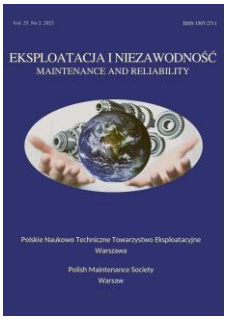


Eksploatacja i Niezawodnosc – Maintenance and Reliability
Volume 28 (2026), Issue 3
journal homepage: http://www.ein.org.pl

Article citation info:
Pang Z, Hou D, Real-time industrial sensor fault diagnosis via global correlation modeling and feature attention-optimized network, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2026: 28(3) http://doi.org/10.17531/ein/218333

Real-time industrial sensor fault diagnosis via global correlation modeling and feature attention-optimized network

Indexed by: Web of Science Group

Zhaolei Pang[a,*], Dali Hou[a]

[a] Shanghai Institute of Technology, China

Highlights

- Developed a model for multi-sensor fault diagnosis that utilizes an MIC and attention-based association network.
- Prevents historical interference and enables precise detection through a dynamic MLP designed for single-step, point-level prediction.
- Proposed correlation compensation aims to eliminate abnormal data interference, ensuring reliable results.

Abstract

To address diagnostic delays and high false-alarm rates in multi-sensor systems, this paper proposes MIC-GAEAN, a lightweight network for anomaly detection. The framework constructs a global sensor correlation network using the Maximal Information Coefficient (MIC) to capture complex dependencies, which is dynamically refined via a lightweight attention mechanism. Notably, training excludes the target sensor's own historical data to prevent information leakage and enhance generalization. An adaptive MLP then performs point-level anomaly detection through single-step prediction. A novel Correlation Compensation Mechanism further uses healthy sensor data to set theoretical norms, distinguishing faults from normal variations and reducing false alarms. Validated on industrial data, MIC-GAEAN demonstrates high real-time accuracy, efficiency, and suitability for data-scarce, long-interval monitoring, offering a reliable solution for industrial system reliability.

Keywords
correlation compensation, fault diagnosis, attention mechanism, maximal information coefficient, MIC-GAEAN

1. Introduction

1.1. Research background and challenges

In modern industry, where systems are large-scale and process variables are highly coupled [1], sensor networks have become the backbone for real-time monitoring and control, enabling precise data acquisition in complex processes across sectors such as manufacturing and energy transportation [2]. These systems are inherently dynamic, with sensor states interdependent and influenced by various factors, including environmental fluctuations (e.g., temperature, humidity), shifts in operational conditions (e.g., load variations), and external noise [3]. The importance of sensor reliability is highlighted by studies that show undetected sensor faults can significantly reduce system efficiency and increase maintenance costs [4,5].

The complexity of sensor networks introduces unique challenges. First, the interdependency among sensors creates a "domino effect" where a localized fault can propagate through the network, compromising the global system stability [6]. Second, nonstationary operational conditions (e.g., startup, shutdown, or mode transitions) create nonlinear data patterns that traditional methods find difficult to model [7]. Third, external disturbances (e.g., electromagnetic interference)

(*) Corresponding author.
E-mail addresses: Z. Pang (ORCID: 0009-0007-3385-7501) 17352181420@163.com, D. Hou (ORCID: 0009-0000-2981-2174) houdali@sit.edu.cn

frequently contaminate sensor data, resulting in false alarms or delayed fault detection [8]. Furthermore, online devices under varying operating conditions face challenges such as data drift and data imbalance, which further complicate real-time monitoring and fault diagnosis [9].

Our models are designed to achieve precise, single-timestep forecasting—a capability that can be extended to reconstruct complete time series—without the computational burden of repeatedly extracting global spatiotemporal dependencies. Their core efficiency stems from operating on a pre-identified, sparse correlation structure within the sensor network, typically retaining only the most influential associations for a target variable. By excluding the target sensor's own data from the model's input features to prevent information leakage, and through adaptive grouping and training, these frameworks distill a highly lightweight detector. This detector performs anomaly detection at the point level by fitting the behavior of correlated sensors and is not unduly influenced by historical data patterns, thereby achieving robust, high-speed inference. Furthermore, this approach is particularly suitable for monitoring equipment with long sampling intervals, where it is challenging to accumulate sufficient time-series data in the short term for reconstruction-based anomaly detection. The model enables rapid correlation-based temporal prediction and effective anomaly identification even under such data-scarce conditions.

Our contributions are: (1) MIC-GAEAN constructs a global mutual information network using the Maximal Information Coefficient (MIC), overcoming the limitations of localized analysis and neglected nonlinear correlations to quantify complex dependencies; (2) It employs a lightweight feature attention mechanism to dynamically optimize the mutual information network, ensuring strong model interpretability while maintaining high dynamic adaptability; (3) Its "Correlation Compensation" leverages healthy sensor data to establish theoretical norms, minimizing false alarms.

1.2. Limitations of existing methods

Traditional fault diagnosis techniques primarily focus on single-sensor analysis, which does not utilize the rich contextual information embedded in sensor networks. For example, methods based on FMECA and RPN [2] rely on predefined failure modes and qualitative risk assessments, which are static and unable to adapt to changing operational scenarios. Data-driven approaches, such as LSTM-based models [10], excel at capturing temporal dependencies within individual sensor streams but overlook spatial correlations between sensors, resulting in incomplete fault representations [11]. Moreover, these univariate sequential models are often highly susceptible to abrupt changes and prevalent industrial noise, which frequently results in high false positive rates. Although auxiliary mechanisms like data credibility assessment can be incorporated to mitigate such sensitivity [12], they typically introduce substantial computational overhead, further diminishing the practicality of these solutions for real-time, resource-constrained industrial deployments.

A key limitation of traditional methods is their susceptibility to variations in operational conditions. Studies have shown that traditional algorithms experience a substantial decline in diagnostic accuracy when transitioning between different operational modes. This occurs because they treat each sensor in isolation, failing to distinguish between normal fluctuations caused by global condition changes and abnormal deviations that indicate faults [13].

To systematically address the complex interdependencies in multi-source data and achieve precise root-cause diagnosis, the fusion of multi-source features for anomaly detection has become a prominent trend. Within this paradigm, Graph Neural Networks (GNNs) have emerged as a key technology, as they can explicitly model dependencies among variables through graph structures, thereby effectively extracting both local and structural correlations [14]. However, conventional GNN methods often rely on prior knowledge to construct static graph structures, which not only incurs high computational costs but also struggles to adapt to dynamically changing environments during long-term monitoring. Research indicates that the performance of such static models degrades significantly under dynamic operational conditions Industrial wireless sensor networks: Challenges, design principles, and technical approaches. To overcome this limitation, dynamic graph models such as GDCMAD have been proposed to emphasize the importance of modeling evolving relational structures. Nevertheless, these approaches often introduce higher computational complexity and structural uncertainty, and their

trade-off between real-time performance and robustness warrants further exploration.

Regarding analytical methods for quantifying associations in multi-source time-series data, classic correlation coefficients, such as the Pearson correlation coefficient for quantifying linear relationships and the Spearman rank correlation coefficient for measuring monotonic associations, are established and computationally efficient statistical tools. However, they fail to effectively detect and quantify complex association patterns that extend beyond their presupposed linear or monotonic assumptions. In recent years, attention mechanisms have become increasingly prevalent in time-series modeling. By enabling models to capture intricate feature relationships, they demonstrate strong capabilities in simulating complex temporal dependencies among multiple correlated variables [15,16]. For instance, FGEAD utilizes self-attention to enhance the learning of information behaviors and interaction patterns, facilitating the extraction of graph structures from multivariate data [17]. However, association analysis based on attention mechanisms suffers from weak interpretability and typically demands substantial data and computational resources for training.

In this case, existing frameworks often struggle with high-dimensional sensor data, as they either ignore inter-sensor relationships, are limited by the complexities of modeling combined linear and nonlinear uncertainties in such systems [18,19], or require substantial computational resources for analysis. To address these issues, the proposed method utilizes the Maximal Information Coefficient [20] to construct a correlation network. A lightweight feature attention mechanism dynamically optimizes the weights of this network, ensuring its interpretability and efficiency while enabling the dynamic modeling of both global sensor correlations and local temporal patterns. Building upon this refined correlation network, an adaptive multilayer perceptron is constructed for data prediction. This architecture ensures streamlined and reliable input features, leading to accurate predictions and highly efficient model performance.

2. Theoretical analysis

Based on real-time sensor data, this paper proposes the MIC-GAEAN model, a lightweight fault diagnosis framework centered on sparse correlation structure and single-step prediction. Its core logic is: first, using the Maximal Information Coefficient (MIC) [21] to construct a global sensor correlation network, quantifying complex linear and nonlinear interdependencies. A lightweight feature attention mechanism then dynamically optimizes this network's weights to adapt to operational changes. Subsequently, an adaptive multi-layer perceptron (MLP) is constructed based on this refined correlation network to perform anomaly detection. This architecture is designed to reduce computational complexity while enhancing detection accuracy. Crucially, the model excludes the target sensor's own historical data during training, learning instead to predict its state solely from the behavior of its correlated sensor, preventing information leakage and enhancing generalization. When a significant deviation between prediction and observation is detected, a correlation compensation mechanism leverages healthy sensor data to establish a theoretical norm, correcting anomalies and isolating faults to prevent propagation. This design enables precise point-level anomaly detection that distinguishes fault-induced deviations from normal systemic fluctuations, achieving high accuracy with minimal computational overhead. The approach is particularly effective in data-scarce scenarios with long sampling intervals, where traditional reconstruction-based models are less applicable. The diagnostic approach can be structured into a system workflow, as illustrated in Figure1.

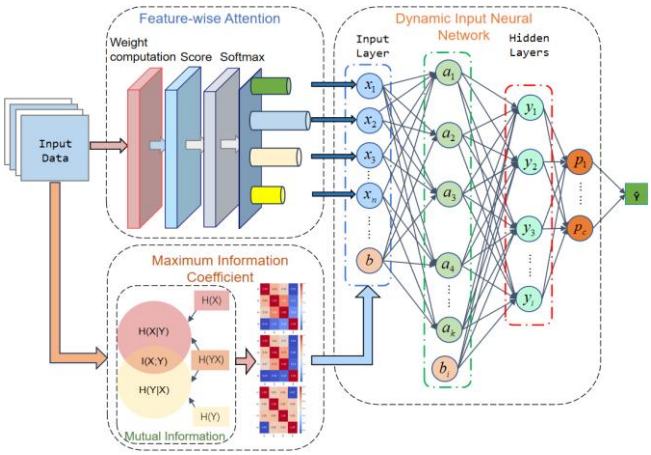


Figure 1. Diagnostic system model structure.

2.1. Correlation analysis module

A correlation analysis module is integrated into the model to construct a sensor association network. By identifying relationships among different features and highlighting those most relevant to fault diagnosis, this module establishes

a structured sensor network that supports subsequent feature weighting and model optimization.

Common methods for evaluating sensor correlations and anomaly levels involve calculating correlation coefficients, such as the widely used Pearson and Spearman coefficients [22] However, these methods exhibit limitations when applied to complex sensor data: the Pearson coefficient assumes linearity and fails to capture nonlinear relationships, while the Spearman coefficient, though capable of handling nonlinearity, depends on rank-based measures and may miss higher-order dependencies [21].

In contrast, the Maximal Information Coefficient (MIC) serves as a more powerful correlation analysis tool, capable of capturing both linear and nonlinear relationships. By maximizing mutual information without assuming specific functional forms or relying on ranks, MIC offers greater flexibility and accuracy in complex datasets. It comprehensively characterizes variable dependencies—particularly useful for data with nonlinear and high-order structures. The core approach involves partitioning the data space into grids, measuring mutual information within each, and normalizing these values to evaluate inter-variable associations [21] As a generalized correlation measure, mutual information (MI) effectively quantifies the shared information and dependence between variables [23]. The specific steps for calculating MIC are as follows:

2.1.1. Scatter plot gridding

For given discrete variables $X = \{x_1, x_2, \dots, x_n\}$ and $Y = \{y_1, y_2, \dots, y_n\}$, a scatter plot is constructed on a 2D plane. The scatter plot is divided into a grid of i columns and j rows, where each grid cell represents a specific range of values for X and Y.

The specific steps are as follows:

Boundary Setting: First, determine the value ranges of X and Y. Let X_{min}, X_{max} and Y_{min}, Y_{max} denote the minimum and maximum values of X and Y, respectively.

Grid Division: Based on X_{min}, X_{max} and Y_{min}, Y_{max}, divide the value range of X into i equal parts and that of Y into j equal parts, forming an i × j grid.

2.1.2. Grid mutual information calculation

After completing the gridding, the information amount in each grid cell is calculated based on the data points within the grid, and the maximum mutual information is computed using these values as the basis for correlation analysis. The calculation of mutual information $I(X;Y)$ relies on the concept of information entropy, which measures the uncertainty of a system and provides deeper insights into the relationships between variables [24]. For a random variable $X = (x_1, x_2, \dots, x_n)$, the information entropy is calculated as:

$$H(X) = \sum_{i=1}^{n} p(x_i) \log p(x_i) \quad (1)$$

In the formula, the variable $p(x_i)$ denotes the probability mass function that the random variable X takes the value x_i, and b is the base of the logarithm, typically 2 (in bits) or e, depending on the unit of information.

The formula for calculating mutual information $I(X;Y)$ is:

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log_b \frac{p(x,y)}{p(x)p(y)} \quad (2)$$

where $p(x,y)$ is the joint probability distribution of X and Y, and $p(x)$, $p(y)$ are their respective marginal probability distributions.

2.1.3. Maximum mutual information calculation

To ensure comparability of mutual information values, normalization is performed. The normalized maximum mutual information ranges between 0 and 1, indicating the relative dependency between variables. The normalization formula is:

$$MIC(X;Y) = \frac{I(X;Y)}{I_{max}} \quad (3)$$

where I_{max} is the maximum mutual information value obtainable under the given dataset and grid division, typically derived through theoretical analysis or simulated data. A normalized mutual information value $MIC(X;Y)$ closer to 1 indicates a stronger correlation between variables.

2.2. Correlation compensation module

In traversal-based fault diagnosis, abnormal sensor data often causes severe interference, as each sensor serves not only as the object of diagnosis but also as a feature and judgment basis in the entire diagnostic process. If a sensor's data is abnormal, the predicted value corresponding to its feature will deviate from the theoretical normal level. To address this, a correlation compensation strategy is proposed. Unlike outlier correction methods such as Imputation [25], Moving Average [26], and Linear Regression Prediction [27], this method mitigates the impacts of numerical mutations and historical data by correcting

outliers based on the dynamic correlations of associated sensors.

The method constructs a fault pre-evaluation model during the model initialization phase. It uses operational data from associated sensors to fit the theoretical normal values of abnormal sensors and gradually replaces the abnormal data. This enhances the stability and accuracy of internal system diagnosis, effectively coping with data interference in complex environments. The following formulas express the core idea:

$$x_{new}^{(n)} = g(x_{related}^{(1)}, x_{related}^{(2)}, \cdots, x_{related}^{(n)}) \quad (4)$$

In the formula, the variable $\hat{y}^{(n)}$ denotes the predicted output value for the $n-th$ sample (sensor value), $x_{related}^{(n)}$ represents the input sensor data for the $n-th$ sample, and f_{BP_n} denotes the model function based on a backpropagation network. Abnormal Data Replacement:

$$\hat{y}^{(n)} = f_{BP_n}(x^{(n)}) \quad (5)$$

In the formula, the function g denotes a correlation-based fitting function that synthesizes

the m correlated sensor observations to reconstruct the theoretical value $x_{new}^{(n)}$. It operates as a multivariate approximation operator, whose mathematical form depends on the statistical properties of correlated data and fitting criteria (e.g., least squares or interpolation)

Input to the next fault diagnosis network:

$$\hat{y}_{new}^{(n)} = f_{BP_n}(x_{new}^{(n)}) \quad (6)$$

Where $y_{new}^{(n)}$ is the new predicted output of the previous fault diagnosis model, $x_{new}^{(n)}$ is the replaced abnormal data, and f_{BP_n} denotes the n-th sensor diagnosis model.

2.3. Traversal fault diagnosis module

The Traversal Fault Diagnosis Module comprises two core sub-modules: the Adaptive MLP trained via the Adam optimizer and the Dynamic Feature Attention Module. This architecture is designed to execute efficient and accurate point-level anomaly detection based on the pre-identified sparse correlation structure. The Adaptive Multi-Layer Perceptron (MLP) serves as the primary anomaly detector. Its key dynamic characteristic is twofold: first, its input feature dimension is dynamically determined by the preceding correlation network, which selects only the most influential peer sensors for the target, thereby reducing computational complexity; second, its effectiveness is enhanced by the dynamic feature attention mechanism which refines the correlation weights in real-time [28]. The Adam (Adaptive Moment Estimation) optimizer is employed to train this Adaptive MLP. It leverages an adaptive learning rate mechanism that computes individual learning rates for each parameter from estimates of the first and second moments of the gradients. This approach ensures stable and efficient convergence of the model by dynamically adjusting the step size during training, which is particularly beneficial for processing the sparse and dynamically weighted features in this diagnostic framework. This overall architecture effectively decomposes the global correlation network into multiple highly correlated sub-networks, which reduces analytical complexity and enhances diagnostic accuracy.

2.3.1. Adaptive MLP sub-module trained with adam optimizer

This sub-module utilizes the Adam (Adaptive Moment Estimation) optimizer for efficient and stable training. Adam is chosen for its proven efficacy in deep learning, as it synergistically incorporates the benefits of both AdaGrad (effective for sparse gradients) and RMSProp (robust to non-stationary gradient statistics). By computing adaptive learning rates for individual parameters based on estimates of the first and second moments of the gradients, Adam facilitates rapid convergence and robust performance. This enables the MLP to effectively learn from the features refined by the preceding Dynamic Feature Attention Module, resulting in a model with high diagnostic accuracy and reliable training stability.

The workflow consists of four succinct steps:

(i) Initialization: Model parameters θ (weights and biases) are initialized. The Adam optimizer's first and second moment vectors, m_0 and v_0.

(ii) Forward/Backward Pass: For a mini-batch, the forward pass yields predictions, the loss $\mathcal{L}$ is computed, and backpropagation calculates the gradient $g_t = \nabla_\theta \mathcal{L}(\theta_{t-1})$.

(iii) Parameter Update: Adam updates its exponential moving averages of the gradient (m_t, for the first moment) and the squared gradient (v_t, for the second moment). To counteract the initial zero-bias of these moving averages, bias-corrected estimates $\hat{m}_t$ and $\hat{v}_t$ are computed. Finally, the model parameters are updated using these corrected estimates. The complete update step for the parameters θ is compactly given by:

$$\theta_t = \theta_{t-1} - \alpha \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t} - \epsilon} \quad (7)$$

Where the bias-corrected moments are defined as $\hat{m}_t = m_t/(1-\beta_1^t)$ and $\hat{v}_t = v_t/(1-\beta_2^t)$, with $m_t = \beta_1 m_{t-1} + (1-\beta_1)g_t$ and $v_t = \beta_2 v_{t-1} + (1-\beta_2)g_t^2$. Here, α is the learning rate, β_1, β_2 are the decay rates for the moment estimates, and ϵ is a small constant for numerical stability. All operations are performed element-wise.

(iv) Iteration: Steps (ii)-(iii) repeat across epochs, refining $\theta$to minimize $\mathcal{L}$ and learn the diagnostic mapping.

2.3.2. Diagnostic sub-module with dynamic MLP and feature attention

This sub-module integrates a lightweight feature attention mechanism with a dynamic Multi-Layer Perceptron (MLP) to form the core diagnostic unit. The key dynamic aspect of this MLP is that its input feature dimension is not fixed but is adaptively determined by the sparse correlation network, corresponding precisely to the number of highly correlated peer sensors identified for the current target. This design significantly reduces computational complexity compared to a standard MLP with a fixed, full-network input dimension. Concurrently, the integrated feature attention mechanism dynamically recalibrates the importance weights of the input features (from these selected correlated sensors) in real-time, ensuring the model focuses on the most relevant signals under varying operational conditions [29,30]. This overall architecture effectively decomposes the global correlation network into multiple highly correlated sub-networks, which reduces analytical complexity and enhances diagnostic accuracy.

The specific workflow is as follows:

Dynamic Input Construction & Attention Weighting: For a target sensor, its pre-identified set of correlated sensors (from the MIC-based network) forms the input feature vector $\mathbf{x}$.

$$\mathrm{s} = \mathrm{W}_2 \cdot \mathrm{ReLU}(\mathrm{W}_1 \cdot \mathbf{x} + \mathrm{b}_1) + \mathrm{b}_2 \quad (8)$$

where $\mathrm{W}_1 \in R^{h \times d}$, $\mathrm{b}_1 \in R^h$, $\mathrm{W}_2 \in R^{h \times d}$, and $\mathrm{b}_2 \in R^d$ are learnable parameters,

d is the input feature dimension, and h is the hidden dimension of the attention network.

A lightweight attention network first processes $\mathbf{x}$ to compute a normalized weight vector $\boldsymbol{\alpha}$:

$$\alpha = \mathrm{softmax}(\mathrm{s}) = \frac{\exp(s_i)}{\sum_{j=1}^{d} \exp(s_i)} \quad (9)$$

Here, $\alpha \in R^d$ denotes the attention weight vector, where each element α_i represents the importance of the $i-th$ feature. The summation in the denominator is taken over all feature dimensions from 1 to d, and s_i is the $i-th$ element of the score vector s. The weighted feature vector is:

$$\tilde{\mathrm{x}} = \mathrm{x} \odot \alpha \quad (10)$$

where $\tilde{\mathrm{x}}$ is the refined feature vector after attention weighting, and $\odot$ denotes the Hadamard (element-wise) product.

Dynamic MLP-based Diagnosis: The weighted feature vector $\tilde{\mathrm{x}}$ is then fed into a compact MLP whose input layer size is dynamically aligned with the dimension of $\tilde{\mathrm{x}}$. This MLP learns to map the weighted correlated sensor readings to an accurate prediction of the target sensor's expected normal value.

$$\hat{y} = MLP_{dynamic}(\tilde{x}\,; \Theta) \quad (11)$$

where Θ represents the MLP parameters optimized initially by the Adam module.

Anomaly Judgment: The final prediction $\hat{y}$ is compared against the actual observed value. A significant deviation triggers the correlation compensation mechanism, indicating a fault.

Parameter Optimization via Loss Minimization: The MLP parameters θ_{MLP} are optimized by minimizing the discrepancy between the predicted value $\hat{y}$ and the ground-truth sensor reading y. This discrepancy is quantified using the Mean Squared Error (MSE) loss function:

$$\mathcal{L}_{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \quad (12)$$

where N is the number of training samples. The Adam optimizer iteratively updates θ_{MLP} to minimize $\mathcal{L}_{MSE}$, thereby enabling the MLP to accurately learn the normal operational mapping.

Variable Summary:

x: Input feature vector of dimension d.

s: Unnormalized attention score vector of dimension d.

α: Attention weight vector after softmax normalization. $\alpha_i \in [0, 1]$ and $\alpha_i = 1$.

$\tilde{\mathrm{x}}$: Output feature vector after applying attention weights.

W_1, b_1: Weight matrix and bias vector of the first linear layer in the attention network.

W_2, b_2: Weight matrix and bias vector of the second linear layer in the attention network.

d: Dimensionality of the input feature vector.

3. Evaluation

To validate the diagnostic performance of the fault diagnosis model under actual operating conditions, this section utilizes sensor data collected from a natural gas transmission station in August 2024. The data integrates key parameters such as temperature, pressure, flow rate, and pressure difference for empirical verification of system faults. During the data reception and preprocessing phase, the system imports time-series data collected by the gas station sensors, performs basic cleaning and normalization to ensure comparability, selects representative sensors, and plots their time-domain waveforms. Through these multi-dimensional operational data, the diagnostic accuracy and robustness of the model in complex environments are evaluated, providing theoretical support for practical applications.

Four representative sensors were selected and assigned identifiers as follows: inlet pressure, pipeline temperature, pressure regulator inlet pressure, and pipeline pressure. These four sensors collectively form a closed-loop monitoring chain for pressure regulation within the station. The three pressure nodes systematically reflect the process flow of natural gas: from the inlet pressure (initial state), through the regulator inlet pressure (key control point), to the pipeline pressure (final state), while the pipeline temperature provides essential thermodynamic state information along the process. The corresponding sensor designations are listed below:

Sensor-1(PT1003)——A1

Sensor-2(TT1402)——A2

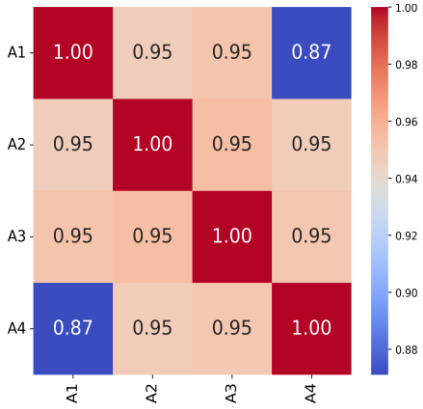
Sensor-3(PT5002)——A3

Sensor-4(PT1102)——A4

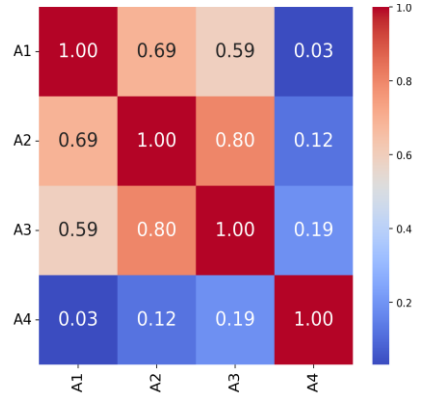
Hardware: Intel® Core™ i5-9300H CPU @ 2.40GHz + NVIDIA GeForce GTX 1660 Ti 6GB GPU; Software: Python 3.8.20, CUDA 12.1. Training parameters: Learning rate r=0.001, 60 epochs, Adam optimizer (Lr=0.001).

3.1. Performance comparison of the correlation analysis module

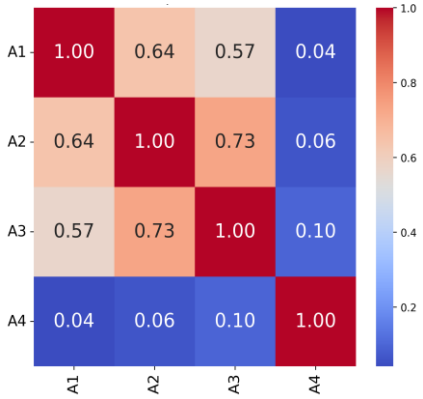
During data evaluation, the proposed method employs the Maximal Information Coefficient (MIC) to analyze sensor associations. Experimental comparisons with conventional methods confirm MIC's superior effectiveness in this framework. Calculations indicate that although linear relationships exist among sensors, they are often obscured by significant noise and interference. The resulting correlation matrices, shown in Figure 2(a), 2(b), 2(c), visually demonstrate MIC's advantage over traditional coefficients in complex sensor systems. MIC adapts well to both linear and nonlinear characteristics while resisting noise interference, enabling more accurate identification of latent relationships under varying conditions. It exhibits significantly higher sensitivity to sensor correlations than other coefficients, confirming its better performance in such models through experimental validation.



(a) MIC Matrix



(b) Pearson matrix



(c) Spearman matrix

Figure 2. Confusion matrix of correlation.

This study also employs the three aforementioned correlation analysis methods to construct association network models and conduct sensor fault diagnosis, yielding comparative results of model performance and accuracy as shown in Figure 3 and Figure 4.

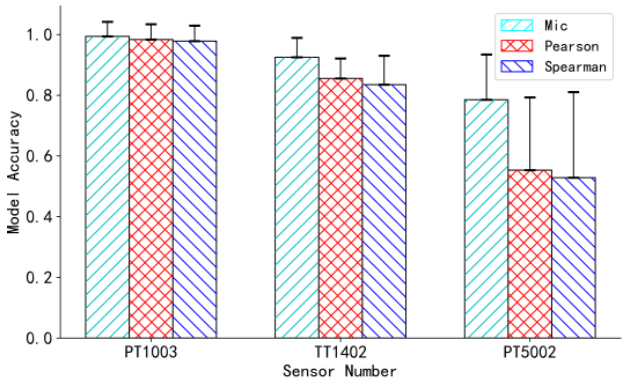


Figure 3. Mean and standard deviation of model accuracy for three methods.

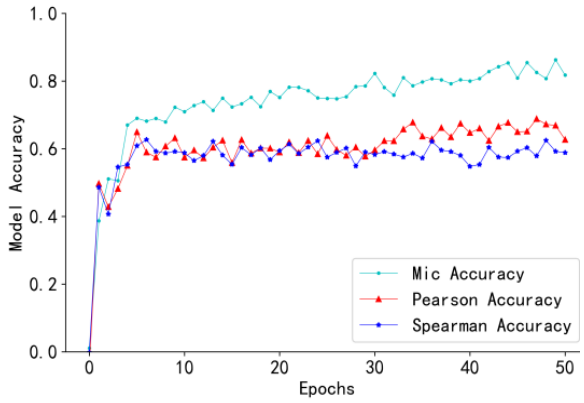


Figure 4. Effect of correlation coefficients on model accuracy.

As shown in Figure 3, the model using the Maximal Information Coefficient (MIC) achieves high accuracy and stable fault diagnosis performance across different scenarios for Sensor-1(PT1003) to Sensor-3(PT5002). In contrast, models based on Spearman and Pearson correlation coefficients exhibit significant accuracy fluctuations and poor stability under varying equipment characteristics and operating conditions. For Sensor-3(PT5002), the Spearman-based model attains an average accuracy of 0.528 with a standard deviation of 0.282, while the Pearson-based model reaches 0.553 with 0.24 deviation. The MIC model significantly outperforms both, with an average accuracy of 0.785 and a standard deviation of 0.149. According to Figure 4, although the Spearman model converges quickly, its accuracy remains low and improves little after initial convergence. The Pearson model shows better convergence and higher accuracy within the same iteration span, yet still underperforms compared to the MIC model. The slower convergence of the MIC model results from its more complex association network and integration of additional features, which ultimately contribute to its superior performance after full optimization.

In summary, the MIC-based model demonstrates higher accuracy, greater stability, and stronger robustness than those using Pearson or Spearman correlations. Therefore, employing MIC to construct the sensor association network enables more efficient and reliable fault diagnosis.

3.2. Validation of the anomaly correction module

The correlation compensation method is applied to handle abnormal sensor data in the case study. A pre-evaluation model performs preliminary fault diagnosis and data accuracy assessment using Equation (4). Detected anomalies are then compensated via Equations (5) and (6), with results cyclically fed back into the model for verification.

The inlet pipeline natural gas temperature (Sensor-2, TT1402) was selected for this comparative analysis. As a thermodynamic variable independent of the station's internal control loop, it synthetically reflects the combined effects of upstream gas source conditions, environmental interference, pipeline transmission, and thermal effects from the pressure regulation process. Furthermore, its relatively high data volatility leads to more pronounced experimental outcomes. Therefore, Sensor-2 (TT1402) was chosen to compare different anomaly correction methods, including the moving average and quartile methods for benchmarking.

The Sum of Absolute Deviations (SAD) is calculated as the total absolute difference between the health degree scores under normal and abnormal states within the sensor group.A lower SAD value indicates smaller deviation and higher fitting accuracy, meaning the assessed state is closer to the normal operational condition**.** This metric quantifies operational deviations, with the comparative results presented in Figure 5.

From Figure5, the total absolute error calculations for the task of Sensor-2(TT1402) show that the original abnormal data is 120.79. Correlation compensation optimization reduces it to 95.29, moving average optimization reduces it to 104.09, and quartile optimization reduces it to 97.12. This indicates that the

equipment health assessment using the correlation compensation method is closest to the normal operation state. It demonstrates minimal overall impact on the sensor network. This is because the correlation compensation method dynamically corrects abnormal data by constructing the sensor network, reducing interference from fluctuations and other disturbances.

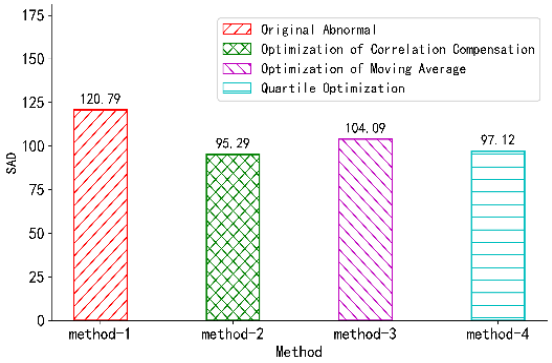


Figure 5. Comparative analysis of sum of absolute deviations for sensor data anomaly handling methods.

The advantages of the correlation compensation method are visualized through health degree line charts under different correction methods. Taking specific data collection points as examples, Figure 6 displays the health score curves for sensors with normal health scores, abnormal health scores, and those processed by three anomaly correction methods. Compared with the normal health curve, abnormal data not only significantly reduces the health score of the abnormal sensor but also affects the health scores of multiple sensor points. This occurs because in the correlated diagnosis system, each sensor serves as a target for fault diagnosis and as a basis for diagnosing other sensors. Abnormal sensors can lower the health scores of associated sensors.

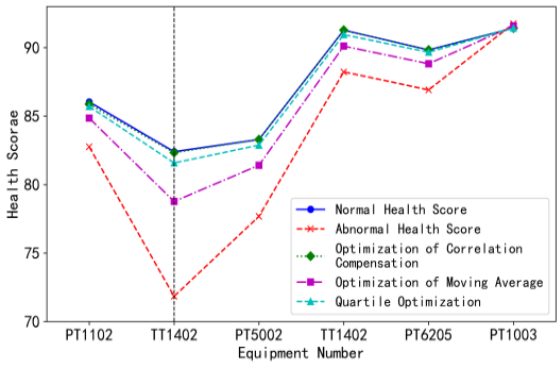


Figure 6. Comparison of sensor health degree curves for different anomaly correction methods.

Taking the point with index Sensor-2(TT1402) in Figure 6 as an example, the normal health score is 82.40. Correlation compensation optimization results in 82.33, quartile optimization results in 81.58, moving average optimization results in 78.76, and the abnormal health score is 71.84. Quartile and moving average optimizations effectively reduce the impact of abnormal data on normal sensors. However, the correlation compensation method is superior—its corrected score curve almost coincides with the normal operation curve. This further confirms that the correlation compensation method can significantly reduce interference from abnormal sensor data to other sensors in this system.

3.3. Comparative analysis of correlation-based and history-dependent models for predicting data transients

This model isolates the features of the target timestep for prediction, meaning its output is derived exclusively from the feature information of that specific timestep, thereby avoiding interference from the target sensor's own data and historical information. To validate the advantages of this design over traditional single time-series models reliant on historical data, the classic CNN-LSTM model is selected for comparison. Data from Sensor-1(PT1003), Sensor-2(TT1402), and Sensor-4(PT1102) are selected for effectiveness verification using a dataset segmentation method. Meanwhile, a traditional single-point time-series prediction model is constructed to perform health assessment on the same sensor tasks, with model accuracy calculated using Equation (13). The comparative results are shown in Figure 7.

$$Accuracy = \frac{1}{N}\sum_{i=1}^{N} I\frac{|\hat{y}_i - y_i|}{|y_i|} \leq threshold \qquad (13)$$

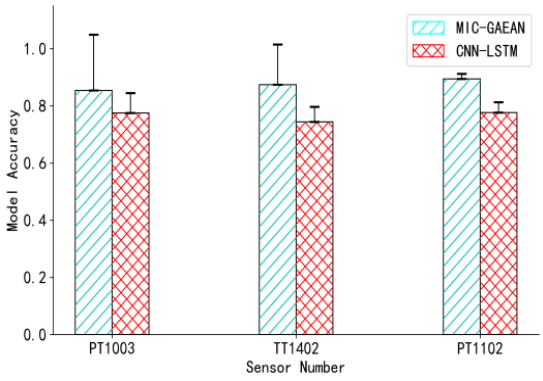


Figure 7. Comparison of model accuracy.

As shown in Figure 7, the two fault diagnosis modes exhibit

similar accuracies under stable sensor operation, with the correlated fault diagnosis model demonstrating slightly higher accuracy and stability in most experimental scenarios compared to the traditional single-point time-series prediction method. Further research and comparison are conducted on normal jumps in sensor data caused by system function adjustments and load changes—common issues in sensor systems. Segments with data jumps are extracted from Sensor-1(PT1003), Sensor-2(TT1402), and Sensor-4(PT1102), as shown in Figure 8, where red dashed lines mark positions corresponding to numerical mutations.

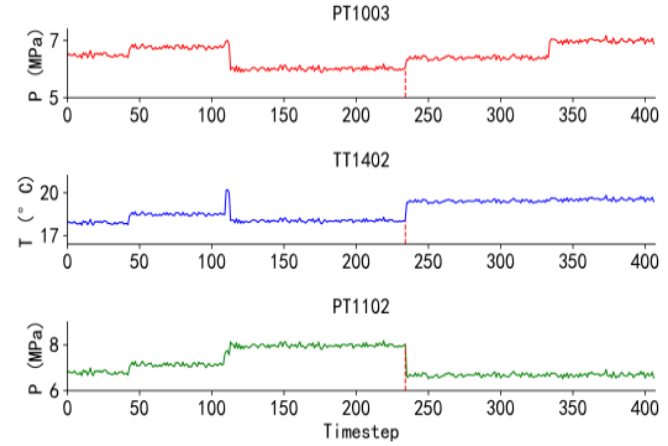


Figure 8. Data jump band.

For the normal jump data in Figure 8, both correlated prediction and single-point time- series models (CNN-LSTM) are used for analysis. A comparison between predicted and actual values is shown in Figure 9. Sensor values are showed in Table 1.

Table 1. Performance comparison at normal jump points for key sensors.

	MIC-GAEAN	Actual Data	CNN-LSTM
PT1003	6.08	6.07	4.88
TT1402	17.29	8.17	9.66
PT1102	6.66	7.78	6.37

The results show that the correlated prediction model outperforms the single-point time-series CNN-LSTM model in predicting jump points across all three sensors, with predictions closely matching the actual values. In contrast, the single-point model relies heavily on historical trends and struggles to accommodate fluctuations caused by load changes or operational shifts. By dynamically incorporating data from related sensors, the correlated model more effectively captures and predicts abrupt data variations.

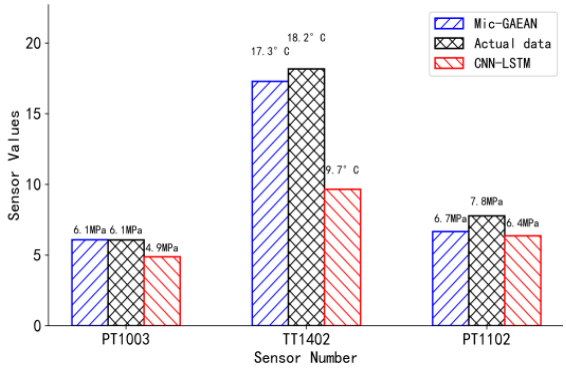


Figure 9. Comparison of model predictions at normal jump points with traditional methods.

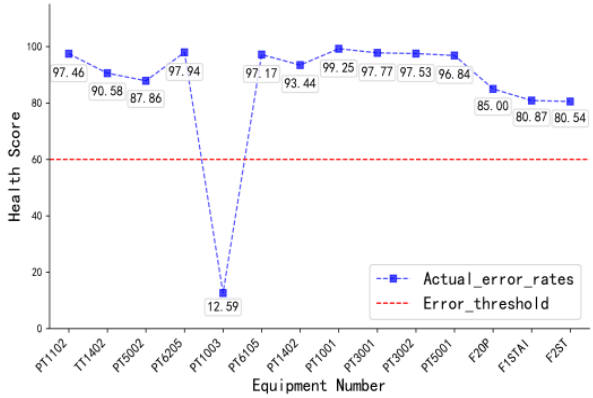


Figure 10. Sensor health score.

3.4. Model Diagnostic Performance

A case study validates the MIC-GAEAN workflow. Its health-index curve (Figure 10) shows the Sensor-1 (PT1003) score falling below the anomaly threshold, confirming detection sensitivity. Correlation compensation isolates the anomaly, preventing propagation to other sensors and enabling reliable early warning.

4. Ablation Study

As the MIC-based correlation network and the dynamic MLP constitute the foundational framework essential for core functionality, they are excluded from this ablation study. The experiments here focus specifically on evaluating the contributions of two key, potentially optional components: the attention mechanism and the Adam optimizer. The analysis proceeds through three sequential parts. We construct model in three stages: Group A (Base): The foundational framework containing only the MIC-based correlation network and the dynamic MLP.

Group B (Base + Adam): Extends Group A by employing

the Adam optimizer to replace a standard optimizer (SGD), isolating the gain from advanced adaptive optimization.

Group C (Full): Further extends Group B by integrating the dynamic attention mechanism, demonstrating the additional benefit of adaptive feature weighting.

An ablation study was conducted on 17 normally operating sensors from a gas transmission station to evaluate the contribution of each core component. The models were trained with a learning rate (Lr) of 0.001 for 60 epochs. The performance metrics, including the R^2 score at the mid-point (epoch 30) and the final statistics, are summarized below in Table 2.

Table 2. Results of the Ablation Study.

Metric	Group A(Base)	Group B (+Adam)	Group C Full)
R^2 Score-Mid	0.3801	0.7182	0.7747
R^2 Score-Full	0.5523	0.8857	0.8777
R^2-Mean-Full	-0.6704	0.1476	0.2796
R^2-Std-Full	2.4330	1.8166	1.6960

The analysis reveals the distinct impact of each added component:

Effect of Adam Optimizer (Group B vs. A): Introducing the Adam optimizer resulted in a profound improvement. The final R^2 score surged from 0.5523 to 0.8857. Notably, the mean R^2 shifted from a negative value (-0.6704) to a positive one (0.1476), indicating the model learned a valid predictive function. The standard deviation also decreased, suggesting more stable training.

Effect of Attention Mechanism (Group C vs. B): Adding the attention mechanism further refined the model's behavior. While the peak final R^2 score experienced a marginal decrease (from 0.8857 to 0.8777), this is offset by significant gains in stability and robustness. The mean R^2 increased by 89% (from 0.1476 to 0.2796), and the standard deviation decreased, confirming that the attention mechanism delivers more consistent and reliable predictions across different sensors, which is critical for a diagnostic system.

In conclusion, Adam is crucial for achieving high predictive accuracy, whereas the attention mechanism is key for ensuring robust and stable performance.

5. Model Performance Comparison

To validate the advantages of the proposed MIC-GAEAN framework, a comprehensive evaluation was conducted against several established deep learning benchmarks: KAN-AD, Transformer-LSTM (Trans-LSTM), and CNN-LSTM. The experiment expanded the dataset by incorporating sensor data collected on May 26 and June 26 from a natural gas transmission station, combined with the original data. From this consolidated dataset, 17 normally operating sensors were selected for evaluation. All models were trained for 60 epochs with a learning rate of 0.001. Performance was measured using the R^2 score and Mean Squared Error (MSE), with metrics recorded at the mid-point (epoch 30) and the end of training. The single-epoch computation time was also recorded. As shown in Table 3.

Table 3. Comparative Model Performance on Sensor Data Prediction.

Metric	KAN-AD	Trans-LSTM	CNN-LSTM	MIC-GAEAN
R^2 Score-Mid	0.8733	0.6579	0.5521	0.8047
MSE-Mid	0.0032	0.0081	0.0123	0.0059
R^2 Score-Full	0.8977	0.8110	0.6286	0.8756
MSE-Full	0.0026	0.0045	0.0100	0.0036
Epoch-Time(s)	13.94	4.61	3.88	0.58

The proposed MIC-GAEAN demonstrates a strong and efficient performance profile:

Superiority over Classic Architectures: MIC-GAEAN significantly outperforms the classic CNN-LSTM, with a final R^2 score 39% higher (0.8756 vs. 0.6286) and an 85% faster iteration speed.

Advantage over a Contemporary Benchmark: It also maintains a clear advantage over the contemporary Transformer-LSTM model, achieving an 8% higher final R^2 score while being nearly 8 times faster per epoch.

Balance of Accuracy and Extreme Efficiency: Compared to the highly accurate KAN-AD model, MIC-GAEAN's final R^2 score is marginally lower (a 2.5% relative difference). However, it accomplishes this with a striking 24-fold reduction in per-epoch computation time (0.58s vs. 13.94s). This exceptional efficiency is critical for real-time industrial diagnostics.

In conclusion, while KAN-AD achieves the higher accuracy, the MIC-GAEAN framework offers a better balance, delivering highly competitive prediction performance ($R^2 > 0.87$) with superior computational efficiency, making it particularly suitable for resource-conscious or time-sensitive deployment

environments.

6. Conclusion

This study proposes MIC-GAEAN, a lightweight fault diagnosis framework that integrates a Maximal Information Coefficient (MIC)-based global correlation network with a dynamic attention mechanism. By constructing sparse sensor groups and training independent sub-models that exclude the target sensor's own history, the method enables precise, single-step anomaly detection while preventing information leakage. The introduced correlation compensation mechanism further enhances reliability by leveraging healthy sensor data to mitigate fault propagation. Validation using real-world natural gas station data confirms the framework's accuracy and efficiency, particularly in data-scarce environments with long sampling intervals.

Future work will focus on extending the model to online learning settings, integrating it with graph neural networks for deeper relational reasoning, incorporating combined frequency- and time-domain analyses to better capture dynamic patterns, and enhancing explainability and cross-domain transferability for broader utility in complex engineering systems.

References

1. Zhao Q, Liu X, Li H, Han Y. Multi-Connected Temporal Encoding Graph Neural Network for Complex Industrial Fault Diagnosis. Eksploatacja i Niezawodność – Maintenance and Reliability 2026; 28(1). https://doi.org/10.17531/ein/208666.
2. Chauhan S, Vashishtha G, Zimroz R. Analysing recent breakthroughs in fault diagnosis through sensor: A comprehensive overview. CMES-Computer Modeling in Engineering & Sciences 2024; 141(3). https://doi.org/10.32604/cmes.2024.055633.
3. Sharma AB, Golubchik L and Govindan R. Sensor faults: Detection methods and prevalence in real-world datasets. ACM Transactions on Sensor Networks (TOSN) 2010; 6(3): 1–39. https://doi.org/10.1145/1754414.1754419.
4. Xie X, Wang J, Han Y, Li W, Knowledge graph-based in-context learning for advanced fault diagnosis in sensor networks. Sensors 2024; 24(24): 8086. https://doi.org/10.3390/s24248086.
5. Verhelst J, Van Ham G, Saelens D, Helsen L, Economic impact of persistent sensor and actuator faults in concrete core activated office buildings. Energy and Buildings 2017; 142: 111–127, https://doi.org/10.1016/j.enbuild.2017.02.052.
6. Mahapatro A, Khilar P M. Fault diagnosis in wireless sensor networks: A survey. IEEE Communications Surveys & Tutorials 2013; 15(4): 2000–2026. https://doi.org/10.1109/ICCS.2018.00011.
7. Ni K, Ramanathan N, Chehade M N H, Balzano L, Nair S, Zahedi S, Kohler E, Pottie G, Hansen M, Srivastava M. Sensor network data fault types. ACM Transactions on Sensor Networks (TOSN) 2009; 5(3): 1–29. https://doi.org/10.1145/1525856.1525863.
8. Jombo G, Zhang E and Lu N. Sensor fault detection and diagnosis: Methods and challenges. In 12th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes. IFAC Proceedings 2024.
9. Lei X, Lu N, Jiang B, Wang C, Chen C. A Multi-scale Attention Mechanism Diagnosis Method with Adaptive Online Updating Based on Deep Learning under Variable Working Conditions. Maintenance Reliability-Eksploatacja i Niezawodność 2025; 27(1). https://doi.org/10.17531/ein/192975.
10. Greff K, Srivastava RK, Koutnuık J, Steunebrink B R, Schmidthuber J. Lstm: A search space odyssey. IEEE transactions on neural networks and learning systems 2016; 28(10): 2222–2232. https://doi.org/10.1109/TNNLS.2016.2582924.
11. Ji Z, Ye S, Ma X. Sparse coding inspired lstm and self-attention integration for medical image segmentation. IEEE Transactions on Image Processing 2024; 33: 6098–6113. https://doi.org/10.1109/TIP.2024.3482189.
12. Fei H, Xiao F, Li G H, Sun L J, Wang R C. An anomaly detection method of wireless sensor network based on multi-modals data stream. Chinese Journal of Computers 2017; 40(8): 1829–1842. https://doi.org/10.11897/SP.J.1016.2017.01829.
13. Zhang J, Luo W, Dai Y. Integrated diagnostic framework for process and sensor faults in chemical industry. Sensors 2021; 21(3): 822. https://doi.org/10.3390/s21030822.
14. He S, He W, Du M, Jiang X, Dong Y. GDCMAD: Graph-based dual-contrastive representation learning for multivariate time series anomaly detection. Information Sciences 2025; 122790. https://doi.org/10.1016/j.ins.2025.122790.
15. Huang Y, Jin Y, Tao K, Xia K, Gu J, Yu L, Li H, Du L, Chen C. MTS-Net: dual- enhanced positional multi-head self-attention for 3D CT diagnosis of May-Thurner syndrome. Biomedical Signal Processing and Control 2026; 112: 108618. https://doi.org/10.1016/j.bspc.2025.108618.

16. Lai Q, Chen P. Unveiling node relationships for traffic forecasting: A self-supervised approach with MixGT. Information Fusion 2025; 120: 103070. https://doi.org/10.1016/j.inffus.2025.103070.
17. Xu W, Wang H, Xie Z. Explainable multivariate time series anomaly detection by feature graph structure learning. Information Sciences 2025; 122789. https://doi.org/10.1016/j.ins.2025.122789.
18. Francis D, Sun F. A comparative analysis of mutual information methods for pairwise relationship detection in metagenomic data. BMC bioinformatics 2024; 25(1): 266. https://doi.org/10.1186/s12859-024-05883-7.
19. Prosvirin A E, Piltan F, Kim J M. Blade rub-impact fault identification using autoencoder- based nonlinear function approximation and a deep neural network. Sensors 2020; 20(21): 6265. https://doi.org/10.3390/s20216265.
20. Kinney J B, Atwal G S. Equitability, mutual information, and the maximal information coefficient. Proceedings of the National Academy of Sciences 2014; 111(9): 3354–3359. https://doi.org/10.1073/pnas.1309933111.
21. Reshef D N, Reshef Y A, Finucane H K, Grossman S R, Mcvean G, Turnbaugh P J, Lander E S, Mitzenmacher M, Sabeti P C. Detecting novel associations in large data sets. Science 2011; 334(6062): 1518–1524. https://doi.org/10.1126/science.1205438.
22. De Winter J C, Gosling S D, Potter J. Comparing the pearson and spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. Psychological methods 2016; 21(3): 273. https://doi.org/10.48550/arXiv.2408.15979.
23. Duncan T E. On the calculation of mutual information. SIAM Journal on Applied Mathematics 1970; 19(1): 215–220. https://doi.org/10.1137/0119020.
24. Nagel D, Diez G, Stock G. Accurate estimation of the normalized mutual information of multidimensional data. The Journal of Chemical Physics 2024; 161(5). https://doi.org/10.1063/5.0217960.
25. Donders A R T, Van Der Heijden G J, Stijnen T, Moons K G M. A gentle introduction to imputation of missing values. Journal of clinical epidemiology 2006; 59(10): 1087–1091. https://doi.org/10.1016/j.jclinepi.2006.01.014.
26. Hyndman R J. Moving averages. In International encyclopedia of statistical science. Springer, 2011. pp. 866–869. https://doi.org/10.1007/978-3-642-04898-2_380.
27. Cai T T, Hall P. Prediction in functional linear regression. The Annals of Statistics 2006; 34(287): 2159–2179. https://doi.org/10.1214/009053606000000830.
28. Huang J, Yan X. Related and independent variable fault detection based on kpca and svdd. Journal of Process Control 2016; 39: 88–99. https://doi.org/10.1016/j.jprocont.2016.01.001.
29. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L, Polosukhin I. Attention is all you need. Advances in neural information processing systems 2017; 30. https://doi.org/10.48550/arXiv.1706.03762.
30. Qin X, Wang Z, Bai Y, Xie X, Jia H. FFA-Net: Feature fusion attention network for single image dehazing. In Proceedings of the AAAI Conference on Artificial Intelligence 2020; 34(7): 11908–11915. https://doi.org/10.48550/arXiv.1911.07559.