Information Sciences 648 (2023) 119496

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Semi-supervised adversarial discriminative learning approach for intelligent fault diagnosis of wind turbine

Te Han [a,b], Wenzhen Xie [c,*], Zhongyi Pei [d,e]

[a] Center for Energy and Environmental Policy Research, Beijing Institute of Technology, Beijing, 100081, China
[b] School of Management and Economics, Beijing Institute of Technology, Beijing, 100081, China
[c] Department of Energy and Power Engineering, Tsinghua University, Beijing, 100084, China
[d] National Engineering Research Center for Big Data Software, Beijing, 100084, China
[e] School of Software, Tsinghua University, Beijing, 100084, China

ARTICLE INFO

Keywords:
Wind turbine
Data-driven
Intelligent diagnosis
Semi-supervised adversarial learning
Metric learning

ABSTRACT

Wind turbines play a crucial role in renewable energy generation systems and are frequently exposed to challenging operational environments. Monitoring and diagnosing potential faults during their operation is essential for improving reliability and reducing maintenance costs. Supervised learning using data-driven techniques, particularly deep learning, offers a viable approach for developing fault diagnosis models. However, a significant challenge in practical wind power equipment lies in the scarcity of annotated samples required to train these models effectively. This paper proposes a semi-supervised fault diagnosis approach specifically designed for wind turbines, aiming to address this challenge. Initially, a semi-supervised deep neural network is constructed using adversarial learning, where a limited set of annotated samples is used in conjunction with a vast amount of unannotated samples. The health status features present in the unannotated samples are leveraged to capture a generalized representation of the underlying features. Subsequently, a metric learning-guided discriminative features enhancement technique is employed to improve the separability of different manifolds, thereby enhancing the performance of the semi-supervised training process. By employing this methodology, it becomes possible to develop a fault diagnosis model with superior accuracy using only a limited amount of annotated samples. Comprehensive fault diagnosis experiments were conducted on a wind turbine fault dataset, revealing the efficacy and superiority of the presented methodology.

1. Introduction

Wind energy for power generation has seen tremendous growth in recent decades around the world. The wind turbine is the critical energy conversion device in wind power generation. However, the mechanical structure of wind turbine is complex, and its working conditions are also harsh and complex. Therefore, mechanical failure of wind turbine often occurs (including bearings, coupling, pedestal, etc.). Health monitoring and diagnosis are essential for enhancing the dependability of wind turbines, minimizing operation and maintenance expenses, and preventing severe accidents [1]. At the heart of this process lies the development of

* Corresponding author.
E-mail address: xwz18@tsinghua.org.cn (W. Xie).

https://doi.org/10.1016/j.ins.2023.119496
Received 22 May 2023; Received in revised form 20 July 2023; Accepted 7 August 2023

Available online 14 August 2023

an intelligent diagnosis model that can automatically learn fault feature representation from monitoring data and make informed decisions [2,3].

In light of the swift advancements made in artificial intelligence (AI) in recent years, there have been numerous endeavors aimed at leveraging this technology in the area of intelligent fault diagnosis. AI-assisted diagnosis approaches can significantly reduce the reliance on prior knowledge of physical fault mechanisms [4,5]. As industries generate vast amounts of big data, conventional manual methods for feature extraction and classification in intelligent fault diagnosis have become increasingly complex [6]. To address this, deep learning methods are gaining popularity due to their ability to learn features adaptively [7–9]. Deep learning models, with their multi-layered structures, facilitate hierarchical feature representation, enabling more accurate fault diagnosis decisions [10,11]. These models have already demonstrated successful applications in practical industrial scenarios, including fault diagnosis [12–14], prognosis and remaining life prediction of wind turbine components [15].

Undoubtedly, deep learning provides a new framework for intelligent fault diagnosis in wind power generation devices. However, there are challenges associated with the practical implementation of these methods. One notable challenge is the requirement for a significant number of annotated samples to train deep learning models for fault diagnosis [16]. In the case of industrial wind turbine equipment, acquiring a sufficient number of annotated fault samples is often a challenging task. Many wind turbine companies have limited fault data accumulated for critical components, and even when extensive data are collected by sensors, manually labeling them is resource-intensive. Consequently, the majority of the data remains unannotated, with only a few samples assumed to be annotated. Therefore, effectively utilizing the available unannotated practical data remains crucial for improving the performance of training models. To this end, semi-supervised learning has a potential prospect in the practical diagnosis applications.

In the field of semi-supervised learning, several common mainstream methods have emerged, including self-training [17,18], graph-based methods such as label propagation [19], and generative methods such as generative adversarial network (GAN) [20] or variational autoencoders (VAE). Recently, researchers have been exploring and proposing adaptive methods specifically tailored for semi-supervised model training in the context of fault diagnosis [21–24]. Chen et al. introduced a semi-supervised random forest within the framework of graph-based methods to effectively diagnose gearbox faults in industrial systems [25]. Zhang et al. proposed a VAE based semi-supervised diagnosis network, which was evaluated for fault diagnosis performance using different numbers of annotated samples [26]. Moradi et al. presented a semi-supervised deep model tailored for extracting a health indicator from structural health monitoring (SHM) data, with a particular focus on monitoring fatigue-induced loading in materials [27]. Zhou et al. designed a semi-supervised GAN approach to effectively leverage the rich fault features present in unannotated data, thereby achieving accurate diagnosis of gear faults in scenarios with limited supervision [28].

Undoubtedly, maximizing the utilization of health status features present in both the limited annotated data and abundant unannotated data is crucial for enhancing the diagnostic capability of models. Although numerous approaches have been designed to leverage unannotated data to improve learning performance, the fundamental problem of sampling bias in semi-supervised learning has received limited attention in the existing literatures. Indeed, the observed distribution of the finite number of annotated samples may diverge from the true underlying distribution [29]. In this context, the integration of annotated and unannotated sample distributions through alignment emerges as a potent approach in semi-supervised learning. Adversarial learning, primarily employed in transfer learning to facilitate cross-domain knowledge transfer by minimizing domain shift and extracting domain-invariant features [30,31], has also demonstrated its effectiveness in semi-supervised learning. Wang et al. firstly considered the sampling bias problem by proposing a semi-supervised learning approach that incorporates adversarial distribution alignment [29]. Si et al. conducted a study in which they employed a tightly coupled approach of adversarial learning and semi-supervised learning to align the distributions between annotated and unannotated data, resulting in improved accuracy for 3D action recognition [32]. Mayer et al. employed adversarial semi-supervised learning to enhance the performance across a range of computer vision tasks [33]. While adversarial-based semi-supervised learning shows great potential, it is important to acknowledge the limited research conducted on its application in the domain of fault diagnosis. As a result, there is a promising opportunity to investigate the integration of annotated and unannotated samples through adversarial learning, aiming to enhance the diagnostic capabilities of models specifically for wind turbine fault diagnosis.

Semi-supervised learning is premised on the manifold assumption, which posits that samples belonging to the same health condition are distributed within a shared manifold [34]. Unlike supervised training, where the distributions of different health conditions can typically be separated, semi-supervised learning leverages unannotated samples to better exploit the underlying structure of data manifold. However, when handling the diagnosis task with extremely limited annotated samples, it is difficult to achieve efficient decision boundary of different manifolds/distributions for deep learning model. The effectiveness of semi-supervised learning may deteriorate. To mitigate this issue, it is crucial to capture a more discriminative decision boundary in deep models. Metric learning techniques hold great promise in facilitating this objective by increasing the distance between the samples from different manifolds in the deep feature embedding space, and reduce the distance between the samples from the same manifold [35,36]. In this manner, the learned deep models are more discriminative for the wind turbines with different faults.

Based on the problems mentioned above, a semi-supervised approach using discriminative features and adversarial learning is introduced for the purpose of detecting faults in wind turbines. The proposed semi-supervised approach can achieve highly accurate diagnosis decisions of wind turbine faults with only limited labeled samples, which shows strong robustness and generalization ability. The major contributions of this research are outlined below:

1) An adversarial semi-supervised learning framework is introduced, which allows the full utilization of health status information contained in unlabeled samples during the training process, leading to the extraction of a more generalized feature representation and improved diagnostic performance.

2) By incorporating triplet loss-based metric learning, our proposed semi-supervised learning framework enhances feature clustering within the same category and improves feature separation between different categories. This approach effectively assists the learning process, resulting in superior diagnostic performance, especially in scenarios with limited availability of labeled samples.

The remainder of this article is organized as follows: Section 2 outlines the problem statement and introduces the proposed methodology, while Section 3 provides a detailed account of the wind turbine experiments conducted. The result analysis is conducted and discussed in Section 4. Section 5 summarizes the conclusions drawn from the research.

2. Methodology

2.1. Problem statement

The primary focus of this study is on developing intelligent diagnosis method for wind turbines. A semi-supervised approach is proposed to establish the intelligent model with limited supervisions. Specifically, few labeled training data $\mathbf{X}_l = \{(\mathbf{x}_1^l, y_1^l), \ldots, (\mathbf{x}_n^l, y_n^l)\}$ from different health conditions are assumed to be available. $\mathbf{x}_i^l$ is a monitoring sample from wind turbine, and y_i^l is the corresponding health condition, namely the label. In addition to the limited labeled samples, the massive monitoring data are unlabeled, denoted as $\mathbf{X}_u = \{\mathbf{x}_1^u, \ldots, \mathbf{x}_m^u\}$. Generally, the $l \ll u$. The proposed semi-supervised approach aims to learn the diagnosis model by comprehensively utilizing the $\mathbf{X}_l$ and $\mathbf{X}_u$. The diagnostic model, which comprises a multi-layered feature extractor $F(\cdot, \theta_f)$ and a diagnostic classifier $C(\cdot, \theta_c)$, is trained to facilitate the diagnosis of wind turbine faults. The θ_f and θ_c are the parameters of diagnosis model, which need to be optimized during model training. With the trained model, the final diagnosis decision of the input sample can be achieved, that is, $C(F(\mathbf{x}_i))$.

To leverage the health information embedded in $\mathbf{X}_u$ from unlabeled data, the adversarial learning is conducted between the $\mathbf{X}_l$ and $\mathbf{X}_u$, so that regularizing the deep feature extractor $F(\cdot, \theta_f)$ to capture the generalized feature representation. Meanwhile, the metric learning is applied to enhance the compactness of samples from same health category and the separability of samples from different health categories in the deep feature space. The more discriminative features can be learned in this step with the supervisions of $\mathbf{X}_l$. By integrating the two considerations, the proposed semi-supervised fault diagnosis approach for wind turbines is anticipated to yield superior results, particularly when only a limited number of annotated training samples are available.

2.2. Semi-supervised fault diagnosis framework for wind turbine

This article introduces a novel semi-supervised adversarial discriminative (SSAD) network for wind turbine fault diagnosis and system maintenance in scenarios with limited annotated samples. The framework, as depicted in Fig. 1, encompasses the following key steps:

Step1: Collect the vibration data of the critical components of wind turbine systems by sensors and data acquisition system.

Step2: Pre-process the collected vibration data and divide them into samples for model training. The training samples contain the limited annotated samples and massive unannotated samples.

Step3: Establish the SSAD network, and train the SSAD network with both the annotated and unannotated samples.

Step4: Deploy the trained network for real-time monitoring. Use the testing samples to verify the trained model, and analyze the model's performance.

Step5: Collect the diagnosis decisions of wind turbine health conditions. Conduct maintenance and health management according to the diagnosis decisions.

The architecture of the SSAD network is shown in Fig. 1. Table 1 presents its detailed structure parameters. The SSAD network is comprised of three parts: namely feature extractor $F(\cdot, \theta_f)$, diagnosis classifier $C(\cdot, \theta_c)$ and discriminator $D(\cdot, \theta_d)$, to implement the semi-supervised training and finally diagnose the health conditions of the wind turbine. The input training data for the networks include the limited labeled sample $\mathbf{X}_l$ and massive unlabeled sample $\mathbf{X}_u$. Obviously, the diagnosis ability of the model is obtained by using the $\mathbf{X}_l$. The basic training process follows the commonly used deep learning methods [7,37]. To ensure efficient feature extraction, the feature extractor contains five convolutional and pooling layers (from Conv_1&Pooling_1 to Conv_5&Pooling_5). After that, the feature embeddings are flattened and mapped to the health condition label by classifier. To avoid overfitting with limited labeled samples, the classifier adopts a simple structure with fewer trainable parameters.

2.3. Adversarial learning-based semi-supervised model

In the proposed diagnosis approach, the adversarial learning is adopted to capture the generalized feature representation by leveraging the health condition information from both annotated and unannotated data. Taking the binary classification as an example (health and fault), the illustrations of semi-supervised training based on adversarial learning are shown in Fig. 2. $F(\mathbf{X}_l)$ and $F(\mathbf{X}_u)$ denote the distributions of annotated and unannotated data, respectively. Before adversarial learning, the supervised training under extremely limited labeled samples may not be able to acquire an ideal decision boundary due to the lack of adequate health status information. The decision boundary learned from limited labeled samples may not adequately capture the feature distribution of the extensive unlabeled samples, i.e., $F(\mathbf{X}_u)$, leading to the unsatisfactory diagnosis performance. The adversarial learning stage is incorporated to effectively leverage the health information present in both labeled and unlabeled samples, thus obtaining a more generalized feature representation.

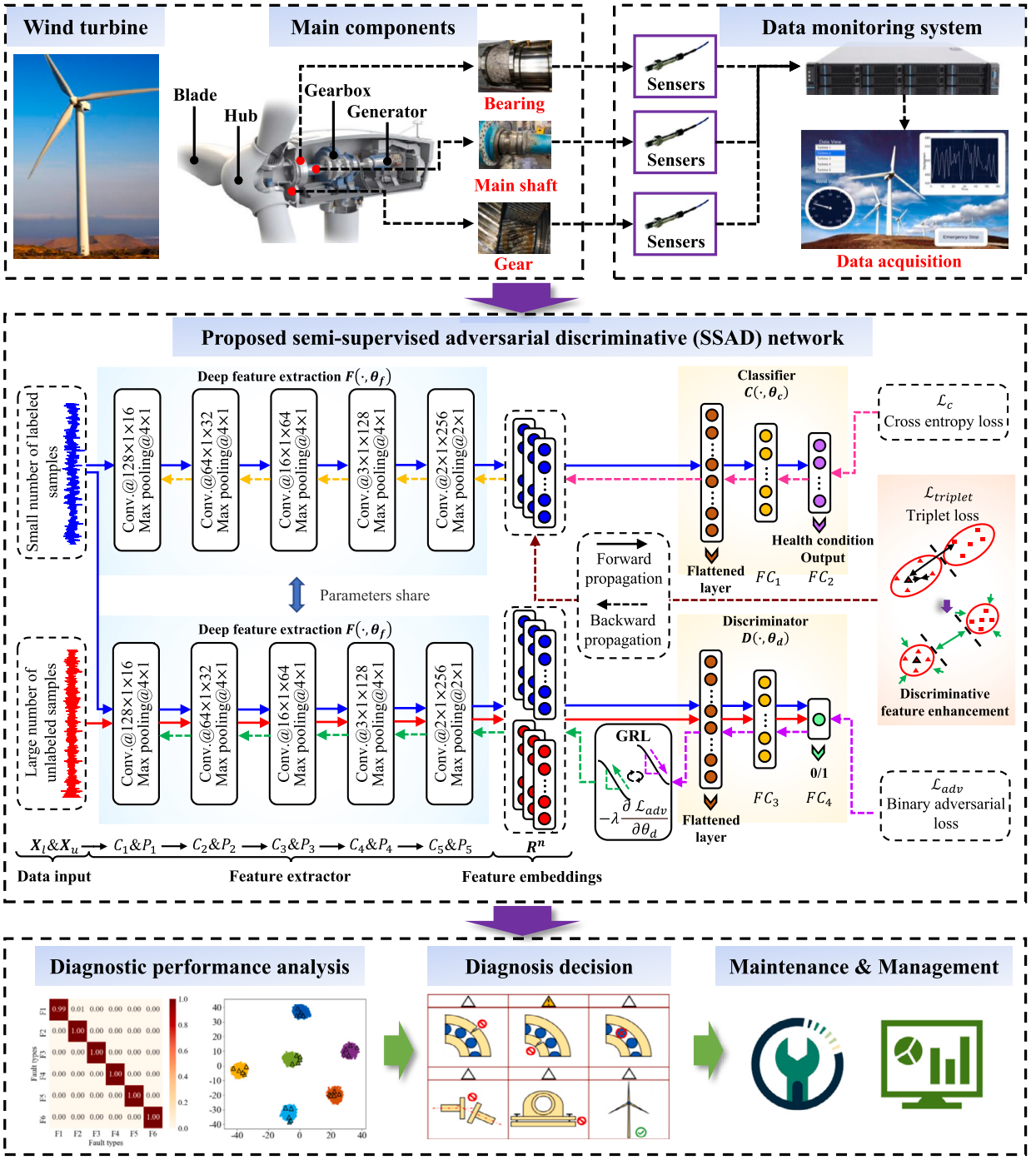

Fig. 1. Proposed semi-supervised fault diagnosis framework for wind turbine.

To achieve the above goal, another binary discriminator $D(\cdot,\theta_d)$ is constructed. The θ_d is the network parameters of the discriminator. The adversarial learning is conducted between the D and F by regularizing the training of F. Given a sample $\mathbf{x}_i$, the health status features can be extracted by F, denoted as $F(\mathbf{x}_i)$. During adversarial learning, the D aims to distinguish the source of the extracted features from labeled data $\mathbf{X}_l$ or unlabeled data $\mathbf{X}_u$, while the F is updated to generate more generalized features to confuse the D. In the manner, the F is trained with both of the annotated and unannotated data. The output is described as below [30,31]:

$$\mathcal{L}_{adv}\left(\mathbf{x}_i^l,\mathbf{x}_i^u,F,D\right)=-\mathbb{E}_{\mathbf{x}_i\sim\mathbf{X}_l}\left[\log D\left(F\left(\mathbf{x}_i\right)\right)\right]-\mathbb{E}_{\mathbf{x}_i\sim\mathbf{X}_u}\left[\log\left(1-D\left(F\left(\mathbf{x}_i\right)\right)\right)\right]. \tag{1}$$

Table 1
Parameters of the SSAD network.

Module	Layers	Parameters size	Activation function
Feature extractor	Input layer	Length of sample	/
$F(\cdot,\theta_f)$	Conv_1&Pooling_1	Convolution: $128\times1\times16$	ReLU
		Max pooling: 4×1(zero padding)	
	Conv_2&Pooling_2	Convolution: $64\times1\times32$	
		Max pooling: 4×1(zero padding)	ReLU
	Conv_3&Pooling_3	Convolution: $16\times1\times64$	
		Max pooling: 4×1(zero padding)	ReLU
	Conv_4&Pooling_4	Convolution: $3\times1\times128$	
		Max pooling: 4×1(zero padding)	ReLU
	Conv_5&Pooling_5	Convolution: $2\times1\times256$	
		Max pooling: 2×1(zero padding)	ReLU
Classifier	Fully-connected_1	128	ReLU
$C(\cdot,\theta_c)$	Fully-connected_2	N_c	Softmax
Discriminator	Fully-connected_3	128	ReLU
$D(\cdot,\theta_d)$	Fully-connected_4	1	Softmax

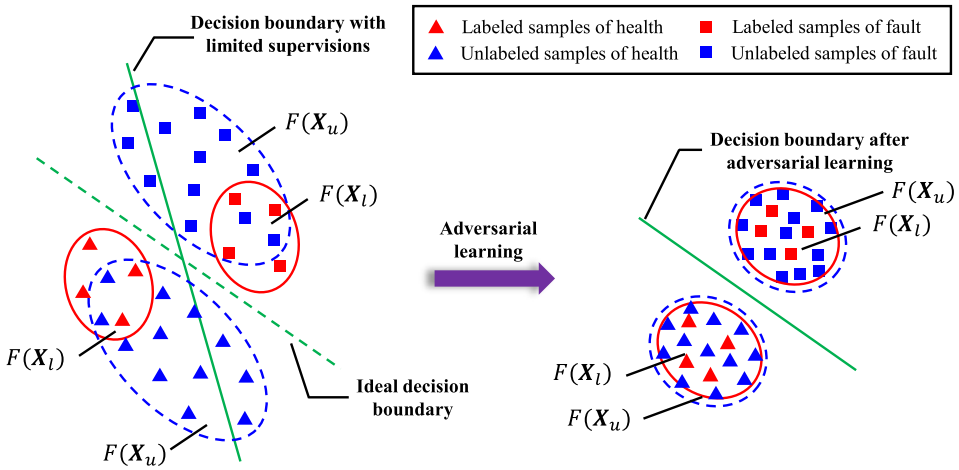

Fig. 2. The illustrations of semi-supervised training based on adversarial learning.

The general formulation of adversarial learning can be derived as below:

$$\begin{aligned}&\min_{D}\mathcal{L}_{adv}\left(\mathbf{x}_i^l,\mathbf{x}_i^u,F\right),\\&\max_{F}\mathcal{L}_{adv}\left(\mathbf{x}_i^l,\mathbf{x}_i^u,D\right).\end{aligned}\tag{2}$$

More details about the network architecture and parameters optimization will be described in Section 2.4.

Adversarial learning plays a crucial role in achieving marginal distribution alignment between annotated and unannotated samples [33]. In scenarios where annotated and unannotated data are acquired from the same health condition, it is generally assumed that their distributions exhibit similarity. However, the limited size of annotated data can lead to deviations in the empirical distribution, resulting in a disparity from the true distribution. This phenomenon, known as empirical distribution mismatching, highlights the discrepancy between the distributions of annotated and unannotated data. In such cases, the utilization of adversarial learning, which incorporates both annotated and unannotated data, becomes pertinent. This approach aims to mitigate the discrepancy between the empirical and true distributions by aligning their respective distributions. By leveraging the additional information present in the unannotated data, the model can enhance its robustness and improve its generalization ability to unseen data. The essence of adversarial learning lies in enabling the model to effectively capture the underlying data distribution, thereby strengthening its capacity to generalize to unseen examples.

2.4. Metric learning-based optimization

While the proposed semi-supervised training with adversarial learning allows for the utilization of the feature information present in large amounts of unannotated data, the negative effects may occur when the feature distribution boundary between different health conditions is not clear. For instance, the feature extractor F may capture the shared but useless features for the annotated health samples and unannotated fault samples, and successfully confuse the discriminator D during adversarial training. To address this

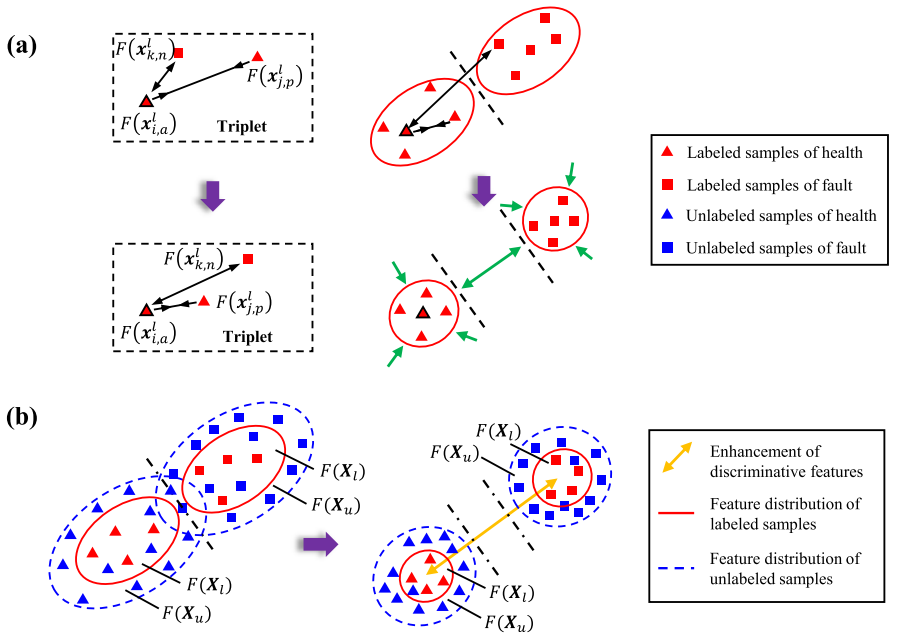

Fig. 3. The illustrations of discriminative features enhancement. (a) Triplet loss optimization, (b) The effects for semi-supervised training.

issue, a metric learning-based discriminative features enhancement is introduced to enhance the performance of semi-supervised training.

The basic idea of metric learning-based discriminative features enhancement is to minimize the discrepancy between samples with the same health condition, while increase the distance of samples with the different categories [38]. More discriminative features for different health conditions can be captured in this way. As shown in Fig. 3(a), a triplet containing an anchor sample from a certain health condition, a positive sample $\mathbf{x}_{j,p}^{l}$ from the same health condition, and a negative sample $\mathbf{x}_{k,n}^{l}$ from a different health condition, is constructed. After processing by feature extractor F, the desired discriminative features in the deep feature space are formulated as:

$$\left\| F\left(\mathbf{x}_{i,a}^{l}\right) - F\left(\mathbf{x}_{j,p}^{l}\right) \right\|_2 + \alpha < \left\| F\left(\mathbf{x}_{i,a}^{l}\right) - f\left(x_{k,n}^{l}\right) \right\|_2, \tag{3}$$

where α is the margin. It can also be interpreted as the measure of dissimilarity between samples from different labels. Supposing N triplets from the $\mathbf{X}_l$, the triplet loss can be formalized as:

$$\mathcal{L}_{triplet}\left(\mathbf{x}_i^l, \mathbf{y}_i^l, F\right) = \sum_{i=1}^{N} \max\left(\left[\left\| F\left(\mathbf{x}_{i,a}^{l}\right) - F\left(\mathbf{x}_{j,p}^{l}\right) \right\|_2 - \left\| F\left(\mathbf{x}_{i,a}^{l}\right) - F\left(\mathbf{x}_{k,p}^{l}\right) \right\|_2 + \alpha \right], 0 \right). \tag{4}$$

Optimizing the triplet loss is capable of learning more discriminative features, and thus ensuring a more clear decision boundary for further classification and diagnosis.

Note that the aforementioned triplet loss optimization is only conducted in a supervised manner. Considering the manifold assumptions, the embedded features of annotated and unannotated samples with the same health condition distribute in the same manifold. When the triplet loss optimization is performed on labeled samples, the learned features are more discriminative, and the manifolds/distributions associated with different health conditions are more separable. As illustrated in Fig. 3(b), such optimization can also enhance the performance of the semi-supervised learning, leading to a more discriminative boundary between different health conditions.

2.5. Training of the SSAD network

In the training process of SSAD network, the cross entropy is utilized as the optimization objective for the fault diagnosis network. The formulation is given as follows:

$$\mathcal{L}_{cls}\left(\mathbf{x}_i^l, \mathbf{y}_i^l, F, C\right) = -\sum_{n=1}^{N_c} \mathbb{1}\left(n = \mathbf{y}_i^l\right) \log C\left(F\left(\mathbf{x}_i^l\right)\right), \tag{5}$$

where N_c is the number of defined health conditions that need to be identified.

Table 2
The descriptions of different health conditions of wind turbine.

health conditions	label
Health	0
Back bearing pedestal loosen fault	1
Roller fault of bearing	2
Inner race fault of bearing	3
Outer race fault of bearing	4
Coupling misalignment fault	5

In addition to the commonly used supervised training, the triplet loss optimization is also conducted to update the parameters of deep feature extractor, so as to achieve more discriminative features. The loss function used for supervised learning is defined as follows:

$$\mathcal{L}_{supervised} = \mathcal{L}_{cls}\left(\mathbf{x}_i^l, \mathbf{y}_i^l, F, C\right) + \beta \cdot \mathcal{L}_{triplet}\left(\mathbf{x}_i^l, \mathbf{y}_i^l, F\right), \tag{6}$$

where β is the parameter to trade-off between cross entropy loss and triplet loss. It is selected from the range of $[0.05, 0.1]$ in this study.

With the aforementioned semi-supervised training using adversarial learning, the mixture of annotated samples $\mathbf{X}_l$ and unannotated samples $\mathbf{X}_u$ is processed by feature extractor $F(\cdot,\theta_f)$ and the discriminator $D(\cdot,\theta_d)$. The adversarial learning between $F(\cdot,\theta_f)$ and $D(\cdot,\theta_d)$ aims to regularize the $F(\cdot,\theta_f)$ to obtain the generalized feature representation. The adversarial loss is described in Eq. (1), and the optimization object is formulized in Eq. (2). The $D(\cdot,\theta_d)$ tries to minimize the adversarial loss $\mathcal{L}_{adv}$ discriminating between samples from the annotated dataset $\mathbf{X}_l$ and the unannotated dataset $\mathbf{X}_u$, while the feature extractor $F(\cdot,\theta_f)$ is trained to maximize the $\mathcal{L}_{adv}$, so that confusing the discriminator $D(\cdot,\theta_d)$.

By integrating the Eq. (1) and Eq. (6), the final optimization objective is formalized as:

$$\begin{aligned} &\min_C \mathcal{L}_{cls}\left(\mathbf{x}_i^l, \mathbf{y}_i^l, F\right), \\ &\min_D \mathcal{L}_{adv}\left(\mathbf{x}_i^l, \mathbf{x}_i^u, F\right), \\ \max_F &-\mathcal{L}_{cls}\left(\mathbf{x}_i^l, \mathbf{y}_i^l, C\right) - \beta \cdot \mathcal{L}_{triplet}\left(\mathbf{x}_i^l, \mathbf{y}_i^l\right) + \mathcal{L}_{adv}\left(\mathbf{x}_i^l, \mathbf{x}_i^u, D\right). \end{aligned} \tag{7}$$

To note, the gradient inversion layer (GRL) is inserted between $F(\cdot,\theta_f)$ and $D(\cdot,\theta_d)$ during the adversarial training process [30]. The GRL is used to change the sign of $\mathcal{L}_{adv}\left(\mathbf{x}_i^l, \mathbf{x}_i^u, F, D\right)$ when back propagating it to the $F(\cdot,\theta_f)$. The basic formula of the GRL is:

$$R(x) = x, \frac{\mathrm{d}R(x)}{\mathrm{d}x} = -\lambda \boldsymbol{I}, \tag{8}$$

where $\boldsymbol{I}$ denotes the identity matrix.

Note that the triplet loss optimization is used to achieve more discriminative features. The optimization with respect to adversarial loss is conducted to realize semi-supervised learning and learn more generalized feature representation. On the one hand, using the triplet loss optimization alone may suffer from overfitting due to the insufficient health status information contained in the limited labeled samples. On the other hand, when the adversarial learning is used alone, the negative alignment of feature distributions between annotated and unannotated samples may degrade the semi-supervised learning. It is necessary to separate the manifolds from different health conditions for the semi-supervised adversarial learning. Therefore, by comprehensively considering the two optimization objectives, the proposed SSAD network can effectively enhance the diagnostic performance of model within a semi-supervised learning framework.

3. Wind turbine fault diagnosis experiments

3.1. Fault experiments of rotor systems in wind turbine

To demonstrate the efficacy of the proposed semi-supervised approach, the experimental verification is carried out using the fault dataset from the wind turbine test rig in Tsinghua University [39]. This test rig could conduct experiments on rotor systems faults of the wind turbine. The wind turbine experimental platform is shown in Fig. 4(a). The wind wheel is connected to the generator through the transmission chain of the wind turbine. The electricity from the generator is stored in an accumulator. The test rig mainly consists of the main bearing, bearing pedestals, rotor, coupling, and generator. As listed in Table 2, health and five fault conditions are considered: health (H), bearing pedestal loosen fault (F1), roller fault of bearing (F2), inner race fault of bearing (F3), outer race fault of bearing (F4) and coupling misalignment (F5). More detailed descriptions and illustrations of different health conditions are shown Fig. 4(b). The vibration data for health monitoring and diagnosis are collected by two acceleration sensors placed on the bearing pedestals. It should be noted that the sampling frequency used in the experiments is 20 kHz. The original monitored vibration waveforms are depicted in Fig. 5, whose characteristics are complicated. Observing the original signals alone, it can be challenging to identify the health conditions.

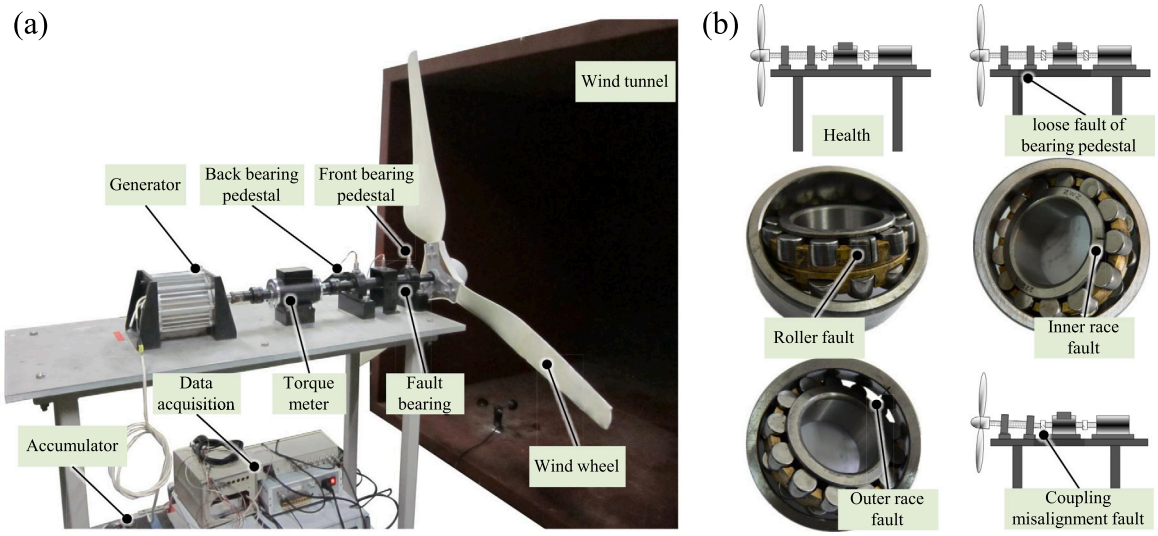

Fig. 4. The wind turbine test rig. (a) Its structure, (b) Details of the health conditions.

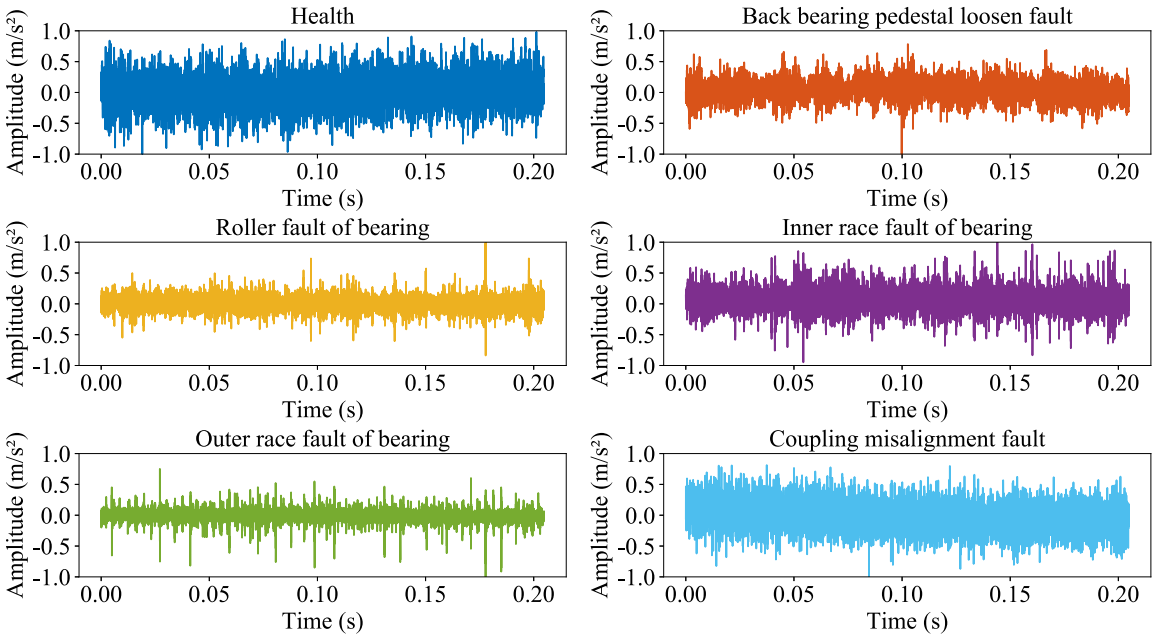

Fig. 5. The monitored vibration signals under different health conditions.

To analyze the diagnostic performance of the semi-supervised SSAD network, wind turbine vibration data under different wind speeds of around 11.5 m/s (seen working condition) and 6.9 m/s (unseen working condition) were used for experimental analysis. The model was trained under the seen condition, while the test was conducted under both seen and unseen working conditions to validate the performance and the generalization ability. The original monitored data from 6 different health conditions are segmented into samples, each containing 4096 sampling points. The proposed method is evaluated through 6 semi-supervised diagnosis experiments, each with a different number of annotated samples, to test its performance and robustness. The numbers of labeled samples in the 6 experiments are 1000, 500, 200, 90, 60 and 30, respectively. During the semi-supervised learning process, another 2000 unannotated samples are also utilized. The trained model is finally tested by another 1000 samples on both known and unknown working conditions, respectively. Ten independent trials are conducted for each experiment with randomly selected training and testing samples.

Table 3
The descriptions of different health conditions of gearbox.

Gear				Bearing						Shaft		Label
32T	96T	48T	80T	IS:IS	ID:IS	OS:IS	IS:OS	ID:OS	OS:OS	Input	Output	
G	G	G	G	G	G	G	G	G	G	G	G	0
C	G	E	G	G	G	G	G	G	G	G	G	1
G	G	E	G	G	G	G	G	G	G	G	G	2
G	G	E	Br	B	G	G	G	G	G	G	G	3
C	G	E	Br	In	B	O	G	G	G	G	G	4
G	G	G	Br	In	B	O	G	G	G	Im	G	5
G	G	G	G	In	G	G	G	G	G	G	Ks	6
G	G	G	G	G	B	O	G	G	G	Im	G	7

[1] S: input shaft; :IS: input side; ID: idler shaft; OS: output shaft; :OS: output side.
[2] G: good; C: chipped; E: eccentric; Br: broken; B: ball; In: inner race; O: outer race; Im: imbalance; Ks: keyway sheared.

3.2. Fault experiments of gearbox

The research presented in this study focuses on the important task of gearbox fault diagnosis, which is crucial for ensuring the safe and efficient operation of wind turbines [40]. To further enhance the significance of the research, a case study on gearbox fault diagnosis is incorporated using the open access gearbox fault dataset from the international Prognostics and Health Management Society. This case study provides valuable insights into the effectiveness of the proposed semi-supervised approach in addressing gearbox-related issues in wind turbines.

The experiments consider eight different health conditions, as listed in Table 3, with detailed fault descriptions available in [41,42]. The experiments are conducted under varying rotating speeds, including 30 Hz, 40 Hz, and 50 Hz for the seen working condition, and 35 Hz and 45 Hz for the unseen working condition. Each sample in the dataset consists of 4096 sampling points. Six semi-supervised diagnosis experiments are performed, employing different numbers of annotated samples (1000, 500, 200, 90, 60, and 30) to evaluate the performance and robustness of the proposed method. Additionally, 2000 unannotated samples are utilized during the semi-supervised learning process. By benchmarking against the gearbox fault dataset, the effectiveness of the developed approach can be compared, contributing to advancements in the field of wind turbine fault diagnosis and ultimately improving the reliability and efficiency of wind turbine systems.

3.3. Comparative studies

The selected comparative methods can be categorized into two classes: a) supervised methods, including backboneNet, backboneNet-D and semi-supervised methods, including label propagation (LP) [25], self-training (ST) [43], mean teacher (MT) [44], FixMatch [45], maximum mean discrepancy (MMD) [46], correlation alignment (CORAL) [47] and backboneNet-A [39].

In the realm of supervised learning methods, the backboneNet represents the backbone network of the proposed SSAD method, which consists of the deep feature extractor $F(\cdot,\theta_f)$ and diagnostic classifier $C(\cdot,\theta_c)$. Additionally, we integrate the backbone network with triplet loss optimization to introduce BackboneNet-D, which aims to evaluate the effectiveness of utilizing discriminative features. Regarding semi-supervised learning methods, comparative approaches including LP, ST, MT, FixMatch, MMD, CORAL and BackboneNet-A, are utilized to demonstrate the superiority of the proposed SSAD. LP and ST are widely used semi-supervised learning techniques, wherein the support vector machine (SVM) is adopted as the fundamental classifier in their training process. LP leverages label propagation from labeled to unlabeled data based on label consistency assumptions, while ST incorporates confident predictions from unlabeled data into the labeled dataset for retraining. Shallow classifiers like SVM are often integrated with LP and ST. Moreover, MT and FixMatch are acknowledged as sophisticated and widely recognized semi-supervised learning techniques, where MT introduces an average teacher model to improve performance through label smoothing, while FixMatch combines labeled and unlabeled data using weak labels and a consistency loss function to enhance model performance. To tackle the distribution alignment between annotated and unannotated data, two well-established deep domain adaptation methods, MMD and CORAL, are also incorporated as comparative approaches in the semi-supervised setting. Furthermore, we investigate the integration of the backbone network with adversarial learning, resulting in the development of the BackboneNet-A approach. BackboneNet-A is trained using both annotated and unannotated data, thereby demonstrating the effectiveness of semi-supervised learning.

The influence of adversarial learning and metric learning in this study is analyzed through ablation experiments, which involve comparisons among the BackboneNet, BackboneNet-D, BackboneNet-A, and SSAD. The proposed SSAD network adopts the remaining hyperparameters settings, where the learning rate is 0.01, the trade-off parameter β is 0.05. It should be noted that the learning rates are selected through optimization experiments, while the selection principle and the optimal range of the trade-off parameter β are elucidated in Section 4.3. Comprehensive experimental details are listed in Table 4. To evaluate the diagnostic performance of different methods comprehensively, three indexes, namely accuracy, F1-score, and recall, are selected for analysis.

Table 4
The descriptions of experiments settings.

Methods	No. of labeled training samples	No. of unlabeled training samples	Training category	No. of testing samples in seen/unseen working conditions	No. of trials
LP	1000/500/200/90/60/30	2000	Semi-supervised	1000/1000	10
ST	1000/500/200/90/60/30	2000	Semi-supervised	1000/1000	10
MT	1000/500/200/90/60/30	2000	Semi-supervised	1000/1000	10
FixMatch	1000/500/200/90/60/30	2000	Semi-supervised	1000/1000	10
MMD	1000/500/200/90/60/30	2000	Semi-supervised	1000/1000	10
CORAL	1000/500/200/90/60/30	2000	Semi-supervised	1000/1000	10
BackboneNet	1000/500/200/90/60/30	0	Supervised	1000/1000	10
BackboneNet-D	1000/500/200/90/60/30	0	Supervised	1000/1000	10
BackboneNet-A	1000/500/200/90/60/30	2000	Semi-supervised	1000/1000	10
Proposed SSAD	1000/500/200/90/60/30	2000	Semi-supervised	1000/1000	10

Table 5
The comparisons of accuracy (%) for different methods in case 1.

Methods	No. of labeled training samples					
	1000[a]	500	200	90	60	30
LP	95.5/89.9[b]	93.2/88.4	91.7/87.5	88.5/86.0	82.4/84.6	72.0/74.2
ST	94.1/87.0	91.0/85.6	89.8/85.7	85.2/79.5	79.0/79.0	69.0/70.0
MT	99.3/97.1	99.0/86.8	95.0/86.7	89.0/83.3	76.3/67.3	52.2/40.1
FixMatch	99.8/95.6	99.1/91.8	98.2/91.3	95.6/80.6	72.7/58.8	58.9/53.1
MMD	98.7/98.4	98.8/97.4	92.7/86.2	62.1/56.4	56.8/48.6	50.2/47.0
CORAL	97.7/94.2	97.3/92.4	86.2/77.4	61.1/55.3	57.1/47.8	52.7/49.0
BackboneNet	98.6/94.3	96.3/87.2	86.7/75.9	61.0/54.2	47.5/40.0	36.8/35.4
BackboneNet-D	100.0/99.9	99.9/99.5	99.8/99.5	93.1/83.4	73.9/64.6	47.9/45.3
BackboneNet-A	100.0/99.9	100.0/99.9	100.0/99.8	96.6/96.1	63.8/62.2	42.1/40.5
SSAD	100.0/100.0	100.0/100.0	100.0/100.0	99.9/99.8	99.9/99.7	93.1/91.7

[a] represents the number of labeled samples used during model training.
[b] the results for the testing samples from the known and unknown wind speeds, respectively.

4. Results analysis

4.1. Comparisons of overall diagnosis results

4.1.1. Case 1: diagnostic results of rotor systems faults in wind turbine

Tables 5–7 present the average accuracy, F1-score, and recall of ten trials for the six experiments. When utilizing 1000 or 500 labeled samples, most of the methods achieve average diagnostic accuracies exceeding 90%. Notably, the proposed SSAD method achieves perfect identification of all health conditions. Similar trends can be observed for the F1-score and recall in Tables 6 and 7. However, when the number of annotated samples used for training is limited, such as with only 30 annotated samples, the diagnostic performance of comparative methods rapidly deteriorates, as indicated in Table 5. For instance, the performance of MMD and CORAL is greatly diminished, resulting in an average diagnostic accuracy of only around 50%. This outcome can be attributed to the fundamental principle of MMD and CORAL, which focuses on aligning the distributions between annotated and unannotated data. Consequently, the extremely scarce labeled samples fail to accurately represent the true data distribution, thereby significantly reducing the effectiveness of distribution alignment techniques.

It is important to highlight that the accuracies of the MT, FixMatch, backboneNet, backboneNet-D, and backboneNet-A methods exhibit a significant decline as the number of labeled samples decreases. Deep learning methods are particularly susceptible to overfitting in scenarios involving extremely limited labeled samples. Specifically, the average diagnostic accuracy of backboneNet reaches only 36.8% when utilizing 30 labeled samples. In comparison, backboneNet-D and backboneNet-A exhibit improvements of approximately 11% and 5% in diagnostic accuracy, respectively. By contrast, the proposed SSAD method leverages discriminative feature enhancement and semi-supervised adversarial learning to achieve both generalized feature representation and discriminative decision boundaries across different manifolds. This enables correct utilization of the information contained in the extensive set of unlabeled samples. Consequently, the SSAD method maintains an accuracy level of over 90% even under extremely limited labeled sample conditions.

To further investigate the influence of the number of labeled samples on model performance, Figs. 6–8 illustrate the curves of diagnostic accuracy, F1-score, and recall as a function of the labeled training sample count, accompanied by their corresponding confidence intervals represented by standard deviation intervals. In general, all models exhibit a declining trend in diagnostic performance as the number of labeled samples decreases. It is noteworthy that the comparative methods are particularly sensitive to the number of labeled samples. With a reduction in the number of labeled training samples, a significant degradation in performance can be observed across all nine comparative methods, particularly for deep models such as backboneNet, backboneNet-D, MMD, and CORAL. In contrast, the proposed SSAD method in this study consistently maintains an average diagnostic accuracy exceeding 90%

Table 6
The comparisons of F1-score (%) for different methods in case 1.

Methods	No. of labeled training samples					
	1000	500	200	90	60	30
LP	95.4/89.9	93.5/88.7	92.1/87.9	88.7/85.8	83.0/84.8	72.4/74.8
ST	93.9/86.6	91.4/85.9	89.9/85.5	85.3/79.6	79.4/79.8	69.2/70.2
MT	99.3/98.6	99.0/85.8	95.1/86.7	88.8/82.9	82.5/75.3	49.8/37.8
FixMatch	99.8/95.8	99.2/91.5	98.2/91.0	95.6/79.5	72.7/56.6	58.8/53.6
MMD	98.4/98.9	98.8/97.3	95.5/86.7	70.1/66.7	67.6/67.2	62.6/58.9
CORAL	98.2/95.4	99.0/95.1	84.8/75.3	65.3/62.6	69.6/65.8	62.4/57.3
BackboneNet	99.8/97.4	99.2/96.9	93.8/91.9	74.4/74.9	62.6/63.2	61.1/62.3
BackboneNet-D	100.0/100.0	100.0/99.8	99.9/99.4	94.3/93.1	82.1/81.2	66.5/68.7
BackboneNet-A	100.0/100.0	100.0/100.0	100.0/100.0	94.0/94.6	71.5/72.7	61.5/62.3
SSAD	100.0/100.0	100.0/100.0	100.0/100.0	100.0/100.0	99.9/99.9	93.4/93.4

Table 7
The comparisons of recall (%) for different methods in case 1.

Methods	No. of labeled training samples					
	1000	500	200	90	60	30
LP	95.4/89.0	93.0/86.5	91.2/85.0	87.6/85.0	80.3/82.4	68.3/69.9
ST	93.8/85.3	90.4/84.2	89.0/76.8	84.2/76.8	77.3/76.8	66.1/66.9
MT	99.5/98.2	99.0/87.5	95.1/87.2	88.7/83.2	83.9/77.0	52.7/42.7
FixMatch	99.8/95.8	99.1/92.2	98.2/91.6	95.6/80.2	73.1/60.6	57.8/54.9
MMD	98.6/99.0	98.9/98.2	94.1/87.9	69.3/67.0	68.4/68.4	62.7/59.0
CORAL	98.2/95.5	99.2/95.6	82.6/75.7	64.1/62.7	71.2/67.9	61.5/56.8
BackboneNet	99.8/96.2	99.2/96.2	94.3/90.8	74.7/73.3	62.9/62.7	61.8/62.5
BackboneNet-D	100.0/100.0	100.0/99.7	99.9/99.1	93.8/92.0	82.0/79.3	67.4/68.5
BackboneNet-A	100.0/100.0	100.0/100.0	100.0/100.0	93.9/94.6	71.3/72.3	61.2/61.4
SSAD	100.0/100.0	100.0/100.0	100.0/100.0	100.0/100.0	99.9/99.9	93.3/93.2

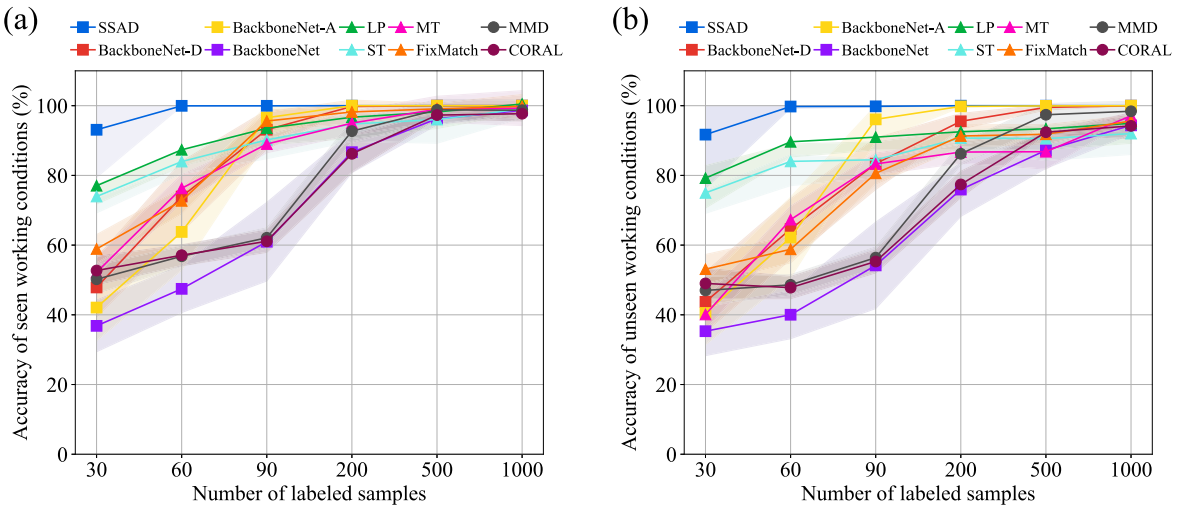

Fig. 6. The diagnostic accuracies versus the number of annotated training samples in case 1, (a) seen working condition, (b) unseen working condition.

even when the number of annotated samples falls below 60. By effectively combining discriminative features and semi-supervised adversarial learning, the SSAD method mitigates the performance degradation resulting from inadequate supervised information.

4.1.2. Case 2: diagnostic results of gearbox faults

The diagnostic results of gearbox faults presented in Tables 8–10 exhibit similar conclusions, aligning with the observations made in Case 1. When utilizing 1000 labeled samples, most of the semi-supervised methods achieve diagnostic accuracies averaging above 90%. However, the performance of comparative methods rapidly deteriorates when the number of annotated samples is limited, such as with 500 and 200 samples. In the conducted ablation experiments, it can be also observed that backboneNet, backboneNet-D, and backboneNet-A are highly sensitive to the number of labeled samples, displaying significant performance degradation when

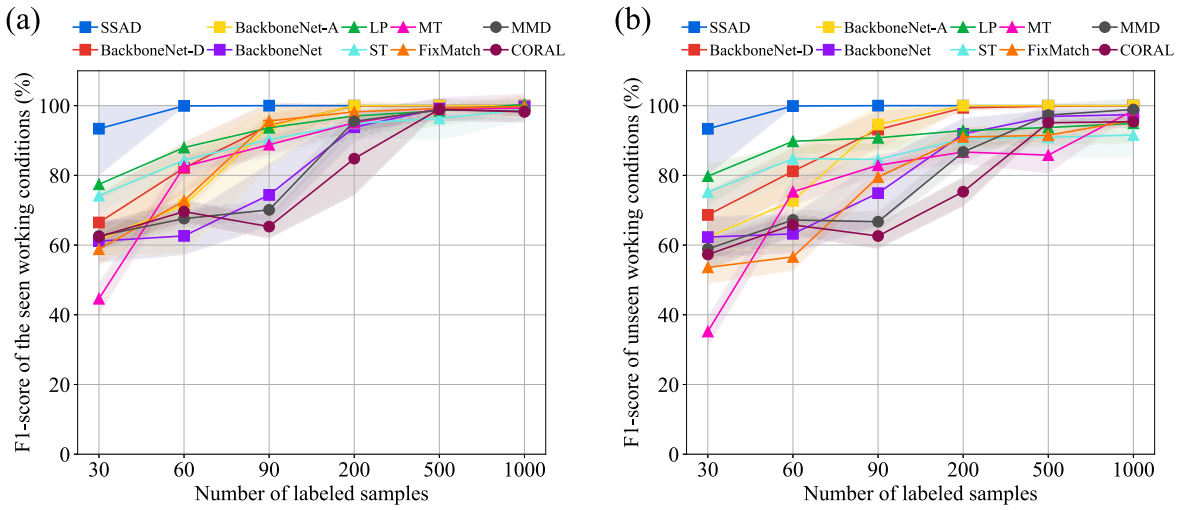

Fig. 7. The F1-scores versus the number of annotated training samples in case 1, (a) seen working condition, (b) unseen working condition.

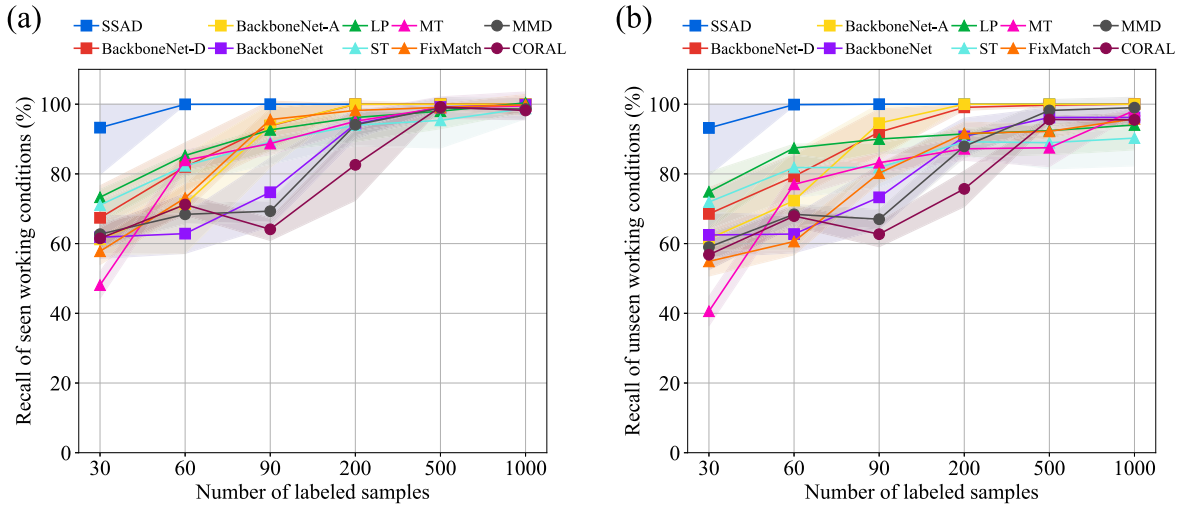

Fig. 8. The recalls versus the number of annotated training samples in case 1, (a) seen working condition, (b) unseen working condition.

the labeled samples decrease below 200. In contrast, the proposed SSAD method demonstrates superior performance under such constraints of limited labeled samples.

This observation underscores the crucial role played by labeled samples in training models in a semi-supervised learning setting. Labeled samples provide explicit supervision and guide the learning process, enabling the model to accurately capture the underlying patterns and make informed predictions. Insufficient labeled samples can lead to a lack of representative information, impeding the model's ability to generalize effectively and resulting in diminished diagnostic accuracy. In this context, the SSAD method effectively mitigates the adverse effects of limited labeled samples, showcasing its robustness and suitability for such scenarios.

4.2. Analysis of detailed diagnosis results in ten trials and different working speeds

Due to space limitations, the subsequent results analysis focuses specifically on the diagnostic results of rotor system faults in wind turbines, as discussed in case 1. The results of diagnostic accuracy in 10 trials under seen working conditions are further presented. Taking the experiments of 60 and 30 labeled training samples as examples, the results are illustrated in Fig. 9. Since the training and testing samples are chosen randomly in each trial, the differences of diagnostic performance may exist in different trials. Fig. 9(a) demonstrates that when only 60 labeled training samples are available, the proposed SSAD method exhibits consistently high diagnostic performance across 10 trials. Even with the extremely limited labeled training samples, the SSAD method could still achieve the highest accuracy among all the methods, and maintain an average accuracy of more than 90%, showing a good robustness characteristic.

Table 8
The comparisons of accuracy (%) for different methods in case 2.

Methods	No. of labeled training samples					
	1000	500	200	90	60	30
LP	82.5/67.1	78.2/63.0	58.7/50.5	46.6/39.4	42.8/35.1	25.5/27.5
ST	90.1/74.1	77.6/67.0	64.6/59.7	45.1/42.4	45.4/42.6	28.2/25.8
MT	96.2/79.5	93.0/78.3	75.1/61.6	45.2/43.2	37.5/37.7	22.1/20.7
FixMatch	97.2/79.8	95.6/76.6	73.8/63.3	48.9/45.9	39.5/36.4	30.0/26.0
MMD	95.5/81.9	82.6/71.0	55.8/46.5	45.2/35.1	36.9/34.3	26.1/24.6
CORAL	91.5/73.3	81.1/73.0	61.0/52.3	40.8/34.0	40.6/36.2	25.5/25.2
BackboneNet	87.3/69.6	78.7/60.9	54.0/51.0	37.8/32.5	30.6/26.8	28.3/25.3
BackboneNet-D	99.1/80.8	96.6/72.5	67.1/50.4	42.2/40.8	41.7/38.1	31.1/28.4
BackboneNet-A	98.6/83.8	96.3/79.5	60.2/51.8	41.4/38.2	41.2/36.3	28.3/32.0
SSAD	99.9/88.2	99.4/80.3	79.6/61.5	49.5/48.6	44.0/33.7	37.2/31.9

Table 9
The comparisons of F1-score (%) for different methods in case 2.

Methods	No. of labeled training samples					
	1000	500	200	90	60	30
LP	81.8/66.0	78.1/62.5	58.9/49.7	46.6/39.2	42.5/32.8	25.3/27.6
ST	90.3/73.1	77.9/65.8	64.6/58.8	45.1/41.3	45.6/41.5	27.5/25.0
MT	96.3/78.8	93.2/78.1	75.6/62.1	43.8/42.3	37.5/35.7	21.3/19.4
FixMatch	97.1/78.9	95.6/75.8	73.6/62.2	49.3/45.9	39.0/36.3	28.2/23.4
MMD	95.5/81.1	82.6/70.2	55.4/45.5	43.5/32.6	35.1/31.0	25.6/24.1
CORAL	91.4/72.0	81.2/71.3	60.6/50.3	40.1/33.4	38.9/34.7	21.3/20.2
BackboneNet	87.6/69.6	79.1/60.9	54.1/51.0	37.2/31.5	29.1/25.2	28.3/24.0
BackboneNet-D	99.0/80.9	96.5/71.1	65.2/47.3	41.0/39.6	41.3/37.1	31.7/26.8
BackboneNet-A	98.6/83.2	96.3/78.8	60.2/50.0	41.2/38.3	40.8/35.8	27.8/32.1
SSAD	99.9/88.0	99.4/78.7	79.4/62.1	48.1/45.1	43.6/33.6	37.4/30.1

Table 10
The comparisons of recall (%) for different methods in case 2.

Methods	No. of labeled training samples					
	1000	500	200	90	60	30
LP	82.5/66.8	78.2/62.8	59.4/50.0	46.6/39.5	43.1/34.9	25.2/27.0
ST	90.1/73.7	77.7/66.6	64.3/59.6	44.9/41.5	45.2/41.3	28.0/24.8
MT	96.3/79.0	93.3/78.0	75.3/61.7	44.8/43.2	37.5/37.6	22.5/20.8
FixMatch	97.1/78.9	95.6/75.6	73.6/62.3	49.0/45.8	39.5/36.0	29.9/24.5
MMD	95.5/81.0	82.5/70.4	55.2/46.2	45.0/34.0	36.3/33.4	25.8/25.3
CORAL	91.5/71.8	81.2/72.5	60.6/51.7	39.8/33.4	39.7/35.2	25.1/24.4
BackboneNet	87.5/70.1	79.0/61.4	54.1/51.3	38.0/32.7	30.6/27.0	28.2/25.2
BackboneNet-D	99.1/80.5	96.5/72.5	66.2/50.5	41.7/40.7	41.7/37.8	31.3/28.3
BackboneNet-A	98.6/84.2	96.3/79.4	60.3/51.8	41.9/38.4	41.3/36.7	28.9/32.9
SSAD	99.9/88.2	99.4/80.3	79.6/61.8	48.6/45.2	44.1/34.1	37.6/31.6

In addition to analyzing the effects of randomized trials, data at different wind speeds may also influence training and test results. Therefore, We further used four sets of data from different wind speeds for training, and tested on the data with speed 6.9 m/s to verify the robustness of the SSAD network. Fig. 10 shows the results. The S1-S5 denotes the wind speed of 11.5 m/s, 12.7 m/s, 10.4 m/s, 9.2 m/s and 8.1 m/s, respectively. For each wind speed, we conducted 10 random randomized trials. All labeled samples are set to 30. Fig. 10(a) and Fig. 10(b) present the diagnostic performance under seen (speed of S1-S5) and unseen (speed of 6.9 m/s) working conditions, respectively. The proposed SSAD network outperforms the other 9 methods in terms of diagnostic performance on average, which further verifies its robustness under different wind speeds.

4.3. Effect of the trade-off parameter on the diagnostic performance

In this work, the triple loss and cross-entropy loss are combined for the optimization process during the supervised training. Therefore, it is necessary to discuss the effect of the trade-off parameter β on the proposed semi-supervised learning framework. Fig. 11 shows the average diagnostic accuracy, the F1-score, and the Recall indicators under the β value from 0.01 to 0.1. Only 30 labeled samples are used here for further verifying the diagnostic performance under both insufficient labeled samples and different β values. Moreover, under each β value, 10 randomized trials were conducted to eliminate the influence of data quality.

As shown in the figure, the blue curve shows the diagnostic results for the seen conditions. The average diagnostic accuracy is higher than 90%, while the F1-score and the recall value are higher than 85%, showing a satisfying diagnostic performance. The

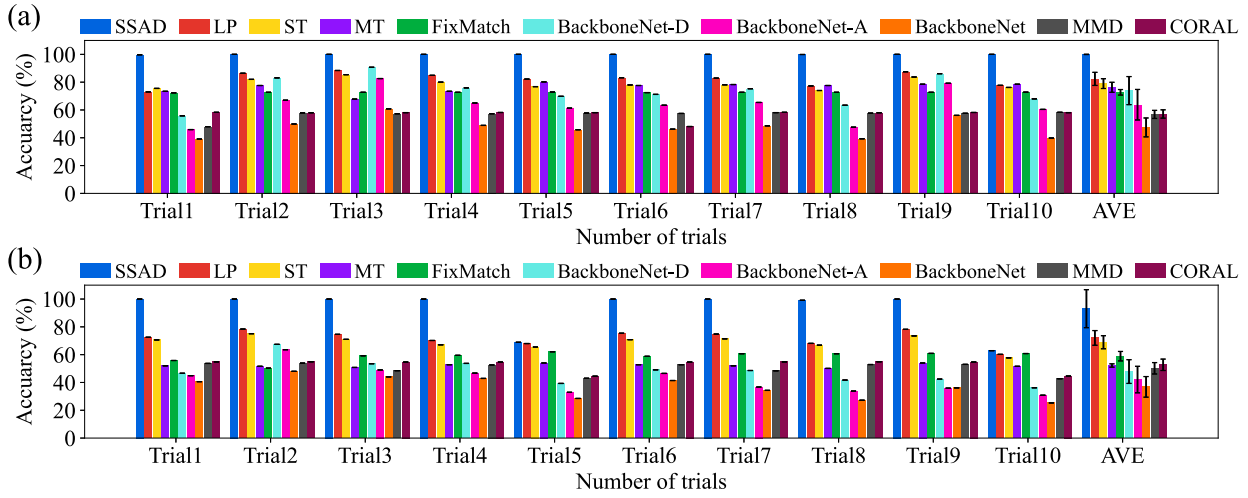

Fig. 9. The results of diagnostic accuracy under 10 trials, (a) 60 labeled training samples, (b) 30 labeled training samples.

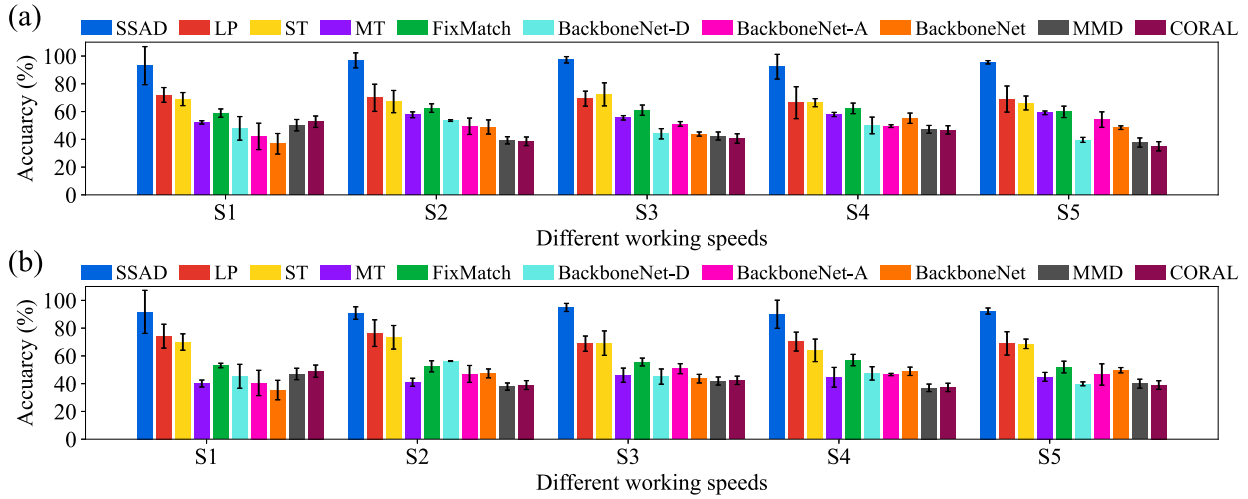

Fig. 10. The results of diagnostic accuracy under 5 different wind speeds, (a) diagnostic accuracy under seen working condition, (b) diagnostic accuracy under unseen working condition.

red curve shows the diagnostic results for the unseen conditions. The average accuracy exceeds 90%, while the F1-score and the recall value are higher than 85%, which are similar to the results under the seen conditions. Therefore, the proposed model has a strong generalization ability and the β value selected in Section 2.4 is reasonable. However, it is also worth noting that different distributions of data will lead to different diagnostic complexity and difficulty, so the selection of beta should be based on specific cases.

4.4. Feature visualization in ablation experiments

To further illustrate the mechanism of the semi-supervised adversarial learning and discriminative features enhancement during the training process, t-SNE is utilized to visualize features of $F(\cdot,\theta_f)$, so as to obtain the feature distribution of annotated and unannotated samples in two-dimensional space.

Fig. 12 shows the feature distribution obtained by the backboneNet, backboneNet-D, backboneNet-A and SSAD methods under 1000 labeled training samples, while the corresponding confusion matrices for the testing samples are presented in Fig. 13. Compared with the backboneNet, the backboneNet-D method adopts discriminative features enhancement, which effectively reduces the intra-category distance and increases the inter-category distance. As shown in Fig. 12, the feature distributions between the categories F3 and F4 appear slight partial aliasing for backboneNet, which may lead to inaccurate diagnosis. On the contrary, the feature distributions obtained by the backboneNet-D are more discriminative. For the backboneNet-A method, semi-supervised adversarial learning can leverage both labeled and unlabeled samples to achieve distribution alignment and obtain a more generalized feature

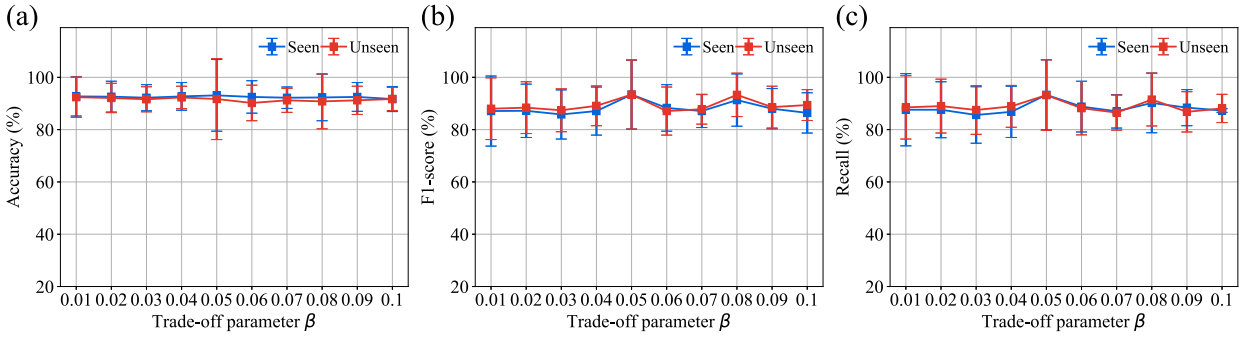

Fig. 11. The average diagnostic accuracy, the F1-score, and the Recall indicators under different β values.

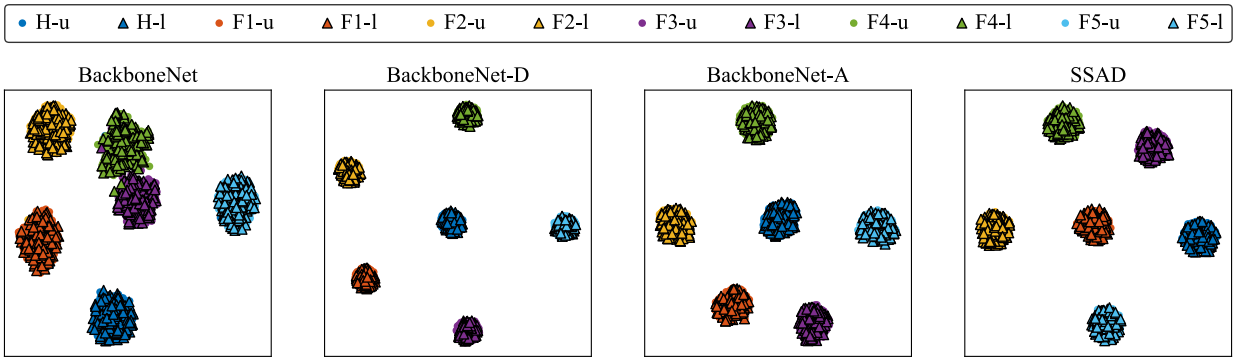

Fig. 12. Feature distributions obtained by the backboneNet, backboneNet-D, backboneNet-A and SSAD methods under 1000 labeled training samples. "-u" represents the unlabeled samples while "-l" represents the labeled samples.

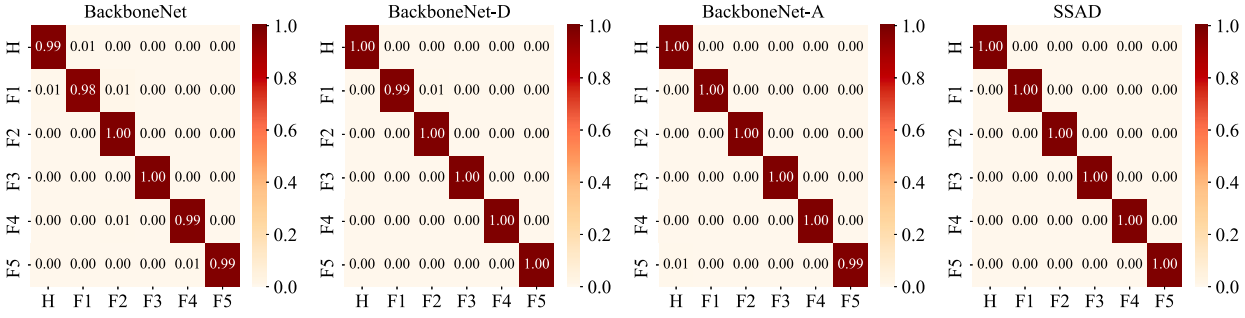

Fig. 13. Confusion matrices obtained by the backboneNet, backboneNet-D, backboneNet-A and SSAD methods under 1000 labeled training samples.

representation, thus obtaining inherent features of the same manifold and reducing the risk of overfitting. As shown in Fig. 12, the semi-supervised adversarial learning can obtain more separable feature distributions compared with the backboneNet.

However, it is also worth emphasizing that gaps appear between the performance of these four methods when there are not sufficient annotated samples for training. Fig. 14 shows the feature distribution obtained under 30 labeled training samples, while the corresponding confusion matrices are presented in Fig. 15. The feature distributions obtained by the backboneNet appear massive overlapping and the diagnostic accuracy is all reduced to less than 65% for each health condition. Although compared with the backboneNet, the backboneNet-D can still increase the inter-category distance to a certain extent, a satisfactory decision boundary is hard to obtain at this time. According to the confusion matrix obtained by the backboneNet-D, the accuracy for part of categories is reduced to about 20%. Understandably, the extremely limited labeled training samples may not reflect the manifold characteristics of the entire health category, and the supervised trained model is easily overfitting.

For the backboneNet-A, the diagnostic performance with 30 labeled samples is also unsatisfactory. The visualization in Fig. 14 indicates that there is overlapping in the feature distributions of the categories H, F1 and F5, leading to suboptimal identification of these three categories by the model. Same phenomenon can be observed for F2 and F4. Thus, it can be found that the effect of adversarial learning is limited to the alignment of the marginal feature distributions between annotated and unannotated samples, while ignores the conditional distribution for accurately diagnosis. Insufficient labeled samples for capturing accurate discriminative

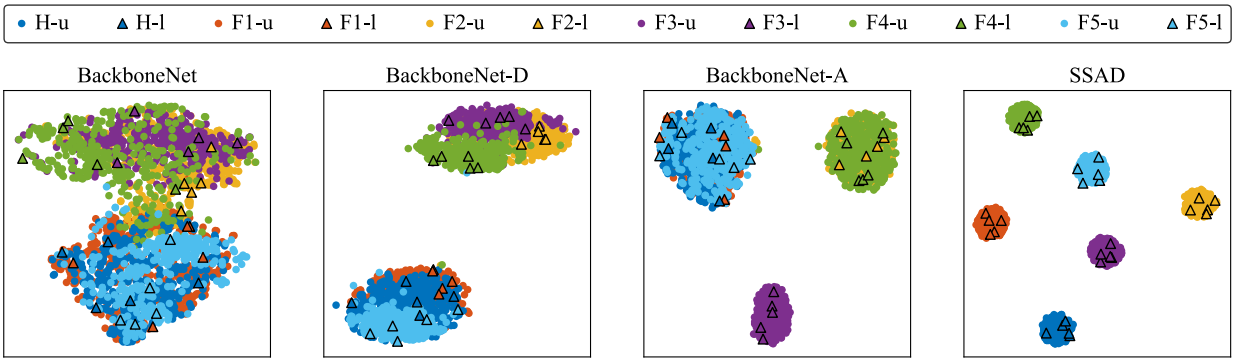

Fig. 14. Feature distributions obtained by the backboneNet, backboneNet-D, backboneNet-A and SSAD methods under 30 labeled training samples. "-u" represents the unlabeled samples while "-l" represents the labeled samples.

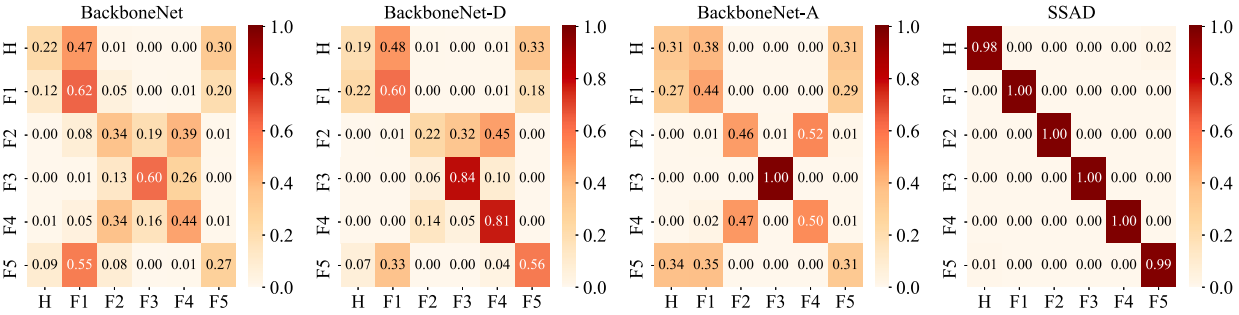

Fig. 15. Confusion matrices obtained by the backboneNet, backboneNet-D, backboneNet-A and SSAD methods under 30 labeled training samples.

features of most unlabeled samples from various categories can result in adversarial learning misaligning feature distributions. This, in turn, could lead to a significant increase in the model's misclassification rate.

When the number of annotated samples decreases from 1000 to 30, the SSAD method exhibits satisfactory discriminative feature distributions and achieves over 90% diagnostic accuracy, thereby demonstrating robustness and excellent diagnostic performance. Specifically, the utilization of semi-supervised training through adversarial learning effectively leverages the feature information contained in unlabeled samples, thus mitigating the degradation in performance resulting from the scarcity of labeled samples. Furthermore, the incorporation of triplet loss optimization enhances the discriminative nature of features belonging to different categories and effectively mitigates feature misalignment during adversarial learning. By jointly considering the benefits of triplet loss and adversarial learning, the SSAD network demonstrates superior diagnostic performance, even when confronted with an exceedingly limited number of labeled training samples.

5. Conclusions

The study presents a novel approach for wind turbine fault diagnosis in situations where labeled training data is limited. The proposed SSAD network incorporates unlabeled samples to learn a more generalized feature representation with adversarial training. Additionally, the network uses triplet loss optimization to increase the discriminability of features across different health categories. By combining these techniques, the SSAD network achieves superior diagnostic performance in semi-supervised learning scenarios. The experimental results proved the efficacy of the proposed SSAD network, maintaining more than 90% accuracy even when the annotated samples are extremely limited. Overall, the proposed SSAD network presents a promising tool for wind turbine fault diagnosis with limited supervision.

The utilization of real-world wind turbine data for validation is crucial to assess the performance and practicality of the proposed method. Collaborations with industry partners are planned to access and analyze such data, aiming to validate the effectiveness of our method in real-world scenarios. The ultimate goal is to integrate our approach into existing fault diagnosis systems, contributing to the field of wind turbine maintenance and fault diagnosis.

CRediT authorship contribution statement

Te Han: Conceptualization, Data curation, Methodology, Resources, Supervision, Validation, Writing – review & editing. **Wenzhen Xie:** Formal analysis, Investigation, Software, Validation, Writing – original draft. **Zhongyi Pei:** Funding acquisition, Project administration, Software, Validation, Writing – review & editing.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (Grant No. 72201152) and the National Key Research and Development Project (Grant No. 2021YFB1715200).

References

[1] H.D.M. de Azevedo, A.M. Araújo, N. Bouchonneau, A review of wind turbine bearing condition monitoring: state of the art and challenges, Renew. Sustain. Energy Rev. 56 (2016) 368–379.
[2] H. Pan, H. Xu, J. Zheng, J. Tong, Non-parallel bounded support matrix machine and its application in roller bearing fault diagnosis, Inf. Sci. 624 (2023) 395–415.
[3] O. García Peyrano, J. Vignolo, R. Mayer, M. Marticorena, Online unbalance detection and diagnosis on large flexible rotors by SVR and ANN trained by dynamic multibody simulations, J. Dyn. Monit. Diagn. 1 (3) (2022) 139–147.
[4] T. de Oliveira Nogueira, G.B.A. Palacio, F.D. Braga, P.P.N. Maia, E.P. de Moura, C.F. de Andrade, P.A.C. Rocha, Imbalance classification in a scaled-down wind turbine using radial basis function kernel and support vector machines, Energy 238 (2022) 122064.
[5] J. Yu, Y. Rui, D. Tao, Click prediction for web image reranking using multimodal sparse coding, IEEE Trans. Image Process. 23 (5) (2014) 2019–2032.
[6] H. Pan, H. Xu, J. Zheng, J. Su, J. Tong, Multi-class fuzzy support matrix machine for classification in roller bearing fault diagnosis, Adv. Eng. Inform. 51 (2022) 101445.
[7] J. Jiao, M. Zhao, J. Lin, K. Liang, A comprehensive review on convolutional neural network in machine fault diagnosis, Neurocomputing 417 (2020) 36–63.
[8] J. Zhang, Y. Cao, Q. Wu, Vector of locally and adaptively aggregated descriptors for image feature representation, Pattern Recognit. 116 (2021) 107952.
[9] J. Yu, M. Tan, H. Zhang, Y. Rui, D. Tao, Hierarchical deep click feature prediction for fine-grained image recognition, IEEE Trans. Pattern Anal. Mach. Intell. 44 (2) (2019) 563–578.
[10] C. Zhang, D. Hu, T. Yang, Anomaly detection and diagnosis for wind turbines using long short-term memory-based stacked denoising autoencoders and xgboost, Reliab. Eng. Syst. Saf. 222 (2022) 108445.
[11] T. Han, Y.-F. Li, Out-of-distribution detection-assisted trustworthy machinery fault diagnosis approach with uncertainty-aware deep ensembles, Reliab. Eng. Syst. Saf. 226 (2022) 108648.
[12] L. Zhang, Q. Fan, J. Lin, Z. Zhang, X. Yan, C. Li, A nearly end-to-end deep learning approach to fault diagnosis of wind turbine gearboxes under nonstationary conditions, Eng. Appl. Artif. Intell. 119 (2023) 105735.
[13] R. Rahimilarki, Z. Gao, N. Jin, A. Zhang, Convolutional neural network fault classification based on time-series analysis for benchmark wind turbine machine, Renew. Energy 185 (2022) 916–931.
[14] Y. Qin, X. Wang, J. Zou, The optimized deep belief networks with improved logistic sigmoid units and their application in fault diagnosis for planetary gearboxes of wind turbines, IEEE Trans. Ind. Electron. 66 (5) (2019) 3814–3824.
[15] A. Saleh, M. Chiachío, J.F. Salas, A. Kolios, Self-adaptive optimized maintenance of offshore wind turbines by intelligent Petri nets, Reliab. Eng. Syst. Saf. 231 (2023) 109013.
[16] J. Yao, T. Han, Data-driven lithium-ion batteries capacity estimation based on deep transfer learning using partial segment of charging/discharging data, Energy 271 (2023) 127033.
[17] B. Li, J. Wang, Z. Yang, J. Yi, F. Nie, Fast semi-supervised self-training algorithm based on data editing, Inf. Sci. 626 (2023) 293–314.
[18] J. Zhang, J. Yang, J. Yu, J. Fan, Semisupervised image classification by mutual learning of multiple self-supervised models, Int. J. Intell. Syst. 37 (5) (2022) 3117–3141.
[19] M. Yan, S.C. Hui, N. Li, Dml-pl: deep metric learning based pseudo-labeling framework for class imbalanced semi-supervised learning, Inf. Sci. 626 (2023) 641–657.
[20] H. Gammulle, S. Denman, S. Sridharan, C. Fookes, Fine-grained action segmentation using the semi-supervised action gan, Pattern Recognit. 98 (2020) 107039.
[21] Y. Feng, J. Chen, T. Zhang, S. He, E. Xu, Z. Zhou, Semi-supervised meta-learning networks with squeeze-and-excitation attention for few-shot fault diagnosis, ISA Trans. 120 (2022) 383–401.
[22] Z. Wang, J. Xuan, T. Shi, A novel semi-supervised generative adversarial network based on the actor-critic algorithm for compound fault recognition, Neural Comput. Appl. (2022) 1–19.
[23] M. Zheng, J. Man, D. Wang, Y. Chen, Q. Li, Y. Liu, Semi-supervised multivariate time series anomaly detection for wind turbines using generator scada data, Reliab. Eng. Syst. Saf. 235 (2023) 109235.
[24] J. Zhuang, M. Jia, Y. Cao, X. Zhao, Semi-supervised double attention guided assessment approach for remaining useful life of rotating machinery, Reliab. Eng. Syst. Saf. 226 (2022) 108685.
[25] S. Chen, R. Yang, M. Zhong, Graph-based semi-supervised random forest for rotating machinery gearbox fault diagnosis, Control Eng. Pract. 117 (2021) 104952.
[26] S. Zhang, F. Ye, B. Wang, T.G. Habetler, Semi-supervised bearing fault diagnosis and classification using variational autoencoder-based deep generative models, IEEE Sens. J. 21 (5) (2021) 6476–6486.
[27] M. Moradi, A. Broer, J. Chiachío, R. Benedictus, T.H. Loutas, D. Zarouchas, Intelligent health indicator construction for prognostics of composite structures utilizing a semi-supervised deep neural network and SHM data, Eng. Appl. Artif. Intell. 117 (2023) 105502.
[28] K. Zhou, E. Diehl, J. Tang, Deep convolutional generative adversarial network with semi-supervised learning enabled physics elucidation for extended gear fault diagnosis under data limitations, Mech. Syst. Signal Process. 185 (2023) 109772.
[29] Q. Wang, W. Li, L.V. Gool, Semi-supervised learning by augmented distribution alignment, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 1466–1475.
[30] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, V. Lempitsky, Domain-adversarial training of neural networks, J. Mach. Learn. Res. 17 (1) (2016) 1–35.
[31] X. Yu, Z. Zhao, X. Zhang, C. Sun, B. Gong, R. Yan, X. Chen, Conditional adversarial domain adaptation with discrimination embedding for locomotive fault diagnosis, IEEE Trans. Instrum. Meas. 70 (2021) 1–12.

[32] C. Si, X. Nie, W. Wang, L. Wang, T. Tan, J. Feng, Adversarial self-supervised learning for semi-supervised 3D action recognition, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII 16, Springer, 2020, pp. 35–51.
[33] C. Mayer, M. Paul, R. Timofte, Adversarial feature distribution alignment for semi-supervised learning, Comput. Vis. Image Underst. 202 (2021) 103109.
[34] J.E. Van Engelen, H.H. Hoos, A survey on semi-supervised learning, Mach. Learn. 109 (2) (2020) 373–440.
[35] C. Wang, C. Xin, Z. Xu, A novel deep metric learning model for imbalanced fault diagnosis and toward open-set classification, Knowl.-Based Syst. 220 (2021) 106925.
[36] J. Wu, C. Sun, C. Zhang, X. Chen, R. Yan, Deep clustering variational network for helicopter regime recognition in HUMS, Aerosp. Sci. Technol. 124 (2022) 107553.
[37] H. Wang, Z. Liu, T. Ai, Long-range dependencies learning based on non-local 1d-convolutional neural network for rolling bearing fault diagnosis, J. Dyn. Monit. Diagn. (2022), https://doi.org/10.37965/jdmd.2022.53.
[38] Z. Xu, K. Zhao, T. Zhang, C. Fu, M. Yan, Z. Xie, X. Zhang, G. Catolino, Effort-aware just-in-time bug prediction for mobile apps via cross-triplet deep feature embedding, IEEE Trans. Reliab. 71 (1) (2022) 204–220.
[39] T. Han, C. Liu, W. Yang, D. Jiang, A novel adversarial learning framework in deep convolutional neural network for intelligent diagnosis of mechanical faults, Knowl.-Based Syst. 165 (2019) 474–487.
[40] M. Hilbert, W.A. Smith, R.B. Randall, The effect of signal propagation delay on the measured vibration in planetary gearboxes, J. Dyn. Monit. Diagn. 1 (1) (2022) 9–18.
[41] G.D. PHM, PHM data challenge 2009, https://www.phmsociety.org/competition/PHM/09, 2009.
[42] P. Liang, C. Deng, J. Wu, Z. Yang, J. Zhu, Z. Zhang, Single and simultaneous fault diagnosis of gearbox via a semi-supervised and high-accuracy adversarial learning framework, Knowl.-Based Syst. 198 (2020) 105895.
[43] D. Yarowsky, Unsupervised word sense disambiguation rivaling supervised methods, in: 33rd Annual Meeting of the Association for Computational Linguistics, 1995, pp. 189–196.
[44] A. Tarvainen, H. Valpola, Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results, Adv. Neural Inf. Process. Syst. 30 (2017).
[45] K. Sohn, D. Berthelot, N. Carlini, Z. Zhang, H. Zhang, C.A. Raffel, E.D. Cubuk, A. Kurakin, C.-L. Li, Fixmatch: simplifying semi-supervised learning with consistency and confidence, Adv. Neural Inf. Process. Syst. 33 (2020) 596–608.
[46] Y. Li, Y. Song, L. Jia, S. Gao, Q. Li, M. Qiu, Intelligent fault diagnosis by fusing domain adversarial training and maximum mean discrepancy via ensemble learning, IEEE Trans. Ind. Inform. 17 (4) (2021) 2833–2841.
[47] J. Si, H. Shi, T. Han, J. Chen, C. Zheng, Learn generalized features via multi-source domain adaptation: intelligent diagnosis under variable/constant machine conditions, IEEE Sens. J. 22 (1) (2022) 510–519.