IOP Publishing Measurement Science and Technology
Meas. Sci. Technol. **31** (2020) 055601 (11pp) https://doi.org/10.1088/1361-6501/ab64aa

Deep learning-based cross-domain adaptation for gearbox fault diagnosis under variable speed conditions

Jaskaran Singh[1,2], **Moslem Azamfar**[1], **Abhijeet Ainapure**[1] and **Jay Lee**[1]

[1] NSF Industry/University Cooperative Research Center on Intelligent Maintenance Systems (IMS), University of Cincinnati, Cincinnati, OH 45221-0072, United States of America
[2] Department of Mechanical Engineering, Thapar Institute of Engineering and Technology, Patiala 147004, Punjab, India

E-mail: singh2jn@ucmail.uc.edu, jaskaran.singh@thapar.edu, azamfamm@mail.uc.edu, ainapuar@mail.uc.edu and lj2@ucmail.uc.edu

Received 9 September 2019, revised 20 November 2019
Accepted for publication 20 December 2019
Published 6 February 2020

Abstract

Existing intelligent gearbox fault diagnosis approaches have two shortcomings: (a) their performance is mostly confined to manual handcrafted features, and (b) they follow a general assumption that the distribution of the data in the source domain (labeled data on which the model is trained) is similar to the target domain (unlabeled data on which the model is tested), which might not be the case in real-world applications. Substantial human expertise and domain knowledge is required for manual feature extraction, and moreover, deploying the same model for a target domain whose distribution is different from the source domain would lead to poor generalization. Since deep learning methods can automatically learn high dimensional feature representations from raw measurement data, this paper proposes a novel deep learning-based domain adaptation (DA) method for gearbox fault diagnosis under significant speed variations. A deep convolutional neural network is used as the main architecture. The paper proposes to minimize the summation of cross-entropy loss (between the labeled source domain data) and maximum mean discrepancy loss (between the labeled source and unlabeled target datasets) simultaneously to adapt the source domain model to be applied in the target domain. The proposed deep learning DA approach is evaluated using experimental data from a gearbox under variable speeds and multiple health conditions. An appropriate benchmarking with both traditional machine learning methods and other DA methods demonstrate the superiority of the proposed method.

Keywords: deep learning, domain adaptation, gearbox health monitoring, fault diagnosis

(Some figures may appear in colour only in the online journal)

 © 2020 IOP Publishing Ltd Printed in the UK

1. Introduction

The gearbox is a key mechanical component in any modern rotating machinery and its performance is closely related to the operation safety. To guarantee continuous operation without service interruption, real-time monitoring methods for gearbox fault diagnosis have received considerable attention [1, 2]. The Industry 4.0 initiative has enabled increased industrial assets [3–6]. Various intelligent algorithms have been proposed for fault diagnosis in the past decade, such as *K*-nearest neighbor [7], self-organizing maps (SOMs) [8], artificial neural networks [9], support vector machines [10], etc. Since these methods require pre-research expertise to extract meaningful fault features from raw time domain vibration data, effective machinery fault detection and diagnosis is quite challenging under considerably complex working environments [11, 12]. Hence, automatic extraction of fault features from gearbox measurement signals without human interference is a necessary requirement [13]. For this reason, in recent years, deep learning methods have attracted greater attention from academia and industry, since they can automatically learn features from raw measurement data without relying on the identification of features by resorting to expert knowledge. However, evolving conditions in industrial environments such as variable operating conditions of a machine, maintenance interventions, upgrading technological and operational plans, etc. may deteriorate the fault detection performance of the trained algorithms [14, 15].

Considering the gearbox fault detection and diagnosis, for instance, the training samples (source domain) for developing the classifier model might be collected under one shaft speed/motor load. Nevertheless, the data in the actual testing conditions (target domain) might be from a gearbox operating under a different motor load/shaft speed. In this case, even though the system (gearbox) remains unchanged and so do the categories of possible fault types, the application environment (operating conditions) has changed. This simply implies that the feature distributions of both datasets are now dissimilar. Such a scenario, where the distribution of training data always differs from that of testing data, is very likely to be observed in practical industrial applications [16]. As a result, the intelligent diagnosis model built using the training samples (source domain data) will lead to poor generalization effects when applied directly to the target domain data. The difficulty faced by the source domain classifier model to correctly classify target domain samples happens due to the domain-shift problem [17]. This domain-shift problem is illustratively explained in figure 1. It can be observed from figure 1 that the source domain classifier model will have a tendency to misclassify the unlabeled data from a new operating condition (target domain) as it has not observed similar patterns in the training period. The fundamental reason for the observation of the domain-shift phenomenon in figure 1 is that the training and testing data are not following the same distribution or are not a part of the same feature space.

An intuitive solution to the above problem is to retrain the data-driven model by adding data under the new operating condition to the training set. Doing so implies that intelligent algorithms require periodic retraining of the tuned model. However, this solution is infeasible in real-world applications, since limited data might be available under new operating conditions, and often the fault data may be unlabeled. Since industrial machines are rarely run to failure to avoid catastrophic accidents, it is difficult to obtain labeled fault data from industrial machines. Hence, domain adaptation (DA) methods are implemented to tackle this problem of different feature distributions (domain shift). DA methods aim to adapt a classifier built using source domain data (labeled information) for use on a different but related target domain (unlabeled information). The fundamental working principle in most of the DA methods is to align both the distributions (source and target) in intermediate feature space.

Recently, many studies have been conducted in the development of DA algorithms, particularly in the area of object recognition, image classification, feature learning and natural language processing [18]. Zhang *et al* [19] implemented adaptive batch normalization (AdaBN) coupled with CNN to improve the DA ability on raw vibration signals for REB fault diagnosis. Zhang *et al* [20] proposed training a CNN using the concept of batch normalization with a very small batch size value, almost equal to the number of health classes. The authors validated the DA technique on the bearing dataset and observed that the generalization ability of the model trained with small batch size is higher. Guo *et al* [21] proposed a transfer learning method that consists of a 1D-CNN module for automatic feature learning and a DA part that maximizes domain recognition error and minimizes the probability distribution distance. The performance of their method was validated on transfer tasks between three different bearing datasets: CWRU [22], IMS [23], and Railway Locomotive [24]. Lu *et al* [25] used a deep neural network for feature learning and designed a DA framework that only uses the source domain data and normal category data from the target domain to accomplish DA tasks. In order to avoid loss of information from faulty samples, a weight regularization term was incorporated in the model. In the final stage, a SVM classifier was trained on transferable features from only source data and used for classification tasks on data from the target domain. Wen *et al* [26] proposed a DA architecture based on a three-layer sparse auto-encoder for automatic feature learning from raw vibration signal and minimizing the maximum mean discrepancy (MMD) term between the features from training and testing domains. The proposed DA approach was evaluated on a CWRU bearing dataset and reported a higher prediction accuracy compared to other methods. In [27], an acoustic spectral imaging technique was used to convert time-domain acoustic emission signals to representative images for different health conditions in low-speed bearings. The images are used in a CNN-based transfer learning model for predicting labels of target domain dataset. Cao *et al* [28] developed a deep CNN-based transfer learning method for gearbox fault diagnosis. In their approach, raw time domain vibration signal is converted to gray scaled images and used as an input to the CNN model. In their approach the first '*n*' layers of the CNN network are trained on the source dataset and fine-tuned on the target transfer tasks. Meanwhile the last (*m*-*n*) layers are trained using data from the new task. To the best of the

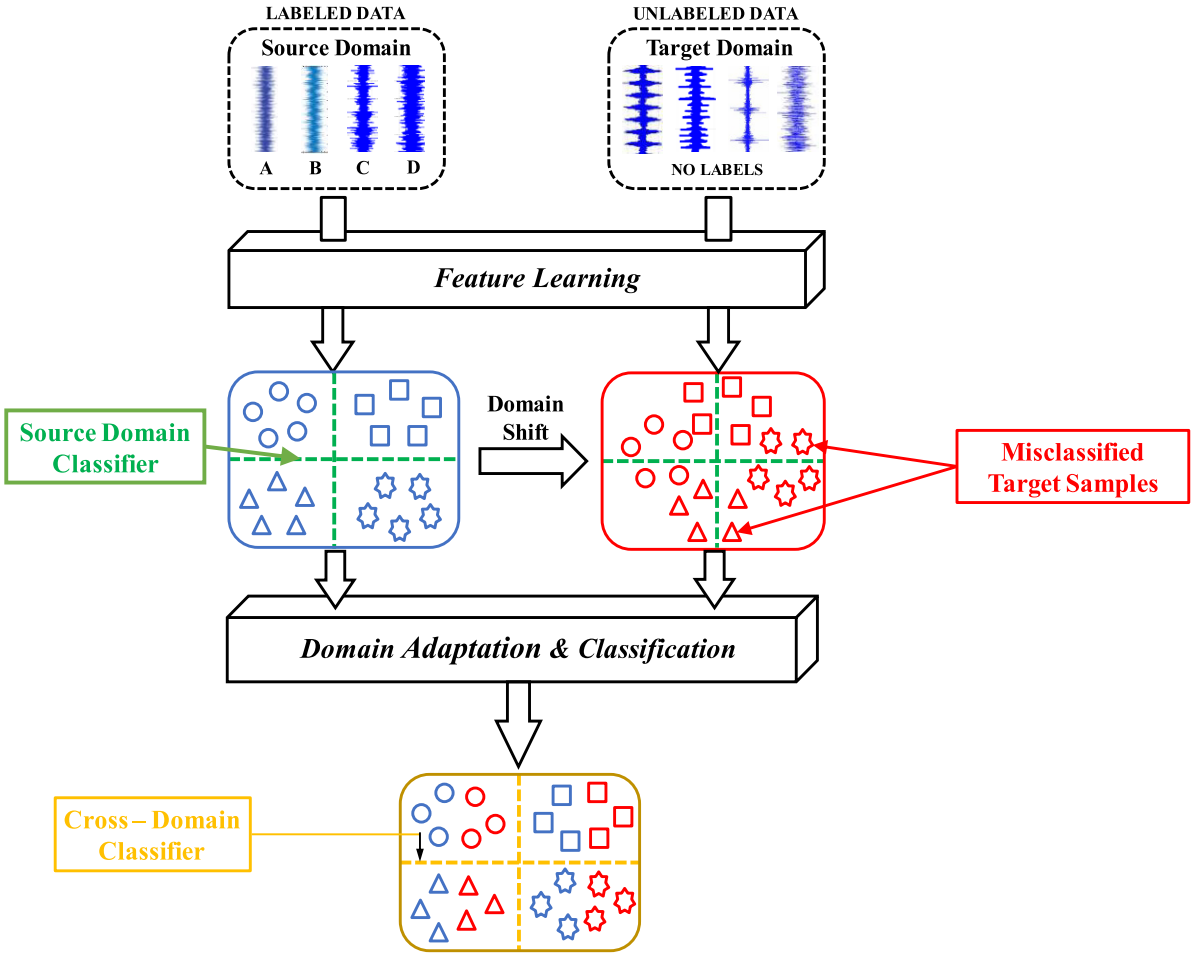

Figure 1. Illustrative example of domain shift and the DA technique.

authors' knowledge there is no study on the effectiveness of CNN-based transfer learning methods for gearbox fault diagnosis under significant variations in operating speed using vibration measurement signals.

Although some studies have been conducted investigating the use of DA algorithms for fault detection and diagnosis of rotating machinery components in industrial applications in recent years, most of the conducted studies have been developed particularly for rolling element bearings. Further, the majority of these studies have validated their proposed approach on a CWRU bearing dataset considering motor load for the transfer tasks. Studies considering significant variation in operating speeds for transfer tasks are very scarce. To the best of the authors' knowledge, there is no research exploring the effectiveness of transfer learning/DA techniques for gearbox fault detection and diagnosis. Therefore, this study explores the use of a deep learning-based cross-domain adaptation method for gearbox fault diagnosis under significant variations in operating speed using vibration measurement signals.

This study proposes a CNN based cross-domain adaptation method (CNN-CDA) for gearbox fault diagnosis to solve the above challenges. The core idea is to first use CNN as a feature extractor to obtain high-level feature representations from the raw time domain vibration data and then subsequently minimize the distribution discrepancy between labeled source domain and unlabeled target domain data for the cross-domain adaptation task. This paper proposes to use raw time-domain vibration data as an input to the CNN model for automatic feature extraction. A unified domain adaptation approach is proposed to accurately classify unlabeled target domain data by minimizing the summation of cross-entropy loss (between labeled source domain data) and MMD between labeled source domain and unlabeled target domain samples. Network architecture and tuning parameters are investigated to provide high classification accuracy in gearbox fault diagnosis under significant speed variations.

The rest of the paper is organized as follows. Section 2 gives details of some preliminaries. The domain adaptive CNN-MMD model is detailed in section 3. Section 4 discusses the details of the experimental set-up. In section 5, the developed model is evaluated on data from a gearbox test rig. Section 6 presents conclusions and future work.

2. Preliminaries

2.1. Domain-shift problem

To better understand the domain-shift problem, let us introduce a basic mathematical notation for the cross-domain adaptation

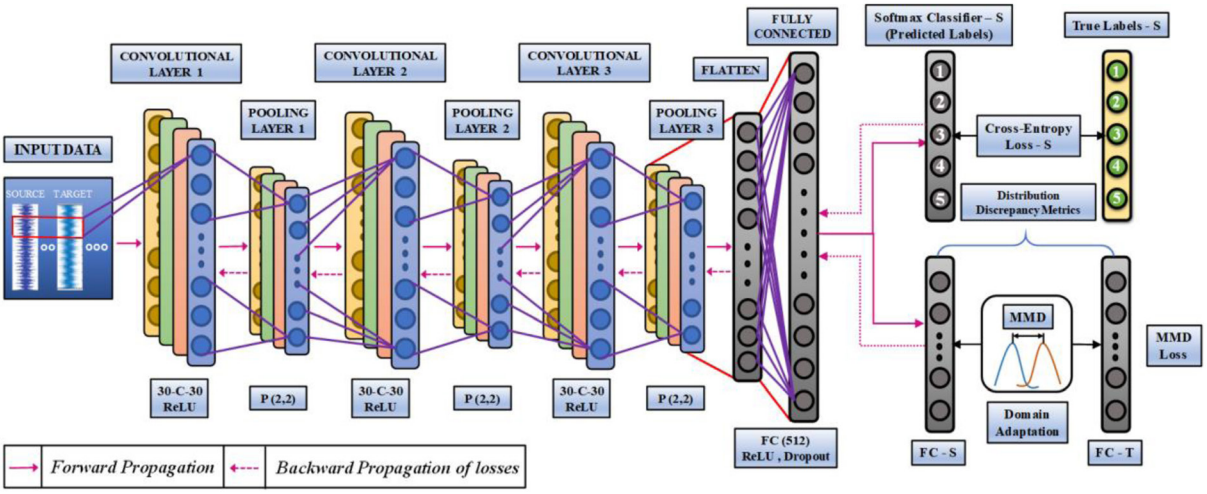

Figure 2. Proposed CNN-CDA model framework.

task. Let χ be a feature space, X be a particular sample, and $P(X)$ be a marginal probability distribution. Let $D = \{\chi, P(X)\} \infty$ be a domain, where $X = \{x_1, x_2 \ldots\ldots x_n\} \epsilon \chi$ and x_i is the ith feature term. The source classifier in figure 1 represents a conventional fault identification problem which focusses on determining the occurring fault types when data from the source and target domains are assumed to follow the same distribution. However, in practical real-world industrial applications, the performance of the source domain classifier is poor as the training samples (source domain) used for building the model cannot be generalized well to the testing samples (target domain). Therefore, we focus on studying situations where the source and target domains follow different data distributions. In such a case the classifier trained on the source domain data would not be directly applied to the target domain.

$D_S = \{x_{Si}, y_{Si}\}$ represents the source domain with labeled training data; $x_{Si} \in \chi_S$ represents a data sample and $y_{Si} \in Y_S$ represents the corresponding label of x_{si}. $D_T = \{x_{Ti}\}$ represents the target domain with unlabeled testing data; where x_{Ti} represents a data sample in χ_T. It must be noted that in this study $\chi_S \neq \chi_T$ and $P(X_s) \neq P(X_T)$ and thus the central idea is on adapting a deep learning model trained on D_S (with labeled fault information) to a data belonging to D_T (with unlabeled fault information). After the DA, we expect that the model trained on D_S would be able to recognize health conditions from data samples belonging to D_T. This study follows some basic assumptions:

i. The number and type of health labels are the same in both the source domain and target domain, i.e. $Y_S = Y_T$.
ii. While model training phase, labeled samples are available from the source domain, only unlabeled samples are available from the target domain.
iii. The source and target domains are related to each other and only differ in their respective data probability distributions.

2.2. Convolutional neural network

The architectural framework of CNN is depicted in figure 2. In general, the architecture majorly consists of three layers: (1) convolutional layer; (2) pooling layer; and (3) fully-connected (FC) layer. The convolutional layer contains a number of filters which are convolved with the raw data. The main function of the convolutional layer is to identify important features or feature maps (set of weights) from the given set of input data. The idea behind using a number of filters is that different filters would detect/identify a different set of features when convolved with the input data. Assuming $\boldsymbol{x} = \{x_1, x_2 \ldots\ldots x_N\}$, where N denotes the length of sequential signal input, the convolution operation can be interpreted as a multiplication operation between a filter kernel $w, w \in R^N$, and a concatenation vector representation $x_{i:i+D-1}$, which is given as

$$x_{i:i+D-1} = x_i \oplus x_{i+1} \oplus \ldots. \oplus x_{i+D-1} \tag{1}$$

where $x_{i:i+D-1}$ represents the window of D length sequential signal starting from index i to index $i + D - 1$ and $\oplus$ concatenates each data sample into longer embedding. The final convolution operation becomes

$$z_i = \phi\left(w^T x_{i:i+D-1} + b\right) \tag{2}$$

where ϕ is the activation function, b is the bias term and z_i is the learned feature of the filter kernel$\boldsymbol{w}$. Hence, the feature map of the jth filter can be denoted as

$$z_j = \left[z_j^1, z_j^2, \ldots., z_j^{N-D+1}\right] \tag{3}$$

with corresponding lengths of $x_{1:D}, x_{2:D+1}, \ldots. x_{N-D+1:N}$.

The extracted features are then passed on to the next layer. i.e. the pooling layer (usually follows a convolution layer), which attempts to reduce the spatial size of the representation from the first layer. It can be seen as a down-sampling layer (lower resolution representation), which reduces the feature

dimensions of the input. The max-pooling function is applied with pooling length g as

$$h_j = \left[h_j^1, h_j^2, \ldots . h_j^s\right] \tag{4}$$

$$h_j^t = \max\left(z_j^{(t-1)g+1}, z_j^{(t-1)g21} \ldots . z_j^{tg}\right) \tag{5}$$

where h_j is a s-dimensional vector, which is the output of the pooling layer applied to the jth feature map. Several alternating convolutional and pooling layers are followed by one or several FC layers. Finally, the result of FC layers is input to a Softmax or sigmoid function classifier.

2.3. Maximum mean discrepancy (MMD)

In general, MMD represents the distance between two probability distributions as a distance between mean embeddings of features in a reproducing kernel Hilbert space. In this paper, MMD is used to evaluate the domain discrepancy between labeled source domain samples and unlabeled target domain samples. Given two probability distributions P and Q on χ, MMD is defined as [29]

$$\mathrm{MMD}\left(\mathcal{F}, P, Q\right) = \sup_{f\in\mathcal{F}}\left(E_{x_1\sim P}\left[f\left(x_1\right)\right]\right) - E_{x_2\sim Q}\left[f\left(x_2\right)\right] \tag{6}$$

where $\mathcal{F}$ is a class function $f : \chi \to \mathcal{H}.\mathcal{H}$ denote reproducing Kernel Hilbert space (RKHS). $\boldsymbol{X}^1 = \left\{x_i^{(1)}\right\}_{i=1}^{n_1}$ and $\boldsymbol{X}^2 = \left\{x_i^{(2)}\right\}_{i=1}^{n_2}$ are data vectors drawn from data space D_1 and D_2, respectively. Based on the fact that f is in the unit ball in a universal RKHS, equation (1) can be rewritten as

$$D\left(\boldsymbol{X}^1, \boldsymbol{X}^2\right) = \left\| \frac{1}{n_1}\sum_{i=1}^{n_1}\varphi\left(x_i^{(1)}\right) - \frac{1}{n_2}\sum_{i=1}^{n_2}\varphi\left(x_j^{(2)}\right)\right\|_{\mathcal{H}} \tag{7}$$

where $\varphi(\cdot)\,\chi \to \mathcal{H}$ is referred to the feature space map. Practical evaluation of MMD is done by employing the kernel method, which aids in evaluating the distance between the distributions of high-level learned features between different domains via

$$\begin{aligned} D_k\left(\boldsymbol{X}^1, \boldsymbol{X}^2\right) &= \left[\frac{1}{n_1^2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_1}\left\langle\varphi\left(x_i^{(1)}\right), \varphi\left(x_j^{(1)}\right)\right\rangle\right. \\ &\left. + \frac{1}{n_2^2}\sum_{i=1}^{n_2}\sum_{j=1}^{n_2}\left\langle\varphi\left(x_i^{(2)}\right), \varphi\left(x_j^{(2)}\right)\right\rangle + \frac{2}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\left\langle\varphi\left(x_i^{(1)}\right), \varphi\left(x_j^{(2)}\right)\right\rangle\right]^{\frac{1}{2}} \\ &= \left[\frac{1}{n_1^2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_1}k\left(x_i^{(1)}, x_j^{(1)}\right) + \frac{1}{n_2^2}\sum_{i=1}^{n_2}\sum_{j=1}^{n_2}k\left(x_i^{(2)}, x_j^{(2)}\right) + \frac{2}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}kk\left(x_i^{(1)}, x_j^{(2)}\right)\right]^{\frac{1}{2}} \end{aligned} \tag{8}$$

where $D_k\left(\boldsymbol{X}^1, \boldsymbol{X}^2\right)$ is the unbiased estimation of $D\left(\boldsymbol{X}^1, \boldsymbol{X}^2\right)$. $k(\cdot,\cdot)$ is a kernel function to compute the inner product in a higher dimensional space, i.e. $k(x, y) = \varphi(x), \varphi(y)$.

3. Proposed method

The flowchart detailing the deep learning-based cross-domain adaptation methodology is depicted in figure 3. Raw time-domain vibration data is acquired from a gearbox at different operating speeds and for different health conditions. The acquired data is partitioned to prepare training samples which consist of labeled source domain training samples and unlabeled target domain training samples. A CNN is constructed and the weights and biases are initialized. Labeled source domain training samples are given as an input to the CNN model and the classification error is minimized in order to train the network. Next, in order to draw the unlabeled target domain data distribution closer to the labeled source domain data distribution, the distribution discrepancy between both

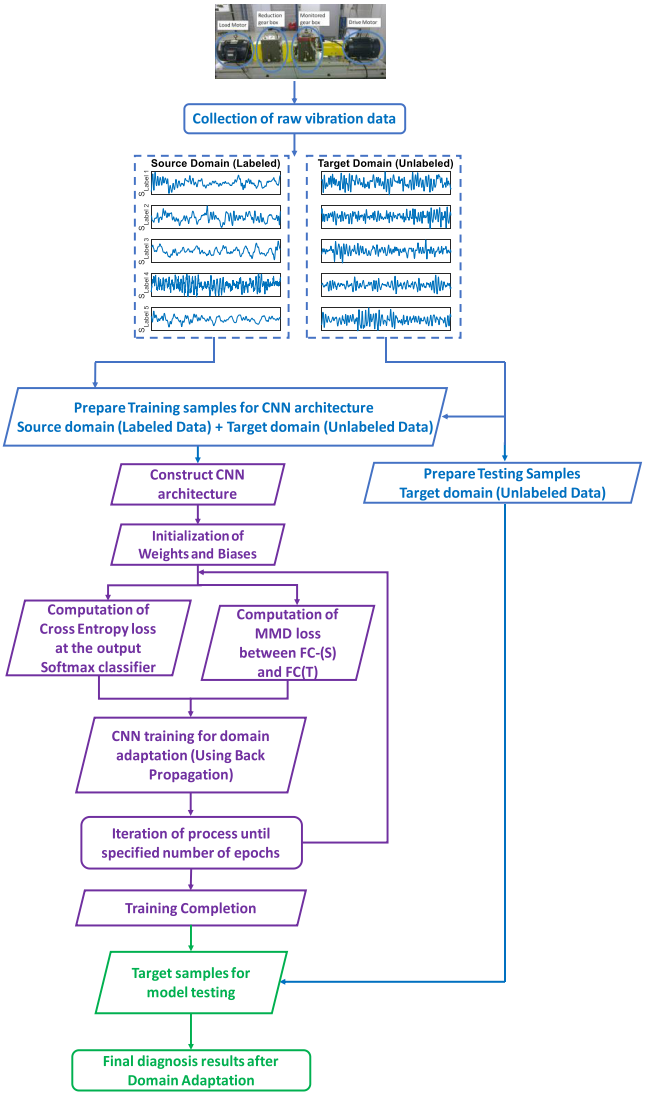

Figure 3. Flow chart of the proposed CNN-CDA method.

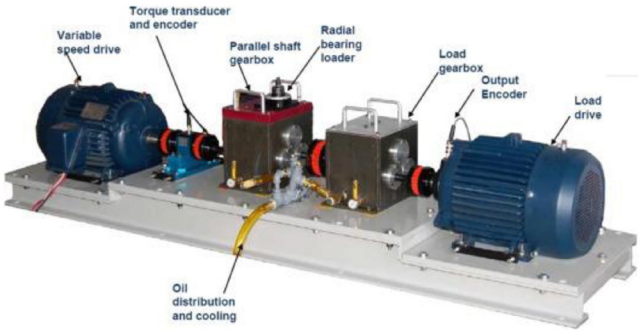

Figure 4. Schematic of the test rig.

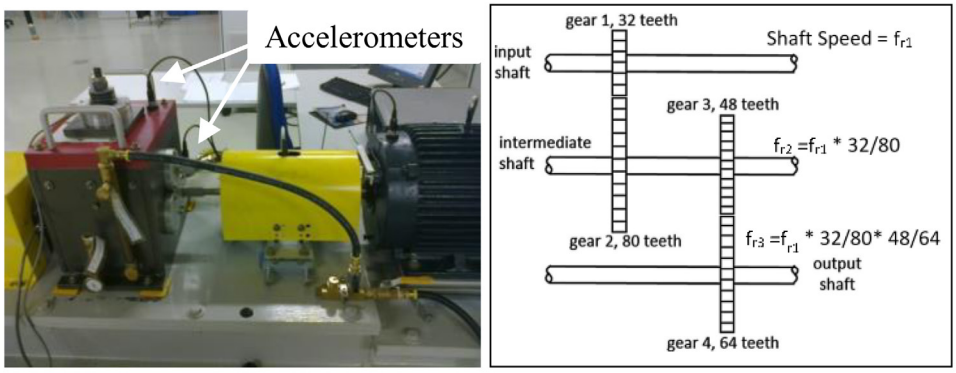

Figure 5. (a) Position of accelerometers; (b) description of gears in the monitored gearbox.

these sets of data is minimized using the multi-kernel MMD. The multi-kernel MMD function detailed in the previous section will be evaluated using the training samples. The summation of the MMD distance and classification loss between the source and target distributions will be minimized to effectively generalize the classifier trained from source domain to target domain. The two losses are explained in detail below:

3.1. Categorical cross-entropy loss (classification loss)

The basic expectation of a CNN architecture is to ensure precise classification. Hence, it uses an objective function to minimize the classification error on the source domain labeled data during the training phase. In general, a source dataset would consist of n_s pairs: $\{(x_{1s}, t_{1s}), (x_{2s}, t_{2s}), \ldots .. (x_{n_s}, t_{n_s}), \}$, where $x_{is} \in R^{\gamma}$ denotes the γ-dimensional network target, and $y_{is} \in R^{C}$ its output. In other words, the objective is to classify γ-dimensional input x_{is} to one of the K health condition categories; the standard Softmax regression loss can be described as follows:

$$L_{\text{classfication}}(\theta, x_s, y_s) = \frac{1}{n_s}\sum_{i=1}^{n_s}\sum_{k=1}^{K} 1\{y_{si} = k\} \log(y_{sik}) \quad (9)$$

where, $1\{\cdot\}$ is a binary indicator function that detects whether the ith training pattern returns 1 if the condition is true, and 0 if the condition if false; indicating whether it belongs to the kth category. K is the number of categories, $\theta = [\theta_1, \theta_1, \ldots \theta_K]^T$ denotes the SoftMax model parameters, and y_{sik} is the predicted output probability distribution for ith observation belonging to class k.

3.2. Distribution discrepancy loss (MMD loss)

As stated in [30], variability in kernels may embed different probability distributions in different RKHSs; hence, an appropriate kernel choice would ensure low testing error of MMD. Therefore, in this study, a mixture of RBF kernels (implying multi-kernel MMD) is utilized:

$$k(x^{\mathrm{S}}, x^{\mathrm{T}}) = \sum_{i=1}^{N_k} k_{\sigma i}(x^{\mathrm{S}}, x^{\mathrm{T}}). \quad (10)$$

$k_{\sigma i}$ represents Gaussian kernel with bandwidth parameters σ_i. Based on trial and error from various experiments, it was found that a mixture of three kernels at simple default bandwidth values of 2, 4, and 8 provided good results. For simplicity purposes, the weights were kept equal. The optimization function to minimize the distribution discrepancy is formulated as follows:

$$L_{\text{MMD}} = \text{MMD}_k\left(\boldsymbol{X}_{\boldsymbol{S}}^{\boldsymbol{f(FC)}}, \boldsymbol{X}_{\boldsymbol{T}}^{\boldsymbol{f(FC)}}\right) \quad (11)$$

where $\boldsymbol{X}_{\boldsymbol{S}}^{\boldsymbol{f(FC)}}$ and $\boldsymbol{X}_{\boldsymbol{T}}^{\boldsymbol{f(FC)}}$ denotes the FC layer representation for the source and target samples respectively and MMD_k represents the multi-kernel MMD between the source and target domain data evaluated on the FC layer representations.

3.3. Overall objective function

By integration the categorical cross entropy loss (equation (9)) with distribution discrepancy loss (equation (11)), the overall objective function becomes

$$\min L = \alpha L_{\text{MMD}} + \beta L_{\text{Classification}} \quad (12)$$

where α and β are the regularization parameters. The default value ofα is '1' and the value of and β is '0' for the without DA case and is '1' for the DA case in this paper.

4. Experimental test set-up

4.1. Test bench

Experimental data are acquired using a Spectra Quest gearbox prognostic simulator (GPS) [31], which is displayed in figures 4 and 5.

GPS is composed of two gearboxes and two electrical motors. Both the drive and load motors are three-phase induction motors having two pair of poles. The rated power output of both motors is 10 Hp. The monitored gearbox is composed of four spur gears (figure 5(b)). Two PCB accelerometers (Model 60811A11 Industrial ICP) were installed to capture vibration in X and Y direction (figure 5(a)). The data was recorded using a computer with a National Instruments acquisition card (NI 4472 series) at a sampling rate of 50 kS s^{-1}.

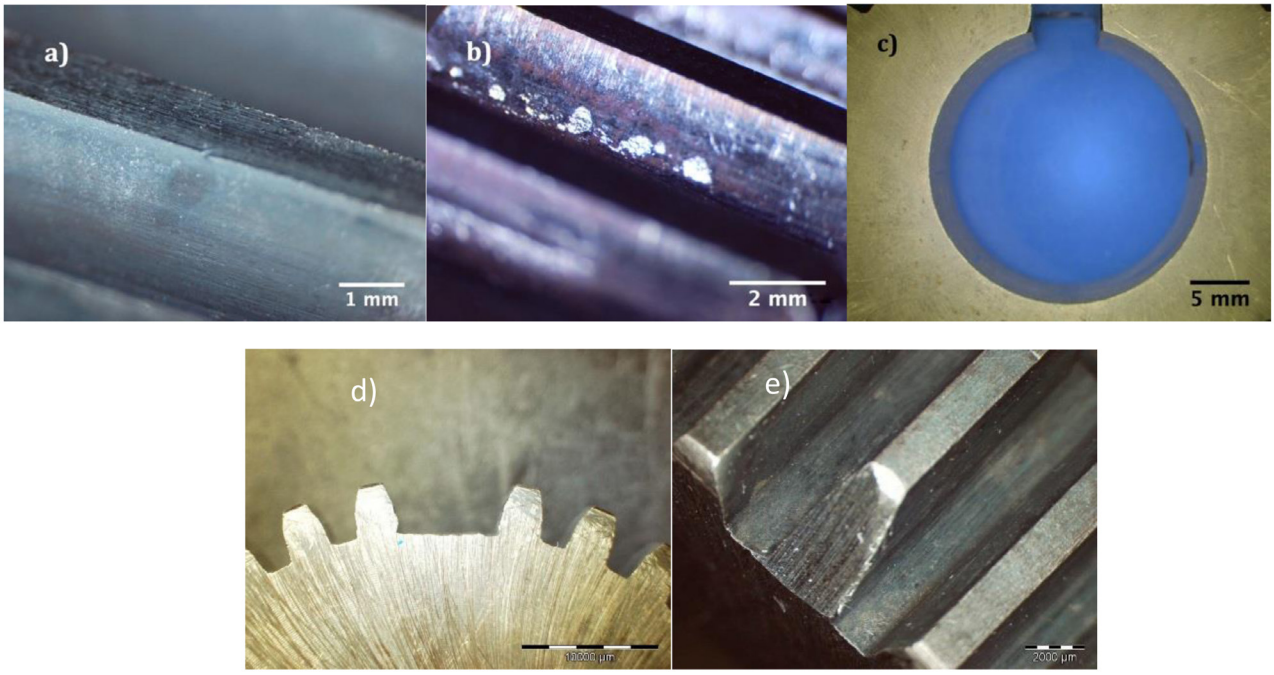

Figure 6. Representative health conditions on the tested gear. (a) Healthy gear. (b) Gear with pitting. (c) Gear with eccentricity. (d) Missing tooth. (e) Chipped tooth.

Table 1. Transfer tasks conducted in this study.

Transfer task	Source domain (speed)	Target domain (speed)
$T_{500–1000}$	500	1000
$T_{500–1500}$	500	1500
$T_{1000–500}$	1000	500
$T_{1000–1500}$	1000	1500
$T_{1500–500}$	1500	500
$T_{1500–1000}$	1500	1000

4.2. Dataset description

The focus is set on analyzing different speeds especially in the no-load condition. The maximum speed that the GPS test bench can reach is 1500 rpm. The minimum selected speed is 500 rpm. One more intermediary speed, 1000 rpm, was selected for the DA task. During the experiments, for each test a different fault type gear was inserted at the place of gear 1 of the first gearbox (the 32 teeth gear in figure 5). Apart from a healthy gear, four different types of faults were considered: eccentric tooth, chipped tooth, missing tooth, and pitting on tooth. The representative fault conditions of the considered gears are depicted in figure 6.

Table 1 represents the details of the transfer tasks considered for analysis. N_{source} and N_{target} represent the respective source and target domain samples for each health condition under one speed. For simplicity, $N_{\text{source}} = N_{\text{target}}$ is considered in this study. The raw data sequence is divided into N_{input} sequential points.

For conducting an effective and robust analysis, ten trials are carried out for each algorithm for diagnosing the gearbox data. The average testing accuracy is reported for all reported experimental results. Our implementation is carried out in the

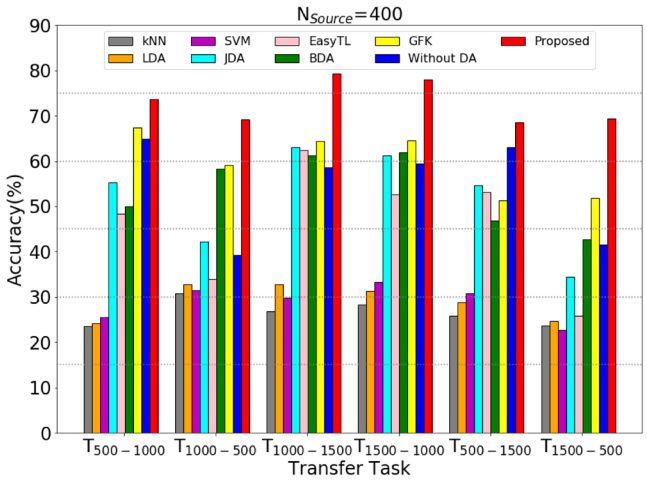

Figure 7. Analysis of the six transfer tasks using different DA methods. $N_{\text{Source}} = 400$ is used.

Tensorflow platform in Windows operating system on Intel® Xeon® W-2145 CPU @ 3.7 GHz processor running at 32 GB RAM and GPU parallel computing (NIVIDIA GeFoce TX 2080 Ti).

5. Results and discussion

This section presents the results of the proposed CNN-CDA method on the gearbox dataset. To validate the reliability, effectiveness and robustness of the proposed method a comparison with three state of the art techniques is carried out. The three approaches are as follows: (a) machine learning methods (SVM, kNN and LDA) based on manually extracted vibration features, (b) transfer learning methods (JDA [32], BDA [33], EasyTL [34], GFK [35]) based on manually extracted

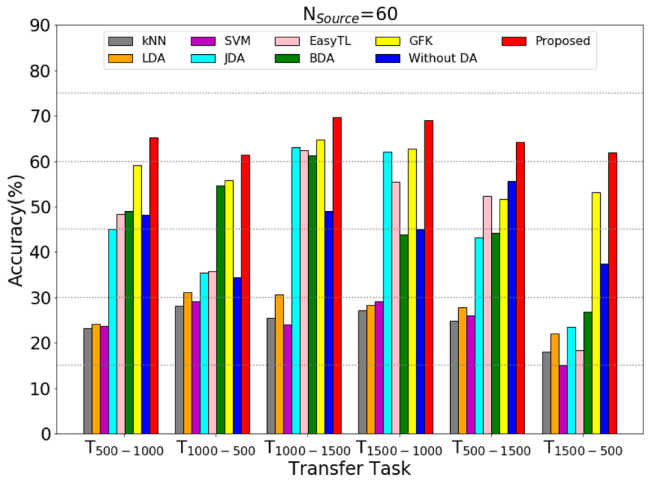

Figure 8. Analysis of the six transfer tasks using different DA methods. $N_{\text{Source}} = 60$ is used.

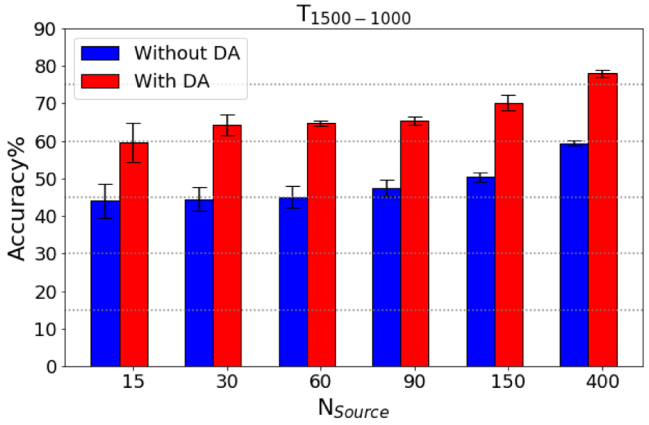

Figure 9. Cross domain testing accuracy for task $T_{1500–1000}$ using different amounts of labeled source domain data.

vibration features, (c) CNN-based feature extraction without cross-domain adaptation (only the classification loss in equation (12) is considered). In (c), the model trained on source domain data is directly used for testing (without DA).

5.1. DA results

Results from all six transfer tasks, i.e. $T_{500–1000}$, $T_{1000–500}$, $T_{1000–1500}$, $T_{1500–1000}$, $T_{500–1500}$ and $T_{1500–500}$, are presented in figure 7. By default, $N_{\text{input}} = 2000$ is used in this particular study. The results from figure 6 indicate that the proposed CNN-based cross-domain adaptation methodology achieves the highest testing diagnosis performance among all techniques in different tasks. It must be noted that the high testing diagnosis performance in the tasks $T_{500–1000}$ and $T_{1000–500}$, $T_{1000–1500}$ and $T_{1500–1000}$, $T_{500–1500}$ and $T_{1500–500}$ indicates the bidirectional effectiveness of the proposed approach. Higher testing accuracies in certain specific tasks relative to other tasks, such as $T_{1000–1500}$, $T_{1500–1000}$, indicates the inherent closeness by nature of the respective source and target domains. Overall, high cross-domain adaptation testing diagnosis performance confirms the effectiveness of the proposed CNN-CDA model.

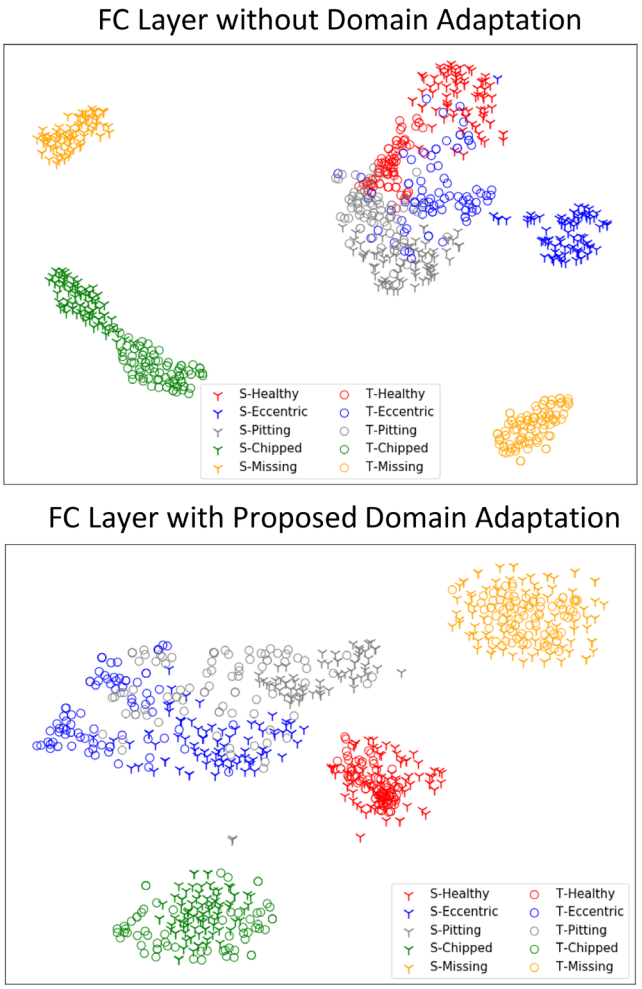

Figure 10. Feature visualization in the FC layer for the $T_{1500–1000}$ task.

As discussed in section 1, since the feature distributions of both datasets (source and target) are dissimilar, the traditional machine learning models (SVM, kNN, LDA) built using the training samples (source domain data) lead to poor generalization when applied directly to the target domain data. In contrast, all representative methods for DA techniques perform significantly better than the traditional machine learning algorithms. GFK has the best performance among them all, followed closely by the BDA method in general. The results from the EasyTL approach are competitive but the results from tasks $T_{500–1500}$ and $T_{1500–500}$ indicate that the approach does not perform well bidirectional between domains. Furthermore, the without DA method achieves reasonably good testing diagnosis results, which are better than the EasyTL approach in most transfer tasks.

5.2. Low amount of labeled source domain data

Figure 8 depicts the results from all transfer tasks for a lower amount of the labeled source domain data used for training (60, respectively).

The results from figure 8 validate the effectiveness and robustness of the proposed CNN-CDA approach by

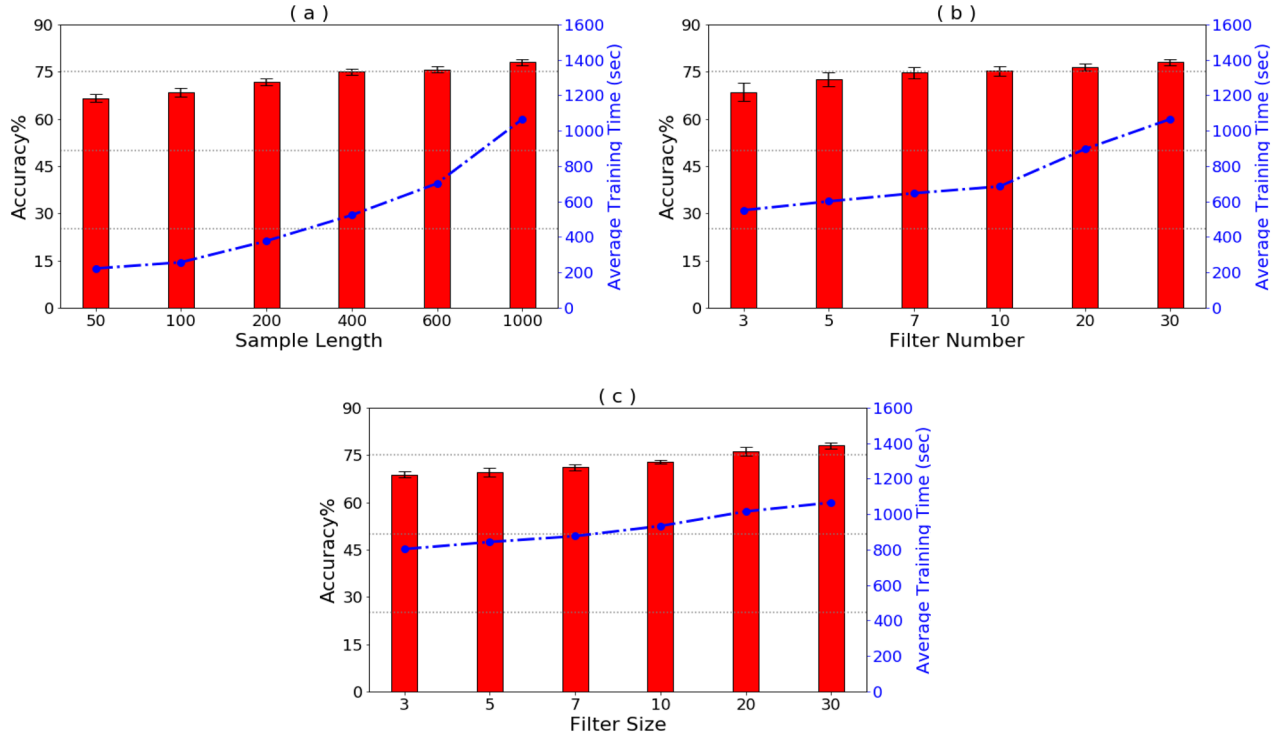

Figure 11. Effects of (a) sample length, (b) the convolutional filter number, and (c) the convolutional filter size on the testing diagnosis accuracy in task $T_{1500–1000}$.

demonstrating larger cross-domain diagnosis improvements in comparison with other methods when a lower amount of labeled source domain data is presented to the proposed CNN-CDA model. The results from figure 8 indicate the ability of the proposed CNN-CDA method to deal with the overfitting problem, which in general is a common issue encountered while implementing intelligent algorithms in situations of insufficient training data. Hence, the proposed method depicts high potential to be extended to other real-world industrial applications that have low availability of labeled data in practice. To further illustrate the performance of the proposed CNN-CDA, figure 9 depicts the results from Transfer task $T_{1500–1000}$ with increasing amount of the labeled source domain data (N_{Source}). The results are compared with the without DA method and clearly indicate that increasing the amount of labeled source domain data increases the cross-domain testing diagnosis performance. The findings are consistent with the general notion about deep learning methodologies that a higher amount of samples used in training phase typically leads to improvement in the network performance.

5.3. Visualization of learned representation

In general, the FC layer captures the most important fault features and thus these features are passed on to the Softmax layer for fault classification. Therefore, visualization of high-dimensional features in the FC layer are presented in this section. The *t*-SNE algorithm was used for visualizing the high-dimensional data representation. Samples are mapped from the original feature space into a 2D space map. Figure 10 depicts the 2D plots of learned representations in the FC layer using the proposed method and the without DA approach for both domain (source and target) for task $T_{1500–1000}$.

It can be observed that for the without DA method, though the learned feature representations with the same health conditions cluster together, they do so independently in both domains. Either (a) separation between respective health labels of both domains is high and learned feature representations from the same health class of two domains are projected into the far-off regions (e.g. missing tooth health condition), or (b) in the target domain, certain class label samples overlap into the regions of other classes (healthy eccentric and pitting condition). This scenario necessitates the proposed idea of DA. In contrast, the features after cross-domain adaptation are very well separated for three health conditions (healthy, missing tooth and chipped tooth) and samples from both domains practically overlap each other. For these three conditions (healthy, missing tooth and chipped tooth), the data samples in both domains that belong to the same health class group and overlap in the region of the feature space. However, the data samples from both of the remaining two health conditions (eccentric gear and pitting) seem to overlap. In conclusion, for the given gearbox dataset, the proposed methodology is able to detect a fault precisely, but with a lower level of fault diagnosis performance.

5.4. Effects of parameters

The section discusses the effect of input sample length (N_{input}), the convolutional filter number and the convolutional filter size on the testing diagnosis accuracy of the proposed CNN-CDA model. The transfer task $T_{1500–1000}$ is used for illustration and

the results are depicted in figure 11. In general, larger N_{input} is generally expected to provide higher diagnosis performance and the same is validated by the results of figure 11(a). Since larger filter size and a large number of convolutional filters in each layer improve the learning capability of CNN, increasing these parameters provides significant improvements in testing accuracies. Furthermore, increasing filter number has a more remarkable effect on testing accuracy in comparison to filter size. Overall, higher values of the above listed parameters provide better results, but at the same time the computational burden increases exponentially (average computational time). Hence, a reasonable tradeoff between the computational burden and the diagnosis accuracy needs to be made while choosing these parameters.

6. Conclusions

A novel deep learning-based cross-domain adaptation methodology for gearbox fault detection and diagnosis is proposed in this paper. The summation of cross-entropy loss (between labeled source domain data) and MMD loss (between labeled source and unlabeled target datasets) is minimized simultaneously to adapt the source domain model to be directly applied on the target domain data samples. Performance of the proposed approach is evaluated on experimental data from a gearbox test rig. The influence of network architecture and tuning parameters of the CNN model are comprehensively investigated to evaluate the performance of gearbox fault diagnosis under significant speed variations.

By comparing the testing diagnosis performance of the proposed CNN-CDA method with other state-of-the-art intelligent machine learning algorithms, DA algorithms and deep learning algorithms, the superiority of the proposed technique is validated and verified. Conventional intelligent algorithms fail to accurately predict the gearbox health condition when tested directly on data from a new domain. It was observed that the conventional DA methods were somewhat able to learn domain-invariant features under different operating speeds, and for this reason, perform better relative to traditional machine learning algorithms in the new domain. However, the proposed method achieves the highest testing accuracy in all reported experimental results. In particular, the proposed method reported highest accuracies with a smaller labeled training dataset relative to other techniques. Moreover, the results demonstrate the ability of the proposed method to deal with the overfitting problem, which is a common issue encountered while implementing intelligent algorithms in situations of insufficient training data.

Even though the proposed method achieves a satisfactory performance, there is a minor drawback that exists in the assumption of having a balanced training dataset over different gearbox health conditions. In industrial applications, though it is a straightforward task to acquire healthy data, it can be quite challenging to obtain data for different fault classes. Therefore, future studies will focus on extending the use of the proposed methodology to an imbalanced training dataset. Further, the current research would be extended to explore the use of a new loss function and to further improve the fault diagnosis performance.

Acknowledgments

The authors would like to acknowledge Inaki Bravo-Imaz from Intelligent Information Systems, IK4—TEKNIKER (Spain) for arranging the experimental data sets.

ORCID iDs

Jaskaran Singh https://orcid.org/0000-0002-0333-5354
Moslem Azamfar https://orcid.org/0000-0001-7642-5003
Abhijeet Ainapure https://orcid.org/0000-0003-2924-6322

References

[1] Liang X, Zuo M J and Feng Z 2018 Dynamic modeling of gearbox faults: a review *Mech. Syst. Signal Process.* **98** 852–76
[2] Pan Y, Hong R, Chen J, Singh J and Jia X 2019 Performance degradation assessment of a wind turbine gearbox based on multi-sensor data fusion *Mech. Mach. Theory* **137** 509–26
[3] Lee J, Davari H, Singh J and Pandhare V 2018 Industrial artificial intelligence for industry 4.0-based manufacturing systems *Manuf. Lett.* **18** 20–23
[4] Lee J, Singh J and Azamfar M 2019 Industrial artificial intelligence (arXiv: 1908.02150)
[5] Lee J, Azamfar M and Singh J 2019 A blockchain enabled cyber-physical system architecture for Industry 4.0 manufacturing systems *Manuf. Lett.* **20** 34–9
[6] Negri E, Ardakani H D, Cattaneo L, Singh J, Macchi M and Lee J 2019 A digital twin-based scheduling framework including equipment health index and genetic algorithms *IFAC-PapersOnLine* (New York: Elsevier) pp 43–8
[7] Wang D 2016 *K*-nearest neighbors based methods for identification of different gear crack levels under different motor speeds and loads: revisited *Mech. Syst. Signal Process.* **70–1** 201–8
[8] Cheng G, Cheng Y, Shen L, Qiu J and Zhang S 2013 Gear fault identification based on Hilbert–Huang transform and SOM neural network *Measurement* **46** 1137–46
[9] Rafiee J, Arvani F, Harifi A and Sadeghi M H 2007 Intelligent condition monitoring of a gearbox using artificial neural network *Mech. Syst. Signal Process.* **21** 1746–54
[10] Chen F, Tang B and Chen R 2013 A novel fault diagnosis model for gearbox based on wavelet support vector machine with immune genetic algorithm *Measurement* **46** 220–32
[11] Singh J, Darpe A K and Singh S P 2017 Bearing damage assessment using Jensen–Rényi divergence based on EEMD *Mech. Syst. Signal Process.* **87** 307–39
[12] Singh J, Darpe A K and Singh S P 2018 Rolling element bearing fault diagnosis based on over-complete rational dilation wavelet transform and auto-correlation of analytic energy operator *Mech. Syst. Signal Process.* **100** 662–93
[13] Jing L, Zhao M, Li P and Xu X 2017 A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox *Measurement* **111** 1–10
[14] Yang B, Lei Y, Jia F and Xing S 2019 An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings *Mech. Syst. Signal Process.* **122** 692–706

[15] Pandhare V, Singh J and Lee J 2019 Convolutional neural network based rolling-element bearing fault diagnosis for naturally occurring and progressing defects using time-frequency domain features *2019 Prognostics System and Health Management Conf.* (IEEE) pp 320–26
[16] Witten I H, Frank E, Hall M A and Pal C J 2016 *Data Mining: Practical Machine Learning Tools and Techniques* (San Mateo, CA: Morgan Kaufmann)
[17] Mar C V 2017 Domain adaptation for visual applications: a comprehensive survey (arXiv:1702.05374)
[18] Lu J, Behbood V, Hao P, Zuo H, Xue S and Zhang G 2015 Transfer learning using computational intelligence: a survey *Knowledge-Based Syst.* **80** 14–23
[19] Zhang W, Peng G, Li C, Chen Y and Zhang Z 2017 A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals *Sensors* **17** 425
[20] Zhang W, Li C, Peng G, Chen Y and Zhang Z 2018 A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load *Mech. Syst. Signal Process.* **100** 439–53
[21] Guo L, Lei Y, Xing S, Yan T and Li N 2018 Deep convolutional transfer learning network: a new method for intelligent fault diagnosis of machines with unlabeled data *IEEE Trans. Ind. Electron.* **66** 7316–25
[22] CWRU 2000 Case Western Reserve University Bearing Data Center Website (http://csegroups.case.edu/bearingdatacenter/home)
[23] Qiu H, Lee J, Lin J and Yu G 2006 Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics *J. Sound Vib.* **289** 1066–90
[24] Lei Y 2016 *Intelligent Fault Diagnoiss and Remaining Useful Life Estimation of Rotating Machinery* (New York: Butterworth-Heinemann)
[25] Lu W, Liang B, Cheng Y, Meng D, Member S and Yang J 2017 Deep model based domain adaptation for fault diagnosis for fault diagnosis *IEEE Trans. Ind. Electron.* **64** 2296–305
[26] Wen L, Gao L and Li X 2017 A new deep transfer learning based on sparse auto-encoder for fault diagnosis *IEEE Trans. Syst. Man, Cybern. Syst.* **49** 136–44
[27] Hasan J, Islam M M M and Kim J 2019 Acoustic spectral imaging and transfer learning for reliable bearing fault diagnosis under variable speed conditions *Measurement* **138** 620–31
[28] Cao P E I, Zhang S and Tang J 2018 Preprocessing-free gear fault diagnosis using small datasets with deep convolutional neural network-based transfer learning *IEEE Access.* **6** 26241–53
[29] Borgwardt K M, Gretton A, Rasch M J, Kriegel H, Smola A J and Scho B 2006 Integrating structured biological data by kernel maximum mean discrepancy *Bioinformatics* **22** 49–57
[30] Gretton A, Sriperumbudur B, Sejdinovic D, Strathmann H and Pontil M 2012 Optimal kernel choice for large-scale two-sample tests *Adv. Neural Inf. Process. Syst.* **xx** 1205–13
[31] Bravo-Imaz I, Ardakani H D, Liu Z, García-Arribas A, Arnaiz A and Lee J 2017 Motor current signature analysis for gearbox condition monitoring under transient speeds using wavelet analysis and dual-level time synchronous *averaging Mech. Syst. Signal Process.* **94** 73–84
[32] Long M, Wang J, Ding G, Sun J and Yu P S 2013 Transfer feature learning with joint distribution adaptation *Proc. IEEE Int. Conf. Computer Vision* pp 2200–7
[33] Wang J, Chen Y, Hao S, Feng W and Shen Z 2017 Balanced distribution adaptation for transfer learning *2017 IEEE Int. Conf. Data Mining* pp 1129–34
[34] Wang J, Chen Y, Yu H, Hunag M and Yang Q 2019 Easy transfer learning by exploiting intra-domain structures *IEEE International Conference on Multimedia and Expo (ICME)* pp 1210–5
[35] Gong B, Shi Y, Sha F and Grauman K 2012 Geodesic flow kernel for unsupervised domain adaptation *2012 IEEE Conf. Computer Vision Pattern Recognition* pp 2066–73