International Journal of Computer Science and Information Technology
ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 4, Number 2, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v4n2.25
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Research on Multi-Sensor Fusion Fault Diagnosis Method Based on Spatiotemporal Attention Mechanism

Tianrui Chu [1, 2, *], Zhixuan Wang [1, 2],

[1] College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou Gansu 730050, China

[2] Key Laboratory of Gansu Advanced Control for Industrial Processes, Lanzhou Gansu 730050, China

*Corresponding Author: ws078118@163.com

ABSTRACT

As industrial systems become increasingly complex, real-time monitoring and intelligent management of industrial equipment have become imperative. However, the limitations in coverage and accuracy of single sensors make it challenging to comprehensively characterize the operational state of equipment, leading to reduced system reliability and increased pressures on data transmission and storage. To address these challenges, this study presents a novel fault diagnosis method based on multi-sensor fusion using a spatio-temporal attention mechanism. Initially, one-dimensional convolutional neural networks (1D-CNN) are employed to extract features from raw signals, effectively capturing local characteristics and ensuring the integrity and validity of fault signals. Subsequently, the spatiotemporal attention mechanism adjusts the feature weights based on the temporal and spatial correlations of different sensors, as well as their respective importance, thereby capturing the spatio-temporal dependencies across multiple sensors and enhancing the efficacy of information fusion. Finally, the proposed method is validated through experiments on a nickel flash smelting furnace system. The results demonstrate that the method achieves a fault diagnosis accuracy exceeding 97.78%, significantly enhancing fault detection and decision-making performance.

KEYWORDS

Multi-sensor fusion; Fault diagnosis; One-dimensional convolutional neural network; Spatio-temporal attention mechanism

1. INTRODUCTION

As modern industrial systems become increasingly complex, operational safety has become an intrinsic performance metric of control systems. Any potential risks or faults in industrial equipment can result in significant economic losses and even pose threats to human health and safety. Therefore, timely and accurate fault diagnosis is critical for ensuring the reliable and efficient operation of industrial systems.

As the fundamental components for information acquisition, sensors enable the monitoring of the operational status of industrial equipment. However, previous studies have predominantly relied on data from single sensors for condition monitoring. This approach faces limitations in terms of coverage and installation location, making it challenging to capture comprehensive information from complex industrial systems. Single-sensor data can lead to misjudgments regarding the health status of industrial equipment if the sensor itself malfunctions. Moreover, environmental conditions and operational variations can significantly affect the measurements collected by a single sensor, resulting in inaccurate fault diagnosis. To address these challenges, the focus of research has shifted toward fault diagnosis based on multi-sensor data fusion. By integrating multiple sensors, a broader range of fault information can be captured. Thus, the development and exploration of multi-sensor data fusion techniques for fault diagnosis are of great significance for ensuring the efficient and healthy operation of industrial equipment.

Currently, fault diagnosis techniques based on multi-sensor information fusion can be broadly categorized into three types: data-level fusion, feature-level fusion, and decision-level fusion. Data-level fusion integrates the raw data from multiple sensors directly, retaining more original information. However, this approach typically has high computational complexity and demands significant system resources [1]. Feature-level fusion focuses on merging the features extracted from each sensor, reducing data dimensionality and computational complexity while enhancing the effectiveness of feature representation. This method has demonstrated a favorable balance of performance in practical applications. In contrast, decision-level fusion synthesizes the independent diagnostic results from each sensor to make a comprehensive decision. While robust, this method depends on the independent diagnosis of each sensor, which can lead to information redundancy and reduced processing efficiency [2]. Additionally, decision-level fusion struggles to capture the spatio-temporal correlations between sensors, making it challenging to optimize fault diagnosis at a finer granularity [3]. As a result, research generally indicates that feature-level fusion offers greater advantages in multi-sensor information processing due to its ease of implementation and effective trade-off between diagnostic accuracy and computational burden [4]. Wang et al. [5] proposed a multi-resolution multi-sensor fusion network model based on deep learning for motor fault diagnosis, using multi-scale analysis of motor vibration and stator current signals. Cui et al. [6] employed multivariate complex mode decomposition to decompose complex-valued signals from multiple directions and extracted multiple orbit features to reflect the system's condition. They constructed fusion feature images to achieve feature-level fusion of multi-sensor information. Liu et al. [7] proposed an improved multi-channel graph convolutional network for rotating machinery diagnosis, constructing graph data for each sensor and designing a parallel graph data processing framework to realize multi-channel feature fusion. However, most of these methods are limited to a certain depth of features, potentially resulting in insufficient representation of fault information and leading to information loss in the final fusion.

With the continuous advancement of deep learning technologies, numerous scholars have extensively explored the application of feature-level fusion methods based on deep learning in the fault diagnosis of rotating machinery. Wang et al. [8], Xie et al. [9], and Gong et al. [10] introduced CNN-based feature-level fusion models to enhance fault diagnosis using vibration signals. In addition to CNN-based feature-level fusion methods, Chen et al. [11] developed a feature fusion method for bearing fault diagnosis, involving the use of sparse autoencoders and deep belief networks. In [12], a one-dimensional convolutional LSTM was employed to fuse vibration signals, improving fault diagnosis accuracy by integrating information from multiple sensors via LSTM. Analyzing these methods reveals that feature-level fusion techniques based on deep learning have indeed made significant progress in the fault diagnosis of rotating machinery, particularly through the integration and fusion of multi-sensor signals using models such as convolutional neural networks (CNN). However, traditional deep learning models tend to focus primarily on extracting local features, often failing to fully capture the spatio-temporal correlations between different sensors. To address this issue, recent research has increasingly incorporated attention mechanisms to better capture the spatio-temporal dependencies among multiple sensors. In [13], an adaptive sparse attention network was proposed, which dynamically focuses on dispersed local fault information in real-time. This method demonstrated improved training efficiency and greater interpretability. In [14], the combination of dense convolutional blocks with a spatial attention mechanism enhanced the model's feature extraction capabilities while reducing the required data volume, enabling the recognition of varying degrees of bearing damage. Chen et al. [15] employed multi-dimensional data fusion, attention mechanisms, and multi-task learning to diagnose faults in gas sensor arrays. By introducing attention mechanisms, they effectively captured critical information within sensor data, improving diagnostic accuracy and robustness, especially in noisy environments and under varying operational conditions. Li et al. [16] proposed a motor fault diagnosis model based on multi-channel signal fusion and an efficient channel attention (ECA) mechanism, accurately identifying motor fault patterns by fusing data from multiple sensors and incorporating the ECA mechanism, significantly improving fault recognition accuracy.

In response to this, this study proposes a fault diagnosis method that combines 1D-CNN with a spatio-temporal attention mechanism, referred to as the Temporal-Spatial Attention One-Dimensional Convolutional Neural Network (TAS-1DCNN). First, the 1D-CNN is employed to extract features from the raw signals of the sensors, refining local characteristics within the signals to ensure the completeness and effectiveness of fault information. Next, by incorporating the spatio-temporal attention mechanism, the feature weights are adaptively adjusted based on the temporal and spatial correlations and importance of different sensors. This approach captures both local and global dependencies among sensors, allowing for a deeper exploration of the spatio-temporal information across multiple sensors. Ultimately, this method enables the effective fusion of multi-sensor data, significantly improving fault diagnosis accuracy and enhancing the robustness of the system.

The remainder of this study is organized as follows: Section 2 presents the theoretical background on one-dimensional convolutional neural networks and attention mechanisms; Section 3 introduces the spatio-temporal attention mechanism network for multi-sensor fusion and the multi-sensor diagnostic model; Section 4 provides experimental validation of the proposed method; finally, conclusions are drawn in Section 5.

2. THEORETICAL BACKGROUND

2.1. 1D-CNN

1D-CNN is a deep learning model specifically designed for processing one-dimensional time series data and is widely used in signal processing and fault diagnosis. The structure of 1D-CNN, as shown in Fig. 1, consists of convolutional layers, pooling layers, and fully connected layers. These layers work together to extract features and classify the input signals. It effectively captures local temporal features while reducing data dimensionality and computational complexity. Compared to traditional methods, 1D-CNN has the advantage of automatically extracting features without the need for complex manual feature engineering, significantly improving model accuracy and robustness. In multi-sensor fusion tasks, 1D-CNN also performs exceptionally well, further enhancing fault detection performance by processing multi-source data.

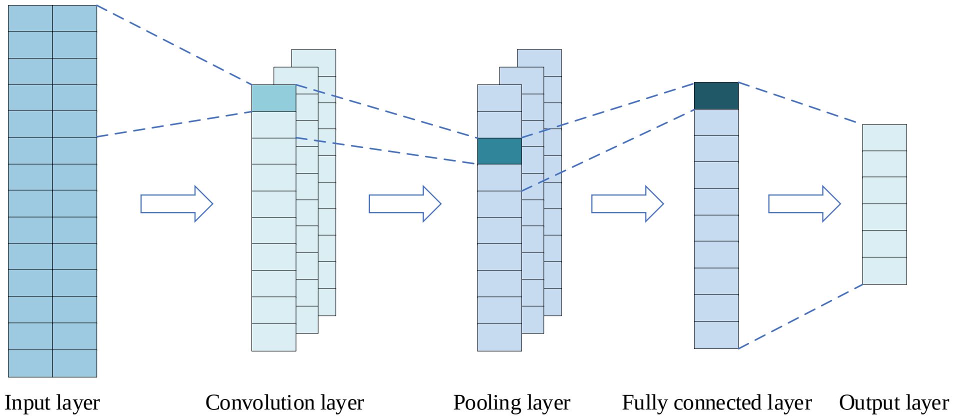


Figure 1. 1D CNN structure diagram

The convolutional layer is the fundamental building block of a 1D-CNN, responsible for extracting local features from the input data. The convolution operation can be expressed as follows:

$$\hbar_j^{\ell} = f\left(\sum_{i\in Fj} \hbar_i^{\ell-1} \bullet k_{ij}^{\ell} + b_j^{\ell}\right) \tag{1}$$

Where $\hbar_j^{\ell}$ represents the elements of the j-th input feature F_j in the ℓ-th layer. layer. k, b, and $f(\cdot)$ represent the convolutional kernel, bias, and the nonlinear activation function of the convolutional layer, respectively. A commonly used activation function is ReLU, which is defined as follows:

$$\mathrm{ReLU}(\hbar) = \max(0, \hbar) \tag{2}$$

The pooling layer is an essential component of 1D-CNN, responsible for reducing the spatial dimensions of the feature maps through downsampling while preserving key information. It decreases complexity, alleviates the computational burden, and extracts important features from the input data. The pooling layer operation is defined as follows:

$$\hbar_j^{\ell+1} = Pooling\left(\hbar_j^{\ell}\right) \tag{3}$$

Where $\hbar_j^{\ell+1}$ represents the elements of the j-th input feature F_j in the $\ell+1$-th layer, and $Pooling(\cdot)$ denotes the pooling operation. The fully connected layer integrates the local features extracted by the convolutional and pooling layers to generate the final output, which is defined as follows:

$$\hbar^{\ell} = g\left(w^{\ell}\hbar^{\ell-1} + b^{\ell}\right) \tag{4}$$

Where w^{ℓ}, b^{ℓ}, and $g(\cdot)$ represent the weights, bias, and nonlinear activation function of the fully connected layer, respectively.

2.2. Attention Mechanism

The attention mechanism, originating from the fields of deep learning and artificial intelligence, enhances the performance of machine learning models by mimicking the human ability to focus attention. This mechanism assigns different weights to various parts of the input based on their importance, generating a weighted sum of the input information. The model then uses this weighted sum as input, helping it focus more effectively on critical information. Typically, these weights are automatically learned and adjusted by the neural network. Assuming the input data and query vector are denoted as $D=(d_{\mu}, \mu=1,2,\cdots,n)$ and q, the attention score of d_i can be defined as follows:

$$s_{\mu} = softmax\left(S\left(q, d_{\mu}\right)\right) = \frac{\exp\left(f\left(q, d_{\mu}\right)\right)}{\sum_{\psi=1}^{n} \exp(f(q, d_{\psi})} \tag{5}$$

Where the scoring function $S(\cdot)$ calculates the similarity between the query q and the input element d_{μ}, and the softmax function transforms these scores into a probability distribution, ensuring that the total sum of the weights aaa equals 1, i.e., $\sum_{\mu=1}^{n} s_{\mu} = 1$. The weighted sum is then computed using the weights s_{μ} and the corresponding input values d_{μ}, as shown in the following formula:

$$context = \sum_{i=1}^{n} s_{\mu} d_{\mu} \tag{6}$$

Where *context* represents the key information that the model focuses on.

3. METHODOLOGY FRAMEWORK

3.1. Temporal Attention Mechanism

The temporal attention mechanism is a deep learning technique specifically designed for time series data. Its core idea is to assign different weights to different time points, enabling the model to focus on the moments that are most influential to the task. Traditional deep learning models often treat all time points equally, overlooking the potential key changes in the time series. The temporal attention mechanism addresses this issue by applying weighted processing to the time points, which is particularly effective in fault diagnosis for industrial equipment. It helps identify critical time points in sensor data, often revealing abnormal conditions or impending faults in the equipment. This significantly improves the accuracy of predictive maintenance and fault diagnosis. Furthermore, incorporating the temporal attention mechanism enhances the model's flexibility and accuracy in handling time series data. The proposed temporal attention mechanism in this study consists mainly of a self-attention module (Self-Attention, SA) and a weighted average module (Weighted Average, WA), with the specific structure illustrated in Fig. 2.

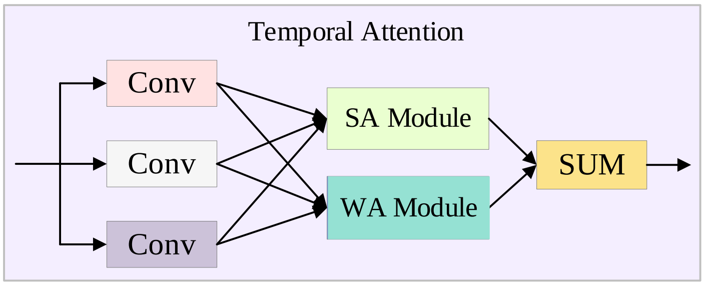


Figure 2. Temporal Attention Mechanism Module

The self-attention module captures global dependencies by calculating the similarity between elements within the input sequence. It typically employs dot-product attention, which calculates the similarity between the query, key, and value to generate attention weights. The structure of the self-attention module is shown in Fig. 3.

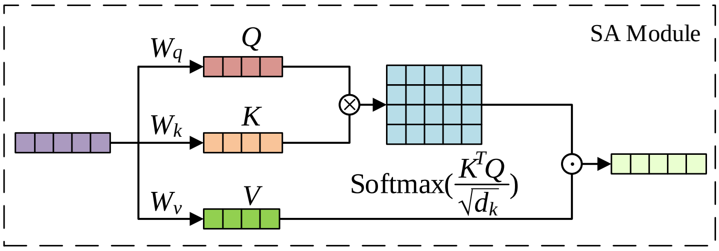


Figure 3. Self-Attention Module

The self-attention mechanism allows each element of the input sequence to be weighted based on the other elements in the sequence, enabling each element to gather information from different positions within the sequence. By calculating the relationships between elements in the sequence and then applying the softmax function, attention weights are obtained. These attention weights can be computed across different time steps, allowing the integration and interaction of information from all time steps.

Each input sequence $X = \{x_1, \ldots, x_n\}$ in the attention mechanism consists of nnn elements, where x_i represents the i-th element in the sequence. For each element, the query Q, key K, and value V are calculated separately:

$$\begin{cases} Q = x_i * W_q \\ K = x_i * W_k \\ V = x_i * W_v \end{cases} \tag{7}$$

In Eq. (7), W_q, W_k, and W_v represent the weight matrices that perform linear transformations on Q, K, and V, respectively. These parameter matrices are unique to each layer of the model.

The calculation of attention weights involves performing a dot product operation between Q and K, followed by division by a normalization factor, and finally applying the softmax function for normalization. This can be expressed as follows:

$$\alpha_i = \text{Softmax}(\frac{K^T Q}{\sqrt{d_k}}) \tag{8}$$

In Eq. (8), d_k represents the dimensionality of K, and K^T denotes the transpose of the matrix. In the Softmax function, the result of the dot product is divided by ggg to scale the dot product, preventing the attention weights from becoming too small or too large, which improves computational efficiency. Finally, the weighted sum is computed by performing a weighted summation of the attention weights and v, resulting in the following weighted sum:

$$H_i = sum(\alpha_i * V_j), j = 1, 2, \dots, n \tag{9}$$

In Eq. (9), H_i represents the i-th element of the output sequence, V_j denotes the j-th element of the value sequence, and α_i refers to the attention weight of the i-th element.

Convolutional weighted averaging effectively extracts local features, smooths the data, and reduces noise levels, thereby clarifying patterns. The structure of the weighted averaging module is shown in Fig. 4.

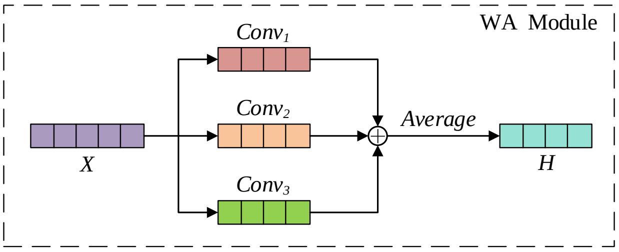


Figure 4. Weighted Average Module

The weighted averaging calculation process is as follows:

$$H_i' = \frac{1}{3}\sum_{i=1}^{3} Conv_i \tag{10}$$

Where H_i' represents the output of the weighted averaging, and $Conv_i$ refers to the result obtained by applying a specific convolutional kernel through the convolution operation.

In summary, the final output feature representation of the temporal attention mechanism is given as:

$$Z_i = \mathrm{ReLU}(H_i + H_i') \tag{11}$$

In Eq. (11), $\mathrm{ReLU}(\cdot)$ represents the Rectified Linear Unit, which performs an element-wise operation that sets all negative values to zero while retaining positive values. This operation facilitates the flow of information within the network, helps mitigate the vanishing gradient problem, and enhances the model's learning performance.

3.2. Spatial Attention Mechanism

The spatial attention mechanism is a deep learning technique that dynamically assigns weights based on the importance of features at different spatial positions. This helps the model focus on key areas, suppress noise, and improve both efficiency and accuracy. In industrial equipment fault diagnosis, the spatial attention mechanism captures critical spatial features, enhancing diagnostic accuracy. In image processing, it highlights essential regions of the image, improving recognition and classification performance. Additionally, the spatial attention mechanism is applied in fields such as natural language processing and audio processing, where it dynamically focuses on the most important parts of the input data, thus boosting the model's ability to handle complex data. The structure of the spatial attention mechanism is shown in Fig. 5.

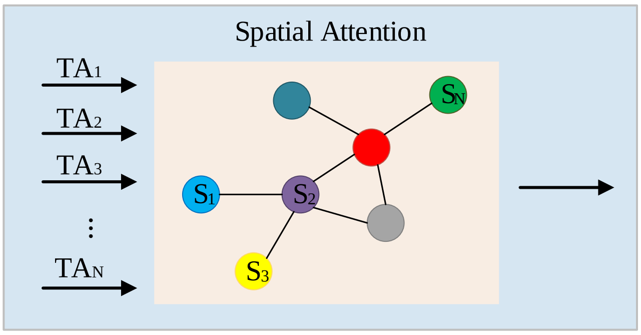


Figure 5. Spatial Attention Mechanism Modul

For the spatial attention module, the input consists of a series of node features $h = \{h_1, h_2, \ldots h_N\}, h_i \in R^F$, where there are N nodes, and each node has F-dimensional features. To retain sufficient expressive power, the input features need to be mapped to a higher-dimensional feature space. Each node i and j must undergo at least one linear transformation to calculate the corresponding attention coefficient for each node, as shown in Eq. (12):

$$e_{ij} = \alpha(Wh_i, Wh_j) \tag{12}$$

Where W is a random weight matrix, and eee represents the influence coefficient of node i on node j. To simplify the calculation and facilitate comparison of the attention coefficients, the *softmax* function is applied to normalize the influence coefficients of all neighboring nodes j for node i, as shown in Eq. (13).

$$\alpha_{ij} = \mathrm{Softmax}(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in N_i} \exp(e_{ik})} \tag{13}$$

Where k represents the neighboring nodes of node i. The final attention coefficient is computed by incorporating a non-linear function, *LeakyReLU*, before normalization, as shown in Eq. (14):

$$\alpha_{ij} = \mathrm{Softmax}(\mathrm{LeakyReLU}(e_{ij})) = \frac{\exp(\mathrm{LeakyReLU}(\alpha(Wh_i, Wh_j)))}{\sum_{k \in N_i} \exp(\mathrm{LeakyReLU}(\alpha(Wh_i, Wh_j)))} \tag{14}$$

Incorporating a non-linear function helps the model learn more complex data representations, preventing information loss and enabling the model to better handle outliers and noise. This improves the model's robustness and enhances the gradient flow throughout the entire network, leading to better overall performance.

3.3. Multi-sensor Spatio-temporal Attention Mechanism Network

The overall fault diagnosis framework proposed in this study is depicted in Fig. 6. Initially, one-dimensional time series data collected from various sensors are processed through a 1D-CNN, which extracts key features via its convolutional, batch normalization (BN), and pooling layers. These layers capture temporal patterns and local dependencies within the data. Subsequently, a temporal attention mechanism is applied to the features extracted by the 1D-CNN, enabling the model to concentrate on the most relevant features in the time series for fault diagnosis, thus improving the model's sensitivity to critical time points. Following this, a spatial attention mechanism is employed to analyze and highlight spatial correlations among the sensors, preserving significant features while disregarding less pertinent ones. The resulting feature data is then flattened and passed to the output module, where fault classification is performed using a Softmax function to yield diagnostic results.

By integrating the preliminary features extracted by the 1D-CNN with the enhanced features provided by the temporal and spatial attention mechanisms, classification algorithms are employed to identify and classify different fault types in the equipment. This approach effectively addresses the complexity of multi-sensor data, significantly improving the accuracy and efficiency of fault diagnosis.

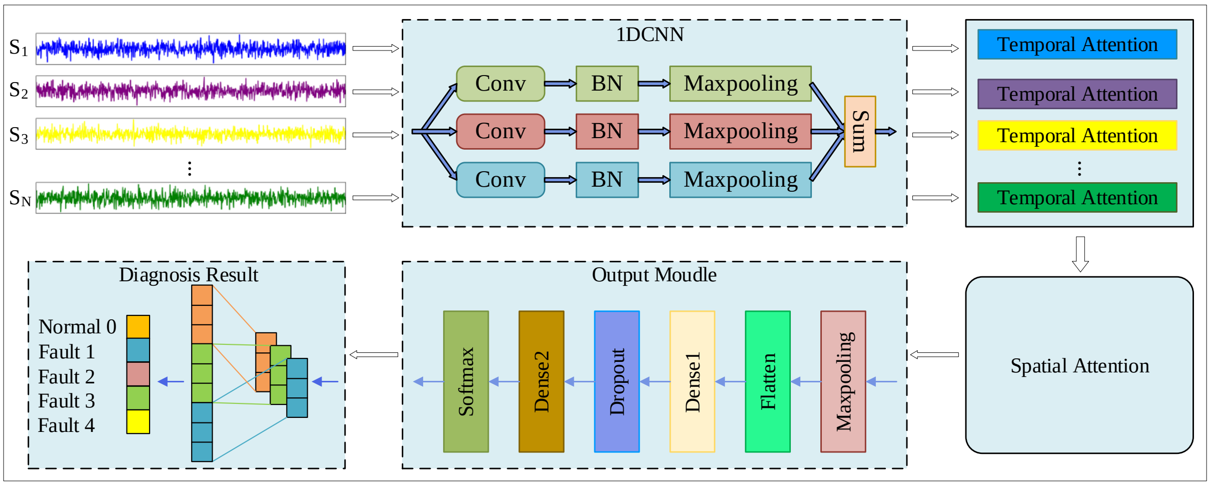


Figure 6. Fault diagnosis overall structure diagram

To achieve efficient fault diagnosis, data is first collected from various sensors and subjected to a series of preprocessing operations, including data cleaning, standardization, and noise reduction, to ensure that the input data is standardized and free of noise. Subsequently, the preprocessed data is fed into a 1D-CNN for training, where critical features are extracted, and local dependencies within the sequences are captured. A temporal attention mechanism is then introduced to identify significant time points in the data, enabling the network to focus on the most relevant information for fault diagnosis. Additionally, a spatial attention mechanism is employed to emphasize key spatial features, further enhancing the network’s sensitivity to fault-related patterns and improving its discriminatory capabilities. After extracting features and integrating them using spatio-temporal attention mechanisms, the fused features are utilized to train a multi-sensor diagnostic network. The network's parameters are iteratively optimized to improve its performance. Following the completion of training, the model's accuracy and generalizability are validated using a test dataset. Finally, the diagnostic results are derived from the network's analysis and output. The integration of spatio-temporal attention mechanisms into the fault diagnosis workflow not only improves diagnostic accuracy and efficiency but also enhances the system's adaptability to complex and dynamically changing environments. This approach offers a robust solution for intelligent fault diagnosis in modern industrial systems.

4. EXPERIMENTAL VERIFICATION

4.1. Nickel Flash Smelting System

The method proposed in this study is applied to the fault diagnosis of the fan in a nickel smelting process flash furnace, with the overall structure illustrated in Fig. 7. To ensure the safe operation of the flash smelting system and to prevent issues such as reduced desulfurization efficiency, equipment damage, or safety accidents due to fan malfunctions, the system utilizes multiple sensors to monitor the fan's operational status. Based on the collected data, the system adjusts the fan speed or initiates an emergency shutdown. Given the diverse types of mechanical equipment involved, relying solely on a single signal source for fault diagnosis is insufficient. The complementary use of multi-source signals, such as vibration, acoustic, and temperature data, can significantly enhance the accuracy of fault diagnosis. To validate the effectiveness of the proposed method, data from three types of sensors were selected for simulation testing in the diagnostic model.

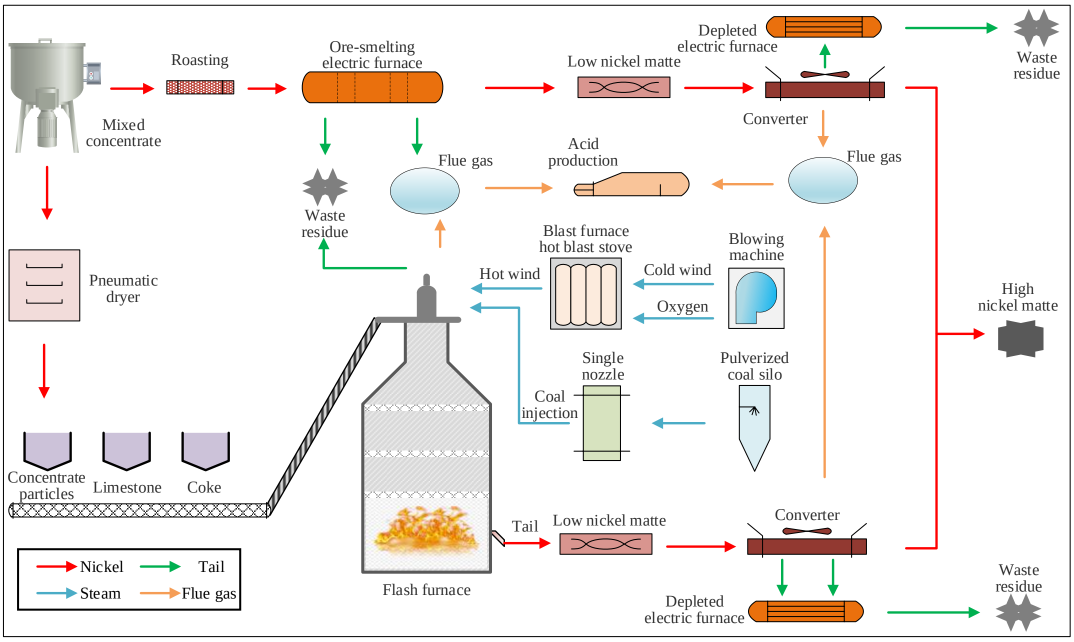


Figure 7. Nickel flash smelting process flow diagram

4.1.1. Experimental data.

To ensure diversity in the experimental data, five different operating conditions were simulated, with corresponding signals collected using vibration, sound, and temperature sensors. The experimental data was obtained through random sampling from the original dataset to ensure objectivity in the results. For each fault type, 4,000 samples were collected for vibration, sound, and temperature, respectively, totaling 20,000 vibration, 20,000 sound, and 20,000 temperature signals. The data was split into training and test sets at a ratio of 4:1, with 75% of the samples used for training and the remaining 25% for testing, which was used for model learning and validation. The sampled data is shown in Table 1.

Table 1. Sample data.

Fault Category	Total Samples	Training Samples	Test Samples	Label
normal	4000	3000	1000	0
bearing fault	4000	3000	1000	1
gearbox fault	4000	3000	1000	2
blade fault	4000	3000	1000	3
generator fault	4000	3000	1000	4

Common gear faults include wear, tooth chipping, tooth breakage, misalignment, and skew, while bearing faults typically involve wear, fatigue, pitting, deposits, and eccentricity. Motor faults are generally characterized by winding burnouts, brush wear, bearing failures, mechanical part damage, and unstable operation. To analyze these fault types, vibration sensors were employed to capture vibration signals under each fault condition. The experiment simulated both normal operation and five distinct fault scenarios, with a sampling frequency of 100 Hz. For each condition, 4,700 data points were collected, resulting in a total of 28,200 data points for each fault category. The detailed sampling data is presented in Table 2.

Table 2. Sample data for each type of fault.

Gearbox Fault	Bearing Fault	Generator Fault	Training Samples	Test Samples	Label
normal	normal	normal	3760	940	0
wear	wear	winding burnout	3760	940	1
shedding	fatigue	brush wear	3760	940	2
tooth breakage	shedding	bearing failure	3760	940	3
Eccentric wear	sediment	mechanical component damage	3760	940	4
skew	eccentricity	unstable motor operation	3760	940	5

4.1.2. Model Training and Parameters.

The multi-sensor fusion model proposed in this study is based on a spatio-temporal attention mechanism and primarily consists of three components: 1D-CNN, temporal attention mechanism, and spatial attention mechanism. Each 1D-CNN comprises three scale branches, all of which consist of convolutional layers, batch normalization layers, and max pooling layers with identical parameter settings. ReLU activation functions are employed in all convolutional layers, with "same" padding applied. The model outputs fault classification results via a Softmax function, with five output neurons corresponding to the five fault types. To prevent overfitting, a dropout rate of 0.5 is used. Given the large volume of data, the model is trained using mini-batches, where the choice of batch size impacts both training efficiency and accuracy. In this study, the batch size was adjusted and a learning rate of 0.001 was set to optimize the training process. The model was developed using the TensorFlow framework with Python 3.7 and executed on a Windows 10 platform. Detailed network structure parameters are presented in Table 3.

Table 3. Detailed parameters of the network model

Layer	Layer Type	Output shape	Parameters	BN	Activation function
1	Input layer	(None, 200, 1)	0	N	
2	Conv1d	(None, 198, 64)	256	Y	ReLU
3	Conv1d_1	(None, 97, 128)	24704	Y	ReLU
4	Conv1d_2	(None, 46, 256)	98560	Y	ReLU
5	Conv1d_3	(None, 21, 256)	196864	Y	ReLU
6	Attention1	(None, None, 32)	70496	N	
7	Attention2	(None, 3, 5)	170	N	
8	Flatten	(None, 15)	0	N	
9	Dense1	(None, 128)	2048	N	ReLU
10	Dense2	(None, 64)	8256	N	ReLU
11	Out dense	(None, 5)	325	N	Softmax

4.1.3. Evaluation metrics.

The efficiency of the proposed method is evaluated using accuracy, precision, recall, specificity and F1-score. These metrics are defined by Eqs. (15)-(19), respectively.

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + TN + FP} \tag{15}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{16}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{17}$$

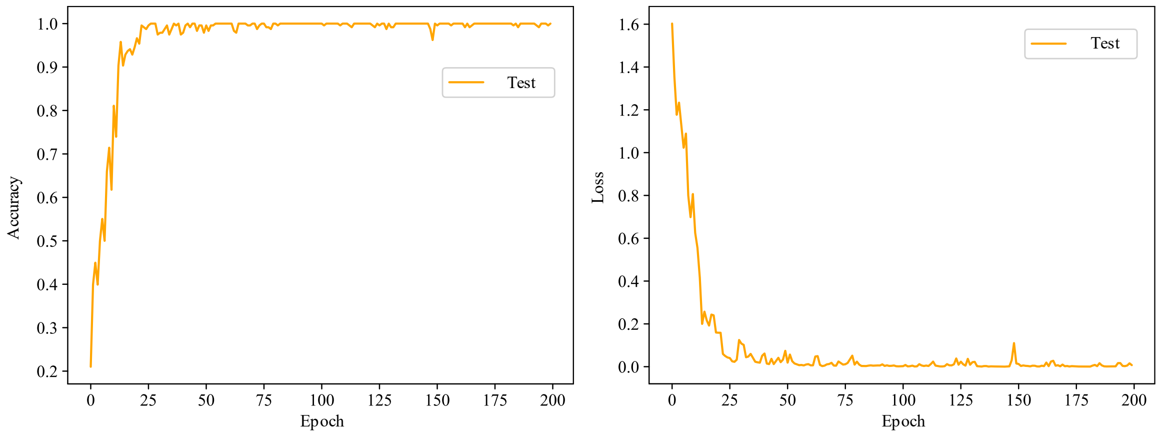
$$\text{Specificity} = 1 - \frac{FP}{TN + FP} \tag{18}$$

$$\text{F1} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{19}$$

Where FN represents the number of false negatives, where the model incorrectly classifies positive samples as negative, and FP represents the number of false positives, where the model incorrectly classifies negative samples as positive. TN denotes true negatives, where the model correctly classifies negative samples, and TP denotes true positives, where the model correctly classifies positive samples. The F1-score is the harmonic mean of precision and recall. A higher F1-score indicates a better balance between precision and recall.

4.2. Experimental Results and Analysis

In a single experiment, the variation curves of accuracy and loss values are illustrated in Fig. 8 (a) and (b).



(a) (b)

Figure 8. Model accuracy and loss values

To validate the model's capability in diagnosing different fault types, a confusion matrix was generated for the diagnostic results on the test set during a single experiment, as shown in Fig. 9. The values in the matrix represent the number and proportion of correctly predicted samples. In categories 0, 2, 3, and 4, there were 54, 45, 42, and 52 accurate predictions, respectively, resulting in a recognition accuracy of 100%. In category 1, there were 44 accurate predictions, yielding a recognition accuracy of 97.78%.

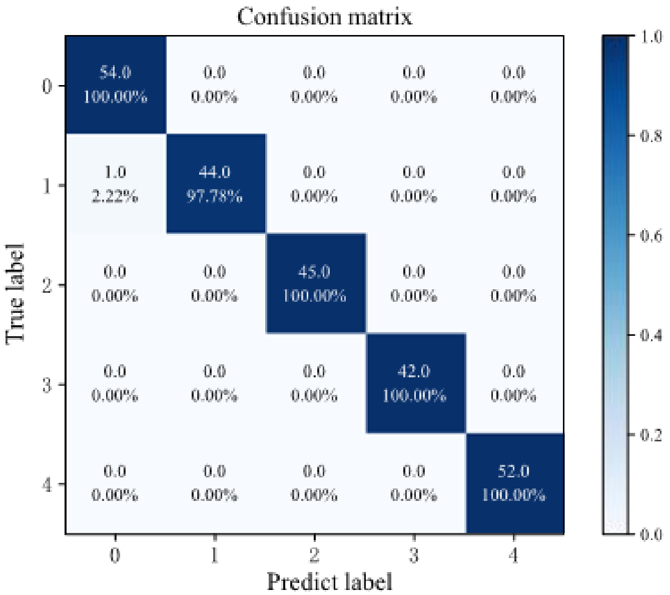


Figure 9. Confusion matrix of the model

Based on the confusion matrix, precision, recall, specificity, and F1-score for each operating condition were calculated, as shown in Table 4. The model achieved average precision, recall, specificity, and F1-score values of 99.64%, 99.56%, 99.89%, and 99.59%, respectively, demonstrating exceptional overall performance and robust diagnostic capabilities for all fault types. In terms of precision, fault categories 1, 2, 3, and 4 reached a maximum precision of 100%, whereas fault type 0 exhibited the lowest precision at 98.2%, indicating a minimal occurrence of false positives. For recall, fault types 0, 2, 3, and 4 attained a recall of 100%, while fault type 1 recorded the lowest recall at 97.8%, which reflects a reduced incidence of false negatives. Regarding specificity, fault types 1, 2, 3, and 4 achieved 100% specificity, whereas fault type 0 had the lowest specificity at 99.5%, highlighting the model's effectiveness in identifying negative samples. Concerning the F1-score, fault types 1, 3, and 4 attained perfect scores of 100%, while fault types 0 and 2 had the lowest

F1-scores at 99.1% and 98.9%, respectively. This indicates a commendable balance between precision and recall for these fault types.

Table 4. Results of classification evaluation metrics for the test set

Category	Precision	Recall	Specificity	F1-score
0	0.982	1.000	0.995	0.991
1	1.000	0.978	1.000	0.989
2	1.000	1.000	1.000	1.000
3	1.000	1.000	1.000	1.000
4	1.000	1.000	1.000	1.000
Average value	0.9964	0.9956	0.9989	0.9959

To enhance the understanding of the classification performance of convolutional neural networks and attention mechanisms at various layers in handling flash furnace fault states, the t-SNE algorithm was utilized to transform the model's output signals into a two-dimensional representation for visualization. The resulting visualization is presented in Fig. 10. The figure indicates that the features are effectively separated and clustered, exhibiting a robust clustering effect. This results in a clear distinction among the five fault states, characterized by a pronounced linear decision boundary.

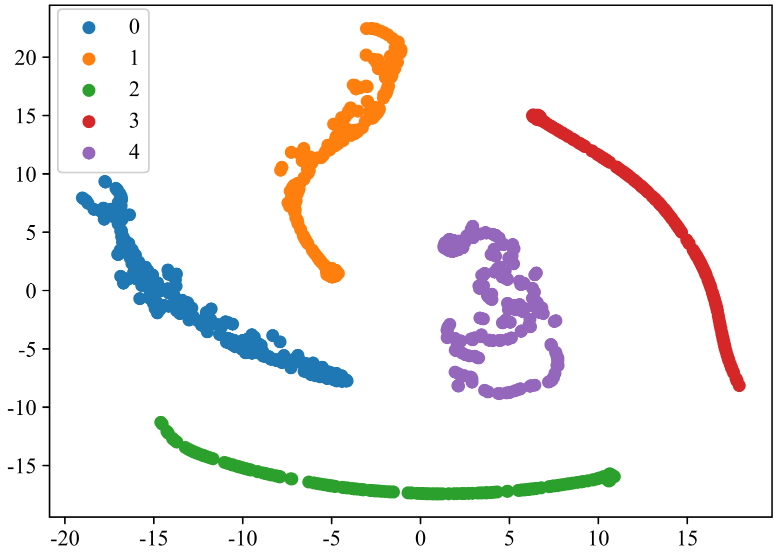
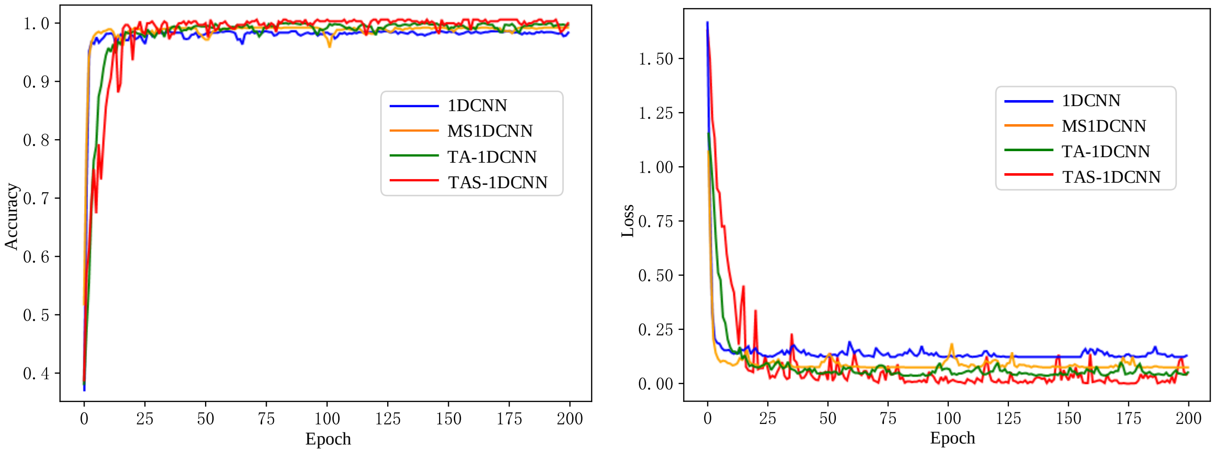


Figure 10. Visualization results of the model

4.3. Comparative Analysis of Experiments

To validate the effectiveness of the network model presented in this study, a comparative analysis was conducted among the following models: 1DCNN, Multi-Scale One-Dimensional Convolutional Neural Network (MS1DCNN), Temporal Attention One-Dimensional Convolutional Neural Network (TA-1DCNN), and the proposed TAS-1DCNN model. Based on the experimental design, each model's testing accuracy and loss values were obtained on the test set, as illustrated in Figures 11(a) and (b). The 1DCNN model exhibited the lowest accuracy and the highest loss value. In comparison to the MS1DCNN and TA-1DCNN models, the TAS-1DCNN model demonstrated varying degrees of improvement in accuracy and reduction in loss. The model proposed in this study achieved the highest diagnostic accuracy and the lowest loss value. These results indicate that the TAS-1DCNN model effectively enhances fault diagnosis performance.



(a) (b)

Figure 11. Comparison of loss and accuracy among different models

The experimental results demonstrate that the proposed method effectively integrates the advantages of 1D-CNN and spatiotemporal attention mechanisms, achieving faster convergence and higher classification stability. This leads to superior accuracy and lower loss values in the model. Evaluation metrics, including accuracy, recall, precision, specificity, F1-score, and model training time, were used to assess each method through 10 experimental trials, with the averages taken as the final results. The detailed data are shown in Table 5.

Table 5. The average results of metrics for different models

Model	Accuracy/%	Precision/%	Recall/%	Specificity/%	F1-score/%	Training time/s
1DCNN	97.36	97.91	98.03	99.48	97.92	460.26
MS1DCNN	98.01	98.07	98.15	99.47	98.03	545.16
TA-1DCNN	98.48	98.96	98.94	99.74	98.94	588.78
TAS-1DCNN	99.62	99.64	99.56	99.89	99.59	595.51

Table 5 presents the comparative results between the proposed fault diagnosis method and three other benchmark methods. As shown in Table 5, the recognition accuracies of the 1DCNN, MS1DCNN, TA-1DCNN, and TAS-1DCNN models are 97.36%, 98.01%, 98.48%, and 99.62%, respectively. The 1DCNN model exhibits a relatively low recognition accuracy and suboptimal performance, while the recognition accuracies of the MS1DCNN and TA-1DCNN models are slightly lower than that of the TAS-1DCNN model. The proposed model achieves an average accuracy improvement of 2.26%, 1.61%, and 1.14% over different models on the test set, indicating that the TAS-1DCNN model can extract features more comprehensively and effectively. Although the training time increases due to the larger network parameters, the model demonstrates higher diagnostic accuracy, enhancing fault recognition efficiency.

The experimental results indicate that the TAS-1DCNN model yields the best overall diagnostic performance. For the TAS-1DCNN model, the precision rates for fault states 0, 1, 2, 3, and 4 all exceed 95%. In contrast, the precision rates for the single-fault diagnoses of the 1DCNN, MS1DCNN, and TA-1DCNN models are mostly lower than those of the TAS-1DCNN model. From the analysis in Table 5, it can be observed that the recall and F1-scores follow the same trend. Therefore, the TAS-1DCNN model can more effectively isolate individual faults compared to the other models. For different faults, there are variations in precision, recall, and F1-scores, yet the TAS-1DCNN model consistently achieves relatively high values across all metrics, significantly improving fault diagnosis accuracy. Based on the above analysis, it is evident that the TAS-1DCNN model is more suitable for fault diagnosis in complex industrial equipment.

5. SUMMARY

This study presents a fault diagnosis approach based on the fusion of multisensor information using a spatio-temporal attention mechanism. Initially, raw data from multiple sensors are fed into a 1D-CNN for feature extraction, efficiently capturing critical temporal features. Next, a temporal attention mechanism is applied to dynamically uncover potential correlations between hidden features and target features, focusing on the most relevant portions of the time series for fault diagnosis. Finally, a spatial attention mechanism further explores the spatial dimension of these features, adaptively focusing on important information while bypassing less relevant data, leading to more precise and efficient fault diagnosis. Experimental comparisons show that the TAS-1D-CNN method achieves higher diagnostic accuracy.

REFERENCES

[1] Xiao X, Li C, He H, et al. Rotating machinery fault diagnosis method based on multi-level fusion framework of multi-sensor information. Information Fusion,Vol 113, pp. 102621, 2025.

[2] Zhang Q, Wei Y, Han Z, et al. Multimodal fusion on low-quality data: A comprehensive survey. arXiv preprint arXiv:240418947,Vol, 2024.

[3] Yang J, Gao T, Zhang H, et al. A multi-sensor fault diagnosis method for rotating machinery based on improved fuzzy support fusion and self-normalized spatio-temporal network. Measurement Science and Technology,Vol 34, pp. 125112, 2023.

[4] Zeng N, Wu P, Wang Z, et al. A small-sized object detection oriented multi-scale feature fusion approach with application to defect detection. IEEE Transactions on Instrumentation and Measurement,Vol 71, pp. 1-14, 2022.

[5] Wang J, Fu P, Zhang L, et al. Multilevel information fusion for induction motor fault diagnosis. IEEE/ASME Transactions on Mechatronics,Vol 24, pp. 2139-50, 2019.

[6] Cui X, Wu Y, Zhang X, et al. A novel fault diagnosis method for rotor-bearing system based on instantaneous orbit fusion feature image and deep convolutional neural network. IEEE/ASME Transactions on Mechatronics,Vol 28, pp. 1013-24, 2022.

[7] Yang C, Liu J, Zhou K, et al. An improved multi-channel graph convolutional network and its applications for rotating machinery diagnosis. Measurement,Vol 190, pp. 110720, 2022.

[8] Xing Z, Liu Y, Wang Q, et al. Multi-sensor signals with parallel attention convolutional neural network for bearing fault diagnosis. AIP Advances,Vol 12, 2022.

[9] Xie Y, Zhang T. Fault diagnosis for rotating machinery based on convolutional neural network and empirical mode decomposition. Shock and Vibration,Vol 2017, pp. 3084197, 2017.

[10] Gong W, Chen H, Zhang Z, et al. A novel deep learning method for intelligent fault diagnosis of rotating machinery based on improved CNN-SVM and multichannel data fusion. Sensors,Vol 19, pp. 1693, 2019.

[11] Chen Z, Li W. Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network. IEEE Transactions on instrumentation and measurement,Vol 66, pp. 1693-702, 2017.

[12] Hao S, Ge F-X, Li Y, et al. Multisensor bearing fault diagnosis based on one-dimensional convolutional long short-term memory networks. Measurement,Vol 159, pp. 107802, 2020.

[13] Jiang J, Li H, Mao Z, et al. A digital twin auxiliary approach based on adaptive sparse attention network for diesel engine fault diagnosis. Scientific reports,Vol 12, pp. 675, 2022.

[14] Plakias S, Boutalis Y S. Fault detection and identification of rolling element bearings with Attentive Dense CNN. Neurocomputing,Vol 405, pp. 208-17, 2020.

[15] Huang P, Wang Q, Chen H, et al. Gas Sensor Array Fault Diagnosis Based on Multi-Dimensional Fusion, an Attention Mechanism, and Multi-Task Learning. Sensors,Vol 23, pp. 7836, 2023.

[16] Miao Z, Feng W, Long Z, et al. Motor Fault Diagnosis Using Attention-Based Multisensor Feature Fusion. Energies,Vol 17, pp. 4053, 2024.