Transport and Communications Science Journal, Vol. 75, Issue 01 (01/2024), 1110-1124

Transport and Communications Science Journal

FAULT CLASSIFICATION OF THE ROLLING BEARING BASED ON MINIMAX ENTROPY DOMAIN ADAPTION AUGMENTED WITH SIGNAL GENERATION ALGORITHM

Van Trang Phung[1], Thanh Lich Nguyen[2*]

[1]Viettel High Technology Industries Corporation, No 380 Lac Long Quan Street, Hanoi, Vietnam

[2]University of Transport and Communications, No 3 Cau Giay Street, Hanoi, Vietnam

ARTICLE INFO

TYPE: Research Article
Received: 01/08/2023
Revised: 18/11/2023
Accepted: 20/11/2023
Published online: 15/01/2024
https://doi.org/10.47869/tcsj.75.1.1
* *Corresponding author*
Email: lichnt@utc.edu.vn

Abstract. Rolling bearing faults have been capturing substantial research attention as they are the root causes of malfunctions in mechatronics systems than any other factors. The detection of rolling bearing faults in the early stage is therefore a mandatory requirement demanded by reliable industrial plants. To release the dependence of diagnostic methods on human expertise and system's understanding, this work proposes a fault classification method for rolling bearings that is based on a deep learning framework. The framework consists of a minimax entropy domain adaptation algorithm augmented with a signal generalization algorithm. The function of the signal generalization algorithm is to reduce the domain shift between training and testing datasets that are often obtained experimentally from different working conditions. The generalized signal is then represented in the form of Fourier series whose coefficients contain intrinsic information that associated with different types of bearing faults. A convolutional neural network extracts the hidden information of bearing faults buried in the Fourier coefficients and then categorises the working condition of the bearing under test. By combining the advantages of both signal processing techniques in the frequency domain and the minimax entropy domain adaptation, the novel diagnostic framework is able to detect bearing faults from different working conditions. The effectiveness of the proposed diagnostic algorithm is experimentally verified by two case studies that were prepared with different types and levels of bearing faults.

Keywords: Fast Fourier transform, bearing faults, deep learning, semi-supervised domain adaptation.

1. INTRODUCTION

According to the IEEE 1983-85 motor survey and the Electric Power Research Institute (EPRI) sponsored survey, bearing failures are responsible for more than 40% of all medium-size induction motor malfunctions [1]. The severity of incipient bearing faults has therefore attracted extensive research attention aiming to develop effective fault diagnostic methods, which can provide great contributions in terms of cost reduction for maintenance and productivity improvement [2-6]. The state-of-the-art methods for bearing condition monitoring can be categorized into three main streams, including model-based, signal-based, and data-driven-based methods [7]. The idea behind the first stream lies in tracking the changes/derivations of either state variables, model parameters or system's outputs from expected values to tell whether or not the system has a fault [8-9]. The primary advantages of the model-based diagnostic methods are associated with the requirement of only a small amount of data to detect the fault and the fact that it can be implemented online. However, the model-based methods require a rich understanding of the system, from which an accurate input-output model must be formed and used in the fault detection procedure. For complex systems, the model-based methods are inapplicable as any input-output models are often unavailable.

The signal-based fault diagnostic methods are based on the analysis of the output signals where vibration, sound, speed, force/torque, motor current and magnetic density are widely used. Since a fault within the system can cause changes in the system output signals, it is possible to analyze the signals to extract intrinsic information that has correlation with the bearing faults [7]. The most well-known signal-based diagnostic method is based on motor current signals (MCS) [10-13]. According to the MCS method, if there is a failure in the bearings of a machine, the air gap flux is modulated and the stator currents are generated at predictable frequencies which are also known as characteristic fault frequencies. The bearing damages can be detected if there are noticeable harmonics in the spectrum of the stator currents at the characteristic fault frequencies. The MCS method was proven successful in detecting single-point defects on the inner raceway, outer raceway, balls or cage of rolling bearings.

Recently, deep learning has been investigated extensively to release the dependency of data-driven methods on the human's a priori understanding of the system. One can find a plenty of research on deep learning algorithms for fault classification focusing on the development of either a new model structure, transferable knowledge or optimization of the model's parameters [14-16]. It is widely accepted that the multilayer structure included in the deep learning framework can extract hidden discriminative features in the raw signal and construct the relationship between the extracted features and the fault labels, thus ensuring a high precision fault classification. However, there are still two critical issues associated with the deep learning-based methods [18]. First, a strong domain shifts between the training and testing data makes the fault classifier difficult to generalize well on the testing data. The domain shift is commonly found in industrial scenarios as a faulty bearing may be operated at different working conditions regarding the motor speed, load torque, and external forces [19]. Second, a deep learning model often requires a large number of labeled samples in the training process. Unfortunately, in the real situation, only a small portion of samples are labeled due to the cost of data acquisition process and the requirement of human expertise.

From the advantages and disadvantages of the available bearing faults diagnostic methods, there is an open question on how we can build up a framework that can use the intrinsic

information obtained from signal processing techniques and overcome the two critical issues of deep learning algorithms as aforementioned. Inspired by the idea of semi-supervised feature learning and the fast Fourier transform, we propose a new framework for bearing fault diagnosis as depicted in Figure 1. The highlight of the framework lies in the combination of a signal generalization algorithm (SGA) aiming to reduce the domain shift and a semi-supervised algorithm following the minimax entropy (MME) domain adaptation [20].

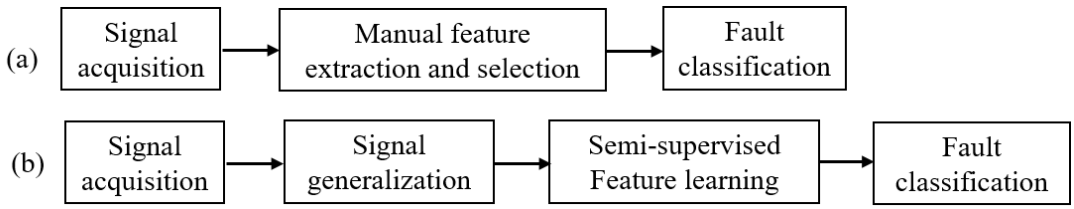

Figure 1. Fault diagnostic framework: (a) the traditional approach, and (b) the proposed framework.

The contributions of this paper can be summarized as follows:

1) Following the calculation of the characteristic fault frequencies, we show how the motor speed affects the domain distribution of the measured vibration signal and why the vibration signal contains elemental information about the bearing damages.

2) A new diagnostic framework composing of a SGA and a semi-supervised domain adaptation algorithm is proposed. The SGA aims to reduce the domain shift between the training and testing data. To do so, the raw signal is scaled and then represented in terms of Fourier series. The Fourier coefficients that contain inherent information about the characteristic fault frequencies are used as the input of the semi-supervised algorithm where a MME is adopted.

3) Two case studies are used to verify the effectiveness of the proposed framework. First, the performance of the MME domain adaptation is tested with and without the use of the SGA. Then, the MME is compared with related work for diagnosing the rolling bearing faults.

From herein, the paper is organized as follows. Section 2 is dedicated to the signal generalization algorithm, where the insight into the characteristic fault frequencies is also elaborated. Section 3 focuses on the MME model for the bearing fault classification, followed by experimental results given in Section 4. Finally, some conclusions are stated in Section 5.

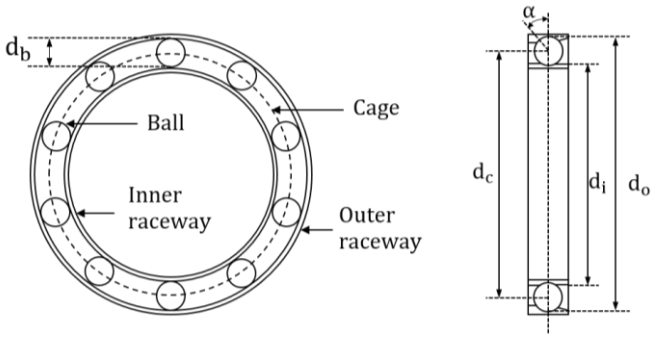

Figure 2. Graphical representation of a rolling bearing: the axial view (left) and the cross-section view (right).

2. SIGNAL GENERATION ALGORITHM (SGA)

In this section, the theoretical fundamental of the characteristic fault frequencies associated with different types of bearing faults is presented. Then a signal generalization algorithm for time-dependent signals is proposed.

2.1 Characteristic fault frequencies of rolling bearings

Figure 2 shows the structure of an angular rolling bearing. There are four main elements in the bearing, namely an inner raceway, an outer raceway, a cage and balls. The load angle is denoted by α. $d_\mathrm{b}, d_\mathrm{c}, d_\mathrm{i}, d_\mathrm{o}$ are the diameter of the balls, the cage, the inner raceway and the outer raceway, respectively [21].

Under the assumption that the outer raceway is fixed inside the bearing housing, the characteristic fault frequencies on the outer raceway f_ORF, on the inner raceway f_IRF, on the balls f_BF and on the cage f_CF are given as follows [22]:

$$f_\mathrm{ORF} = \frac{z}{2} \cdot f_\mathrm{n} \cdot \left(1 - \frac{d_\mathrm{b}}{d_\mathrm{c}} \cdot \cos\alpha\right) \quad (1)$$

$$f_\mathrm{IRF} = \frac{z}{2} \cdot f_\mathrm{n} \cdot \left(1 + \frac{d_\mathrm{b}}{d_\mathrm{c}} \cdot \cos\alpha\right) \quad (2)$$

$$f_\mathrm{BF} = \frac{d_\mathrm{c}}{2d_\mathrm{b}} \cdot f_\mathrm{n} \cdot \left(1 - \frac{d_\mathrm{b}^2}{d_\mathrm{c}^2} \cdot \cos^2\alpha\right) \quad (3)$$

$$f_\mathrm{CF} = \frac{1}{2} \cdot f_\mathrm{n} \cdot \left(1 - \frac{d_\mathrm{b}}{d_\mathrm{c}} \cdot \cos\alpha\right) \quad (4)$$

where z is the number of bearing balls and f_n is the rotational frequency of the inner raceway.

It can be seen from Eq. (1) – Eq. (4) that the characteristic fault frequencies depend on the mechanical dimension of the bearing, on the load angle, and on the inner raceway rotational speed. To have a deeper insight into the characteristic fault frequencies, we examine the benchmark dataset KAt [23]. To prepare the KAt dataset, an accelerometer was used to measure the housing acceleration $a(t)$ of a bearing whose parameters are given in Table 1.

Table 1. Parameters of the bearings for KAt dataset

Parameter	Symbols	Value
Cage's diameter	d_c	28.55 mm
Number of balls	z	8
Load angle	α	0^0
Ball's diameter	d_b	6.75 mm

According to Eq. (1), the outer raceway fault frequency corresponding to the motor speed $n_1 = 900\ \mathrm{min}^{-1}$ ($f_\mathrm{1n} = 15$ Hz) and $n_2 = 1500\ \mathrm{min}^{-1}$ ($f_\mathrm{2n} = 25$ Hz) is $f_\mathrm{1ORF} = 45.81$ Hz and $f_\mathrm{2ORF} = 73.35$ Hz, respectively. The spectrum of the vibration signal associated with KA04 is shown in Figure 3(a). It is observed that the amplitudes of the harmonics at the characteristic fault frequency and their multiples are significant compared to other frequencies.

$$f_\mathrm{ORF}^k = k \cdot f_\mathrm{ORF}, \quad k = 1, 2, 3 \ldots \quad (5)$$

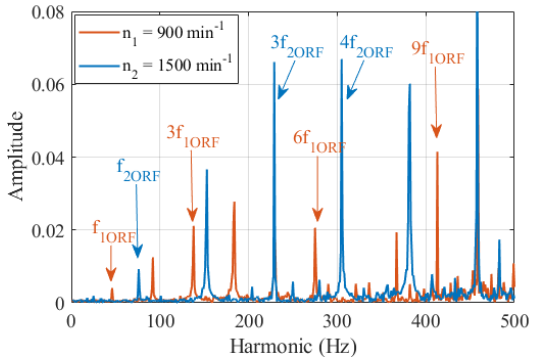

Figure 3(a). Characteristic fault frequency corresponding to an outer raceway defect (KA04-KAt: making interpolation of one sample with 64,000 data points).

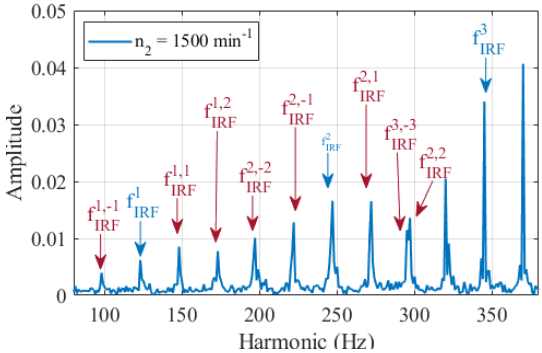

Figure 3(b). Characteristic fault frequency and the sideband effect corresponding to an inner raceway defect (KI18-KAt: making interpolation of one sample with 64,000 data points).

The characteristic fault frequencies are observed at the round up or round down of the theoretical frequencies when they are not integers. An increase in the motor speeds shifts the fault frequency and its multiples from the left to the right.

For the case of an inner raceway defect, experiment KI18 is considered. The spectrum of the vibration signal with the motor speed $n_2 = 1500\ \text{min}^{-1}$ is depicted in Figure 3(b). According to Eq. (2), the inner raceway fault frequency is $f_{\text{IRF}} = 123.64$ Hz. Figure 3(b) shows that harmonics at f_{IRF} and their multiples are clearly visible.

$$f_{\text{IRF}}^{k} = k \cdot f_{\text{IRF}}, \quad k = 1, 2, 3 \ldots \tag{6}$$

The inner raceway effect has a distinctive phenomenon related to the sideband effect resulting from a so-called load zone distribution. It was proven that when a radial load is applied to a bearing, the rolling elements are not equally loaded [24]. When the inner raceway defect rotates through the load zone, an impulse is generated every time a ball passes the damaged area. The intensity of the shock will be at its greatest magnitude if the defect is located at the point of maximum radial force. Therefore, the most intense impulse is generated on every turn of the machine, assuming that the inner raceway synchronously rotates with the shaft of the driven machine. As the modulated signal produces spectral components at the sum- and difference- frequencies, the sidebands around $k \cdot f_{\text{IRF}}$ will be appear at any frequencies given by

$$f_{\text{IRF}}^{k,v} = k \cdot f_{\text{IRF}} + v \cdot f_{\text{n}}, \quad k = 1, 2, 3 \ldots, v = \pm 1, \pm 2 \ \ldots \tag{7}$$

The sideband effect is clearly observed in Figure 3(b). The spectral analysis shown in Figure 3(a) and Figure 3(b) proves that the vibration signal contains elemental information about the bearing damages that are hidden in terms of characteristic fault frequencies. The fault frequencies are therefore widely used in diagnostic methods. In this research work, we propose a diagnostic model that utilizes the spectral information of the vibration signal rather than the raw data.

2.2. Signal Generalization Algorithm (SGA)

Our key idea lies in the signal generalization algorithm aiming to reduce the domain shift of time-dependent raw signals measured at different motor speeds. Without the loss of

generality, supposing that we need to generalize vibration signals $v_1(t)$ and $v_2(t)$ associated with motor speeds $n_1(\text{min}^{-1})$ and $n_2(\text{min}^{-1})$, respectively. The signal generalization algorithm is in Table 2.

There are some important considerations while conducting the SGA. First, in theory, N_{FFT} is no more than $1/2 \cdot N_s \cdot \min(k_1, k_2)$. N_{FFT} is a hyper-parameter and is chosen empirically. The higher the value of N_{FFT}, the more harmonics and hence the more features to be considered, and the more cubersome the deep learning model is. Second, $\Delta n_{\max}$ decides how much domain shift is remained in the generalized signals. $\Delta n_{\max}$ should be as small as possible but no need to be zero because a domain adaptation algorithm can handle effectively a small domain shift. In real scenarios where the motor is operated at hundreds revolution per minute, $\Delta n_{\max} \leq 100\ \text{min}^{-1}$ can be selected. Third, the FFT algorithm in Step 3 can be carried out with a sliding-window method [25], especially in the case N_{FFT} is small compared with $1/2 \cdot N_s \cdot \min(k_1, k_2)$. This feature enables the SGA to be conducted online. And finally, the algorithm generalizes raw signals in terms of frequency but the amplitude. This means that the spectral amplitude of the generalized signals can differ from that in the raw signals as a result of the signal scaling in Step 2.

Table 2. Signal generalization algorithm.

<table>
<tr><td>Input:</td><td>Raw signals $\{v_1(t)\}_{t=1}^{N_s}$ and $\{v_2(t)\}_{t=1}^{N_s}$ corresponding to motor speed n_1, n_2; N_s is the number of data points in one sample; The number of Fourier coefficients to be considered N_{FFT}; the maximum speed difference after generalization $\Delta n_{\max}$</td></tr>
<tr><td>Output:</td><td>Fourier coefficients $\{c_{1sg}\}_{n=1}^{N_{\text{FFT}}}$ and $\{c_{2sg}\}_{n=1}^{N_{\text{FFT}}}$ of the generalized signal $v_{1sg}(t)$ and $v_{2sg}(t)$ of $v_1(t)$ and $v_2(t)$, respectively.
$\{c_{1sg}\}_{n=1}^{N_{\text{FFT}}} \xleftarrow{\text{FFT}} v_{1sg}(t) \xleftarrow{\text{signal scaling}} v_1(t)$
$\{c_{2sg}\}_{n=1}^{N_{\text{FFT}}} \xleftarrow{\text{FFT}} v_{2sg}(t) \xleftarrow{\text{signal scaling}} v_2(t)$</td></tr>
<tr><td colspan="2">Algorithm procedure:</td></tr>
<tr><td>Step 1:</td><td>- Find a divisor n_{1sg} of n_1 and n_{2sg} of n_2 such that $\lvert n_{1sg} - n_{2sg} \rvert \leq \Delta n_{\max}$.
- Define k_1, k_2 such that
$n_1 = k_1 \cdot n_{1sg};\ n_2 = k_2 \cdot n_{2sg}$</td></tr>
<tr><td>Step2:</td><td>Scaling the raw signals:
for i = 1: N_s
for j = 1: k_1
$v_{1sg}(i \cdot k_1 + j) = v_1(i);$
end
for j = 1: k_2
$v_{2sg}(i \cdot k_2 + j) = v_2(i);$
end
end</td></tr>
<tr><td>Step 3:</td><td>Taking Fourier interpolation of the generalized signals
$\{c_{1sg}\}_{n=1}^{N_{\text{FFT}}} \xleftarrow{\text{FFT}} v_{1sg}(t)$
$\{c_{2sg}\}_{n=1}^{N_{\text{FFT}}} \xleftarrow{\text{FFT}} v_{2sg}(t)$</td></tr>
</table>

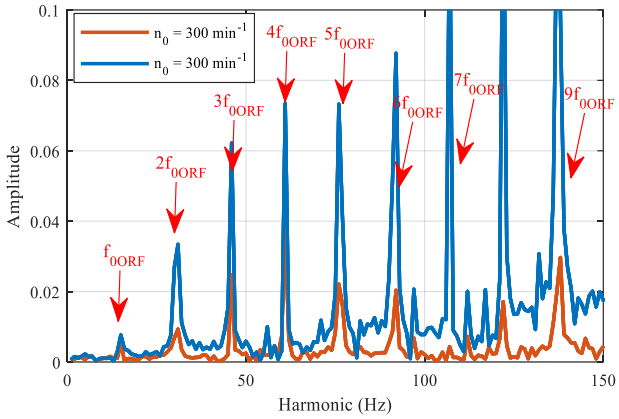

Figure 4. Generalized characteristic fault frequencies corresponding to an outer raceway defect at two different motor speeds (KA04-KAt).

Let's take experiment KA04 in the dataset KAt as an example. N_s is set to 64,000 data points, meaning that a sample of vibration raw signal measured in 1 (s) with the sampling frequency of 64 kHz is considered. N_{FFT} is set to 2500 and $n_{1sg} = n_{2sg} = n_0 = 300$ min^{-1}. The gain k_1 and k_2 are 3 and 5, respectively. In this case, $\Delta n_{\max}$ can be any arbitrary positive real number. The spectrum content of the generalized signals is shown in Figure 4. It is clearly observed there is no domain shift in the spectral content as in this case $n_{1sg} = n_{2sg}$ is hold. It is also realized that the spectral content of the raw vibration signal corresponding to $n_1 = 900$ min^{-1} (the bright orange in Figure 3(a)) is scaled in the factor of $k_1 = 3$ (the bright orange in Figure 4). The same phenomenon is observed with the raw signal measured at $n_2 = 1500$ min^{-1}.

3. MINIMAX ENTROPY DOMAIN ADAPTION

In this section, the minimax entropy domain adaptation is presented. We verify the data cleaned by the proposed processing algorithm on various domain adaptation approaches. However, we focus on the minimax entropy strategy [20] for semi-supervised domain adaptation, in which the classification performance on the target domain is significantly improved by only using a minority of labeled target samples during the training process. By using the minimax entropy strategy, the representations of source and target domains can be aligned without requiring any domain adaptation component, such as a domain discriminator; thus, it can reduce the model complexity and computation time. The minimax strategy is divided into two specific stages. In the first stage, the model, including a feature extractor and a classifier, is trained on rich labeled samples from the source domain and a minority of labeled samples from the target domain as follows:

$$\mathcal{L}_{\text{Cls}} = \mathbb{E}_{\left(x_i^{\text{l}}, y_i^{\text{l}}\right) \sim \mathcal{D}_{\text{l}}} \sum_{k=1}^{K} 1_{\left[k=y_i^{\text{l}}\right]} \log\left(F\left(G\left(x_i^{\text{l}}\right)\right)\right) \tag{8}$$

where F(.) and G(.) are the classifier and feature extractor, respectively, $\left(x_i^{\text{l}}, y_i^{\text{l}}\right)$ is a pair sample and label in the labeled set, $\mathcal{D}_{\text{l}} = \mathcal{D}_{\text{S}} \cup \mathcal{D}_{\text{tl}}$, where $\mathcal{D}_{\text{S}}$ is the set of source data, and $\mathcal{D}_{\text{tl}}$ is the set of labeled target data. The role of this stage is to estimate the prototypes represented by the

weights of the classifier. Then, the minimax strategy is conducted in the second stage. To be specific, the feature extractor is trained to minimize entropy on the unlabeled target data, in which the discriminative target representations can be achieved by clustering around the estimated prototypes. In contrast, the classifier is trained to maximize entropy on the unlabeled target data for domain adaptation, which encourages the feature extractor to generate unlabeled features similar to prototype generation. The entropy is computed on the unlabeled target samples as follows:

$$H = -\mathbb{E}_{(x_i^{\mathrm{tu}})\sim\mathcal{D}_{\mathrm{tu}}} \sum_{k=1}^{K} p(y=k|x_i^{\mathrm{tu}}) \log \sum_{k=1}^{K} p(y=k|x_i^{\mathrm{tu}}) \tag{9}$$

Where $p(y=k|x_i^{\mathrm{tu}})$ represents the probability of the sample x_i^{tu} to be predicted as class *k*. This probability is calculated by using a similarity-based between the extracted feature and each element of weight vectors as follows:

$$p(y=k|x_i^{\mathrm{tu}}) = \mathrm{softmax}\left(\frac{1}{T}\mathbf{W}^{\mathrm{T}}\left(G(x_i^{\mathrm{tu}})\right)\right) \tag{10}$$

where T is a temperature, $\mathbf{W} = [\mathbf{w_1}, \mathbf{w_2}, \dots, \mathbf{w_K}]$ is a matrix including weight vectors of the cosine classifier, in which each $\boldsymbol{w_i}$ represents the certain prototype of the *i*-th class. This process is depicted in Figure 5.

4. EXPERIMENTS

In this section, experimental results are presented to verify the effectiveness of the proposed fault classification framework. In our model, a five-layer 1-D CNN [19] was employed for the feature extractor, while a fully connected neural network was used for fault classification, as shown in Figure 6.

4.1. Case study 1: KAt

a) Dataset description

KAt dataset was collected from a specific test rig as presented in [23]. There are 32 experiments for testing healthy bearings, artificially damaged bearings, and accelerated damaged bearings. Each experiment has 20 files where each file contains 256,000 vibration data points resulting from a 4 s – 64 kHz data acquisition procedure. The damages in the faulty bearings are located in either the inner or the outer raceways, meaning that the KAt dataset has three classes. In addition, the test rig was operated under four conditions that differed from each other in terms of the motor speed, load torque and radial force. The four operating conditions are denoted by A, B, C and D, whose parameter settings are given in Table 3.

To generate the data samples, the overlapping sliding windows procedure aiming to segment any time-dependent signals is used [19]. The window size and the shifting window are set to 16000 and 4000, respectively. The higher the window size, the better noise-free of the samples, but the more computational burden of the fault classification algorithm. The value of the shifting window is inversely proportional to the number of samples considered in the

training and testing phases. Generally, a higher number of samples provide more information for the fault classification algorithm, yet it requires more memory.

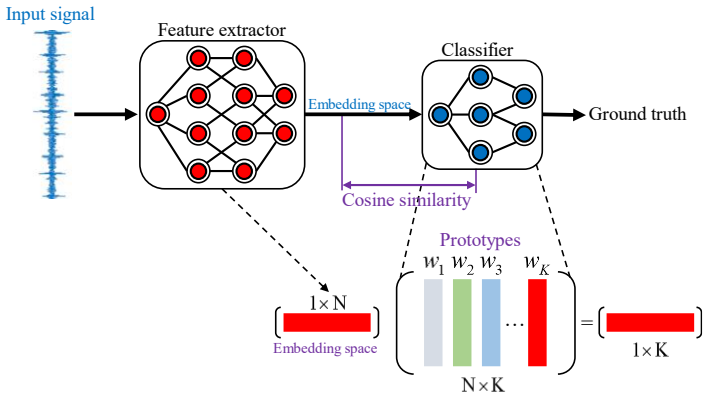

Figure 5. Framework of the Minimax Entropy Domain Adaptation for fault diagnostic.

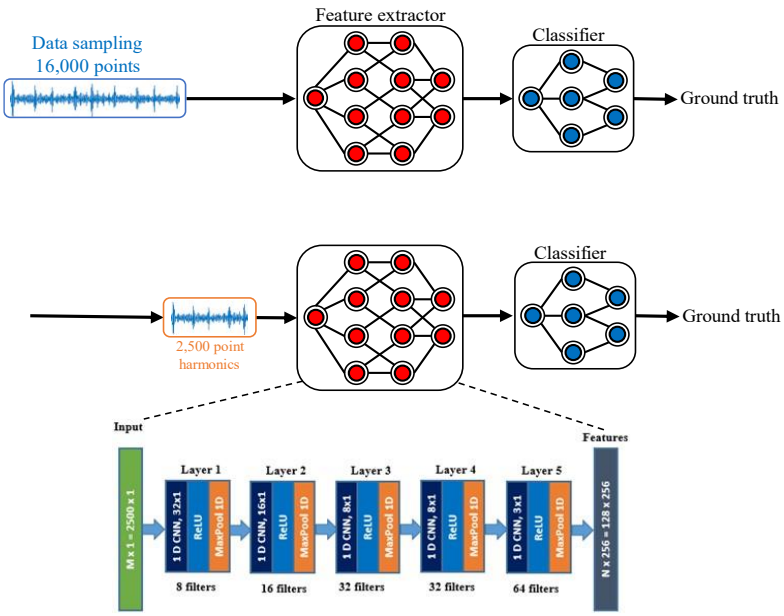

Figure 6. Five layers 1D CNN feature extractor and Data preparation for KAt.

As aforementioned, there are two types of faulty bearings: artificial and accelerated faulty bearing damages. This research work focuses on real accelerated damaged cases. For each working condition, one experiment for a healthy bearing, five experiments for inner raceway faults and five experiments for outer raceway faults are taken into account. In this configuration, there are 1200, 6000, 6000 samples for the healthy, the inner raceway faults and the outer raceway faults, respectively. Therefore, the total number of samples for each working condition A, B, C and D is 13200.

It is observed in Table 3 that there are two operating speeds of the motor: 900 min^{-1} and 1500 min^{-1}. The signal generalization algorithm can be conducted in a similar way as already presented in Section 2.2. The number of considered harmonics $N_{\text{FFT}} = 2500$ that is much lower than the sample's size. The graphical representation of the data preparation is depicted in Figure 7.

Table 3. Working condition of KAt dataset.

Working condition (domains)	Motor speed (min^{-1})	Load torque (Nm)	Radial force (N)
A	900	0.7	1000
B	1500	0.1	1000
C	1500	0.7	400
D	1500	0.7	1000

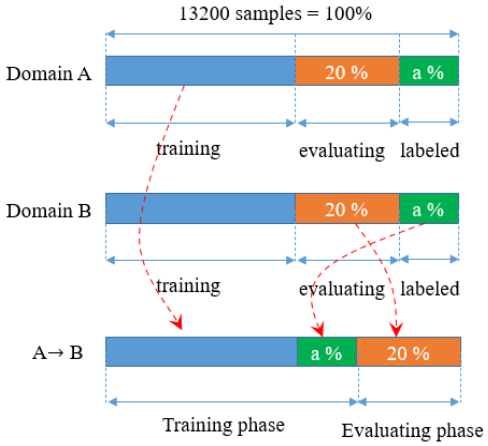

Figure 7. Data splitting for Case study 1 (KAt).

b) Experimental results

Experiments were conducted to highlight the effectiveness of the signal generalization and of the MME algorithms. To evaluate the performance of the SGA, a comparison between the fault classification frameworks with and without the signal generalization is conducted. Figure 7 shows the data split for each working condition, or each domain, where the number of examples for the evaluation process is set to 20 %. The number of labeled samples in each domain is set to a %. The remaining samples are for the training dataset. Figure 7 also demonstrates an example of the cross-domain A → B where A is the source domain and B is the target domain. It is realized that the training phase uses the training samples in domain A and the labeled samples in domain B, while the evaluation process utilizes the only samples in domain B.

Table 4 shows the evaluation of 12 cross-domain experiments without (RAW) and with the signal generalization algorithm (SGA) for five cases: $a = 1$ %, $a = 0.7$ %, $a = 0.5$ %, $a = 0.3$ %, and $a = 0.1$ %. Each experiment is conducted three times to minimize the effect caused by data bias and inter-domain discrepancy. The average accuracy is calculated and given in column AVG. Overall, the SGA ensures an average accuracy considerably higher than that with raw signal: 78.71 % compared with 74.49 % for the case $a = 0.1$ % and 96.37 % compared with 88.80 % for the case $a = 1$%. It is observed that the proposed framework provides a quite high average accuracy (86.22 %) even though there are just 40 samples ($a = 0.2$ %) with labels in the target domain used in the training phase. In this case, 5 samples belong to the healthy bearing and 20 samples belong to the inner and outer raceway defects. When $a = 0.1$ %, the number of samples for the healthy condition becomes one. This is the main reason why the algorithm accuracy is lowered as one sample hardly represents a whole class of 1200 samples.

Table 3 shows that domain A differs from the others in terms of motor speed. As already mentioned, domain discrepancies caused by the changing motor speed considerably affect the accuracy of the fault classification algorithm. This phenomenon is confirmed again in Table 4 that the RAW experiments achieve lower accuracy for six cross-domain experiments related to domain A. To realize the effectiveness of the SGA, we denote A_{AVG} as the average accuracy of six cross-domain experiments related to domain A. Table 4 demonstrates that the SGA provides

a higher average accuracy A_{AVG} in all cases respecting the value of a. The value of A_{AVG} associated with the SGA is at least 9.84 % ($a = 0.5$ %) higher than that related to the RAW experiments.

In addition to the algorithm precision, the training time, the computational burden and the size of the model are important factors when the fault classification framework is considered in practical application. It is obvious that the SGA imposes the processor to carry out the signal scaling and the fast Fourier transform algorithm, but it reduces the size of the data samples greatly. For the above experiment, the sample size of the SGA is $N_{\mathrm{FFT}} = 2500$ that is much smaller than the original sample size of 16000.

Table 4. KAT: Evaluation of the fault classification algorithm with and without the signal generation.

$N_t^{labeled}$	Method	$A \rightarrow B$	$A \rightarrow C$	$A \rightarrow D$	$B \rightarrow A$	$B \rightarrow C$	$B \rightarrow D$	$C \rightarrow A$	$C \rightarrow B$	$C \rightarrow D$	$D \rightarrow A$	$D \rightarrow B$	$D \rightarrow C$	A_{AVG}	AVG
132	SGA	87.37	93.34	92.16	93.10	99.98	100	94.79	99.97	98.23	97.54	100	99.94	93.05	96.37
(a=1%)	RAW	85.05	75.25	87.71	71.49	98.85	100	65.60	100	100	85.32	100	96.36	78.4	88.80
92	SGA	70.49	91.02	94.47	94.24	99.97	100	91.83	99.99	99.78	91.69	100	99.97	88.96	94.45
(a=0.7%)	RAW	81.46	66.78	81.73	89.15	98.13	100	54.44	100	100	87.96	100	98.56	76.92	88.18
66	SGA	78.88	86.11	99.93	89.92	100	100	83.79	99.62	100	79.64	100	99.99	86.38	93.15
(a=0.5%)	RAW	72.69	72.94	81.91	81.45	98.01	100	67.58	100	100	82.67	100	98.15	76.54	87.95
40	SGA	88.21	67.42	84.64	75.81	84.15	100	72.67	99.99	93.81	75.85	100	92.10	77.43	86.22
(a=0.3%)	RAW	70.25	62.75	83.05	76.03	97.01	100	54.47	100	100	63.01	100	96.92	68.26	83.62
13	SGA	72.71	66.50	75.50	66.23	75.93	100	53.76	93.36	90.42	53.56	100	96.55	64.71	78.71
(a=0.1%)	RAW	56.81	51.45	57.43	53.42	87.37	100	54.73	100	96.53	46.97	100	89.16	53.47	74.49

The effectiveness of the MME method is verified by comparing its performance with that of the CDAN [26], DANN [27], and DAN [28]. It is well-known that CDAN, DANN and DAN are unsupervised domain adaptation methods while MME is a semi-supervised one. To make the comparison fair and meaningful, the datasets used to train and verify the MME, CDAN, DANN, and DAN are exactly the same. The comparison was also conducted for three cases: $a = 0.1$ %, $a = 0.3$ % and $a = 0.5$ %. It should be mentioned again that each experiment is carried out three times. The average accuracy is computed and shown in Figure 8. It is observed that the MME ensures a considerably higher accuracy than the three competing approaches. Compared with CDAN, the second-best method, the MME's accuracy surpasses 8.65 % (93.15 % compared with 84.50 %) for the case a = 0.5 % and 0.97% (78.71 % compared with 77.74 %) for the case a = 0.1 %.

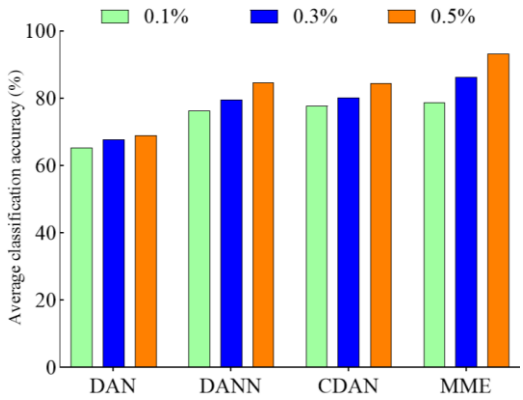

Figure 8. Performance of the MME, CDAN, DANN and DAN for KAt dataset.

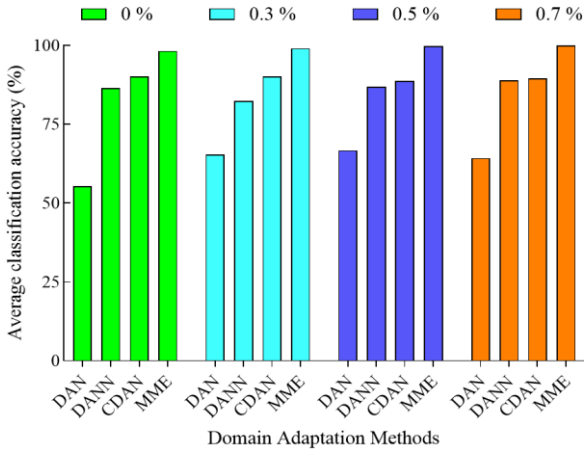

Figure 9. Performance of the MME, CDAN, DANN and DAN for CWRU dataset.

4.2. Case study 2: Case Western Reserve University Dataset (CWRU)

a) Dataset description

The CWRU dataset [29] was collected from the drive end equipped with a 2-hp electric motor. The faulty bearings were managed to have damages at the inner raceway (IF), outer raceway (OF), or ball bearings (BF). The test-bench was operated at specific load conditions ranging from 0 to 3 horsepower (HP), meaning that there are four domains denoted by E, F, G, H as shown in Table 5.

The sampling rate for the data acquisition was set to 12 kHz. Table 5 shows that each domain (E, F, G, H) has 10 data subsets according to 10 working conditions of the bearings: one healthy and three faulty levels for the IR, OR and BF. Each data subset was obtained from a 10 (s) data acquisition procedure, resulting in 120,000 data points for each data subset. Similarly to case study 1, the overlapping sliding window procedure is applied, where the size of each sample is 12,000 data points and the shifting window is 500 data points. In this configuration, each domain has 2160 samples. The data splitting procedure is similar to that as depicted in Figure 7.

One important characteristic of the CWRU dataset is associated with a stable motor speed that ranges from 1720 min^{-1} to 1797 min^{-1}. Due to the small motor speed variation, the domain

Table 5. Working condition of CWRU dataset.

Working condition (domains)	Load torque (hp)	Fault types	Fault size (inches)
E	0	Normal, IF, OF, BF	0.007, 0.014,0.021
F	1	Normal, IF, OF, BF	0.007,0.014,0.021
G	2	Normal, IF, OF, BF	0.007,0.014,0.021
H	3	Normal, IF, OF, BF	0.007,0.014,0.021

shift originated from the motor speed is insignificant, thus Steps 1 and 2 of the SGA can be omitted. The domain adaptation is now handled merely by the deep learning algorithm.

b) Experimental results

Experiments were also conducted with the CWRU dataset to verify the performance of the proposed fault classification framework. To examine the effectiveness of the SGA, 12 cross-domain experiments based on the MME algorithm with either the raw data or the SGA were implemented. The average accuracy of the experiments is given in Table 6. It is clearly observed that the fault classification accuracy is high even in the cases where there are just a few labeled samples in the target domain used in the training phase. The uncritical domain shift in the CWRU is one of the main reasons for the result. Extensive experiments show that the MME method augmented with the SGA ensures a 100% accuracy when $a \geq 1$ %. For the case $a = 1$ % there are only 22 labeled samples in the target domain to be used in the training process. Despite the small number of labeled samples, the framework still provides a 100% fault classification accuracy. When $a < 1$ %, the proposed framework achieves the average accuracy of at least 98.5% for the case $a = 0$ %, or in other words, the MME works in the unsupervised domain adaptation setting. Compared with the utilization of the raw samples whose size is two times as large as the SGA (12000 data points compared with 6000 Fourier coefficients) and the training time is 1.6 times longer (68.8 s with the raw signal compared with 41.2 s of the SGA), the accuracy of the proposed framework is slightly lower when $a < 1\%$. In these scenarios, the number of samples $N_t^{\text{labeled}} < 22$. The number of samples for the healthy condition can be 0 or 1 that hardly represents the whole normal class.

Table 6. CWRU: Evaluation of the fault classification algorithm with and without the signal generation.

$N_t^{labeled}$	**Method**	$E \rightarrow F$	$E \rightarrow G$	$E \rightarrow H$	$F \rightarrow E$	$F \rightarrow G$	$F \rightarrow H$	$G \rightarrow E$	$G \rightarrow F$	$G \rightarrow H$	$H \rightarrow E$	$H \rightarrow F$	$H \rightarrow G$	AVG
22	SGA	100	100	100	100	100	100	100	100	100	100	100	100	100
(a=1%)	RAW	100	100	100	97.5	100	99.77	100	100	99.09	98.4	100	100	99.56
15	SGA	100	100	100	100	100	100	100	100	100	100	100	100	100
(a=0.7%)	RAW	100	100	100	96.81	100	100	100	99.31	98.86	98.63	100	100	99.47
11	SGA	100	100	96.77	100	100	100	100	100	100	100	100	100	99.73
(a=0.5%)	RAW	100	100	100	95.30	100	100	100	100	98.18	99.54	100	100	99.42
6	SGA	100	100	100	88.96	100	100	100	100	100	100	100	100	99.08
(a=0.3%)	RAW	100	100	100	94.77	100	99.77	100	100	99.31	99.09	100	100	99.41
0	SGA	100	100	79.01	100	100	100	100	100	100	100	100	100	98.25
(a=0%)	RAW	100	100	100	95.04	100	100	100	100	97.86	99.77	100	100	99.39

The second experiment is dedicated to the performance comparison of the MME, CDAN, DANN and DAN. The average accuracies of the 12 cross-domain experiments are shown in Figure 9. Compared with CDAN, the second-best method, the MME's accuracy surpasses at least 8.06 % for the case the four methods work in unsupervised condition (a = 0 %). The MME together with the SGA ensure a very high accuracy of 98.25 % even for the case a = 0 %. This fact confirms the effectiveness of the proposed diagnostic framework.

5. CONCLUSIONS

In this paper, a novel diagnostic framework for the detection of rolling bearing faults was presented. The framework is composed of a signal generalization algorithm and a semi-supervised deep learning algorithm based on the minimax entropy domain adaptation. The signal generalization algorithm (SGA) processes the raw data in terms of rescaling and

representing it in the form of Fourier series. The SGA was proven successfully to reduce the domain shift between the source and the target domains, especially in the case the datasets were obtained from different operational speed of the electrical motor, thus enhancing the prediction accuracy of the deep learning algorithm based on the MME with less computational time and less memory required for the implementation of the diagnostic framework. Experimental results with two published datasets confirm the outstanding performance of the proposed diagnostic framework.

ACKNOWLEDGMENT

This research is funded by University of Transport and Communications (UTC) under grant number T2023-CK-001.

REFERENCES

[1]. W. F. Braun, B. G. Douglass, C. R. Heising, D. O. Koval, P. O'Donnell, EEE Recommended Practice for the Design of Reliable Industrial and Commercial Power Systems (Gold Book), IEEE Std 493-1997 IEEE Gold Book, (1998) 1–464. https://doi.org/10.1109/IEEESTD.1998.89291
[2]. F. Immovilli, M. Cocconcelli, A. Bellini, R. Rubini, Detection of Generalized-Roughness Bearing Fault by Spectral-Kurtosis Energy of Vibration or Current Signals, IEEE Trans. Ind. Electron., 56 (2009) 4710–4717. https://doi.org/10.1109/TIE.2009.2025288
[3]. T. Wang, Z. Liu, G. Lu, J. Liu, Temporal-Spatio Graph Based Spectrum Analysis for Bearing Fault Detection and Diagnosis, IEEE Trans. Ind. Electron., 68 (2021) 2598–2607. https://doi.org/10.1109/TIE.2020.2975499
[4]. M. Kuncan, An Intelligent Approach for Bearing Fault Diagnosis: Combination of 1D-LBP and GRA, IEEE Access, 8 (2020) 137517–137529. https://doi.org/10.1109/ACCESS.2020.3011980
[5]. D. Huang, J. Yang, D. Zhou, G. Litak, Novel Adaptive Search Method for Bearing Fault Frequency Using Stochastic Resonance Quantified by Amplitude-Domain Index, IEEE Trans. Instrum. Meas., 69 (2020) 109–121. https://doi.org/10.1109/TIM.2019.2890933
[6]. S. Nath, J. Wu, Y. Zhao, W. Qiao, Low Latency Bearing Fault Detection of Direct-Drive Wind Turbines Using Stator Current, IEEE Access, 8 (2020) 44163–44174. https://doi.org/10.1109/ACCESS.2020.2977632
[7]. X. Dai, Z. Gao, From Model, Signal to Knowledge: A Data-Driven Perspective of Fault Detection and Diagnosis, IEEE Trans. Ind. Inform., 9 (2013) 2226–2238. https://doi.org/10.1109/TII.2013.2243743
[8]. S. Zhang et al., Model-Based Analysis and Quantification of Bearing Faults in Induction Machines, IEEE Trans. Ind. Appl., 56 (2020) 2158–2170. https://doi.org/10.1109/TIA.2020.2979383
[9]. M. Ojaghi, M. Sabouri, and J. Faiz, Analytic Model for Induction Motors Under Localized Bearing Faults, IEEE Trans. Energy Convers., 33 (2018) 617–626. https://doi.org/10.1109/TEC.2017.2758382
[10]. D. T. Hoang, H. J. Kang, A Motor Current Signal-Based Bearing Fault Diagnosis Using Deep Learning and Information Fusion, IEEE Trans. Instrum. Meas., 69 (2020) 3325–3333. https://doi.org/10.1109/TIM.2019.2933119
[11]. F. Dalvand, S. Dalvand, F. Sharafi, M. Pecht, Current Noise Cancellation for Bearing Fault Diagnosis Using Time Shifting, IEEE Trans. Ind. Electron., 64 (2017) 8138–8147. https://doi.org/10.1109/TIE.2017.2694397
[12]. E. Elbouchikhi, V. Choqueuse, F. Auger, M. E. H. Benbouzid, Motor Current Signal Analysis Based on a Matched Subspace Detector, IEEE Trans. Instrum. Meas., 66 (2017) 3260–3270. https://doi.org/10.1109/TIM.2017.2749858
[13]. A. Naha, A. K. Samanta, A. Routray, A. K. Deb, Low Complexity Motor Current Signature Analysis Using Sub-Nyquist Strategy With Reduced Data Length, IEEE Trans. Instrum. Meas., 66 (2017) 3249–3259. https://doi.org/10.1109/TIM.2017.2737879

[14]. L. Wen, X. Li, L. Gao, Y. Zhang, A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method, IEEE Trans. Ind. Electron., 65 (2018) 5990–5998. https://doi.org/10.1109/TIE.2017.2774777

[15]. P. Cao, S. Zhang, J. Tang, Preprocessing-Free Gear Fault Diagnosis Using Small Datasets With Deep Convolutional Neural Network-Based Transfer Learning, IEEE Access, 6 (2018) 26241–26253. https://doi.org/10.1109/ACCESS.2018.2837621

[16]. T. Han, C. Liu, W. Yang, D. Jiang, Deep transfer network with joint distribution adaptation: A new intelligent fault diagnosis framework for industry application, ISA Trans., 97 (2020) 269–281. https://doi.org/10.1016/j.isatra.2019.08.012

[17]. R. Zhang, H. Tao, L. Wu, Y. Guan, Transfer Learning With Neural Networks for Bearing Fault Diagnosis in Changing Working Conditions, IEEE Access, 5 (2017) 14347–14357. https://doi.org/10.1109/ACCESS.2017.2720965

[18]. G. Xu, M. Liu, Z. Jiang, W. Shen, C. Huang, Online Fault Diagnosis Method Based on Transfer Convolutional Neural Networks, IEEE Trans. Instrum. Meas., 69 (2020) 509–520. https://doi.org/10.1109/TIM.2019.2902003

[19]. M. Ragab et al., Adversarial Multiple-Target Domain Adaptation for Fault Classification, IEEE Trans. Instrum. Meas., 70 (2021) 1–11. https://doi.org/10.1109/TIM.2020.3009341

[20]. K. Saito, D. Kim, S. Sclaroff, T. Darrell, K. Saenko, Semi-supervised Domain Adaptation via Minimax Entropy, ArXiv190406487 Cs, (2019), Accessed: Feb. 26, 2022. [Online]. Available: http://arxiv.org/abs/1904.06487

[21]. J. R. Stack, T. G. Habetler, R. G. Harley, Fault classification and fault signature production for rolling element bearings in electric machines, IEEE Trans. Ind. Appl., 40 (2004) 735–739. https://doi.org/10.1109/TIA.2004.827454

[22]. S. A. McInerny, Y. Dai, Basic vibration signal processing for bearing fault detection, IEEE Trans. Educ., 46 (2003) 149–156. https://doi.org/10.1109/TE.2002.808234

[23]. C. Lessmeier, J. Kimotho, D. Zimmer, W. Sextro, Condition Monitoring of Bearing Damage in Electromechanical Drive Systems by Using Motor Current Signals of Electric Motors: A Benchmark Data Set for Data-Driven Classification, Jul. 2016.

[24]. F. B. Oswald, E. V. Zaretsky, J. V. Poplawski, Effect of Internal Clearance on Load Distribution and Life of Radially Loaded Ball and Roller Bearings, Tribol. Trans., 55 (2012) 245–265. https://doi.org/10.1080/10402004.2011.639050

[25]. E. T. S. Calvo, Diagnostics of Rotor Asymmetries in Inverter-Fed, Variable Speed Induction Machine. University of Siegen: Logos Verlag Berlin GmbH, (2009).

[26]. M. Long, Z. Cao, J. Wang, M. I. Jordan, Conditional Adversarial Domain Adaptation, ArXiv170510667 Cs, (2018), Accessed: Apr. 21, 2022. [Online]. Available: http://arxiv.org/abs/1705.10667

[27]. Y. Ganin et al., Domain-Adversarial Training of Neural Networks, ArXiv150507818 Cs Stat, (2016), Accessed: Apr. 21, 2022. [Online]. Available: http://arxiv.org/abs/1505.07818

[28]. M. Long, Y. Cao, J. Wang, M. I. Jordan, Learning Transferable Features with Deep Adaptation Networks, ArXiv150202791 Cs, (2015), Accessed: Apr. 30, 2022. [Online]. Available: http://arxiv.org/abs/1502.02791

[29]. W. A. Smith, R. B. Randall, Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study, Mech. Syst. Signal Process., 64–65 (2015) 100–131. https://doi.org/10.1016/j.ymssp.2015.04.021