IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 29, 2021, accepted February 22, 2021, date of publication March 1, 2021, date of current version May 14, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3062865

An Abnormal Data Processing Method Based on An Ensemble Algorithm for Early Warning of Wind Turbine Failure

QIANG ZHAO[1], KUNKUN BAO[1], ZHENFAN WEI[1], YINGHUA HAN[2], AND JINKUAN WANG[3]

[1]School of Control Engineering, Northeastern University at Qinhuangdao, Qinhuangdao 066004, China
[2]School of Computer and Communication Engineering, Northeastern University at Qinhuangdao, Qinhuangdao 066004, China
[3]College of Information Science and Engineering, Northeastern University, Shenyang 110819, China

Corresponding author: Kunkun Bao (1801875@stu.neu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant U1908213, in part by the Fundamental Research Funds for the Central Universities under Grant N182303037, in part by the Colleges and Universities in Hebei Province Science Research Program under Grant QN2020504, and in part by the Foundation of Northeastern University at Qinhuangdao under Grant XNB201803.

ABSTRACT In the modeling process of fault warning models, modeling data plays an important role, the quality of which affects the performance of the model. The manual selection of modeling data according to fault records is time-consuming and makes it difficult to guarantee the high quality of the data because of inconsistencies, errors, and losses of records in the fault log file. For this reason, the present study proposes a framework of abnormal data processing based on an unsupervised serialization ensemble algorithm, which considers the high dimensional characteristics of operational data and the relationship between known low dimensional variables. Meanwhile, the influence of modeling data using different data processing methods on fault prediction performance is studied. The improved stacked autoencoder (I-SAE) based on the idea of partial data reconstruction is proposed to learn the high dimensional characteristics of operational data, which can enhance the separability of normal and abnormal operational data. The improved density-based spatial clustering of applications with noise (DBSCAN) clustering algorithm based on the density ratio is developed to handle sparse abnormal data with data imbalance characteristics. Finally, the case analysis results demonstrate that the abnormal data processing model proposed in this article has better performance than other methods, and the performance of the fault prediction model can be effectively enhanced by improving the quality of modeling data.

INDEX TERMS Density ratio, early warning, serialization ensemble, stacked autoencoder.

I. INTRODUCTION

In recent years, with the deterioration of the environment via pollution and the depletion of traditional energy, clean and renewable wind energy has attracted worldwide attention and developed rapidly. However, wind turbines are generally installed in remote areas (such as mountains, seas, and plateaus), where they operate in harsh weather environments for a long time [1]. These harsh operating environments induce frequent malfunctions of wind turbines, which increases unscheduled downtime and repair costs. Hence, to reduce repair costs and improve the reliability of wind turbines, the development of fault diagnosis technology has aroused great interest in the wind energy industry in recent years.

Most approaches on the fault diagnosis and condition monitoring of wind turbines have been reported, which can be divided into model-based, knowledge-based, and data-driven methods [2], [3]. Depending on an accurate mathematical model, model-based techniques can be an effective fault diagnosis method, such as observer-based techniques [4], Kalman filter and estimators [5], and particle filter [6]. However, obtaining an accurate model is challenging in practice because of the complexity of the wind turbine system [7]. Knowledge-based techniques, including signal processing [8] (such as vibration signal, strain measurement, and acoustic

The associate editor coordinating the review of this manuscript and approving it for publication was Francesco Tedesco.

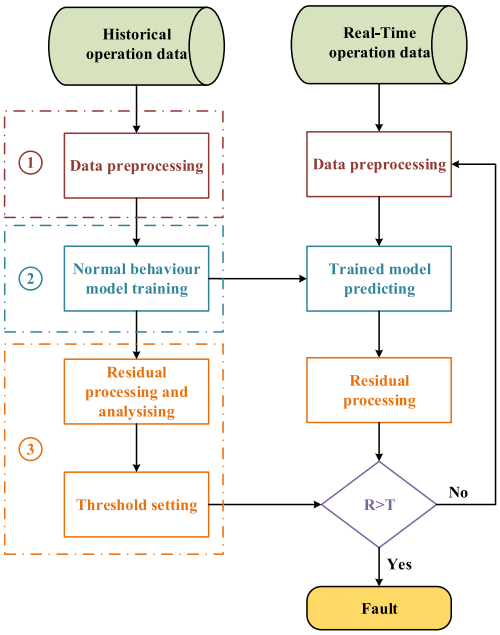


FIGURE 1. The general framework of wind turbine condition monitoring.

emission) and fuzzy logic methods [9], rely on deep expert knowledge. However, the difficulty of expert knowledge acquisition and expensive measurement equipment hinder the application of the knowledge-based method in practice. Data-driven methods monitor the operation status of wind turbines by mining the information in the operation data, which mainly involve machine learning and data mining methods [10], [11], such as artificial neural network, deep learning, Gaussian regression, and support vector machine. Because it does not need complex mathematical models and profound expert knowledge, data-driven methods have been developed rapidly in recent years. Supervisory control and data acquisition (SCADA) data, which contain abundant operational status information, has been widely proved to be effective in the field of wind turbine fault research [12], [13]. In the research on imminent fault early warning using SCADA data, methods based on the normal behavioral model are the most commonly used modeling methods [14]. Fig. 1 shows the general framework of wind turbine condition monitoring based on the normal behavioral model. Generally, the three aspects (modeling data, modeling method, and threshold setting method) in parts 1, 2, and 3 of Fig. 1 have an impact on the performance of fault early warning models. At present, most of research is focused on modeling methods (such as deep belief networks (DBN) [10], stacked autoencoder (SAE) [15], and convolutional neural networks (CNN) [16])) and threshold setting methods (adaptive threshold [10], [17], new residual index [18]–[20], and application of Mahalanobis distance [21], [22]) to improve the performance of fault diagnosis and condition monitoring. However, research on the influence of modeling data for fault prediction performance is rare and needs to be further developed. Sun *et al.* [23] analyzed the influence of different training sample data on the accuracy of wind turbine fault diagnosis, including current SCADA data, historical SCADA data from the studied wind turbine, and the current SCADA data from a similar wind turbine. Therefore, this article studies the influence of modeling data quality on fault prediction performance, enriches the knowledge system of fault prediction research.

In normal behavior-based modeling, modeling data plays an important role, the quality of which directly affects the performance of fault prediction models [24], [25]. The SCADA data collected from the wind turbine system contain many abnormal data, which are the data beyond the pattern of normal status. There are various reasons for abnormal data, including communication facility error, extreme weather conditions, human interference, and wind turbine fault. Depended on the intrinsic characteristics, the abnormal data can be divided into three types: the missing abnormal data, the isolated abnormal data, and the fault abnormal data. The missing abnormal data are usually caused by communication equipment failure, which are manifested as the absence of some attributes. The isolated abnormal data are the data far away from the normal data and are usually caused by extreme weather conditions, communication noise, and human factors. They usually appear as sparse data far away from normal operation data in SCADA data. The fault abnormal data are usually caused by the degradation and failure of wind turbine components (such as sensors, generators, and gearboxes) and wind curtailment. They usually appear as dense data different from normal operation patterns in SCADA data. Due to the obvious characteristics of missing attributes, missing abnormal data can be easily identified. In this article, isolated abnormal data and fault abnormal data are the main processing objects. When the abnormal data are mixed into the training samples, the normal behavior model can learn not only the expected normal state information but also the unexpected abnormal state information, which may lead to the phenomenon of missing an alarm. Therefore, it is important and necessary to deal with the abnormal data of the modeling data.

The operational data of wind turbines are typical multi-variable data that have complex high dimensional characteristics due to the coupling between wind turbine subsystems. However, because of the simplicity of existing modeling data processing methods, including manual operation based on the log files of fault records [26]–[28] and simple processing based on prior knowledge (such as deleting the data of a physical variable beyond a certain threshold [27], [29], [30]), it is difficult to identify abnormal data effectively. Moreover, due to the inconsistencies, losses, and errors of fault records, it is difficult to obtain high-quality modeling data even by manual data processing according to log files [31], [32]. Therefore, it is of great significance to develop an effective unsupervised abnormal data processing method for modeling data.

It is challenging to handle abnormal data for high dimensional data, especially using unsupervised methods.

Generally, the methods for processing high dimensional abnormal data based on an unsupervised model include the clustering method in high dimensional space [21], the dimensional reduction fusion clustering method [33], [34], and the anomaly scoring method based on reconstruction errors. Due to the influence of the 'curse of dimensionality', clustering directly in high dimensional space provides poor performance and has a high computational cost [35]. Although dimensional reduction makes clustering operations easier, the lack of prior knowledge in unsupervised methods leads to the separation of dimensional reduction and the clustering process, which makes it difficult to ensure that the features extracted from dimensional reduction are suitable for anomaly data detection [36], [37]. The method of abnormal scoring based on the reconstruction errors depends on the precondition that the normal data are much more numerous than the abnormal data, which is easy to accomplish in the operation data of wind turbines. On this basis, prior knowledge based on the method of abnormal scoring is that the reconstruction error of the model to the normal data is small by reconstructing the operational data, whereas the reconstruction error of the abnormal data is relatively large. Therefore, this kind of method is suitable for the data processing of the wind turbine operational data. However, one drawback of this method is that the abnormal data processing model has a weak distinction between normal data and abnormal data, which means that it is easy to cause confusion between them. In this article, the idea of partial data reconstruction is introduced to solve this problem. Besides, due to the lack of prior knowledge of the unsupervised method, it is difficult to discriminate all abnormal data. According to the characteristics of wind turbine operational data, the relationship between some low dimensional variables is known, such as the wind speed-power curve. Based on this, we can consider some known low dimensional variable relationships as prior knowledge to further improve the performance of abnormal data processing.

In this article, we propose an unsupervised abnormal data identification algorithm to obtain high-quality modeling data and study the effect of those data for fault prediction performance through relevant cases. The main innovations of this article are as follows: (a) An unsupervised sequential-ensemble algorithm for wind turbine abnormal data processing is proposed, which considers both the high dimensional characteristics of operational data and the known low dimensional variable relationships. (b) In the process of abnormal data identification considering the high dimensional characteristics of the operational data, the method of anomaly scoring based on the reconstruction errors is introduced as a discriminate criterion for abnormal data. The improved stacked autoencoder (I-SAE) based on the idea of partial data reconstruction is proposed to learn the complex high dimensional characteristics from the operational data and enhance the separability between normal data and abnormal data. In addition, a multi-model fusion strategy using voting criteria is introduced to enhance the robustness of the proposed unsupervised model. (c) Different algorithms are used for different types of abnormal data in the process of identification while considering the relationship of low dimensional variables. The improved density-based spatial clustering of applications with noise (DBSCAN) clustering algorithm based on the idea of the density ratio is proposed to handle sparse abnormal data and solve the problem of poor clustering performance caused by data imbalance. The Otsu method based on maximizing the variance between classes is used to eliminate stacked abnormal data. (d) Through the model comparison based on different abnormal data processing methods, we study the influence of abnormal data processing methods on the performance of the fault warning model.

The organizational structure of this article is as follows: (a) The overall framework of the wind turbine fault warning model based on abnormal data processing in this study is provided, and the algorithms involved in the proposed model are explained in Section 2; (b) the experimental results of abnormal data processing based on the proposed model and fault prediction are presented in Section 3; (c) relevant conclusions and future work are drawn in Section 4.

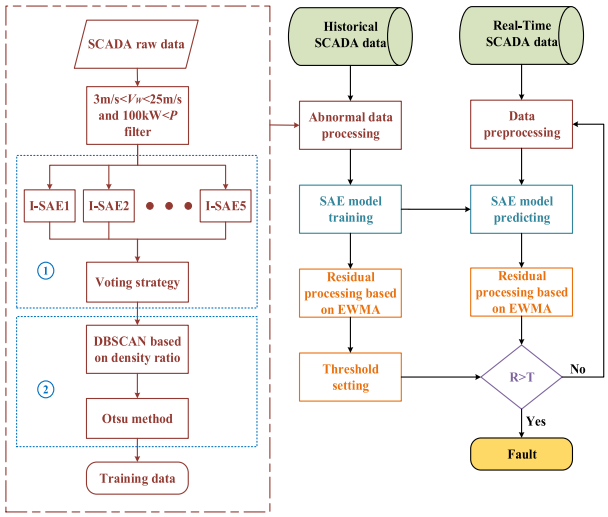


FIGURE 2. The overall framework of the wind turbine fault warning model proposed in this article.

II. OVERALL FRAMEWORK AND RELATED METHODOLOGY OF THE WIND TURBINE FAULT WARNING MODEL

In this article, an unsupervised ensemble framework for the abnormal operational data identification of wind turbines is proposed, and the influence of modeling data based on abnormal data processing methods on fault prediction performance is studied. Fig. 2 presents the overall framework of the fault warning model proposed in this article. This framework consists of two parts: the offline modeling process and the online monitoring process.

- **offline modeling process.**
  1) **acquisition of modeling data**

     To improve the quality of the modeling data, this article focuses on developing an abnormal data processing model based on an unsupervised sequential-ensemble algorithm according to the characteristics of wind turbine operational data. The abnormal data processing model comprehensively takes into account the high dimensional characteristics of operational data and prior knowledge based on the known low dimensional variable relationships.

     The wind turbine is in operation between cut-in wind speed and cut-out wind speed. The data outside this range should be filtered out. The cut-in wind speed and cut-out wind speed of the wind turbine used in this study are 3 m/s and 25 m/s respectively, which is determined by the design of the wind turbine. Besides, due to the frequent startup and shutdown processes of wind turbines in low wind speed ranges, the data in the low wind speed range are different from the normal operation data. Therefore, to obtain high-quality normal operation data, the data in the low wind speed range should be filtered out. To solve this problem, a lower limit of output power can be set to filter the data. According to relevant references [23], [30] and operation experience, the appropriate power value - 100kW is selected as the lower limit of output power for 1.5MW wind turbines. Consequently, in the framework of abnormal data processing, as shown in Fig. 2, we filter some data based on this criterion: 3 m/s $< V_W <$ 25 m/s and 100 kW $< P$.

     As shown in the first part of the abnormal data processing in Fig. 2, the I-SAE with enhanced separability is proposed to learn the high dimensional characteristics between the operational data, which can identify abnormal data by modeling the relationship between high dimensional data. To avoid the difficulty of setting discriminant parameters and improve the robustness of the model, the model with different discriminant parameters is used to process the operational data, and the processing results are fused by voting criteria.

     As shown in the second part of the abnormal data processing in Fig. 2, the quality of the modeling data is further improved based on the prior knowledge of the known low dimensional relationships. We integrated the density ratio into the DBSCAN clustering algorithm to identify the sparse abnormal data and avoid the erroneous identification of the original DBSCAN algorithm caused by the wind speed data imbalance. Faced with stacked abnormal data, the simple and effective Otsu algorithm based on maximizing the variance between classes is introduced. Finally, the processed data are used as training samples of the normal behavioral model.
  2) **modeling process**

     The selection of the normal behavioral model is of great significance to the performance of fault prediction. As mentioned before, SAE, as a deep learning model, has strong data-mining abilities. In the proposed framework of fault warning, the general SAE is used to capture the complex nonlinear relationship between SCADA data. In the training process of the model, the SCADA data processed by the proposed abnormal data processing model is used as the input of the SAE model for reconstruction training.
  3) **threshold setting**

     After the training of the SAE model, the reconstruction error calculated by input and reconstruction output is taken as a monitoring index. Secondly, to prevent the influence of noise in SCADA data, a common residual smoothing algorithm, exponentially weighted moving average (EWMA), is used to deal with reconstruction error. In the last, the threshold is calculated based on the mean and variance of the reconstructed error.

- **online monitoring process.**

  After the offline modeling process is completed, online condition monitoring can be carried out. In the real-time monitoring process, SCADA data are transmitted to the SAE model trained in the offline process. Similarly, the reconstruction error of real-time SCADA data is calculated and processed by EWMA smoothing algorithm. Comparing the reconstruction error with the threshold value set in the offline modeling process, the fault alarm information is sent out when the reconstruction error exceeds the threshold.

A. ABNORMAL DATA PROCESSING CONSIDERING HIGH DIMENSIONAL CHARACTERISTICS

1) STANDARD AUTOENCODER (AE)

The standard autoencoder (AE) is a three-layer feedforward neural network with a symmetrical structure, which means that its input neurons have the same number as its output neurons. As shown in Fig. 3, the general structure of the standard AE is composed of two parts: the encoder and the decoder. In the process of AE training, tied weights, that is, transposing the relationship between the weight of the encoder (W) and the decoder (W^T) is a common technique to reduce the training parameters by half, accelerate the training, and reduce the risk of over-fitting. Using nonlinear mapping, the encoder maps the data onto a latent representation to learn the high dimensional characteristics of the operational data. Given the input $X = [X_1, X_2, \cdots, X_N]$, the hidden layer feature h extracted by encoding process is shown in (1), where

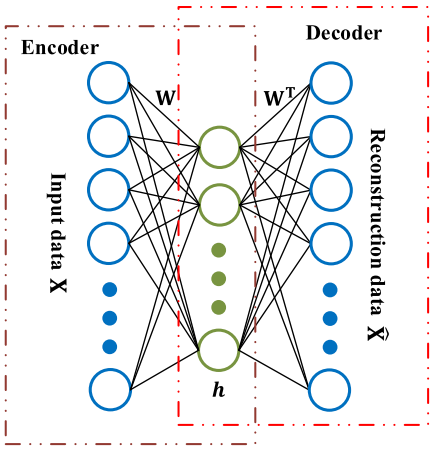


FIGURE 3. The general structure of standard AE.

N is the number of training samples:

$$h = f(WX + b) \tag{1}$$

where b represents the bias vectors of the encoding processes, and h represents the hidden layer feature extracted by AE neural network.

Conversely, the decoder remaps the hidden layer representation to the original data to achieve the purpose of unsupervised feature extraction. Equation (2) represents the reconstruction result of the input data:

$$\hat{X} = g(W^T h + c) \tag{2}$$

where c represents the bias vectors of the decoding processes, and $g(\cdot)$ and $f(\cdot)$ are the activation functions of the corresponding neurons. $\hat{X}$ is the reconstructed value of the input sample.

AE can be regarded as a non-linear feature extraction model based on unsupervised learning, which means that there is no need for label participation in the training process. A combination of encoder and decoder is the training method used for the unsupervised AE model. Minimizing the loss function via the back-propagation (BP) algorithm can be used to train the AE network. The loss function of the SAE neural network is shown in (3):

$$\arg\min_{W,b,c} loss = \frac{1}{N}\sum_{i=1}^{N} \| X_i - \hat{X}_i \|^2 \tag{3}$$

where X_i and $\hat{X}_i$ represent the original training data and reconstruction values, respectively. And N is the number of training samples.

2) STACKED AUTOENCODER MODEL (SAE)

With the breakthrough of neural network technology, deep neural networks have been widely used in many fields because of their stronger feature extraction abilities. The structure of SAE formed by stacking multiple standard AEs is shown in Fig. 4. This structure can learn the deeper high dimensional characteristics of operational data and extract the finer features. Unlike standard AE, it is difficult to train SAE directly using the BP algorithm due to the problem of the gradient vanishing because of the deep structure. To this end, using layer-by-layer pretraining to obtain better parameter initialization can effectively solve the problem of gradient vanishing. Based on this idea, the training process of SAE is generally divided into two stages-the pretraining process and the fine-tuning process. In the pretraining process, multiple standard AEs of different structures are trained in turn and spliced into the structure, as shown in Fig. 4, where the output of the previous AE hidden layer is taken as the input of the next AE. After pretraining, the whole SAE is fine-tuned using the BP algorithm.

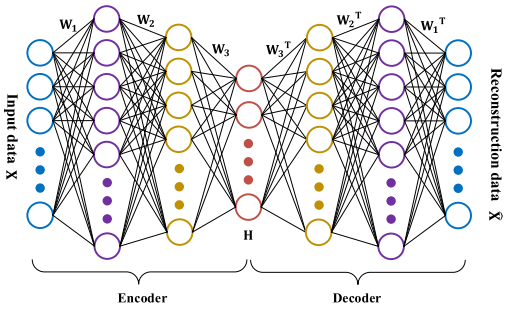


FIGURE 4. The general structure of SAE.

3) I-SAE MODEL WITH ENHANCED SEPARABILITY

Due to the presence of more normal data than abnormal data, for SAE, the reconstruction error of normal operational data is less than that of abnormal data, which is the principle of SAE for abnormal data processing. To enhance SAE's abnormal data identification abilities, according to research in [38], the idea of partial data reconstruction based on the discrimination process is introduced to improve the separability of SAE for normal and abnormal wind turbine operational data. Therefore, combined with the characteristics of operational data, a discrimination process based on the dynamic threshold is added before batch training to select some samples as training data, which has a low reconstruction error. In the training process of random gradient descent, given a small batch sample $X = [X_1, X_2, \cdots, X_M]$, the dynamic threshold is set in (4):

$$threshold = \frac{1}{M}\sum_{i=1}^{M} \| X_i - \hat{X}_i \|^2 * \alpha \tag{4}$$

where α is defined as a discriminant parameter and M is the number of small batches training samples.

Before batch training, a discrimination process is needed to further screen the training data. First, the reconstruction error of each sample in the batch is calculated by forwarding propagation, and then the samples with reconstruction

errors greater than the dynamic threshold are used as training samples.

Besides, to reduce the influence of manually setting the discriminant parameter α, an ensemble of different parameter setting models based on voting criteria is proposed, which increases the robustness and reliability of the unsupervised algorithm. As shown in Algorithm 1, the implementation of abnormal data processing considering high dimensional characteristics is as follows.

Algorithm 1 I-SAE With Enhanced Separability

Input: Filtered (based on 3 m/s $< V_W <$ 25 m/s and 100 kW $< P$) and standardized SCADA data: $X = [X_1, X_2, \cdots, X_N]$, N is the number of the samples; Discriminant parameter sets $\upsilon = \{1.2, 1.25, 1.3, 1.35, 1.4\}$; Learning rate μ; Batch size η; Pre-training epochs ξ_p; Fine-tuning epochs ξ_f; Structure of the model and the number of layer β.

Output: Abnormal data processing results considering the high dimensional characteristics

for $m = 1$ to 5 **do**
  set $\alpha = \upsilon[m]$.
  Model initialization: Initialization of weights and bias.
  Model pre-training:
    for $i = 1$ to β **do**
      for $j = 1$ to ξ_p **do**
        for $k = 1$ to N/η **do**
          Training the standard autoencoder of each hidden layer in turn.
        end for
      end for
    end for
  Model fine-tuning:
    for $i = 1$ to ξ_f **do**
      for$j = 1$ to N/η **do**
        Calculate the average value (E_{mean}) of the reconstruction error of the batch.
        Choose the samples whose reconstruction error is less than $E_{mean} * \alpha$ as training samples.
        Training model based on gradient descent method.
      end for
    end for
  All samples are predicted by the trained model. And calculate the reconstruction error (E_{X_i}) of each sample and the mean value (E_{mean}) of reconstruction error of all samples.
  If $E_{X_i} > E_{mean} * \alpha$, X_i is defined as abnormal data in model m.
end for
Voting criteria:
  If a sample appears two times in the abnormal results of five models, the sample is marked as abnormal data.

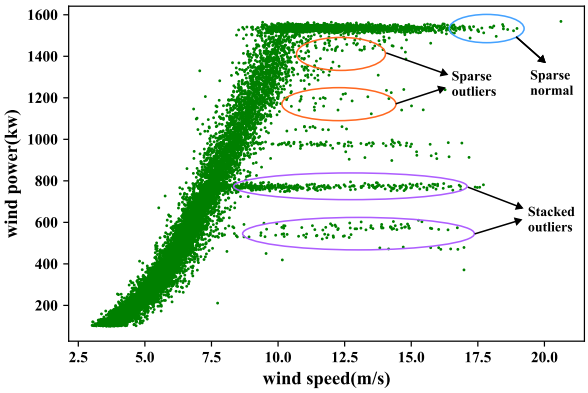


FIGURE 5. The distribution of abnormal data in the $V_W - P$ scatters.

B. ABNORMAL DATA PROCESSING CONSIDERING THE RELATIONSHIP OF LOW DIMENSIONAL VARIABLES

Although SAE can learn the high dimensional characteristics of operational data and identify abnormal data, it remains difficult to identify all abnormal data accurately due to a lack of prior knowledge. To further improve the performance of abnormal data processing, we eliminate abnormal data based on known low dimensional variable relationships. Generally, there are two kinds of common abnormal data types in low dimensional data relationships: sparse abnormal data and stacked abnormal data. As shown in Fig. 5, sparse abnormal data are randomly distributed around the normal data, which are usually caused by sensor failures, communication noise, and some uncontrollable factors. Because of the distribution characteristics of sparse abnormal data, it is suitable to use the density clustering method for processing. Stacked abnormal data are usually caused by wind turbine failure, wind curtailment commands, and communication failures [39]. It is difficult to handle stacked abnormal data by using the density clustering method. Therefore, in this study, the Otus algorithm is considered to distinguish the normal cluster from the abnormal cluster. In addition, according to expert knowledge, the gearbox oil temperature and gearbox shaft temperature are limited to 75°C and 80°C, respectively [12].

1) DBSCAN BASED ON DENSITY RATIO

DBSCAN is a common clustering algorithm based on density, which can distinguish the sparse data from the dense data [40]. For the traditional DBSCAN algorithm, the clustering performance is better when the density of the normal data and abnormal data are significantly different. However, the operational data of wind turbines show unbalanced characteristics because the data for high wind speeds are far less common than those for low wind speeds. As shown in Fig. 5, the marked sparse normal data have the same density as the sparse abnormal data, which will cause the sparse normal data to be deleted by mistake using the traditional DBSCAN algorithm. To eliminate the impact of the wind speed imbalance

on reducing the performance of the clustering algorithm, inspired by the research in [41], the algorithm based on the density ratio is proposed to handle sparse abnormal data. To describe the algorithm based on the density ratio, the following definitions are given. Consider the following sample data set: $D = [X_1, X_2, \cdots, X_N]$.

Definition 1 ($\epsilon-$Neighborhood): The $\epsilon-$neighborhood of data point X_i is defined as $N_\epsilon(X_i) = \{X_j \in D \mid distance(X_i, X_j) \leq \epsilon\}$, where the $distance(\cdot)$ is usually European distance.

Definition 2 (Density Ratio): Define the ϵ_1-neighborhood of data point X_i on the wind speed attribute as $N_{\epsilon_1}(X_i\_V_W) = \{X_j \in D \mid abs(X_i\_V_W, X_j\_V_W) \leq \epsilon_1\}$, where $abs(\cdot)$ is the absolute value function, and $X_i\_V_W$ is the projection of X_i on the wind speed attribute. The density ratio of data point X_i is defined as

$$\rho_{X_i} = \frac{N_\epsilon(X_i)}{N_{\epsilon_1}(X_i\_V_W)}. \tag{5}$$

Definition 3 (Core Object): If the ρ_{X_i} of data point X_i is greater than ρ, then X_i is called the core object.

Definition 4 (Directly Density-Reachable): When X_j is in the $\epsilon-$neighborhood of X_i and X_i is the core object, then the data point X_j is directly density-reachable from the core object X_i.

Definition 5 (Density-Reachable): If there is a sample dataset $\{P_1, P_2, \cdots, P_T\}$ that satisfies $P_1 = X_i$, $P_T = X_j$, and P_{t+1} is directly density-reachable from data point P_t, then data point X_j is density-reachable from data point X_i.

Definition 6 (Density-Connected): If there is a core object X_k that satisfies both X_i and X_j are density-reachable from data point X_k, then X_i and X_j density are density-connected.

As shown in the above definition, compared with the traditional DBSCAN algorithm, the improved DBSCAN algorithm defines the core object by its density ratio. In the definition of the density ratio shown in (5), the performance of abnormal data processing is improved by considering the unbalanced characteristics of wind speed. The clustering process of the DBSCAN algorithm based on the density ratio is the same as that of the traditional DBSCAN algorithm, which is used to find the sample set with the maximum density-connected value. The detailed process of the DBSCAN algorithm based on the density ratio is shown in Algorithm 2.

2) OTSU ALGORITHM BASED ON WIND SPEED PARTITIONS

The Otsu algorithm is a common image threshold segmentation technology based on variance [42]. As shown in (6), the optimal threshold X_k can be found to segment two types of data with different distributions by maximizing σ:

$$\sigma = (\frac{1}{k}\sum_{i=1}^{k} X_i - \frac{1}{N-k}\sum_{i=k+1}^{N} X_i)^2 \tag{6}$$

where σ represents the variance between classes and k is the dividing point of different classes.

Algorithm 2 Abnormal Data Processing Considering the Relationship of Low Dimensional Variables

Input: SCADA data after processing based on I-SAE with enhanced separability
Output: SCADA data processed by abnormal data

DBSCAN algorithm based on density ratio:
  Definition: Let the data class $id = \{-1, 0, 1, 2, 3, \cdots\}$, where the -1 represents unclassified, 0 represents the noise class, and $\{1, 2, 3, \cdots\}$ represents different class of data clusters. The dataset of the corresponding class is $\{class(-1), class(0), class(1), \cdots\}$.
  Initialization: set ϵ, ρ, $id = 1$, $class(id) = \phi$, adding all data to $class(-1)$. Here ϕ is empty set.
  while $class(-1) \neq \phi$ **do**
    Select a datum X_i as the current datum to be processed in $class(-1)$.
    while true **do**
      if X_i is the core point **then**
        Remove X_i from $class(-1)$ and add it to $class(id)$.
        Add all points in $\epsilon-$ neighborhood of X_i to auxiliary dataset C.
      else
        Remove X_i from $class(-1)$ and add it to $class(0)$.
      end if
      if $class(-1) = \phi$ or $C = \phi$ **then**
        break
      else
        Select a datum X_i as the current datum to be processed in C.
      end if
    end while
    $id = id + 1$, $class(id) = \phi$.
  end while
  return the class with the largest number of datasets.
Otsu algorithm based on wind speed partition (Take $V_W - P$ as an example):
  for i=3:18:0.6 **do**
    Get data of $i < V_W < i + 0.6$ and make $P = sorted(data[windpower])$.
    through $\max \sigma = (\frac{1}{k}\sum_{i=1}^{k} P_i - \frac{1}{N-k}\sum_{i=k+1}^{N} P_i)^2$, find P_K.
    if $abs(\frac{1}{K}\sum_{i=1}^{K} P_i - \frac{1}{N-K}\sum_{i=K+1}^{N} P_i) > 300$kW **then**
      Mark P_i as abnormal data, where P_i is less than P_K.
    end if
  end for

As shown in Fig. 5, taking the 9-9.6m/s interval of the $V_W - P$ scatters as an example, it can be seen that the normal data and the stacked abnormal data have different distributions.

Therefore, the method of maximizing the variance between classes can effectively distinguish the stacked abnormal data from the normal data. The process of stacked abnormal data processing based on maximizing the variance between classes is shown in Algorithm 2.

C. ESTABLISHMENT OF THE WIND TURBINE FAULT WARNING MODEL AND THRESHOLD SETTING

As mentioned before, the stacked autoencoder is used as the fault warning model, with an autoencoder in each layer and the value of the hidden layer as input to the next layer to learn the inherent characteristics of the normal operational data. It should be noted that in the process of establishing the normal behavioral model, general SAE is used instead of I-SAE with enhanced separability because the original SCADA data were processed with abnormal data.

1) SMOOTHING PROCESS OF THE RECONSTRUCTION ERRORS

Due to the noise influence of SCADA data and the uncertain factors in the modeling process, the prediction residual often has a large peak value, which may cause a false alarm. Therefore, a common data smoothing algorithm, EWMA, is adopted in this article. Equation (7) is the statistical variable Z_t of the EWMA algorithm:

$$Z_t = \lambda E_t + (1-\lambda) Z_{t-1} \tag{7}$$

$$E_t = \| X_t - \hat{X}_t \|^2 . \tag{8}$$

where λ is the smoothing coefficient. E_t represents the original reconstruction error at time t. X_i and $\hat{X}_i$ represent the original training data and reconstruction values, respectively.

2) THRESHOLD SETTING PROCESS

Equation (9) and (10) provide the mean value and variance expression of the statistical variable Z_t and the setting of the corresponding control line, respectively. In the process of real-time fault early warning, when the statistical variable Z_t is greater than the control line $UCL(t)$, the system is currently in an abnormal operational status.

$$\mu_{Z_t} = \mu_{E_t}, \sigma_{Z_t} = \frac{\sigma_{E_t}^2}{n} \frac{\lambda}{2-\lambda} [1-(1-\lambda)^{2t}] \tag{9}$$

$$UCL(t) = \mu_{Z_t} + L\sigma_{Z_t} = \mu_{E_t} + L\sigma_{E_t} \sqrt{\frac{\lambda[1-(1-\lambda)^{2t}]}{(2-\lambda)n}} \tag{10}$$

where $\mu_{(\cdot)}$ and $\sigma_{(\cdot)}$ represent the mean and variance of the corresponding variables, respectively. L is the control line parameter, and n is the number of samples.

III. CASE STUDY AND CONTRAST ANALYSIS

A. DATA DESCRIPTION

To verify the effectiveness of the proposed abnormal data processing model and fault warning model, the SCADA data at 10 min intervals from 1 February 2014 to 12 June 2014 were collected from a wind farm in the Shandong Province of China. This wind farm contains 33 wind turbines of the same type with a rated capacity of 1.5 MW, all of which are equipped with a SCADA system. Generally, the SCADA system is equipped with many sensors (such as temperature, power, current, and wind speed), which are distributed in subsystems to monitor the operational status of the wind turbines. Therefore, SCADA data provide massive information on the health status of wind turbines. Table 1 shows the 15 variables used in this study and their corresponding descriptions.

TABLE 1. Description of the adopted variables related to the operational status of the wind turbine.

Parameter	Variable	Units
Wind speed	V_W	m/s
Wind power	P	kW
Generator speed	V_G	m/s
Rotor speed	V_R	m/s
U_1 current	I_1	A
U_2 current	I_2	A
U_3 current	I_3	A
Ambient temperature	T_A	°C
Nacelle temperature	T_N	°C
Gearbox oil temperature	T_O	°C
Gearbox shaft 1 temperature	T_{G1}	°C
Gearbox shaft 2 temperature	T_{G2}	°C
Gearbox bearing A temperature	T_{GA}	°C
Gearbox bearing B temperature	T_{GB}	°C
U_1 winding temperature	T_W	°C

B. ABNORMAL DATA PROCESSING

1) PRELIMINARY DATA PROCESSING

As mentioned before, to avoid the influence of the random characteristics of low wind speed data on the modeling, the SCADA data are preprocessed based on this criterion: 3 m/s $< V_W <$ 25 m/s and 100 kW $< P$. The relationship between wind speed and wind power after preprocessing of the three wind turbines is shown in Fig. 6. It can be seen from the wind speed and wind power relationship diagram that there are many abnormal data among the wind turbine's operational data, which will distort the normal relationships between the SCADA variables in the established model. For this reason, we developed an unsupervised ensemble algorithm to handle abnormal data.

2) ABNORMAL DATA PROCESSING CONSIDERING HIGH DIMENSIONAL CHARACTERISTICS

Due to the coupling between the subsystems of the wind turbine, the SCADA data essentially have high dimensional characteristics. Therefore, the relationships between the SCADA variables should be considered when modeling to analyze abnormal data. In this article, an I-SAE with enhanced separability is proposed to model SCADA data.

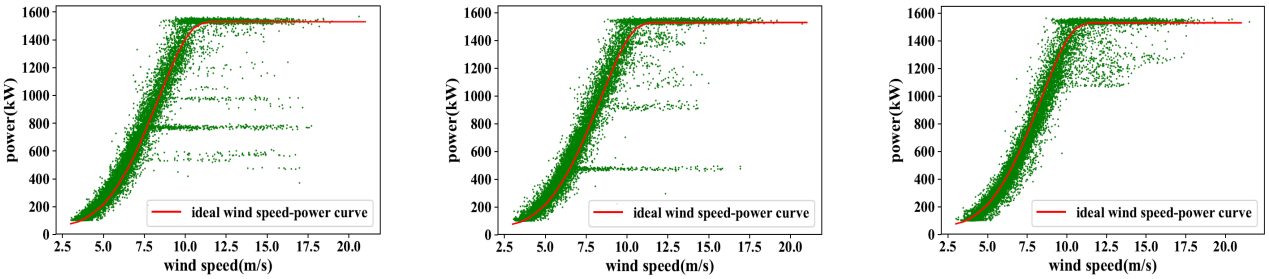


FIGURE 6. The pretreatment results based on 3 m/s < V_W < 25 m/s and 100 kW < P from (a) wind turbine #1, (b) wind turbine #2, and (c) wind turbine #3.

TABLE 2. Hyperparameters settings related to the proposed model.

Model	Parameter	Values
The model of abnormal data processing	Learning rate μ	0.01
	Batch size η	1000
	Pre-training epochs ξ_p, Fine-tuning epochs ξ_f	5, 1000
	The structure of I-SAE	15-100-50-25-50-100-15
	Discriminant parameter sets υ	$\{1.2, 1.25, 1.3, 1.35, 1.4\}$
	Voting criteria parameters	2
	$\epsilon-$neighborhood and density ratio of $V_W - P$	0.02, 0.04
	$\epsilon-$neighborhood and density ratio of $V_W - V_G$	0.02, 0.03
	$\epsilon-$neighborhood and density of $V_R - P$	0.01, 20
	Number of partitions used by the Otsu algorithm	20
The model of fault prediction	Learning rate μ	0.01
	Batch size η	1000
	Pre-training epochs ξ_p, Fine-tuning epochs ξ_f	5, 1000
	The structure of SAE	15-100-50-25-50-100-15
	Smoothing parameters of the EWMA algorithm λ	0.1
	Control line parameters L	4

The relevant parameters of the model are shown in Table 2. Fig. 7 shows the abnormal data processing results of wind turbine #1 based on the I-SAE model with enhanced separability. In addition, to demonstrate the superiority of the proposed model, SAE without enhanced separability is used as the comparison model in this study, and the results are shown in Fig. 8.

Based on the results of Fig. 7, the I-SAE model with enhanced separability can identify the vast majority of abnormal data. For example, in Fig. 7(a)-(c), the sparse and stacked abnormal data far from the normal cluster are identified. In Fig. 7(d)-(e), abnormal high-temperature data and low-temperature data are also identified. It can be seen from Fig. 8 that many abnormal data in the SAE model without enhanced separability were not identified, such as the marked areas in Fig. 8(a)-(e). From a comparison between Fig. 7 and Fig. 8, the proposed I-SAE with enhanced separability shows better performance in wind turbine abnormal data processing. This is because the selection process of training data is added to the training process of the SAE neural network in the I-SAE model. By constantly updating and selecting samples with small reconstruction error as training data, the I-SAE model is more in line with the normal operational data characteristics. Thus, the performance of abnormal data processing has been greatly improved. However, although this method can effectively identify a large number of abnormal data, a few abnormal data remained unidentified, such as the marked areas in Fig. 7(a)-(c). Therefore, this study considers the relationship of the known variables as prior knowledge to further improve the performance of abnormal data processing.

3) ABNORMAL DATA PROCESSING CONSIDERING THE RELATIONSHIP OF LOW DIMENSIONAL VARIABLES

For sparse abnormal data, the improved DBSCAN algorithm based on the density ratio is proposed to solve the problem of wind speed imbalance. According to prior knowledge, $V_W - P$, $V_W - V_G$, and $V_R - P$ are considered as two-dimensional variable research objects. Among them, $V_W - P$ and $V_W - V_G$ use the improved DBSCAN algorithm based on the density ratio proposed in this article, and $V_R - P$ uses the standard DBSCAN algorithm based on density to process. The model-related parameter settings are shown in Table 2.

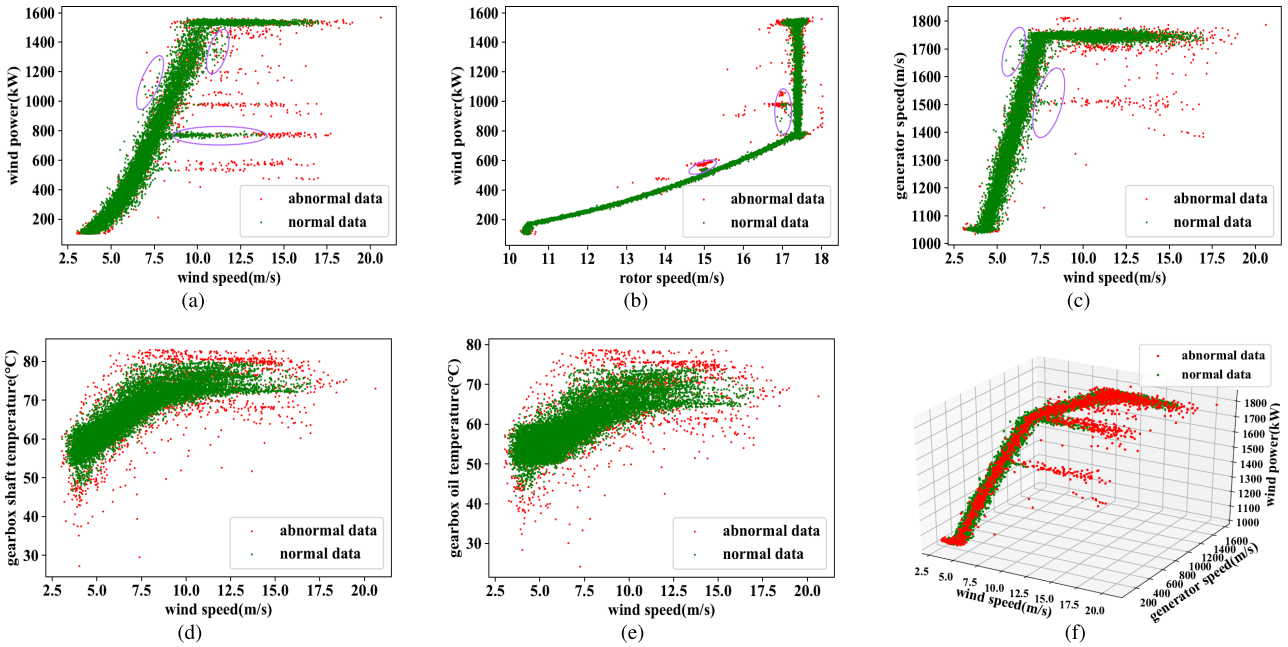


FIGURE 7. The result of abnormal data processing based on I-SAE with enhanced separability for wind turbine #1 (a) $V_W - P$, (b) $V_R - P$, (c) $V_W - V_G$, (d) $V_W - T_{G1}$, (e) $V_W - T_O$, and (f) $V_W - V_G - P$.

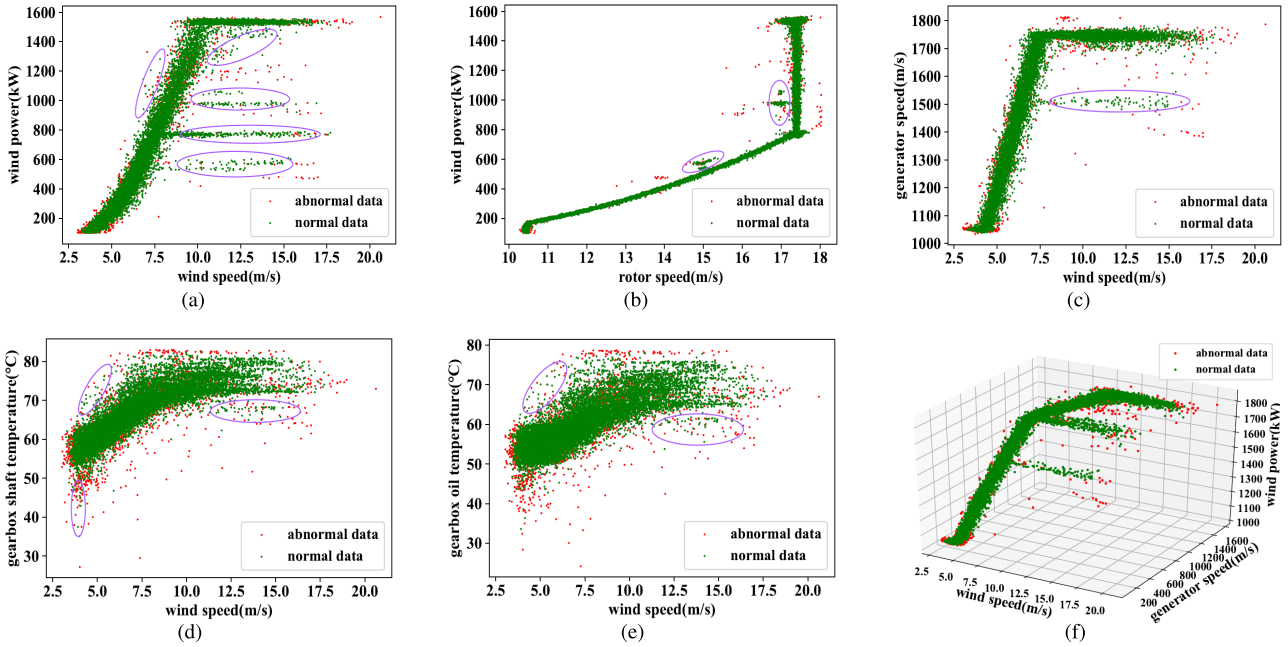


FIGURE 8. The result of abnormal data processing based on SAE with without enhanced separability for wind turbine #1 (a) $V_W - P$, (b) $V_R - P$, (c) $V_W - V_G$, (d) $V_W - T_{G1}$, (e) $V_W - T_O$, and (f) $V_W - V_G - P$.

Fig. 9 shows the results of DBSCAN based on density and those of DBSCAN based on the density ratio of the original data and the data processed by considering the characteristics of the high dimensional data. As can be seen from Fig. 9, the DBSCAN algorithm based on density causes data misjudgment. Some normal data are identified as abnormal data due to the imbalanced characteristics of wind speed, such as the marked areas in Fig. 9(b) and Fig. 9(d). Conversely, it can be seen from Fig. 9(a) and Fig. 9(c) that the method based on the density ratio can effectively solve this problem. In the improved DBSCAN algorithm, the concept of density ratio is introduced to solve the problem of wind speed imbalance. The density ratio based on wind speed data can make the difference between the density ratio of the normal sparse data

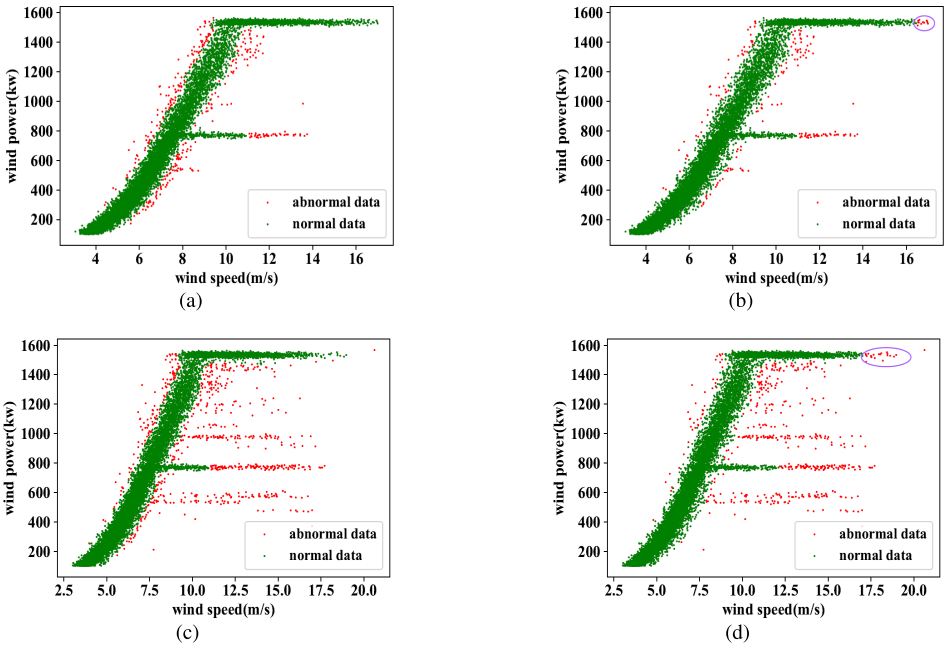


FIGURE 9. The processing results for wind turbine #1 (a) DBSCAN based on density ratio from the data processed by considering the characteristics of high-dimensional data, (b) DBSCAN based on density from the data processed by considering the characteristics of high-dimensional data, (c) DBSCAN based on density ratio from the raw data, and (d) DBSCAN based on density from the raw data.

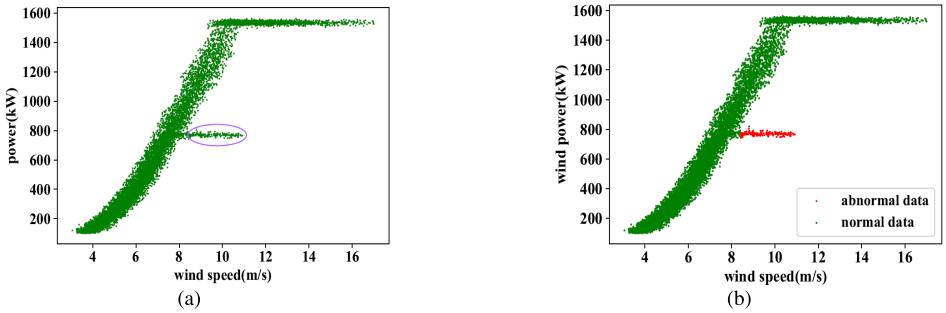


FIGURE 10. (a) The $V_W - P$ diagrams waiting to handle stacked abnormal data. (b) the result processed by OTSU algorithm for wind turbine #1.

part and normal dense data part is small, and the density ratio of sparse abnormal data has an obvious difference. Therefore, the DBSCAN algorithm based on density ratio does not cause misjudgment. On the contrary, due to the influence of wind speed imbalance, the density of the normal sparse data part is similar to that of the abnormal sparse data part, which leads to the phenomenon of misjudgment when using the traditional DBSCAN algorithm.

For stacked abnormal data (such as the marked area in Fig. 10(a)), the Otsu algorithm based on maximizing the variance between classes is proposed. As can be seen from Fig. 10(b), the Otsu algorithm can accurately distinguish the stacked abnormal data from the normal data. Fig. 11 shows the final results of the unsupervised ensemble algorithm proposed in this article.

In addition, to show the superiority of the unsupervised ensemble model proposed in this article, only the abnormal data processing considering the low dimensional relationship is shown in Fig. 12. As shown in the marked area in Fig. 12(d)-(e), there are a large number of isolated abnormal data that were not identified. The reason is that only a few ideal variable relation curves can be obtained, which makes it impossible to deal with abnormal data effectively for the most ideal variable relation curves that cannot be obtained. As is shown in Fig. 13, a dimensional reduction fusion clustering method based on reference [34] is used for comparison. From the result, only a small number of abnormal data is identified in this method. Comparing Fig. 7, Fig. 11, Fig. 12, and Fig. 13, it can be seen that the abnormal data processing algorithm proposed in this article has the best performance. In the comparison algorithm, the proposed abnormal data processing model based on serialization ensemble algorithm achieves the optimal performance, because it considers not only the high-dimensional characteristics of the operational

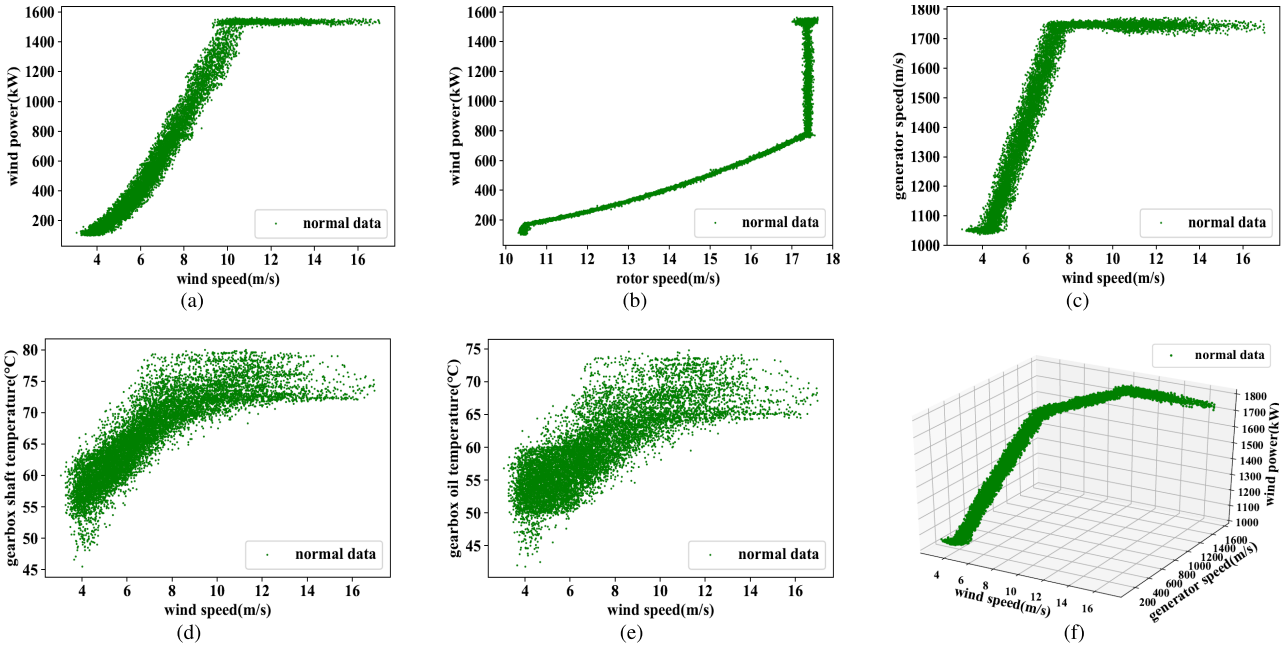


FIGURE 11. High dimensional data processing combined with low dimensional data processing for wind turbine #1 (a) $V_W - P$, (b) $V_R - P$, (c) $V_W - V_G$, (d) $V_W - T_{G1}$, (e) $V_W - T_O$, and (f) $V_W - V_G - P$.

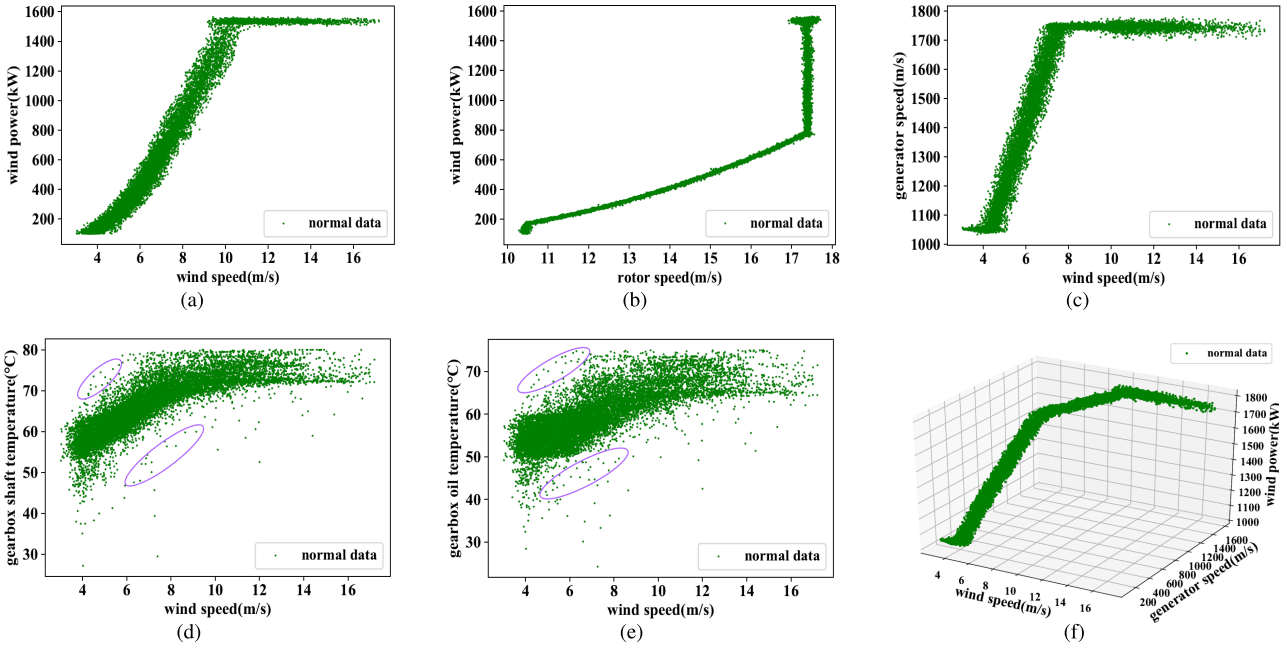


FIGURE 12. The abnormal data processing result only considering the low dimensional relationship for wind turbine #1 (a) $V_W - P$, (b) $V_R - P$, (c) $V_W - V_G$, (d) $V_W - T_{G1}$, (e) $V_W - T_O$, and (f) $V_W - V_G - P$.

data but also the known low dimensional variable relationships. The advantage of the unsupervised algorithm is that it does not need prior knowledge of the wind turbine which is difficult to obtain, and it can capture the high-dimensional characteristics of SCADA variables. In the unsupervised data processing algorithm which only considers the high dimensional characteristics of operational data, some abnormal data are not recognized due to the lack of prior knowledge. In the method only considering the low dimensional relationship, it is also difficult to obtain excellent performance because it does not consider the high dimensional characteristics of operational data. In the dimensional reduction fusion clustering method based on reference [34], the performance is also relatively poor. The separation between dimensional

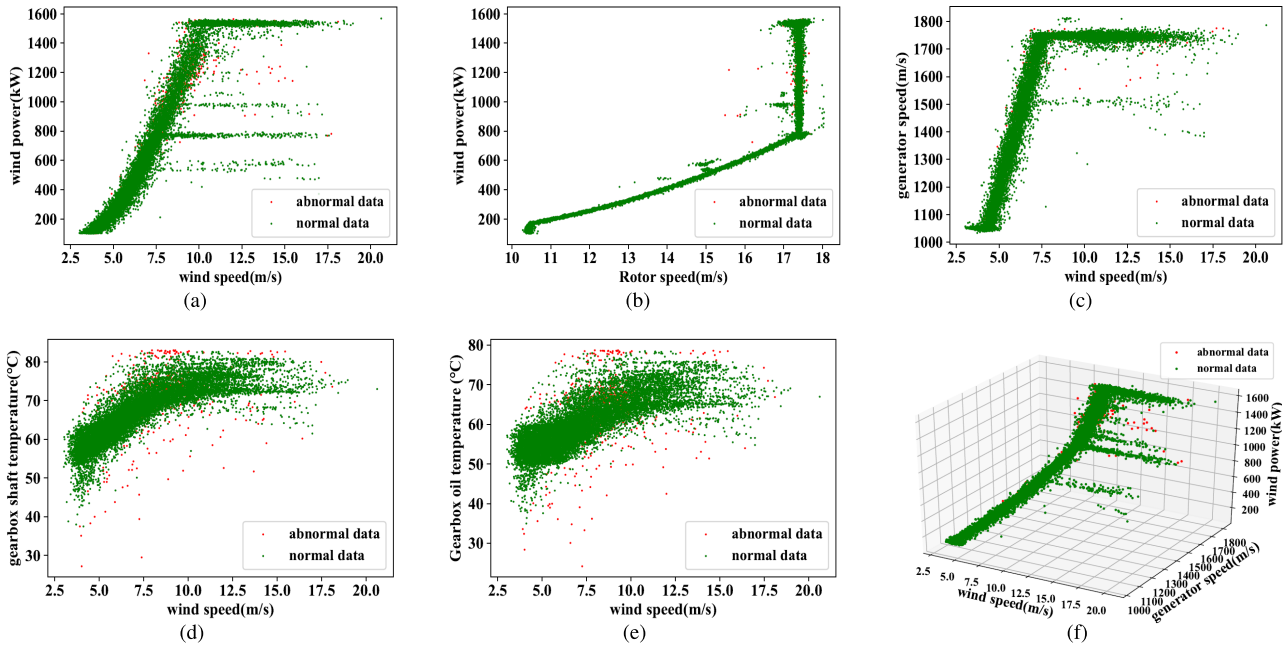


FIGURE 13. The abnormal data processing result based on the method in reference [34] for wind turbine #1 (a) $V_W - P$, (b) $V_R - P$, (c) $V_W - V_G$, (d) $V_W - T_{G1}$, (e) $V_W - T_O$, and (f) $V_W - V_G - P$.

TABLE 3. Description of the relevant fault information.

Wind turbines	Time	Alarm message
#1	2014/5/30 14 : 26	Converter fault and Inverter fault
#9	2014/5/20 16 : 17	Pitch system fault, Gearbox type fault, and Generator type fault
#12	2014/4/25 18 : 23	Gearbox type fault, Gearbox oil temperature over limit fault

reduction and clustering process makes it difficult to obtain excellent performance in wind turbine abnormal data processing.

C. FAILURE PREDICTION CASES

To study the impact of abnormal data processing on the performance of potential fault early warning, three fault cases were studied, the details of which are shown in Table 3. In this study of fault early warning, the impact of modeling data quality is analyzed based on the comparison of models with different abnormal data processing methods, including (a) the data processed by the proposed algorithm, (b) the data processed by the low dimensional abnormal data processing algorithm, (c) the data processed by the high dimensional abnormal data processing algorithm, and (d) the data processed by the criterion: 3 m/s $< V_W <$ 25 m/s and 100 kW $< P$. The corresponding methods are used to deal with 33 wind turbines respectively. All the operational data after processing are used as the training data, and then the operational data for a period of time before the fault alarm are used as the testing data. The number of modeling data based on the above four different abnormal data processing models is 349003, 394114, 371415, and 428542 respectively. The relevant parameters of the fault early warning model are shown in Table 2. The predicted results of the three faults are shown in Fig. 14-16. The time performance results of failure prediction cases are collected in Table 4, where the fault A-C is shown in Table 3, and methods A-D represent the abnormal data processing method proposed in this article, the low dimensional abnormal data processing algorithm, the high dimensional abnormal data processing algorithm, and the criterion of 3 m/s $< V_W <$ 25 m/s and 100 kW $< P$, respectively.

The monitoring process based on the four different abnormal data processing models before the converter fault and inverter fault alarm in wind turbine #1 is shown in Fig. 14. In this fault case, 712 data points before the fault alarm are used for research. To show the current operational state, some normal operational state data are used for comparison. In Fig. 14, the last 712 data points of the 1000 data points are the studied data, while the first 288 data points are the data of the normal operational state. In Fig. 14(a), the model based on the data processed by the method proposed in this article detected the fault 598 data points in advance (4 days, 3 hours, and 40 minutes). The normal operational state and abnormal operational state provided better discrimination. In Fig. 14(b),

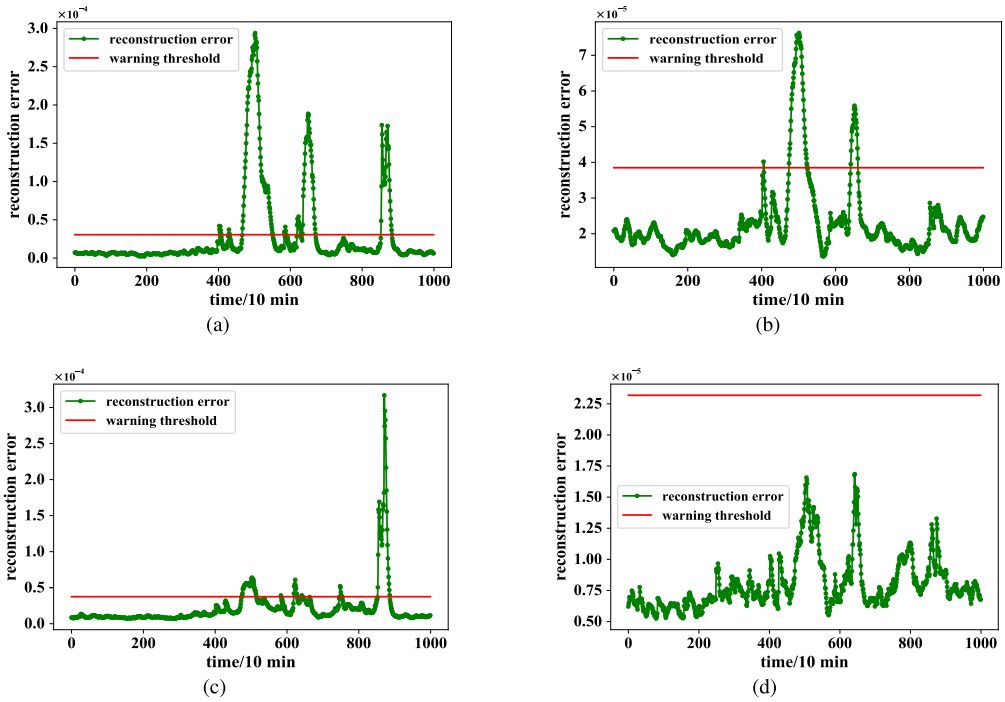


FIGURE 14. Fault warning results of #1 based on (a) the data processed by the proposed algorithm, (b) the data processed by the low dimensional abnormal data processing algorithm, (c) the data processed by the high dimensional abnormal data processing algorithm, and (d) the data processed by the criterion: $3\ m/s < V_W < 25\ m/s$ and $100\ kw < P$.

the model based on the data processed by the low dimensional abnormal data processing algorithm first detected the fault at the 405th data point, and then the fault was detected again at the 474th data point. However, the fault only lasted for one moment at the 405th data point, which could be mistakenly identified as a noise impact. In addition, the abnormal state around the 860th data point was not detected. This may be because similar abnormal data were mixed into the training data, thus making it difficult to distinguish between the normal state and abnormal state. In Fig. 14(c), the model based on the data processed by the high dimensional abnormal data processing algorithm detected the fault 525 data points in advance (3 days, 15 hours, and 30 minutes). This result is 12 hours and 10 minutes later than the method proposed in this article. As shown in Fig. 14(d), the model based on the data processed by the criterion (3 m/s $< V_W <$ 25 m/s and 100 kW $< P$) did not detect faults because a large number of abnormal data in the training data led to a large variance in the threshold setting making it difficult to distinguish between the normal and abnormal operational states.

The monitoring processes based on different abnormal data processing models before the pitch system fault, gearbox type fault, and generator type fault alarm in wind turbine #9 are shown in Fig. 15. Similarly, in Fig. 15, the last 474 data points of the 800 data points are taken as the studied data, while the first 326 data points are the data of normal operational state. In Fig. 15(a), the model based on the data processed by the method proposed in this article detected the fault 293 data points in advance (2 days and 50 minutes). The normal and abnormal operational states offered better discrimination. As shown in Fig. 15(b), the model based on the data processed by the low dimensional abnormal data processing algorithm first detected a fault at the 547th data point (1 day, 18 hours, and 10 minutes in advance). As shown in Fig. 15(c), the model based on the data processed by the high dimensional abnormal data processing algorithm detected the fault 293 data points in advance (2 days and 50 minutes). Although this model has the same time as the model proposed in this article in Fig. 15(a), it can be seen from the EWMA that the model proposed in this article identified more abnormal states and that the normal operational state and abnormal operational data had higher separability. In Fig. 15(d), the model based on the data processed by the criterion (3 m/s $< V_W <$ 25 m/s and 100 kW $< P$) also did not detect faults.

The monitoring processes based on different abnormal data processing models for gearbox type faults and gearbox oil temperature over-limit fault alarms in wind turbine #12 are shown in Fig. 16. In Fig. 16, the last 500 data points of the 800 data points are the studied data, and the first 300 data points are the data of the normal operational state. In Fig. 16(a), the model based on the data processed by the method proposed in this article detected the fault 383 data points in advance (2 days, 15 hours, and 50 minutes). As shown in Fig. 16(b), the model based on the data processed by the low dimensional abnormal data processing algorithm first detected a fault at the 418th data point (2 days, 15 hours,

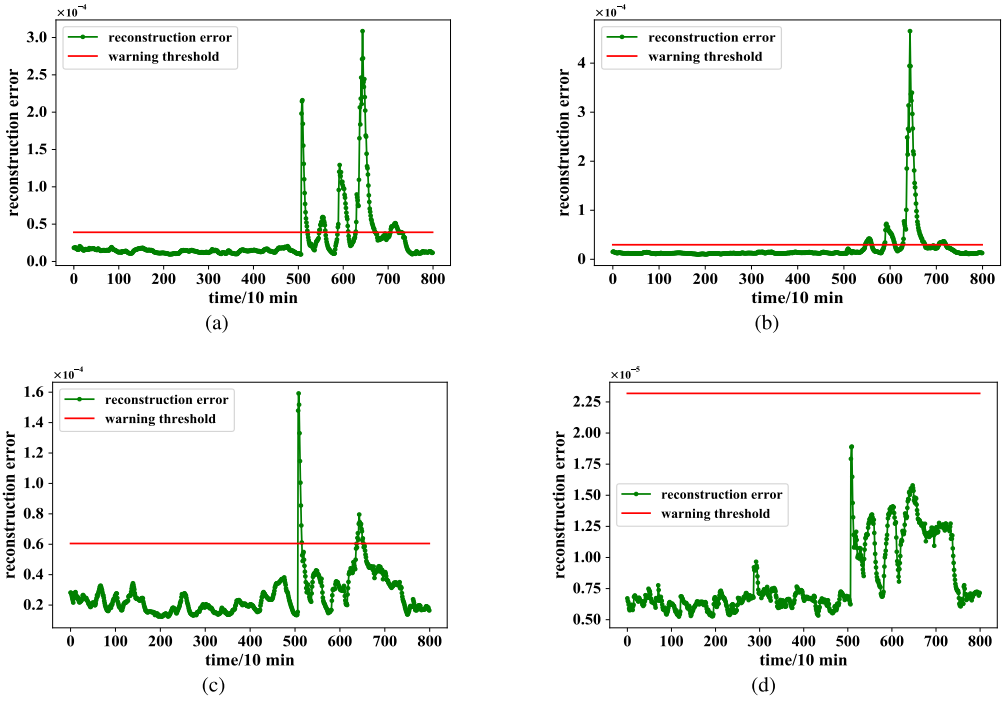


FIGURE 15. Fault warning results of #9 based on (a) the data processed by the proposed method, (b) the data processed by the low dimensional abnormal data processing algorithm, (c) the data processed by the high dimensional abnormal data processing algorithm, and (d) the data processed by the criterion: $3\ m/s < V_W < 25\ m/s$ and $100\ kw < P$.

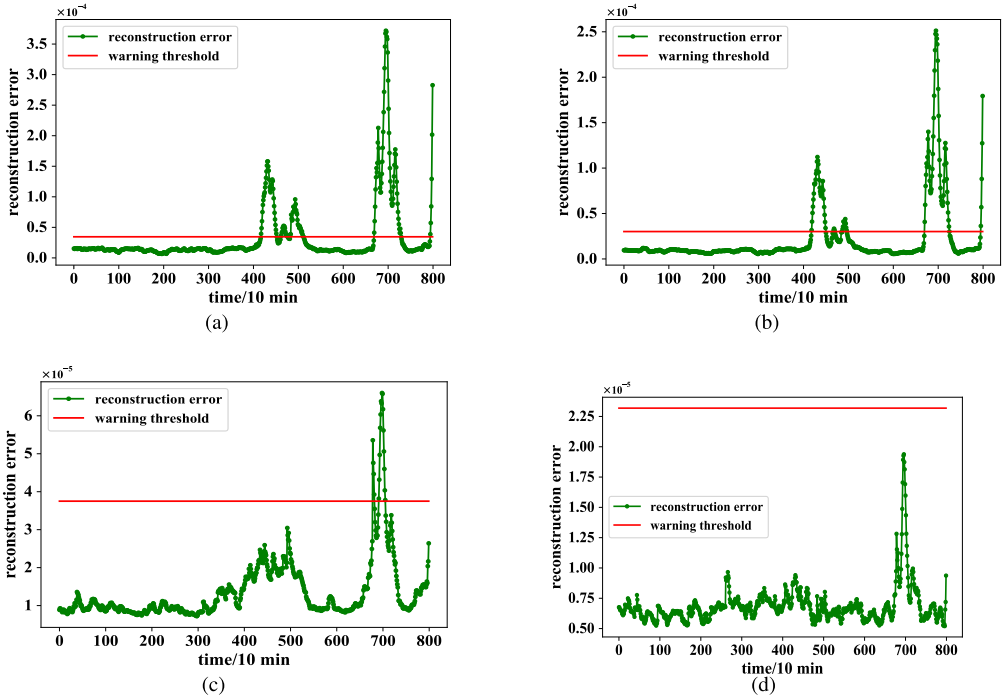


FIGURE 16. Fault warning results of #12 based on (a) the data processed by the proposed method, (b) the data processed by the low dimensional abnormal data processing algorithm, (c) the data processed by the high dimensional abnormal data processing algorithm, and (d) the data processed by the criterion: $3\ m/s < V_W < 25\ m/s$ and $100\ kw < P$.

and 40 minutes in advance). Through a comparison between Fig. 16(a) and Fig. 16(b), it can be seen that the method proposed in this article is able to identify more abnormal operational states. As shown in Fig. 16(c), the model based on the data processed by the high dimensional abnormal data processing algorithm detected the fault 122 data points in

TABLE 4. Fault prediction results of different fault types.

Fault type	Method A	Method B	Method C	Method D
Fault A	4 days, 3 hours, and 40 minutes	4 days, 3 hours, and 10 minutes	3 days, 15 hours, and 30 minutes	0 day
Fault B	2 days and 50 minutes	1 day, 18 hours, and 10 minutes	2 days and 50 minutes	0 day
Fault C	2 days, 15 hours, and 50 minutes	2 days, 15 hours, and 40 minutes	20 hours and 20 minutes	0 day

advance (20 hours and 20 minutes) which is 1 day, 19 hours, and 30 minutes later than the method proposed in this article. In Fig. 16(d), the model based on the data processed by the criterion of (3 m/s $< V_W <$ 25 m/s and 100 kW $< P$) did not detect faults.

IV. CONCLUSION AND FUTURE WORK

The acquisition of high-quality modeling data plays an important role in the establishment of the normal behavioral model, which affects the early detection performance of potential faults. To guarantee the quality of modeling data and reduce the time needed to manually select modeling data, this article proposes an abnormal data processing model based on an unsupervised serialization ensemble algorithm. In addition, the influence of the quality of modeling data on the performance of fault prediction is studied. Based on the above case studies, the following conclusions can be drawn: (a) Based on a comparison of different abnormal data processing methods, including the unsupervised ensemble method proposed in this article, the method only considering the high dimensional characteristics of data and the method only considering the known low dimensional variable relationships, the abnormal data processing method proposed in this article shows the best performance. (b) According to the analyses of fault early warning cases based on four different modeling datasets, it can be seen that the performance of the fault prediction model can be effectively enhanced by improving the quality of modeling data.

REFERENCES

[1] X. Long, P. Yang, H. Guo, and W. U. Xiwen, "Review of fault diagnosis methods for large wind turbines," *Power Syst. Technol.*, May 2017, pp. 3480–3491.

[2] G. Jiang, P. Xie, H. He, and J. Yan, "Wind turbine fault detection using a denoising autoencoder with temporal information," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 1, pp. 89–100, Feb. 2018.

[3] M. Rezamand, M. Kordestani, R. Carriveau, D. S.-K. Ting, and M. Saif, "A new hybrid fault detection method for wind turbine blades using recursive PCA and wavelet-based PDF," *IEEE Sensors J.*, vol. 20, no. 4, pp. 2023–2033, Feb. 2020.

[4] P. F. Odgaard and J. Stoustrup, "Fault tolerant control of wind turbines using unknown input observers," *IFAC Proc. Volumes*, vol. 45, no. 20, pp. 313–318, Jan. 2012.

[5] X. Wei, M. Verhaegen, and T. van Engelen, "Sensor fault detection and isolation for wind turbines based on subspace identification and Kalman filter techniques," *Int. J. Adapt. Control Signal Process.*, vol. 24, no. 8, pp. 687–707, Dec. 2010.

[6] F. Cheng, L. Qu, and W. Qiao, "Fault prognosis and remaining useful life prediction of wind turbine gearboxes using current signal analysis," *IEEE Trans. Sustain. Energy*, vol. 9, no. 1, pp. 157–167, Jan. 2018.

[7] M. Kordestani, M. Saif, M. E. Orchard, R. Razavi-Far, and K. Khorasani, "Failure prognosis and Applications—A survey of recent literature," *IEEE Trans. Rel.*, early access, Sep. 17, 2019, doi: 10.1109/TR.2019.2930195.

[8] Z. Liu, X. Wang, and L. Zhang, "Fault diagnosis of industrial wind turbine blade bearing using acoustic emission analysis," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 9, pp. 6630–6639, Sep. 2020.

[9] B. T. Thumati, M. A. Feinstein, and S. Jagannathan, "A model-based fault detection and prognostics scheme for Takagi–Sugeno fuzzy systems," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 4, pp. 736–748, Aug. 2014.

[10] H. Zhao, H. Liu, W. Hu, and X. Yan, "Anomaly detection and fault analysis of wind turbine components based on deep learning network," *Renew. Energy*, vol. 127, pp. 825–834, Nov. 2018.

[11] M. Rezamand, M. Kordestani, M. E. Orchard, R. Carriveau, D. S.-K. Ting, and M. Saif, "Improved remaining useful life estimation of wind turbine drivetrain bearings under varying operating conditions," *IEEE Trans. Ind. Informat.*, vol. 17, no. 3, pp. 1742–1752, Mar. 2021.

[12] Q. Zhao, K. Bao, J. Wang, Y. Han, and J. Wang, "An online hybrid model for temperature prediction of wind turbine gearbox components," *Energies*, vol. 12, no. 20, p. 3920, Oct. 2019.

[13] P. Qian, D. Zhang, X. Tian, Y. Si, and L. Li, "A novel wind turbine condition monitoring method based on cloud computing," *Renew. Energy*, vol. 135, pp. 390–398, May 2019.

[14] G. Helbing and M. Ritter, "Deep learning for fault detection in wind turbines," *Renew. Sustain. Energy Rev.*, vol. 98, pp. 189–198, Dec. 2018.

[15] L. Wang, Z. Zhang, J. Xu, and R. Liu, "Wind turbine blade breakage monitoring with deep autoencoders," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 2824–2833, Jul. 2018.

[16] Z. Kong, B. Tang, L. Deng, W. Liu, and Y. Han, "Condition monitoring of wind turbines based on spatio-temporal fusion of SCADA data by convolutional neural networks and gated recurrent units," *Renew. Energy*, vol. 146, pp. 760–768, Feb. 2020.

[17] H. Zhao, H. Liu, H. Liu, and Y. Lin, "Condition monitoring and fault diagnosis of wind turbine generator based on stacked autoencoder network," *Autom. Electr. Power Syst.*, vol. 42, no. 11, pp. 102–108, Jan. 2018.

[18] Y. Zhang, H. Zheng, J. Liu, J. Zhao, and P. Sun, "An anomaly identification model for wind turbine state parameters," *J. Cleaner Prod.*, vol. 195, pp. 1214–1227, Sep. 2018.

[19] C. Yang, J. Liu, Y. Zeng, and G. Xie, "Real-time condition monitoring and fault detection of components based on machine-learning reconstruction model," *Renew. Energy*, vol. 133, pp. 433–441, Apr. 2019.

[20] P. Cross and X. Ma, "Model-based and fuzzy logic approaches to condition monitoring of operational wind turbines," *Int. J. Autom. Comput.*, vol. 12, no. 1, pp. 25–34, Feb. 2015.

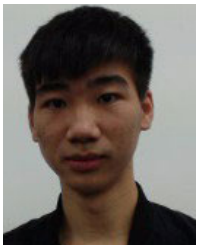
[21] P. Bangalore, S. Letzgus, D. Karlsson, and M. Patriksson, "An artificial neural network-based condition monitoring method for wind turbines, with application to the monitoring of the gearbox," *Wind Energy*, vol. 20, no. 8, pp. 1421–1438, Aug. 2017.

[22] P. Bangalore and L. B. Tjernberg, "An artificial neural network approach for early fault detection of gearbox bearings," *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 980–987, Mar. 2015.

[23] P. Sun, J. Li, C. Wang, and X. Lei, "A generalized model for wind turbine anomaly identification based on SCADA data," *Appl. Energy*, vol. 168, pp. 550–567, Apr. 2016.

[24] M. Schlechtingen and I. F. Santos, "Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection," *Mech. Syst. Signal Process.*, vol. 25, no. 5, pp. 1849–1875, Jul. 2011.
[25] A. Zaher, S. D. J. McArthur, D. G. Infield, and Y. Patel, "Online wind turbine fault detection through automated SCADA data analysis," *Wind Energy*, vol. 12, no. 6, pp. 574–593, Sep. 2009.
[26] G. Helbing and M. Ritter, "Power curve monitoring with flexible EWMA control charts," in *Proc. Int. Conf. Promising Electron. Technol. (ICPET)*, Oct. 2017, pp. 124–128.
[27] L. Wang, Z. Zhang, H. Long, J. Xu, and R. Liu, "Wind turbine gearbox failure identification with deep neural networks," *IEEE Trans. Ind. Informat.*, vol. 13, no. 3, pp. 1360–1368, Jun. 2017.
[28] E. Lapira, D. Brisset, H. Davari Ardakani, D. Siegel, and J. Lee, "Wind turbine performance assessment using multi-regime modeling approach," *Renew. Energy*, vol. 45, pp. 86–95, Sep. 2012.
[29] A. Kusiak and A. Verma, "Analyzing bearing faults in wind turbines: A data-mining approach," *Renew. Energy*, vol. 48, pp. 110–116, Dec. 2012.
[30] J. Li, X. Lei, H. Li, and L. Ran, "Normal behavior models for the condition assessment of wind turbine generator systems," *Electr. Power Compon. Syst.*, vol. 42, no. 11, pp. 1201–1212, Aug. 2014.
[31] K. Leahy, C. Gallagher, P. O'Donovan, K. Bruton, and D. O'Sullivan, "A robust prescriptive framework and performance metric for diagnosing and predicting wind turbine faults based on SCADA and alarms data with case study," *Energies*, vol. 11, no. 7, p. 1738, Jul. 2018.
[32] J. Tautz-Weinert and S. J. Watson, "Challenges in using operational data for reliable wind turbine condition monitoring," in *Proc. Int. Soc. Offshore Polar Eng. (ISOPE)*, 2017, pp. 25–30.
[33] Z. Sun and H. Sun, "Stacked denoising autoencoder with density-grid based clustering method for detecting outlier of wind turbine components," *IEEE Access*, vol. 7, pp. 13078–13091, 2019.
[34] T. Amarbayasgalan, B. Jargalsaikhan, and K. Ryu, "Unsupervised novelty detection using deep autoencoders with density based clustering," *Appl. Sci.*, vol. 8, no. 9, p. 1468, Aug. 2018.
[35] A. Zimek, E. Schubert, and H.-P. Kriegel, "A survey on unsupervised outlier detection in high-dimensional numerical data," *Stat. Anal. Data Mining*, vol. 5, no. 5, pp. 363–387, Oct. 2012.
[36] C. Song, Y. Huang, F. Liu, Z. Wang, and L. Wang, "Deep auto-encoder based clustering," *Intell. Data Anal.*, vol. 18, no. 6S, pp. S65–S76, Dec. 2014.
[37] X. Guo, L. Gao, X. Liu, and J. Yin, "Improved deep embedded clustering with local structure preservation," in *Proc. IJCAI*, 2017, pp. 1753–1759.
[38] Y. Xia, X. Cao, F. Wen, G. Hua, and J. Sun, "Learning discriminative reconstructions for unsupervised outlier removal," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1511–1519.
[39] H. Long, L. Sang, Z. Wu, and W. Gu, "Image-based abnormal data detection and cleaning algorithm via wind power curve," *IEEE Trans. Sustain. Energy*, vol. 11, no. 2, pp. 938–946, Apr. 2020.
[40] Y. Ren, X. Liu, and W. Liu, "DBCAMM: A novel density based clustering algorithm via using the mahalanobis metric," *Appl. Soft Comput.*, vol. 12, no. 5, pp. 1542–1554, May 2012.
[41] Y. Zhu, K. M. Ting, and M. J. Carman, "Density-ratio based clustering for discovering clusters with varying densities," *Pattern Recognit.*, vol. 60, pp. 983–997, Dec. 2016.
[42] T. Y. Goh, S. N. Basah, H. Yazid, M. J. A. Safar, and F. S. A. Saad, "Performance analysis of image thresholding: Otsu technique," *Measurement*, vol. 114, pp. 298–307, Jan. 2018.

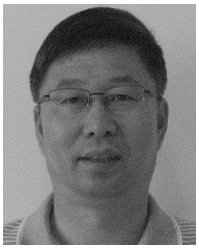
QIANG ZHAO received the Ph.D. degree in navigation guidance and control from the College of Information Science and Engineering, Northeastern University, Shenyang, China, in 2017. He is currently a Lecturer of measurement and control technology and instrumentation program with the School of Control Engineering, Northeastern University at Qinhuangdao. His research interests include electricity theft detection, fault diagnosis of renewable energy generation systems, and optimal operation of smart grid.



KUNKUN BAO received the B.S. degree from the Electromechanic Engineering College, Nanyang Normal University. He is currently pursuing the master's degree with the School of Control Engineering, Northeastern University at Qinhuangdao, China. His research interests include industrial big data analysis and condition monitoring and fault diagnosis of wind turbines.

ZHENFAN WEI received the B.S. degree from the College of Electronic Information Engineering, Hebei University, China, in 2020. He is currently pursuing the master's degree with the School of Control Engineering, Northeastern University at Qinhuangdao, China. His research interests include planning and operation of coupled power systems and transportation systems.

YINGHUA HAN received the Ph.D. degree from the College of Information Science and Engineering, Northeastern University, Shenyang, China, in 2008. She is currently a Professor with the School of Computer and Communication Engineering, Northeastern University at Qinhuangdao. Her research interests include edge computing and edge cloud collaborative application, industrial artificial intelligence, industrial big data mining and applications, and optimal control of smart grids.



JINKUAN WANG received the M.S. degree from Northeastern University, Shenyang, China, in 1985, and the Ph.D. degree from the University of Electro-Communications, Chofu, Japan, in 1993. In 1990, he joined the Institute of Space Astronautical Science, Japan, as a Special Member. In 1994, he was an Engineer with the Research Department, COSEL, Japan. Since 1998, he has been a Professor with the College of Information Science and Engineering, Northeastern University. His research interests include intelligent control, adaptive array, wireless sensor networks, and optimal operation of smart grids.