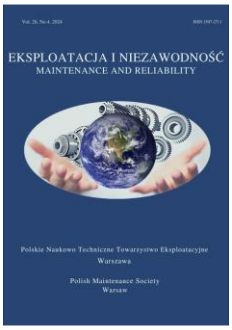


Eksploatacja i Niezawodnosc – Maintenance and Reliability
Volume 26 (2024), Issue 4
journal homepage: http://www.ein.org.pl

Article citation info:
Yang C, Li P, Lang X, Wu J, Wang H, A novel NMF-DiCCA deep learning method and its application in wind turbine blade icing failure identification, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2024: 26(4) http://doi.org/10.17531/ein/190381

A novel NMF-DiCCA deep learning method and its application in wind turbine blade icing failure identification

Indexed by: Web of Science Group

Chuangyan Yang[a], Peng Li[a,*], Xun Lang[a], Jiande Wu[b], Hancheng Wang[a]

[a] school of information science and engineering, Yunnan University, China
[b] the School of Engineering, and Yunnan Key Laboratory of Intelligent Systems and Computing, Yunnan University, China

Highlights

- The improved DiCCA algorithm based on NMF is used to extract the dynamic latent variables of nonlinear data.
- SSAE is utilized for extracting latent structural features, while GRU is employed to capture temporal correlation hidden features.
- Feature fusion using attention mechanism combines SSAE and GRU to achieve feature characterization.

Abstract

Wind turbine blade icing data has the characteristics of multi-source and multi-variable. It is difficult to characterize and identify the icing failure with multi-scale features. In this paper, a novel Non-negative Matrix Factorization-Dynamic-inner Canonical Correlation Analysis (NMF-DiCCA) based on Gated Recurrent Unit (GRU) and Stack Sparse Autoencoder (SSAE) algorithm is proposed to solve this problem. Firstly, using NMF instead of Singular Value Decomposition(SVD) decomposition method in DiCCA algorithm, the NMF-DiCCA is applied to obtain the dynamic latent variable of time serie. Secondly, the latent structure features S of dynamic latent variable is captured by SSAE. Thirdly, the temporal correlation hidden feature H of dynamic latent variable is extracted by GRU. Finally, the attention weight distribution between latent structure S and temporal correlation hidden feature H is integrated using the attention mechanism, and the fusion feature is reconstructed using the improved SSAE(ISSAE) based on GRU and SSAE. For the first time, dynamic latent variable analysis and deep learning representation algorithm are applied to icing failure identification of wind turbine blade, and effective results are obtained.

Keywords
wind turbine blade icing, NMF-DiCCAdynamic latent variable, deep learning representation, improved SSAE

1. Introduction

Among all kinds of new energy, wind energy has attracted extensive attention from all over the world due to its advantages such as huge reserves, no pollution, and wide distribution[1][2][3]. As the Wind power industry continues to develop, it provides renewable energy across the globe, the reliable, safe, and stable running of wind turbines has extremely crucial practical significance in recent years[4]. Therefore, ensuring the stable electric power system has a positive impact on the sustainable development ability of society. Wind turbine blades are vital in wind energy systems[5]. According to statistics, using a doubly-fed wind turbine as an example, wind turbine blades account for the largest proportion of the cost structure, amounting to up to 23.58%[6]. The operating efficiency of the determines the capabilities and power of the wind turbine installation[7]. However, a significant number of wind blades will freeze under the impact of terrible running

(*) Corresponding author.
E-mail addresses: C. Yang (ORCID: 0000-0002-7532-6836) yangchuangyan@mail.ynu.edu.cn, P. Li (ORCID: 0000-0001-9432-8995) lipeng@ynu.edu.cn, X. Lang (ORCID: 0000-0001-7380-6935) langxun@zju.edu.cn, J. Wu (ORCID: 0000-0003-2780-4925) jdwu@ynu.edu.cn, H. Wang (ORCID: 0009-0006-1944-456X) wanghancheng@mail.ynu.edu.cn.

environmental and climatic circumstances, which will not only cause significant economic losses but also serious safety accidents[8]. Therefore, one of the development trends is to achieve accurate detection of wind power blade icing failure[9].

In order to achieved the icing detection, the large amount of data is get, such as external environment, electrical and mechanical information[10]. The data-driven approach is often employed to transform the wind turbine blade icing detection issue into a classification or clustering problem, followed by forecasting the likelihood of wind turbine blade icing or accomplishing abnormality detection[11]. The data-driven approach detects icing conditions by mining common features of icing conditions from a large amount of training history data. It is applicable to academic and industrial circles because it does not need prior historical experience and relies on operational data. The blade icing detection approach, specifically based on deep neural networks, has attracted considerable interest, particularly as an end-to-end solution[12]. Hacıefendioğlu [13] and Yuan [14] proposed the icing detection method based on the fully connected neural networks. But the test accuracy of the model is low. Consequently, He [15] proposed a feature representation method based on SAE, the learning ability of wind turbine icing detection model is improved. But it is difficult to capture the nonlinear interactive effects between significantly different parameters. M.Priyatharishini[16] proposed a deep learning identification method using SSAE networks in integrated circuit industry. The SSAE model controls the sparsity of its hidden layer, allowing it to adeptly learn accurate sample feature expressions even in challenging scenarios, efficiently reducing the dimensionality of input information[17]. On the other hand, SSAE can extract spatial scale data features, but the SSAE model's accuracy will be affected as a result of the redundant features. Recently, Xiang[18] introduced a neural network-driven long-term memory model aimed at extracting features from multivariate data and capturing the relationships between input and output variables. However, the computational parallelism of long short-term algorithm is poor. Aim to slove the problem, Kong [19] presents a wind turbine blade icing detection technique using Convolutional Neural Network (CNN) and GRU, effectively identifying the wind turbine's operational state. However, it is challenging to determine which features are relevant and which are not for a deep learning network model[20]. Deep learning models have gained widespread usage in wind turbine blade icing detection to extract impactful features and mitigate the influence of redundant information on subsequent detection processes.

To sum up, significant constraints persist in both the extraction of diagnostic features for wind turbine blade icing and the construction of models. Many current ice detection methods rely on singular models, which inherently possess limited capacity to comprehend the intricate characteristics of the blade icing process and exhibit diminished generalization abilities[18]. Many researchers have suggested in recent years that the attention mechanism will employ the information released by the encoder and decoder to calculate the fault contribution weight coefficient matrix, enabling effective information to be extracted [21]. Therefore, in order to explore the internal relationship between different features, we combined with SSAE and GRU models to fuse the multiple features of wind turbines based on the attention mechanism. Finally, the neural network is trained so that it can accurately tell what state wind turbines are in.

Additionally, alongside the extensive volume of data, an essential characteristic of wind turbines blade data is the multitude of variables. While deep learning networks excel at learning and computing associations within vast datasets, they encounter challenges in performing efficient feature selection in the presence of numerous variables[22]. This can impact the accuracy of fault detection and diminish the model's interpretability. To evaluate the correlation between two sets of data to achieve feature extraction. An improved fault detection method for industrial processes is proposed to solve the problem that traditional Canonical correlation analysis (CCA) is not effective in dealing with early multiplicative fault detection[23]. Following this, to address non-Gaussian challenges encountered in industrial processes, a novel non-Gaussian process fault detection method has been introduced. This method amalgamates generalized canonical correlation analysis with stochastic algorithms, thereby combining the strengths of CCA and stochastic approaches [24]. Nevertheless, these approaches overlook the dynamic nature of the process and are exclusively applicable to static processes. To solve this problem, Dynamic CCA (DCCA) is developed, combining traditional

CCA with data augmentation model [25]. In addition, in order to avoid the shortcomings of direct matrix augmentation, the Dynamic Inner canonical correlation analysis (DiCCA) algorithm is proposed to extract past prediction methods that maximized canonical correlation, and to maintain both dimensionality reduction and order of prediction[26-27].

DiCCA utilizes Singular Value Decomposition (SVD) to decompose data, enabling the extraction of pertinent features. This process aims to capture the internal structure and correlations within the data, especially in scenarios where low rank approximation is necessary. [28]. Non-Negative Matrix Factorization(NMF) is another approach used to decrease matrix rank. It breaks down the data matrix into the product of non-negative matrices [29-30]. NMF is more effective for non-negative data and nonlinear structures, whereas SVD performs better with linear and Gaussian data. Given that real production process data often exhibit nonlinear and non-stationary traits, this paper opts for the NMF method over SVD decomposition.

In summary, this research provided three contributions:

1) The improved DiCCA algorithm based on NMF is used to extract the dynamic latent variables of nonlinear data.

2) SSAE is utilized for extracting latent structural features, while GRU is employed to capture temporal correlation hidden features.

3) Feature fusion using attention mechanism combines SSAE and GRU to achieve feature characterization.

4) A wind farm open data set is used to verify the effectiveness of the proposed method.

2. Related Worked

2.1. Dynamic-inner Canonical Correlation Analysis

From the data matri $\mathbf{X}=[x_1\ x_2\ \dots\ x_{s+N}]^T$, we form the subsequent smaller matrices:

$$X_i=[x_{i+1}\ x_{i+2}\ \dots\ x_{i+N}]^T \text{ for } i=0,1,\dots, \tag{1}$$

and the respective subsets of latent scores $t_i=\mathbf{X}_i\mathbf{w} \in P^N$ for $i=0,1,\dots,s$. where $\|\mathbf{w}\|^2=1$, the latent factor t_i is presumed to adhere to a normal distribution without autocorrelation, s is orderuncorrelated in time for a high enough.

Then defines t_s as current block vector with other t_i as the time-lagged vectors. The prediction $\hat{t}_i$ of t_s can be written as follows based Autoregressive (AR) model:

$$\hat{t}_s=\sum_{i=1}^{s}\beta_i\mathbf{t}_{s-i}=\sum_{i=1}^{s}\beta_i\mathbf{X}_{s-i}\,\mathbf{w} \tag{2}$$

DiCCA aims to optimize the correlation among the latent variables t_i and its prediction $\hat{t}_i$ as follows:

$$\max_{w,\beta}\frac{\mathbf{t}_s{}^T\hat{\mathbf{t}}_s}{\|\mathbf{t}_s\|\left\|\hat{\mathbf{t}}_s\right\|} \tag{3}$$

where $\|\boldsymbol{t}_s\|^2=1$, $\left\|\hat{\boldsymbol{t}}_s\right\|^2=1$. Applying Lagrange multipliers to (3) and using (2) get.

$$L=\mathbf{w}^T\mathbf{X}_s^T\sum_{i=1}^{s}\beta_i\mathbf{t}_{s-i}\,\mathbf{w}+\frac{1}{2}\lambda_1(1-\mathbf{w}^T\mathbf{X}_s\mathbf{w})+\frac{1}{2}\lambda_2\big(1-\mathbf{w}^T\sum_{i=1}^{s}\beta_i\mathbf{t}_{s-i}\sum_{i=1}^{s}\beta_i{}^T\mathbf{t}_{s-i}^T\,\mathbf{w}\big) \tag{4}$$

Taking the derivatives of L with respect to β and $\boldsymbol{w}$ and equating them to zero:

$$\frac{\partial L}{\partial\beta}=0,\frac{\partial L}{\partial\mathbf{w}}=0 \tag{5}$$

Calculate $\max_{w,\beta} J=\boldsymbol{t}_s{}^T\hat{\boldsymbol{t}}_s$; to ensure for convergence. To extract the next DLV, examining the next DLV involves applying the same iteration procedure to the subsequent deflated **X** matrix: $\mathbf{X}=\mathbf{X}-\mathbf{t}\mathbf{P}^T$, where the loading vector $\mathbf{P}=\mathbf{X}^T\mathbf{t}/\mathbf{t}^T\mathbf{t}$.

To deal with the rank-deficiency of **X**, a full-rank decomposition is perform follows: $\mathbf{X}=\mathbf{U}\mathbf{C}$. Where **U** has full colum rank and **C** has full row rank.The latent variable scores will be: $\mathbf{t}=\mathbf{X}\mathbf{w}=\mathbf{U}\mathbf{C}\mathbf{w}$.

SVD and NMF are two commonly used matrix decomposition techniques that can be employed to uncover hidden structures or features within a data matrix, considering matrix **U** as the data matrix **X**.

SVD breaks down a matrix into the product of three matrices:$\boldsymbol{X}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$,where **U** and **V** are orthogonal matrices, the matrix **Σ**, comprising singular values, is diagonal and encapsulates the features and configuration of matrix **X**. In this decomposition, matrix **U** encompasses the left singular vectors of **X**, matrix **V** holds the right singular vectors, and Σ contains the singular values. Through SVD, one can identify the feature vectors and singular values that maximize the representation of information present in the original matrix.

2.2. SSAE deep neural network

Using stacked Autoencoder can effectively extract data features, but the generalization of stacked AE is poor. In order to improve the feature extraction capability of SAE, sparsity is introduced to constrain the hidden layer. The sparse constraint principle: active units is employed to represent features in the hidden layer, get a sparse representation from a single encoder. The SAE model is shown in Fig. 1

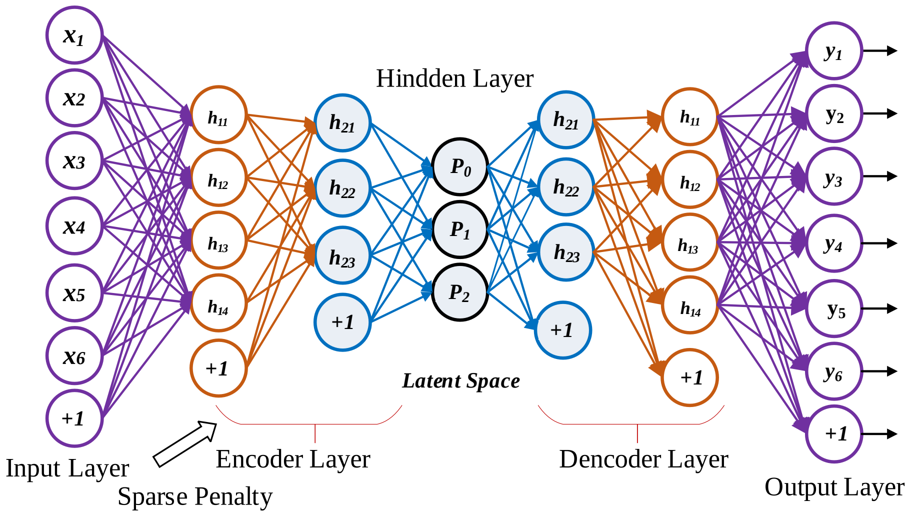


Fig.1. The SSAE model.

A regular term $\Omega_{Sparsity}$ Sparsity is added to construct Sparsity. The sparsity parameter ρ is used to constrain the sparsity of hidden layer. The *KL* divergence is used to evaluate the distribution between ρ and the estimate of , as follows:

$$\rho_i = \frac{1}{N} \sum^{\substack{j=1\\N}} h(\mathbf{w}_i^T x_j + b_i) \tag{6}$$

$$\hat{\rho}_i = \frac{1}{N} \sum^{\substack{j=1\\N}} h(\mathbf{w}_i^T \rho_j + b_i) \tag{7}$$

$$\Omega sparsity = \sum_{i=1}^{D(1)} KL(\rho \parallel \hat{\rho}_i) = \sum_{i=1}^{D(1)} \rho \, log(\frac{\rho}{\hat{\rho}_i}) + (1 - \rho) log(\frac{1-\rho}{1-\hat{\rho}_i}) \tag{8}$$

where N is the number of training samples, x_j is the jth training data, x is the ith row of the weight matrix w, and b_i is the ith entry of the bias vector. When the output value of a neuron is close to 1, it is considered to be active. On the contrary, if it is close to zero, I t is considered inhibited. The cost function as follows:

$$\Omega weights = \frac{1}{2} \sum_{l=1}^{L} \sum_{j=1}^{N} \sum_{i=1}^{K} (\mathbf{w}_{ji})^2 \tag{9}$$

the L is the number of hidden layers, K is the output data length of the hidden layer, and N is the input data length of the hidden layer[31], and weight attenuation term, as follows:

$$co\,s\,t\,function = \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K} (x_{kn} - \hat{x}_{kn})^2 + \lambda * \Omega sparsity + \beta * \Omega weights \tag{10}$$

where β is the sparse penalty factor, which λ is the weight attenuation coefficient.

Hierarchical training of stacked SAE can activate the first layer features and represent the second layer features. Ensure that previous features are stable and ignore interactions with subsequent features. The output dimension reduction information contains the essential features of the original data and removes the interference in the high-dimensional signal.

2.2. The feature extraction based on GRU

The hidden target information is extracted efficiently by GRU. The GRU contains update gates and reset gates[32]. The straightforward architecture of GRU significantly reduces runtime without compromising prediction accuracy[33]. The GRU structure is shown in Fig. 2.

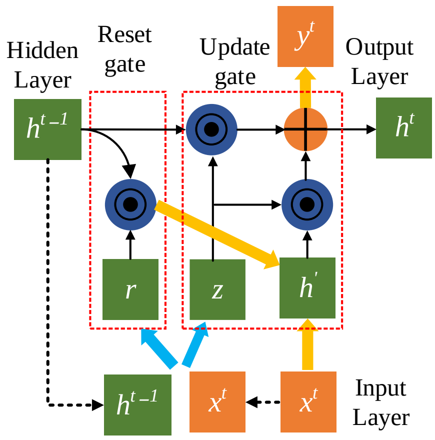


Fig.2. Schematic diagram of GRU structure.

The input layer is $\mathbf{X}=\{x_1,x_2,\ldots,x_t\}$. The hidden state h^{t-1} encapsulates information from the preceding node. Where z and r represent the update gate and reset gate, respectively.

$$h^{t-1\prime} = h^{t-1} \odot r \quad (11)$$

The $h^{t-1\prime}$ is concatenated with the input x_t. $\odot$ is the Hadamard Product, which is the multiplication of the corresponding elements in the operation matrix. $\oplus$ represents matrix addition operation[34], update the expression as follows:

$$h^{t-1\prime} = (1-z) \odot h^{t-1} + z \odot h' \quad (12)$$

3. The proposed method

A novel algorithm named NFM-DiCCA-ISSAE is proposed to solve the identified problems[35]. The structure of the proposed model is shown in Fig.3

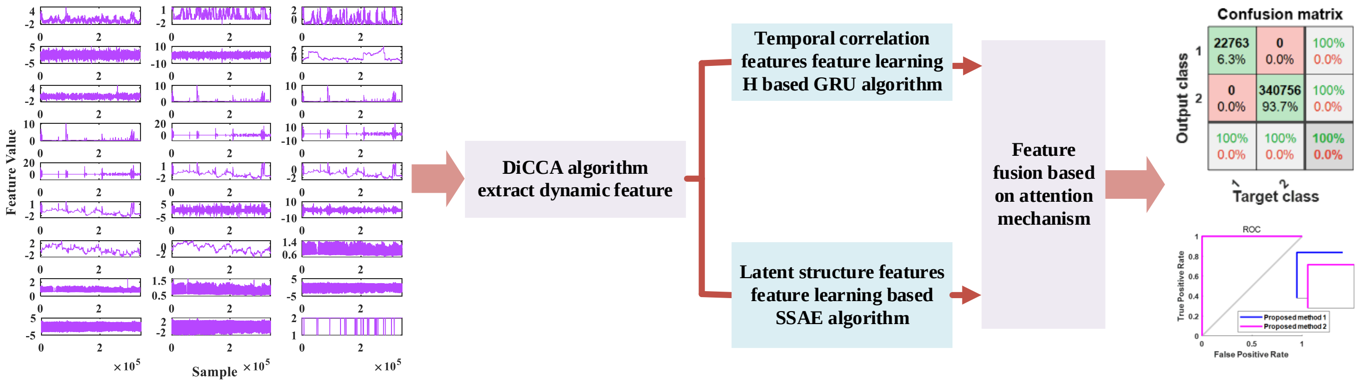


Fig.3. The structure of the proposed model.

Dynamic latent variable feature extraction: the algorithm replaces the traditional SVD decomposition in the DiCCA algorithm with NMF. The NMF-DiCCA method is used to process the time series data and extract the dynamic potential variables from it. These variables capture the temporal evolution within the data set, which is crucial for understanding the underlying dynamic properties.

Latent structure features feature learning: After extracting dynamic potential variables, the SSAE method is used to reveal the internal structural features of these variables, which are marked as *S*. This second stage involves the use of SSAE techniques to identify underlying patterns and structures within the dynamic latent variables obtained in the previous NMF-DiCCA step.

Temporal correlation features feature learning: revealing these structural features enriches the understanding of the inherent characteristics of the data.The final stage involves the use of gated cycle units (GRU) to capture temporal correlations and hidden features within dynamic latent variables. By utilizing GRUs, the model captures the time dependencies and hidden patterns present in dynamic latent variables.

Feature fusion: the attention weight distribution between the structure feature S and the time-varying hidden feature H is calculated using the attention mechanism. This attention mechanism is helpful to reconstruct and fuse the features based on the calculated attention weight distribution, so as to obtain a finer representation.

3.1. Dynamic latent variable feature extraction based NMF-DiCCA algorithm

In practical applications, when it comes to dealing with non-negative and nonlinear data, NMF shows its unique advantages. Compared with other methods such as SVD, NMF is more suitable for capturing nonlinear structures and non-negative features. Its non-negative constraint enables NMF to better reflect the local characteristics and nonlinear relations of data in many cases.

NMF decomposes a non-negative matrix into the product of two other non-negative matrices: **X**≈**WH**, where **W** and **H** are non-negative matrices, and their multiplication approximates the original matrix **X**. NMF is commonly used in text mining, image processing, and other fields. It is suitable for data that necessitates non-negativity constraints and often provides more intuitive interpretability.

3.2. The feature set is established based on SSAE and GRU

There is class imbalance in wind turbine blade icing detection, and the traditional method of resampling is usually used to reconstruct the data to solve the class imbalance problem[35-37]. However, resampling data is highly random, it is easy to cause important sample information missing or sample space overlap leading to overfitting.

To avoid the problems, features obtained through feature learning can provide clear classification boundaries. A set of features with good discriminative ability, and can alleviate the problem of unbalanced data. New features are extracted from original features by function mapping, it is assumed that there are n original features (or attributes) represented as A_1, A_2, …, A_n, another feature set can be obtained, it denoted as B_1, B_2, …, $B_m(m<n)$. The feature extraction formula is expressed as $B_i=f_i(A_1, A_2, \ldots, A_n)$, $i\in[1,m]$, and f is the corresponding function mapping. Then, the new features are used to replace the original features, and finally m features are obtained. The SSAE model is sensitive to the input and can extract more hidden information, so it is easier to obtain reasonable classification boundaries. However, GRU has low sensitivity to input and can learn robust information from data and has stable expression ability.

Therefore, feature set is generated by combining two independent neural networks, SSAE and GRU, which capture the global, stable and hidden features of the original data. The different features obtained are combined together for better expressiveness.

3.3. Feature fusion method based on attention mechanism

The SSAE is essentially a nonlinear dimensionality reduction method, which compresses the original redundant high-dimensional dataset to effectively extract data features. However, it cannot ensure that the feature set after dimensionality reduction contains highly similar information to the original data. Although the GRU network has poor ability to capture global effective features, more comprehensive local hidden information can be obtained by directly adding the linear dependence between the current state and the previous data poin[34]. Therefore, a feature fusion method with attention mechanism is proposed. Attention mechanisms can direct the attention of neural networks to valuable features. Based on how well they work together, the attention layer may change the weights of individual neurons in order to make their connections stronger[38-39]. The steps are as follows:

Step1: The similarity between H and K(K is the key node of the original signal input) is calculated by using the dot product method, denoted by f: $f(H,K_i)=H^TK_i$, i=1,2,…,m.

Step2: The similarity is normalized, weight matrix a_i is get, as follows:

$$a_i = soft\,max(\frac{f(\mathbf{H},\mathbf{K}_i)}{\sqrt{d_k}} \tag{13}$$

where d is the variance.

Step3: For the scattered weight a_i, all the values $\mathbf{V}$ are weighted and summed to get the Attention vector, as follows:

$$Attention = \sum_{i=1}^{m} a_i\,\mathbf{S}_i \tag{14}$$

3.4. Flowchart of the proposed method

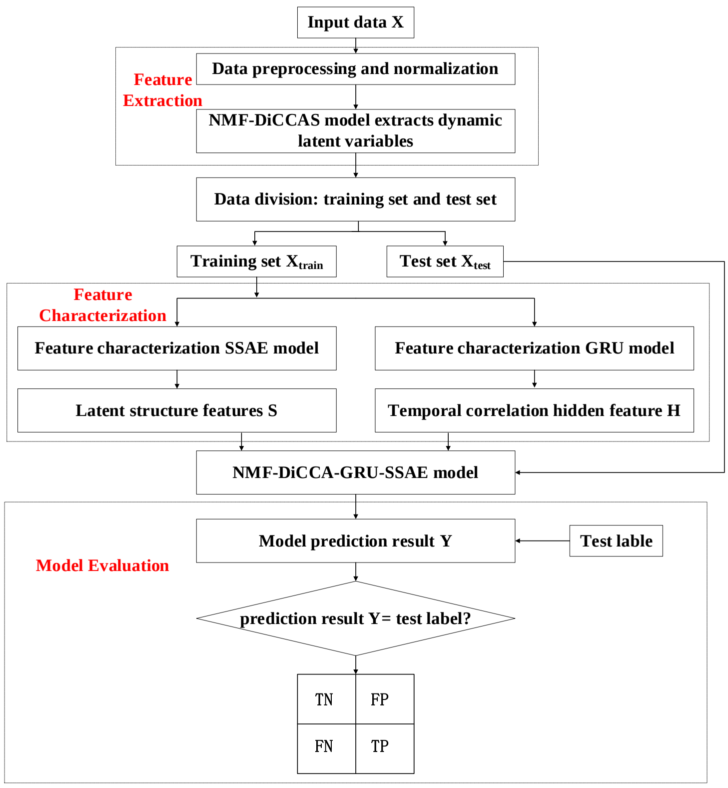


Fig.4. Flowchart of the proposed method.

The flowchart of the proposed method is shown in Fig.4.

Step1: The input data **X** is normalized pre-processed, and the dynamic latent variable features are extracted using the NFM-DICCA model.

Step2: Select the appropriate data partitioning ratio and divide the dynamic latent variable features into training set and test set.

Step3: Input the training set into the SSAE model and the GRU model respectively, and extract the latent structure feature S and the temporal correlation hidden feature H respectively.

Step4: Feature S and feature H are fused using the attention mechanism.

Step5: Feature S and feature H are fused by the attention mechanism, and a trained NMF-DiCCA-ISSAE model is established.

Step6: Input the test set into the trained model NMF-

DiCCA-ISSAE, and get the prediction result **Y**.

Step7: Determine whether the prediction result **Y** matches the test label? And obtain the confusion matrix.

4. Experimental Analysis

4.1. Data description

The SCADA data used in this study originate from the platform http://www.industrial-bigdata.com, which contains 26 consecutive numerical variables screened by WT manufacturers. The public dataset consists of the SCADA system data of two direct drive wind turbines A1(15#) and A2(21#) of a wind farm during the winter ice period [40-41], including 26 characteristics as shown in Table 1. A detailed explanation of the data is shown in Table 2. In this experimental, we select 80% of all samples for offline modeling and the remaining 20% for online monitoring.

Table 1. Data acquisition by SCADA system.

Serial Number	Feature	Serial Number	Feature
1	Wind speed	14	Temperature of nacelle
2	Temperature of pitch motor	15	Pitch angle of blade 1
3	Temperature of pitch motor 2	16	Pitch angle of blade 2
4	Temperature of pitch motor 3	17	Pitch angle of blade 3
5	Generator speed	18	Temperature of ng5 1
6	The active power of grid-side	19	Temperature of ng5 2
7	Average wind direction in 25 s	20	Temperature of ng5 3
8	Wind direction	21	Pitch speed of blade 1
9	Acceleration in x-direction	22	Pitch speed of blade 2
10	Acceleration in y-direction	23	Pitch speed of blade 3
11	Yaw position	24	DC of charger of ng5 1
12	Temperature of environment	25	DC of charger of ng5 2
13	Yaw speed	26	DC of charger of ng5 3

Table 2. Detailed Introduction of data.

Data type	Acquisition time	Total samples	Feature dimension
A1 (15#)	2015 Nov. 1st-2016 Jan. 1st	363,519	26
A2 (21#)	2015 Nov. 1st-2015 Nov. 30th	179,124	26

4.2. Evaluation Index

The confusion matrix is the most fundamental, intuitive, and straightforward approach to determining the accuracy of a classification model. The number of observations classified by the classification model into the negative and positive classes is counted. The following evaluation indicators are used to evaluate how effectively the classification algorithm is performing, as shown in Table 3.

Table 3. Detailed introduction of data.

Evaluation index	Equation	descriptive meaning
Accuracy (*ACC*)	$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$	the ratio of true cases to all cases
Precision(*PRE*)	$Precision = \frac{TP}{TP+FP}$	The percentage of correct positive samples in total positive samples
Sensitivity(*SEN*)	$Sensitivity = \frac{TP}{TP+FN}$	the probability that the results of TP is correctly classified
Specificity(*SPE*)	$Specificity = \frac{TN}{TN+FP}$	the number of negatives that are correctly classified

The precision and recall output results are combined by the F1-score. The F1-score is between [0,1], If the output is close to 1, it is the best output of the model; otherwise, close to 0 is the worst output of the model, given by equation (15).

$$F_1 = \frac{2TP}{2TP+FP+FN} \tag{15}$$

The False Positive Rate (FPR) and True Positive Rate (TPR) values, respectively, are the ordinate and abscissa of the Receiver Operating Characteristic(ROC). The ROC curve for each image relates to a specific model. The points of the ROC curve consist of FPR and TPR. To create the ROC curve, various points made up of [FPR, TPR] is obtained under different model thresholds [0,1].

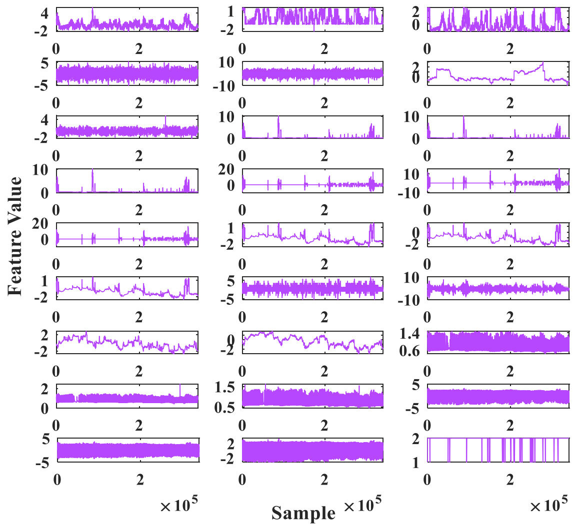
$$TPR = \frac{TP}{FN+TP} \tag{16}$$

$$FPR = 1 - \frac{FP}{FP+TN} = 1 - recall_{negative} \tag{17}$$

4.3. Test process and result analysis

The 26 characteristic time-domain waveforms of A1(15#) wind turbine blade is taken from the SCADA system[41], as shown in Fig.5. The 26 features in the Figure 5 correspond to the wind turbine blade data collected in Table 1

(*a*). The photo of wind turbine's frozen blades.



(*b*). The 26 features time-domain waveforms of A1(15#) wind turbine blades.

Fig.5. The wind turbine blade data.

It can be seen in Fig. 5, the wind turbine blades icing data set is collected with multi-source features, and its feature curves present the characteristics of jump and global instability. It is difficult to comprehensively characterize the real operating conditions of the wind turbine with multi-source characteristics. Therefore, the icing failure cannot be directly identified from the time domain waveform of feature set.

To solve this problem, feature learning is used to extract the features of wind turbines to improve the accuracy of ice detection. A series of methods AE、SAE、SSAE、GRU、ISSAE、DiCCA-ISSAE and NMF-DiCCA-ISSAE are used to extract useful information. According to the training set and the test set, three groups of comparative experiments are divided. The hyperparameters of the NMF-DiCCA-ISSAE Model are shown in Table 4.

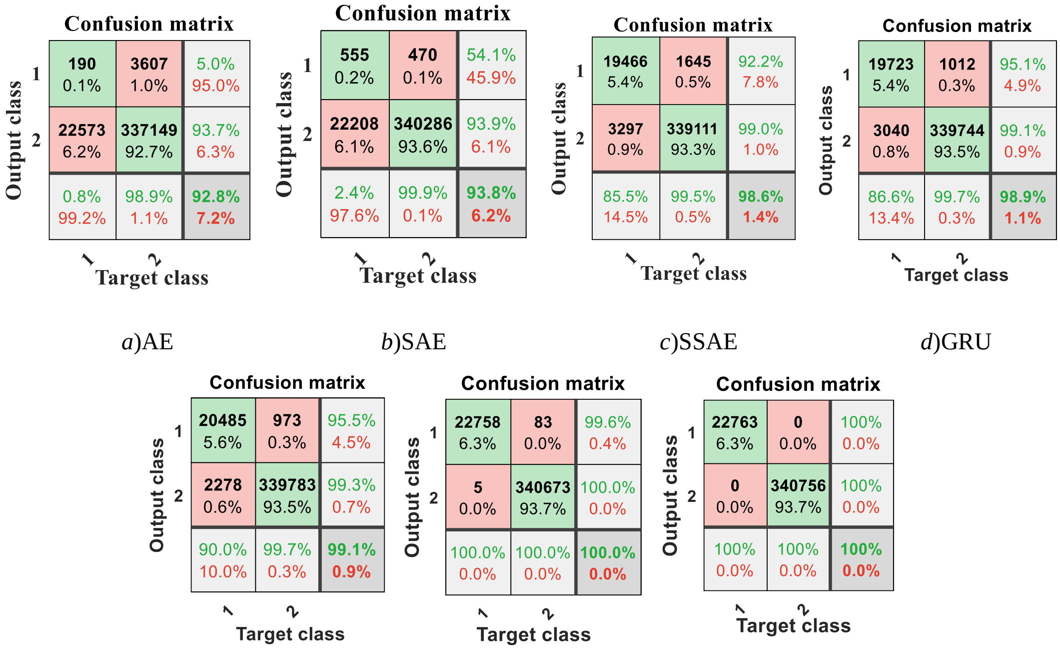
Table 4. Hyperparameters of the NMF-DiCCA-ISSAE Model.

Parameter	Value
Hidden Size	50
Encoder Function	satlin
Decoder Function	purelin
L2 Regularization	0.01
Sparsity Regularization	0.01
Sparsity Proportion	0.1

4.3.1. Experimental verification 1:

Considering the data set size, the data set is divided into two mutually exclusive sets using the set-aside method: the training set and the validation set, training set: test set = 8:2. All the data of A1(15#) wind turbine blades are input into the NMF-DiCCA-ISSAE model for training to get the trained model, Then all the data sets are input into the trained model as a test set to get the test results.

The results of the confusion matrix obtained by different methods are shown in Fig.6. From Fig.6*a*) to Fig.6*g*), it can be seen that different feature extraction methods have different detection rates.

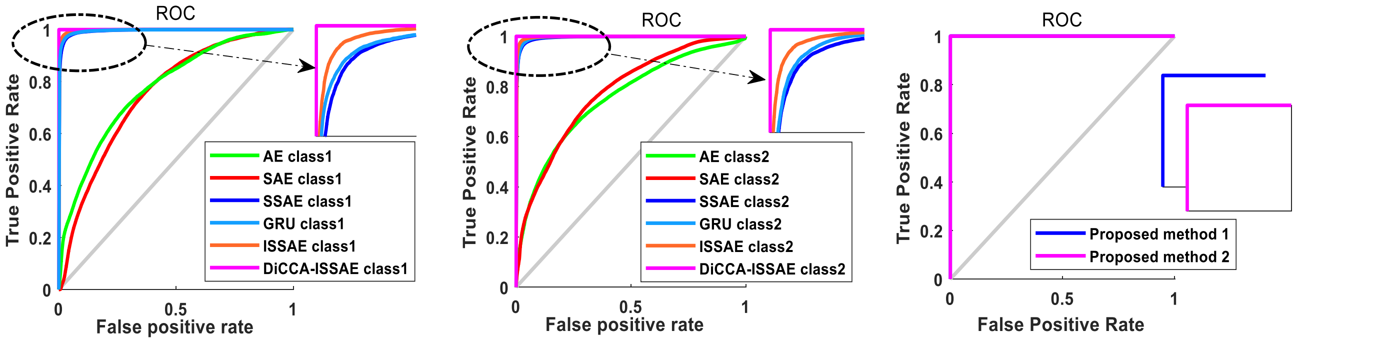


a)AE *b*)SAE *c*)SSAE *d*)GRU

e)ISSAE(GRU-SSAE) *f*) Dicca-ISSAE *g*)NMF-DiCCA-ISSAE

Fig.6. The confusion matrix obtained by different methods.

The ROC results obtained by different methods are shown in Fig.7. The ROC curve visualizes a model's performance. On the ROC curve, the horizontal axis represents the False Positive Rate (FPR), indicating the proportion of negative samples incorrectly classified as positive—a measure of false positives. The vertical axis represents the True Positive Rate (TPR), signifying the proportion of positive samples correctly identified as positive—a measure of true positives. This graph offers a comprehensive view of the model's performance across various thresholds. Fig.7*a*) shows the ROC curve for the class 1, showing the performance of the model in identifying or classifying the class 1. Typically, the closer the ROC curve aligns with the upper left corner, the superior the model's performance within that category. This alignment indicates higher True Positive Rates and lower False Positive Rates, reflecting better discriminative ability and accuracy of the model. Fig.7*b*) shows the ROC curve for class 2, showing the performance of the model in identifying or classifying the class 2. Similarly, the proximity of the curve to the top-left corner signifies better model performance within that category. This positioning suggests enhanced predictive accuracy, with higher true positive rates and lower false positive rates, reflecting the model's superior ability to discriminate between classes. Fig.7*c*) shows the ROC curve of the proposed improved method, showing high accuracy and reliability in terms of positive and negative samples.



a)ROC of class 1 *b*) ROC of class 2 *c*)Proposed method(NMF-DiCCA-ISSAE)

Fig.7. The ROC results obtained by different methods.

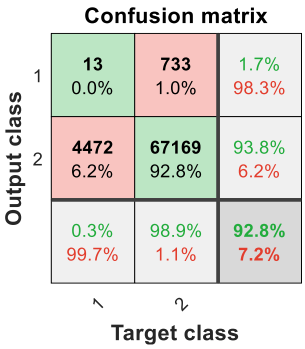
Table 5. Evaluation indicators of different methods for A1(15#).

A1(15#)	TPR	FPR	Accuracy	Precision	Sensitivity	Specificity	F_1 Score
AE	0.937249	0.05004	0.927982	0.989415	0.937249	0.05004	0.962626
SAE	0.938736	0.541463	0.937615	0.998621	0.938736	0.541463	0.967753
SSAE	0.990371	0.922079	0.986405	0.995172	0.990371	0.922079	0.992766
GRU	0.991131	0.655196	0.988279	0.99703	0.991131	0.655196	0.994072
GRU-SSAE	0.99334	0.954656	0.991057	0.997145	0.99334	0.954656	0.995239
DiCCA-ISSAE	0.999985	0.996366	0.999758	0.999756	0.999985	0.996366	0.999871
NMF-DiCCA-ISSAE	**1**	**1**	**1**	**1**	**1**	**1**	**1**

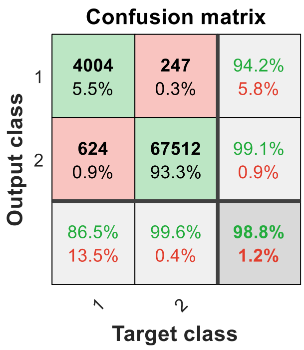
In order to quantify the advantages of different methods, the indicators(TPR、FPR、Accuracy(ACC)、Precision(PRE)、Sensitivity(SEN)、Specificity(SPE)、 F_1 Score) described in section 4.2 is used for evaluation, the results are shown in Table 5. It can be seen from Table 4 that the proposed method has achieved superior performance in different indicators.

4.3.2. Experimental verification 2:

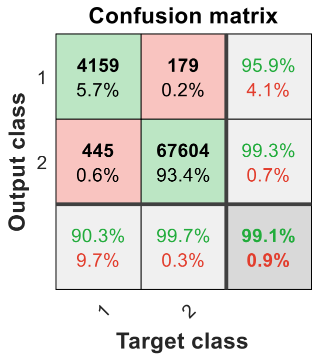
In order to verify the influence of different partitioning rules on the stability of the training set and the test set. The fixed model does not change, only how the training set and the test set are divided. The data of A1(15#) wind turbine blades are divided according to training set: test set = 8:2. 80% of the training set is input to the NMF-DiCCA-ISSAE model for training to get the trained model, and then 20% of the test set is input to the trained model to get the test result. The results of the confusion matrix obtained by different methods are shown in Fig.8. From Fig.8*a*) to Fig.8*g*), It can be seen that different feature extraction methods have different detection rates.



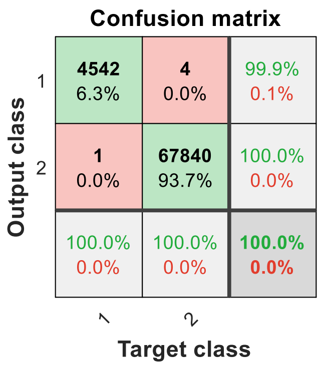
a)SAE



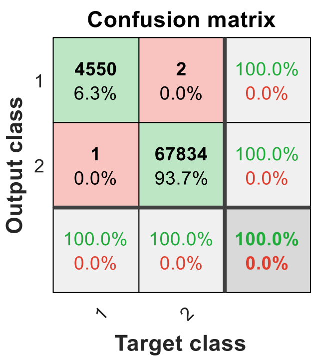
b)SSAE



c)ISSAE



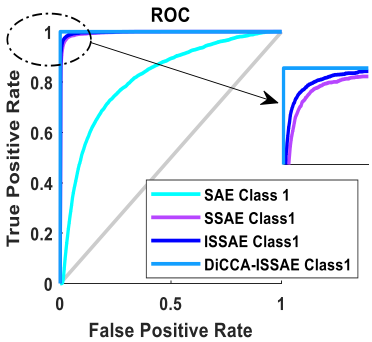
d) DiCCA-ISSAE



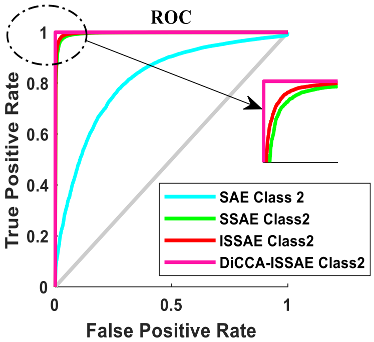
e) proposed method

Fig.8. The confusion matrix obtained by different methods.

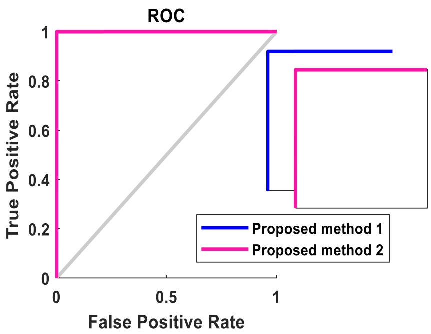
The ROC results obtained by different methods are shown in Fig.9. Fig.9*a*) shows the ROC curve for Class 1, showing the performance of the model in identifying or classifying Class 1 samples. Fig.9*b*) shows the ROC curve for Class 2, showing the performance of the model in identifying or classifying Class 2 samples.. Fig. 9*c*) shows the ROC curve of the improved method in this paper, which shows high accuracy and reliability on both positive and negative samples.



a)ROC of class 1



b) ROC of class 2



c)Proposed method(NMF-DiCCA-ISSAE)

Fig.9. The ROC results obtained by different methods.

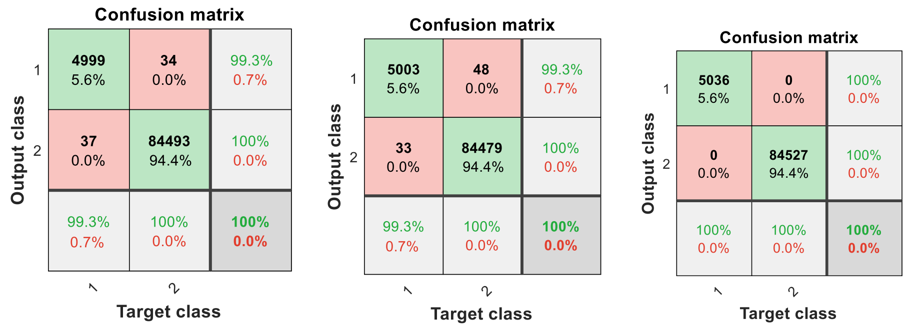
Table 6. Evaluation indicators of different methods for A1(15#).

A1(15#)	TPR	FPR	Accuracy	Precision	Sensitivity	Specificity	F_1 Score
SAE	0.937578	0.017426	0.928094	0.989205	0.937578	0.017426	0.962699
SSAE	0.990842	0.941896	0.987967	0.996355	0.990841	0.941896	0.99359
ISSAE	0.993460	0.958737	0.991379	0.99735	0.993460	0.958737	0.99540
DICCA-ISSAE	0.999852	0.999120	0.999541	0.999410	0.999852	0.999120	0.999631
NMF-DiCCA-ISSAE	**0.999985**	**0.999560**	**0.999959**	**0.999970**	**0.999985**	**0.999561**	**0.999978**

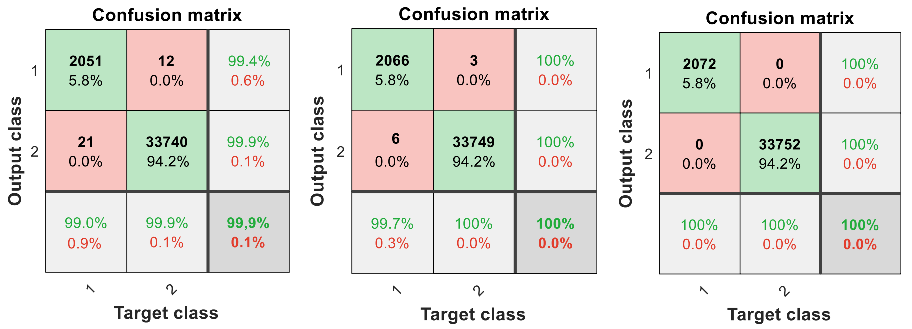
In order to quantify the advantages of different methods, indexes TPR, FPR, ACC, PRE, SEN, SPE and F_1 are used for evaluation, and the results are shown in Table 6. As can be seen from Table 5, the monitoring of the proposed model is the most effective, and the proposed model is not sensitive to the division of different data sets.

4.3.3. Experimental verification 3:

The validation results of the proposed method on A2(21#) wind turbine blades are also the most effective and the most accurate. The results of the confusion matrix obtained by different methods are shown in Fig.10. From Fig.10*a*) to Fig.10*f*), It can be seen that the detection rate of the proposed methods better than other methods. The evaluation indicator pairs are shown in Table 7, as can be seen from Table 7, the proposed method is the most superior regardless of how the training set and the test set are divided.



a)GRU-SSAE(5:5) *b*)DiCCA-ISSAE(5:5) *c*)NMF-DiCCA-ISSAE(5:5)



d)GRU-SSAE(8:2) *e*)DiCCA-ISSAE(8:2) *f*)NMF-DiCCA-ISSAE(8:2)

Fig.10. The confusion matrix obtained by different methods.

Table 7. Evaluation indicators of different methods A2(21#).

A2(21#)	method	TPR	FPR	Accuracy	Precision	Sensitivity	Specificity	F_1 Score
training set= test set	GRU-SSAE	0.999562	0.993245	0.999207	0.999598	0.999562	0.993245	0.999580
	DiCCA-ISSAE	0.999610	0.990497	0.999096	0.999432	0.999610	0.990497	0.999521
	NMF-DiCCA-ISSAE	**1**	**1**	**1**	**1**	**1**	**1**	**1**
training set: test set = 8:2	GRU-SSAE	0.999378	0.994183	0.999079	0.999644	0.999378	0.994183	0.999511
	DiCCA-ISSAE	0.999822	0.998550	0.999749	0.999911	0.999822	0.998550	0.999867
	NMF-DiCCA-ISSAE	**1**	**1**	**1**	**1**	**1**	**1**	**1**

5. Conclusion

In order to improve the accuracy of multi-source feature extraction and ice detection. It extracts unsupervised features by calculating the similarity between classification layer prediction and matching sample labels, and integrates the fault category information. The experimental results show that:

1)The application of NMF enhanced the performance of DiCCA model. Compared with SVD, NMF is more suitable for processing nonlinear data.

2)SSAE effectively extracted latent structural features; GRU successfully captured time-dependent hidden features. It better to comprehensive characterization of wind turbine blade icing characteristics

3) Experimental results show the effectiveness and reliability of the proposed method

Funding：

This work was supported in part by the National Natural Science Foundation of China(62163036)(Corresponding author: Peng Li);Major Project of Science and Technology of Yunnan Province (202402AD080001)(Third author:Xun Lang);Xingdian Talent Support Program - Industrial Innovation Talent Project (XDYC-CYCX-2022-0010) (Corresponding author: Peng Li);Yunnan Province Young and Middle-aged Academic Technology Leaders Training Program(202105AC160094) (Corresponding author: Peng Li);Yunnan Fundamental Research Projects under Grant(202301AT070277)(Third author:Xun Lang).

References

1. Meng D, Yang S, Jesus A, Fazeres-Ferradosa T, Zhu S. A novel hybrid adaptive Kriging and water cycle algorithm for reliability-based design and optimization strategy: Application in offshore wind turbine monopile. Computer Methods in Applied Mechanics and Engineering 2023; 412: 116083. https://doi.org/10.1016/j.cma.2023.116083.
2. Meng D , Yang S , Jesus A, Zhu S. A novel Kriging-model-assisted reliability-based multidisciplinary design optimization strategy and its application in the offshore wind turbine tower. Renewable energy 2023; 203, 407-420. https://doi.org/10.1016/j.renene.2022.12.062.
3. Meng, D., Wang, H., Yang, S., Lv, Z., Hu, Z. Fault Analysis of Wind Power Rolling Bearing Based on EMD Feature Extraction. CMES-Computer Modeling in Engineering and Sciences 2022; 130(1), 543–558. https://doi.org/10.32604/cmes.2022.018123.
4. Tao T, Liu Y, Qiao Y. Wind turbine blade icing diagnosis using hybrid features and Stacked- XGBoost algorithm. Renewable Energy 2021; 180: 1004-1013. https://doi.org/10.1016/j.renene.2021.09.008.
5. Liu L, Guan D, Wang Y, Ding C, Wang M, Chu M. Data-Driven Prediction of Wind Turbine Blade Icing, 2021 China Automation Congress , Beijing, China 2021; 5211-5216, https://doi: 10.1109/CAC53003.2021.9727866.
6. Yi H, Jiang Q. Discriminative feature learning for blade icing fault detection of wind turbine. Measurement Science and Technology 2020; 31(11). https://doi.10.1088/1361-6501/ab9bb8.
7. Xu C, Fan S, Liu Y, Liu X, Huang L. Wind turbine blade icing detection: a federated learning approach. Energy 2022; 254:124441. https://doi.org/10.1016/j.energy.2022.124441.
8. Chen W, Qiu Y. Diagnosis of wind turbine faults with transfer learning algorithms. Renewable Energy 2021; 163:2053-2067 https://doi.org/10.1016/j.renene.2020.10.121.
9. Yang X, Ye T. Diagnosis of Blade Icing Using Multiple Intelligent Algorithms. Energies 2020; 13(11):2975. https://doi:10.3390/en13112975.
10. Yoon T Kang Y. Proper orthogonal decomposition of continuum-dominated emission spectra for simultaneous multi-property measurements. Energy 2022; 254:124458. https://doi.org/10.1016/j.energy.2022.124458.
11. Zhang S, Lang Z. SCADA-data-based wind turbine fault detection: A dynamic model sensor method. Control Engineering Practice 2020; 102. https://doi. 10.1016/j.conengprac.2020.104546.
12. Jing L, Zhao M, Li P. A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox.Measurement 2017; 111:1-10. https://doi.org/10.1016/j.measurement.2017.07.017.
13. Kemal H, Hasan B. Intelligent ice detection on wind turbine blades using semantic segmentation and class activation map approaches based on deep learning method, Renewable Energy 2022; 182:1-16. https://doi.org/10.1016/j.renene.2021.10.025.
14. Yuan B, Wang C. WaveletFCNN: A Deep Time Series Classification Model for Wind Turbine Blade Icing Detection. ArXiv, 2019.
15. He Z, Shi T, Xuan J. Milling tool wear prediction using multi-sensor feature fusion based on stacked sparse autoencoders. Measurement 2022; 190:110719. https://doi.org/10.1016/j.measurement.2022.110719.
16. Priyatharishini M. A deep learning based malicious module identification using stacked sparse autoencoder network for VLSI circuit reliability. Measurement 2022; 194:111055. https://doi.org/10.1016/j.measurement.2022.111055.
17. Zhan X, Liu Z. A novel method of health indicator construction and remaining useful life prediction based on deep learning. Eksploatacja i Niezawodnosc-Maintenance and Reliability 2023.
18. Lei J, Liu C. Fault diagnosis of wind turbine based on Long Short-term memory networks. Renewable Energy 2019; 133:422-432. https://doi.org/10.1016/j.renene.2018.10.031.
19. Xiang L, Wang P. Fault detection of wind turbine based on SCADA data analysis using CNN and LSTM with attention mechanism[J]. Measurement 2021; 175(8):109094. https://doi:10.1016/J.MEASUREMENT.2021.109094.

20. Kong Z, Tang B. Condition monitoring of wind turbines based on spatio-temporal fusion of SCADA data by convolutional neural networks and gated recurrent units. Econpapers 2020; 146: 760-768, https://doi.org/10.1016/j.renene.2019.07.033.
21. Zhang J, Kong X, Cheng L, Qi H, Yu M. Intelligent fault diagnosis of rolling bearings based on continuous wavelet transform-multiscale feature fusion and improved channel attention mechanism. Eksploatacja i Niezawodność – Maintenance and Reliability 2023;25(1):16. https://doi:10.17531/ein.2023.1.16.
22. Yang S, Meng D,Wang H, Yang C. A novel learning function foradaptive surrogate-model-basedreliability evaluation. Phil. Trans. R.Soc 2024; 382: 20220395.http://doi.org/10.1098/rsta.2022.0395.
23. Xu J, Xiang L. Stacked Sparse Autoencoder (SSAE) for Nuclei Detection on Breast Cancer Histopathology Images. IEEE Transactions on Medical Imaging 2016; 35(1):119-130. https://doi: 10.1109/TMI.2015.2458702.
24. Chen Z, Zhang K, Ding S X. Improved canonical correlation analysis-based fault detection methods for industrial processes. Journal of Process Control 2016; 41:26-34. https://doi.org/10.1016/j.jprocont.2016.02.006.
25. Chen Z, Steven. Canonical correlation analysis-based fault detection methods with application to alumina evaporation process. Control Engineering Practice 2016;46:51-58. https://doi.org/10.1016/j.conengprac.2015.10.006.
26. Dong Y, Qin S. Dynamic-Inner Canonical Correlation and Causality Analysis for High Dimensional Time Series Data. IFAC-PapersOnLine 2018; 51(18):476-481. https://doi.org/10.1016/j.ifacol.2018.09.379.
27. Févotte, Cédric, Idier, Jérôme. Algorithms for nonnegative matrix factorization with the beta-divergence.Neural Computation 2010;23(9):2421 - 2456. https://doi: 10.1162/NECO_a_00168.
28. Yang C, Nie K, Qiao J, Danlei Wang. Robust echo state network with sparse online learning. Information Sciences 2022; 594. https://doi.org/10.1016/j.ins.2022.02.009.
29. Hei Z, Sun B, Wang G, Lou Y, Zhou Y. Multi-feature spatial distribution alignment enhanced domain adaptive method for tool condition monitoring. Eksploatacja i Niezawodność – Maintenance and Reliability 2023;25(4). https://doi:10.17531/ein/171750.
30. Y. Dong, Y. Liu S, Qin J. Efficient Dynamic Latent Variable Analysis for High-Dimensional Time Series Data. in IEEE Transactions on Industrial Informatics 2020; 16(6): 4068-4076, https://doi: 10.1109/TII.2019.2958074.
31. Wu Y, Zhang Y, Zou X. Estimated date of delivery with electronic medical records by a hybrid GBDT-GRU model. ISA transactions2022. https://doi.org/10.1038/s41598-022-08664-5.
32. Zheng G, Sun W, Zhang H, Zhou Y, Gao C. Tool wear condition monitoring in milling process based on data fusion enhanced long short-term memory network under different cutting conditions. Eksploatacja i Niezawodność – Maintenance and Reliability 2021;23(4):612-618. https://doi:10.17531/ein.2021.4.3.
33. Lyu Y, Zhang Q, Chen A, Wen Z. Interval Prediction of Remaining Useful Life based on Convolutional Auto-Encode and Lower Upper Bound Estimation. Eksploatacja i Niezawodność – Maintenance and Reliability 2023;25(2). https://doi:10.17531/ein/165811.
34. Li Y, Li Y. A homotopy gated recurrent unit for predicting high dimensional hyperchaos. Communications in Nonlinear Science and Numerical Simulation 2022; 115. https://doi.org/10.1016/j.cnsns.2022.106716.
35. Dai G, Wang X. MRGAT: Multi-Relational Graph Attention Network for knowledge graph completion. Neural Networks 2022; 154: 234-245. https://doi.org/10.1016/j.neunet.2022.07.014.
36. Sun C, Zhang Y, Huang G, Lin Liu, Xiaochen Hao. A soft sensor model based on long&short-term memory dual pathways convolutional gated recurrent unit network for predicting cement specific surface area. ISA Transactions 2022. https://doi.org/10.1016/j.isatra.2022.03.013.
37. Zhang H, Song C. Image-Model-Based Fault Identification for Wind Turbines Using Feature Engineering and MuSnet . IEEE Transactions on Industrial Informatics 2022;18(10): 6592-6601. https://doi: 10.1109/TII.2022.3157748.
38. Wang H, Li P, Lang X, FTGAN: A Novel GAN-Based Data Augmentation Method Coupled Time–Frequency Domain for Imbalanced Bearing Fault Diagnosis. IEEE Transactions on Instrumentation and Measurement 2023;72:1-14. https://doi: 10.1109/TIM.2023.3234095.
39. Dong Y, Qin, S. New Dynamic Predictive Monitoring Schemes Based on Dynamic Latent Variable Models. Industrial and Engineering Chemistry Research 2020;59(6), 2353-2365. Advance online publication. https://doi.org/10.1021/acs.iecr.9b04741.
40. Tong R, Li P, Gao L, Lang X, A. Miao and X. Shen. A Novel Ellipsoidal Semisupervised Extreme Learning Machine Algorithm and Its Application in Wind Turbine Blade Icing Fault Detection. IEEE Transactions on Instrumentation and Measurement 2022;71:1-16.

https://doi: 10.1109/TIM.2022.3205920.

41. Tong R, Li P, Lang X, Liang J, Cao M. A novel adaptive weighted kernel extreme learning machine algorithm and its application in wind turbine blade icing fault detection. Measurement 2021;185:110009. https://doi.org/10.1016/j.measurement.2021.110009.