities, its complexity brings the risk of overfitting, thereby increasing computational overhead. This study conducts ablation testing using data coverage as the indicator, and the results are shown in Figure 8.

Figures 8 (a) and (b) show the test results of the research model on the GPOMD and GPFD

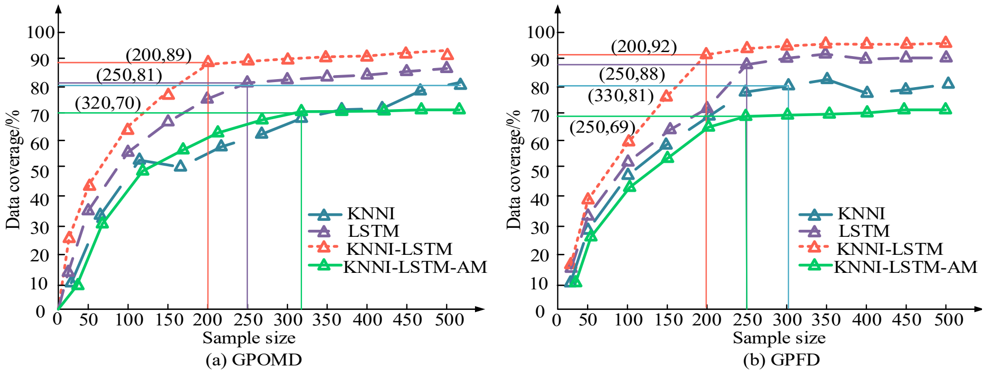


Figure 8: Ablation test results

datasets. The maximum improvement in performance is 11% in the GPOMD dataset and 12% in the GPFD dataset. After combining AM, the KNNI-LATM-AM model in the GPOMD dataset has a maximum coverage of 89% for pipeline data, and the KNNI-LATM-AM model in the GPFD dataset has a maximum coverage of 92% for pipeline data. After combining KNNI with LSTM, the data coverage and prediction accuracy are significantly improved, because KNNI effectively fills in the missing data and ensures the completeness of the inputs to the LSTM model. When AM is added, the model performance is further improved. Especially in the processing of long sequence data, AM can help the model to focus on the key time steps, thus improving the prediction accuracy. As a result, the complete model combining KNNI, LSTM, and AM performs best in the ablation experiments, verifying the effectiveness of the combination of these modules. This study introduces similar data management models for comparison, such as Support Vector Regression (SVR), Random Forest Regression (RFR), and Transformer models. The results of testing based on data classification error are shown in Figure 9.

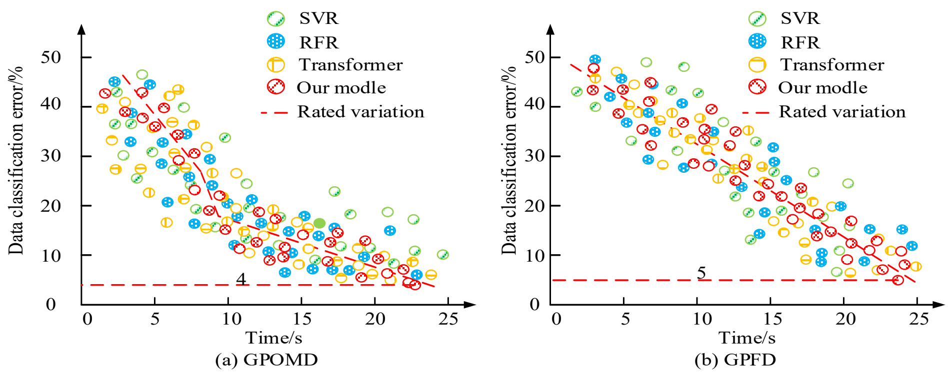


Figure 9: Comparison results of data classification errors for different models

Figures 9 (a) and 9 (b) compare the classification errors of four models in GPOMD and GPFD. With the extension of testing time, the performance improvement of data classification errors for SVR and RFR is relatively small, with the lowest errors being 8% and 7%, respectively. The data classification error of Transformer is relatively small, with a minimum classification error of 5%. Its poor fit with the rated classification line indicates that although the model can achieve lower data classification errors, and there is still room for improvement compared to the classification error values at the rated time. The data classification error of the research method is close to the rated value, and its lowest data classification errors under GPOMD and GPFD are 4% and 5%, respectively, which once again verifies the strong effectiveness and robustness of the method.

### 4.2 Simulation testing of IDM prediction model for gas pipeline

This study takes 8 pipeline projects, including Scenario 1 (Chengdu central urban pipeline), Scenario 2 (suburban industrial pipeline), Scenario 3 (mountainous pipeline), and Scenario 4 (pipeline near rivers), as the testing background. The data collection time is set to 6 months, with measurements taken every hour. The obtained data are centrally processed through operations such as data cleaning, outlier screening, denoising, and normalization. After processing, similar methods are introduced to compare the accuracy of data risk prediction in 8 scenarios, such as DBN, ELM, and GRNN models. Figure 10 shows the comparison results.

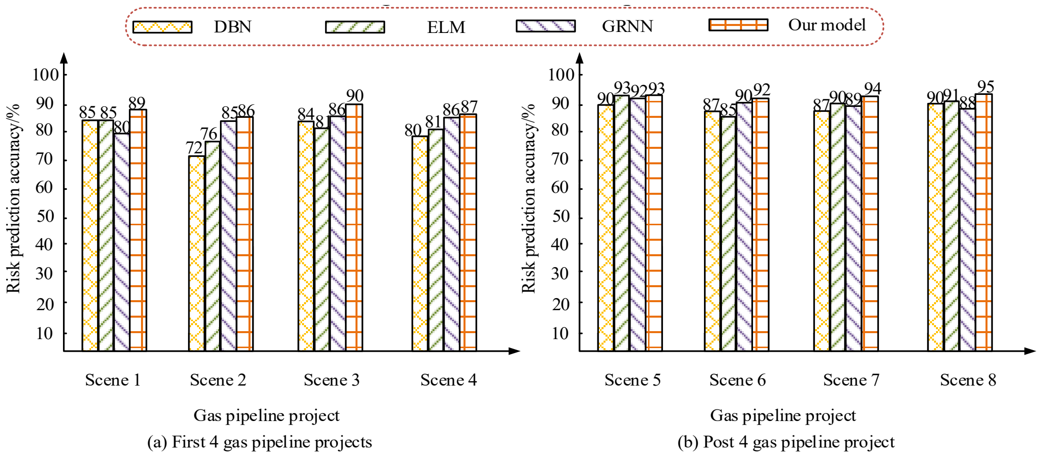


Figure 10: Comparative results of modeled risk prediction in multiple scenarios

Figure 10 (a) and Figure 10 (b) show a comparison of risk prediction for the first four and last four gas pipeline scenarios. In Figure 10 (a), the research models all show high accuracy, at 89%, 76%, 90%, and 87%. The prediction accuracy of DBN in these scenarios is 85%, 72%, 84%, and 80%, slightly lower than the research model. In Figure 10 (b), the research model still maintains high prediction accuracy, with the highest values being 95%, 90%, 94%, and 95%. Although DBN, ELM, and GRNN models have shown some competitiveness in certain scenarios, the overall research models demonstrate higher prediction accuracy in 8 scenarios. The superior performance of the model proposed in the study is mainly due to its temporal data processing capability and the introduction of the AM. LSTM can effectively learn the temporal patterns of gas pipelines and AM helps the model to focus on the key data points in diverse scenarios, which improves the model's generalization capability and adaptability. This study continues to test with computation time as the indicator, as shown in Figure 11.

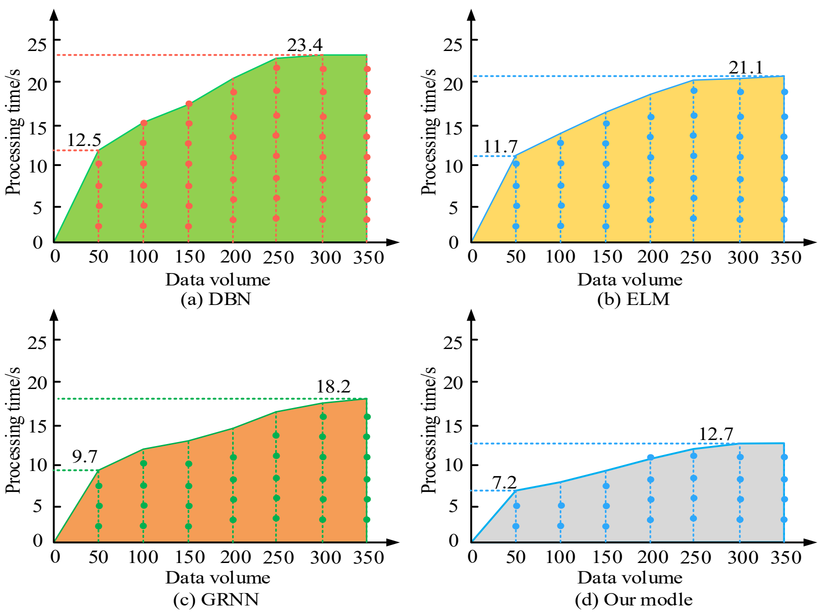


Figure 11: Runtime comparison results for different models

Figures 11 (a) to (d) show a comparison of the running time of DBN, ELM, GRNN, and the research model. Overall, the processing time of all models increases with the increase of data volume, but the magnitude and speed of growth vary. The processing time of DBN and ELM is relatively long, especially when the data volume is large, with processing times reaching 23.4s and 21.1s, respectively. The processing time of GRNN is slightly better, but it still takes 18.2s when the data volume reaches its maximum. In contrast, the research model shows a significant advantage in processing time, with a processing time of only 12.7s even at maximum data volume. However, although the present model performs better in terms of efficiency, there are still some potential limitations. First, the complexity of the LSTM layer and the introduction of the AM increase the computational burden, and the computational efficiency of the model may be affected as the amount of data increases further. For this reason, the introduction of model compression techniques, distributed computing, and parallel processing can be considered in the future to ensure the execution efficiency of the system under large-scale pipeline networks. Especially when it comes to large-scale pipeline networks spanning hundreds of thousands of kilometers, data partitioning and node level parallel computing will be key to improving system efficiency. Finally, this study tests pipeline temperature, humidity, and pressure using precision (P), recall (R), and F1 values as indicators. Table 3 shows the specific results.

Table 3: Multi-indicator test results

| Targets | Model | P/% | R/% | F1/% | Resource consumption rate/% | $P$-value |
|---|---|---|---|---|---|---|
| Pipe temperature | DBN | 91.33 | 90.27 | 90.80 | 37.66 | 0.032 |
| | ELM | 92.47 | 91.36 | 91.92 | 34.18 | 0.041 |
| | GRNN | 93.58 | 89.64 | 91.61 | 24.96 | 0.027 |
| | Research model | 95.71 | 93.21 | 94.46 | 18.94 | 0.006 |
| Pipe humidity | DBN | 89.61 | 88.74 | 89.18 | 34.12 | 0.037 |
| | ELM | 88.54 | 89.63 | 89.09 | 26.58 | 0.044 |
| | GRNN | 90.67 | 90.85 | 90.76 | 21.26 | 0.027 |
| | Research model | 93.28 | 92.76 | 93.02 | 19.34 | 0.003 |
| Pipeline pressure | DBN | 89.54 | 88.11 | 88.83 | 35.47 | 0.029 |
| | ELM | 89.69 | 88.96 | 89.33 | 28.69 | 0.035 |
| | GRNN | 91.23 | 91.27 | 91.25 | 19.67 | 0.019 |
| | Research model | 94.58 | 93.67 | 94.13 | 16.33 | 0.005 |

In Table 3, among the test results of multiple indicators, the research model outperforms other models in predicting pipeline temperature, humidity, and pressure. In pipeline temperature prediction, the P, R, and F1 values of the research model are 93.21%, 91.44%, and 92.31%, respectively, with a resource consumption rate of 18.94%, significantly lower than other models. In pipeline humidity prediction, the new model has P, R, and F1 values of 90.68%, 91.02%, and 90.85%, and a resource consumption rate of 19.34%, demonstrating higher prediction accuracy and lower resource consumption. In contrast, although DBN, ELM, and GRNN models perform well in certain indicators, they are generally inferior to research models, especially in terms of resource consumption, which is significantly higher. The above results once again validate the effectiveness and superiority of the research model in GPDM and risk prediction. To further verify the statistical significance of the model performance, the study introduces the $P$-value as a significance test index. The results of the $P$-value show that the differences between all the compared models and the proposed model are statistically significant ($P<0.05$), thereby substantiating the superiority of the proposed model in the prediction of gas pipeline data. Especially in the prediction of pipeline pressure, the F1 value of research model is 94.13%, which is significantly higher than the 91.25% of the GRNN model, and its corresponding $P$-value is 0.019, which indicates that the difference between the models possesses significance. In addition, it is necessary to consider the security of data during transmission and storage. For this purpose, encryption technology and distributed data storage schemes, such as blockchain technology, can be introduced to ensure that data privacy is not violated.

## 5 Conclusion

A gas pipeline intelligent management and prediction model based on the combination of KNNI, LSTM, and AM was proposed in the study. Experimental results demonstrated that the model exhibited favorable performance in terms of coverage, classification error, and processing time for gas pipeline data, with a prediction accuracy of 95% and a minimum processing time of 12.7 seconds.

These findings validated its efficacy in practical applications. There are three main contributions of the study. Firstly, data integrity is enhanced by KNNI. Secondly, LSTM handles complex temporal dependencies. Thirdly, the AM improves the focus on key temporal information, which results in higher prediction accuracy and efficiency. This model can be applied to pipeline systems in other industries, such as oil and water supply, providing technical support to improve risk prediction. The widespread impact on pipeline management is that this model provides more efficient and intelligent management tools for natural gas companies and related industrial systems. This model can provide early warning of potential faults, reduce accident risks, optimize resource utilization, reduce maintenance costs, and improve management efficiency. Intelligent management systems can significantly improve traditional manual monitoring methods, especially in large-scale and complex environments. While the model shows promise for improving pipeline safety and efficiency, future work should focus on optimizing performance for more complex environments and longer time series. Moreover, further validation of the model's applicability is required in a greater number of industrial scenarios. Extending its application to other industrial pipeline systems, such as the fusion processing of diverse sensor data, will also represent a significant area of future research.

# References

[1] Orkomi A A, Bora K M. Traffic Induced Noise Pollution Prediction Using Traffic Noise Model (Tnm) In Rasht, Iran. Environmental Engineering & Management Journal (EEMJ), 2021, 20(9): 1415-1423.

[2] Tang M, Cao J, Fan Z, Zhang D, Pandelic Ionut. A Data-Driven Assessment Model for Metaverse Maturity. International Journal Of Computers Communications & Control, 2024, 19(4): 288-300.

[3] Raska P, Ulrych Z, Baloun J. Using Adaptive Neural Networks For Optimising Discrete Event Simulation. International Journal of Simulation Modelling (IJSIMM), 2024, 23(2): 227-238.

[4] Fang X, Wang Y. Modelling And Simulation Of Purchasing Decision By Data-Driven Demand. International Journal of Simulation Modelling (IJSIMM), 2024, 23(2): 371-381.

[5] Wang S L, Zhang X. Optimization Strategies And Simulation Of Integrated Management In Supply Chains. International Journal of Simulation Modelling (IJSIMM), 2024, 23(3): 543-554.

[6] Kim B C, Kim H, Moon Y, Lee G, Wun, D. End-to-end digitization of image format piping and instrumentation diagrams at an industrially applicable level. Journal of Computational Design and Engineering, 2022, 9(4): 1298-1326.

[7] Odili P O, Daudu C D, Adefemi A, Ekemezie I O , Usiagu G S. Integrating advanced technologies in corrosion and inspection management for oil and gas operations. Engineering Science & Technology Journal, 2024, 5(2): 597-611.

[8] Khan U, Pao W, Sallih N, Hassan F. Flow Regime Identification in Gas-Liquid Two-Phase Flow in Horizontal Pipe by Deep Learning. Journal of Advanced Research in Applied Sciences and Engineering Technology, 2022, 27(1): 86-91.

[9] Ma T, Hu C, Wang L, Ma L, Mao Z, Xuan H. An intelligent feature recognition method of natural gas pipelines based on shapelet and blending fusion model. Measurement Science and Technology, 2023, 35(1): 15004-15009.

[10] Huang Y, Peng H, Wen L, Xing T. Using digital technologies to plan and manage the pipelines network in city. IET Smart Cities, 2023, 5(2): 95-110.

[11] Kumar R, Kumar P, Kumar Y. Multi-step time series analysis and forecasting strategy using ARIMA and evolutionary algorithms. International Journal of Information Technology, 2022, 14(1): 359-373.

[12] Hosseini S A, Masoudi H, Sajadiye S M, mEhdizadeh S A. Nitrogen estimation in sugarcane fields from aerial digital images using artificial neural network. Environmental Engineering & Management Journal (EEMJ), 2021, 20(5): 713-723.

[13] Sadeghi J, Oghabi M, Sarvari H, Sabeti MS, Kashefi H, Chan D. Identification and prioritization of seismic risks in urban worn-out textures using fuzzy delphi method. Environmental engineering and management journal, 2021, 20(6): 1035-1046.

[14] Ahmed S, Nielsen I E, Tripathi A, Siddiqui S, Ramachandran R, Rasool G. Transformers in time-series analysis: A tutorial. Circuits, Systems, and Signal Processing, 2023, 42(12): 7433-7466.

[15] Singh D, Alam Q. Is tourism expansion the key to economic growth in India? An aggregate-level time series analysis. Annals of Tourism Research Empirical Insights, 2024, 5(2): 100126-100131.

[16] Ashry I, Mao Y, Wang B. A review of distributed fiber–optic sensing in the oil and gas industry. Journal of Lightwave Technology, 2022, 40(5): 1407-1431.

[17] Yerdelen C, Abdelkader M. Long-term trend analysis of annual precipitation for short periodicities. Environmental Engineering & Management Journal (EEMJ), 2021, 20(11): 1721-1729.

[18] Dorais S. Time series analysis in preventive intervention research: A step-by-step guide. Journal of Counseling & Development, 2024, 102(2): 239-250.

[19] Ghashghaie M, Eslami H, Ostad-Ali-Askari K. Applications of time series analysis to investigate components of Madiyan-rood river water quality. Applied Water Science, 2022, 12(8): 202.

[20] Hosseini S, Vaferi B. Determination of methanol loss due to vaporization in gas hydrate inhibition process using intelligent connectionist paradigms. Arabian Journal for Science and Engineering, 2022, 47(5): 5811-5819.

[21] Mahmoodi S, Najafi A A, Ghods K. A Multi-Objective Water Cycle Algorithm for the BI-Objective Multi-Mode Project Resource Renting Problem. Economic Computation & Economic Cybernetics Studies & Research, 2024, 58(1): 104-119.

[22] Ferrara M, Ciano T. Improving the Interpretability of Asset Pricing Models by Explainable AI: A Machine Learning-based Approach. Economic Computation & Economic Cybernetics Studies & Research, 2024, 58(4): 5-19.

[23] Ihueze C C, Okafor C E, Obuka S N, Abdulrahman J, Onwurah U O. Integrity and cost evaluation of natural fibers/HDPE composite tailored for piping applications. Journal of Thermoplastic Composite Materials, 2023, 36(1): 372-398.

[24] Tan Z, Zhang X, Xie B, Wang D, Liu B, Yu T. Fast learning optimiser for real-time optimal energy management of a grid-connected microgrid. IET Generation, Transmission & Distribution, 2022, 16(1): 2977-2987.

[25] Velazquez-Mena A, Benitez-Perez H, Rodríguez-Martínez R C DICOMIST: An methodology for Performing Distributed Computing in Heterogeneous ad hoc Networks. International Journal of Computers Communications & Control, 2024, 19(4): 6526-6527.

[26] Kim J S, Kang D S. A study on fire data augmentation from video/image using the similar-label and F-guessed method. Computer Science and Information Systems, 2024, 21(2): 645-661.

[27] Hema D D, Kumar K A. Novel optimised deep learning approach for an efficient traffic state prediction based on CAE-ICCDCS-GRU model. International Journal of Bio-Inspired Computation, 2024, 23(2): 80-98.

[28] Zhang Yan, Ke Yang. ECW-EGNet: Exploring Cross-ModalWeighting and edge-guided decoder network for RGB-D salient object detection. Computer Science and Information Systems, 2024, 21(3):947-969.

[29] Medrano Cesar L, Diego Alonso G C, Lafarga Octavio V, Cervantes J. MCClusteringSM: An approach for the Multicriteria Clustering Problem based on a Credibility Similarity Measure. Computer Science & Information Systems, 2024, 21(3): 1147-1117.

[30] Shi Y, Xue X, Xue J, Qu Y. Fault Detection in Nuclear Power Plants using Deep Leaning based Image Classification with Imaged Time-series Data. International Journal of Computers Communications & Control, 2022, 17(1): 1-5.

[31] Özcan G E, Karacı A, Enez K. Application of deep neural networks in modeling the capture of Ips sexdentatus in pheromone trap. Environmental Engineering & Management Journal (EEMJ), 2022, 21(1): 105-116.

[32] Barbitoff Y A, Abasov R, Tvorogova V E, Glotov A S, Predeius A V. Systematic benchmark of state-of-the-art variant calling pipelines identifies major factors affecting accuracy of coding sequence variant discovery. BMC genomics, 2022, 23(1): 155-157.

[33] Gasparin A, Lukovic S, Alippi C. Deep learning for time series forecasting: The electric load case. CAAI Transactions on Intelligence Technology, 2022, 7(1): 1-25.

[34] Yang L, Wang R, Zhou Y, Liang J, Zhao K, Burleigh S C. An Analytical Framework for Disruption of Licklider Transmission Protocol in Mars Communications. IEEE Transactions on Vehicular Technology, 2022, 71(5): 5430-5444.

[35] Zhang, D., et al. "Deep learning-based fault diagnosis for oil and gas pipeline networks." IEEE Transactions on Industrial Informatics, 2021, 17(8), 5605-5614.

[36] Rad D, Kiss C, Paraschiv N. Automated comparative predictive analysis of deception detection in convicted offenders using polygraph with random forest, support vector machine, and artificial neural network models. Studies in Informatics and Control, 2024, 33(3): 39-48.

[37] Wang, H., et al. "A survey on deep learning for time series forecasting." ACM Computing Surveys, 2023, 55(2), 1-37.

[38] Li L, Chung W. Application of Artificial Intelligence in Visual Communication of Green Urban Rural Integration Landscape Design. Ecological Chemistry and Engineering, 2024, 31(4): 583-597.

Member since 2012
JM08090

This journal is a member of, and subscribes to the principles of,
the Committee on Publication Ethics (COPE).
https://publicationethics.org/members/international-journal-computers-communications-and-control

*Cite this paper as:*

Mei, C.; Lu, S.; Song, Z.; Li, H.; Feng, Z.; Liu, J. (2025). A Time Series-based Data Modelling Approach for Natural Gas Pipeline Intelligent Management System, *International Journal of Computers Communications & Control*, 20(4), 6813, 2025.
https://doi.org/10.15837/ijccc.2025.4.6813