

## RESEARCH ARTICLE

# A Multi-View Spatio-Temporal Feature Fusion Approach for Wind Turbine Condition Monitoring Based on SCADA Data

HONG WANG<sup>1</sup>, HUI XIE, SHUWEI LIU, SONGSONG SONG, AND WEI HAN

School of Physics and Electronic Engineering, Hebei Normal University for Nationalities, Chengde 067000, China

Corresponding author: Wei Han (hwly2006@163.com)

This work was supported in part by the Science Research Project of Hebei Education Department under Grant BJK2022062, in part by the Science and Technology Research and Development Program of Chengde City under Grant 202102A067, in part by the Clean Energy (Carbon Peaking and Carbon Neutrality) Industry Research Institute of Chengde under Grant 202205B090, in part by the Doctoral Starting up Foundation of Hebei Normal University for Nationalities under Grant DR2021001, and in part by Chengde Science and Technology Program under Grant 202305B100.

**ABSTRACT** Condition monitoring of wind turbines is critical for increasing the reliability of the turbines and reducing their operation and maintenance costs. Supervisory control and data acquisition (SCADA) systems have been widely regarded as a promising technique to monitor the health status of turbines due to their abundance and cost-effective operation data. However, SCADA data are fundamentally multivariate time series with inherent spatio-temporal correlations. Therefore, it is still difficult to extract such correlations and then accurately identify the health status. This paper proposes a novel multi-view spatio-temporal feature fusion approach (MVSTCNN) based on convolutional neural networks (CNN) for condition monitoring of wind turbines. Specifically, multiple CNN modules with convolutional kernels of varying sizes are designed to extract correlations among several sensor variables and the temporal dependency concealed in each variable in parallel. A main advantage of the proposed method is its capacity to capture multiscale local information and global information simultaneously in both temporal and spatial dimensions, which improves the performance of condition monitoring. Real SCADA data from a wind farm is utilized to evaluate the effectiveness and superiority of the proposed approach. The SCADA data experiments demonstrate that the proposed approach is effective for early fault detection in wind turbines.

**INDEX TERMS** Condition monitoring, convolutional neural network, multi-view spatio-temporal feature fusion, SCADA data, wind turbine.

## I. INTRODUCTION

As a clean, pollution-free, and environmentally friendly renewable energy source, wind energy has become a major source of sustainable energy production and the reduction of greenhouse gas emissions. Wind energy has attracted worldwide attention and has enormous development potential. As a result, a massive number of wind turbines have been deployed both onshore and offshore. However, the majority of wind turbines are installed in distant locations and subjected to extreme weather and complex operating

conditions. Therefore, it is easy to cause frequent failures, eventually resulting in high downtime detection and operation and maintenance costs. Hence, advanced condition monitoring methods are in high demand to detect impending faults in wind turbines so as to prevent economic losses and accelerate the growth of the wind industry [1].

Numerous condition monitoring technologies for wind turbines have been developed to date, with the most common being vibration analysis [2], acoustic analysis [3], lubrication oil analysis [4], and others. However, these methods are limited in practical wind farm applications due to the requirement for additional measurement sensors and data acquisition equipment. Alternately, without more hardware investments,

The associate editor coordinating the review of this manuscript and approving it for publication was Senthil Kumar<sup>1</sup>.

supervisory control and data acquisition (SCADA) is considered a feasible and economical method for wind turbine condition monitoring and has received extensive attention in recent years. SCADA systems are now popularly equipped on most large-scale commercial wind turbines. A typical SCADA system contains fundamental state information and provides a large number of monitoring parameters connected with the operation condition of wind turbines, such as power, wind speed, current, and voltage, among others [5]. Therefore, the running status of wind turbines can be recognized and incipient failures can be detected by fully utilizing the plentiful SCADA data and extracting the useful hidden characteristics.

In previous studies, various machine learning and statistical algorithms, including artificial neural networks [6], [7], support vector machines [8], [9], cointegration analysis [10], performance curve-based methods [11], [12], Gaussian processes [13], [14], and nonlinear state estimation technology [15], have been proposed for condition monitoring of wind turbines using SCADA data. Obviously, due to the large amount of SCADA data and the typical shallow network structure, these traditional methods are difficult to handle and learn the complicated nonlinear relationships among complex monitoring information, resulting in limited detection performance. In recent years, deep learning, as a new emerging technology, has received increased attention in the domain of wind turbine condition monitoring because of its powerful ability to process large amounts of data and capture hierarchical and deep feature representation [16]. Deep learning-based approaches aim to stack multiple nonlinear processing layers in hierarchical designs to extract important and abstract information from data, which is well suited for analyzing SCADA data with high nonlinearity and correlations. Some academics have concentrated on using deep learning networks to monitor the condition of wind turbines. For instance, Wang et al. [17] developed a deep neural network-based condition monitoring model and identified the early faults of wind turbine gearboxes. Encalada-Dávila et al. [18] utilized a gated recurrent unit neural network to monitor the conditions of the wind turbine main bearing and realized incipient fault detection. Zhao et al. [19] proposed a deep automatic encoder (DAE) network to monitor wind turbines and implement early warning of fault components. Wang et al. [20] used DAEs to identify impending wind turbine blade breakage faults. Zhu et al. [21] introduced a convolutional neural network (CNN) and a long short-term memory network into the condition monitoring of wind turbine gearbox bearings and identified their impending faults.

The abovementioned deep learning applications have proven to be quite effective in feature learning and representation, and have improved anomaly detection results. However, due to the complicated spatial-temporal correlations that are inherent in the massive SCADA data, it is still a challenge

to effectively monitor the conditions of wind turbines. Specifically, as complicated electromechanical systems, wind turbines are composed of multiple subsystems and components, including gearboxes, generators, blades, principal shafts, bearings, etc. [22]. There are dependencies and interactions between different subsystems and components [23]. As a result, multiple relevant SCADA condition variables reflecting the health status of wind turbines are highly correlated, which means that SCADA data exhibit typical spatial correlation characteristics. Additionally, each condition variable is essentially a time series that will vary over time because of the influence of external operational environments. In other words, the measured values of each variable at different time points have strong temporal dependence, implying that SCADA data present representative temporal correlation characteristics. To address this challenge, inspired by the powerful nonlinear feature learning and representation capability of CNN, this paper proposes a CNN-based multi-view spatio-temporal feature fusion method, named MVSTCNN, to capture the complex spatio-temporal correlations concealed in SCADA multivariate time series from multiple perspectives. In particular, the innovative network can extract features on both temporal and spatial scales in parallel, and multi-scale local feature learning and global feature learning are also taken into account concurrently, allowing for the depth and comprehensive extraction of spatio-temporal features. The main contributions of this paper are summarized as follows.

(1) A new MVSTCNN network is presented to mine the spatio-temporal correlations inherent in SCADA data, and then a condition monitoring model using healthy data is developed. The MVSTCNN network does not only perform parallel spatio-temporal feature extraction but also captures and integrates local and global features on both temporal and spatial scales, which can improve condition monitoring performance.

(2) Multi-scale local temporal and spatial feature learning models are designed, respectively. The interactive and complementary temporal dependences and spatial correlations of SCADA data are extracted from multiple scales by setting multiple one-dimensional temporal and spatial convolution kernels with different sizes, enhancing the feature extraction capability of the proposed MVSTCNN network.

(3) Real SCADA datasets are employed to evaluate the performance of the proposed condition monitoring method, and contrast experiments are conducted.

The remainder of this article is organized as follows. Section II goes through the theoretical background of the MVSTCNN method. Section III presents the proposed SCADA data-based MVSTCNN condition monitoring framework. In Section IV, the effectiveness of the proposed monitoring method is demonstrated by a case study with actual SCADA data. Finally, conclusions are provided in Section V.

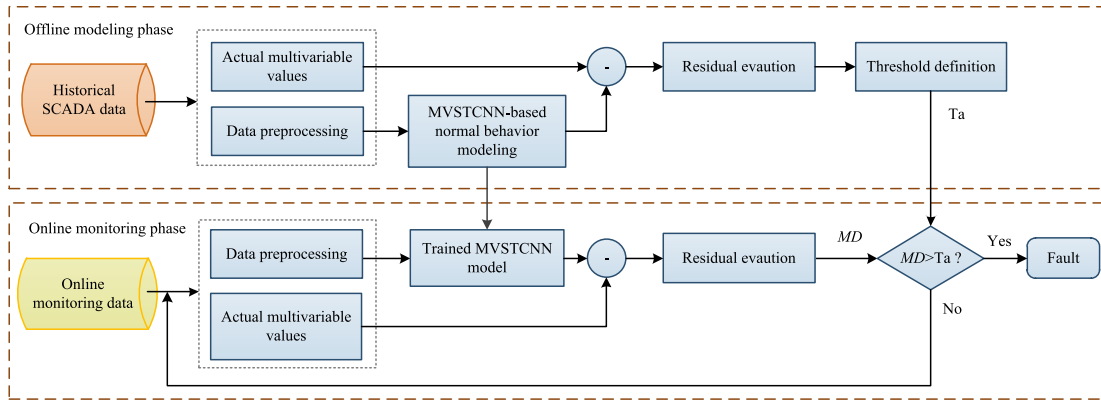


FIGURE 1. Flowchart of the proposed condition monitoring framework.

## II. OVERVIEW OF CONVOLUTIONAL NEURAL NETWORKS

CNNs are a specialized kind of multi-layer feed-forward artificial neural networks, which were originally proposed by LeCun et al. [24] for handwritten digit recognition. Inspired by biological neurology, CNNs are designed to imitate the behavior of the mammalian visual cortex. Compared with traditional fully connected neural networks, the key characteristics of CNNs are shared weights and translation invariance [25]. Due to their powerful capacity to automatically extract features, CNNs have been successfully used in a variety of challenging research domains, including computer vision, image classification, and natural language processing. A typical CNN is mainly composed of convolutional layers, pooling layers, and fully connecting layers. The convolutional layer extracts local features from input data by performing convolution operations with convolution kernels. Since only local parameters need to be computed, the convolution operation can dramatically optimize the number of parameters and make the learning layer simpler [26]. The convolution operation is defined as

$$x_j^l = f\left(\sum_i x_i^{l-1} * k_{ij}^l + b_j^l\right) \quad (1)$$

where  $x_j^l$  refers to  $j$ th feature map at the  $l$ th layer,  $x_i^{l-1}$  represents the  $i$ th input feature map at  $(l - 1)$ th layer,  $k_{ij}^l$  is the convolution kernel connecting  $i$ th input feature map with  $j$ th feature map,  $b_j^l$  is the bias term,  $*$  denotes the convolution operation, and  $f(\cdot)$  is a nonlinear activation function that can improve the expression ability of CNN. In this study, Rectified Linear Unit (ReLU) is selected as the activation function, which is expressed as  $f(x) = \max(0, x)$ .

Following that, the characteristics learned from the convolutional layer are fed into the pooling layer, thus preserving the most important information and enhancing computational efficiency. In the pooling process, maximum pooling or average pooling is often applied for pooling operations, and the maximum pooling is chosen in the paper to obtain the local maximum value. Next, local information is integrated by the

fully connecting layer, and each neuron in this layer has complete connectivity to all neurons in the previous layer.

## III. PROPOSED MVSTCNN CONDITION MONITORING FRAMEWORK

Fig. 1 depicts the overall flowchart of the presented framework for wind turbine condition monitoring. In practical applications, the majority of SCADA monitoring data are collected when wind turbines are operating normally, while faulty data are generally rare or even difficult to obtain. Therefore, the basic idea behind the proposed framework is that normal historical SCADA data are adopted to construct the normal behavior monitoring model. To this end, this section introduces a MVSTCNN method that aims to discover latent spatio-temporal correlations in normal operation data. Notably, the proposed condition monitoring scheme is based on the analysis of multivariate residuals between actual measurements and the predicted outputs from the well-trained normal behavior model. The changes in residuals can indicate the health state of wind turbines, indicating normal or potential anomalies. When wind turbines are in healthy operating conditions, low residual values are usually yielded because normal test data can well match the learned normal model. Conversely, test data with high residual values is identified as a fault or anomaly. To be specific, the detailed procedures for MVSTCNN-based wind turbine condition monitoring are described as follows.

(1) Offline modeling phase: Historical SCADA data from a healthy period are first obtained. Then, data preprocessing is required to enhance the condition monitoring performance. Further, the MVSTCNN method is applied to capture the normal behaviors of wind turbines and obtain sophisticated feature representations. At last, based on residuals, the monitoring indicator is determined and the alarm threshold is calculated for the early fault warning of wind turbines.

(2) Online monitoring phase: First, the newly collected measurements are preprocessed in the same manner as in the offline phase. The data are then sent to the well-trained MVSTCNN model to automatically capture spatio-temporal

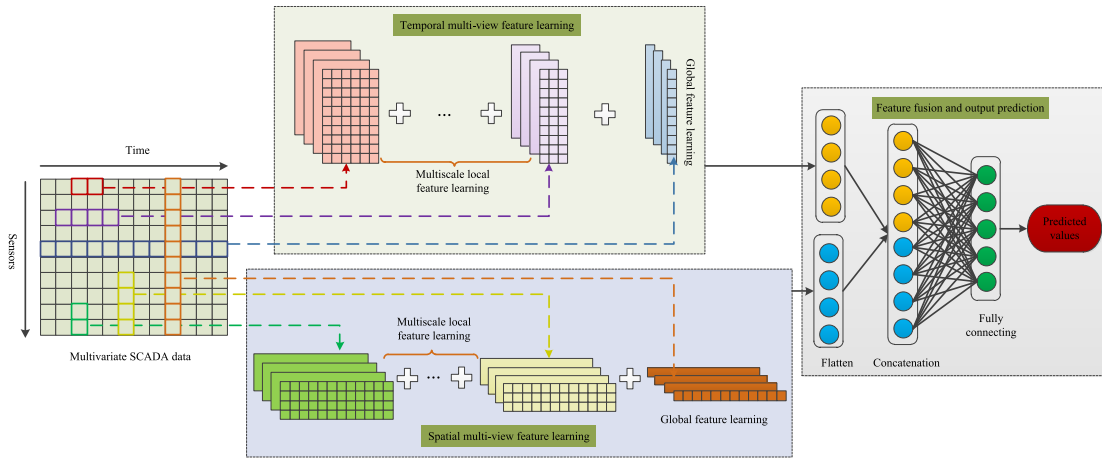


FIGURE 2. Architecture of the proposed MVSTCNN method.

characteristics and obtain corresponding residuals. After that, the residuals are further processed and compared with the defined threshold to identify the health status of wind turbines.

### A. VARIABLE SELECTION

In the wind turbine SCADA system, there are over one hundred sensor monitoring variables, include not only parameters that describe the operating conditions of wind turbines, like generator speed and power, but also parameters that represent the states of subsystems or components, such as temperature, voltage, current, etc. But not all of the variables are favorable for the establishment of the condition monitoring model. If all of these variables are input to the model at the same time, it will increase the computational complexity while decreasing the prediction accuracy. Hence, the selection of input variables should be carefully considered to screen out some important and valuable variables. In order to compress the amount of data and promote detection performance, the Pearson correlation coefficient is employed to assess the correlation between various monitoring variables in this paper. Variables with a strong correlation will be retained, whereas variables with a low or even no correlation will be discarded. The operation is represented as

$$r = \frac{N \sum XY - \sum X \sum Y}{\sqrt{N \sum X^2 - (\sum X)^2} \sqrt{N \sum Y^2 - (\sum Y)^2}} \quad (2)$$

where  $N$  denotes the variable sample size,  $X$  and  $Y$  are sensor variables.

In view of the fact that multiple monitoring variables in SCADA data have different units and value ranges, it makes sense to execute a data normalization step prior to input variable selection. The purpose is to rescale all variables to a specific range such that each variable contributes equally to the correlation coefficient calculation. According to [27], the

formula is as follows

$$\gamma_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)} \quad (3)$$

where  $x_{ij}$  denotes the  $i$ th measured value of variable  $j$ ,  $\min(x_j)$  and  $\max(x_j)$  denote the minimum and maximum values of variable  $j$ , respectively.  $\gamma_{ij}$  is the normalized value with a range of  $[0, 1]$ .

### B. PROPOSED MVSTCNN METHOD

The overall architecture of the proposed MVSTCNN method is presented in Fig. 2. A key property of this approach is that it automatically mines the spatial and temporal correlation information implied in complex SCADA data from different perspectives. Generally, the proposed architecture is mainly divided into three parts: temporal multi-view feature learning, spatial multi-view feature learning, and feature fusion and output prediction. The details of the proposed method are presented below.

#### 1) TEMPORAL MULTI-VIEW FEATURE LEARNING

To facilitate understanding, let matrix  $\mathbf{X} \in R^{S \times T}$  be input samples, with  $S$  representing the number of sensor variables and  $T$  denoting sampling points. It is well known that SCADA data itself is a time series, with each sensor variable changing over time. In order to better explore the temporal correlation inherent in each sensor variable, a temporal multi-view feature learning module based on the one-dimensional CNN is designed. This module is inspired by the inception structure [28] and consists of temporal multiscale local feature learning and temporal global feature learning. On the one hand, multiple one-dimensional temporal convolution kernels with different sizes are set in parallel to capture complementary and interactive local characteristics at different scales. In this subsection, according to the complexity of the model and experimental results, there are three scale channels, which are named CNN1, CNN2, and CNN3, respectively. As shown in Fig. 3, these CNN channels have three convolutional

layers followed by a pooling layer. In particular, CNN1, CNN2, and CNN3 each have convolution kernel sizes of  $1 \times 2$ ,  $1 \times 3$ , and  $1 \times 4$ , which are designed to extract data features between two, three, and four time points, respectively. Notably, it is challenging for a common CNN to capture the relationship between two data points that are far apart in the lower layers when the size of the convolution kernel is relatively small [29]. On the other hand, in order to overcome this shortcoming, the size of the temporal convolution kernel is set equal to the sampling points in terms of each sensor variable, i.e.,  $1 \times T$ , to mine the features between time points 1 and  $T$ . This can also be viewed as global feature learning in the temporal dimension and is named CNN4. Note that batch normalization (BN) layers are added to the four modules to decrease the number of parameters and prevent overfitting [30]. The characteristics yielded by these modules are then cascaded along the time axis. In this case, temporal multi-view information can be learned and obtained.

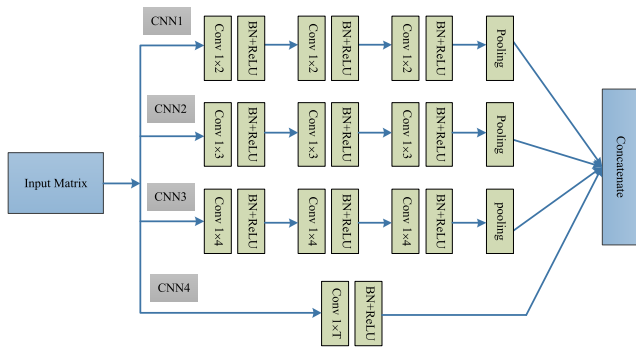


FIGURE 3. The structure of the temporal multi-view feature learning network.

### 2) SPATIAL MULTI-VIEW FEATURE LEARNING

In order to deeply mine the spatial correlations between various sensor variables, a spatial multi-view feature learning network is constructed in this subsection. Similar to temporal multi-view feature learning, one-dimensional CNN is adopted in this network, which includes two parts: spatial multiscale local feature learning and spatial global feature learning. However, the difference is that several CNNs with spatial convolution kernels are designed. In other words, convolution operations only slide along the spatial dimension of multivariate SCADA time series. In the aspect of spatial multiscale local feature learning, three different scale channels, including CNN5, CNN6, and CNN7, are used to capture interactive and rich characteristics between multiple sensor variables in parallel. The convolution kernel sizes of the three channels are  $2 \times 1$ ,  $3 \times 1$ , and  $4 \times 1$ , respectively, with the intention of extracting correlations between two variables, three variables, and four variables. Likewise, each channel is composed of three convolutional layers in series and a subsequent pooling layer. Additionally, to execute spatial global feature extraction, the length of the one-dimensional spatial convolution kernel is set to the number of sensor variables.

This implies that the convolution kernel with a size of  $S \times 1$  fuses variable 1 to variable  $S$  at each point in time to learn the characteristics between all variables. And this learning phase is represented by the CNN8 module. During the feature extraction process, the BN layer is also applied after each convolutional layer. Finally, the spatial correlations generated by each CNN module are cascaded along the variable axis for further condition monitoring. The schematic diagram of the spatial multi-view feature learning network is displayed in Fig. 4.

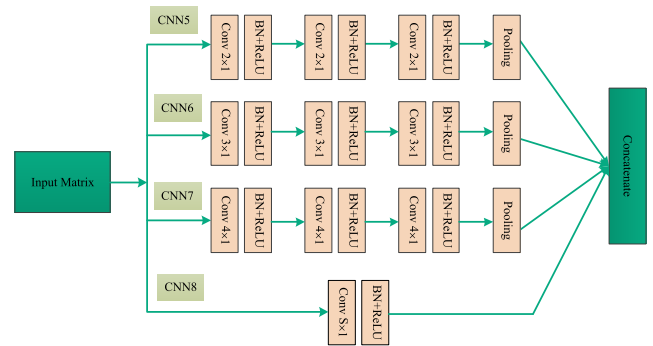


FIGURE 4. Schematic diagram of the spatial multi-view feature learning network.

### 3) FEATURE FUSION AND OUTPUT PREDICTION

Because the temporal and spatial multi-view characteristics learned in the first two stages have inconsistent dimensions, they are separately fed into the flatten layer to produce the consistent dimension. The transformed features are then cascaded, and the resulting multi-view spatio-temporal information is taken as the input of the next fully connecting layer. A regression output layer that has neurons with the same number of sensor variables follows immediately behind the fully connecting layer and is used for prediction. In particular, similar to multilayer perceptron networks, all neurons in the fully connecting layer and its adjacent layers are globally connected. For the training of the proposed model, the mean squared error between the predicted values and the actual values is employed as the loss function, which is optimized by the stochastic gradient descent with momentum algorithm. Given  $N$  training samples, the loss function is described as

$$H = \frac{1}{N} \sum_{i=1}^N (Y' - Y)^2 \quad (4)$$

where  $H$  represents the loss function,  $Y'$  and  $Y$  denote the predicted values and actual values, respectively.

### C. FAULT DETECTION APPROACH

The establishment of the MVSTCNN normal behavior model is intended to constantly monitor the running state of wind turbines and identify forthcoming early breakdowns. This is important for preventing major failures and improving the

reliability of wind turbines. As mentioned above, the analysis of the residual between the actual measurements and the estimated values underlies the fault detection method in the study. The MVSTCNN approach takes the multivariate SCADA data from  $t_1$  to  $t_T$  as input  $\mathbf{X}$ , and the target output  $\mathbf{Y}$  is the values at the next data point  $t_{T+1}$  of the time series. Setting the input and output in such a way will aid the model in discovering more implicit information, thus enhancing its generalization ability and prediction performance [31]. The input and output can be defined as

$$\mathbf{X} = \begin{bmatrix} x_{t_1}^1 & x_{t_2}^1 & \cdots & x_{t_T}^1 \\ x_{t_1}^2 & x_{t_2}^2 & \cdots & x_{t_T}^2 \\ \vdots & \vdots & \vdots & \vdots \\ x_{t_1}^S & x_{t_2}^S & \cdots & x_{t_T}^S \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} x_{t_{T+1}}^1 \\ x_{t_{T+1}}^2 \\ \vdots \\ x_{t_{T+1}}^S \end{bmatrix}' \quad (5)$$

where  $x_{t_i}^j$  is the data of  $j$ th sensor variable at  $i$ th point in time.

To identify the anomalies in wind turbines effectively, it is essential to select an advanced statistical method to determine the monitoring indicator. As a unit-less distance metric method, Mahalanobis distance (MD) has the advantage of considering the correlation of variables and transforming multivariate data into a univariate distance value. MD has been successfully applied to identify wind turbine anomalies [6], [32]. Hence, the MD measure is adopted in this study to derive the monitoring indicator for anomaly detection. The monitoring value  $MD_i$  of the  $i$ th sample data is expressed as follows

$$MD_i = \sqrt{(\mathbf{E}_i - \boldsymbol{\mu})\mathbf{C}^{-1}(\mathbf{E}_i - \boldsymbol{\mu})^T} \quad (6)$$

where  $\mathbf{E}_i$  denotes the residual vector of the  $i$ th sample,  $\boldsymbol{\mu}$  and  $\mathbf{C}^{-1}$  are the mean value and inverse covariance matrix of residuals of all samples, respectively.

The threshold for anomaly detection is calculated based on the MD values for residuals obtained in the validation phase. At this stage, wind turbines are operating normally, and there is no abnormal behavior. The  $MD_j$  of the  $j$ th sample for the validation data is defined as follows

$$MD_{refj} = \sqrt{(\mathbf{E}_{refj} - \boldsymbol{\mu}_{ref})\mathbf{C}_{ref}^{-1}(\mathbf{E}_{refj} - \boldsymbol{\mu}_{ref})^T} \quad (7)$$

where  $\mathbf{E}_{refj}$  represents the residual vector of the  $j$ th sample of the validation data.  $\boldsymbol{\mu}_{ref}$  and  $\mathbf{C}_{ref}^{-1}$  are the mean value and the inverse covariance matrix of validation residuals, respectively.

According to equation (7), the monitoring indicator values of all validation data can be achieved. Next, kernel density estimation (KDE) is used to evaluate the probability density distribution of these indicator values to determine the anomaly detection threshold. KDE is a common nonparametric estimation technique that has been widely used in the domains of anomaly detection and process monitoring [33], [34].

Assuming  $x$  is a random variable, the estimated probability density function is given as

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (8)$$

where  $x_i$  is the given data sample,  $n$  is the number of samples,  $h$  is the bandwidth parameter, and  $K(\cdot)$  is the kernel function.

The Gaussian kernel function is adopted in this paper and is described as

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} \quad (9)$$

The known probability density function can then be used to calculate the fault detection threshold  $T_a$  at a given confidence level  $\alpha$  by

$$\hat{f}_h(x < T_a) = \int_{-\infty}^{T_a} \hat{f}_h(x) dx = \alpha \quad (10)$$

In the process of condition monitoring, if the monitoring indicator value  $MD_i$  passes the threshold  $T_a$ , an alert signal will be generated. This can push operators to pay attention to the operating status of wind turbines and take the necessary precautions to prevent serious failures.

#### IV. CASE STUDY

In this section, the proposed approach is applied to the actual generator condition monitoring, and the implementation results and comparative experiments are displayed and analyzed in detail.

##### A. DATA DESCRIPTION

The SCADA data for this study came from an actual wind farm in Inner Mongolia, China. This wind farm comprises over 100 identical wind turbines with a nominal power of 1.5 MW. SCADA systems have been installed on all wind turbines, and SCADA data is sampled at 30-s intervals. These SCADA data log a variety of sensor measurements related to the operation condition of wind turbines, such as active power, generator speed, the temperature of components, etc. For the majority of the turbines, the SCADA dataset was available from July 1 to September 23, 2014.

In order to accurately identify potential faults, the training data for normal behavior modeling should cover all normal operation zones of turbines as much as possible. According to SCADA system records, there was no abnormal behavior of generators in turbines 6, 17, 24, 33, 34, 49, and 53 from July to September. Therefore, the healthy SCADA data collected from these turbines are considered for modeling. Whereas turbine 28 suffered a generator speed anomaly failure on August 8, and the data prior to the failure are taken as the testing data to verify the detection performance of the model.

##### B. MODEL DEVELOPMENT

In order to capture the normal behavior of the generator, several valuable input variables should be carefully chosen. Due to the close relationship between generator speed and the

health of the generator, the variables most relevant to generator speed are determined as the input for modeling in this paper. The Pearson correlation coefficients are calculated, and all data are linearly normalized to the range of [0, 1] prior to the calculation. Table 1 provides the selected variables.

TABLE 1. Selected modeling variables.

No.	Sensor Variable	Unit	Correlation Coefficient
1	Generator speed	r/min	1
2	Gearbox speed	r/min	1
3	Converter side speed	r/min	0.9954
4	Gearbox front bearing temperature	°C	0.7695
5	Gearbox rear bearing temperature	°C	0.7389
6	Wind speed	m/s	0.7078
7	Current phase C	A	0.7010
8	Current phase B	A	0.6982
9	Converter side torque	Nm	0.6961
10	Current phase A	A	0.6923
11	Active power	kW	0.6804
12	Gearbox oil temperature	°C	0.6563
13	Generator torque	Nm	0.6284

It should be noted that moving window processing is used to generate the multivariate time series input matrix combining temporal and spatial information. The historical healthy operating data is separated into a group of fragments with a moving time window of 1 hour without overlap, which indicates that there are 120 data points included in the moving window. A total of 1868 samples are obtained by preprocessing, and each one is continuous in time. Then, 1400 samples are chosen at random as the training data, with the remainder serving as the validation data. As mentioned in Section III-B, several different CNN modules are designed in the proposed approach to effectively extract the spatio-temporal characteristics of normal behavior. In addition, there are a series of other parameters that also need to be set in the model training process. Table 2 presents the detailed network structure settings.

The monitoring indicator values for the validation data are derived based on the established MVSTCNN model. Fig. 5 presents the histogram and estimated probability density function of the monitoring values using the KDE technique.

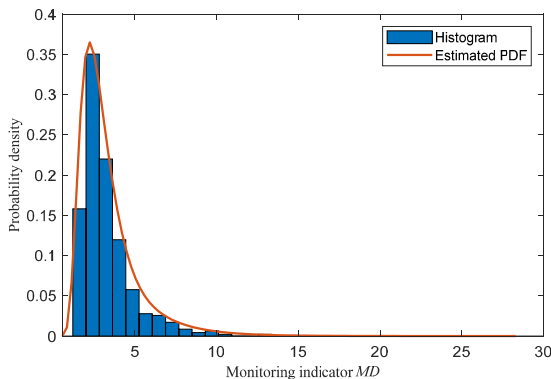


FIGURE 5. Histogram and estimated PDF with KDE for monitoring values obtained from the validation data.

TABLE 2. Details of the MVSTCNN framework.

Describe	Parameter Setting
Input	13×120
Mini-batch size	5
Initial learning rate	0.0001
Max epochs	20
Temporal multi-view feature learning	[1×2, filter 16, conv]×3, stride 1×1
	[1×2, maxpool], stride 1×2
	[1×3, filter 16, conv]×3, stride 1×1
	[1×2, maxpool], stride 1×2
	[1×4, filter 16, conv]×3, stride 1×1
Spatial multi-view feature learning	[1×2, maxpool], stride 1×2
	[1×120, filter 16, conv], stride 1×1
	[2×1, filter 16, conv]×3, stride 1×1
	[2×1, maxpool], stride 2×1
	[3×1, filter 16, conv]×3, stride 1×1
Output	[2×1, maxpool], stride 2×1
	[4×1, filter 16, conv]×3, stride 1×1
	[2×1, maxpool], stride 2×1
Output	13

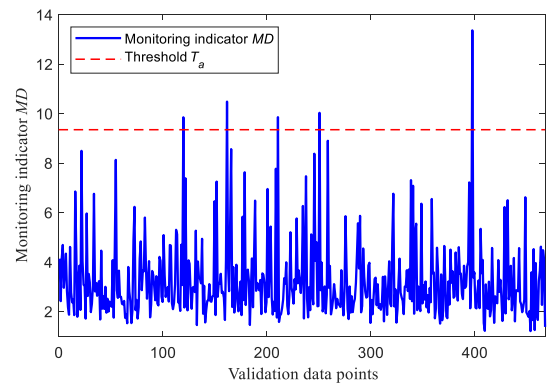
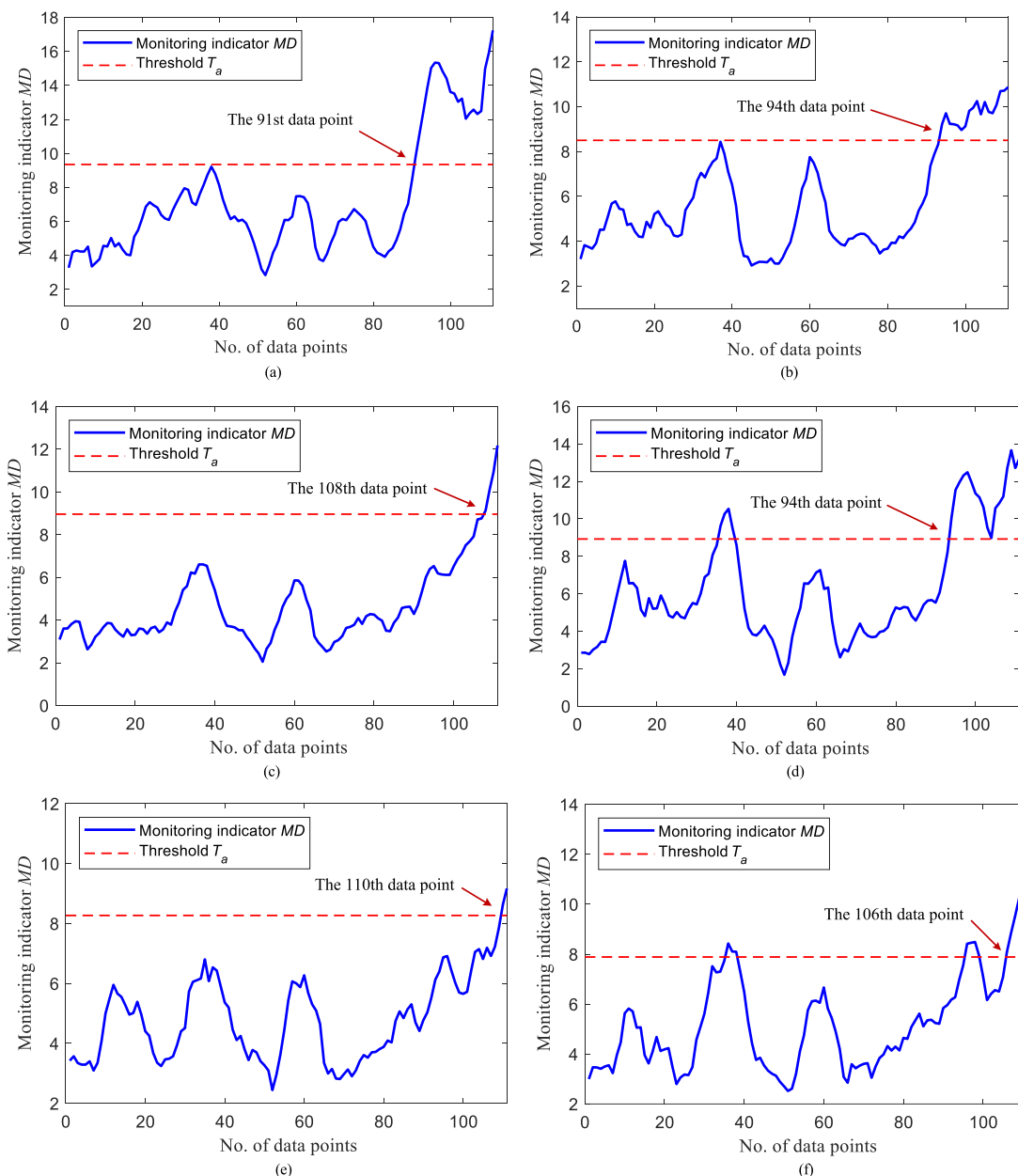


FIGURE 6. The monitoring values of validation data and corresponding threshold.

As can be seen, the estimated probability density function fits the actual distribution well, making it appropriate for further determination of the fault detection threshold according to (10). The monitoring indicator values for the validation data and the corresponding threshold are shown in Fig. 6. In this paper, the confidence level is set at 99%.

C. ABNORMAL DETECTION RESULT

According to the event logs of the wind turbine, Turbine 28 suffered a generator speed abnormal fault on August 8, 2014. The data from August 4 to 8 before this event are set as testing data for abnormal detection. The already-identified modeling variables are first selected from the test samples. As with the preprocessing of the healthy data, the selected data are then rescaled and split into matrices with temporal and spatial dimensions to construct corresponding test inputs and outputs. The residuals are derived based on the predicted values of the testing data with the trained MVSTCNN model. The monitoring indicator values are further computed to monitor the condition of the generator. In order to continuously reflect the trend of the monitoring values and eliminate the impact of random interference, a moving average calculation



**FIGURE 7. Condition monitoring results of different models (a) MVSTCNN (b) MSTCNN (c) MTCNN (d) MSCNN (e) STCNN (f) SSCNN.**

is used for statistical analysis. In this paper, a moving window with a size of 5 is adopted through experiments. Fig. 7(a) presents the monitoring result with the MVSTCNN model.

As illustrated in Fig. 7(a), the monitoring indicator value surpasses the fault threshold  $T_a$  in the 91st data point and continues to fluctuate above the threshold for a period. Considering the moving window size is 5, the threshold is actually exceeded in the 96th data point. According to analysis, an abnormality is detected 19 hours ahead of the speed anomaly failure of the generator.

To verify the ability of the MVSTCNN approach, comparative experiments are conducted on multiple CNN models with

different structures, including spatio-temporal multiscale CNN (MSTCNN), temporal multiscale CNN (MTCNN), spatial multiscale CNN (MSCNN), temporal single-scale CNN (STCNN), and spatial single-scale CNN (SSCNN). All of these models have the same structure and settings as the proposed method, except that certain specific modules are discarded. In terms of the MSTCNN, it takes into account the local temporal and spatial correlations on multiple scales simultaneously. Whereas the MTCNN and MSCNN only extract multiscale local features from temporal and spatial dimensions, respectively. For the latter two methods, local dependencies on a single scale are considered. The results are displayed in Figs. 7(b)–(f).



As can be seen from Figs. 7(c) and (e), compared to the MD value of the STCNN model, which first continuously crosses the  $T_a$  in the 110th data point, the MTCNN exceeds the  $T_a$  in the 108th data point, and both are maintained for a period. We can conclude that the MTCNN is able to detect the upcoming failure two hours before the STCNN. It can be seen from Fig. 7(d) that the MD value crosses the  $T_a$  in the 36-39th data points, whereas after the 39th point, the MD value drops below the  $T_a$  and continues for a period. Until the 94th point, the MD exceeds the  $T_a$  again and remains. From Fig. 7(f), we can observe that similar to the monitoring result of the MSCNN, the MD of the SSCNN model comes across the  $T_a$  at points 36-38 and 96-98, respectively. Nevertheless, the starting point that consistently exceeds the  $T_a$  appears at the 106th data point. Combining Figs. 7(d) and (f), we can consider that the noise and disturbance interfere with the monitoring results and lead to false alarms before the anomaly is actually detected. Whereas from the perspective of truly identifying the potential faults, the MSCNN model can be 12 hours ahead of the SSCNN model. From the result of Fig. 7(b), we can observe that the MD fluctuates below the  $T_a$  before the 94th data point and then continuously passes the  $T_a$ . It shows that the MSTCNN is capable of detecting the fault 16 hours in advance, but 3 hours behind the proposed MVSTCNN shown in Fig. 7(a). The comparative monitoring results of different models are summarized in Table 3.

**TABLE 3. Monitoring performance of different models.**

Model	Early warning time (h)	Times of false alarm
STCNN	0	0
SSCNN	4	6
MTCNN	2	0
MSCNN	16	4
MSTCNN	16	0
Proposed	19	0

From the comparative experiment, we can observe that the performance of the multiscale feature extraction models MTCNN and MSCNN outperforms the STCNN and SSCNN that only consider a single scale because of their capability to learn interactive and complementary temporal and spatial features. For the MSTCNN model, it incorporates both temporal and spatial information at different scales, so it can detect anomalies earlier than the MTCNN and avoid false alarms compared with the MSCNN. In particular, compared with the MSTCNN model, the proposed MVSTCNN further integrates global information that takes into account the correlations of all time points and all sensor variables, it thus can extract more valuable features and provide earlier warning for impending faults. In this case, there is more time available for operators to take appropriate measures to prevent major failures.

## V. CONCLUSION

This paper presents an innovative multi-view spatio-temporal feature fusion approach called MVSTCNN for monitoring the operating status of wind turbines. This model is based on

convolutional neural networks and extracts the inherent temporal and spatial information from SCADA multivariate time series in a parallel manner. One main advantage of the proposed method is the design of the multiscale local temporal and spatial feature learning modules to extract rich and complementary spatio-temporal features. The other important contribution is the inclusion of the global feature extraction modules, which capture the correlations of all time points and all sensor variables. The proposed model can enhance feature extraction ability and improve condition monitoring performance. In order to effectively identify impending faults in wind turbines, the MD of the residuals is computed as the monitoring indicator. And the fault threshold is determined using the monitoring indicator during normal operation.

The performance of the proposed approach is verified using SCADA data from a real wind farm. Compared with other CNN-based models, the proposed method has the ability to capture more valuable information, and it is more effective to detect anomalies without generating false alarms and achieve earlier warning. This means that the MVSTCNN model has considerable potential in practical wind turbine condition monitoring applications, which can guarantee the reliable operation of wind turbines and lessen economic losses. However, the current work has its limitations. To continuously reflect the operating state of wind turbines, extensive and long-term SCADA data needs to be acquired. In the next work, massive amounts of available SCADA data will be collected for long-term health monitoring. Moreover, it makes sense to focus on fault isolation methods to discover the underlying cause of the defect, thereby offering decision-making for wind turbine maintenance.

## REFERENCES

- [1] W. Udo and Y. Muhammad, "Data-driven predictive maintenance of wind turbine based on SCADA data," *IEEE Access*, vol. 9, pp. 162370–162388, 2021.
- [2] S. Koukoura, J. Carroll, A. McDonald, and S. Weiss, "Comparison of wind turbine gearbox vibration analysis algorithms based on feature extraction and classification," *IET Renew. Power Gener.*, vol. 13, no. 14, pp. 2549–2557, Oct. 2019.
- [3] S. Souza, P. van Lieshout, A. Perera, T.-H. Gan, and B. Bridge, "Determination of the combined vibrational and acoustic emission signature of a wind turbine gearbox and generator shaft in service as a pre-requisite for effective condition monitoring," *Renew. Energy*, vol. 51, pp. 175–181, Mar. 2013.
- [4] P. Sanchez, D. Mendizabal, K. Gonzalez, C. R. Zamarreño, M. Hernaez, I. R. Matias, and F. J. Arregui, "Wind turbines lubricant gearbox degradation detection by means of a lossy mode resonance based optical fiber refractometer," *Microsyst. Technol.*, vol. 22, no. 7, pp. 1619–1625, Jul. 2016.
- [5] J. Dai, Y. Tan, W. Yang, L. Wen, and X. Shen, "Investigation of wind resource characteristics in mountain wind farm using multiple-unit SCADA data in chenzhou: A case study," *Energy Convers. Manage.*, vol. 148, pp. 378–393, Sep. 2017.
- [6] P. Bangalore, S. Letzgus, D. Karlsson, and M. Patriksson, "An artificial neural network-based condition monitoring method for wind turbines, with application to the monitoring of the gearbox," *Wind Energy*, vol. 20, no. 8, pp. 1421–1438, Aug. 2017.
- [7] Z. Luo, C. Liu, and S. Liu, "A novel fault prediction method of wind turbine gearbox based on pair-copula construction and BP neural network," *IEEE Access*, vol. 8, pp. 91924–91939, 2020.
- [8] C. Yang, J. Liu, Y. Zeng, and G. Xie, "Real-time condition monitoring and fault detection of components based on machine-learning reconstruction model," *Renew. Energy*, vol. 133, pp. 433–441, Apr. 2019.

- [9] Y. Yang, A. Liu, H. Xin, and J. Wang, "Fault early warning of wind turbine gearbox based on multi-input support vector regression and improved ant lion optimization," *Wind Energy*, vol. 24, no. 8, pp. 812–832, Aug. 2021.
- [10] P. B. Dao, W. J. Staszewski, T. Barszcz, and T. Uhl, "Condition monitoring and fault detection in wind turbines based on cointegration analysis of SCADA data," *Renew. Energy*, vol. 116, pp. 107–122, Feb. 2018.
- [11] Y. Hu, Y. Xi, C. Pan, G. Li, and B. Chen, "Daily condition monitoring of grid-connected wind turbine via high-fidelity power curve and its comprehensive rating," *Renew. Energy*, vol. 146, pp. 2095–2111, Feb. 2020.
- [12] F. Bilendo, H. Badihi, N. Lu, P. Cambron, and B. Jiang, "A normal behavior model based on power curve and stacked regressions for condition monitoring of wind turbines," *IEEE Trans. Instrum. Meas.*, vol. 71, 2022, Art. no. 3520013.
- [13] Y. Li, S. Liu, and L. Shu, "Wind turbine fault diagnosis based on Gaussian process classifiers applied to operational data," *Renew. Energy*, vol. 134, pp. 357–366, Apr. 2019.
- [14] N. Chen, R. Yu, Y. Chen, and H. Xie, "Hierarchical method for wind turbine prognosis using SCADA data," *IET Renew. Power Gener.*, vol. 11, no. 4, pp. 403–410, Mar. 2017.
- [15] P. Guo, D. Infield, and X. Yang, "Wind turbine generator condition-monitoring using temperature trend analysis," *IEEE Trans. Sustain. Energy*, vol. 3, no. 1, pp. 124–133, Jan. 2012.
- [16] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mech. Syst. Signal Process.*, vol. 115, pp. 213–237, Jan. 2019.
- [17] L. Wang, Z. Zhang, H. Long, J. Xu, and R. Liu, "Wind turbine gearbox failure identification with deep neural networks," *IEEE Trans. Ind. Informat.*, vol. 13, no. 3, pp. 1360–1368, Jun. 2017.
- [18] Á. Encalada-Dávila, L. Moyón, C. Tutivén, B. Puruncajas, and Y. Vidal, "Early fault detection in the main bearing of wind turbines based on gated recurrent unit (GRU) neural networks and SCADA data," *IEEE/ASME Trans. Mechatronics*, vol. 27, no. 6, pp. 5583–5593, Dec. 2022.
- [19] H. Zhao, H. Liu, W. Hu, and X. Yan, "Anomaly detection and fault analysis of wind turbine components based on deep learning network," *Renew. Energy*, vol. 127, pp. 825–834, Nov. 2018.
- [20] L. Wang, Z. Zhang, J. Xu, and R. Liu, "Wind turbine blade breakage monitoring with deep autoencoders," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 2824–2833, Jul. 2018.
- [21] A. Zhu, Q. Zhao, T. Yang, L. Zhou, and B. Zeng, "Condition monitoring of wind turbine based on deep learning networks and kernel principal component analysis," *Comput. Electr. Eng.*, vol. 105, Jan. 2023, Art. no. 108538.
- [22] W. Qiao and D. Lu, "A survey on wind turbine condition monitoring and fault diagnosis—Part I: Components and subsystems," *IEEE Trans. Ind. Electron.*, vol. 62, no. 10, pp. 6536–6545, Oct. 2015.
- [23] W. Yang, R. Court, and J. Jiang, "Wind turbine condition monitoring by the approach of SCADA data analysis," *Renew. Energy*, vol. 53, pp. 365–376, May 2013.
- [24] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [25] C.-L. Liu, W.-H. Hsiao, and Y.-C. Tu, "Time series classification with multivariate convolutional neural network," *IEEE Trans. Ind. Electron.*, vol. 66, no. 6, pp. 4788–4797, Jun. 2019.
- [26] L. Xiang, X. Yang, A. Hu, H. Su, and P. Wang, "Condition monitoring and anomaly detection of wind turbine based on cascaded and bidirectional deep learning networks," *Appl. Energy*, vol. 305, Jan. 2022, Art. no. 117925.
- [27] H.-H. Yang, M.-L. Huang, C.-M. Lai, and J.-R. Jin, "An approach combining data mining and control charts-based model for fault detection in wind turbines," *Renew. Energy*, vol. 115, pp. 808–816, Jan. 2018.
- [28] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 1–9.
- [29] L. Chen, G. Xu, Q. Zhang, and X. Zhang, "Learning deep representation of imbalanced SCADA data for fault detection of wind turbines," *Measurement*, vol. 139, pp. 370–379, Jun. 2019.
- [30] C. Szegedy and S. Ioffe, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. ICML*, Lille, France, 2015, pp. 448–456.
- [31] Z. Kong, B. Tang, L. Deng, W. Liu, and Y. Han, "Condition monitoring of wind turbines based on spatio-temporal fusion of SCADA data by convolutional neural networks and gated recurrent units," *Renew. Energy*, vol. 146, pp. 760–768, Feb. 2020.
- [32] G. Jiang, P. Xie, H. He, and J. Yan, "Wind turbine fault detection using a denoising autoencoder with temporal information," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 1, pp. 89–100, Feb. 2018.
- [33] J. Chen, J. Li, W. Chen, Y. Wang, and T. Jiang, "Anomaly detection for wind turbines based on the reconstruction of condition parameters using stacked denoising autoencoders," *Renew. Energy*, vol. 147, pp. 1469–1480, Mar. 2020.
- [34] P. P. Odiwei and Y. Cao, "Nonlinear dynamic process monitoring using canonical variate analysis and kernel density estimations," *IEEE Trans. Ind. Informat.*, vol. 6, no. 1, pp. 36–45, Feb. 2010.



**HONG WANG** received the Ph.D. degree in control science and engineering from Yanshan University, Qinhuangdao, China, in 2021.

She is currently a Lecturer with the School of Physics and Electronic Engineering, Hebei Normal University for Nationalities, Chengde, China. Her research interests include condition monitoring of wind turbines, data mining, and deep learning.



**HUI XIE** received the Ph.D. degree in condensed matter physics from Jilin University, Changchun, China, in 2020.

She is currently a Lecturer with the School of Physics and Electronic Engineering, Hebei Normal University for Nationalities, Chengde, China. Her research interests include machine learning and data mining.



**SHUWEI LIU** received the master's degree in electrical engineering from Xinjiang University, Urumqi, China, in 2018.

He is currently a Lecturer with the School of Physics and Electronic Engineering, Hebei Normal University for Nationalities, Chengde, China. His research interest includes fault diagnosis in power systems.



**SONGSONG SONG** received the Ph.D. degree in power engineering and engineering thermophysics from Beijing University of Technology, Beijing, China, in 2017.

She is currently an Associate Professor with the School of Physics and Electronic Engineering, Hebei Normal University for Nationalities, Chengde, China. Her research interests include machine learning and equipment fault diagnosis.



**WEI HAN** received the Ph.D. degree in condensed matter physics from the Institute of Physics, Chinese Academy of Sciences, Beijing, China, in 2013.

He is currently a Professor with the School of Physics and Electronic Engineering, Hebei Normal University for Nationalities, Chengde, China. His research interests include equipment fault diagnosis and performance evaluation.

...