Abnormal Vibration Detection of Wind Turbine Based on Temporal Convolution Network and Multivariate Coefficient of Variation

Jun Zhan[a,*], Chengkun Wu[b], Xiandong Ma[c], Canqun Yang[d], Qiucheng Miao[a], Shilin Wang[e]

[a]*College of Computer Science, National University of Defense Technology, Changsha 410073, China*

[b]*State Key Laboratory of High Performance Computing, College of Computer Science, National University of Defense Technology, Changsha 410073, China*

[c]*Engineering Department, Lancaster University, Lancaster, LA1 4YW, UK*

[d]*National Supercomputing Centre of Tianjin, China*

[e]*Beijing Goldwind HuiNeng Technology Co. Ltd., Beijing, China*

Abstract: A working wind turbine generates a large amount of multivariate time-series data, which contain abundant operation state information and can predict impending anomalies. The anomaly detection of the wind turbine nacelle that houses all of the generating components in a turbine have been challenging due to its inherent complexities, systematic oscillations and noise. To address these problems, this paper proposes an unsupervised time-series anomaly detection approach, which combines deep learning with multi-parameter relative variability detection. A normal behavior model (NBM) of nacelle vibration is firstly built upon training normal historical data of the supervisory control and data acquisition (SCADA) system in the high-resolution domain. To better capture the temporal characteristics and frequency information of vibration signals, the vibration spectrum vector is integrated with the multivariate time-series data as inputs and the spectrum-embedded temporal convolutional network (SETCN) is then used to extract latent features. The anomalies are detected through a multi-variate coefficient of variation (MCV) based anomaly assessment index (AAI) of relative variability among vibration residuals and environment parameters of the nacelle. The approach considers the time-series characteristics of input data and preserves the spatio-temporal correlation between variables. Validations using data collected from real-world wind farms demonstrate the effectiveness of the proposed approach.

Keywords: Abnormal detection, wind turbine, supervisory control and data acquisition (SCADA), multi-variate coefficient of variation (MCV).

1 Introduction

As one of the most important renewable energy generation technologies, wind power technology has been paid more attentions by researchers. According to statistics, the global cumulative installed capacity has reached 621GW by 2020 [1]. However, due to the harsh working environment and changing working conditions, abnormalities occur frequently, which brings great challenges to maintenance [2]. Thus, it is necessary to develop new anomaly detection methods to upgrade the maintenance mode and reduce maintenance costs [3].

* Corresponding author

Email address: zhanjun20@nudt.edu.cn (Jun Zhan)

At present, the monitoring of wind farms mainly relies on the Supervisory Control and Data Acquisition (SCADA) system or Condition Monitoring System (CMS) [4]. Compared with the CMS system which needs high-frequency signal acquisition for off-line analysis [5][6], the acquisition frequency of the SCADA system is usually $1Hz$ (sampled per second) or $0.00167Hz$ (averaged and then sampled per 10 minutes) [7], which can directly collect from the wind turbine control system and realize long-term online monitoring and data storage. It has become a common monitoring system for wind farms [8][9][10][11]. Therefore, the detection of wind turbine anomalies based on the existing SCADA data will be the low-cost and most efficient means to help us realize intelligent operation and maintenance. The SCADA system monitors hundreds of related condition parameters of wind turbines, such as wind speed, generator speed, power and temperature, which contain a wealth of operation state information of the turbines. Among these variables, the vibration signal can directly reflect the operation status and has become an interesting research topic [12][13][14][15]. However, wind turbines represents a complex electromechanical system, resulting in widespread non-stationary changes in operating parameters. Consequently, fault information may be covered by non-stationary changes, resulting in a high rate of misjudgment of anomaly detection [16].

Statistical and machine learning methods have been used in the data-based vibration abnormal detection of wind turbines. Guo Peng et al. [17] applied the nonlinear state estimation technique (NSET) to the vibration modeling of the wind turbine tower, by constructing a reasonable process memory matrix to represent dynamic process of the normal operation of the turbines. Wu Xin et al. [18] studied the fault characteristics of the wind turbine gearboxes, and proposed an anomaly detection method based on echo state network (ESN) and dynamic threshold scheme. Jin Xiaohang et al. [19] proposed an ensemble approach to detect anomalies and diagnose faults in wind turbines. In this method, historical data collected from healthy wind turbines are used to establish a Mahalanobis space as a reference space to model their normal behaviors. By comparing the predicted behavior from the training model with the reference space, anomalies can be detected. However, these methods smooth the training data, which may lose many statistical characteristics of the signal. The detection method using original $1Hz$ signal of SCADA system can be more effective [20]. Compared with the method based on the 10-minute averaged data, those high-resolution SCADA data preserve the statistics and frequency characteristics of the original signal and thus has a higher accuracy. Jun OGATA et al. [21] introduced a method of vibration time-frequency feature extraction based on Fourier local autocorrelation. However, this method not only requires overall understanding of mechanical and other associated domain knowledge, but also cannot cope with a huge amount of historical data to extract complex nonlinear dynamic information; therefore the application is greatly limited.

Naturally, vibration signals are time series, presenting temporal dynamics and nonlinear characteristics due to the interaction and dependency between different subsystems or components [22]. To deal with such complex data, the deep neural network algorithm could provide an effective solution. Chao Guo et al. [23] studied a new high-speed train (HST) failure analysis using deep belief network (DBN) to automatically learn features of raw data. Shao Haidong et al. [24] studied the vibration abnormality detection method of rotating machinery under variable working conditions, and proposed a fault diagnosis method based on deep autoencoder (DAE) feature learning. Ke Yan et al. [25] studied the impact of training data enhancement on model performance, and the chiller fault diagnosis are realized through the optimized VAE network. For the gearboxes with frequent failures, Jianbo Yu et al. [26] proposed a new DNN, one-dimensional residual convolutional autoencoder (1-DRCAE) to learn features from vibration signals directly in an unsupervised-learning way. A. E. Elsaid et al. [27] developed a recurrent neural network (RNN) capable of predicting aircraft engine vibrations by using long short-term memory (LSTM) neurons. These methods prove that deep learning algorithms are advantageous in the

field of vibration anomaly detection. However, most of these methods have mainly focused on either individual statistical characteristics or time series characteristics[28][29][30].

To overcome the above problems, this paper proposes a condition monitoring method based on the combination of temporal convolutional network (TCN) and MCV to perform spatio-temporal fusion of SCADA data. The research objective focuses on the high dynamic vibration of the wind turbine nacelle of $1Hz$ (original time series data), which has been considered to be more challenging than analyzing the inertial data such as the temperature and pressure of wind turbine. The main contributions of this paper are summarized as follows:

- The input data of the model is enhanced by embedding the frequency spectrum of the vibration signal into the multivariate time series data as domain knowledge, which improves the visibility and effectiveness of the SCADA data, and makes the model more suitable for actual application scenarios.
- The deep TCN network is applied to model with the high dynamic original high-resolution SCADA nacelle vibration data in $1Hz$, rather than using the data after down-sampling or smoothing. To our best knowledge, this is the first attempt of the deep learning method in this field. It can retain the temporal and frequency characteristics of the signals, which helps to improve the accuracy of the NBM.
- MCV is utilized to fuse nacelle acceleration residual and environmental parameters and construct AAI with higher sensitivity for detecting small drift of vibration signals.

The rest of the paper is organized as follows. Section II introduces the proposed detection approach including overall framework, SETCN model and MCV based AAI. Section III presents case studies for vibration anomaly detection of wind turbines. The conclusions and possible future work are given in Section IV.

2 Proposed Anomaly Detection Approach

2.1 Overall framework

The study of this paper is focused on the turbine nacelle because it houses all of the generating components in a turbine. Analysis and detection of nacelle vibration has been challenging due to its inherent complexities, systematic oscillations and noise. Firstly, historical data under healthy conditions is used to model normal behavior. The frequency spectrum of the vibration data is extracted and embedded into the time-series data to form new training data. The major goal is to represent the energy of a vibration signal in both temporal and frequency dimensions at the same time, so as to help the network in extracting the characteristic information with limited data and computational resources better [31][32][33][34][35]. Then, these embedded vectors are further inputted into the deep TCN model to filter out high-value information and produce output through a fully connected layer. The whole process maps the multi-dimensional mixed time-frequency data to the vibration signal by searching for the optimal non-linear function expression. MCV is subsequently introduced to calculate the relative variability between the output signal of the NBM and the actual signal. Finally, the alarm threshold is determined through the training data. For the proposed method, the temporal dependency and inter-correlations of data are considered simultaneously to accurately identify the operation state of the wind turbine.

From a data perspective, anomalies of wind turbines are defined as the situation in which the observation data obviously deviates from its historical healthy data. These anomalies only account for a

small proportion of the data, which are often hidden in large amounts of data and hard to be found. Due to the small number of abnormal samples and the difficulty of collection, it is hard to use supervised algorithms to identify anomalies. However, it is easy to collect healthy data which are usually collected from the newly built or overhauled turbines after stable operation for more than one year. The characteristics of the healthy conditions can be extracted from a large amounts of SCADA historical data to establish a NBM, and the alarm threshold of the observation data is estimated to realize unsupervised anomaly detection, which is very useful for realizing wind turbine online condition monitoring. During operation of the wind turbines, the real-time data collected are inputted into these models after the same processing as the training process. When the calculated AAI exceeds the threshold obtained in the healthy state, the observed signal is considered to have undergone abnormal changes, indicating that a failure or potential failure is pending to occur. The overall scheme framework of anomaly detection method proposed in this paper is depicted in **Figure 1**. Three major stages are involved:

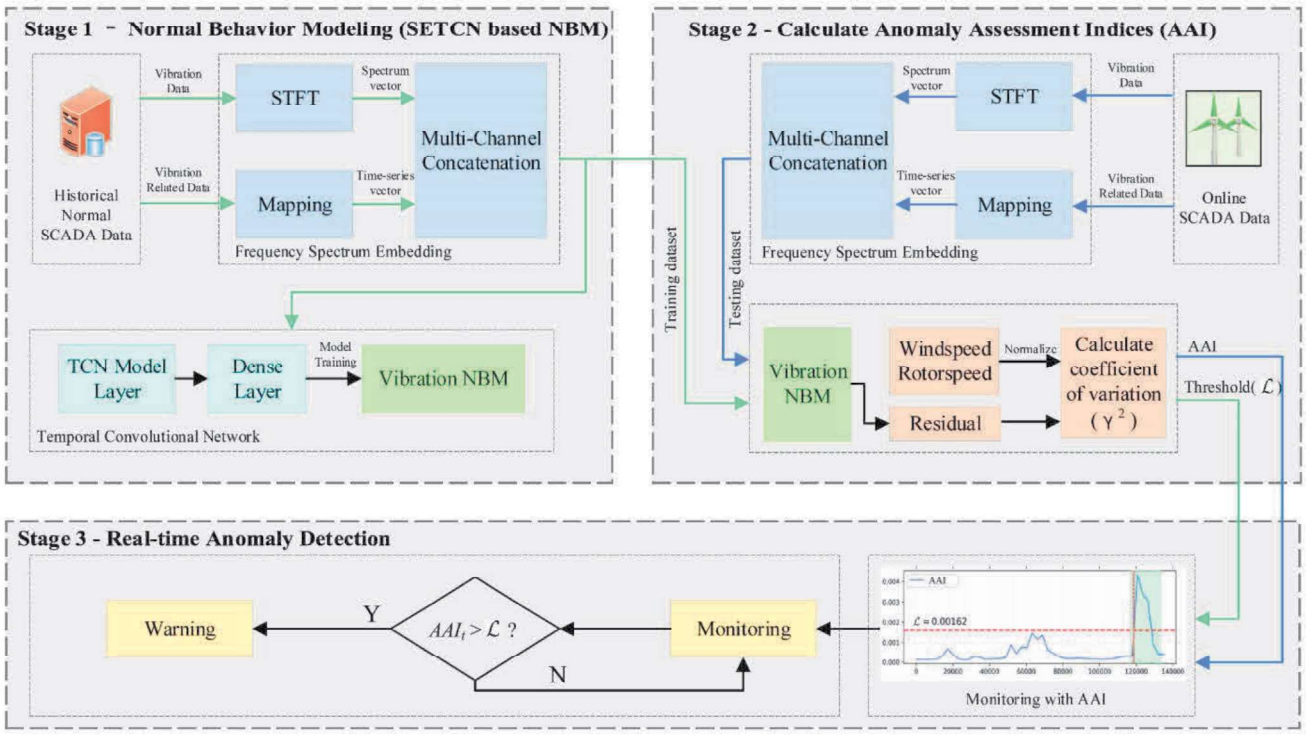

Figure 1 Temporal-spectrum fusion anomaly detection framework for wind turbines

- **Stage 1, Normal behavior modeling:** The data stably operating for more than one year after new construction and overhaul are selected as normal data. Then the data preprocessed by the sliding window with the length of $N = 120$ samples are regarded as the input of the NBM of nacelle vibration. To make full use of the time-frequency characteristics of the vibration signal to improve the accuracy of the model, we transform the vibration signal into a spectrum vector with Short-Time Fourier Transform (STFT) and embed it into the time series to enhance the training data. Finally, the regression training is conducted based on the constructed TCN prediction model using the vibration signals in the X (horizontal) and Y (vertical) directions at the next time moment as the target value, and the NBM can be gained.
- **Stage 2, Calculate anomaly assessment index:** In this stage, the real-time data which are preprocessed in the same way as the offline training stage are delivered to the NBM for

prediction, and the predicted value is then compared with the actual value to obtain the residual signal. The MCV based AAI is calculated by using the residuals in the X and Y directions, combined with the current wind speed. Taking into account the impact of environmental conditions, the AAI is not static, which fluctuates within a certain range. With the same method, the proposed threshold of AAI under healthy conditions can also be calculated by using the training dataset.

- **Stage 3, Real-time anomaly detection:** If the value of AAI sequence obtained from the online detection process exceeds the maximum AAI value in the normal state, an alarm of anomaly will be produced, which provides a basis for operation and maintenance personnel to perform the troubleshooting and maintenance. The data used in this paper are from the yaw system. Driven by the yaw system, nacelle of wind turbine rotates keeping vane on the windward side, which is conducive to the capture of wind energy by the wind turbine. The main components of the yaw system are the yaw motors and gear-driven bearings. The fault of this system will cause abnormal vibration of the nacelle.

2.2 SETCN Normal behavior model

2.2.1 Dilated Causal Convolutions and TCN Residual Module

In this paper, the TCN which is based on the model in reference [36] is adopted to model nacelle vibration signals. For the analysis of time series data, the most commonly used neural network is the RNN, such as LSTM, GRU[37], which can employ the internal memories to process input time series. However, the TCN, a member of the convolutional neural network (CNN) family, performs better than the RNN in processing long sequences of inputs [38]. Temporal convolutional network (TCN), first proposed by Lea et al. in 2016, is a deep learning approach for extracting time series data features. It is widely used in fields like time series prediction [39], probability prediction [40], traffic prediction [41]. In terms of structure, TCN uses one-dimensional causal convolution and dilated convolution as standard convolution layer, encapsulates each two such convolution layers and identical mapping into a residual module, and then stacks the depth network by the residual module. The convolution layer of TCN combines causal convolution and dilated convolution. The purpose of using causal convolution is to ensure that the prediction of previous time steps will not use future information, because the output of time step T will only be obtained according to the convolution operation on T-1 and its previous time steps. The purpose of using dilated convolution is to increase the size of the receptive field. Dilated convolutions are identical to regular one-dimensional convolution, except that a new parameter called "dilation rate" is added to the convolution layer, which controls the spacing of values when the convolution kernel processes data. In the convolution process, not all positions in the convolution window are involved in the calculation, but certain holes will be left based on the dilation rate, and the operation will be completed once these hole positions are filled with zero. With the more layers, more holes need to be filled in the convolution window.

The vibration signals are modelled with input parameters of high-resolution ($1Hz$) SCADA data without smoothing. As a result, the network needs to deal with long historical data sequence, which results in a complicated network structure and heavy calculation burden. Dilated causal convolution of TCN network can solve this problem, which can make receptive field grow exponentially. For a 1-D sequence input $x \in \mathbb{R}^n$ and convolution kernel $f: \{0, \dots, k-1\} \to \mathbb{R}$, the convolution at the moment t is defined as:

$$F(t) = (x *_d f)(t) = \sum_{i=0}^{k-1} f(i) \cdot x_{t-d \cdot i} \quad (1)$$

where d is the dilation factor representing step length among the convolution kernels. When $d = 1$, a dilated convolution becomes a regular convolution. k is the size of convolution kernel, and $t - d \cdot i$ indicates the direction of the past. The basic structure of network is shown in **Figure 2**.

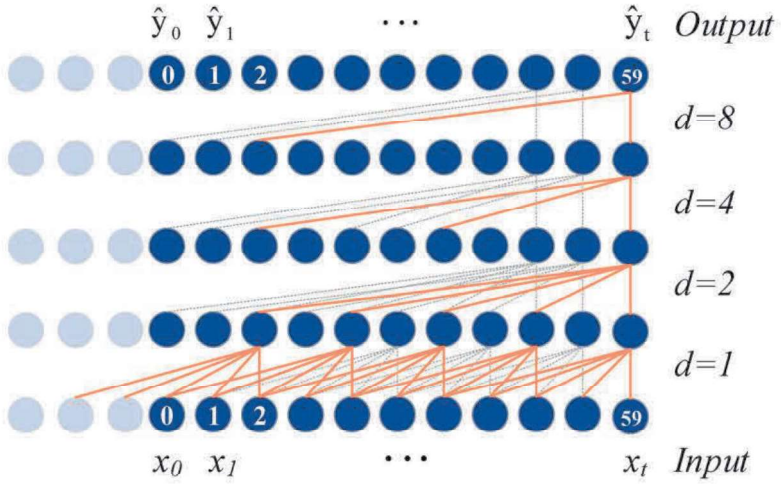

Figure 2 Illustration of the dilated causal convolution with 4-layer

More layers of the TCN network need always be stacked in the course of actual implementation. As a result, it is necessary to include a residual connection in the output of each TCN layer to reduce learning complexity while avoiding gradient exploding or vanishing. The basic residual block structure of the TCN network is shown in **Figure 3**, which is composed of two dilated causal networks and each of them includes a kernel operation module consisting of a dilated causal convolution, a nonlinear activation function (ReLU), a weight normalization, and a dropout regularization. In the meantime, a kernel network of $1 * 1$ is used to connect between input Z_{i-1} and output Z_i. After the addition of residual connection, the TCN module can be expressed as:

$$Z_i = \mathcal{F}(Z_{i-1}, \{W_i\}) + Conv_{1*1}(Z_{i-1}) \quad (2)$$

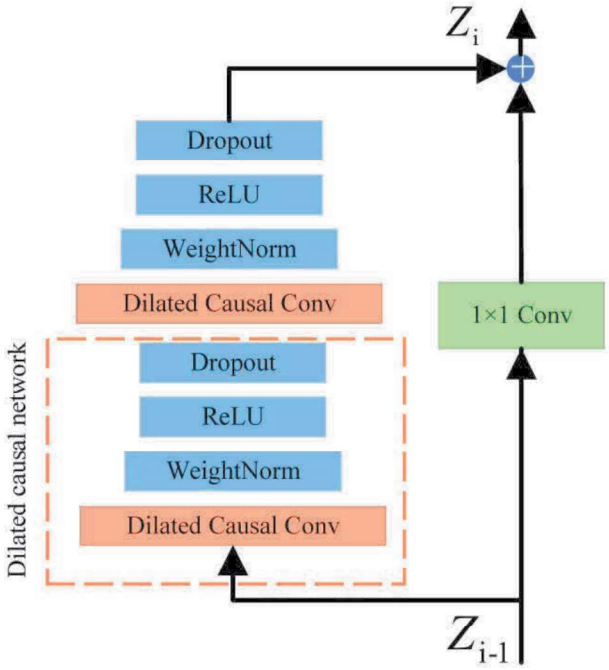

Figure 3 Illustration of the residual block structure of TCN

2.2.2 Model Structure of SETCN

The overview of SETCN architecture is shown in **Figure 4**. The input vector I is applied to extract temporal depth features through the TCN network. At the end of the network, a dense layer is added as the output layer. The dense layer is included because we need to reduce the dimension of the tensor of TCN output to 1 in order for the network's final output dimension to match the dimension of the vibration signal. Since there are many convolution layers stacked in the network, in order to help the network converge and reduce the risk of gradient exploding or vanishing, the initialization process before training is very important. Significantly, the input I in this study is concatenation by the multivariate time-series R collected by SCADA and the spectrum vector P of the vibration signal after passing through the STFT. In general, the spectrum vector generated by STFT is a 2D-vector, and its dimension is affected by the window size and step size what we have chosen. Therefore, to realize the embedding of the spectrum we intercept P the same sequence length with R. I is denoted as:

$$\mathrm{I} = \mathrm{Concat.}(R, P) = \begin{bmatrix} r_{t_0}^1 & r_{t_1}^1 & & r_{t_N}^1 \\ r_{t_0}^2 & r_{t_1}^2 & \cdots & r_{t_N}^2 \\ r_{t_0}^\alpha & r_{t_1}^\alpha & & r_{t_N}^\alpha \\ \vdots & & \ddots & \vdots \\ p_{t_0}^1 & p_{t_1}^1 & & p_{t_N}^1 \\ p_{t_0}^2 & p_{t_1}^2 & \cdots & p_{t_N}^2 \\ p_{t_0}^\beta & p_{t_1}^\beta & & p_{t_N}^\beta \end{bmatrix} \tag{3}$$

After obtaining I, a sliding window with a width of w and a length of l is used to extract data accordingly, which generate a set of subsequences of length l. In the above denotation, N represents the total length of the dataset. Therefore, after sliding window is processed with step size of 1, $(N - l) + 1$

set of subsequences are obtained. Through the model, the vibration value at time $t+1$ is predicted by the subsequence at time t and recorded as $O=[v_{t+1}]$. α represents the dimension of the multivariate time series R, and β denotes the spectral dimension of the vibration signal after STFT. The total dimension of the concatenated input I is $w=\alpha+\beta$. In this paper, we select 6 SCADA parameters and set the STFT window length $nperseg=60$, $noverlap=59$ to obtain the spectrum with dimension 31, therefore, $\alpha=6$, $\beta=31$.

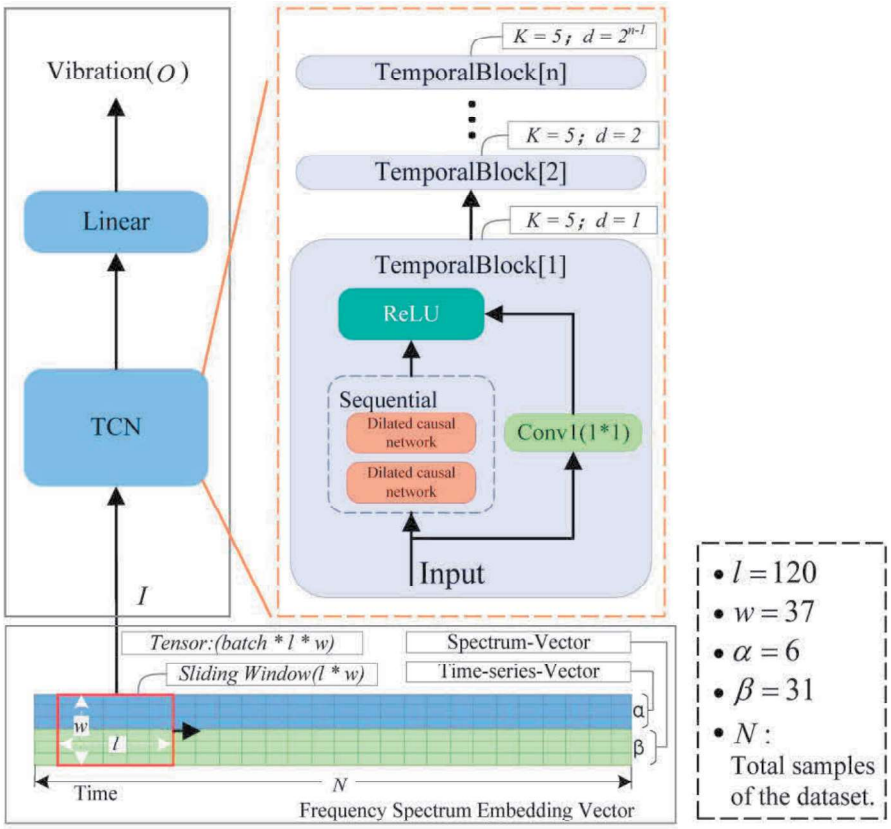

Figure 4 Illustration of the SETCN architecture

2.2.3 Model evaluation

When different models are compared and analyzed, their Mean Squared Error (MSE), Mean Absolute Error (MAE) and Coefficient of Determination (R^2) are calculated.

$$MAE=\frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i-y_i| \tag{4}$$

$$MSE=\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i-y_i)^2 \tag{5}$$

$$R^2=1-\frac{\sum_{i=1}^{n}(\hat{y}_i-y_i)^2}{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2} \tag{6}$$

In the above formula, $\hat{y}$ represents the predicted value of the model, y represents the true value of the measurement, and $\bar{y}$ represents the mean value of the true value y. The more MAE, MSE tending to 0, the better the model effect and the higher the accuracy. R^2 is the ratio of the Sum of Squares for regression (SSR) to the Sum of Squares for total (SST), reflecting explainable ratio of the independent

variable to the dependent variable. The closer the R^2 value is to 1, the better the model fit.

2.3 MCV based AAI

2.3.1 Multivariate coefficient of variation (MCV)

The residual between the predicted value and the true value is clearly affected by the prediction results and cannot be directly used for anomaly identification. Hence, setting an abnormal discriminant indicator is extremely important to realize an early warning of the failure. Some researches have proposed the evaluation indicators, such as the entire residual(RE) [42], Mahalanobis distance(MD) [43][44]. However, these methods only take a single variable and thus ignore the interaction among variables so that small abnormal drifts cannot be effectively detected. Therefore, it is necessary to propose an abnormal discriminant indicator based on the multivariate analysis.

Multivariate coefficient of variation was firstly put forward by Reyment (1960). Thereafter various definitions were proposed by Van Valen [45], Voinov VG and Nikulin MS [46]. For the method defined by Nikulin, the robustness becomes the strongest when there are disturbances or noise pollution in the sample data. Meanwhile, asymptotic variance of sample estimator is not affected by dimensional changes. Therefore, this paper adopts multivariate coefficient of variation as an anomaly assessment index.

For random samples X which follows a normal distribution with a length of n and dimension of p, assuming that its mean value is μ, covariance matrix is Σ, that is, for $X \sim N(\mu, \Sigma)$, where $X_i^T = (X_{i1}, X_{i2}, \dots X_{ip})$, $i \in (1, n)$, $\mu^T = (\mu_1, \mu_2, \dots \mu_p)$, the coefficient of variation can be computed as:

$$\gamma = (\mu^T \ \Sigma^{-1} \ \mu)^{-\frac{1}{2}} \tag{7}$$

In the calculation process, we can use the sample mean $\bar{X}$ and sample covariance matrix S to estimate μ and Σ respectively:

$$\bar{X} = \left(\frac{1}{n}\sum_{j=1}^{n} X_{1j} \dots \frac{1}{n}\sum_{j=1}^{n} X_{pj}\right)^T \tag{8}$$

$$S = \frac{1}{n-1}\sum_{j=1}^{n} \left(X_j - \bar{X}\right)\left(X_j - \bar{X}\right)^T \tag{9}$$

Therefore, the estimated value of the coefficient of variation $\hat{\gamma}$ can be given by:

$$\hat{\gamma} = (\bar{X}^T S^{-1} \bar{X})^{-\frac{1}{2}} \tag{10}$$

To achieve a real-time anomaly detection, the newly collected data need to be handled continuously. Hence, we use a new time window with length L to obtain sample data sequentially. In the meantime, for the convenience of calculation and anomaly observation, the AAI calculated with a sliding window length L is monitored and noted as $\hat{\gamma}_t^2$.

$$AAI(L) = \hat{\gamma}_t^2 = \frac{1}{(\bar{X}^T S^{-1} \bar{X})} \tag{11}$$

2.3.2 Abnormal warning threshold

When an anomaly occurs in wind turbines, the changes of mean value and covariance of samples result in significant change of AAI. Therefore, an appropriate threshold needs to be set so that the fault can be detected reliably before a serious failure of nacelle vibration occurs. In this paper, the maximum value of AAI by the training dataset of NBM in wind turbines is selected as the threshold:

$$\mathcal{L} = \max\{\hat{\gamma}_{01}^2, \hat{\gamma}_{02}^2, \cdots, \hat{\gamma}_{0T}^2 | T = 1,2,3 \cdots\} \quad (12)$$

where $\hat{\gamma}_{0j}^2$ denotes the AAI at the moment j covering a window length L under the normal condition of wind turbines.

Compared with the traditional method by which all wind turbines are set to the same fixed threshold at the manufacturing factory, the proposed method takes into account the inherent differences in terms of the performance of wind turbines and calculates the threshold by using operating data in healthy conditions, which is beneficial to reduce false alarms.

3 Experiments and Results

In this section, the proposed approach is verified by applying the data from real wind farms. The experiment is composed of two parts. In the first part, the performance of the model is evaluated under the healthy dataset. The second part elaborates how to effectively distinguish the abnormal behavior by using the proposed method. The relevant experimental code is developed based on python3.7. TCN, LSTM, and GRU are implemented using PyTorch 1.6.0, and all calculations are performed on a 64-bit linux operation system installed on a computer with an NVIDIA Tesla V100 GPU. The source code is available at https://github.com/zhanjun717/SETCN_MCV_AD/tree/master.

3.1 Selection of Input Parameters

To accurately model the vibration of the wind turbine nacelle, we need to determine the inputs of the model by analyzing the key parameters causing the vibration. Wind turbines can be divided into double fed induction generator (DFIG) and direct-drive wind power generator (DWPG) according to their structures. The data in this paper comes from the DWPG, the structure of which mainly includes blades, hub, generator, nacelle, and tower. Among them, the blades convert the absorbed wind energy into mechanical energy for driving the generator to generate electrical energy. Compared with DFIG, the hub of DWPG is directly connected to the generator rotor, eliminating the speed-increasing gearbox. The nacelle is positioned on the slewing bearing at top of the tower on which an anemometer measuring wind speed and direction is installed. The nacelle also accommodates the power converter and control cabinet. The load acting on the blades is transmitted to the nacelle through the hub and generator and finally acts on the tower, resulting in the vibration of various structural parts. Different from inertial variables such as temperature and pressure, the vibration can reflect the instantaneous characteristics related to the operating state of the nacelle in real-time. If the component connected to the nacelle fails, the vibration signal will change immediately. Therefore, although the frequency of the vibration signal collected by SCADA is low, a long-term online monitoring can still be realized, which is of great significance.

The vibration sensors are usually installed in the nacelle and can collect the nacelle acceleration signals in the horizontal direction (X-direction) and vertical direction (Y-direction) of the nacelle. The time interval is generally 1 second or 10 minutes (note SCADA data sampled at 1 second, i.e., $1Hz$, are used in this paper). The factors that cause turbine vibration are complicated. The vibration magnitude is related to the hub, yaw system, generator, and other excitation sources as well as the design of the nacelle structure. Since the inherent characteristics of the structure are relatively stable, this paper does not consider the changes of the inherent characteristics, rather mainly analyzes the excitation source that causes the vibration.

Among the parameters collected by SCADA, the variables related to vibration mainly include ‘Wind speed’, ‘Hub speed’,’ Pitch motor current’, ‘Hydraulic brake’, ‘Generator torque’, ‘Cabin temperature’and ’Power’ [47][48]. Therefore, this paper selects these parameters as the inputs of the model.

As an example, **Figure 5** shows the operating parameters from wind turbine. During the day, the working conditions of the wind turbine are constantly changing as the wind speed changes. Among them, mode A represents the working condition that the output power of the wind turbine reaches the rated value and remains unchanged while mode B represents the shutdown of turbine operation. When the blade angle is retracted to the stop position, the output power is 0. Mode C represents the power generation below the rated power. During this time period, the wind turbine is running with maximum power point tracking (MPPT). It can be seen from the **Figure 5** that the change of working conditions is the main factor causing the vibration change. When the wind turbine is stopped, the aerodynamic load of the wind on the blades will still cause large nacelle vibration. Under healthy working conditions, these parameters will maintain a relatively stable non-linear relationship. However, once a fault occurs, the relationship among them will mutate and deviate from the healthy mode, which provides a basis for subsequent abnormal detection.

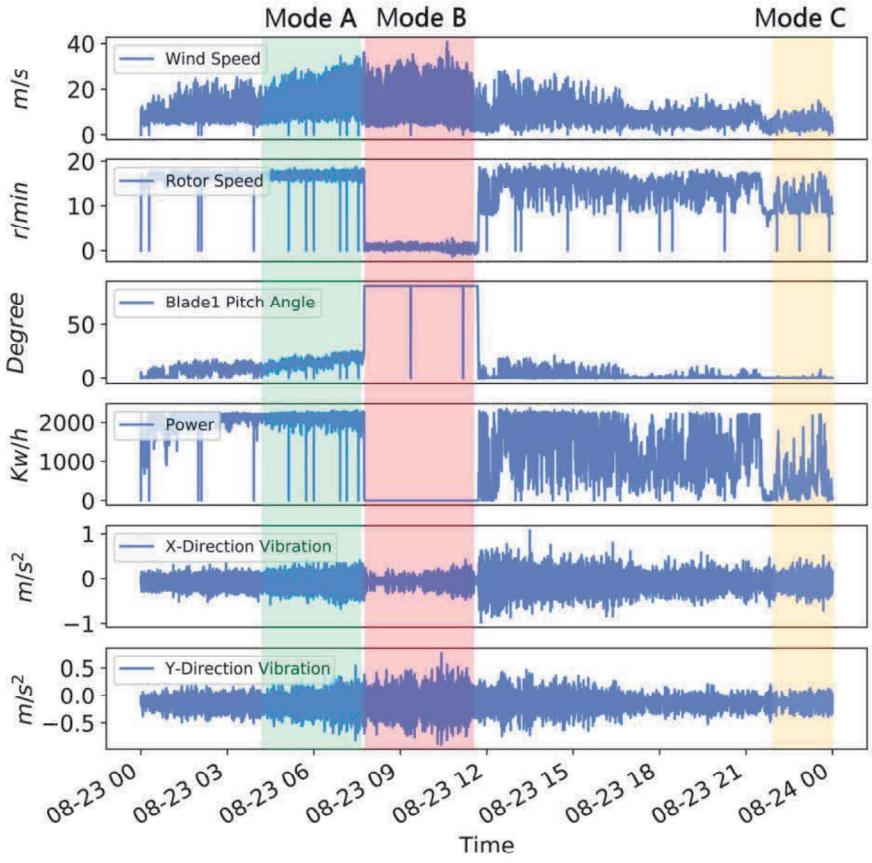

Figure 5 The vibration signals in the x and y directions of the nacelle and its related variables under different working conditions

3.2 Validation on healthy wind turbines

3.2.1 Normal Behavior Data

The SCADA data used in this study were collected from five real wind farms located in different regions of China. As shown in **Table 1**, the data contain historical data in the healthy condition, labelled as WT1, WT2, WT3, WT4 and WT5, which are used to verify the feature extraction capabilities of the proposed NBM under different working conditions and wind turbine types. The other two datasets

labelling WT6, WT7 (details will be given in the next subsection) contain the data collected from historical healthy condition as well as two different abnormal operation patterns, which are used to compare the performance of the proposed NBM. It should be noted that although the signals of these two wind turbines are abnormal, none of them trigger the currently set alarm threshold, which poses a potential threat to the equipment.

Table 1 Dataset from Healthy Wind Turbines

NO.	Working condition	Rate	Describe
WT1	Location: Mountain Environment: Hot and Humid Power: 2MW	1 second	Healthy condition
WT2	Location: Gobi Environment: Cold, Sand wind Power: 2MW	1 second	Healthy condition
WT3	Location: Flatlands Environment: Warm, Rainy Power: 2MW	1 second	Healthy condition
WT4	Location: Coastal Environment: Salt spray, hot Power: 2MW	1 second	Healthy condition
WT5	Location: Offshore Environment: Salt fog, typhoon Power: 5MW	1 second	Healthy condition

3.2.2 Selection of SETCN model hyperparameters

As mentioned before, TCN can obtain different receptive fields through stacking, thereby affecting the model performance. To determine the best structure of the SETCN model for NBM tasks, numerous model structures were compared. **Table 2** shows the changing trend of model performance indicators under different hyperparameters. It can be seen from the table that when the layers of the TCN model are 8 and the number of hidden layer nodes is 32, the model achieves the best overall performance. After that, the increasing of layers causes performance degradation, which may be caused by overfitting owing to too many model parameters. At the same time, with the increase of network complexity, the training time has also increased dramatically. Therefore, a structure with 8 layers and 32 hidden units is finally adopted. In our study, SCADA data with high resolution contain both temporal and frequency information. Too small convolution kernel may ignore temporal domain information, whereas too large convolution kernel may have difficulty in acquiring frequency domain information. Furthermore, the convolution kernel with an even number cannot guarantee to remain the size of input and output of feature map unchanged although padding is added symmetrically. Therefore, the kernel size is set to 5.

Table 2 The results of NBM under different network structures

	Hidden Layer	Hidden Units	MSE	MAE	R2	Time cost(s/epoch)
SETCN-1	4	16	0.0011376	0.0203212	0.74016	151
SETCN-2	8	32	0.000704287	0.0142384	0.76608	252
SETCN-3	16	64	0. 00068965	0. 0139541	0.76903	8795

3.2.3 Comparison with other methods

There are many other models that have been applied to anomaly detection. This subsection makes a detailed comparison in terms of multiple performance indicators between the SETCN method and these

models. The compared models mainly include echo state network(ESN) [18], multi-layer neural network(MLP) [18], multi-layer LSTM, multi-layer GRU [49], convolutional bi-directional LSTM networks(CNN+LSTM) [50], transformer [51] and base-TCN [36]. In these models, the base-TCN is used to validate the efficiency of spectrum embedding, hence it only accepts the original vibration signal as input, and uses the same network topology and superparameters as the SETCN in the experimental stage. To ensure the objectivity of the comparison results, the experiment process uses the same WT1 dataset, and the training set and test set duration are 10 days and 1 day respectively.

As can be seen from Table 3 that ESN and MLP show the worst results. This is because these methods mainly consider the statistical characteristics of the data and ignore the temporal characteristics of the vibration signal. Furthermore, the indicators show that the TCN algorithm performs the best, and the coefficient of determination R^2 reaches 0.7661.

Table 3 Effect of NBM under different algorithms

Algorithms	MSE	MAE	R2
Echo State Network (ESN) [18]	0.00227118	0.0288218	0.4837
Multi-layer NN (MLP) [18]	0.00224838	0.0285943	0.4889
Multi-layer LSTM	0.00129546	0.022662	0.7052
Multi-layer GRU [49]	0.00137155	0.0233322	0.6879
CNN+LSTM [50]	0.00149661	0.0232127	0.6595
Transformer (6 layer) [51]	0.0012909	0.0220651	0.7062
Base-TCN [36]	0.00110303	0.0199301	0.749
Proposed SETCN	**0.000704287**	**0.0142384**	**0.7661**

The existing vibration anomaly detection methods based on deep learning can be mainly classified into two categories: time series-based methods and spectral analysis methods [52][53]. Therefore, in our research, we input both the time-series signal and the frequency spectrum of the vibration into the feature extraction model at the same time. The experimental result in the last two lines of **Table 3** shows that the accuracy of the proposed SETCN is higher than that of the basic TCN model without an embedded spectrum under the same network structure. As an example, **Figure 6** and **Figure 7** show the forecasted curves with LSTM and SETCN, respectively. It can be known from the figure that SETCN has a better fitting performance than LSTM. The main reasons are i) for vibrations with high dynamic characteristics, short-term temporal dependence is more important than long-term temporal dependence, and ii) the vibration signal not only contains the time-domain features, but also the high dynamic frequency features that those ordinary inertial signals do not have. The time-frequency characteristics of the signal extracted by STFT help the deep network to learn more information.

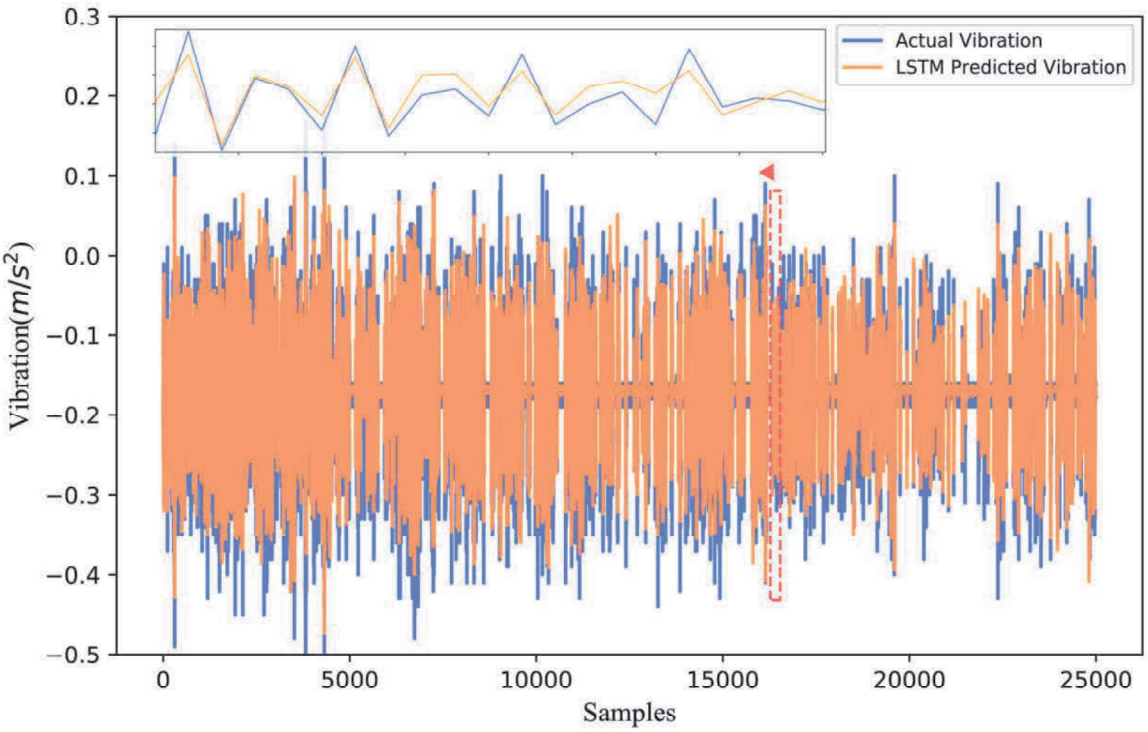

Figure 6 Prediction results of LSTM algorithm

The vibration prediction results for the test set using SETCN are shown in **Figure 7**. Results show that the predicted value can follow the changing trend of the original vibration, and its peak-to-peak value is also closer. When an abnormality occurs, the abnormal data will deviate from the predicted vibration, which will facilitate to detect the occurrence of the abnormality. The abnormality will be verified on the dataset of WT6 and WT7.

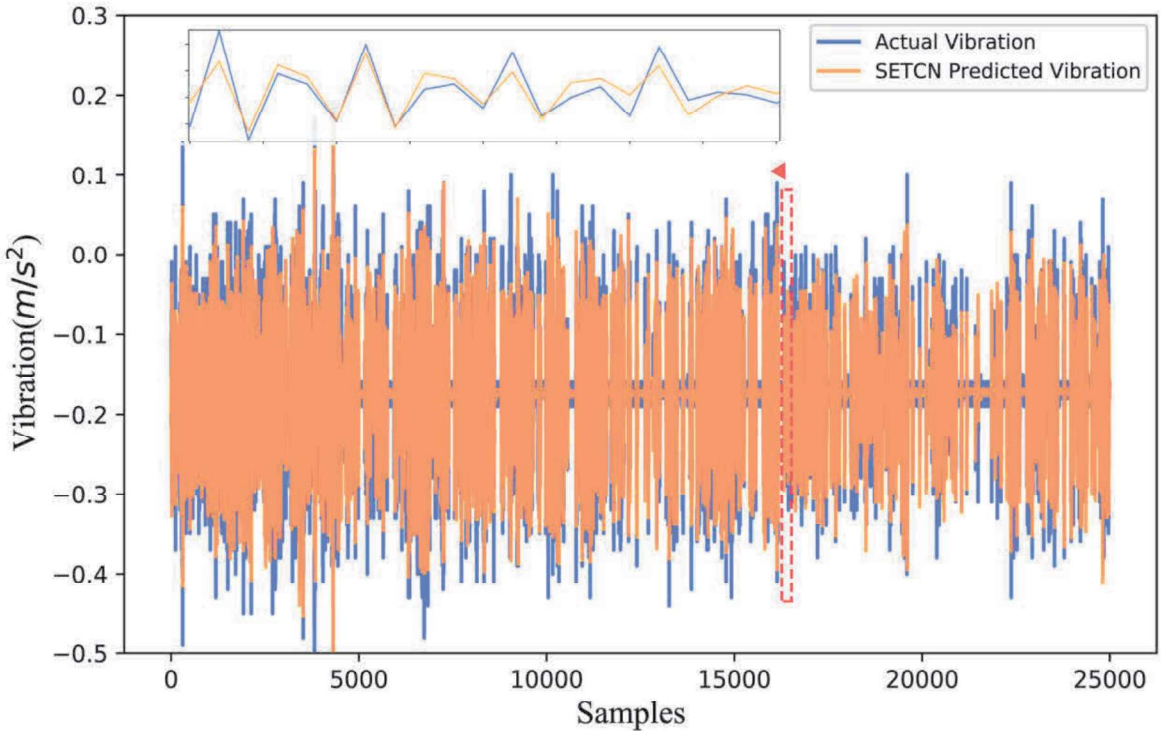

Figure 7 Prediction results of SETCN algorithm

3.2.4 Analysis of normal behavior model under different working conditions

Wind turbines of different types are usually built in areas with heterogeneous environmental conditions, which inevitably result in the difference in operating conditions. To demonstrate the versatility of NBM on different wind turbines, the datasets from different regions of China were collected for comparative experiments. As shown in **Table 4**, the SETCN algorithm is equally effective when applied to different wind turbines.

Table 4 SETCN results under different wind turbines

Dataset	MSE	MAE	R2
WT1	0.000704287	0.0142384	0.7661
WT2	0.000411282	0.0107143	0.7862
WT3	0.000365426	0.00986525	0.7532
WT4	0.000354641	0.01256566	0.7668
WT5	0.00179644	0.0317397	0.7668

3.3 Anomaly Detection

3.3.1 Abnormal Behavior Data

After the NBM based on SETCN is trained offline, the next task is to use the proposed framework to detect the abnormalities of the real-time operating data. We use the historical data from two different wind turbines, WT6 and WT7, as shown in the **Table 5**, to assess the feasibility of the proposed approach. The two turbines were built in 2015 and have been operating in good condition for a period of time. However,

abnormalities occurred in the later period. The nacelle vibration data of WT6 showed abrupt anomalies in Jan. 2018, while the data of WT7 contained periodic anomalies in Jan. 2016. It is worth noting that after the signal changes, none of the existing monitoring systems trigger the predefined fixed threshold, producing an alarm.

Table 5 Data from Abnormal Wind Turbines

NO.	Time	Samples	Rate	Describe
WT6 (Case 1)	Jun. 2017 - Dec. 2017	18358519	1 second	Healthy condition
	Jan. 2018 - Feb. 2018	30818719	1 second	Abrupt abnormality caused by yaw bearing abnormality
WT7 (Case 2)	Dec. 2015 - Feb. 2016	950052	1 second	Healthy condition
	Feb. 2016 - Mar. 2016	777343	1 second	Periodic abnormality caused by hub aerodynamic imbalance

3.3.2 Case 1, Abrupt abnormality

The wind turbine can automatically track the direction of the wind to capture the maximum wind energy. The most important component to realize this process is the yaw system. The yaw system consists of the yaw motors, the yaw bearing, and the yaw brake based on hydraulic equipment. As the overall weight of the nacelle and hub is very large, if the yaw bearing is abnormal (such as cog breaking) during the yaw process, it will cause an abrupt change of the nacelle vibration.

Abrupt abnormality is usually defined as an abnormal state at which the signal suddenly deviates from the original general state. If the deviation is large enough to exceed the existing fixed threshold, it is generally easy to detect and vice versa. **Figure 8** shows the wind speed, vibration, and the AAI curve of WT6 before and after the failure. The light green area in the figure is to show where the abnormality occurs. From the wind speed **Figure 8**(a) and the vibration diagrams **Figure 8**(b) and **Figure 8**(c), it can be seen that before the abnormality occurs, the vibration signal changes with the change of wind speed, and the trend remains stable. When the wind speed exceeds 10m/s, the wind turbine enter the rated power generation mode, the amplitude of vibration reaches the maximum. However, after the abnormality occurs, the vibration behaves differently. Having checked the event logs of SCADA system, it is found that this wind turbine experienced a yaw system abnormality after January 2018. Through the analysis of data, it is found that on January 18th, at the position marked by the red line in the figure, the signal had a sudden change. However, the peak-to-peak value $-1.4m/s^2$ did not exceed the fixed alarm threshold $\pm 1.75m/s^2$ that was preset in the main control program of wind turbines, and therefore no alarm shutdown was triggered.

By using the method proposed, we first select the data from the WT6 in the healthy state to train the NBM, then calculate the MCV based AAI value as the threshold by equation (12). In the subsequent abnormal identification process, the abnormal data is inputted into the trained model, according to equation (11), we choose $L = 3600$ as observed every hour,, the AAI is calculated. **Figure 8** shows AAI is calculated for wind speed, X-direction vibration residual and Y-direction vibration residual. According to equation (12), the maximum AAI threshold $\mathcal{L} = 0.00162$ is obtained. **Figure 8**(d) shows the change of AAI value. It can be seen from the figure that at the sample 118990, the proposed method detects the abnormality, which represents 10:00 on Jan. 18, 2018 when the signal changes.

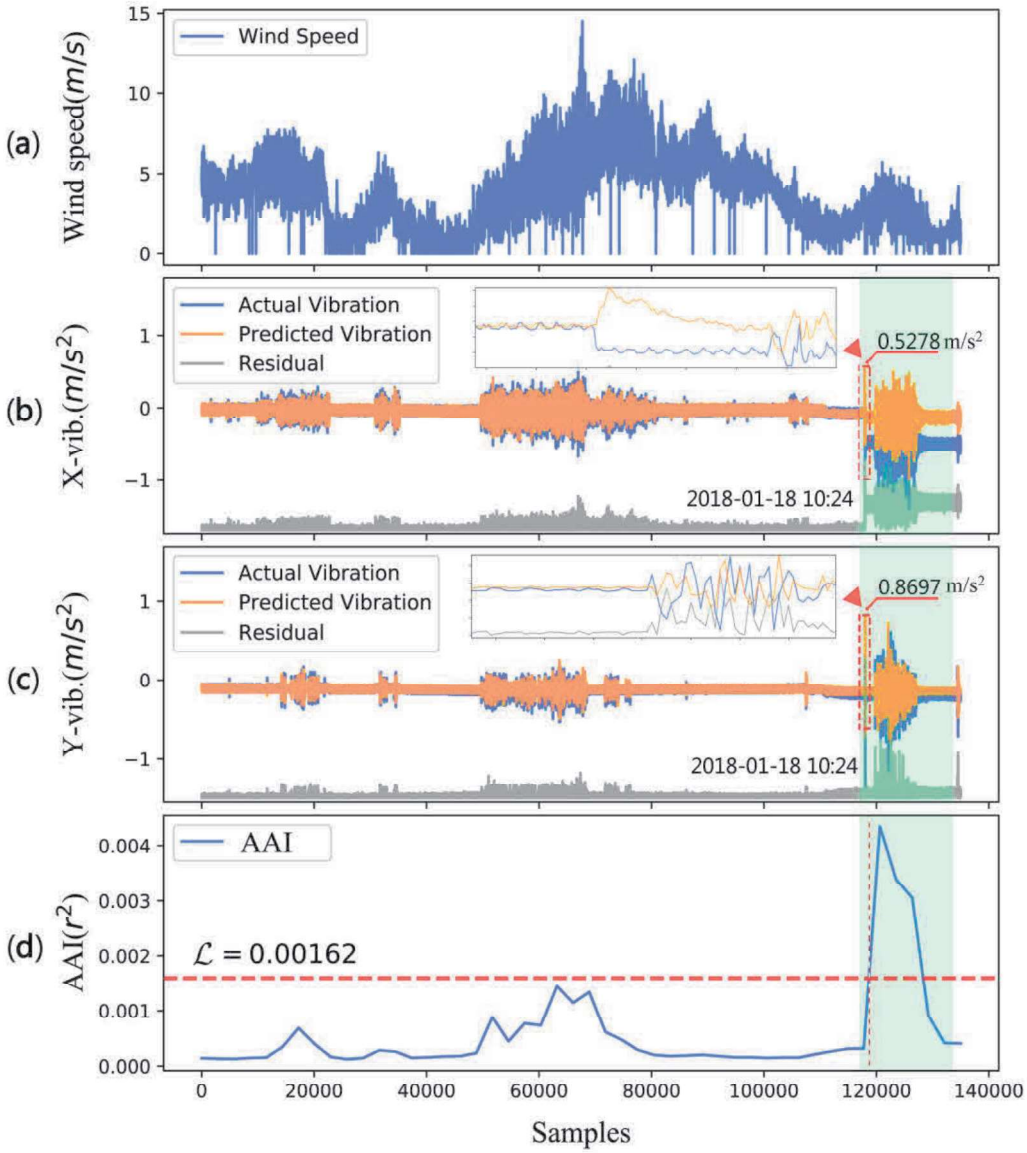

Figure 8 Abrupt vibration detection result

3.3.3 Case 2, Periodic Abnormality

The data, in this case, comes from a wind turbine with an aerodynamically unbalanced hub. As a rotating power generation equipment, periodic abnormality from the vibration of the wind turbine nacelle often occurs due to impeller imbalance. This kind of periodic abnormality is difficult to detect by the traditional threshold methods because of the weak signal in the early stage of the fault. As time goes on, this abnormality will further deteriorate and cause catastrophic damage to wind turbines.

Figure 9 shows the wind speed, vibration, and the AAI curve of the data of WT7 before and after the failure. The light green area in the figure is periodic vibration abnormality occurred. It can be seen that before the anomaly occurrence, the AAI maintains fluctuations within a certain range during which we calculate the maximum AAI threshold $\mathcal{L} = 0.0051$ based on the training data. As can be seen from

Figure 9(b) and **Figure 9**(c), the wind turbine began to experience periodic abnormal vibrations on Feb. 14, 2016, however, the peak-to-peak value did not exceed the fixed alarm threshold $\pm 1.75\ m/\ s^2$ and trigger an alarm to shutdown the turbine operation. With the proposed method, this weak abnormal signal can be accurately detected. **Figure 9**(d) shows that through the AAI observed every hour, this abnormality can be detected almost at the same time when appearing.

The above cases demonstrate that the proposed method can quickly detect the weak abnormal changes in the vibration signals of the wind turbines, which are often submerged in complex interference. The NBM establishes a highly dynamic non-linear relationship between the vibration signal and these key signals causing the vibration through the deep learning network so that the environmental interference can be eliminated, thus enhancing anomaly detection. The proposed AAI considers signal changes over temporal domain as well as the direct coupling among the signals, which can further improve the detection accuracy. In these two cases, the processing time of 3600 sampling points per hour is about 0.6 seconds, which can ensure a real-time detection of anomalies.

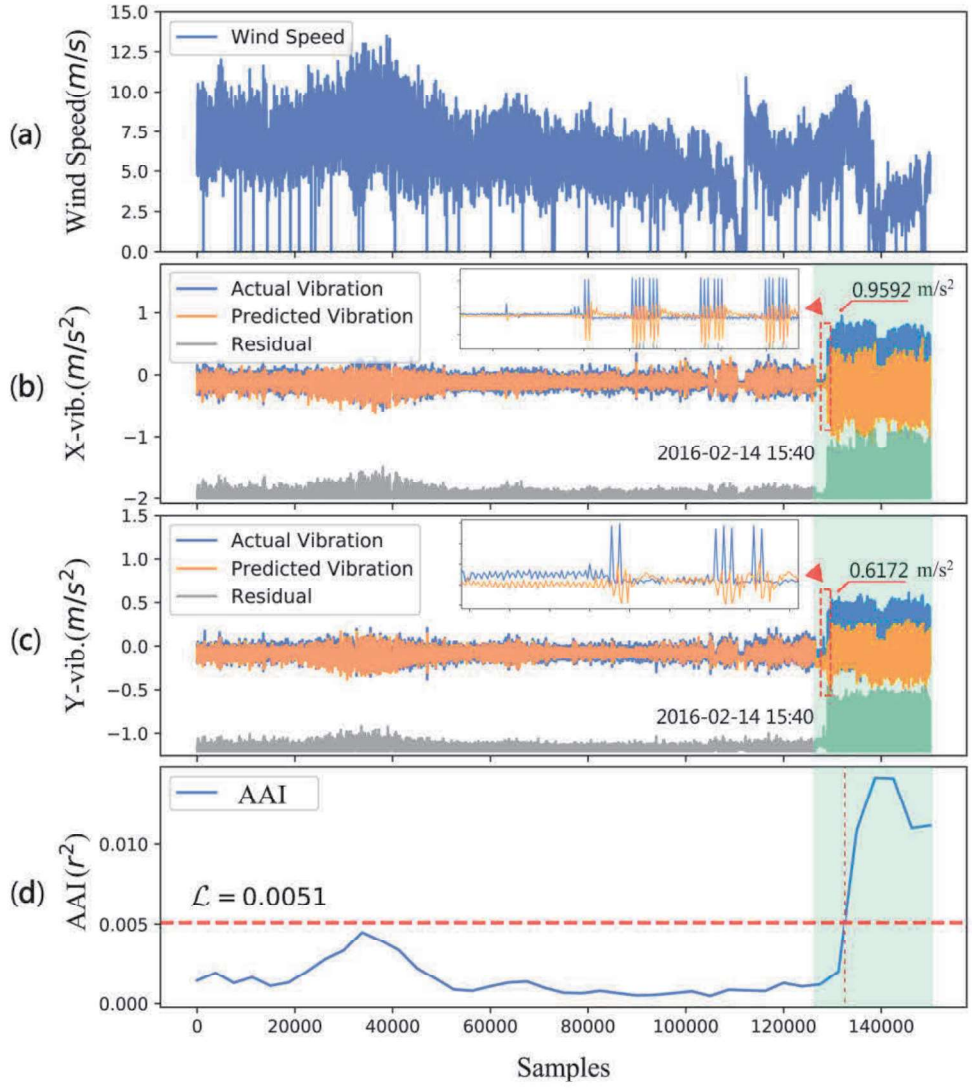

Figure 9 Periodic vibration detection result

4 Conclusion

In this paper, we present an anomaly detection framework of wind turbines considering nacelle vibration signals. The first step is to construct a NBM of wind turbines based on SETCN. This is novel when compared with currently prevailing time series vibration prediction methods with SCADA data. Through an input data augmentation in the training phase to learn temporal characteristics and spectral characteristics simultaneously, our model outperformed other well-established models and achieved better results on real-word data from wind turbines of different types under different working conditions. For anomaly detection, we proposed a MCV based AAI, which represents a quantified metric for anomaly assessment. AAI outperforms other methods by taking into account relative variability among parameters. Experiments are carried out with real-world wind farm data under complicated operating conditions. The results demonstrate that the proposed approach can deliver effective detection for high dynamic vibration anomaly even in situations where weak abnormal vibrations occur.

The proposed method in this paper has a potential to allow for a real-time monitoring and diagnosis as to whether the wind turbine operates normally or not without much prior knowledge on abnormal behavior detection. This would be significant to ensure the safe operation of wind turbines and improve their electricity production by incorporating the proposed method with the turbine controller. Should the methodology be adopted in wind turbines, it can impact the broad wind industry in a very positive manner. Future work will consider the influence of measurement noise and SCADA sampling frequency on the NBM. A more refined method to define threshold will also be explored to further enhance the universality of the framework.

Acknowledgment

This work is jointly funded by the National Science Foundation of China (U1811462), the National Key R&D project by Ministry of Science and Technology of China (2018YFB1003203), and the open fund from the State Key Laboratory of High Performance Computing (No. 201901-11). The funder CW took part in the formulation and development of methodology, and provided financial support for this study.

References

[1] G. W. E. Council, Global wind report 2019, Tech. rep. (2019).

[2] Q. Jiang, X. Yan, B. Huang, Deep discriminative representation learning for nonlinear process fault detection, IEEE Transactions on Automation Science and Engineering 17 (3) (2020) 1410-1419. doi:10.1109/TASE.2019.2956087 .

[3] M. N. Scheu, L. Tremps, U. Smolka, A. Kohos, F. Brennan, A systematic failure mode effects and criticality analysis for offshore wind turbine systems towards integrated condition based maintenance strategies, Ocean Engineering 176 (MAR.15) (2019) 118-133. doi:10.1016/j.oceaneng.2019.02.048 .

[4] Jannis, Tautz-Weinert, S. J. Watson, Using scada data for wind turbine condition monitoring ?€? a review, IET Renewable Power Generation 11 (4) (2016) 382-394. doi:10.1049/iet-rpg.2016.0248 .

[5] R. Uma Maheswari, R. Umamaheswari, Trends in non-stationary signal processing techniques applied to vibration analysis of wind turbine drive train - a contemporary survey, Mechanical Systems and Signal Processing 85 (2017) 296-311. doi:https://doi.org/10.1016/j.ymssp.2016.07.046 .

[6] R. Sun, Z. Yang, X. Chen, S. Tian, Y. Xie, Gear fault diagnosis based on the structured sparsity time-

frequency analysis, Mechanical Systems and Signal Processing 102 (2018) 346-363. doi:https://doi.org/10.1016/j.ymssp.2017.09.028 .

[7] M. Entezami, S. Hillmansen, P. Weston, M. Papaelias, Fault detection and diagnosis within a wind turbine mechanical braking system using condition monitoring, Renewable Energy 47 (2012) 175-182. doi:https://doi.org/10.1016/j.renene.2012.04.031 .

[8] F. Qu, J. Liu, Y. Ma, D. Zang, M. Fu, A novel wind turbine data imputation method with multiple optimizations based on gans, Mechanical Systems and Signal Processing 139 (2020) 106610. doi:https://doi.org/10.1016/j.ymssp.2019.106610 .

[9] J. P. Salameh, S. Cauet, E. Etien, A. Sakout, L. Rambault, Gearbox condition monitoring in wind turbines: A review, Mechanical Systems and Signal Processing 111 (2018) 251-264. doi:https://doi.org/10.1016/j.ymssp.2018.03.052 .

[10] W. Zhang, X. Ma, Simultaneous fault detection and sensor selection for condition monitoring of wind turbines, Energies 9 (4) (2016) 280. doi:10.3390/en9040280 .

[11] T. Matsui, K. Yamamoto, S. Sumi, N. Triruttanapiruk, Detection of lightning damage on wind turbine blades using the scada system, IEEE Transactions on Power Delivery 36 (2) (2021) 777-784. doi:10.1109/TPWRD.2020.2992796 .

[12] D. Astolfi, L. Scappaticci, L. Terzi, Fault diagnosis of wind turbine gearboxes through temperature and vibration data, International Journal of Renewable Energy Research 7 (2) (2017) 965-976.

[13] S. Xue, I. Howard, Torsional vibration signal analysis as a diagnostic tool for planetary gear fault detection, Mechanical Systems and Signal Processing 100 (2018) 706-728. doi:https://doi.org/10.1016/j.ymssp.2017.07.038 .

[14] C. Zhang, L. Cheng, J. Qiu, H. Ji, J. Ji, Structural damage detections based on a general vibration model identification approach, Mechanical Systems and Signal Processing 123 (2019) 316-332. doi:https://doi.org/10.1016/j.ymssp.2019.01.020 .

[15] P. Gangsar, R. Tiwari, Comparative investigation of vibration and current monitoring for prediction of mechanical and electrical faults in induction motor based on multiclass-support vector machine algorithms, Mechanical Systems and Signal Processing 94 (2017) 464-481. doi:https://doi.org/10.1016/j.ymssp.2017.03.016 .

[16] Y. Hu, C. Zhao, Fault diagnosis with dual cointegration analysis of common and specific nonstationary fault variations, IEEE Transactions on Automation Science and Engineering PP (99) (2019) 1-11. doi:10.1109/TASE.2019.2917580 .

[17] Peng, Guo, David, Infield, Wind turbine tower vibration modeling and monitoring by the nonlinear state estimation technique (nset), Energies 5 (12) (2012) 5279-5293. doi:10.3390/en5125279 .

[18] W. Xin, W. Hong, G. Jiang, X. Ping, X. Li, Monitoring wind turbine gearbox with echo state network modeling and dynamic threshold using scada vibration data, Energies 12. doi:10.3390/en12060982 .

[19] X. Jin, Z. Xu, W. Qiao, Condition monitoring of wind turbine generators using scada data analysis, IEEE Transactions on Sustainable Energy PP (99) (2020) 1-1. doi:10.1109/TSTE.2020.2989220 .

[20] L. Pedro, V. T. Luis, W. Matthias, K. Martin, P. Joachim, Normal behaviour models for wind turbine vibrations: Comparison of neural networks and a stochastic approach, Energies 10 (12) (2017) 1944. doi:10.3390/en10121944 .

[21] M. OGATA, J.and MURAKAWA, Vibration-based anomaly detection using flac features for wind turbine condition monitoring, EWSHM 2016.

[22] H. Sohn, C. R. Farrar, Damage diagnosis using time series analysis of vibration signals, Smart Material Structures 10 (3) (2001) 446-451. doi:10.1088/0964-1726/10/3/304 .

[23] G. Chao, Y. Yan, P. Hong, T. Li, W. Jin, Fault analysis of high speed train with dbn hierarchical ensemble, in: 2016 International Joint Conference on Neural Networks, 2016.

[24] H. Shao, H. Jiang, H. Zhao, F. Wang, A novel deep autoencoder feature learning method for rotating machinery fault diagnosis, Mechanical Systems and Signal Processing 95 (2017) 187-204. doi:10.1016/j.ymssp.2017.03.034 .

[25] K. Yan, J. Su, J. Huang, Y. Mo, Chiller fault diagnosis based on vae-enabled generative adversarial networks, IEEE Transactions on Automation Science and Engineering (2020) 1-9 doi:10.1109/TASE.2020.3035620 .

[26] J. Yu, X. Zhou, One-dimensional residual convolutional autoencoder based feature learning for gearbox fault diagnosis, IEEE Transactions on Industrial Informatics 16 (10) (2020) 6347-6358. doi:10.1109/TII.2020.2966326 .

[27] A. E. Elsaid, T. Desell, F. E. Jamiy, J. Higgins, B. Wild, Optimizing long short-term memory recurrent neural networks using ant colony optimization to predict turbine engine vibration, Applied Soft Computing 73. doi:10.1016/j.asoc.2018.09.013 .

[28] G. Toh, J. Park, Review of vibration-based structural health monitoring using deep learning, Applied Sciences 10 (5) (2020) 1680. doi:10.3390/app10051680 .

[29] T. Wang, Q. Han, F. Chu, Z. Feng, Vibration based condition monitoring and fault diagnosis of wind turbine planetary gearbox: A review, Mechanical Systems and Signal Processing 126 (2019) 662-685. doi:10.1016/j.ymssp.2019.02.051 .

[30] Y. Du, S. Zhou, X. Jing, Y. Peng, H. Wu, N. Kwok, Damage detection techniques for wind turbine blades: A review, Mechanical Systems and Signal Processing 141 (2020) 106445. doi:10.1016/j.ymssp.2019.106445 .

[31] B. Rezaeianjouybari, Y. Shang, Deep learning for prognostics and health management: State of the art, challenges, and opportunities, Measurement 163 (2020) 107929. doi:10.1016/j.measurement.2020.107929 .

[32] Y. Wang, L. Chen, Y. Liu, L. Gao, Wavelet-prototypical network based on fusion of time and frequency domain for fault diagnosis, Sensors 21 (4) (2021) 1483. doi:10.3390/s21041483 .

[33] M. Wu, F. Liu, T. Cohn, Evaluating the utility of hand-crafted features in sequence labelling, arXiv preprint arXiv:1808.09075 doi:10.18653/v1/D18-1310 .

[34] M. Hssayeni, S. Saxena, R. Ptucha, A. Savakis, Distracted driver detection: Deep learning vs handcrafted features, Electronic Imaging 2017 (10) (2017) 20-26. doi:10.2352/ISSN.2470-1173.2017.10.IMAWM-162 .

[35] M. G. Roychowdhury, S.; Diligenti, In image classification using deep learning and prior knowledge, Workshops at the Thirty-Second AAAI Conference on Artificial Intelligence 2018.

[36] V. K. S. Bai, J. Z. Kolter, An empirical evaluation of generic convolutional and recurrent networks for sequence modeling, arXiv preprint arXiv:1803.01271.

[37] O. B. Sezer, M. U. Gudelek, A. M. Ozbayoglu, Financial time series forecasting with deep learning: A systematic literature review: 2005-2019, Applied Soft Computing 90 (2020) 106181.

[38] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, G. D. Hager, Temporal convolutional networks for action segmentation and detection, in: proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 156-165. doi:10.1109/cvpr.2017.113 .

[39] P Lara-Benítez and M Carranza-García and Luna-Romera, J. M. and Riquelme, J. C., Temporal convolutional networks applied to energy-related time series forecasting, applied sciences 10 (7) (2020) 2322. doi:10.20944/preprints202003.0096.v1 .

[40] Y. Chen, Y. Kang, Y. Chen, Z. Wang, Probabilistic forecasting with temporal convolutional neural network, Neurocomputing 399 (2020) 491-501.

[41] W. Zhao, Y. Gao, T. Ji, X. Wan, F. Ye, G. Bai, Deep temporal convolutional networks for short-term traffic flow forecasting, IEEE Access 7 (2019) 114496-114507. doi:Deep temporal convolutional networks for short-term traffic flow forecasting .

[42] Z. Kong, B. Tang, L. Deng, W. Liu, Y. Han, Condition monitoring of wind turbines based on spatio-temporal fusion of scada data by convolutional neural networks and gated recurrent units, Renewable Energy 146. doi:10.1016/j.renene.2019.07.033 .

[43] Q. Peng, X. Ma, P. Cross, An integrated data-driven model-based approach to condition monitoring of the wind turbine gearbox, IET Renewable Power Generation 11 (9) (2017) 1177-1185. doi:10.1049/iet-rpg.2016.0216 .

[44] J. Zhan, R. Wang, L. Yi, Y. Wang, Z. Xie, Health assessment methods for wind turbines based on power prediction and mahalanobis distance, International Journal of Pattern Recognition & Artificial Intelligence 33 (2) (2019) 1951001.1-1951001.17. doi:10.1142/S0218001419510017 .

[45] L. V. Valen, Multivariate structural statistics in natural history, Journal of Theoretical Biology 45 (1) (1974) 235-247. doi:10.1016/0022-5193(74)90053-8 .

[46] M. Nikulin, V. Voinov, Unbiased Estimators and Their Applications, Unbiased estimators and their applications, 1996.

[47] J. Dai, X. Yuan, D. Liu, X. Long, X. Liu, Vibration analysis of large direct drive wind turbine nacelle based on scada system, Acta Energiae Solaris Sinica.

[48] F. Castellani, L. Garibaldi, A. P. Daga, D. Astolfi, F. Natili, Diagnosis of faulty wind turbine bearings using tower vibration measurements, Energies 13 (6) (2020) 1474. doi:10.3390/en13061474 .

[49] Y. Tao, X. Wang, R. S??nchez, S. Yang, Y. Bai, Spur gear fault diagnosis using a multilayer gated recurrent unit approach with vibration signal, IEEE Access 7 (2019) 56880-56889. doi:10.1109/ACCESS.2019.2914181 .

[50] R. Zhao, R. Yan, J. Wang, K. Mao, Learning to monitor machine health with convolutional bi-directional lstm networks, Sensors 17 (2) (2017) 273. doi:10.3390/s17020273 .

[51] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, arXiv.

[52] A. Brandt, A signal processing framework for operational modal analysis in time and frequency domain, Mechanical Systems and Signal Processing 115 (2019) 380-393. doi:https://doi.org/10.1016/j.ymssp.2018.06.009 .

[53]N Günnemann, and J. Pfeffer, Predicting defective engines using convolutional neural networks on temporal vibration signals 74 (2017) 92-102.