Complex & Intelligent Systems (2024) 10:8109–8125
https://doi.org/10.1007/s40747-024-01584-z

ORIGINAL ARTICLE

A novel local feature fusion architecture for wind turbine pitch fault diagnosis with redundant feature screening

Chuanbo Wen[1] · Xianbin Wu[1] · Zidong Wang[2] · Weibo Liu[2] · Junjie Yang[3]

Received: 12 January 2024 / Accepted: 6 July 2024 / Published online: 14 August 2024

Abstract

The safe and reliable operation of the pitch system is essential for the stable and efficient operation of a wind turbine (WT). The pitch fault data collected by supervisory control and data acquisition systems (SCADA) often contain a wide variety of variables, leading to redundant features that interfere with the accuracy of final diagnosis results, making it difficult to meet requirements. Also, the problem of extracting only local features while ignoring global information is present in the feature extraction process using the deep Convolutional Neural Network (CNN) model. To address these issues, the global average correlation coefficient is proposed in this article to measure the correlation between multiple variables in SCADA data. By considering the correlation among multiple variables comprehensively, redundant features are effectively eliminated, enhancing the accuracy of fault diagnosis. Furthermore, a new local amplification fusion architecture network (LAFA-Net) based on multi-head attention (MHA) is introduced. An efficient local feature extraction module, designed to enhance the model's perception of detailed features while maintaining global context information, is first introduced. LAFA-Net integrates the advantages of CNN and MHA, efficiently extracting and fusing valuable features from filtered data for both local and global aspects. Experiments on real pitch fault data demonstrate that the global average correlation coefficient effectively screens out redundant features in the dataset that negatively impact fault diagnosis results, thereby improving diagnosis efficiency and accuracy. The LAFA-Net model, capable of accurately diagnosing multiple types of pitch faults, shows a superior classification effect and accuracy compared to several advanced models, along with a faster convergence speed.

Keywords Fault diagnosis · Multi-head attention · Pearson correlation coefficient · Wind turbine · Deep learning · Redundant feature

Introduction

The installed capacity of wind turbines (WTs) has significantly increased as a crucial method for addressing energy balance and electricity demand. Within this context, the pitch system, a key component of the WT generator, plays a vital role in its operational process. Serving as the core subsystem for regulating WT output power and controlling blade safety braking, the pitch system frequently adjusts to variations in wind speed and direction. This frequent operation leads to a notable fault rate of 35%. Therefore, the development of effective fault diagnosis methods for the pitch system is essential for ensuring safe, low-cost operation and maintenance of the WT.

In recent years, deep learning-based intelligent fault diagnosis has garnered significant attention, which primarily involves two steps: fault feature extraction and fault pattern classification [1, 16, 23, 24, 38–40, 42]. It is noteworthy that an extensive amount of operation and maintenance data is recorded in the Supervisory Control and Data Acquisition (SCADA) systems of WTs [46]. The application of data mining technology to extract fault features from these data has increasingly become a predominant research approach

✉ Chuanbo Wen
chuanbowen@163.com

Zidong Wang
Zidong.Wang@brunel.ac.uk

Junjie Yang
yangjj@sdju.edu.cn

1 School of Electrical Engineering, Shanghai Dianji University, Shanghai 201306, China

2 Department of Computer Science, Brunel University London, Uxbridge, Middlesex UB8 3PH, UK

3 School of Electric Information Engineering, Shanghai Dianji University, Shanghai 201306, China

in WT fault diagnosis. Models frequently employed in this domain include recurrent neural networks (GRUs), Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Support Vector Machines [2, 21, 32]. For instance, a method has been proposed in [12] for comparing turbine behavior predicted by a training model with a reference space, which involves collecting SCADA data to simulate normal WT behavior and establishing the Mahalanobis space as the reference, thereby effectively detecting faults through residual construction. Furthermore, an adaptive multivariate time-series convolutional network has been introduced in [44], which resamples SCADA data over multiple time steps, and the resulting resampling matrix encapsulates multivariate time series information of different states, demonstrating excellent diagnostic performance.

It is worth mentioning that a framework based on degree-mean field theory has been proposed in [43], which constructs a directed graph and implements a fault localization method based on node heterogeneity abstracted from WT components. Moreover, a correlation graph-convolutional neural network approach for WT faults has been suggested [26], which considers the multidimensionality of SCADA data and their weak correlation with faults. This method employs a state tracking strategy to quantify the coupling between state parameters, which realizes a new WT fault diagnosis approach by using the correlation graph-convolutional neural network. While CNNs' advantages in feature extraction are widely acknowledged in the field of data-driven fault diagnosis, they are initially designed for image processing, which introduces certain limitations in this context. Specifically, the fixed kernel size of CNNs' convolutional layers leads to a relatively small perceptual domain and restrictions on global information acquisition, affecting their original performance. This limitation underscores the need to focus on researching SCADA data-based methods to provide effective and reliable SCADA fault diagnosis solutions.

Most wind farm manufacturers currently equip their wind farms with SCADA systems to continuously monitor operational status and performance. These systems record a vast amount of sensory data at intervals ranging from a few seconds to 10 min, covering a wide variety of state variables. However, the availability of fault data is often limited, so effectively screening out heterogeneous feature data can significantly enhance the accuracy of diagnosing pitch faults. Furthermore, with the ongoing advancement of deep learning models, effective feature selection has emerged as an active and expanding area of research, which involves downsizing and interpreting data to identify relevant features.

Feature selection methods can generally be divided into three categories based on their relationship with the model classifier: filter methods, wrapper methods, and embedded methods [11, 18, 50, 52]. Wrapper feature selection evaluates the performance of different feature subsets using machine learning algorithms. This method benefits from considering the interrelationships between features but incurs a high computational cost. Embedded feature selection, on the other hand, integrates feature selection into the model training process, allowing for automatic selection of the most representative feature subset. However, this approach tends to overlook the interrelationship between feature variables. Filter feature selection, evaluated based on statistics between features and target variables, is computationally efficient and well-suited for large-scale feature sets.

The filter feature selection method, which is based on different statistical features of data, computes feature correlations and performs feature selection independently of learning techniques, and this contrasts with the high computational cost of wrapper methods and the reliance on learning algorithms in embedded methods. Filter feature selection offers the advantages of a smaller computational load, wide applicability, and minimal dependence on classifiers. For instance, a multi-objective particle swarm optimization algorithm-based heuristic search feature selection method has been proposed in [22], which addresses the limitations of sequential methods but still faces the challenge of excessive computational demand. To mitigate this issue, an embedded chaotic whale survival algorithm has been introduced [7] by selecting a subset of features with a lower computational cost. While this method reduces computational effort, it does not overcome the inherent limitations of embedded feature selection, such as dependency on classifiers.

The shortcomings of both wrapper and embedded methods are effectively circumvented by filter feature selection, which operates independently of classifiers and learning algorithms. An iterative local Fisher score feature selection method, proposed in [6], aims to minimize local intra-class scatter while maximizing local inter-class scatter, thereby achieving effective feature selection. Furthermore, in [20], a method has been introduced to select a subset of significant features based on Euclidean distance for accurate fault diagnosis, and this approach exemplifies the efficacy of filter feature selection in isolating relevant features without the computational burden or classifier dependency seen in other methods. On the other hand, the Pearson Correlation Coefficient (PCC) is increasingly used in filter feature selection methods to identify feature redundancy by assessing the correlation between two variables [51], which is employed to balance relevance and redundancy in feature selection [34], optimize feature sets for fault diagnosis [49], solve overfitting in prediction models [25], and rank features based on correlation and mutual information [33]. However, these methods often focus only on the correlation between pairs of features, neglecting the collective impact of multiple correlated feature variables in feature screening.

Inspired by the identified limitations of existing methods, in this paper, we propose a novel global average correlation coefficient (GACC) to select effective features from real WT SCADA raw data for the fault diagnosis of the pitch system. This method addresses the deficiency of the PCC (which is limited to measuring the correlation between pairs of features) while maintaining low computational effort and simultaneously considering the correlation among multiple feature classes. Furthermore, to tackle the limitation of CNNs in acquiring global information, we design a local feature amplification extraction module, which enhances the ability of the system to capture local nuances without losing sight of the overall global context. Moreover, a new deep learning architecture is developed to improve the efficient extraction and fusion of local features from multi-fault variables, particularly in scenarios with small sample sizes, and such an architecture ensures that the original global contextual information is preserved while dealing with the intricacies of multiple fault variables.

The key contributions of this article are outlined as follows.

(1) xIntroduction of a new GACC: This coefficient measures the correlation among multiple variables, overcoming the limitation of the PCC which only assesses correlations between pairs of variables. The GACC enhances the accuracy and reliability of screening redundant features.
(2) Design of a local feature amplification module: This module addresses the issue of CNNs losing global context information when extracting local features, with purpose to improve the classification accuracy for multi-class pitch faults in wind turbines.
(3) Development of the LAFA-Net architecture: This new architecture combines the strengths of convolutional networks and Multi-Head Attention (MHA) in learning local and global features, which utilizes MHA's capability in understanding global information dependencies and interactions between features, and efficiently fuses local features of various fault types for effective pitch fault diagnosis.
(4) Effective application of the LAFA-Net model: The proposed LAFA-Net model demonstrates strong performance on real pitch fault datasets, providing valuable insights for the operation and maintenance of wind farms.

The above contributions signify notable advancements in the field of wind turbine fault diagnosis, particularly in enhancing the accuracy and efficiency of fault detection and classification through innovative methods and architectures.

Theoretical background

Convolutional neural network

CNNs are a type of deep learning model renowned for their ability to actively extract features from data, particularly adept at extracting and learning local features from raw data for classification purposes. A standard CNN model typically comprises a series of convolutional layers, pooling layers, and fully connected layers Fig. 1. The primary function of the convolutional layer is to extract localized features. In this layer, different convolutional kernels function as various feature extractors, analogous to filters in signal processing. Without loss of generality, the convolution process can be represented as:

$$a_j^l = f\left(\sum_{i=1}^{M} a_i^{l-1} * w_{ij}^l + b_j^l\right) \tag{1}$$

where a_j^l is the jth feature map of the l layer, M is number of features of the $l-1$ layer, w_{ij}^l is the convolution kernel weights of the ith feature of layer $l-1$ and the jth feature of layer l, b_j^l is the deviation of the jth feature map at the lth layer, $f(\cdot)$ is a nonlinear activation function, which can usually be the ReLU function, the Sigmoid function, or the Tanh function.

In a CNN, the pooling layer reduces feature space dimensionality and improves computational efficiency through down-sampling. There are two primary types of pooling operations: the maximum pooling selects the maximum value from a subset of the input features, preserving major features, and the average pooling calculates the average value of a group of features, smoothing the input and reducing noise. These operations enhance the robustness of the CNN by making it less sensitive to the exact location of features. The pooling operations can be described as follows:

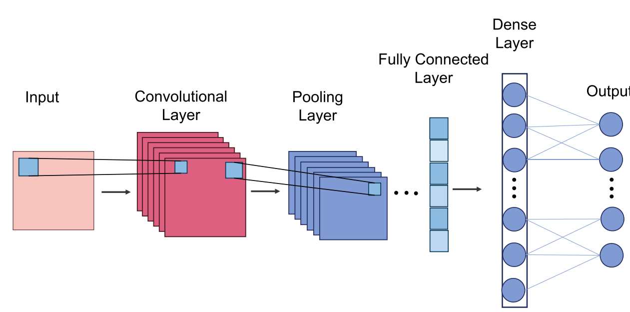


Fig. 1 1DCNN framework

$$P_m = \max_{a_j^l \in S} a_j^l \tag{2}$$

$$P_a = \operatorname{average}_{a_j^l \in S} a_j^l \tag{3}$$

where P_m and P_a are the output of the pooling layers and S is the pooling size of the sliding window. The functions `max` and `average` are employed to take the maximum value and the average value, respectively. The fully-connected layer is usually placed after multiple convolutional or pooling layers, and it maps the multidimensional features extracted from the previous layer into the vector label space in preparation for the subsequent classification task, and finally the `Softmax` function is used to compute the relative probabilities between different classes.

Pearson correlation coefficient

The PCC is mainly applied to measure the correlation between two variables, which is commonly used in the natural and social sciences to assess the degree of correlation between two independent variables, as well as the statistical significance and direction of this correlation. For two variables x and y with sample size N, the generalized PCC is defined as follows:

$$\rho = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}} \tag{4}$$

where x_i and y_i are the ith value in the variables x and y, respectively, and $\bar{x}$ and $\bar{y}$ are the means of x and y respectively. It can be seen from (4) that PCC ranges from -1 to $+1$. When $\rho > 0$, it indicates that there is a positive correlation between x and y; when $\rho < 0$, it indicates that there is a negative correlation between x and y; and when $\rho = 0$, it indicates that there is no correlation between x and y. When it is extended to a set of n variables $\left\{\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n\right\}$. A PCC matrix C is obtained and contains the values of the correlation coefficients for all pairs of variables, which can be described as:

$$C = \begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & \rho_{nn} \end{bmatrix}_{n \times n} \tag{5}$$

where ρ_{ij} $(i, j = 1, 2, \ldots, n)$ is the correlation coefficients between $\tilde{X}_i$ and $\tilde{X}_j$, which can be obtained from (4). It is easy to see that the C is a symmetric matrix and the diagonal elements are all 1.

Multi-head attention

In traditional attention mechanisms, attention weights are computed by calculating the similarity between the query and the key, and then the weights are applied to the value to obtain a weighted representation of the value. The multi-head attention mechanism enhances the expressive power of the model by introducing multiple attention heads and performing independent feature extraction and representation learning in each head. Therefore, the multi-head attention mechanism can be regarded as an extended form of the self-attention mechanism, which adds the ability to parallelize the computation of multiple attention heads on the basis of the original self-attention mechanism, thus further enhancing the model's expressive ability and performance. With different attention heads, the model can simultaneously attend to different feature subspaces, thus better capturing the dependencies between inputs and the interactions between features [27, 30]. Multi-Head Attention (MHA) has been widely applied in the fields of model prediction, speech recognition, sentiment analysis, and target detection [48].

The architectures of MHA and Scaled Dot-Product Attention (SDPA) are shown in Fig. 2. For an input sequence $X = [X_1, X_2, \ldots, X_N] \in \mathrm{R}^{D_x \times N}$, it first needs to be linearly mapped to three different spaces to obtain the query vector $q_i \in \mathrm{R}^{D_k}$, key vector $k_i \in \mathrm{R}^{D_k}$ and value vector $v_i \in \mathrm{R}^{D_v}$ respectively. The linear mapping process for X can be abbreviated as:

$$Q = W_q X \in \mathrm{R}^{D_k \times N} \tag{6}$$

$$K = W_k X \in \mathrm{R}^{D_k \times N} \tag{7}$$

$$V = W_v X \in \mathrm{R}^{D_v \times N} \tag{8}$$

where $W_q \in R^{D_k \times D_x}$, $W_k \in R^{D_k \times D_x}$ and $W_v \in R^{D_v \times D_x}$ are the parameter matrices of the linear mapping. $Q = [q_1, q_2, \ldots, q_N]$, $K = [k_1, k_2, \ldots, k_N]$ and $V = [v_1, v_2, \ldots, v_N]$ are matrices consisting of query vectors, key vectors and value vectors, respectively. D_k is the dimension of the query and key, D_v is the dimension of value, and N is the length of the input tensor. For each q_r, the output can be obtained by using the key-value pair attention mechanism:

$$h_r = \mathtt{att}((K, V), q_r) \tag{9}$$

$$= \sum_{j=1}^{N} \alpha_{rj} v_j = \sum_{j=1}^{N} \mathtt{softmax}(s(k_j, q_r)) v_j \tag{10}$$

where α_{rj} $(r, j = 1, 2, \ldots, n)$ denotes the weight of the rth output attention to the jth input, v_j is the value at a particular position in the value matrix V, and $\mathtt{att}(\cdot)$ is attention operation. The final sequence of output vectors can be represented as:

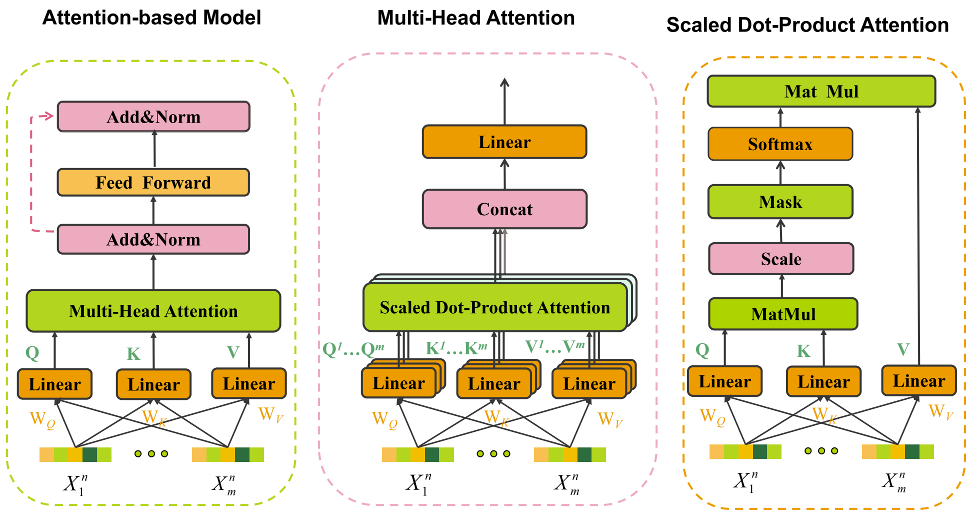


Fig. 2 Structure of multi-head attention

$$H = V \bullet \mathtt{softmax}\left(\frac{K^T Q}{\sqrt{D_k}}\right) \tag{11}$$

where $\mathtt{softmax}(\cdot)$ is the operation of normalization by columns, $\bullet$ is the scaled dot product operation.

After the extension of the self-attention model, we have

$$MultiHead(H) = W_o[head_1, head_2, \ldots head_M] \tag{12}$$

$$head_i = \mathtt{att}(Q_i, K_i, V_i) = \mathtt{att}(W_q^i H, W_k^i H, W_v^i H) \tag{13}$$

where W_o is the output projection matrix, $MultiHead(\cdot)$ is multi-head attention operation, M is number of multi-head attention, and $i = [1, 2, \ldots, M]$.

Proposed method

Feature screening

In WT pitch systems, raw fault data includes diverse variables from multiple sensors. However, not all variables are useful for fault diagnosis, as redundant and irrelevant features can lead to errors. Traditional PCC methods, focused only on pairwise correlations, are insufficient for this multivariate context. Therefore, in this paper, we introduce an extension to the generalized PCC, creating a global correlation coefficient. This new approach evaluates the correlation of one variable with many, overcoming the limitations of traditional PCC and reducing the impact of redundant features on diagnosis accuracy.

For the multi-classification problem of pitch fault diagnosis in wind turbines, there are h distinct outcome states, each corresponding to a different classification result, denoted as Y_l $(l = 1, 2, \ldots, h)$. Let $\tilde{X}^l$ represent the dataset recorded when fault Y_l occurs, and assume that each dataset has a sample size of N. We firstly define the correlation coefficient with respect to X_i based on dataset $\tilde{X}^l$ as follows:

$$g_i^l = \frac{n \cdot \left\|C_i^l\right\|_2 - \sum_{j=1}^{n}\left|\rho_{ij}^l\right|}{\sqrt{n^3}}, l = 1, 2, \cdots, h \tag{14}$$

where C_i^l $(l = 1, 2, \ldots, h)$ is the row vector corresponding to X_i based on dataset $\tilde{X}^l$, $\|C_i^l\|_2$ is the 2-norm of the vector C_i^l and $|\rho_{ij}^l|$ is the absolute value of ρ_{ij}^l. In (14), the PCC of X_i with other variables is formulated as a 2-norm. This approach maintains the correlation characteristics inherent to the original PCC, while also linking X_i with other variables to ascertain the degree of one-to-many correlations. The extension of the PCC in this manner makes the statistic more interpretable and relevant for multivariate analysis. Furthermore, it's important to address the presence of negative values in the PCC. These values represent the direction of correlation between pairs of variables but can lead to the loss of statistical properties when calculating a mean directly. To circumvent this issue, the absolute values of the vector elements are considered. This adjustment ensures that the resulting statistics accurately reflect the strength of the correlations, regardless of their direction, thereby providing a more robust and meaningful analysis for the one-to-many relationships in the context of pitch fault diagnosis.

In a multi-classification context, each state category has its dataset with the same number and type of features, but the contribution of each feature varies across categories. To standardize this in the analysis, the Global Average Correlation Coefficient (GACC) is used, which is calculated by averaging the global correlation coefficients of each variable across all state categories, as shown in the formula:

$$g_i = \frac{\sum_{l=1}^{h} g_{il}}{h} \tag{15}$$

where g_i represents the GACC for a variable, and g_{il} is its global correlation coefficient in the dataset of category l. The GACC values range from 0 to 1, with higher values indicating stronger global correlation. A suitable threshold θ is set for feature selection to identify the most relevant features for pitch fault diagnosis.

Remark 1 In (14), the 2-norm is utilized to link the self-correlation coefficient of a single variable with the correlation coefficients of other variables. This approach is effective for measuring the spatial distance of vectors, thereby better characterizing the variables' significance and their statistical relevance. By averaging the absolute values of these coefficients, the method refines the distance measure obtained via the 2-norm. This allows for a comprehensive synthesis of the correlation between a single variable and multiple variables, facilitating more accurate and meaningful variable screening for analysis or modeling.

Algorithm 1 Algorithm for screening of redundant feature

Require: Original datasets $\tilde{X}^l$ $(l = 1, 2, \cdots, h)$, classification result Y_l, l is the indicator of classification result;
Ensure: Updated variable X
for $i = 1$ to n **do**
  for $l = 1$ to h **do**
    Calculate $C_i^l = [\rho_{i1}^l, \rho_{i2}^l, \cdots, \rho_{in}^l]$ according to dataset $\tilde{X}^l$ and (4);
    Calculate g_i^l according to (14);
  end for
  Calculate g_i according to (15);
  if $g_i \geq \theta$ **then**
    Keep X_i;
  else
    Remove X_i from X;
    Update $X = [X_1, X_2, \cdots, X_{i-1}, X_{i+1}, \cdots X_n]^T$.
  end if
end for

Local feature extraction amplification module

The superiority of CNNs in extracting local features is well-established and has been effectively applied in various fields, including the analysis of sensor data collected by SCADA systems in wind turbines. These systems capture essential fault features, crucial for effective fault diagnosis. Unlike the more complex MHA, the convolutional operation in CNNs can be fine-tuned by altering the convolutional kernel size and stride, thereby adjusting the model's capability to learn local features during data propagation. Given the diverse sensor data in a pitch system, local features at the time of fault occurrence are critical for diagnosing specific fault types. However, while a single CNN model relies on multiple convolutional layers for efficient local feature learning, deep CNN architectures may lead to overfitting, especially with small sample datasets. To address this, we design a Local Feature Extraction Amplification (LFEA) module, which is capable of accurately extracting and enlarging local features without adding complexity to the model. It employs Global Average Pooling (GAP) and shape modifications to avoid overfitting, maintaining the inherent advantages of CNNs in local feature extraction. Moreover, the LFEA module ensures that global context information is preserved even after the local features are amplified.

The LFEA module in the model is designed to efficiently capture and amplify local features while preserving the integrity of the input data. The key components of the LFEA module include (1) the first convolutional layer that uses a convolution kernel size of 3 to precisely capture local features without losing critical information; (2) the second global average pooling layer that averages the extracted features to preserve and amplify key information; and (3) the feature rearrangement and further filtering, where the features are rearranged and further refined through an additional convolutional layer, combined with a multiplication operation with the input. This structure enables the LFEA module to extract essential local information effectively and maintain the completeness of input features, all without significantly increasing the model's computational demands.

LAFA-Net architecture

The advantages of the multi-head attention mechanism on global features include parallelized processing, multi-view focusing, improved model robustness, and integrated feature representations, which allows the model to capture multiple different global feature information simultaneously. The output of the multi-head attention mechanism is a linear combination of the feature representations of multiple heads, which allows the model to synthesize global features through feature fusion at different scales. This type of feature fusion helps to extract richer information and allows for better adaptation to different tasks. These advantages have made the multi-head attention mechanism one of the commonly used techniques in deep learning and have yielded good results in a variety of tasks.

In order to guarantee the effective diagnosis of multiple faults in the pitch system, it is not enough to accurately capture the local information of a single fault feature. On this basis, it is also necessary to have good overall learning ability for global context information. Thus, the local information of individual fault features is effectively integrated with the global information to ensure the accuracy of multi-fault classification and the robustness of the diagnostic model. We

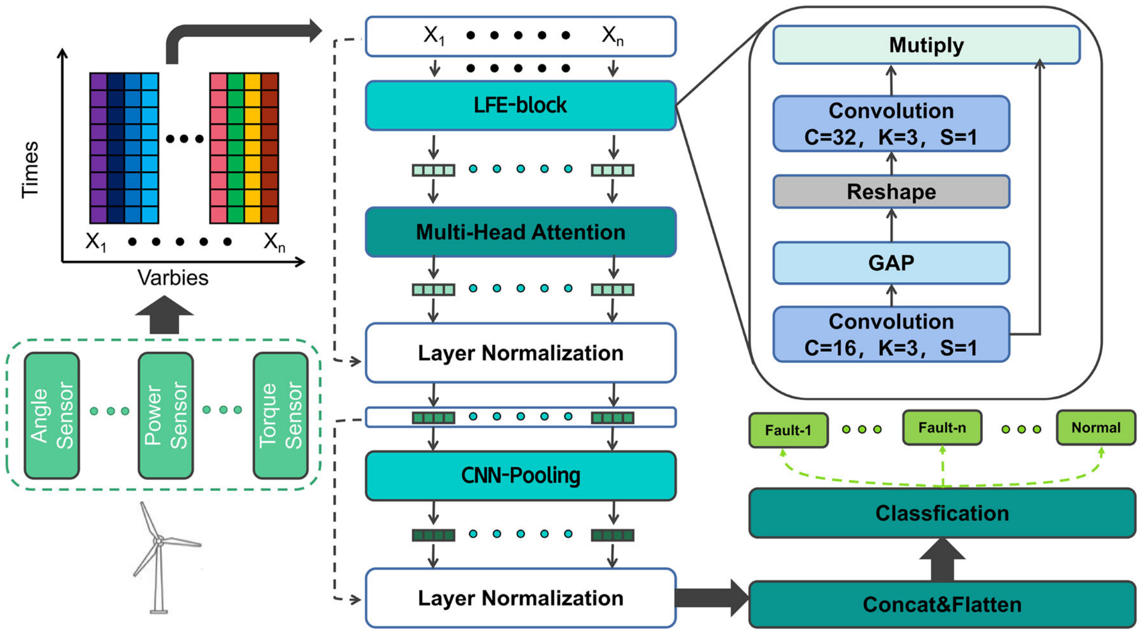


Fig. 3 Structure of the LAFA-Net model

propose a new feature learning architecture for LAFA-Net to effectively integrate local features and global information, refine the learning ability of the network on features, and improve the learning efficiency of the model on valuable features. The core of the model consists of two parts: the LFEA module and the MHA. The LFEA module realizes the accurate capture and amplification of local information, and the MHA realizes the extraction and learning of global information with different focuses. The two components work in conjunction with each other to enhance the model's ability to model input features.

For a given data set, the input is $X^i = \{x_1^i, x_2^i, \ldots, x_m^i\}$ $(i = 1, 2, \ldots, n)$, where x_m^i is the nth data for the mth variable, the corresponding output is $Y = \{Y_1, Y_2, \ldots, Y_h\}$. In the LAFA architectural model of Fig. 3, the inputs are filtered with the feature variables first extracted by the LFEA module to amplify the local features:

$$L(x) = \delta(P(\delta(x_m^n))^T; W) \otimes x_m^n \tag{16}$$

where $\delta(\cdot)$ is the convolution with ReLU being the activation function, $P(\cdot)$ is the global average pooling operation, $\otimes$ is multiplication operation, and W is the weights generated during the convolution operation. It is worth noting that the transpose operation is an essential part of the final amplification of local features and preservation of global information. After the LFEA module, the feature information is fed to the MHA core module to realize the fusion of globally important features, which combines the outputs of multiple attention headers, and then utilizes a linear layer with learnable weights W to aggregate these features to obtain the output of MHA:

$$Z^n = Concat(\zeta(x_1^n), \ldots \zeta(x_m^n)) \tag{17}$$

$$\zeta(x_m^n) = \mathtt{att}(L(x_m^n)W_q^m, L(x_m^n)W_k^m, L(x_m^n)W_v^m) \tag{18}$$

Subsequently, residual concatenation and layer normalization are performed on the output features of the MHA, an operation that effectively speeds up training while the offset of the internal covariates is reduced:

$$S^n = LN(X^n + Z^n) \tag{19}$$

$$LN(\mu) = \frac{\mu - \bar{\mu}}{\sqrt{\sigma^2 + \varepsilon}} \tag{20}$$

where $\bar{\mu}$ and σ^2 are the mean and variance of μ, respectively. Then, the output of layer normalization is fed into the combination module of convolutional and pooling layers for the secondary learning of local features. The feature information is refined by layer normalization and residual connection to complete the fusion of local features and the forward propagation learning process of global information. Finally, the flattening operation and `softmax` classifier are used to realize the classification of different fault types and normal operation states.

Application of LAFA-Net and GACC for pitch fault diagnosis

The application procedure of pitch fault diagnosis for WT on LAFA-Net model is shown in Fig. 4, and the specific details are as follows:

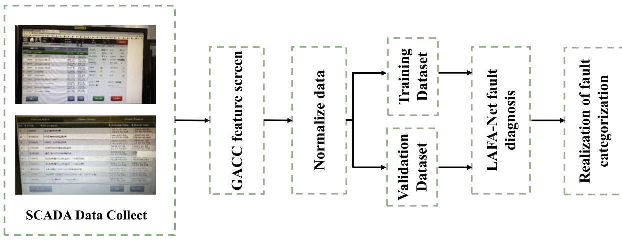


Fig. 4 Application procedure of pitch fault diagnosis

Step 1: Obtain raw sensor data from the turbine SCADA and intercept fault data from the fault log from the time of the fault to reset, as well as normal operation data prior to the fault.

Step 2: The preliminary selected data are screened using the GACC method to eliminate redundant variables in the data.

Step 3: The data retained in the various types of sensors are normalized to limit the data to a certain range and eliminate the adverse effects of outliers on the data itself.

Step 4: The data is randomly disrupted and divided into a training set and a validation set, where 80% is the training set and 20% is the validation set.

Step 5: The LAFA-Net model is first trained with the training set data, and then the validation set is used to test the performance of the model and the accuracy of the diagnosis of the paddle faults.

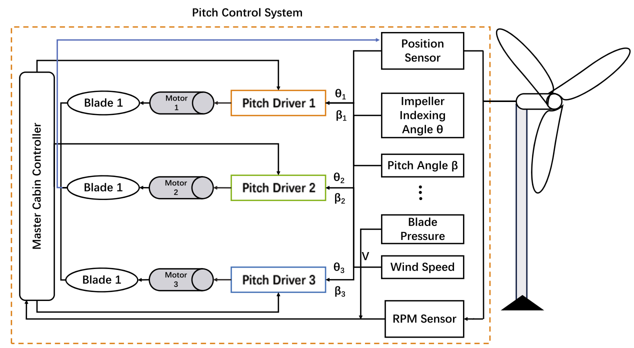


Fig. 5 Pitch control system

Table 1 SCADA data variables

No	Variable name	No	Variable name
1	Generator active power	12	Grid A phase voltage
2	Generator reactive power	13	Grid B phase voltage
3	Grid active power	14	Grid C phase voltage
4	Grid reactive power	15	Grid A Line voltage
5	Paddles set angle	16	Grid B Line voltage
6	Paddle A pitch angle	17	Grid C Line voltage
7	Paddle B pitch angle	18	Grid voltage
8	Paddle C pitch angle	19	Grid frequency
9	Paddle A pitch torque	20	Paddle A pitch press
10	Paddle B pitch torque	21	Paddle B pitch press
11	Paddle C pitch torque	22	Paddle C pitch press

Table 2 Types of operational faults

No	Running status	Label
1	Proportional valves error	0
2	Paris blades misaligned error	1
3	Blade specified position error	2
4	Pitch angle deviation error	3
5	Pitch system low pressure error	4
6	Normal	5

Experiments

In this section, we verify the effectiveness of our method by validating pitch fault data collected by real SCADA. The deep learning framework used in this article is Keras, the backend engine is TensorFlow, the computer processor used is Intel(R) Core i5-5200u-CPU@2.20 Ghz, and Python is used to build CNN models in the Keras environment.

Wind turbine pitch SCADA fault data

The pitch system is the core sub-system for power regulation in a WT and adjusts the pitch angle of the turbine blades in real time according to the wind speed to achieve the maximum output of the turbine. As shown in the Fig. 5, the pitch actuator will match the blade position according to the pressure and wind speed from the blade. The nacelle master controller end action signals to the pitch drive, based on which, the impeller index angle and pitch angle are adjusted, and the change of the blade position is realized.

The real SCADA data used in this article is from a 4MW DC-driven hydraulic WT at a wind farm in Shanghai, China. This data records a variety of types of fault data for the WT pitch system from July 2022 to January 2023, with a total of 18,000 sets of data, mainly including normal data and fault data within 10 s before and after the fault, with data sampled once at 10 ms on average. The operational data includes generator active power, generator reactive power, pitch angle, pitch torque, etc.; the status data includes normal operation and fault status codes and reset time stamp records, etc. The specific data names and variables are shown in Tables 1 and 2.

However, the data collected by the SCADA system cannot be trained immediately. SCADA collects fault data in the following manner. SACDA takes the fault timestamp as a reference when a turbine fault occurs, and then saves the normal data for the 10 s before the fault timestamp and the fault data for the 10 s after. The data recorded after each fault is not always fault data, because the operation status will be reset to normal operation after the fault is removed. At this time, the data recorded during the sampling time is the data during normal operation, which leads to the fact that the data 10 s after the fault is not always considered as fault data. Therefore, using data recorded before and after the 10 s period would result in abnormal correlation of variables, leading to inaccurate diagnostic results. First, the valid fault data is intercepted based on the time stamp of the fault and the reset time. Then, GACC is utilized to screen the features for each type of fault state. It is verified that a θ value set to 0.5 in the pitch fault data satisfies the screened dataset enabling the model accuracy optimal. Among the 22 feature variables in Table 1, five feature variables were excluded after GACC screening were Grid A Line voltage, Grid B Line voltage, Grid B Line voltage, Grid voltage, and Grid frequency. Finally, in order to achieve multi-fault diagnosis and classification, the datasets corresponding to the six different types of states are mixed and randomly disrupted. After all the above is done, 80% of all the final data is divided into training set and 20% into validation set.

Performance metric

The corresponding evaluation metrics need to be adjusted for the multi-fault classification problem. In addition to ac curacy, the F-score is used as an overall classification performance metric as well as a classification metric for individual categories. The F-score takes into account both the precision and recall of a model. By combining these two metrics, the F-score provides a more comprehensive and accurate assessment of a model's classification performance. In particular, in scenarios where a small number of categories are more important, the F score can better reflect the model's classification performance for most categories while focusing on a small number of categories. Therefore, F-score is a commonly used and reliable metric to help assess the classification performance of a model. The calculation formula is given in Table 3, where TP, TN, FP, and FN refer to true positives, true negatives, false positives, and false negatives in class i, respectively. k represents the number of classes being classified.

GACC feature screening experiment

In this subsection, we mainly discuss the effectiveness of the proposed GACC method after feature screening for improving the final accuracy of the model. Four datasets were used for validation respectively. Datasets 1 is the original data without any redundant feature screening, datasets 2 is the data after screening using GACC, and datasets 3 and 4 are the datasets with randomly screened feature variables, respectively. For consistency, the number of feature variables screened is the same as in datasets 2. However, the difference

Table 3 Performance metric and calculated expression

Performance metric	Calculated expression
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$
f1-score	$\frac{2(precision \times recall)}{precision+recall}$
Micro-precision (MIP)	$\frac{\sum_{i=1}^{k} TP_i}{\sum_{i=1}^{k} (TP_i+FP_i)}$
Micro-recall (MIR)	$\frac{\sum_{i=1}^{k} TP_i}{\sum_{i=1}^{k} (TP_i+FN_i)}$
F-score	$\frac{2(MIP \times MIR)}{NIP+MIR}$
Accuracy	$\frac{\sum_{i=1}^{k} (TP_i+TN_i)}{\sum_{i=1}^{k} (TP_i+TN_i+FP_i+FN_i)}$

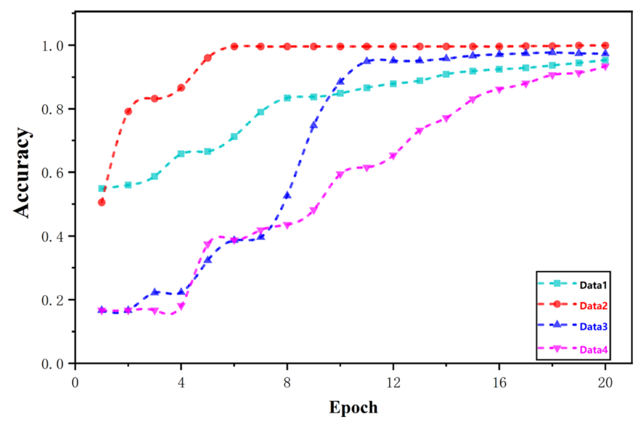


Fig. 6 Experimental accuracy for Dataset 1, Dataset 2, Dataset 3 and Dataset 4

is that the screened variables in datasets 3 contain a portion of the GACC screened variables, while datasets 4 does not.

On the LAFA-Net model, the model accuracies obtained from the four datasets are 0.9533, 0.9956, 0.9767, and 0.9522, respectively. It is not difficult to get the accuracy curves from Fig. 6, which show that the accuracy on datasets 2 is significantly better than that on the other three datasets. The results on datasets 3 can also effectively reflect that the features screened by GACC can effectively improve the model accuracy, because the randomly eliminated variables in datasets 3 contain some of the features screened by GACC.

The validation loss and accuracy curves of the proposed method are given in Fig. 7, from which it can be seen that in the early stage of training presents a higher loss due to the initialization of the model parameters. With the continuation of the training process, the model is rapidly accelerated and gradually converges to a converged state after only 6 epochs, while also reaching a high accuracy. The fluctuation of accuracy and loss in the subsequent process becomes very stable. The results show that our method has good generalization and can achieve a stable high accuracy with fewer training epochs. From the above comparative experimental results, the GAAC feature screening method can effectively elimi-

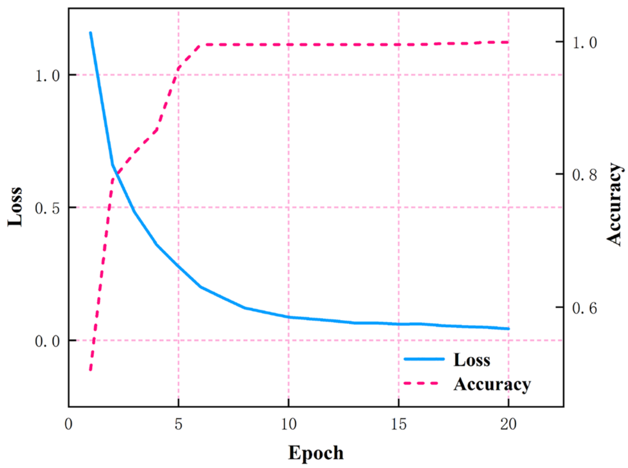


Fig. 7 LAFA-Net model validation loss, accuracy curves

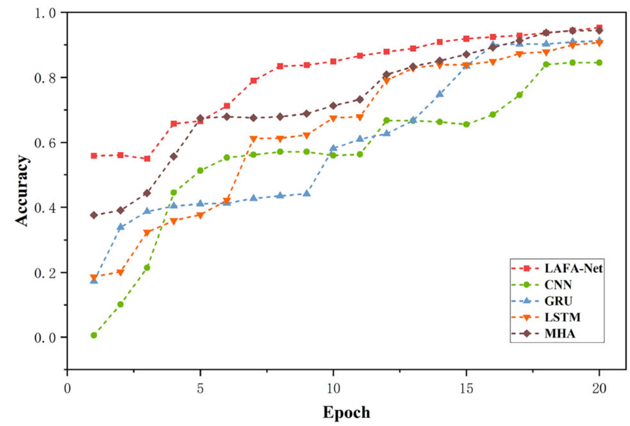


Fig. 8 LAFA-Net, CNN, GRU, LSTM and MHA accuracy curve comparison in DataSets1

nate redundant features to achieve the effect of improving the accuracy of the model.

Comparison with different deep learning methods for GACC

In this subsection, the discussion focuses on the feasibility of GACC's screening results on several different models. We compare the performance of LAFA-Net with GRU [2], LSTM [32], CNN [44] and MHA models. All of these models are more advanced and applicable in deep learning. CNN and MHA are models that are more sensitive to local features and globally valid information, respectively. GRU and LSTM are recurrent neural network-based models, which were initially used in natural language processing, and are able to have good learning classification processing ability for small sample sequence data, can effectively integrate dependencies between information, and do not need to consider the distance problem. These models can be compared with the LAFA model from three perspectives: local features, global information, and processing sequence data. We recorded accuracy, precision, and recall under LAFA-Net and four other baseline models. The results are shown in TABLE IV.

From Table 4, it can be seen that the accuracy and MIP of all the models in datasets 2 (GACC-screened data) have been greatly improved, and the accuracy and precision of the CNN model have been improved by 11.66% and 7.54% compared to dataset 1, reaching 96.22% and 97.19%, respectively. In GRU, LSTM and MHA, the accuracy is improved to 94.33%, 94.38% and 98.33%, and the MIP is improved to 95.96%, 96.16% and 98.64% compared to datasets 1. However, the highest values of accuracy and MIP are still

Table 4 Performance metrics of experimental results for different models on Data1, Data2, Data3 and Data4

Method	CNN (%)	GRU (%)	LSTM (%)	MHA (%)	LAFA-Net (%)
Datasets 1					
Accuracy	84.56	81.22	90.75	94.44	95.33
MIP	89.65	94.21	89.74	95.68	96.03
MIR	84.67	91.56	88.56	94.67	95.00
Datasets 2					
Accuracy	96.33	94.33	94.78	98.33	99.56
MIP	97.19	95.96	96.16	98.64	99.89
MIR	96.67	94.67	95.22	98.56	99.89
Datasets 3					
Accuracy	93.11	92.22	93.56	97.33	97.67
MIP	95.02	94.27	94.80	97.29	97.82
MIR	92.89	91.67	93.22	96.98	97.67
Datasets 4					
Accuracy	90.67	82.44	82.56	91.33	95.22
MIP	93.14	84.63	89.24	94.64	96.35
MIR	91.33	81.22	82.11	93.44	95.33

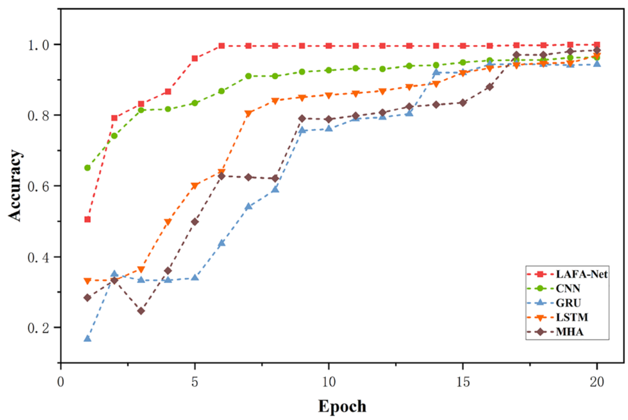
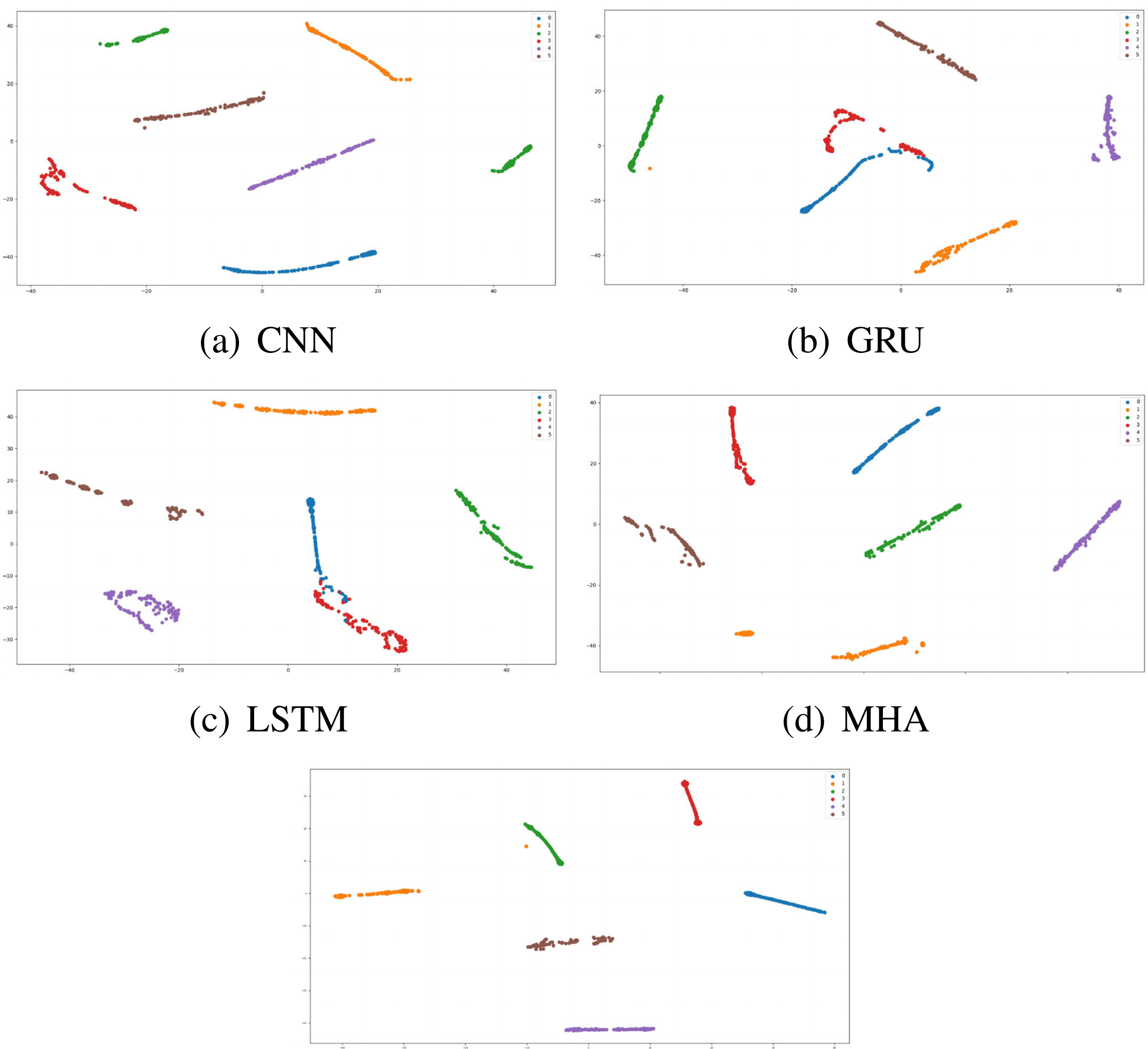


Fig. 9 LAFA-Net, CNN, GRU, LSTM and MHA accuracy curve comparison in DataSets2

in LAFA-Net, 99.56% and 99.89% respectively. It can be seen that our redundant feature screening method is able to effectively improve the accuracy of the final classification results and the overall accuracy of the model, and the advantages of the LAFA-Net model over other models in terms of final classification accuracy and model MIP can also be effectively verified.

In terms of MIR, the average recall of the LAFA-Net model exceeds 95% on all four datasets. This is an indication that LAFA-Net can accurately identify the majority of fault types, especially on dataset 2, where the MIR reaches 99.89%, which is 1.33% higher compared to the highest of 98.56% (MHA) among several other types of models. This indicates that only very few fault data are missed or misiden-



(a) CNN (b) GRU (c) LSTM (d) MHA (e) LAFA-Net

Fig. 10 Categorical scatterplot of experimental results of different pitch fault diagnosis experiments in 2D space tified in the given dataset, which proves that our model achieves better performance in the practical application of pitch fault diagnosis. Figures 8 and 9 show the accuracy curves of five different models on datasets 1 and 2. It can be seen that the proposed method converges significantly faster than several other types of models during the training process and can quickly reach a higher level of accuracy with fewer training rounds. From this result, our method of local feature fusion can effectively improve the training speed of the model, and the full fusion of various types of local features of faults and the combination of the model's own ability to learn global information is an effective means to improve the diagnostic accuracy of the model.

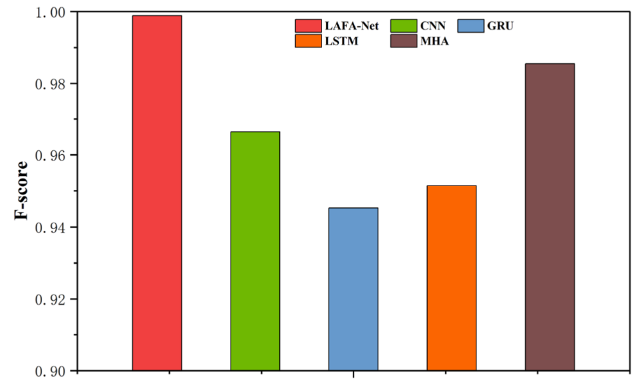


Fig. 11 Overall F1-score of CNN, GRU, MHA, LSTM, LAFA-Net models for pitch fault diagnosis

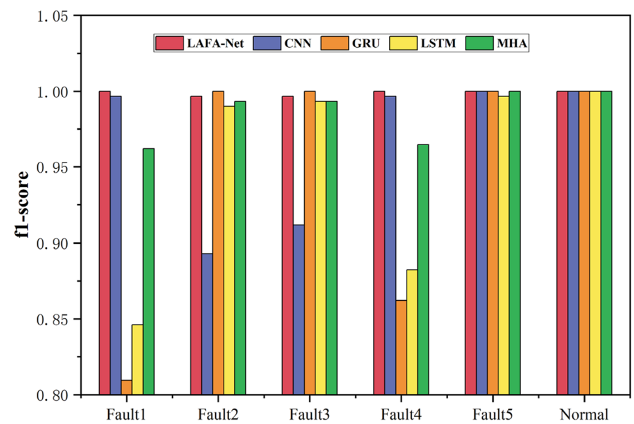


Fig. 12 F1-score of CNN, GRU, MHA, LSTM, and LAFA-Net models for six single type of pitch fault diagnosis

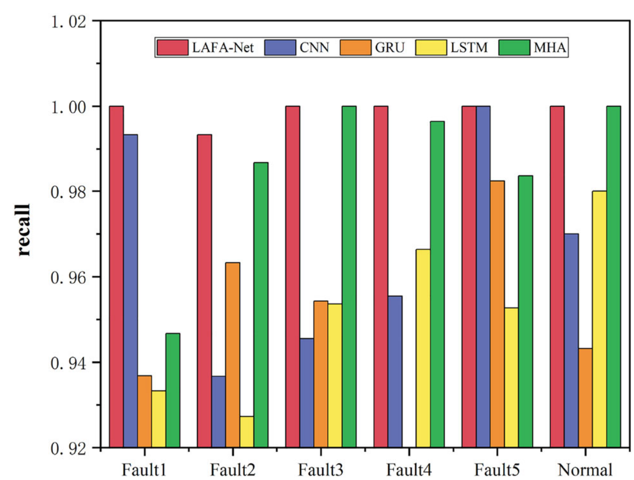


Fig. 13 Recognition recall of CNN, GRU, MHA, LSTM, and LAFA-Net models for six single types of pitch fault diagnosis

Comparison of pitch fault diagnosis results

In this subsection, we analyze and compare the actual classification effects on several different types of models for the pitch fault datasets 2 (data screened by GACC). We use two metrics, F-score and f1-score, to measure the effect of individual models on overall fault classification and single fault classification, and also use t-distributed stochastic neighbor embedding visualization technique to downscale the final data and present the final classification scatter plots. The visualization results of the five models for pitch fault diagnosis results are shown in Fig. 10, where the different colors indicate the different pitch fault types. Obviously, among the output distributions in the visualized scatterplot, LAFA-Net finally has the best classification results, which can accurately classify the distributions of almost all types of faults, and only one data point is neglected.

In MHA and CNN, although a more satisfactory classification result can be achieved, a part of fault states is obviously neglected in the final distribution, which reduces the classification accuracy of the model. However, the classification effect of the output distribution in GRU and LSTM is the worst, and the problem of faults being ignored is improved compared to MHA and CNN, but the phenomenon of being misclassified obviously occurs again. Obviously, the LAFA-Net model can not only effectively identify the fault state and fault characteristics, but also show certain superiority in the final classification in the application of the actual paddle fault diagnosis.

The f1-socre and recall results for single fault classification are shown in Figs. 12 and 13, where we can see that, in terms of both the effectiveness of the classification and the identification of the faults, the LAFA-Net model has a higher F1-socre and recall for each class of faults than several other classes of models. From the overall classification score in Fig. 11, the overall score of our method has reached 0.9989, which is at the highest level among all the models, 5.36% higher than the GRU model with the lowest score, and 1.34% higher than the MHA model with a relatively high score. As for the performance of single fault classification, our method scores on Fault 2 and Fault 3 are both 0.9967, and it scores 1 on several other fault types. The secondary better performer is the MHA model with scores of 0.9619, 0.9933, 0.9934, and 0.9646 on Fault1 through Fault4, respectively. And the other three types of models are significantly lower

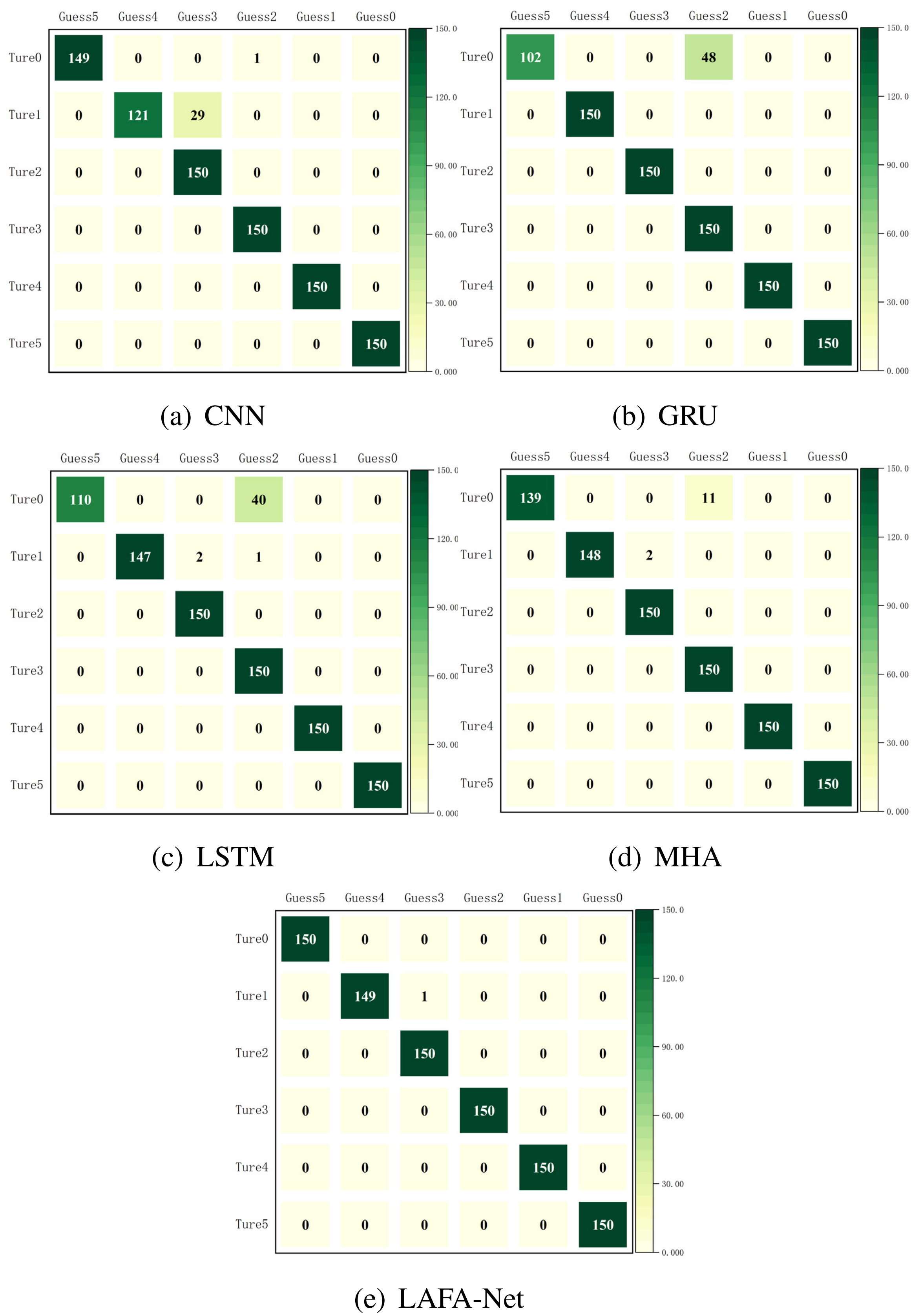


Fig. 14 Pitch fault diagnosis confusion matrices of different methods

than these two types of models in terms of scores. In terms of fault recognition rate, the LAFA-Net model only has a 99.33% recognition rate on Fault2, and the recognition rate of other types of fault states is 100%.

The model classification performance can be seen more intuitively in the confusion matrix, and Fig. 14 shows the confusion matrix of the classification results of the five models on datasets 2. After fusing the local features of a single fault, the classification performance of LAFA-Net on all types of faults reaches a high level. From the confusion matrix results of the LAFA-Net model, it is easy to see that the classification results are basically concentrated on the diagonal elements of the matrix, and only one data point is misclassified. Compared with the classification confusion matrix results of other types of models, the proposed method outperforms other types in overall classification performance and achieves a high classification accuracy. In summary, the above results strongly demonstrate that our method has efficient fault recognition and excellent classification performance for practical fault diagnosis applications, which outperforms other models of the same type.

Comparison with state-of-the-art deep learning methods for pitch fault diagnosis

In this paper, some popular fault diagnosis models are selected for comparative experimental validation in support of demonstrating the advantages of LAFA-Net. The Adam optimization algorithm with learning rate of 0.001 is used in all the training processes. The selected fault diagnosis models include:

(1) MS-Resnet: a multiscale residual network based on 1D-CNN, which improves the feature learning ability of the model by extracting spatial multiscale features from SCADA data, and is applied to fault diagnosis of WTs [9];
(2) SC-1DCNN: a self-calibrated 1D-CNN based on self-calibration, which can effectively learn shared features from data, and consider both marginal and conditional distributions for fault diagnosis models [13];
(3) MF-DCRN: a CNN with a multi-acceptance domain denoising block which is used to enhance the deep features and filter out the interfering feature information. The adaptive feature integration module is embedded in the CNN model to better utilize the extracted information, effectively realizing the fault diagnosis of gearboxes and industrial pumps [35].

Figure 15 shows the experimental accuracy curves of MS-Resnet, SC-1DCNN, MF-DCRN and LAFA-Net on the WT pitch fault dataset. It can be seen that on the basis of the same epochs, LAFA-Net converges significantly faster than the other models and can converge stably at a higher accuracy rate. From the comparison results of F-score, MIR and MIP for fault classification in Fig. 16, our method performs the best among the four types of models while MS-Resnet performs the worst. Both F-score, MIR, and MIP values, LAFA-Net reaches the highest value of 0.9989. In addition, the advantages of LAFA-Net compared to the other several classes of models can be more intuitively seen from the confusion matrix of the classification results in Fig. 17. Among the final results of the four classes of models, only LAFA-Net has more elements clustered on the diagonal in its confusion matrix. Overall, these results are a good validation that in general, our method has a great advantage in fault diagnosis oriented to actual WT pitch fault data.

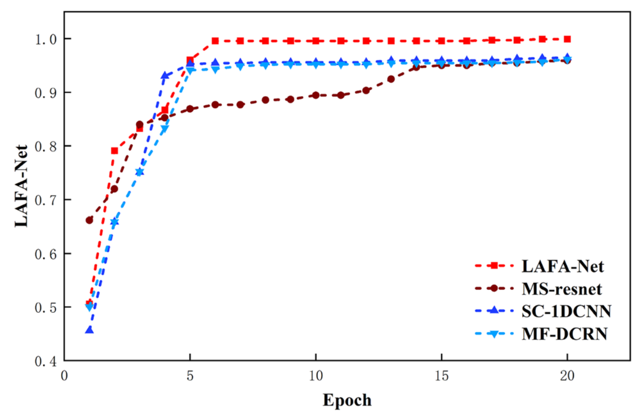


Fig. 15 Accuracy curves of MS-Resnet, SC-1DCNN, MF-DRCN, and LAFA-Net models for six single types of pitch fault diagnosis

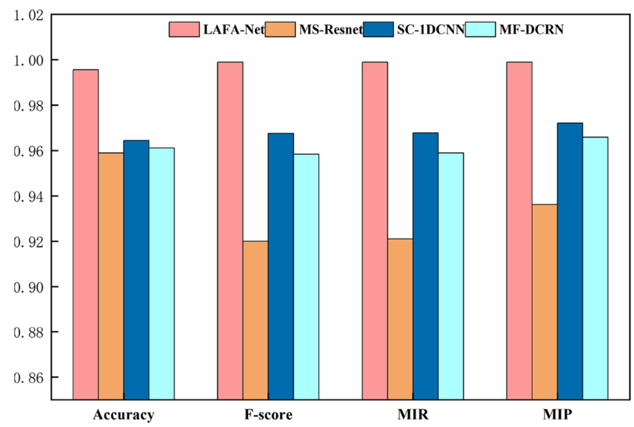


Fig. 16 F-score, MIR and MIP of MS-Resnet, SC-1DCNN, MF-DRCN, and LAFA-Net models for six single types of pitch fault diagnosis

Conclusion

In this paper, we have introduced a novel metric, namely, the Global Average Correlation Coefficient (GACC), for redundant feature screening. The proposed GACC effectively

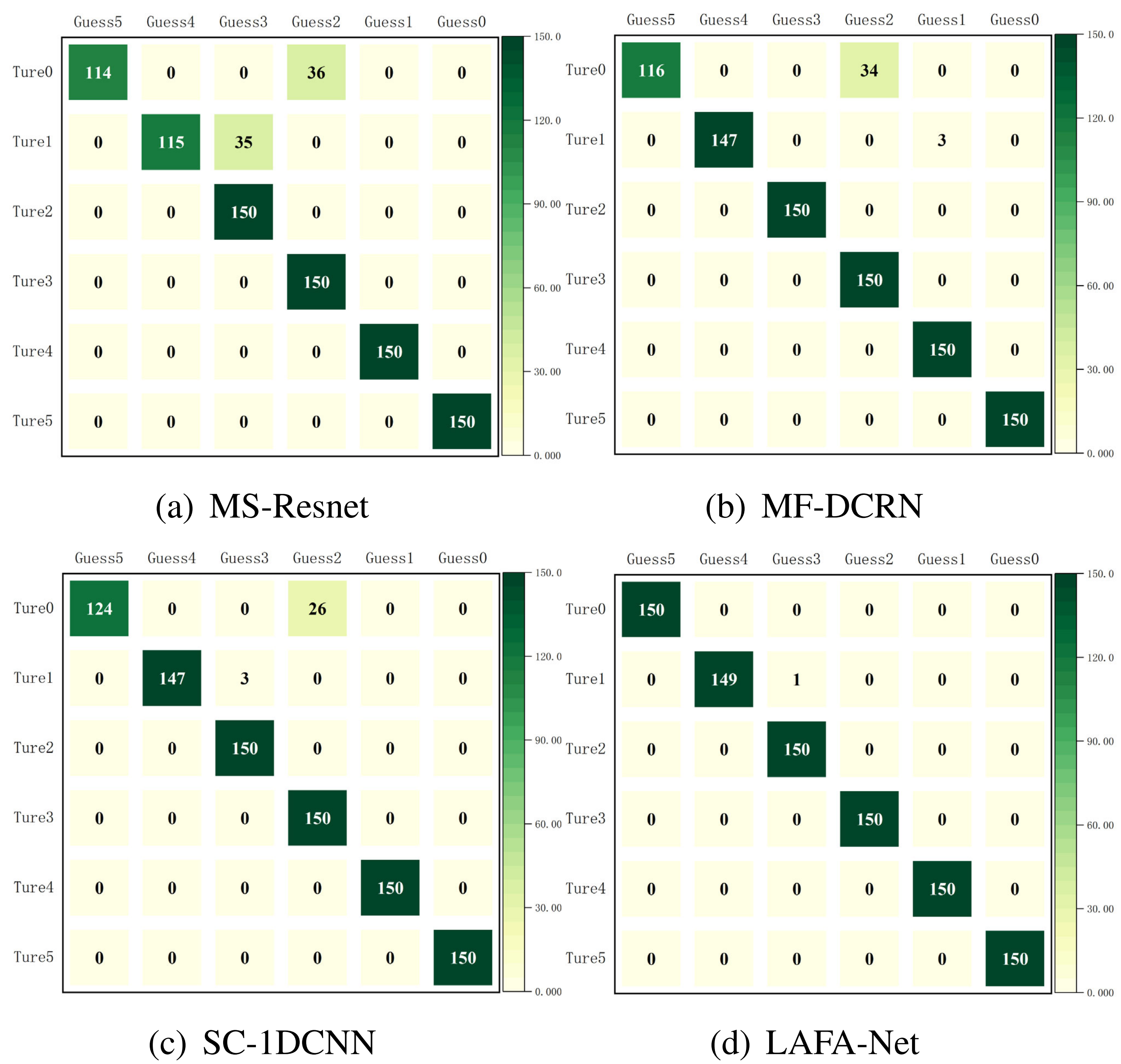


Fig. 17 Pitch fault diagnosis confusion matrices of MS-Resnet, SC-1DCNN, MF-DRCN, and LAFA-Net

synthesizes the degree of correlation between a single variable and other variables in a multivariate system, addressing the limitations of traditional PCC. The usage of GACC-screened datasets enhances the accuracy of various models in pitch fault diagnosis, ensuring that the most effective features in the dataset are utilized to their fullest potential. To address the issue of traditional deep CNNs often neglecting global context information when extracting local features, we have proposed a multi-attention based LAFA-Net and further introduced an LFEA module to comprehensively extract effective features of a single type of pitch fault. The LAFA-Net architecture has effectively integrated the strengths of convolutional networks in local feature learning and MHA in global feature learning, thereby realizing efficient pitch fault diagnosis. It has been demonstrated via comparative experiments with various models and datasets that the GACC-screened dataset has significantly improved the convergence speed and accuracy of fault diagnosis classification. The LAFA-Net model has shown superior accuracy and recognition rates in pitch fault diagnosis compared to several other models. These results have convincingly illustrated the considerable advantage of this new method in practical pitch fault diagnosis applications, which marks a significant advancement in the field. In the future, we aim to (1) apply the proposed LAFA-Net for fault diagnosis of other industrial processes e.g., manufacturing and medical engineering

[4, 10, 28, 36, 41, 47], (2) integrate model-driven methods to improve the detection system based on signal processing, state estimation, and filtering approaches [5, 17, 29, 53–56], (3) develop a new control strategy to adaptively extract patterns and features in a multi-modal manner [8, 14, 37, 45], and (4) employ evolutionary computation algorithms to optimize the hyperparameters of the LAFA-Net [3, 15, 19, 22, 31].

Funding This work was supported in part by the European Union's Horizon 2020 Research and Innovation Programme under Grant 820776 (INTEGRADDE), the Royal Society of the UK, the Alexander von Humboldt Foundation of Germany, the BRIEF Award of Brunel University London in theUK, and the Capacity Building Project of Shanghai Local Colleges and Universities of China under Grant 22010501100.

Data availability The data that support the findings of this study are available from the corresponding author, C. Wen, upon reasonable request.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Dou J, Song Y (2023) An improved generative adversarial network with feature filtering for imbalanced data. Int J Netw Dyn Intell 2(4):100017
2. Encalada-Dávila Á, Moyón L, Tutivén C, Puruncajas B, Vidal Y (2022) Early fault detection in the main bearing of wind turbines based on gated recurrent unit (GRU) neural networks and SCADA data. IEEE/ASME Trans Mech 27(6):5583–5593
3. Fang J, Liu W, Chen L, Lauria S, Miron A, Liu X (2023) A survey of algorithms, applications and trends for particle swarm optimization. Int J Netw Dyn Intell 2(1):24–50
4. Fang J, Wang Z, Liu W, Lauria S, Zeng N, Prieto C, Sikström F, Liu X (2024) A new particle swarm optimization algorithm for outlier detection: industrial data clustering in wire arc additive manufacturing. IEEE Trans Autom Sci Eng 21(2):1244–1257
5. Feng S, Li X, Zhang S, Jian Z, Duan H, Wang Z (2023) A review: state estimation based on hybrid models of Kalman filter and neural network. Syst Sci Control Eng 11(1):2173682
6. Gan M, Zhang L (2021) Iteratively local fisher score for feature selection. Appl Intell 51(8):6167–6181
7. Guha R, Ghosh M, Mutsuddi S, Sarkar R, Mirjalili S (2020) Embedded chaotic whale survival algorithm for filter-wrapper feature selection. Soft Comput 24(17):12821–12843
8. Han F, Liu J, Li J, Song J, Wang M, Zhang Y (2023) Consensus control for multi-rate multi-agent systems with fading measurements: the dynamic event-triggered case. Syst Sci Control Eng 11(1):2158959
9. He Q, Pang Y, Jiang G, Xie P (2021) A spatio-temporal multi-scale neural network approach for wind turbine fault diagnosis with imbalanced SCADA data. IEEE Trans Ind Inform 17(10):6875–6884
10. He G, Hu M, Shen Z (2023) Consensus of switched multi-agents system with cooperative and competitive relationship. Syst Sci Control Eng 11(1):2192008
11. Jiang L, Kong G, Li C (2021) Wrapper framework for test-cost-sensitive feature selection. IEEE Trans Syst Man Cybern Syst 51(3):1747–1756
12. Jin X, Xu Z, Qiao W (2021) Condition monitoring of wind turbine generators using SCADA data analysis. IEEE Trans Sustain Energy 12(1):202–210
13. Li S, Yu J (2022) Deep transfer network with adaptive joint distribution adaptation: a new process fault diagnosis mode. IEEE Trans Instrum Meas 71:3507813
14. Li X, Song Q, Zhao Z, Liu Y, Alsaadi FE (2022) Optimal control and zero-sum differential game for Hurwicz model considering singular systems with multifactor and uncertainty. Int J Syst Sci 53(7):1416–1435
15. Li H, Liu H, Lan C, Yin Y, Wu P, Yan C, Zeng N (2023) SMWO/D: a decomposition-based switching multi-objective whale optimiser for structural optimisation of turbine disk in aero-engines. Int J Syst Sci 54(8):1713–1728
16. Li X, Li M, Yan P, Li G, Jiang Y, Luo H, Yin S (2023) Deep learning attention mechanism in medical image analysis: basics and beyonds. Int J Netw Dyn Intell 2(1):93–116
17. Li X, Song Q, Liu Y, Alsaadi FE (2023) Saddle-point equilibrium for Hurwicz model considering zero-sum differential game of uncertain dynamical systems with jump. Int J Syst Sci 54(2):357–370
18. Liu H, Ren G, Dai T, Zhang D, Xu P, Zhang W, Hu B (2023) Diversity-oriented contrastive learning for RGB-T scene parsing, In: Proceedings of 2nd international conference on sensing, measurement, communication and internet of things technologies, Changsha, China
19. Ma G, Wang Z, Liu W, Fang J, Zhang Y, Ding H, Yuan Y (2023) Estimating the state of health for lithium-ion batteries: a particle swarm optimization-assisted deep domain adaptation approach. IEEE/CAA J Autom Sin 10(7):1530–1543
20. Patel SP, Upadhyay SH (2020) Euclidean distance based feature ranking and subset selection for bearing fault diagnosis. Expert Syst Appl 154
21. Qin W, Luo X, Zhou M (2024) Adaptively-accelerated parallel stochastic gradient descent for high-dimensional and incomplete data representation learning. IEEE Trans Big Data 10(1):92–107
22. Rostami M, Forouzandeh S, Berahmand K, Soltani M (2020) Integration of multi-objective PSO based feature selection and node centrality for medical datasets. Genomics 112(6):4370–4384
23. Shakiba FM, Shojaee M, Azizi SM, Zhou M (2022) Real-time sensing and fault diagnosis for transmission lines. Int J Netw Dyn Intell 1(1):36–47
24. Tong G, Li Q, Song Y (2023) Two-stage reverse knowledge distillation incorporated and self-supervised masking strategy for industrial anomaly detection. Knowl-Based Syst 273:110611
25. Tu T, Su Y, Tang Y, Guo G, Tan W, Ren S (2024) SHFW: second-order hybrid fusion weight—median algorithm based on machining

learning for advanced IoT data analytics. Wirel Netw https://doi.org/10.1007/s11276-023-03395-5 (in press)
26. Wang D, Cao C, Chen N, Pan W, Li H, Wang X (2022) A correlation-graph-CNN method for fault diagnosis of wind turbine based on state tracking and data driving model. Sustain Energy Technol Assess 56:102995
27. Wang C, Wang Z, Liu W, Shen Y, Dong H (2023) A novel deep offline-to-online transfer learning framework for pipeline leakage detection with small samples. IEEE Trans Instrum Meas 72:3503913
28. Wang D, Shi S, Lu J, Hu Z, Chen J (2023) Research on gas pipeline leakage model identification driven by digital twin. Syst Sci Control Eng 11(1):2180687
29. Wang Y-A, Shen B, Zou L, Han Q-L (2023) A survey on recent advances in distributed filtering over sensor networks subject to communication constraints. Int J Netw Dyn Intell 2(2):100007
30. Wang C, Wang Z, Ma L, Dong H, Sheng W (2023) A novel contrastive adversarial network for minor-class data augmentation: applications to pipeline fault diagnosis. Knowl-Based Syst 271:110516
31. Wang Y, Liu W, Wang C, Fadzil F, Lauria S, Liu X (2023) A novel multi-objective optimization approach with flexible operation planning strategy for truck scheduling. Int J Netw Dyn Intell 2(2):100002
32. Xiang L, Wang P, Yang X, Hu A, Su H (2021) Fault detection of wind turbine based on SCADA data analysis using CNN and LSTM with attention mechanism. Measurement 175:109094
33. Xie S, Zhang Y, Lv D, Chen X, Lu J, Liu J (2023) A new improved maximal relevance and minimal redundancy method based on feature subset. J Supercomput 79:3157–3180
34. Xu J, Tang B, He H, Man H (2017) Semisupervised feature selection based on relevance and redundancy criteria. IEEE Trans Neural Netw Learn Syst 28(9):1974–1984
35. Xu Y, Yan X, Sun B, Zhai J, Liu Z (2022) Multireceptive field denoising residual convolutional networks for fault diagnosis. IEEE Trans Ind Electron 69(11):11686–11696
36. Xu X, Lin M, Luo X, Xu Z (2023) HRST-LR: a Hessian regularization spatio-temporal low rank algorithm for traffic data imputation. IEEE Trans Intell Transport Syst 24(10):11001–11017
37. Yu L, Cui Y, Liu Y, Alotaibi ND, Alsaadi FE (2022) Sampled-based consensus of multi-agent systems with bounded distributed time-delays and dynamic quantisation effects. Int J Syst Sci 53(11):2390–2406
38. Yuan Y, Zhang H, Wu Y, Zhu T, Ding H (2016) Bayesian learning-based model-predictive vibration control for thin-walled workpiece machining processes. IEEE/ASME Trans Mech 22(1):509–520
39. Yuan Y, Tang X, Zhou W, Pan W, Li X, Zhang H-T, Ding H, Goncalves J (2019) Data driven discovery of cyber physical systems. Nat Commun 10(1):1–9
40. Yuan Y, Ma G, Cheng C, Zhou B, Zhao H, Zhang H-T, Ding H (2020) A general end-to-end diagnosis framework for manufacturing systems. Natl Sci Rev 7(2):418–429
41. Zeng N, Li X, Wu P, Li H, Luo X (2024) A novel tensor decomposition-based efficient detector for low-altitude aerial objects with knowledge distillation scheme. IEEE/CAA J Autom Sin 11(2):487–501
42. Zeng R, Qin Y, Song Y (2024) A non-iterative capsule network with interdependent agreement routing. Expert Syst Appl 238:122284
43. Zhang K, Tang B, Deng L, Yu X, Wei J (2021) Fault source location of wind turbine based on heterogeneous nodes complex network. Eng Appl Artif Intell 103:104300
44. Zhang G, Li Y, Zhao Y (2023) A novel fault diagnosis method for wind turbine based on adaptive multivariate time-series convolutional network using SCADA data. Adv Eng Inform 57:102031
45. Zhang Y, Zou L, Liu Y, Ding D, Hu J (2023) A brief survey on nonlinear control using adaptive dynamic programming under engineering-oriented complexities. Int J Syst Sci 54(8):1855–1872
46. Zhang Z, Zhou F, Zhang C, Wen C, Hu X, Wang T (2023) A personalized federated learning-based fault diagnosis method for data suffering from network attacks. Appl Intell 53:22834–22849
47. Zhang Y, Lan R, Li X, Fang J, Ping Z, Liu W, Wang Z Class imbalance wafer defect pattern recognition based on shared-database decentralized federated learning framework. *IEEE Trans Instrum Meas*https://doi.org/10.1109/TIM.2024.3395316 (in press)
48. Zheng J, Zhang S, Wang Z, Wang X, Zeng Z (2023) Multi-channel weight-sharing autoencoder based on cascade multi-head attention for multimodal emotion recognition. IEEE Trans Multimed 25:2213–2225
49. Zhong T, Qu J, Fang X, Li H, Wang Z (2021) The intermittent fault diagnosis of analog circuits based on EEMD-DBN. Neurocomputing 436:74–91
50. Zhou H, Guo J, Wang Y (2016) A feature selection approach based on term distributions. SpringerPlus 5(1):1–14
51. Zhou H, Wang X, Zhu R (2022) Feature selection based on mutual information with correlation coefficient. Appl Intell 52:5457–5474
52. Zhu Q, Yang Y (2018) Discriminative embedded unsupervised feature selection. Pattern Recognit Lett 112:219–225
53. Zou L, Wang Z, Geng H, Liu X (2021) Set-membership filtering subject to impulsive measurement outliers: a recursive algorithm. IEEE/CAA J Autom Sin 8(2):377–388
54. Zou L, Wang Z, Shen B, Dong H (2023) Encryption-decryption-based state estimation with multi-rate measurements against eavesdroppers: a recursive minimum-variance approach. IEEE Trans Autom Control 68(12):8111–8118
55. Zou L, Wang Z, Shen B, Dong H (2023) Moving horizon estimation over relay channels: dealing with packet losses. Automatica 155:111079
56. Zou L, Wang Z, Shen B, Dong H, Lu G (2023) Encrypted finite-horizon energy-to-peak state estimation for time-varying systems under eavesdropping attacks: tackling secrecy capacity. IEEE/CAA J Autom Sin 10(4):985–996