MDPI

Communication

Semi-Supervised Transfer Learning Method for Bearing Fault Diagnosis with Imbalanced Data

Xia Zong [1,2], Rui Yang [1,3,*], Hongshu Wang [1,2], Minghao Du [1,2], Pengfei You [1,2], Su Wang [1,2] and Hao Su [1,2]

1   School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China; xia.zong18@student.xjtlu.edu.cn (X.Z.); hongshu.wang18@student.xjtlu.edu.cn (H.W.); minghao.du18@student.xjtlu.edu.cn (M.D.); pengfei.you18@student.xjtlu.edu.cn (P.Y.); su.wang19@student.xjtlu.edu.cn (S.W.); hao.su19@student.xjtlu.edu.cn (H.S.)
2   School of Electrical Engineering, Electronics & Computer Science, University of Liverpool, Liverpool L69 3BX, UK
3   Research Institute of Big Data Analytics, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China
*   Correspondence: r.yang@xjtlu.edu.cn; Tel.: +86-512-88161502

Abstract: Fault diagnosis is essential for assuring the safety and dependability of rotating machinery systems. Several emerging techniques, especially artificial intelligence-based technologies, are used to overcome the difficulties in this field. In most engineering scenarios, machines perform in normal conditions, which implies that fault data may be hard to acquire and limited. Therefore, the data imbalance and the deficiency of labels are practical challenges in the fault diagnosis of machinery bearings. Among the mainstream methods, transfer learning-based fault diagnosis is highly effective, as it transfers the results of previous studies and integrates existing resources. The knowledge from the source domain is transferred via Domain Adversarial Training of Neural Networks (DANN) while the dataset of the target domain is partially labeled. A semi-supervised framework based on uncertainty-aware pseudo-label selection (UPS) is adopted in parallel to improve the model performance by utilizing abundant unlabeled data. Through experiments on two bearing datasets, the accuracy of bearing fault classification surpassed the independent approaches.

Keywords: fault diagnosis; imbalanced data; semi-supervised learning; transfer learning; uncertainty-aware pseudo-label selection

1. Introduction

With the progress of industrialization, rotating machinery has gradually become of great significance and is widely used in industrial applications. However, the working condition of rotating machinery is quite onerous, which always makes them degenerate and abates the machinery service performance [1]. In detail, a specific fault type, the bearing fault, accounts for almost 30% of the faults in rotating machinery [2]. Although traditional fault diagnosis methods based on engineers' ample experience and domain-specific knowledge have shown good performance, rotating machinery has become increasingly sophisticated in recent years, making diagnosing faults more difficult. Moreover, manual fault diagnosis is laborious and time-consuming. Since intelligent diagnostic methods have emerged recently as one of the most advanced and trendy approaches to solve these issues, resorting to intelligent fault diagnosis is a great choice and worthy of research [3]. For implementing signal analysis-based fault diagnosis in practice, the most generally used method is to extract and classify the main features utilizing data preprocessing and classification algorithms [4].

Many artificial intelligence techniques were applied in practical scenarios of industrial manufacturing, including k-nearest neighbor (K-NN) algorithms [5], Bayesian classifiers [6], support vector machines (SVMs) [7], artificial neural networks (ANNs) [8], and deep learning approaches most recently [9]. Among them, the convolutional neural network

(CNN) [10] showed outstanding performance in transfer learning-based fault diagnosis. In 2017, You et al. [11] proposed a CNN combined with support vector regression (SVR) which achieved accuracies of 93.9% and 97.6% for two separate datasets. For most classification models that only use a single dataset, the architecture of a CNN for feature extraction and other artificial intelligence methods for classification can provide high accuracy.

Many successful applications of machine learning algorithms are based on the precondition of a large amount of labeled training data and testing data in the same distribution. The imbalanced and limited data collected from practical systems may lead to low classification accuracies in bearing fault diagnosis for traditional machine learning methods. Meanwhile, the established machine learning model may become unsuitable when dealing with newly acquired data, since such data may not follow the same distribution as the training dataset. Nevertheless, the real-world challenge is the frequent lack of labeling and unhealthy data in bearing fault diagnosis. Under this situation, semi-supervised learning (SSL) can help to alleviate this difficulty by requiring some labeled data [12,13]. SSL is a machine learning task between supervised learning and unsupervised learning. Consistency regularization and pseudo-labeling are two dominant approaches in SSL, and low-density regions of the decision boundaries are a general presumption [14]. Compared with consistency regularization, which often requires numerous augmentation operations, pseudo-labeling can be used in most domains with high accuracy. The key inspiration of SSL is to filter the unlabeled instances with high confidence and use them for training the labeled data for the next iteration.

In some practical applications, even unlabeled data from the same domain can be challenging to obtain. Therefore, transfer learning is a promising technique for overcoming the challenge outlined above, as it is based on transferring knowledge across domains [15,16]. Transfer learning aims to increase model accuracy or reduce the number of labeled samples in the target domain by leveraging knowledge from the source domain [17,18]. In the area of transfer learning-based fault diagnosis, the feature spaces of the source and target domains are usually adopted by the maximum mean discrepancy (MMD) distance [19,20]. According to a review by Pan and Yang [21], the basic approaches to transfer learning can be divided into four categories: instance transfer, feature representation transfer, parameter transfer, and relation knowledge transfer. Moreover, the increasing popularization of deep neural networks prompted researchers to apply them to the subject of transfer learning.

In the beginning, most of the methodologies were based on pretrained recurrent neural networks [22]. When the generative adversarial nets (GAN) approach was first trained for solving transfer problems, it became a hot topic instantly for its remarkable performance. Yaroslav Ganin [23] first introduced an adversarial mechanism into the training of neural networks, known as domain-adversarial neural networks (DANNs). In this study, the learning objective of the network is that the feature generators are designed to help distinguish between the two domains as much as possible while preventing the discriminator from discriminating between the differences in the two domains. In 2019, Yu et al. [24] extended the concept of dynamic distribution adaptation to GAN and presented dynamic adversarial adaptation networks (DAANs) to solve the issue of mismatched contributions of the marginal (global) and conditional (local) distributions between domains. Figure 1 illustrates the different effects of marginal and conditional distributions in transfer learning applications. The marginal distribution influences more when two domains are substantially distinct (Source vs. Target I). In contrast, the conditional distribution should be prioritized when the global distributions are closer (Source vs. Target II).
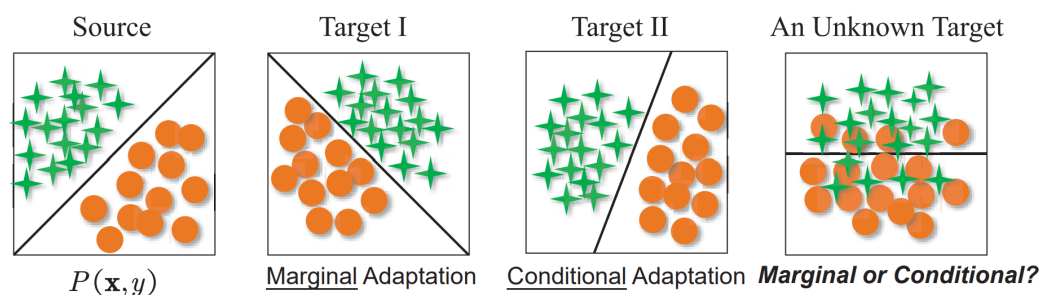
Figure 1. Different effects of marginal and conditional distributions in transfer learning applications [24].

However, to guarantee the success of such domain adaptation methods, there should be abundant labeled data in the domains, which is always impractical in actual working conditions. Collecting enough data also increases the cost of time and effort in fault diagnosis. The data imbalance in diagnosing machinery bearing faults can be outlined in two aspects: the data imbalance between normal and abnormal samples and the insufficient amount of data in settings with different specified external or internal operating parameters. Consequently, this method aims at solving these two problems based on transfer semi-supervised learning. More specifically, semi-supervised learning focuses on the first issue through pseudo-labeling. In contrast, transfer learning addresses the second aspect by transferring knowledge from another different dataset [25,26].

As shown in Figure 2, the traditional pseudo-labeling usually involves feeding a small amount of labeled data into the model for initial training and then feeding the unlabeled data into the model for classification [14]. When the confidence of predicting whether a sample belongs to a class exceeds the predetermined threshold, the sample is given the corresponding pseudo-label. Alternatively, the class in which the maximum confidence of the model prediction belongs is directly selected as the pseudo-label. The pseudo-label is added to the original training dataset as if it is labeled for retraining.
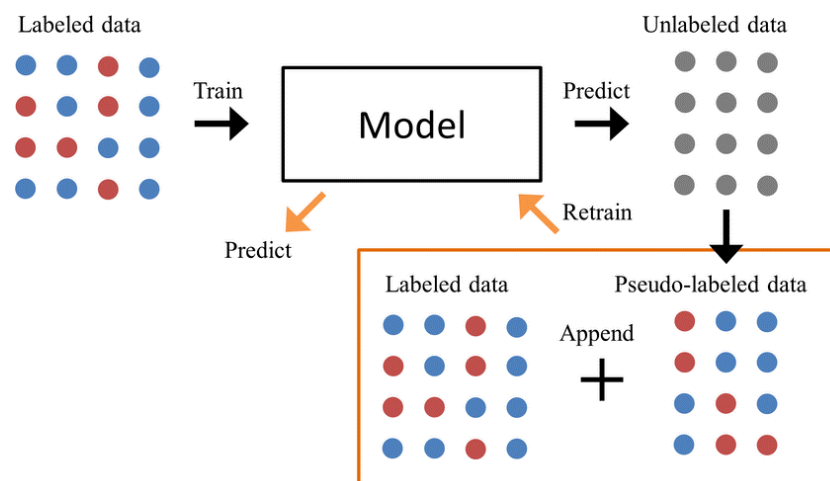

Figure 2. Traditional pseudo-labeling [27].

However, this approach often suffers from the problem that the pseudo-labels have a high confidence level regardless of whether the samples are correctly labeled or not. Suppose massive unlabeled samples are mislabeled and used for training. In that case, this will result in many noisy samples in the training set, which will affect the performance significantly. It is not sufficient to use the confidence of the softmax layer as the only basis for filtering. Uncertainty-aware pseudo-label selection (UPS) is an effective semi-supervised learning framework that introduced negative learning and uncertainty estimation with expected calibration error (ECE) into conventional pseudo-labeling method [14,28]. Its

performance surpasses consistency regularization in many tasks, which is another primary SSL method.

In conclusion, a novel framework, the uncertainty-aware pseudo-label selection (UPS) model with a DANN, is proposed based on the concept of a semi-supervised learning-generative adversarial network to overcome the mentioned problems of imbalanced data. The main contributions of this paper are as follows:

1. A hybrid UPS model with a DANN is proposed with a variable ratio to improve accuracy and robustness;
2. Unlabeled data are labeled with pseudo-labels to enlarge the labeled target dataset;
3. The proposed method is successfully verified in the analysis of the bearing fault diagnosis task on the Case Western Reserve University (CWRU) dataset and Xi'an Jiaotong University-Sumyoung (XJTU-SY) dataset, where the diagnosis accuracy is proven to be higher than other well-known fault diagnosis methods.

The structure of this paper is as follows. Section 2 introduces the data preprocessing and the proposed method based on UPS and a DANN. Section 3 presents the experiments and illustrates the results by comparing them with independent approaches. Section 4 concludes the paper.

2. Materials and Methods

2.1. Data Preprocessing

The short-time Fourier transform (STFT) plays a significant role in preprocessing the raw signal data. A Fourier transform is a traditional method to transform the time domain signal into a frequency domain signal. It has a limitation in that it lacks the temporal resolution for the time domain signals. The STFT applies the window and shifts it so that it has a fixed temporal resolution for the time domain signal, which constructs the spectrogram for the subsequent data input. The main formula of an STFT is

$$X(n_0, \omega) = \sum_{n=-\infty}^{\infty} x(n)w(n_0 - n)e^{-jwn} \tag{1}$$

where $x(n)$ is the discrete signal sequence, $w(n)$ is the analysis window, n_0 is the window center, and ω is a continuous variable-denoting frequency. $X(n_0, \omega)$ is a frequency function of the time section n_0. Then, the window slides to obtain $X(n_0 + s, \omega)$ where s is the hop size and obtains the STFT result of the next section. Lastly, the frequency results are combined in chronological order to form a complete spectrogram. That aside, the window function is a Hann window.

The network will be experimented upon through transferring from the CWRU dataset [29] to the XJTU-SY dataset [30]. Figure 3 shows some examples of the CWRU data. Bearing faults in the CWRU dataset are artificially created in different areas of the bearing, and the data are recorded at different sampling rates. Therefore, the vibration figure tends to be regular and periodic in its amplitude along with the time series.

However, the XJTU-SY dataset contains the full life cycle of bearing degeneration. As shown in Figure 4, demonstrating an example of XJTU-SY with obvious transition characteristics, the whole process can be split into three phases by observing the sudden change between them. The first phase is the normal vibration data of the bearing, so the amplitude usually stabilizes within a low-value range. The second phase is the vibration data when the bearing starts to degenerate. During this phase, the amplitude will fluctuate more heavily and sometimes gradually increase over time. The third phase is the vibration data when it is completely damaged. As a result, the amplitude will continue to rise more markedly, eventually reaching a very high level. Nevertheless, for some of the cases shown in Figure 5, the degenerative process is gradual, while in others it may be sharp. The degeneration of the second phase may not be evident and can therefore be ignored, allowing focusing on the first phase and third phase. The data in the first phase are labeled as normal data, and the third phase's data are labeled as fault data.
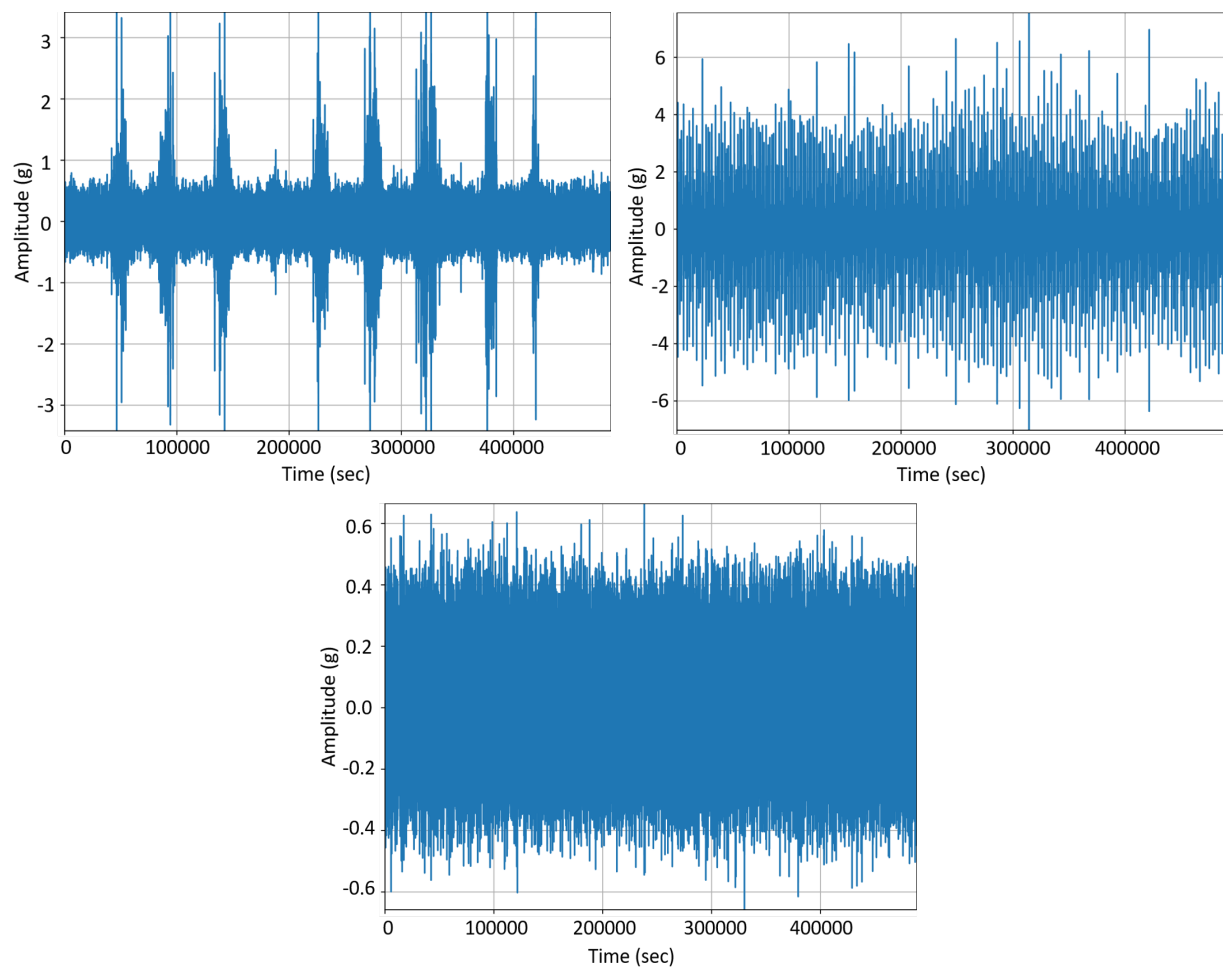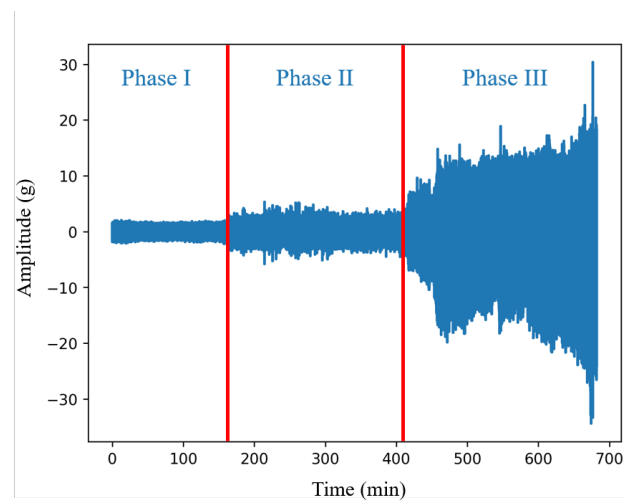
Figure 3. Examples of CWRU dataset.

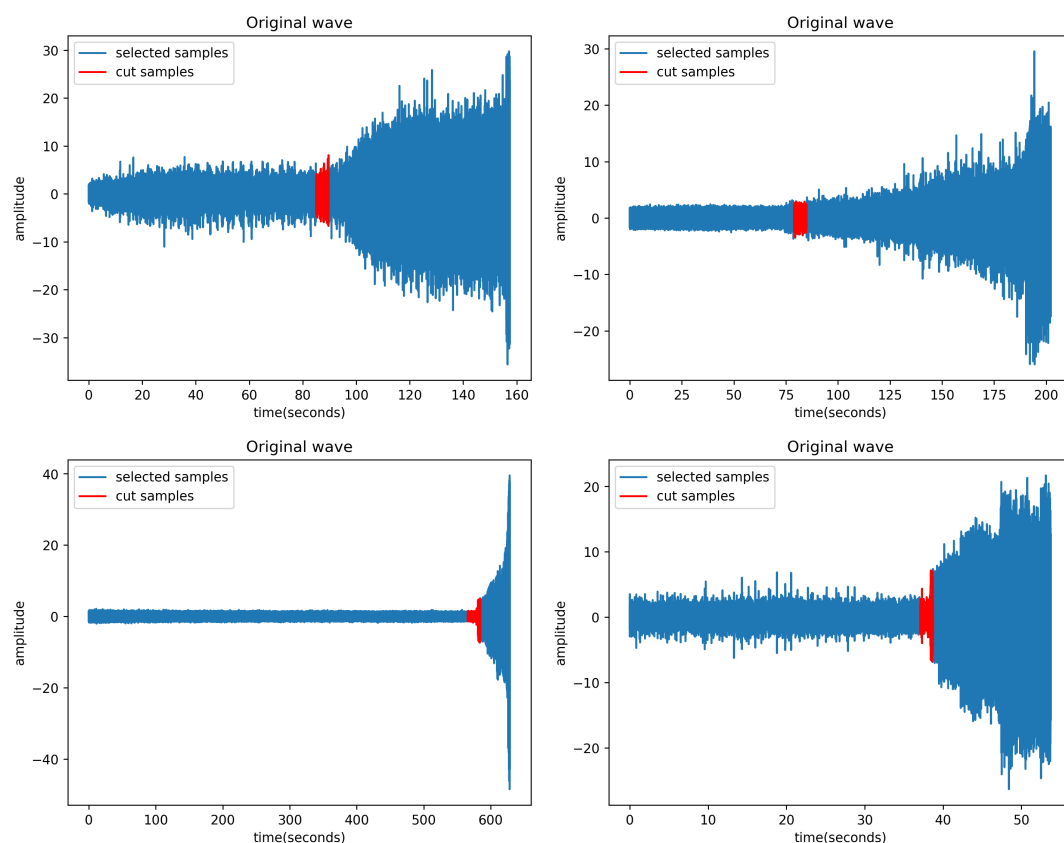**Figure 4.** Full life cycle vibration signals of XJTU-SY dataset.

Figure 5. Samples of separating XJTU-SY phases.

In conclusion, the two datasets both have commonalities and differences which determine why the transfer from CWRU to XJTU-SY was chosen. The common elements are that they are both bearing vibration data and have some of the same fault classes. Nevertheless, compared with the CWRU dataset, the XJTU-SY dataset is closer to the actual working conditions where bearings will gradually degenerate, but the drawback is the small amount of data. The CWRU dataset, by contrast, contains a larger amount of data, but the data are recorded in a different environment, and the bearing faults are artificially created. Therefore, the CWRU dataset was transferred into the form of the XJTU-SY dataset to solve the data imbalance problem. Table 1 demonstrates the differences between the CWRU and XJTU-SY datasets from six perspectives in detail.

Table 1. Comparison of CWRU and XJTU-SY datasets.

	CWRU	XJTU-SY
Working Condition	(1) 1797 rpm (2) 1772 rpm 3) 1750 rpm (4) 1730 rpm	(1) 2100 rpm (35 Hz) and 12 kN (2) 2250 rpm (37.5 Hz) and 11 kN (3) 2400 rpm (40 Hz) and 10 kN
Degeneration Process	No	Yes
Sample Frequency	12 kHz 48 kHz	25.6 kHz
Vibration Signals in Each Sample	Around 122,000	Depends on bearing's lifetime
Fault Element	Inner race, ball, outer race	Inner race, ball, cage, and outer race
Fault type	Single fault element	Multiple fault elements

2.2. Proposed Method Based on UPS and a DANN

2.2.1. Negative Learning

The negative learning (NL) proposed by [31] is used mainly to obtain good initialization of the network for learning with noisy labels. In this approach, a network is first trained by randomly generating negative labels (NL step) and then using that network to selectively generate negative labels using confidence scores (SelNL). The selective positive learning (SelPL) they used also relied on creating positive pseudo-labels based on confidence. Different from the NL in [31], the NL in UPS is designed to include additional unlabeled samples into the training phase and generalize pseudo-labeling for multi-label classification settings. In a trained network, the sample x^i outputs the probability $p^{(i)}$, and $p_c^{(i)}$ refers to the probability of class c. Similar to one-hot encoding in traditional multi-classification problems, it can be converted to a $1 \times C$-dimensional label consisting of that class of labels. Therefore, the pseudo-labels $\tilde{y}_c^{(i)}$ of sample $x^{(i)}$ are computed as follows:

$$\tilde{y}_c^{(i)} = \mathbb{1}\left[p_c^{(i)} \geq \gamma\right] \qquad (2)$$

where $\gamma \in (0,1)$ is the threshold for labels, which is hard to determine. The binary vector represents the pseudo-labels selected as

$$g_c^{(i)} = \mathbb{1}\left[p_c^{(i)} \geq \tau_p\right] + \mathbb{1}\left[p_c^{(i)} \leq \tau_n\right] \qquad (3)$$

When $\tilde{y}_c^{(i)}$ is chosen, $g_c^{(i)} = 0$; otherwise, $g_c^{(i)} = 1$. τ_p is the confidence threshold for positive labels, and τ_n is the confidence threshold for negative ones. Cross-entropy loss is estimated for the samples with selected positive pseudo-labels for single-label classification. When no positive label is selected, negative learning is used with negative cross-entropy loss. The expression of negative learning is

$$L_{\text{NCE}}\left(\tilde{y}^{(i)}, \hat{y}^{(i)}, g^{(i)}\right) = -\frac{1}{s^{(i)}} \sum_{c=1}^{C} g_c^{(i)}\left(1 - \tilde{y}_c^{(i)}\right) \log\left(1 - \hat{y}_c^{(i)}\right) \qquad (4)$$

where $s^{(i)}$ is the number of selected pseudo-labels for sample i. As a result, even if the model is not confident enough about whether the sample belongs to a class, it can help improve the accuracy of the diagnosis by disproving with a low probability that the sample most probably would not belong to a class.

2.2.2. Uncertainty Estimation

The experiment results show that the ECE score has a positive correlation with the prediction uncertainty, which implies that the model with a lower uncertainty is inclined to have a more significant calibration capability [14]. The uncertainty of the output value can be calculated as an alternative confidence level for selecting reliable pseudo-labeled samples. ECE is a standard metric for evaluating the calibration capability of a classifier, which can be obtained as follows:

$$ECE = \sum_{l=1}^{L} \frac{1}{|D|}\left|\sum_{x^{(i)} \in I_l} \max_c \hat{y}_c^{(i)} - \sum_{x^{(i)} \in I_l} \mathbb{1}\left[\arg\max_c \hat{y}_c^{(i)} = \arg\max_c \tilde{y}_c^{(i)}\right]\right| \qquad (5)$$

where the confidence predictions on dataset D are divided into L bins that are evenly spaced, and the samples in a particular bin l are referred to as I_l.

Hence, a more reliable subset of pseudo-labels is used in training by considering both the confidence and uncertainty of a network prediction. Now, Equation (2) can be reformulated as

$$g_c^{(i)} = \mathbb{1}\left[u\left(p_c^{(i)}\right) \leq \kappa_p\right]\mathbb{1}\left[p_c^{(i)} \geq \tau_p\right] + \mathbb{1}\left[u\left(p_c^{(i)}\right) \leq \kappa_n\right]\mathbb{1}\left[p_c^{(i)} \leq \tau_n\right] \qquad (6)$$

where $u(p)$ is the uncertainty of a prediction p while κ_p and κ_n are the uncertainty thresholds.

2.2.3. Model Structure

Figure 6 illustrates the proposed network in this paper: a deep neural network combining a DANN and UPS, where the DANN can support UPS to filter pseudo-labels more robustly and new labels can expand the labeled target data to make the distribution of the source and target domain closer. The DANN model takes all input data for training. The UPS model takes data from the target domain for the pseudo-label selection with uncertainty awareness:

$$CE_{pl} = CE_d \cdot \alpha + CE_s \cdot (1 - \alpha) \tag{7}$$

The parameter α is adaptive and can be learned by gradient descent. CE_d and CE_s are the cross-entropy of the DANN and UPS, and CE_{pl} is the cross-entropy of the pseudo-labels. They will be weighted by α before the cross-entropy layer and then passed to select high-confidence instances as the new samples of labeled target data.
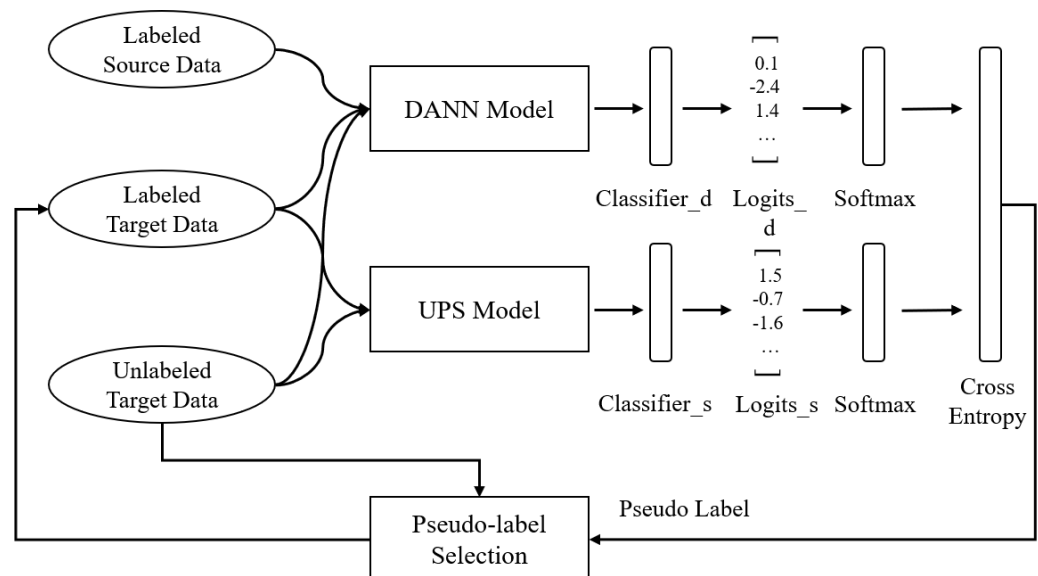

Figure 6. Model structure of proposed network based on DANN and UPS.

3. Experiment and Result Analysis

3.1. Experiment Set-Up

The data from the CWRU dataset is the source-labeled data. The data from the XJTU-SY dataset as target domain consists of only a small amount of labeled data and a large amount of unlabeled data for transfer semi-supervised learning. Each sample of data was split into 240 portions every 2000 data collection spots. The data size in total for the CWRU dataset was 480,000 for each class, and the data size for the XJTU-SY dataset was 672,000. Table 2 shows the selection of data used for training in detail.

The source dataset, created by the Bearing Data Center of Case Western Reserve University (CWRU), is the most widely cited standard dataset for current research on signal processing and fault diagnosis of bearing vibration [29]. It is also considered the primary dataset for training network models and testing network performance. Electro-discharge machining (EDM) was used artificially to induce single-point faults to the test bearings with fault diameters of 7 mils, 14 mils, and 21 mils. Each class of fault diameter was introduced separately at the inner race, ball, and outer race [32]. By changing the bearing diameter, fault location, motor load and speed, and sampling frequency, the experiment generated a variety of valid data in a limited number of practical machines. Considering the balance

of data and the common fault element with the CWRU dataset, three fault labels were selected: normal bearings, inner fault bearings, and outer fault bearings (Table 3).

Table 2. Data selection of three classes.

Class	Label	Source/Target	Labeled or Unlabeled	Data Size
Inner Race	0	Source	Labeled	480,000
		Target	Labeled	120,000
			Unlabeled	552,000
Outer Race	1	Source	Labeled	480,000
		Target	Labeled	120,000
			Unlabeled	552,000
Normal	2	Source	Labeled	480,000
		Target	Labeled	120,000
			Unlabeled	552,000

Table 3. Bearings selected from CWRU dataset.

Diameter	Load (HP)	Motor Speed (rpm)	File Name	Fault Element
0.007″	3	1730	IR007_3	Inner Race
0.014″	3	1730	IR014_3	Inner Race
0.021″	3	1730	IR021_3	Inner Race
0.007″	3	1730	OR007@6_3	Outer Race
0.014″	3	1730	OR014@6_3	Outer Race
0.021″	3	1730	OR021@6_3	Outer Race
-	-	-	Normal_1	-
-	-	-	Normal_2	-
-	-	-	Normal_3	-

The target XJTU-SY bearing dataset [30] was acquired from accelerated degeneration experiments of rolling element bearings with 15 bearings under 3 operating conditions. Due to the different working conditions of different bearings, the service life of the bearings can also vary significantly, which means the data are highly imbalanced. The fault elements include single and multiple points, specifically the inner race, outer race, and cage for a different bearing lifetime. Based on the different degeneration performances, 8 bearings from the XJTU-SY dataset were selected and split accordingly as the target domain and the same three classes of data from the CWRU dataset as the source domain (Table 4).

Table 4. Bearings selected from XJTU-SY dataset.

Bearing	Fault Element	Normal Range	Fault Range
Bearing 2_1_37.5 Hz	Inner	1-452	454-484
Bearing 2_2_37.5 Hz	Outer	1–50	51–159
Bearing 2_4_37.5 Hz	Outer	1–30	31–40
Bearing 2_5_37.5 Hz	Outer	1–120	121–337
Bearing 3_1_40 Hz	Outer	1–2463	2464–2536
Bearing 3_3_40 Hz	Inner	1–340	341–369
Bearing 3_4_40 Hz	Inner	1–1416	1417–1514
Bearing 3_5_40 Hz	Outer	1–10	11–110

3.2. Results and Discussion

In general, the overall accuracy of the proposed method in the test dataset can reach up to 99% after 50 epochs (Figure 7), which indicates the ability to transfer the model. The accuracy here is defined as

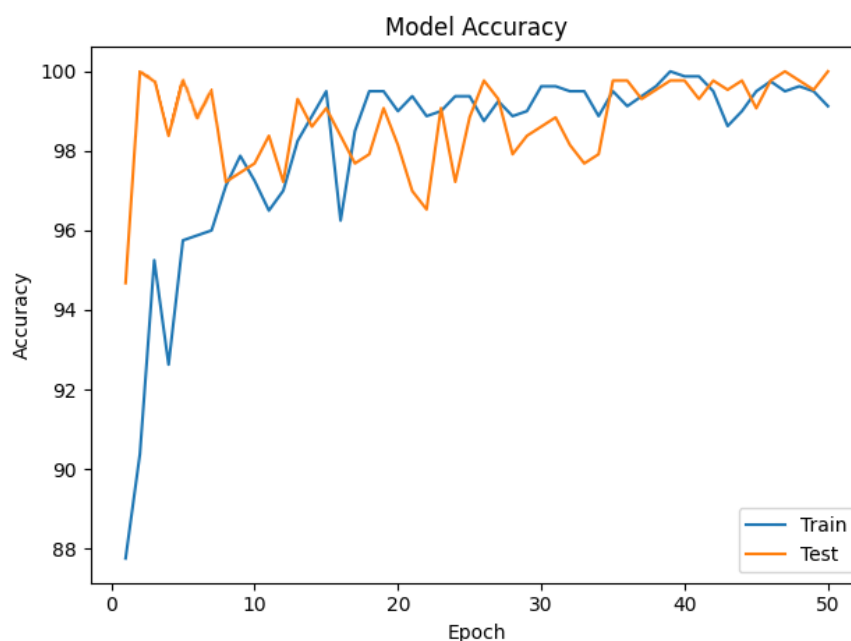$$accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{8}$$

Figure 7. Model accuracy of UPS + DANN.

It is interesting to note that the test accuracy was even higher than the training accuracy at the first 10 epochs. Since there was no data leakage in the validation set, and the split of the training-test data set was completely random, it is speculated that the reason for this phenomenon may be that the noise in the training set was greater than that in the validation set. The data augmentation by pseudo-labeling made the training data more complex than the test data, and the model was not able to fully memorize the training data.

Table 5 illustrates the performances of the proposed method and other methods. Baseline refers to predicting the test data of the XJTU-SY dataset with the model trained by the CWRU dataset directly. This model has no transfer learning to bring the distributions of the source and target domains into proximity and also no pseudo-labeling to extend the imbalanced training data. Therefore, it is noticed that the baseline only preserved a test accuracy of 23%. When it comes to transferring technologies solely by transferring from the CWRU dataset to the XJTU-SY dataset, UPS and the DANN provided average test accuracies of 42% and 56%, respectively. This performance was twice as high as that of the baseline model, which means that both popular transfer learning models can improve the accuracy considerably at first. When using only semi-supervised learning by using UPS to train the pseudo-labels, the average test accuracy showed an increase of up to 76%. This indicates the remarkable power of uncertainty-aware pseudo-labeling and proves the ability of semi-supervised learning in resolving the problem of a deficiency of labeled data. Ultimately, when combining transfer learning and semi-supervised learning, both methods can improve the accuracy of UPS alone, but the final proposed model, UPS + DANN, showed a greater average test accuracy of 96% compared with UPS + DAAN at 90%.

Figure 8 depicts the confusion matrices of six different models, with the rows indicating true labels and columns indicating predicted labels. The percentage of each type of feature is shown in each cell in the confusion matrix. Among the three classes, the precision of the

outer race was the highest in most models, and UPS was incompetent at predicting normal data specifically. Moreover, the performance of the DANN was more average for all classes compared with the DAAN. It can also be observed clearly that UPS + DANN preserved a relatively high accuracy, especially in the inner race and outer race classes.

Table 5. Test accuracy comparison.

Model	Best Test Acc.	Average Test Acc.
Baseline	29.84%	23.45%
DAAN	45.67%	42.33%
DANN	60.72%	56.88%
UPS	84.21%	76.35%
UPS + DAAN	96.43%	90.20%
UPS + DANN	**99.63%**	**96.77%**

(a)
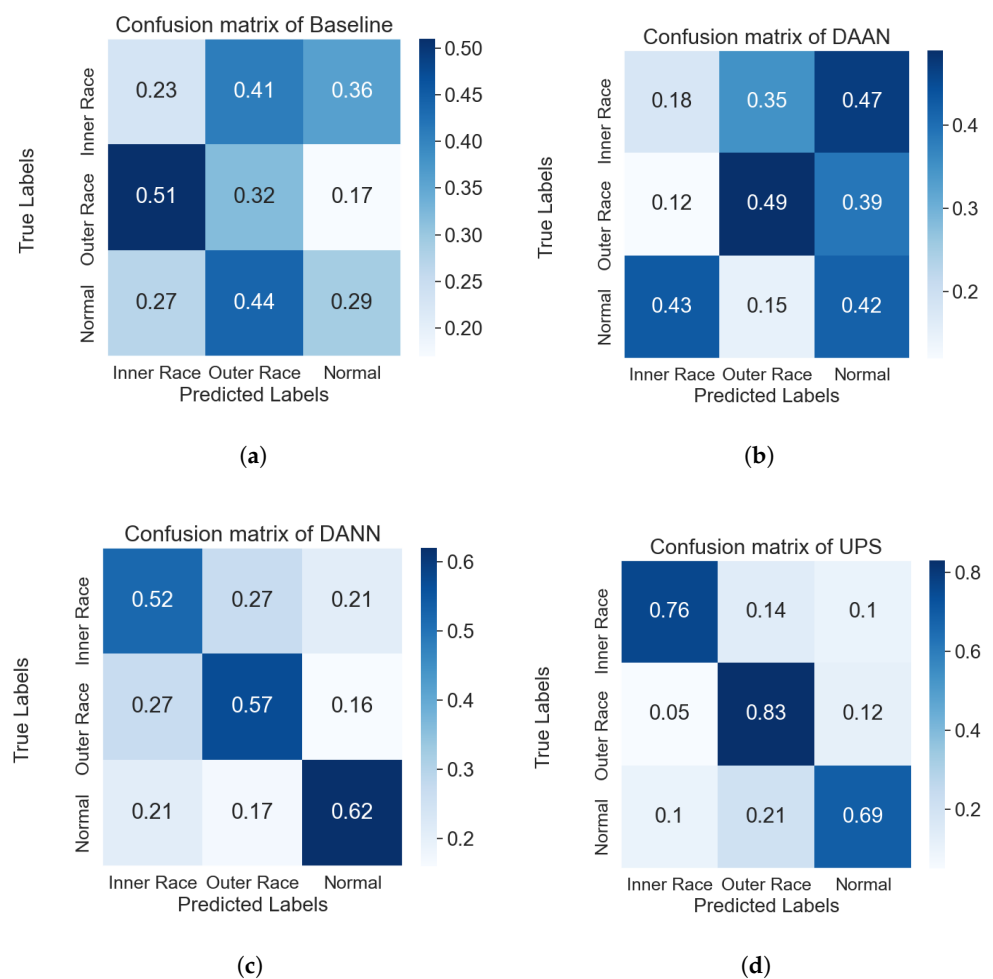
(b)

(c)
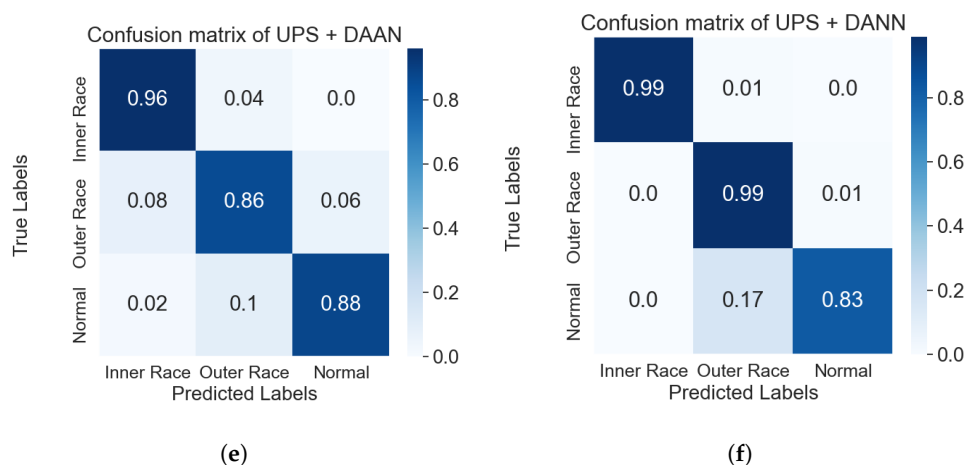
(d)

Figure 8. *Cont.*

(**e**)

(**f**)

Figure 8. Confusion matrices of different models: (**a**) confusion matrix of baseline, (**b**) confusion matrix of DAAN, (**c**) confusion matrix of DANN, (**d**) confusion matrix of UPS, (**e**) confusion matrix of UPS + DAAN, and (**f**) confusion matrix of UPS + DANN.

4. Conclusions

This paper proposes a method based on a DANN and UPS for fault diagnosis of imbalanced machinery bearings. This model combines the advantages of semi-supervised and transfer learning and makes them reinforce each other. Uncertainty-aware pseudo-label selection is used to balance data between labeled and unlabeled. A domain-adversarial neural network complements the target domain via transferring from the source domain. To demonstrate the efficacy of the proposed method, experiments from two different datasets for transfer learning were performed. Compared with the independent approaches, including a DANN, DAAN, and UPS, the outcomes were correspondingly superior. Some further research directions can be undertaken in the future: (1) applying heterogeneous transfer learning to predict the label of the target domain, which never appeared in the source domain, and (2) reducing the proportion of labeled data and testing the robustness of the model repetitively.

Author Contributions: Conceptualization, X.Z., R.Y., H.W., M.D., P.Y., S.W. and H.S.; methodology, X.Z., R.Y., H.W., M.D., P.Y., S.W. and H.S.; software, X.Z., H.W., M.D., P.Y., S.W. and H.S.; validation, X.Z., H.W., M.D., P.Y., S.W. and H.S.; formal analysis, X.Z., H.W., M.D., P.Y., S.W. and H.S.; investigation, X.Z., H.W., M.D., P.Y., S.W. and H.S.; resources, R.Y.; data curation, X.Z., H.W., M.D., P.Y., S.W. and H.S.; writing—original draft preparation, X.Z., H.W., M.D., P.Y., S.W. and H.S.; writing— review and editing, X.Z. and R.Y.; visualization, X.Z., H.W., M.D., P.Y., S.W. and H.S.; supervision, R.Y.; project administration, R.Y.; funding acquisition, R.Y. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Restrictions apply to the availability of these data. Data were obtained from Case Western Reserve University and Xi'an Jiaotong University and are open to access from http://engineering.case.edu/bearingdatacenter/download-data-file (accessed on 16 November 2021) and http://biaowang.tech/xjtu-sy-bearing-datasets (accessed on 16 November 2021) with the permission of Case Western Reserve University and Xi'an Jiaotong University, respectively.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhang, Y.; Yang, R.; Huang, M.; Han, Y.; Wang, Y.; Di, Y.; Su, D.; Lu, Q. A simultaneous fault diagnosis method based on cohesion evaluation and improved BP-MLL for rotating machinery. *Shock Vib.* **2021**, *2021*, 7469691. [CrossRef]
2. Wu, L.; Yao, B.; Peng, Z.; Guan, Y. Fault diagnosis of roller bearings based on a wavelet neural network and manifold learning. *Appl. Sci.* **2017**, *7*, 158. [CrossRef]
3. Yang, R.; Zhong, M. *Machine Learning-Based Fault Diagnosis for Industrial Engineering Systems*; CRC Press: Boca Raton, FL, USA, 2022.
4. Lu, Q.; Yang, R.; Zhong, M.; Wang, Y. An improved fault diagnosis method of rotating machinery using sensitive features and RLS-BP neural network. *IEEE Trans. Instrum. Meas.* **2019**, *69*, 1585–1593. [CrossRef]
5. Wang, D. K-nearest neighbors based methods for identification of different gear crack levels under different motor speeds and loads: Revisited. *Mech. Syst. Signal Process.* **2016**, *70*, 201–208. [CrossRef]
6. Baraldi, P.; Podofillini, L.; Mkrtchyan, L.; Zio, E.; Dang, V.N. Comparing the treatment of uncertainty in Bayesian networks and fuzzy expert systems used for a human reliability analysis application. *Reliab. Eng. Syst. Saf.* **2015**, *138*, 176–193. [CrossRef]
7. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1999.
8. Lei, Y.; Jia, F.; Lin, J.; Xing, S.; Ding, S.X. An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. *IEEE Trans. Ind. Electron.* **2016**, *63*, 3137–3147. [CrossRef]
9. Liu, R.; Yang, B.; Zio, E.; Chen, X. Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mech. Syst. Signal Process.* **2018**, *108*, 33–47. [CrossRef]
10. LeCun, Y.; Boser, B.; Denker, J.; Henderson, D.; Howard, R.; Hubbard, W.; Jackel, L. Handwritten digit recognition with a back-propagation network. In *Advances in Neural Information Processing Systems 2 (NIPS 1989)*; Neural Information Processing Systems: San Diego, CA, USA, 1989; pp. 396–404.
11. You, W.; Shen, C.; Guo, X.; Jiang, X.; Shi, J.; Zhu, Z. A hybrid technique based on convolutional neural network and support vector regression for intelligent diagnosis of rotating machinery. *Adv. Mech. Eng.* **2017**, *9*, 1687814017704146. [CrossRef]
12. Ouali, Y.; Hudelot, C.; Tami, M. An overview of deep semi-supervised learning. *arXiv* **2020**, arXiv:2006.05278.
13. Lee, D.H. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. *Workshop Challenges Represent. Learn. ICML* **2013**, *3*, 896.
14. Rizve, M.N.; Duarte, K.; Rawat, Y.S.; Shah, M. In defense of pseudo-labeling: An uncertainty-aware pseudo-label selection framework for semi-supervised learning. *arXiv* **2021**, arXiv:2101.06329.
15. Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A comprehensive survey on transfer learning. *Proc. IEEE* **2020**, *109*, 43–76. [CrossRef]
16. Wan, Z.; Yang, R.; Huang, M.; Zeng, N.; Liu, X. A review on transfer learning in EEG signal analysis. *Neurocomputing* **2021**, *421*, 1–14. [CrossRef]
17. Wan, Z.; Yang, R.; Huang, M. Deep transfer learning-based fault diagnosis for gearbox under complex working conditions. *Shock Vib.* **2020**, *2020*, 8884179. [CrossRef]
18. Wan, Z.; Yang, R.; Huang, M.; Liu, W.; Zeng, N. EEG fading data classification based on improved manifold learning with adaptive neighborhood selection. *Neurocomputing* **2022**, *482*, 186–196. [CrossRef]
19. Mao, W.; Liu, Y.; Ding, L.; Safian, A.; Liang, X. A new structured domain adversarial neural network for transfer fault diagnosis of rolling bearings under different working conditions. *IEEE Trans. Instrum. Meas.* **2020**, *70*, 1–13. [CrossRef]
20. Wang, X.; Yang, R.; Huang, M. An unsupervised deep-transfer-learning-based motor imagery EEG classification scheme for brain-computer interface. *Sensors* **2022**, *22*, 2241. [CrossRef]
21. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2009**, *22*, 1345–1359. [CrossRef]
22. Yang, R.; Huang, M.; Lu, Q.; Zhong, M. Rotating machinery fault diagnosis using long-short-term memory recurrent neural network. *IFAC-PapersOnLine* **2018**, *51*, 228–232. [CrossRef]

23. Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; Marchand, M.; Lempitsky, V. Domain-adversarial training of neural networks. *J. Mach. Learn. Res.* **2016**, *17*, 2030–2096.

24. Yu, C.; Wang, J.; Chen, Y.; Huang, M. Transfer learning with dynamic adversarial adaptation network. In Proceedings of the 2019 IEEE International Conference on Data Mining (ICDM), Beijing, China, 8–11 November 2019; pp. 778–786.

25. Chen, S.; Yang, R.; Zhong, M. Graph-based semi-supervised random forest for rotating machinery gearbox fault diagnosis. *Control Eng. Pract.* **2021**, *117*, 104952. [CrossRef]

26. Yang, Z.; Yang, R.; Huang, M. Rolling bearing incipient fault diagnosis method based on improved transfer learning with hybrid feature extraction. *Sensors* **2021**, *21*, 7894. [CrossRef] [PubMed]

27. Kim, S.; Lee, Y.; Tama, B.; Lee, S. Reliability-enhanced camera lens module classification using semi-supervised regression method. *Appl. Sci.* **2020**, *10*, 3832. [CrossRef]

28. Naeini, M.P.; Cooper, G.; Hauskrecht, M. Obtaining well calibrated probabilities using Bayesian binning. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–30 January 2015.

29. Case Western Reserve University Bearing Data CenterWebsite. Available online: https://engineering.case.edu/bearingdatacenter (accessed on 5 May 2022).

30. Wang, B.; Lei, Y.; Li, N.; Li, N. A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. *IEEE Trans. Reliab.* **2018**, *69*, 401–412. [CrossRef]

31. Kim, Y.; Yim, J.; Yun, J.; Kim, J. Nlnl: Negative learning for noisy labels. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 101–110.

32. Zhang, R.; Tao, H.; Wu, L.; Guan, Y. Transfer learning with neural networks for bearing fault diagnosis in changing working conditions. *IEEE Access* **2017**, *5*, 14347–14357. [CrossRef]