

Research on elevator door anomaly detection method based on multisource spatial-temporal information fusion

Qingyun Li

*College of Information Science and Engineering, Northeastern University,
Shenyang, China*

Gongning Li

*Key Laboratory of High-Parameter Elevator Intelligent Operation and Maintenance,
State Administration for Market Regulation,
Special Equipment Safety Supervision Inspection Institute, Nanjing, China*

Mukai Wang

China Special Equipment Inspection and Research Institute, Beijing, China

Jiefeng Li and Yongqing Wang

*Key Laboratory of High-Parameter Elevator Intelligent Operation and Maintenance,
State Administration for Market Regulation,
Special Equipment Safety Supervision Inspection Institute, Nanjing, China, and*

Mingxing Jia and Dapeng Niu

*College of Information Science and Engineering, Northeastern University,
Shenyang, China*

Abstract

Purpose – Elevator door system failures are a leading cause of elevator malfunctions, impacting safety and operational efficiency. Existing anomaly detection methods often overlook the relative positional relationships among time-series data sequences. This study aims to propose a novel spatial-temporal information fusion approach to accurately identify abnormal states in elevator door systems.

Design/methodology/approach – This paper develops an integrated spatial-temporal fusion framework for anomaly detection. First, operational time-series data are structured into a node graph using a K-Nearest Neighbor (KNN) based method to model complex inter-sequence interactions. Subsequently, a Graph Convolutional Network (GCN) is employed to extract local dependencies and global spatial information, while a Bidirectional Long Short-Term Memory network (BiLSTM) captures temporal evolutionary characteristics. A multi-source information-driven feature fusion mechanism is then designed to enhance model robustness. The proposed method is experimentally validated using real-world elevator door operating data, including vibration and video-derived signals.

Findings – The experimental results demonstrate that the proposed method effectively identifies abnormal elevator door states, achieving a high detection accuracy of 96.27%. This confirms the framework's effectiveness and reliability in practical scenarios.

Originality/value – This research presents a novel KNN-based graph construction method that captures dynamic dependencies between time-series sequences based on their relative positions. Furthermore, it develops

© Qingyun Li, Gongning Li, Mukai Wang, Jiefeng Li, Yongqing Wang, Mingxing Jia and Dapeng Niu. Published in *Journal of Intelligent Manufacturing and Special Equipment*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at [Link to the terms of the CC BY 4.0 licence](#).

Funding: This work was supported by the Open Project Fund of the Key Laboratory of High-Parameter Elevator Intelligent Operation and Maintenance, State Administration for Market Regulation under Grant No. JSTJ-IOMHL-202506, the National Natural Science Foundation of China under Grant No. 62573101, and Jiangsu Provincial Special Equipment Safety Supervision and Inspection Institute Science and Technology Plan Project under Grant No. KJ(Y)202512.



an integrated framework that concurrently models spatial structural relationships and temporal dynamics, overcoming the limitations of methods that treat these dimensions separately. Finally, it introduces a multi-source feature fusion mechanism that leverages the complementarity of information from different sources and dimensions, significantly enhancing the model's representation capability and robustness under complex operating conditions.

Keywords Elevator door system, Anomaly detection, K-nearest neighbor, Spatial-temporal network, Information fusion

Paper type Research article

1. Introduction

As a critical subsystem for passenger access, elevator door performance significantly impacts both safety and operational efficiency. Statistics indicate that malfunctions in elevator door systems account for over 50% of all elevator failures, constituting a leading cause of passenger entrapments, service disruptions, and even fatal incidents such as shearing injuries and falls (Lan *et al.*, 2021; Wang *et al.*, 2018). Given the structural complexity and variable operating conditions of elevator door systems, performance degradation often manifests as gradual, latent processes that can culminate in sudden failures. Consequently, reliance solely on traditional periodic inspections or reactive repairs proves insufficient for achieving effective fault prediction and precise performance assessment (An *et al.*, 2021). Therefore, the development of intelligent, real-time anomaly detection methods for elevator door systems is crucial for enhancing reliability, ensuring passenger safety, and optimizing maintenance strategies.

In recent years, elevator operation condition monitoring and fault diagnosis techniques have made significant progress (Esteban *et al.*, 2016; Wang *et al.*, 2024). Numerous researchers have evaluated and predicted the overall health state of elevators utilizing vibration analysis, operational data mining, artificial intelligence algorithms, and related approaches. Recent advances in deep learning for fault diagnosis in mechanical and electromechanical systems, such as multi-scale capsule networks (Lu *et al.*, 2025), vision transformers with time-frequency fusion (Xiao *et al.*, 2025a), and novel convolutional attention mechanisms (Wang *et al.*, 2025), have demonstrated strong feature extraction capabilities. However, their direct application to the specific spatio-temporal dependency problem in elevator door systems remains unexplored. For instance, Niu monitored elevator conditions through time-domain and frequency-domain analysis of vibration signals (Niu *et al.*, 2025). Pan proposed an elevator risk assessment method based on an integration of fuzzy comprehensive evaluation and machine learning (Pan *et al.*, 2023). Guo presented a semi-supervised two-stage deep learning network, comprising a feature selector and a classifier, for elevator condition assessment (Guo *et al.*, 2024a). However, anomaly detection and performance evaluation research on door systems, high-incidence areas for elevator failures that directly impact passenger safety, remains relatively scarce. Existing methods assessing the overall elevator state often struggle to capture the subtle performance degradation characteristics unique to door systems. Moreover, a real-time evaluation system capable of assessing the dynamic performance of door systems under complex interaction conditions is notably lacking. Under such conditions, failures often arise from chain reactions and spatio-temporal coupling among multiple component degradations during operation. For example, belt wear may affect door speed, which in turn exacerbates vibration induced by worn sliders. Therefore, research focused on anomaly detection methods for door systems is of significant importance for advancing predictive maintenance and enabling early fault warning in elevators.

Traditional signal processing and shallow learning methods usually rely on handcrafted features and assume fixed temporal or statistical patterns. However, when applied to elevator door systems, such assumptions often become restrictive. Elevator door anomalies frequently exhibit strong nonlinearity, temporal dependency, and structural coupling among signal segments, which are difficult to capture using conventional approaches. Graph convolutional

networks provide a natural way to model such non-Euclidean structural relationships, while deep sequential models enable automatic extraction of hierarchical temporal features. As a result, combining graph convolutional networks with deep temporal models provides a suitable framework for modeling elevator door anomalies.

To address the challenge of capturing the performance characteristics of the door system and the lack of a dynamic evaluation framework, this paper proposes a method that integrates temporal and spatial features from multi-source information to enable accurate detection of elevator door anomalies. This research aims to construct a framework capable of simultaneously capturing spatial and temporal dynamic features and achieving high-precision anomaly state identification. The main contributions of this work are summarized as follows:

- (1) A KNN-based method is proposed for constructing a graph structure from time-series data. This structure characterizes spatial correlations between sensor nodes while effectively capturing dynamic dependencies among sequences based on their relative positions.
- (2) An integrated spatiotemporal information processing framework is developed. Within this framework, graph convolutional networks extract spatial features encompassing both local node dependencies and global spatial topology. Concurrently, bidirectional long short-term memory networks are employed to capture long-term evolutionary patterns and short-term fluctuation characteristics within the temporal dimension.
- (3) To overcome the limitations of single-source signals, a feature fusion mechanism leveraging multi-source information is proposed to enhance representation and detection accuracy. This mechanism fully leverages the complementarity of information from different sources and dimensions, thereby enhancing the model's reliability under complex operating conditions.

The remainder of the paper is organized as follows: [Section 2](#) introduces related work on spatial-temporal and multi-source information fusion. [Section 3](#) details the proposed methodology for multi-source spatiotemporal fusion. [Section 4](#) describes the data acquisition process and presents experimental validation of the proposed method. Finally, [Section 5](#) concludes the paper.

2. Related work

In recent years, with the popularization of industrial Internet of Things (IoT) technology, multi-dimensional data such as vibration, sound, and video of door opening and closing generated during the operation of elevator door systems have been realized to be collected in real time ([Xie et al., 2024](#); [Hsu et al., 2020](#); [Pan et al., 2024](#)). These data contain rich operational state and safety information, which not only reflect the dynamic characteristics of the system evolving over time, but also reflect the interrelationship of the states of the system at different locations. Owing to their advantage in modeling non-Euclidean spatial relationships in sensor networks, graph convolutional networks have found widespread application in areas such as traffic flow prediction and social behavior analysis since Kipf and Welling introduced the semi-supervised framework ([Kipf and Welling, 2016](#)). Yang combined capsule networks with GCN for harmonic drive compound fault diagnosis ([Yang et al., 2023a](#)). Seo proposed a method combining graph autoencoders with graph convolutional networks to model dynamic relationships among subsystems in hydrogen extraction systems, enabling effective anomaly detection and fault diagnosis ([Seo et al., 2024](#)). Chen proposed the SA-GCN framework, which integrates structural analysis prior knowledge with measurement data ([Chen et al., 2022](#)). Xiao proposed MV-TriGCN, a semi-supervised multi-view learning framework that enhances GCN stability and generalization through an improved triplet loss, diverse view graph construction, and stepwise training ([Xiao et al., 2025b](#)). Most of the above studies focus

on modeling the spatial structural features of sensors. However, in practice, sensor data often contain both temporal evolution patterns and spatial dependencies, demonstrating tightly intertwined spatial-temporal characteristics with distinct yet interacting feature dimensions. Therefore, a comprehensive consideration of both temporal and spatial dimensions is needed to more fully explore the data's intrinsic structure and potential semantics.

Furthermore, in the field of intelligent operation and maintenance of industrial equipment, the integration of multi-source sensor data for spatial-temporal modeling has emerged as a core technical approach for condition monitoring and performance evaluation (Wu *et al.*, 2024; Han *et al.*, 2025; Lv *et al.*, 2025). Fathizadan proposed a spatio-temporal anomaly detection framework using convolutional LSTM autoencoders and control charts to effectively identify anomalies in additive manufacturing processes (Fathizadan *et al.*, 2024). Sun performed fault diagnosis through temporal node connection modeling (Sun and Yin, 2025). Guo proposed a time-frequency domain interpretation method for CNNs in bearing fault diagnosis, enhancing model transparency through Grad-CAM and gradient-ascent-based kernel visualization (Guo *et al.*, 2024b). Allen proposed the KESA framework, which combines deep learning with domain knowledge for spatiotemporal fault detection and interpretation in complex industrial processes (Allen *et al.*, 2024). Addressing the limitations of multi-sensor long-term sequence modeling in structural health monitoring, Yang utilized a two-layer attention mechanism to enhance LSTM's ability to capture spatial-temporal dynamic features (Yang *et al.*, 2023b). Zhou realized spatial-temporal joint modeling through the coupling of multi-view graph attention and temporal convolution (Zhou and Wang, 2024). However, the above methods generally define the graph structure based on the static correlations of sensors, while ignoring the dynamic dependencies between time series data points based on their relative temporal positions. This omission of temporal association patterns hinders the model's ability to fully capture the complex spatiotemporal interactions during equipment operation, consequently limiting diagnostic accuracy and robustness.

Although deep learning has significantly promoted the development of intelligent fault diagnosis techniques, methods relying on a single signal source often face a generalization bottleneck due to insufficient feature characterization capability under complex and severe working conditions. For this reason, multi-source heterogeneous data fusion technology has become an important research direction. Currently, fault diagnosis research based on information fusion is conducted at three levels, i.e., the data level, feature level, and decision level. Zhang employed a residual pyramid algorithm to separately fuse acoustic and vibration signals from multiple spatial locations, generating two fused acoustic-vibration signals, and subsequently constructed a multi-source fault feature set (Zhang *et al.*, 2022a). Feng proposed a multisource state space-based method for tool RUL prediction that models multistage degradation using a Wiener process, achieving improved prediction accuracy by integrating historical and real-time monitoring data (Feng *et al.*, 2025a). Xu proposed a dynamic feature selection matrix optimization information integration method to transform a multi-source information system into a single-source information system (Xu and Li, 2025). Feng proposed a diagnostic framework based on vibration, electrostatic, and infrared multi-source heterogeneous data fusion, and achieved online diagnosis of gearbox faults through data conversion and feature fusion (Feng *et al.*, 2025b). While these fusion methods have shown promise in fault diagnosis, they also raise concerns regarding data transformation and feature redundancy, as well as decision ambiguity.

In summary, current research has made significant progress in graph structure construction and multi-source data fusion, especially in modeling spatial structure using GCN and capturing temporal dependence with the help of models such as LSTM. However, most of the existing methods deal with the temporal and spatial dimensions separately, ignoring the interactive characteristics of temporal location and structural relationships among sequences, which makes it difficult to comprehensively portray the dynamic evolution mechanism in equipment operation. To this end, this paper proposes a multi-source information anomaly detection method that fuses temporal and spatial features. This method accurately identifies

abnormal states of the elevator door system by deeply extracting latent temporal and spatial dependencies from the data and effectively incorporating them within a multi-source information fusion framework.

3. Methodology

3.1 Problem formulation and overall architecture

Elevator door anomaly detection focuses on identifying abnormal operating states during door opening and closing processes based on multi-source time series data. In practical operation, door system faults usually do not appear as isolated signal fluctuations, but rather manifest as localized deviations in temporal evolution and abnormal interactions among different signal segments. Such patterns are typically associated with interactions among multiple degraded components during door operation.

Therefore, elevator door anomaly detection is treated as a spatio temporal pattern recognition task, in which both the temporal evolution of individual signals and the dependency relationships among time series segments are considered. Effective detection requires capturing long-term trends and short-term fluctuations in the temporal domain, as well as modeling the structural relationships that reflect similarity and interaction among segments under different operating conditions. Based on this consideration, a spatio temporal modeling framework is developed to support accurate identification of abnormal door operation states.

To address this problem, this study proposes a multi-source spatio temporal information fusion framework for elevator door anomaly detection, as illustrated in Figure 1. The overall framework integrates temporal feature extraction, spatial dependency modeling, and multi-source feature fusion to achieve accurate identification of abnormal door operation states.

Specifically, the proposed framework consists of a temporal modeling module, a spatial modeling module, and a feature fusion and classification module. The input of the framework includes multi-source signals collected during elevator door operation, such as video-derived speed sequences and vibration signals. These signals are first preprocessed and segmented into time series samples, which serve as the basic units for subsequent spatio temporal modeling.

In the temporal modeling module, CNN are employed to extract local temporal features, while BiLSTM are used to capture long term temporal dependencies. In addition, a multi-head attention pooling mechanism is introduced to emphasize time steps that are more sensitive to anomaly characteristics, thereby enhancing the discriminative ability of temporal features.

In parallel, the spatial modeling module aims to characterize the dependency relationships among time series segments. Each segment is treated as a node in the graph, and a K nearest neighbor based graph construction strategy is used to establish similarity connections among nodes. Graph convolutional networks are then applied to the constructed graph to extract spatial features that reflect localized and sample specific interactions associated with abnormal door behavior.

Finally, the temporal features and spatial features extracted from different sources are fused to form a unified spatio temporal representation. This fused representation is fed into a fully connected classifier to output the anomaly category of the elevator door system.

The structure of the fusion module is expressed as:

$$\begin{cases} z_{fused} = ReLU(W_1 [z_{time}; z_{graph}] + b_1) \\ \hat{y} = Softmax(W_2 z_{fused} + b_2) \end{cases} \quad (1)$$

Where z_{fused} represents the fused spatial-temporal feature set. z_{time} represents the multi-source temporal feature set. z_{graph} represents the multi-source spatial feature set. W_i represents the weight, and b_i represents the bias. \hat{y} represents the output probability distribution over each state category.

Through this integrated spatio temporal modeling framework, the proposed method enables effective anomaly detection under complex operating conditions.

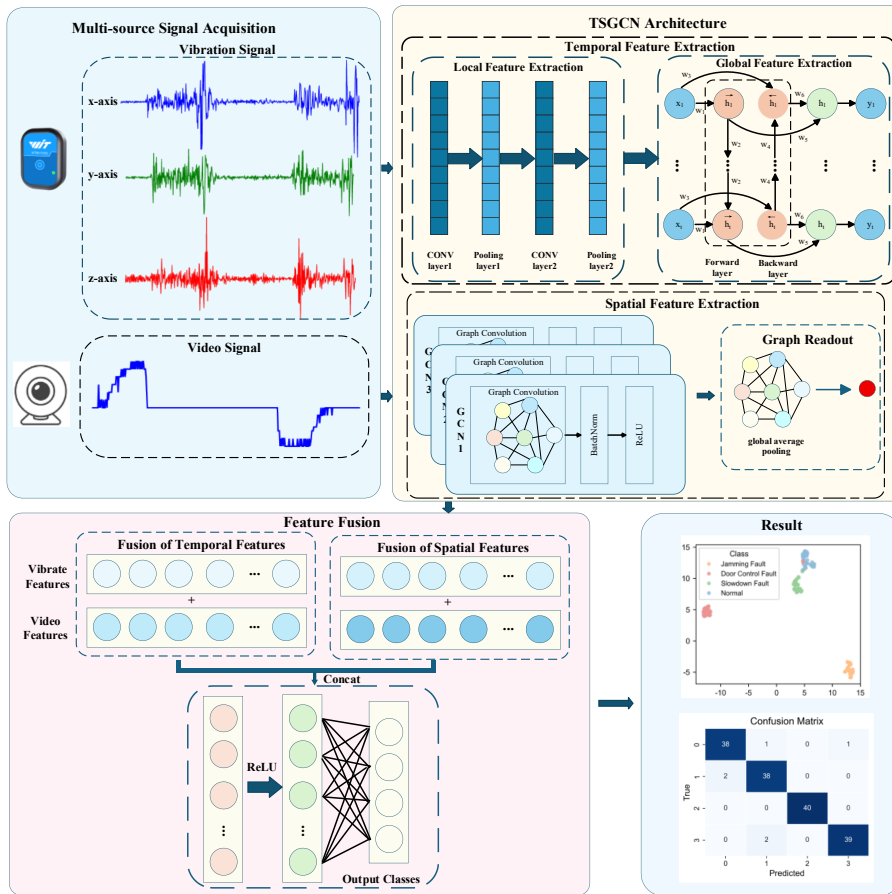


Figure 1. Convolutional network fusion model for spatial-temporal maps. **Source(s):** Figure created by authors

3.2 KNN graph construction

Selecting an appropriate graph construction strategy is crucial for modeling the dynamic interactions in elevator door operation. Existing approaches include fully connected graphs, correlation-based graphs, and physical topology-based graphs. Fully connected graphs introduce dense and redundant connections, which may lead to over-smoothing and reduced discriminative ability in graph convolutional networks (Li *et al.*, 2018). Correlation-based graph construction methods rely on global statistical dependency and typically produce static graphs, which may be insufficient to characterize localized and transient interactions in time-series segments (Wu *et al.*, 2020). Physical topology-based graphs require explicit prior knowledge of sensor layouts or mechanical structures; however, in elevator door systems, the interactions among temporal segments are primarily driven by motion consistency and feature similarity rather than fixed spatial adjacency.

To address these limitations, we employ a KNN based approach to construct a dynamic, sample-specific graph. This method connects each time series segment to its k most similar neighbors in the feature space, effectively capturing localized relationships that are indicative of anomalous patterns. The KNN graph is adaptive to each sample and preserves locality,

making it particularly suitable for detecting the subtle, localized deviations characteristic of elevator door faults (Qi *et al.*, 2021; Seo *et al.*, 2024).

The detailed construction process is as follows.
A time series sample of length T can be represented as:

$$x = [x_1, x_2, \dots, x_T] \quad (2)$$

Before constructing the node graph, global normalization of the sequences is required to eliminate numerical scale differences between different samples. The standardization formula is as follows:

$$\tilde{x}_t = \frac{x_t - \mu}{\sigma} \quad (3)$$

where \tilde{x}_t represents the standardized sample value at time t . x_t denotes the original time series sample value at time t . μ and σ stand for the global mean and standard deviation across all time points, respectively.

Subsequently, the sliding window method is used to extract node features. At time t , a sliding window, consisting of the time point itself and the subsequent $w - 1$ time points, serves as the node input. The features of the t th graph node are then constructed as follows:

$$v_t = [\tilde{x}_t, \tilde{x}_{t+1}, \dots, \tilde{x}_{t+w-1}] \in \mathbb{R}^w \quad (4)$$

This approach effectively captures the dynamic changes of local time slices and enhances the temporal awareness of nodes.

The graph structure data is further constructed based on the node features and the topology of the graph. It can be represented as follows:

$$G = G(V, E, A) \quad (5)$$

where $V \in \mathbb{R}^{n \times d}$ is the node feature matrix, with n being the number of nodes and d being the feature dimension of a single node. E represents the edge set. $A \in \mathbb{R}^{n \times n}$ represents the adjacency matrix of the graph.

To establish structural connections between nodes, the edge set is constructed using the KNN method with Euclidean distance (Qi *et al.*, 2021). For each node i , the k nearest non-self nodes in the feature space are selected to form a neighbor set $N_k(i)$, which takes the following form:

$$N_k(i) = \{j | j \in \underset{j \neq i}{\text{argTop}} - k(\|\tilde{x}_i - \tilde{x}_j\|_2)\} \quad (6)$$

Where \tilde{x}_i and \tilde{x}_j are the normalized node features. $\|\cdot\|_2$ denotes the Euclidean distance. And $\text{Top} - k$ denotes the selection of the k neighboring nodes with the smallest distance.

While constructing edges, a gaussian kernel function is used to determine the edge weights w_{ij} to enhance the model's ability to model neighbor similarity:

$$w_{ij} = \exp\left(-\frac{\|\tilde{x}_i - \tilde{x}_j\|_2}{2\beta^2}\right) \quad (7)$$

where β is the kernel function bandwidth coefficient, which controls the attenuation of adjacency strength. The construction of the time series graph structure using a sliding window and K-nearest neighbors is illustrated in Figure 2.

3.3 Spatio-temporal feature extraction based on TSGCN

Based on the KNN constructed dynamic graph described in Section 3.2, the spatio-temporal feature extraction stage aims to jointly model temporal evolution patterns and dependency relationships among time series segments during elevator door operation. In this work, each time series segment is treated as a graph node, while edges represent similarity based interactions captured by the KNN graph. Under this formulation, elevator door anomaly detection requires simultaneous characterization of temporal dynamics within each node and spatial dependencies across nodes.

To meet this requirement, a temporal spatial graph convolutional network (TSGCN) is employed as the core feature extraction model. The TSGCN integrates temporal modeling modules with graph convolution operations, enabling effective extraction of both time-domain characteristics and structural information embedded in the constructed graph. This design allows the model to capture localized abnormal patterns and their propagation across related segments, which are critical for accurate detection of elevator door anomalies.

3.3.1 Time dimension model. To extract short-term fluctuations and long-term structural features in time series data, a temporal modeling module consisting of CNN, BiLSTM and multi-head attention pooling is proposed. The module sequentially extracts local dynamic patterns by convolution, captures global temporal dependencies by BiLSTM, and weightedly aggregates key time steps by the multi-head attention mechanism, finally generating discriminative sequence-level representations.

This module captures abrupt changes, local patterns, or instantaneous pattern variations in signals within a short period, leveraging the local receptive field sliding window of convolutional neural network. It is well-suited for information with obvious local structures, such as vibration signals and image velocity curves. Two layers of 1-D convolution and pooling operations are used to compress the sequence length and enhance the extraction of local features.

Subsequently, the convolved sequence features are fed into a bidirectional LSTM network to integrate contextual information and establish long-term dependencies in the temporal dimension. Compared to a unidirectional LSTM, BiLSTM employs parallel computation with both forward and backward LSTM units, effectively merging preceding and succeeding information at each time step to capture the signal's global temporal structure more comprehensively.

In BiLSTM, the forward layer computes sequentially from $t = 1$ to T , storing the forward hidden state \vec{h}_t at each time step. Conversely, the backward layer computes in reverse order from $t = T$ to 1, recording the backward hidden state \overleftarrow{h}_t (Schuster and Paliwal, 1997). Finally, the hidden states from both directions are concatenated to form the output representation for the current time step, which is then used for subsequent decision-making or feature fusion.

The computational procedure can be formalized as:

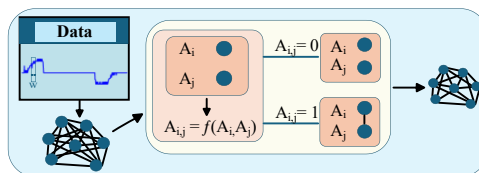


Figure 2. Construction of time series graph structure. **Source(s):** Figure created by authors

$$\begin{cases} \vec{h}_t = f(w_1 x_t + w_2 \vec{h}_{t-1}) \\ \overleftarrow{h}_t = f(w_3 x_t + w_4 \overleftarrow{h}_{t+1}) \\ y_t = g(w_5 \vec{h}_t + w_6 \overleftarrow{h}_t) \end{cases} \quad (8)$$

where x_t represents the input at time step t . $f(\cdot)$ is the activation function. $g(\cdot)$ denotes the output transformation, and w_i is the weight parameters.

To obtain a global representation of the entire sequence, a multi-attention pooling mechanism is introduced to replace the traditional average pooling or maximum pooling. This mechanism learns different attention weights, enabling the selection of the most discriminative time step information for the classification task. The computational process is as follows:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (9)$$

where d_k represents the feature dimension of the key/query in each attention head.

Multiple attention heads compute attention responses in parallel across different subspaces. These responses are then concatenated and projected to obtain a unified representation.

$$\begin{cases} head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \\ MultiHead(Q, K, V) = Concat(head_1, head_2, \dots, head_h)W_o \\ z_{time} = Mean(MultiHead(Q, K, V)) \in R^{B \times D} \end{cases} \quad (10)$$

The overall time dimension model is shown in Figure 3.

3.3.2 *Spatial dimension model.* Spatial dimension models aim to characterize the structural relationships between different nodes within a signal. Given the low dimensionality and limited expressive power of original node features, a linear mapping is employed to increase their dimensionality, as expressed below:

$$X^0 = \phi(X) = W_{proj} \cdot X + b \quad (11)$$

where $W_{proj} \in \mathbb{R}^{D_{hid} \times d_0}$ is the ascending weight matrix and b is the bias vector.

After that, the upscaled features are fed into the GCN to construct the structural relationships between nodes. The adjacency matrix of the graph is $A \in \mathbb{R}^{n \times n}$, where $A_{ij} = 1$ indicates that node i and node j are connected. A symmetric normalization method is used to construct the graph propagation matrix, expressed as follows (Seo et al., 2024):

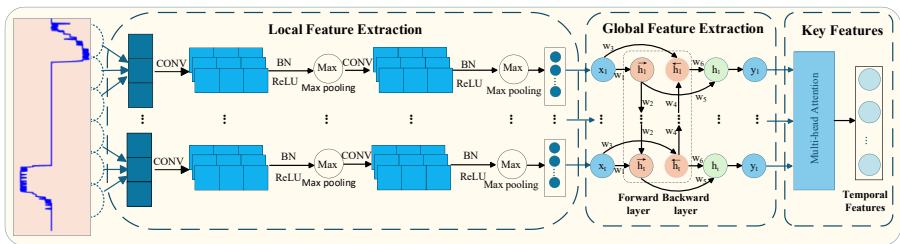


Figure 3. Time dimension model. Source(s): Figure created by authors

$$\begin{cases} \tilde{A} = A + I \\ \tilde{D}_{ii} = \sum_j \tilde{A}_{ij} \\ \hat{A} = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} \end{cases} \quad (12)$$

where I is the identity matrix and \tilde{D} is the corresponding degree matrix.

The propagation and updating process of the l th layer graph convolution is expressed as:

$$X^{(l+1)} = \sigma(\hat{A}X^{(l)}W^{(l)}) \quad (13)$$

where $X^{(l)} \in \mathbb{R}^{N \times d_l}$ denotes the node features of the l th layer. $W^{(l)} \in \mathbb{R}^{d_l \times d_{l+1}}$ is the learnable parameter of the layer, and $\sigma(\cdot)$ is the nonlinear activation function.

The overall spatial dimensional model is shown in Figure 4.

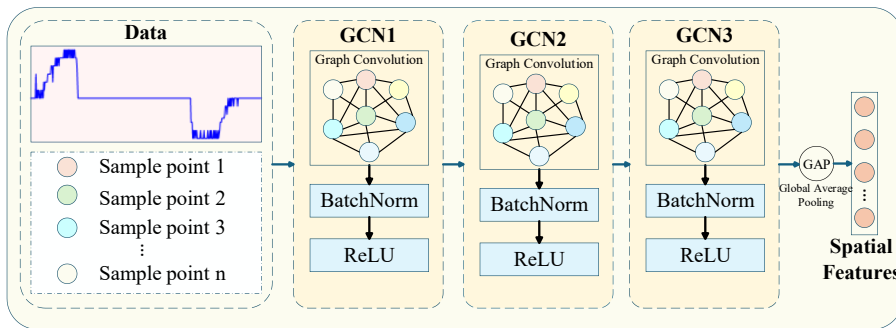


Figure 4. Spatial dimension model. Source(s): Figure created by authors

The pseudocode of TSGCN for anomaly detection is presented in Algorithm 1. During each epoch, the training process is executed, and subsequently, validation is performed. The system saves the model that achieves the best performance on the validation dataset. Upon reaching the maximum epoch, the optimal model is deployed to identify anomaly types within the test set.

Algorithm 1. Spatial-temporal information fusion for anomaly detection

Input: Multi-modal time series data: video sequence X_v , vibration sequence X_{vib} graph structure data: G_v, G_{vib}

Output: Anomaly category prediction results

1. Preprocessing stage:

2. Time series are extracted from video and sensors
3. KNN graph structures G_v, G_{vib} are constructed based on signal features
4. The dataset is divided into training, validation, and test sets

5. Training and validation stage:

6. Network parameters are randomly initialized
 7. Inputs: hidden layer dimension, learning rate lr , batch size B , number of training epochs N
 8. **for** i in training epochs do
 9. A batch of training data is sampled: $(X_v, X_{vib}, G_v, G_{vib})$
 10. Temporal features are extracted using BiLSTM
 11. Application of CNN to extract local temporal patterns
 12. Encoding time series using BiLSTM
 13. Implementation of multi-attention pooling
 14. BiLSTM features of video and vibration are concatenated
 15. Spatial features are extracted using GCN
 16. GCN modules are applied to G_v and G_{vib} respectively
 17. Graph features are extracted through global average pooling
 18. Graph convolutional features of video and vibration are concatenated
 19. Temporal and spatial features are fused
 20. BiLSTM and GCN features are concatenated
 21. The result is input to a fully connected layer classifier
 22. Calculate the crossentropy loss
 23. Backpropagate and update the parameters
 24. Evaluate model performance on a validation set
 25. **End**
 26. **Testing stage:**
 27. **Input:** test set $(X_v, X_{vib}, G_v, G_{vib})$
 28. **Output:** Anomaly type prediction results
- Source(s):** Algorithm created by authors

4. Experimental verification

To validate the effectiveness of the TSGCN-based anomaly detection method, experiments were conducted on the elevator door system dataset. To further evaluate its performance, we selected several classical anomaly detection methods as comparison baselines. Meanwhile, the influence of key parameters in TSGCN on the model performance is deeply analyzed.

4.1 Experimental data setup

4.1.1 *Data acquisition.* The experimental data were collected from elevator doors in normal operation within an apartment building. Faults were artificially introduced to obtain data representing different states, including Normal, Slowdown, Jamming, and Abnormal Door Closing. The elevator doors have a center-opening structure. Vibration data were collected by installing an attitude measurement sensor in the gap between the

landing door and the car door on one side of the car. This sensor integrates a three-axis accelerometer, a three-axis gyroscope, a three-axis angle sensor, and a three-axis magnetometer. The sensor supports data storage and export, with an output frequency ranging from 0.2 Hz to 200 Hz, and its sampling frequency is independently configurable. Video data, capturing the opening and closing process, were collected by deploying a camera in the corner of the car. The video frame rate was 30 fps, while the vibration sensor's sampling frequency was set to 50 Hz. Figure 5 shows the installation location of the acquisition equipment.

We simulated various states on the actual elevator door to obtain normal and abnormal data. Specifically, slowing down was simulated by inserting objects between the door and the column to increase the running friction. Abnormal door closing was simulated by placing obstacles between the two doors so that the door could not close. And jamming abnormal data were collected directly from the actual faulty elevator door.

4.1.2 Data processing. Each collected sample contains a complete door opening and closing operation. Vibration data were selected from the X, Y, and Z axes acceleration signals. The X-axis represents the lateral vibration of the elevator door, the Y-axis represents the vertical vibration, and the Z-axis represents the fore-and-aft vibration. The three-dimensional vibration signals is shown in Figure 6.

The process of video data processing is as follows. First, each video sample is framed, and the edge position of the door panel in each frame is extracted using edge detection algorithms. Then, the pixel displacement between consecutive frames is calculated and converted to physical speed using a calibrated scale factor. An illustrative example of this process is shown in Figure 7, where raw video frames are transformed into a door displacement curve and subsequently into a speed curve.

Based on the extracted door displacement information, the opening and closing running curves of the elevator door are obtained, as shown in Figure 8.

Due to the varying lengths of the collected sample sequences, sequence alignment and padding mechanisms are introduced into the batch processing stage to ensure uniform input dimensions. This leverages LSTM's capability to handle variable-length sequences. Specifically, the samples within each batch are first sorted in descending order according to their sequence lengths. Then, a uniform padding strategy extends all sequences to the length of the longest sequence in the current batch. The padding value is set to zero to avoid interfering with valid features. To preserve the original temporal structure, the actual length information is transmitted to the network as input during sequence packing, eliminating the impact of padding values on model training.



Figure 5. Device acquisition diagram. **Source(s):** Figure created by authors

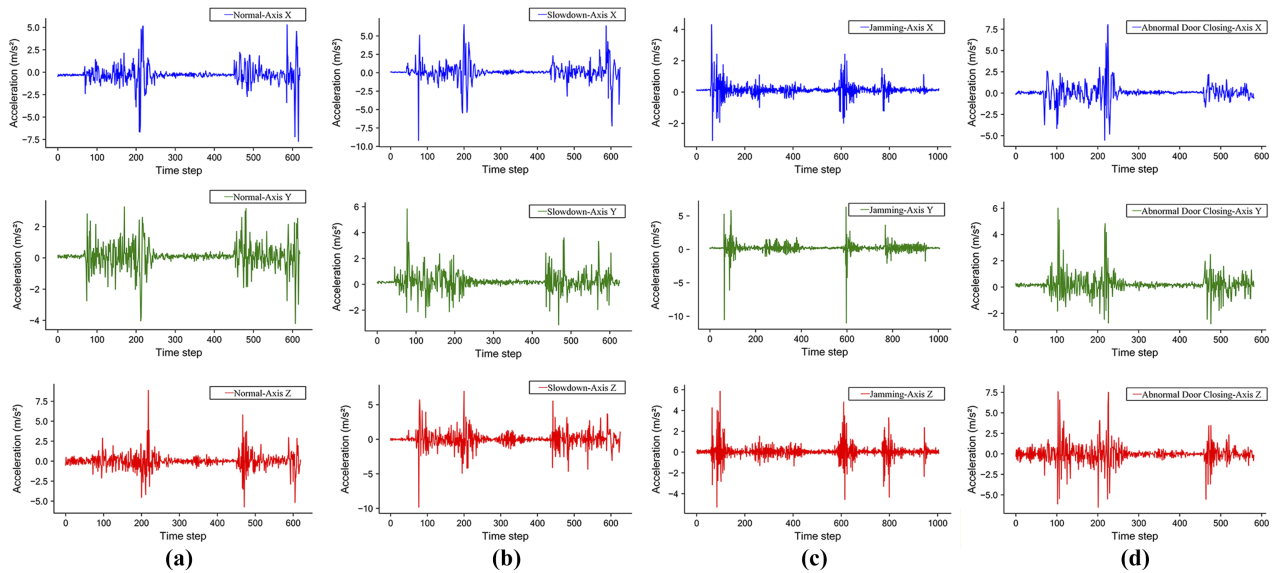


Figure 6. Three-dimensional vibration signals under different states. (a) Normal. (b) Slowdown. (c) Jamming. (d) Abnormal door closing. **Source(s):** Figure created by authors

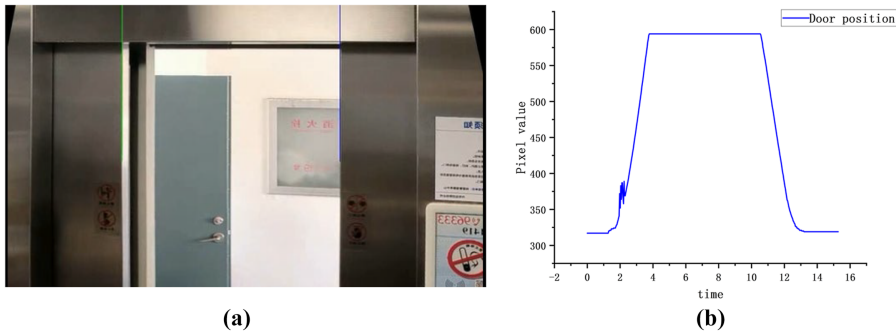


Figure 7. Video processing. (a) Edge position of the door. (b) Door displacement curve. **Source(s):** Figure created by authors

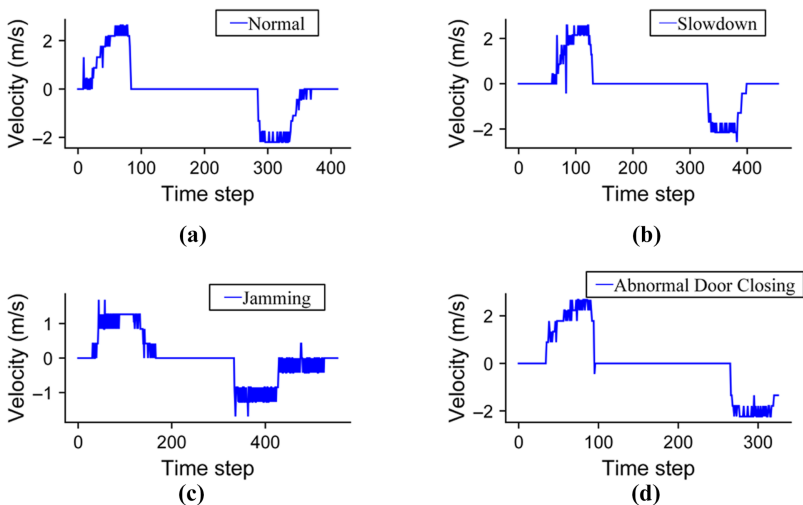


Figure 8. Elevator door operating curves under different states. (a) Normal. (b) Slowdown. (c) Jamming. (d) Abnormal door closing. **Source(s):** Figure created by authors

4.2 Experimental results

4.2.1 Evaluation indicators. The vibration and video datasets comprises 802 samples. In accordance with common practice, 60% of the data was randomly selected to form the training set, with 20% allocated to the validation set and 20% to the test set. The specific number of training, validation and test samples are given in [Table 1](#).

Accuracy and F1 score are used as core metrics to evaluate the model performance. Accuracy represents the proportion of samples correctly predicted by the model out of the total samples, reflecting the overall classification accuracy. The F1 score, integrating precision and recall, is suitable for evaluating performance with imbalanced data or in scenarios where misclassification occurs. The F1 score ranges from 0 to 1, with higher values indicating a better balance between precision and recall. The formulations of the evaluation metrics are as follows:

Table 1. Description of elevator door conditions

Condition	Label	Number of training/validation/testing samples
Normal	0	120/40/40
Slowdown	1	120/40/40
Jamming	2	120/41/40
Abnormal door closing	3	120/40/41

Source(s): Table created by authors

$$\left\{ \begin{array}{l} Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \\ Precision = \frac{TP}{TP + FP} \\ Recall = \frac{TP}{TP + FN} \\ F1 = \frac{2 * Precision * Recall}{Precision + Recall} \end{array} \right. \quad (14)$$

where TP , FP , TN , and FN represent the numbers of true positive, false positive, true negative, and false negative samples, respectively.

4.2.2 Network structure parameterization. To fully leverage multi-source data features for elevator door anomaly detection, we designed a spatiotemporal network architecture incorporating BiLSTM and GCN. The structure and training parameters of each submodule are detailed in [Table 2](#) and [Table 3](#).

Table 2. Structure of the network module

Module name	Functional	Network architecture
BiLSTM	Local feature extraction	Conv1d(k = 3, s = 1, p = 1), BatchNorm1d(64), ReLU() MaxPool1d() Conv1d(k = 3, s = 1, p = 1), BatchNorm1d(128), ReLU() MaxPool1d()
GCN	Contextual relationship Focus on important features Spatial feature	BiLSTM(hidden = 128(Video)/256(Vibrate)) Multi-head Attention(num_heads = 4) GCNConv(hidden = 128(Video)/256(Vibrate)), ReLU() GCNConv(), ReLU() GCNConv(), ReLU(), Global Average Pooling()
Characteristic fusion	Multisource spatial-temporal characterization	Concat() Linear() Dropout(0.3) Linear()

Source(s): Table created by authors

Table 3. Training configuration parameters in the network

Parameter	Set value
Optimizer	AdamW
Initial learning rate	1e-4
Weight decay	5e-2
Scheduling strategy	ReduceLROnPlateau
Loss function	CrossEntropy
Epoch	200
Batch size	16

Source(s): Table created by authors

4.2.3 Experimental results. To validate the proposed method, we conducted experiments on a dataset that we collected from real elevator operations, containing both video and vibration signals. The classification results of the method are shown in the confusion matrix in Figure 9. The performance of the test set as follows: out of 161 test samples, 155 fault categories were correctly detected, achieving a detection accuracy of 96.27%, which demonstrates high overall accuracy. Notably, the recognition rate for jamming faults reached 100%. However, misclassification occurred in the normal state (label 0), slowdown (label 1), and abnormal door closing (label 3). Further analysis revealed that while these three states are distinct, they were all artificially simulated on the same elevator. This may have led to pattern overlap in the sensor signals, consequently hindering the model’s ability to effectively discriminate features. Subsequent research will focus on introducing more types of sensor data to improve the differentiation ability of similar faults.

The features are further visualized using the UMPA method and the results are shown in Figure 10. As the figure shows, the feature distributions for “Normal” and “Slowdown” labels are highly similar and difficult to distinguish accurately, consistent with the confusion matrix results.

4.3 Further discussion

4.3.1 Hyperparameter sensitivity analysis. In order to explore the influence of hyperparameters on model performance, we conducted a hyperparameter sensitivity analysis to determine appropriate values. During KNN graph construction, the number of

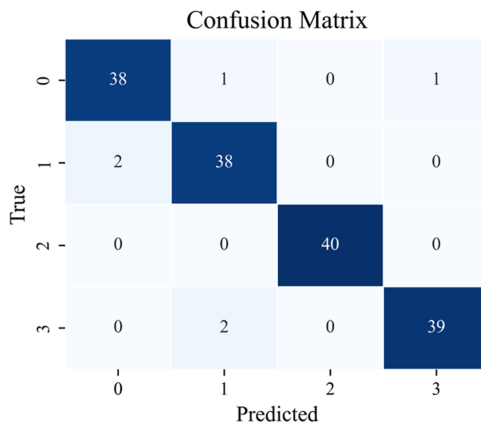


Figure 9. Confusion matrix result. **Source(s):** Figure created by authors

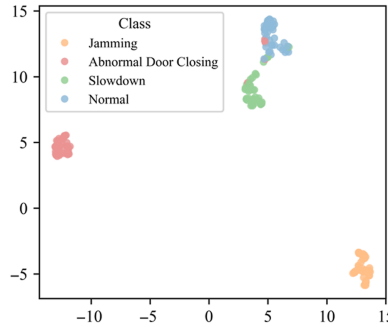


Figure 10. UMPA visualization result. **Source(s):** Figure created by authors

node neighbors (K value) and the distance metric significantly influence the classification results. A K value that is too small results in a sparse graph structure, which, while beneficial for capturing local key point dependencies, may overlook long-range related node information, leading to insufficient information propagation. Conversely, a K value that is too large creates a dense graph structure and increases the information propagation path. This can alleviate the locality limitation of the adjacency structure but may introduce redundant or noisy connections, resulting in over-smoothing or overfitting.

In this study, we consider several commonly used distance metrics to explore their impact on temporal-spatial graph construction, including Euclidean, Manhattan, Chebyshev, and Minkowski distances. Euclidean distance measures the straight-line distance between two points in a multidimensional space and aligns well with the natural continuity of time series data. Manhattan distance, defined as the sum of absolute axial differences, emphasizes edge features and is more sensitive to small perturbations. Chebyshev distance focuses on the maximum coordinate difference, making it suitable for detecting extreme change points. Minkowski distance introduces a tunable parameter p , which reduces to Manhattan distance when $p = 1$, Euclidean distance when $p = 2$, and Chebyshev distance as $p \rightarrow \infty$. Therefore, we specifically analyze the effects of Euclidean, Manhattan, and Chebyshev distances on the experimental results. Based on comparative evaluation, Euclidean distance is selected as the primary metric for subsequent graph construction due to its superior performance and consistency with temporal smoothness.

In addition, a single-factor iterative experiment was conducted on the learning rate and dropout using the controlled variable method to analyze the model's sensitivity to key hyperparameters. Only the target hyperparameters are adjusted in each round of experiments, and the rest of the parameters are kept unchanged. Specifically, the settings were as follows: learning rate $lr \in [10^{-5}, 5 \times 10^{-5}, 10^{-4}, 5 \times 10^{-4}, 10^{-3}]$, and dropout $\in [0.1, 0.2, 0.3, 0.4, 0.5]$.

The model's performance on the validation set was recorded for each set of parameter configurations, and the results are shown in Figure 11. The optimal hyperparameters selected in the experiments are: $K = 5$, $lr = 10^{-4}$, $dropout = 0.3$, and Euclidean distance for the distance metric. As shown in Figure 11(a), when $K < 5$, the graph structure is too sparse due to the insufficient number of neighbors, failing to fully capture the long-range dependencies between nodes. Consequently, information propagation is insufficient, leading to low accuracy. When $K = 6 \sim 7$, the graph structure becomes dense due to the excessive K , introducing a large number of redundant and noisy connections, which results in the over-averaging of node features and the destruction of local key information. When $K > 7$, the model partially alleviates the locality limitation by increasing the number of neighbors, and the accuracy improves. For the distance metric in Figure 11(b), Manhattan distance is more sensitive to noise, resulting in a slight decrease in accuracy, while Chebyshev distance

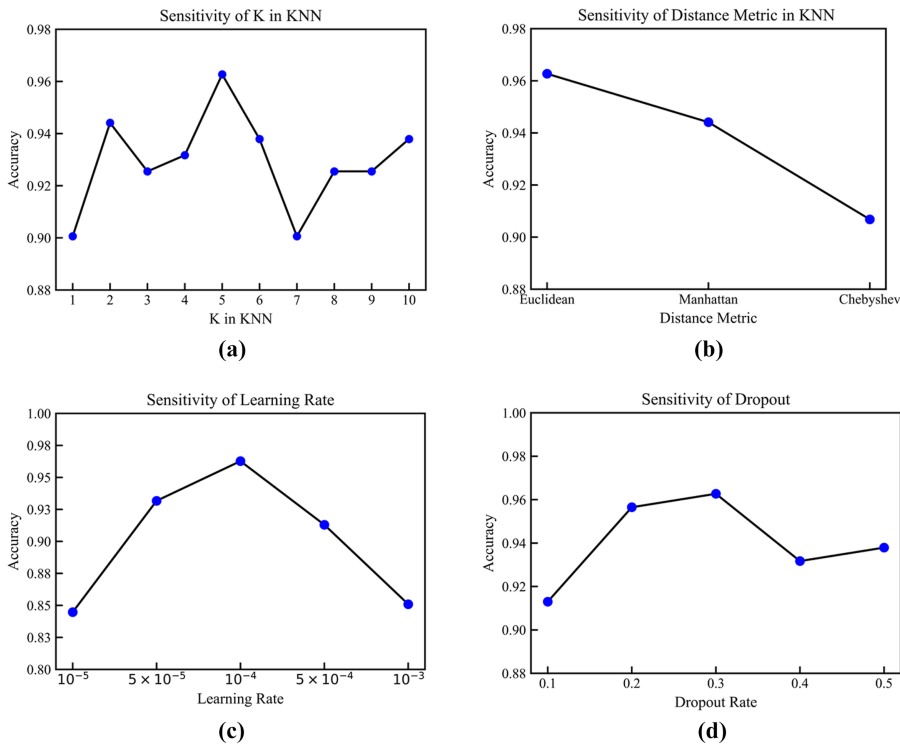


Figure 11. Accuracy comparison of different hyperparameters. (a) Sensitivity of K in KNN. (b) Effect of distance metric on KNN. (c) Sensitivity of learning rate. (d) Sensitivity of dropout. **Source(s):** Figure created by authors

performs poorly because it ignores temporal continuity. Regarding the learning rate parameter in Figure 11(c), values that are too small or too large will hinder model performance. A smaller learning rate leads to slow convergence, while a larger one causes the parameters to update excessively, making the model skip the optimal solution and resulting in unstable performance. As for the dropout in Figure 11(d), an excessively high dropout rate causes the model to underfit, leading to a significant loss of information during the entire training process, while a rate that is too low may not be sufficient to prevent overfitting.

4.3.2 Comparison with common single-sensor anomaly detection methods. Currently, most anomaly detection methods rely on single-type sensor signals, especially vibration signals, to determine the operating status of a system. To verify the effectiveness of the proposed multi-source fusion anomaly detection method, we selected single-sensor modeling methods such as LiConvFormer (Yan et al., 2024), TCN (Zhang et al., 2022b), CNN (Liu et al., 2019), BiLSTM (Abebe et al., 2024), and GCN (Chen et al., 2022) as comparison baselines. Each model was trained and evaluated under the same data conditions, and the performance comparison results are shown in Figure 12.

The numerical results are shown in Table 4. We can find that the proposed model outperforms other methods in all four metrics. It is worth noting that the proposed scheme also contains BiLSTM and GCN models. As shown in Table 4, the BiLSTM and GCN models yield lower detection results for both video and vibration data compared to the proposed TSGCN method. Since the BiLSTM model only analyzes the temporal dependency, it ignores the spatial structure information contained in the time series data. Conversely, the GCN focuses

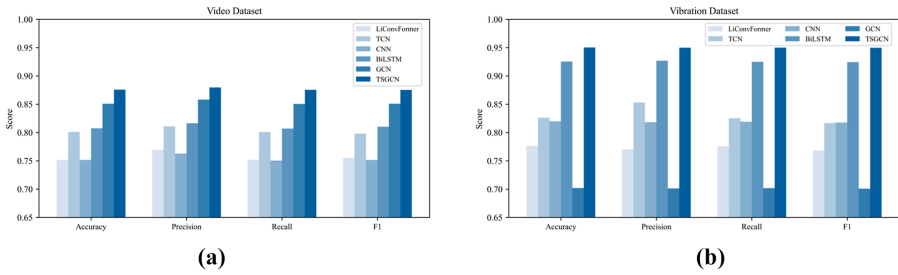


Figure 12. Performance comparison of anomaly detection models under single-sensor conditions. (a) Video data performance comparison. (b) Vibration data performance comparison. **Source(s):** Figure created by authors

Table 4. Performance comparison of anomaly detection models under single sensor conditions

Methods	Results of the comparison of the two datasets											
	Video data		Vibration data		Accuracy		Precision		Recall		F1	
LiConvFormer	75.16%	76.95%	75.17%	75.50%	77.64%	77.01%	77.56%	76.82%				
TCN	80.12%	81.09%	80.09%	79.80%	82.61%	85.29%	82.53%	81.68%				
CNN	75.16%	76.28%	75.06%	75.16%	81.99%	81.83%	81.91%	81.76%				
BiLSTM	80.75%	81.65%	80.70%	81.02%	92.55%	92.69%	92.50%	92.44%				
GCN	85.09%	85.81%	85.06%	85.09%	70.19%	70.12%	70.17%	70.08%				
TSGCN	87.58%	87.97%	87.55%	87.53%	95.03%	94.99%	95.00%	94.97%				

Source(s): Table created by authors

solely on spatial structure information, lacking the ability to capture sequential dependencies between data points. In contrast to traditional schemes, the proposed TSGCN leverages a wider range of information and can selectively prioritize different data sources.

4.3.3 Comparison with non-temporal modeling approaches. To further validate the effectiveness of the spatial-temporal fusion model, we compared the proposed multi-source spatiotemporal graph method with non-spatiotemporal methods that disregard temporal dependencies or spatial structures. These methods typically rely on statistical feature extraction or simple network structures and perform holistic encoding of input signals. They often fail to consider the dynamic evolution of time series or the structural relationships between sequences, which makes it difficult for them to capture key features of fault evolution. The comparative methods include Multilayer Perceptron (MLP) that directly classifies flattened input signals (Rawat et al., 2018), xLSTM enhancing temporal modeling of long sequences with exponential gating and modified memory structure (Beck et al., 2024), mixCNN extracting richer spatial features through a hybrid convolution design with residual connections (Zhao and Jiao, 2023), and ResCISTA-Net extending CISTA by adding residual blocks for better feature extraction (Rao et al., 2024). Table 5 presents their performance comparison.

As shown in Table 5, the TSGCN model achieves the best performance in all four metrics and far outperforms the other compared methods. The MLP model’s performance is the poorest, as it completely disregards the signal’s time-series structure and spatial correlations. While ResCISTA-Net leverages residual blocks to improve low-level feature extraction, it also neglects temporal and spatial structures, hindering its ability to effectively identify complex fault patterns. xLSTM focuses solely on enhancing the capture of long-sequence time dependencies, without considering the spatial arrangement of data points. On the contrary,

Table 5. Comparison of the performance of fusion models and non-spatiotemporal methods in anomaly detection

Methods	Accuracy	Precision	Recall	F1
MLP	64.60%	64.65%	64.70%	64.08%
xLSTM	77.64%	78.88%	77.56%	77.80%
mixCNN	74.53%	74.07%	74.44%	73.73%
ResCISTA-Net	65.84%	65.43%	65.81%	65.50%
TSGCN	96.27%	96.30%	96.28%	96.28%

Source(s): Table created by authors

mixCNN only extracts spatial features and does not introduce dynamic evolution of time series. Both of them fail to effectively fuse the spatial-temporal synergy information of time series signals, which leads to poor extraction of low-resolution anomaly categories.

4.3.4 Comparison with state-of-the-art spatio-temporal models. To provide a comprehensive comparison with existing spatio-temporal modeling approaches, this paper conducts experiments against several representative advanced methods, including MTGNN (Wu *et al.*, 2020), ASTGNN (Guo *et al.*, 2022), STFGNN (Li and Zhu, 2020) and STSGCN (Sofianos *et al.*, 2021). The above models are all representative graph neural network methods in the fields of multivariate time series modeling and spatio-temporal feature learning in recent years, capable of modeling temporal dependencies and structural correlation information from different perspectives.

Among them, MTGNN adaptively learns the graph structure through the graph learning layer and combines temporal convolution for spatio-temporal modeling, making it an advanced model for multivariate time series prediction. ASTGNN introduces an attention mechanism in the convolution of spatio-temporal graphs to dynamically learn the importance weights between different time steps and spatial nodes. STFGNN designed a graph structure that integrates spatio-temporal information and utilized the parallel GCN module to extract spatio-temporal features respectively. STSGCN constructs local spatiotemporal maps and simultaneously captures local spatiotemporal correlations by using dedicated convolutional modules. All models adopted the same data preprocessing procedures and input features as in this study, and the model hyperparameters were all tuned to achieve their best performance. The comparison results are shown in Table 6.

As can be seen from Table 6, the TSGCN model we proposed significantly outperforms the other four advanced spatiotemporal graph models in all evaluation indicators. This performance difference indicates that models relying on global or fixed graph structures have limitations in capturing local and transient abnormal patterns existing in the operation of elevator doors. TSGCN can more effectively represent local spatio-temporal patterns by constructing sampler k-nearest neighbor graphs and jointly modeling spatio-temporal dynamic dependencies, thereby significantly improving the performance of anomaly detection.

Table 6. Performance comparison of state-of-the-art spatio-temporal models for anomaly detection

Methods	Accuracy	Precision	Recall	F1
MTGNN	77.02%	78.03%	77.02%	76.13%
ASTGNN	78.88%	79.30%	78.88%	78.97%
STFGNN	77.02%	80.22%	77.02%	75.66%
STSGCN	66.46%	66.39%	66.46%	66.13%
TSGCN	96.27%	96.30%	96.28%	96.28%

Source(s): Table created by authors

5. Conclusion

This study proposes a novel multi-source spatial-temporal information fusion model for the accurate recognition of elevator door operation states and anomaly detection. Advanced feature engineering techniques are employed across both temporal and spatial domains to comprehensively capture the temporal dynamics of sensor signals as well as their latent structural correlations. Using a dataset gathered through our own data acquisition of elevator door operations, we conducted experiments to systematically analyze the effects of hyperparameters, including the K value and distance metric used in graph construction, on model performance. Furthermore, the proposed method was compared with traditional single-sensor models, methods lacking spatiotemporal modeling, and several representative state-of-the-art spatiotemporal graph based models. The experimental results demonstrate that the proposed multi-source spatial-temporal fusion model outperforms the comparative methods in accuracy and F1 score, validating the effectiveness and advantages of fusing spatial-temporal structures for complex state recognition. In summary, the spatial-temporal graph neural network-based anomaly detection model for elevator door systems, as developed in this paper, exhibits promising performance and application potential. This is achieved through the integration of multi-source information, spatiotemporal dependencies, and graph structure modeling, which enables more effective characterization of localized abnormal patterns compared with existing spatiotemporal approaches. It offers both theoretical support and a methodological framework for multi-modal abnormal state recognition within elevator systems. Future research will focus on further optimizing the model architecture, enhancing its ability to identify abnormal states in scenarios with limited samples, and exploring the incorporation of a wider range of sensor data to improve the model's generalization and robustness.

References

- Abebe, M., Kim, S.Y., Koo, B. and Jeong, H.-S. (2024), "Adaptive signal fusion for swashplate pump fault detection using bidirectional long short-term memory and wavelet scattering transform", *Engineering Applications of Artificial Intelligence*, Vol. 138, 109375, doi: [10.1016/j.engappai.2024.109375](https://doi.org/10.1016/j.engappai.2024.109375).
- Allen, L., Lu, H. and Cordiner, J. (2024), "Knowledge-Enhanced spatiotemporal analysis for anomaly detection in process manufacturing", *Computers in Industry*, Vol. 161, 104111, doi: [10.1016/j.compind.2024.104111](https://doi.org/10.1016/j.compind.2024.104111).
- An, Z., Bai, D., Huang, Y., Ning, W., Deng, Y., Gan, N. and Liu, S. (2021), "Building elevator safety monitoring system based on the BIM technology", *Journal of Physics: Conference Series*, Vol. 1939 No. 1, 012026, doi: [10.1088/1742-6596/1939/1/012026](https://doi.org/10.1088/1742-6596/1939/1/012026).
- Beck, M., Pöppel, K., Spanring, M., Auer, A., Prudnikova, O., Kopp, M., Klambauer, G., Brandstetter, J. and Sepp Hochreiter, S. (2024), "xLSTM: extended long short-term memory", *Advances in Neural Information Processing Systems*, arXiv, Vol. 37, pp. 107547-107603, doi: [10.52202/079017-3417](https://doi.org/10.52202/079017-3417).
- Chen, Z., Xu, J., Peng, T. and Yang, C. (2022), "Graph convolutional network-based method for fault diagnosis using a hybrid of measurement and prior knowledge", *IEEE Transactions on Cybernetics*, Vol. 52 No. 9, pp. 9157-9169, doi: [10.1109/TCYB.2021.3059002](https://doi.org/10.1109/TCYB.2021.3059002).
- Esteban, E., Salgado, O., Iturrospe, A. and Isasa, I. (2016), "Model-based approach for elevator performance estimation", *Mechanical Systems and Signal Processing*, Vols 68-69, pp. 125-137, doi: [10.1016/j.ymssp.2015.07.005](https://doi.org/10.1016/j.ymssp.2015.07.005).
- Fathizadan, S., Ju, F., Lu, Y. and Yang, Z. (2024), "Deep spatio-temporal anomaly detection in laser powder bed fusion", *IEEE Transactions on Automation Science and Engineering*, Vol. 21 No. 4, pp. 5227-5239, doi: [10.1109/TASE.2023.3309927](https://doi.org/10.1109/TASE.2023.3309927).
- Feng, T., Guo, L., Gao, H. and Liu, X. (2025a), "A multisource state space-based tool remaining useful life prediction method considering multistage degradation characteristics", *IEEE Sensors Journal*, Vol. 25 No. 7, pp. 11216-11225, doi: [10.1109/JSEN.2025.3540781](https://doi.org/10.1109/JSEN.2025.3540781).

- Feng, L., Ding, Z., Yin, Y., Wang, Y., Zhang, Q., Liu, X., Yuan, Z. and Li, H. (2025b), "Scraper conveyor gearbox fault diagnosis based on multi-source heterogeneous data fusion", *Measurement*, Vol. 247, 116797, doi: [10.1016/j.measurement.2025.116797](https://doi.org/10.1016/j.measurement.2025.116797).
- Guo, S., Lin, Y., Wan, H., Li, X. and Cong, G. (2022), "Learning dynamics and heterogeneity of spatial-temporal graph data for traffic forecasting", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 34 No. 11, pp. 5415-5428, doi: [10.1109/TKDE.2021.3056502](https://doi.org/10.1109/TKDE.2021.3056502).
- Guo, L., Niu, D., Zhao, J. and Jia, M. (2024a), "Operation condition assessment for elevators based on deep siamese network and t-S semi-supervision model", *IEEE Transactions on Instrumentation and Measurement*, Vol. 73, pp. 1-13, doi: [10.1109/tim.2024.3373804](https://doi.org/10.1109/tim.2024.3373804).
- Guo, L., Gu, X., Yu, Y., Duan, A. and Gao, H. (2024b), "An analysis method for interpretability of convolutional neural network in bearing fault diagnosis", *IEEE Transactions on Instrumentation and Measurement*, Vol. 73, pp. 1-12, doi: [10.1109/TIM.2023.3334350](https://doi.org/10.1109/TIM.2023.3334350).
- Han, P., Huang, Z., Li, W., He, W. and Cao, Y. (2025), "Multi-sensor bearing fault diagnosis based on evidential neural network with sensor weights and reliability", *Expert Systems with Applications*, Vol. 269, 126533, doi: [10.1016/j.eswa.2025.126533](https://doi.org/10.1016/j.eswa.2025.126533).
- Hsu, C.-Y., Qiao, Y., Wang, C. and Chen, S.-T. (2020), "Machine learning modeling for failure detection of elevator doors by three-dimensional video monitoring", *IEEE Access*, Vol. 8, pp. 211595-211609, doi: [10.1109/ACCESS.2020.3037185](https://doi.org/10.1109/ACCESS.2020.3037185).
- Kipf, T.N. and Welling, M. (2016), "Semi-supervised classification with graph convolutional networks", arXiv, doi: [10.48550/arXiv.1609.02907](https://doi.org/10.48550/arXiv.1609.02907).
- Lan, S., Jiang, S., Qiu, J., Wan, Z., Chen, L., Li, G. and Alam, J. (2021), "Statistical analysis of typical elevator accidents in China from 2002 to 2019", *Applied Mathematics and Nonlinear Sciences*, Vol. 6 No. 2, pp. 193-208, doi: [10.2478/amns.2021.2.00003](https://doi.org/10.2478/amns.2021.2.00003).
- Li, M. and Zhu, Z. (2020), "Spatial-temporal fusion graph neural networks for traffic flow forecasting", *In Proceedings of the AAAI Conference on Artificial Intelligence*, arXiv, Vol. 35 No. 5, pp. 4189-4196, doi: [10.1609/aaai.v35i5.16542](https://doi.org/10.1609/aaai.v35i5.16542).
- Li, Q., Han, Z. and Wu, X. (2018), "Deeper insights into graph convolutional networks for semi-supervised learning", *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32 No. 1, doi: [10.1609/aaai.v32i1.11604](https://doi.org/10.1609/aaai.v32i1.11604).
- Liu, X., Zhou, Q., Zhao, J., Shen, H. and Xiong, X. (2019), "Fault diagnosis of rotating machinery under noisy environment conditions based on a 1-D convolutional autoencoder and 1-D convolutional neural network", *Sensors, MDPI AG*, Vol. 19 No. 4, p. 972, doi: [10.3390/s19040972](https://doi.org/10.3390/s19040972).
- Lu, J., Zhang, W., Lu, C., Xiao, G. and Wang, Q. (2025), "A multi-scale convolution capsule network with data augmentation and attention mechanisms for elevator fault diagnosis", *ISA Transactions*, Vol. 167, pp. 1873-1887, doi: [10.1016/j.isatra.2025.09.041](https://doi.org/10.1016/j.isatra.2025.09.041).
- Lv, J., Kim, B.-G., Parameshachari, B.D., Slowik, A. and Li, K. (2025), "Large model-driven hyperscale healthcare data fusion analysis in complex multi-sensors", *Information Fusion*, Vol. 115, 102780, doi: [10.1016/j.inffus.2024.102780](https://doi.org/10.1016/j.inffus.2024.102780).
- Niu, D., Yang, M., Jia, M., Jin, H. and Luo, G. (2025), "Performance evaluation of elevators using a novel hierarchical softmax regression model", *Mechanical Systems and Signal Processing*, Vol. 228, 112429, doi: [10.1016/j.ymssp.2025.112429](https://doi.org/10.1016/j.ymssp.2025.112429).
- Pan, W., Xiang, Y., Gong, W. and Shen, H. (2023), "Risk evaluation of elevators based on fuzzy theory and machine learning algorithms", *Mathematics*, Vol. 12 No. 1, p. 113, doi: [10.3390/math12010113](https://doi.org/10.3390/math12010113).
- Pan, J., Shao, C., Dai, Y., Wei, Y., Chen, W. and Lin, Z. (2024), "Research on fault prediction method of elevator door system based on transfer learning", *Sensors*, Vol. 24 No. 7, p. 2135, doi: [10.3390/s24072135](https://doi.org/10.3390/s24072135).
- Qi, C., Zhang, J., Jia, H., Mao, Q., Wang, L. and Song, H. (2021), "Deep face clustering using residual graph convolutional network", *Knowledge-Based Systems*, Vol. 211, 106561, doi: [10.1016/j.knsys.2020.106561](https://doi.org/10.1016/j.knsys.2020.106561).

- Rao, F., Zeng, M. and Cheng, Y. (2024), "A novel interpretable model via algorithm unrolling for intelligent fault diagnosis of machinery", *IEEE Sensors Journal*, Vol. 24 No. 1, pp. 495-505, doi: [10.1109/jsen.2023.3332755](https://doi.org/10.1109/jsen.2023.3332755).
- Rawat, A.S., Rana, A., Kumar, A. and Bagwari, A. (2018), "Application of multi layer artificial neural network in the diagnosis system: a systematic review", *IAES International Journal of Artificial Intelligence (IJ-AI)*, Institute of Advanced Engineering and Science, Vol. 7 No. 3, p. 138, doi: [10.11591/ijai.v7.i3.pp138-142](https://doi.org/10.11591/ijai.v7.i3.pp138-142).
- Schuster, M. and Paliwal, K.K. (1997), "Bidirectional recurrent neural networks", *IEEE Transactions on Signal Processing*, Vol. 45 No. 11, pp. 2673-2681, doi: [10.1109/78.650093](https://doi.org/10.1109/78.650093).
- Seo, J., Noh, Y., Kang, Y.-J., Lim, J., Ahn, S., Song, I. and Kim, K.C. (2024), "Graph neural networks for anomaly detection and diagnosis in hydrogen extraction systems", *Engineering Applications of Artificial Intelligence*, Vol. 135, 108846, doi: [10.1016/j.engappai.2024.108846](https://doi.org/10.1016/j.engappai.2024.108846).
- Sofianos, T., Sampieri, A., Franco, L. and Galasso, F. (2021), "Space-time-Separable graph convolutional network for pose forecasting", *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Presented at the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, IEEE, pp. 11189-11198, doi: [10.1109/ICCV48922.2021.01102](https://doi.org/10.1109/ICCV48922.2021.01102).
- Sun, K. and Yin, A. (2025), "Multi-sensor temporal-spatial graph network fusion empirical mode decomposition convolution for machine fault diagnosis", *Information Fusion*, Vol. 114, 102708, doi: [10.1016/j.inffus.2024.102708](https://doi.org/10.1016/j.inffus.2024.102708).
- Wang, Q., Leng, Y., Li, D., Zhang, X., Li, R., Zhu, H. and Zhang, H. (2018), "MCU system-based intelligent high-speed elevator door operator fault analysis and research", *IOP Conference Series: Materials Science and Engineering*, Vol. 428, 012028, doi: [10.1088/1757-899x/428/1/012028](https://doi.org/10.1088/1757-899x/428/1/012028).
- Wang, Q., Chen, L., Xiao, G., Wang, P., Gu, Y. and Lu, J. (2024), "Elevator fault diagnosis based on digital twin and PINNs-e-RGCN", *Scientific Reports*, Vol. 14 No. 1, 30713, doi: [10.1038/s41598-024-78784-7](https://doi.org/10.1038/s41598-024-78784-7).
- Wang, Q., Yin, C., She, K., Tong, Q., Lu, G., Zhang, H. and Lu, J. (2025), "Bearing fault diagnosis for variable operating conditions based on KAN convolution and dual branch fusion attention", *Scientific Reports*, Vol. 15 No. 1, 21442, doi: [10.1038/s41598-025-04620-1](https://doi.org/10.1038/s41598-025-04620-1).
- Wu, Z., Pan, S., Long, G., Jiang, J., Chang, X. and Zhang, C. (2020), "Connecting the dots: multivariate time series forecasting with graph neural networks", *In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, arXiv, pp. 753-763, doi: [10.1145/3394486.3403118](https://doi.org/10.1145/3394486.3403118).
- Wu, W., Tan, C., Zhang, S. and Dong, F. (2024), "Joint mining of fluid knowledge and multi-sensor data for gas-water two-phase flow status monitoring and evolution analysis", *Advanced Engineering Informatics*, Vol. 62, 102687, doi: [10.1016/j.aei.2024.102687](https://doi.org/10.1016/j.aei.2024.102687).
- Xie, P., Zhang, L., Li, M. and Qiu, C. (2024), "An elevator door anomaly detection method based on improved deep multi-sphere support vector data description", *Computers and Electrical Engineering*, Vol. 120, 109660, doi: [10.1016/j.compeleceng.2024.109660](https://doi.org/10.1016/j.compeleceng.2024.109660).
- Xiao, G., Yao, J., Zhong, L., Xiao, Z. and Lu, J. (2025a), "MB-ViT: MBConv vision transformer with time-frequency feature fusion for bearing fault diagnosis", *Neural Computing and Applications*, Vol. 37 No. 27, pp. 22801-22825, doi: [10.1007/s00521-025-11509-7](https://doi.org/10.1007/s00521-025-11509-7).
- Xiao, H., Dornaika, F., Charafeddine, J. and Bi, J. (2025b), "Metric learning-enhanced semi-supervised graph convolutional network for multi-view learning", *Information Fusion*, Vol. 124, 103420, doi: [10.1016/j.inffus.2025.103420](https://doi.org/10.1016/j.inffus.2025.103420).
- Xu, W. and Li, Y. (2025), "Enhancing information fusion and feature selection efficiency via the PROMETHEE method for multi-source dynamic decision data sets", *Knowledge-Based Systems*, Vol. 309, 112781, doi: [10.1016/j.knosys.2024.112781](https://doi.org/10.1016/j.knosys.2024.112781).
- Yan, S., Shao, H., Wang, J., Zheng, X. and Liu, B. (2024), "LiConvFormer: a lightweight fault diagnosis framework using separable multiscale convolution and broadcast self-attention", *Expert Systems with Applications*, Vol. 237, 121338, doi: [10.1016/j.eswa.2023.121338](https://doi.org/10.1016/j.eswa.2023.121338).

- Yang, G., Tao, H., Du, R. and Zhong, Y. (2023a), "Compound Fault diagnosis of harmonic drives using deep capsule graph convolutional network", *IEEE Transactions on Industrial Electronics*, Vol. 70 No. 4, pp. 4186-4195, doi: [10.1109/tie.2022.3176280](https://doi.org/10.1109/tie.2022.3176280).
- Yang, K., Ding, Y., Geng, F., Jiang, H. and Zou, Z. (2023b), "A multi-sensor mapping Bi-LSTM model of bridge monitoring data based on spatial-temporal attention mechanism", *Measurement*, Vol. 217, 113053, doi: [10.1016/j.measurement.2023.113053](https://doi.org/10.1016/j.measurement.2023.113053).
- Zhang, K., Gao, T. and Shi, H. (2022a), "Bearing fault diagnosis method based on multi-source heterogeneous information fusion", *Measurement Science and Technology*, Vol. 33 No. 7, 075901, doi: [10.1088/1361-6501/ac5deb](https://doi.org/10.1088/1361-6501/ac5deb).
- Zhang, H., Ge, B. and Han, B. (2022b), "Real-time motor fault diagnosis based on TCN and attention", *Machines*, Vol. 10 No. 4, p. 249, MDPI AG, doi: [10.3390/machines10040249](https://doi.org/10.3390/machines10040249).
- Zhao, Z. and Jiao, Y. (2023), "A fault diagnosis method for rotating machinery based on CNN with mixed information", *IEEE Transactions on Industrial Informatics*, Vol. 19 No. 8, pp. 9091-9101, doi: [10.1109/tii.2022.3224979](https://doi.org/10.1109/tii.2022.3224979).
- Zhou, L. and Wang, H. (2024), "MST-GAT: a multi-perspective spatial-temporal graph attention network for multi-sensor equipment remaining useful life prediction", *Information Fusion*, Vol. 110, 102462, doi: [10.1016/j.inffus.2024.102462](https://doi.org/10.1016/j.inffus.2024.102462).

Corresponding author

Dapeng Niu can be contacted at: niudapeng@neu.edu.cn