

---

# UNSUPERVISED TRANSFER LEARNING ACROSS DIFFERENT DATA MODALITIES FOR BEARING'S SPEED IDENTIFICATION

---

**Diego Nieves Avendano\***, Dirk Deschrijver, Sofie Van Hoecke,

Ghent University - imec

Ghent, Belgium

Corresponding author\* {diego.nievesavendano@ugent.be}

## ABSTRACT

**Context.** Recent advancements in transfer learning have revolutionized predictive maintenance, enabling cross-domain generalization for components with varying characteristics and operating under different conditions. While traditional transfer learning approaches require labeled data in both source and target domains, unsupervised transfer learning aims at a more cost-effective alternative for which only labels are available in the source domain.

**Methods.** This study investigates adversarial transfer learning between two different sensor modalities: vibration and acoustic. The aim is to enable bearing monitoring using microphones, which are, in general terms, cheaper and easier to deploy than vibration sensors; and without the need of having to label data in the target domain. The research goal is to identify the operating speed of a bearing testbed. The source domain data corresponds to vibration measurements taken from an attached sensor, while the target domain uses a microphone array at distance. Artificial Neural Networks are used as the base architecture. Transferability is assessed with two unsupervised adversarial learning techniques: gradient reversal and deep correlation alignment. Their performance is compared to traditional supervised transfer learning via fine-tuning.

**Results.** Experimental results demonstrate that gradient reversal outperforms deep correlation alignment, and is able to achieve results similar to those obtained with supervised transfer learning.

**Conclusion.** These findings highlight the feasibility of speed identification using a microphone array and establish a baseline for future condition monitoring research with such sensors.

### 1. INTRODUCTION

Predictive maintenance (PdM) models of mechanical components are key components of Prognostics and Health Management (PHM) in the industry. Accurate PdM models allow improvements in terms of quality, safety, maintenance scheduling, and cost reduction. While developing monitoring solutions in the industry, an important step is the correct selection of the sensor type. The sensor’s type selection is done based on criteria such as the type of machine to monitor, implementation costs, early detection capabilities, whether the sensors can be attached to the machine or not, among others. Regarding the monitoring of mechanical components, the most common sensor types measure vibration, sound, or temperature. Thanks to the advances in computer vision the use of cameras and infrared cameras has also become widespread. With respect to the sensor’s cost, vibration sensors based on piezoelectric components and MEMS tend to be in the high cost of the solutions; microphones and ultrasonic microphones in the mid-price spectrum; and temperature sensors in the low cost.<sup>1</sup> It is important to note that this is a broad generalization, as specifications concerning bandwidth, noise reduction, packaging, and other characteristics can considerably impact the sensor’s price. Another important aspect is the early detection capability of each of these sensor modalities. As depicted in Fig. 1, vibration sensors and ultrasonic microphones have the best early detection capabilities, whereas temperature and microphone sensors have more limited detection capabilities. Finally, the environment where the machines operate poses limitations on the type of sensor that can be used and potential costs due to sensor’s damage. For example, machines with mechanical components that move or rotate can easily damage sensors and wires. Other environment scenarios pose limitations due to the presence of contaminants, extreme temperatures, high humidity, water, among others. In this respect sensors that do not need to be attached or in proximity to the machines, such as cameras and microphones, provide a better alternative. This research focuses on creating a bridge between vibration sensors and microphones, where the goal is to enable better monitoring capabilities using microphone data by reusing data collected with vibration sensors.

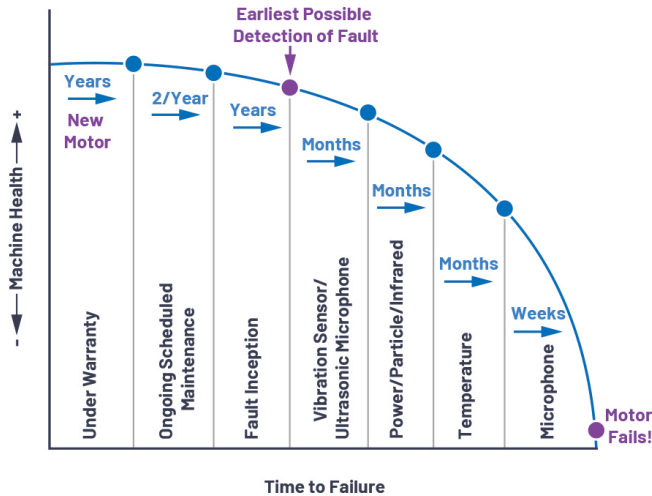


Figure 1: Machine health over time and the sensor types that can detect damage at each stage. Depicted in.<sup>1</sup>

## Unsupervised transfer learning for bearing monitoring

The interest in acoustic sensors as a monitoring solution is based on the advantages that microphones offer over vibration sensors, namely: they are easier to set up; can be installed at safe distances from the machines. In addition, as they do not need to be attached to the machines they can be easily protected from environmental hazards, such as high temperature or humidity; and are cheaper when compared to other sensor types. However, the use of microphones comes with challenges and limitations. The main drawback of microphone solutions is that they capture environmental noise which is irrelevant for monitoring purposes. The noise bandwidth can, in some instances, overlap with the bandwidth of interest for monitoring, and in such scenarios, filtering is not a trivial task. Furthermore, sensors attached to the machine (e.g., vibration) capture the spectra of the actual vibration, whereas microphones capture a distorted representation due to the interference of other possible components, attenuation due to varying distance to the microphone, phase and frequency shifts introduced due to the change in medium, among other factors. Finally, according to industrial reports,<sup>1</sup> acoustic sensors have the most limited capability for detecting damage at early stages.

To harness the potential of acoustic sensors while addressing their shortcomings, this research proposes the use of deep learning models, which can infer high-quality features that overcome the previously mentioned distortions. Furthermore, unsupervised transfer learning techniques are investigated to enable transfer in a scenario where labels are only available for the source domain. This is ideal in scenarios where labeling can be expensive or deployment can be accelerated by avoiding labeling in the target domain. The transferal can be described as training models using high quality vibration data and co-training using unlabeled and lower quality data from microphones.

Two transfer learning (TL) paradigms are evaluated here, namely supervised transfer learning (STL) and unsupervised transfer learning (UTL). The difference between the paradigms is based on whether labels from the target domain can be used during the training phase. The STL approach followed is considered a reference for the performance that could be expected if the labels were available. The methodologies are evaluated using the Smart Maintenance Living Lab (SMLL) dataset created by Flanders Make and imec, which includes vibration, acoustic, and temperature recordings.<sup>2</sup>

The paper is divided as follows: Section 2 reviews works related to PdM tasks using deep learning and transfer learning techniques. Section 3 presents the theory behind fine-tuning which is the STL technique used as a baseline. Section 4 presents the theory behind the unsupervised learning techniques used in this work. Section 5 presents the SMLL dataset and describes the differences between the domains. Section 6 outlines the methodology, which consists of the feature representation, the architecture selection, and the evaluation method. Section 7 presents the results and discussion. Section 8 proposes lines for future work. Finally, Section 9 presents the conclusions and suggestions for future work.

## **2. RELATED WORK**

Over the last few years, machine learning techniques have proven successful in solving various PdM tasks. More specifically, deep learning has garnered attention due to its ability to identify patterns from raw or lightly processed data. The related work reviewed focuses on the deep learning solutions as the transfer learning techniques covered in this work are designed to work for neural network architectures. Deep learning techniques have been successfully used for

## Unsupervised transfer learning for bearing monitoring

tasks such as anomaly detection, fault classification, and remaining useful life estimation.<sup>3,4</sup> This work addresses the problem of identifying the operating speed, a known parameter, and therefore can be used as a metric of how a healthy bearing deviates from previously seen data. The goal is demonstrating that transferal can be done in an unsupervised way. If the task can be successfully performed, the next steps aim towards a PdM task via unsupervised learning by, for example, using the learned embeddings in combination with a one-class classifier or other anomaly detection approaches.

This Section is divided as follows: first, an overview of deep learning approaches for PdM is detailed; followed by a brief overview of supervised transfer learning in the PdM domain; and finally, the unsupervised transfer learning literature.

### 2.1. Deep learning

This section provides an overview of successful applications of deep learning with different data representations within the supervised domain for PdM. Artificial Neural Networks (ANN) have successfully been used on raw time signals for benchmark datasets such as the Case Western Reserve University (CWRU) and the Intelligent Maintenance System (IMS) Bearing Dataset.<sup>5</sup> A feed-forward network that uses raw data as input was able to correctly classify the different faults present in the datasets. However, using feed-forward networks is only possible with relatively small time windows as the dimensionality of the network increases drastically with high-sampling sensors. Other data representation modes, such as spectrograms, are preferred as they provide smaller input spaces. The spectrogram representation has been used in combination with feed-forward networks for tasks such as remaining useful life estimation.<sup>6,7</sup>

Convolutional Neural Networks (CNN) are an evident extension to ANN due to their ability to exploit time and frequency relations of the data, in addition to being more parameter efficient. CNN architectures have become more popular due to their parameter efficiency, as can be observed from the extensive literature surrounding them. For example, CNNs have been used together with temporal data<sup>8</sup> and wavelet transforms<sup>9</sup> for estimating remaining useful life. Raw temporal data, or lightly processed, has been used for fault classification of bearings.<sup>10-13</sup> In these studies the input vector corresponds to the vibrations over time, where in some cases data is rearranged as images in two dimensions in order to use two dimensional kernels. Representation in the frequency domain such as Fourier transform and discrete wavelet transform,<sup>14</sup> continuous wavelet transform,<sup>15</sup> spectrograms,<sup>16,17</sup> and wavelet decomposition<sup>18-20</sup> have also been used for fault classification over some of the most famous benchmarks such as CWRU, IMS, and proprietary datasets. In other cases, custom features over time have been used in combination with CNN architectures for fault classification as well as wear estimation.<sup>21,22</sup> Overall, the literature points to a strong preference for CNNs over other architectures, allowing for a lot of flexibility in the forms of data input representation.

The literature review shows that extensive research has been done in deep learning for PdM, and that diverse architectures of neural networks can be used. This research uses the mel-spectrogram representation in combination with ANNs as it reduces the number of parameters considerably. An alternative in this case would be to use two-dimensional representations, such as wavelet transforms or spectrograms, in combination with architectures such as CNNs or LSTMs,

## Unsupervised transfer learning for bearing monitoring

which would provide additional information about changes over time. However, as this work focuses on the transfer learning approach, the simpler architecture of ANNs was preferred.

### 2.2. Transfer learning in PdM tasks

Transfer learning concerns a group of problems in which the objective is reusing previously acquired knowledge for similar tasks. The TL problem can be phrased as follows: (i) given a group of Domains  $\mathcal{D}$ , each domain contains a dataset with a given set of features; (ii) the features in each domain measure the same properties or similar ones, but the distributions across domains differ; and (iii) TL finds a way to leverage the common information across domains to improve results on the task for any given domain. Most commonly, TL is seen as a problem in which for a first domain there is a considerable amount of (high quality) data and the knowledge of this domain wants to be applied on a new domain, where there is a limited amount of data, and sometimes of lower quality. In other cases it can be seen as reusing a model trained to solve a certain task (e.g. image recognition), or to solve a secondary task (e.g. regression).

Large pre-trained deep learning models have been used as base models for STL. Considerable research has proven that the technique helps in improving performance, and reducing the target dataset requirements.<sup>23,24</sup> Large networks like ResNet<sup>25</sup> and InceptionV3<sup>26</sup> were developed for image classification tasks and trained on the Imagenet dataset, which consists of pictures of real-life objects. The trained versions of these models have been reused for different tasks with different data representations. For example, fault classification of wind turbine gearboxes using wavelet representation as inputs and the pre-trained ResNet as feature map extraction;<sup>27</sup> fault detection in photovoltaic plants using thermal images and the pre-trained ResNet;<sup>28</sup> and wear estimation of cutting tools based on image inputs using pre-trained models such as ResNet, InceptionV3 and AlexNet as basis for a fine-tuning approach.<sup>29</sup> One important point to consider about the previously mentioned approaches is the validity of the pre-trained weights, as they have been trained to recognize objects in real-life images, which may not necessarily translate to representations of spectrograms or other types of images, such as the physical condition of mechanical components. The work by Janssens et al.<sup>30</sup> uses the pre-trained VGG network for fault detection and oil level prediction using infra-red images. Their work showed that models trained for image recognition (VGG) can be adapted for thermal images. Notice that in this scenario the base weights are likely to be relevant as they pertain real world shapes. This example is relevant to the work presented in this paper, with the difference that in the proposed scenario the data does not correspond to photographic images.

Other types of transfer learning approaches have focused on the transferral across two predictive maintenance tasks. This shift can occur in cases where the source and target correspond to similar components with different specifications; the analysis of different fault types;<sup>13,22,31</sup> or discrepancies in data distributions of the same mechanical component due to large heterogeneity.<sup>32,33</sup>

Concerning UTL for PdM, adversarial transfer learning with class weights has been applied across bearings operating under different conditions of the CWRU dataset.<sup>34</sup> Their results proved to have a clear advantage over a baseline with no adaptation and on par with other domain adaptation approaches. A technique that uses CNNs with a training procedure that uses the maximum mean discrepancy to align domains was evaluated on the Case Western Reserve University

## Unsupervised transfer learning for bearing monitoring

(CWRU) bearing dataset<sup>35</sup> and an in-house produced dataset.<sup>36</sup> The transferral here is between different operating loads for the CWRU data and different speeds for their in-house data. The objective is to identify the bearing's condition between healthy and different fault categories. Their results achieve high accuracy in the identification of faults. Notice that the work is done on the CWRU dataset, which despite being a common benchmark is considered easy to solve, with dozens of papers obtaining accuracies above 99% for each fault type and making it hard to assess whether one or other techniques offer some true advantage. An adversarial approach with selective adaptation in more challenging datasets was presented by Deng et al.<sup>37</sup> In their case the adaptation is done between the CWRU dataset and the Paderborn bearing dataset,<sup>38</sup> both of which differ considerably in the type of bearings and operating conditions; and between the CWRU and the XJTU-SY dataset<sup>39</sup> for different bearing and fault types. Their results show the effectiveness of transferring between identical machines but also different ones.

The literature review shows that the TL problem in the context of PdM is focused mainly on transferal across different types of faults and for changes in operating conditions for the same machine, with the notorious exception of the work by Deng et al.<sup>37</sup> Regarding STL, the scenario of changes in non-image modalities for PdM has never been thoroughly researched. Furthermore, in contrast with previous work, this research focuses on the transfer between different data modalities, which has been limited so far.

The proposed TL scenario is interesting as a way of adapting models to changes in the monitoring technology. Potential applicable scenarios include: (i) updates in the monitoring technology, (ii) changes in workshop conditions that prevent sensors from being attached to the machine (e.g., rotating components or high temperatures), and (iii) enabling the use of cheaper and easier-to-deploy technologies (microphones are considerably cheaper than vibration sensors). In summary, transfer learning from vibration signals to acoustics can enable new methods of condition monitoring. This paper reviews transfer learning techniques under the assumption that no labels are available in the target domain.

The literature review identified challenges within the task are the following:

1. Large variability in the source domain. Bearing datasets have often been reported as being small and having considerable heterogeneity within same condition tests. This makes it hard for generalization of models, and poses additional challenges for the TL task. Large heterogeneity has been reported in datasets such as Pronostia,<sup>40</sup> IMS,<sup>41</sup> as well as SMLL.<sup>42</sup>
2. Distortions in the spectrogram due to a medium change. The core idea in this research is that microphones can capture similar information to that of vibration sensors. However, sensor differences, frequency bands, and distortions due to the change of medium can considerably affect the acoustic signal quality.
3. Few datasets suitable for evaluating TL are available. Datasets such as PHM and IMS were not designed for the TL task, and their sample sizes have proven to be challenging. The CWRU bearing dataset is the most commonly used dataset for TL between operating conditions and has been extensively employed. An interesting dataset is the one used by Gebraeel et al., consisting of 36 bearings; however, it is not publicly

## Unsupervised transfer learning for bearing monitoring

available.<sup>6</sup> To the best of our knowledge, no public datasets concerning different data modalities have been developed or released.

### 3. SUPERVISED TRANSFER LEARNING

The fine-tuning method is one of the most widely used approaches for the STL task.<sup>24</sup> This technique is broadly used with deep learning models as their parameters can be easily adapted with respect to the network's performance over the target domain. The fine-tuning takes the following steps: First, a neural network is trained with a large amount of high quality data from the source domain. Then, a portion of the layers is so-called *frozen*, meaning that the weights are fixed, and the training continues using the target domain data. The weights which are frozen are not adapted during this step. Finally, the training is stopped using any preferred strategy, such as detecting the plateau over the validation error.

Most commonly, if the target dataset is small and the number is relatively large, the preferred approach is freezing all the layers except the last few ones. The idea behind it is that the initial layers are able to learn low-level features (e.g. interactions between the frequency bands) whose interactions are potentially shared across the different domains. Given the small dataset at hand (17 independent bearing tests), this is the approach taken.

### 4. UNSUPERVISED TRANSFER LEARNING

An interesting subproblem within transfer learning is the transferability in scenarios where labels are available in the source domain but not in the target domain. This problem can arise in the industry when a sensing technology has to be changed, and labeling new data becomes costly or undesirable.

These techniques use information about the feature distributions in the source and domain and try to align them. One of the earliest examples is the structural correspondence learning.<sup>43</sup> This paradigm was originally proposed for natural language processing tasks and consists of a representation learning algorithm, where auxiliary classification tasks are used to identify a set of features that are informative in both domains. The features are aligned across domains by doing a linear transformation. More recent work has focused on ways to achieve similar results over neural networks that inherently can generate feature subspaces, which is the core idea behind representation learning techniques.

#### 4.1. Gradient reversal

The gradient reversal (ReverseGrad) method uses a neural network with two branches, one of which needs to perform a classification task over the labels of the data (label predictor), while the second branch needs to predict the domain (domain classifier).<sup>44</sup> A special layer, named gradient reversal, is used on the domain classifier branch. This layer leaves the data unchanged during the forward pass and reverses the gradients during propagation. This effectively causes the network to become less effective at classifying the domain, a concept known as domain confusion. Furthermore, it serves as a way to decrease the discrepancy in the feature distributions between domains at intermediate layers. Fig. 2 shows the architecture, notice that the upper branch solves the category classification task and the lower branch the

## Unsupervised transfer learning for bearing monitoring

domain classification task. The outcome of this approach produces feature maps that are agnostic to the domain, which in turns bypasses the need of labeling the target domain data.

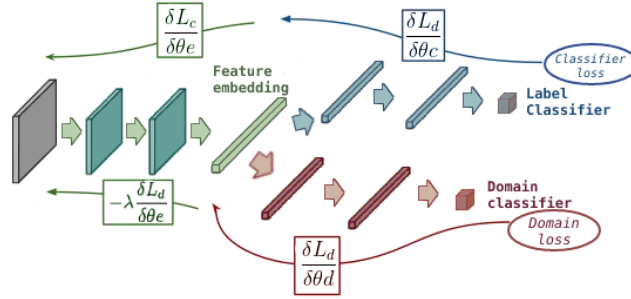


Figure 2: Schematic of the gradient reversal architecture. Figure based on the work by Ganin et al.<sup>44</sup>

In practice gradient reversal can be achieved by either of the following two strategies:

1. Using a function that behaves as a skip layer during inference and applies a  $-\lambda$  multiplier at the domain classifier head during training. The lambda ( $\lambda$ ) parameter serves as a balance between the domain classifier loss and the label predictor loss.
2. Using an adversarial loss in which the domain labels are swapped during training.

According to Ganin et al.,<sup>44</sup> no significant difference was found between both strategies. This paper follows the first strategy. The lambda parameter is adjusted over epochs following Eq. (1), where gamma is a fixed value, and  $p$  represents the epoch. The purpose here is to allow the network to first find relevant features on each domain, and then start the process of domain confusion.

$$\lambda_p = \frac{2}{1 + \exp(-\gamma \cdot p)} \quad (1)$$

### 4.2. Deep correlation alignment

Deep correlation alignment (Coral) is a more recent algorithm that introduces a loss term in the loss function which aligns the second order moments of the data.<sup>45</sup> The advantages compared to ReverseGrad are that it does not require the two branches of the network, which in turn reduces complexity.

Coral loss is defined as the difference between the distances of sample covariances per batch in Eq. (2), where  $C_S$  and  $C_T$  are the covariance matrices of the source and target,  $d$  is the number of features, and  $\|\cdot\|_F^2$  corresponds to the squared matrix Frobenius norm.

$$\ell_{Coral} = \frac{1}{4d^2} \|C_S C_T\|_F^2 \quad (2)$$

## Unsupervised transfer learning for bearing monitoring

The coral loss and task’s loss are balanced following Eq. (3), where  $\lambda$  is a weighting factor, and  $t$  is the number of layers to which the coral loss is applied. The coral loss offers the flexibility of selecting layers where to apply the loss. Similar to gradient reversal, it is important to notice that there is a trade-off between merging clusters from both domains and achieving the task. The work by Sun et al. does not provide guidance on how to determine an adequate  $\lambda$  value for the different layers or its possible effects.<sup>45</sup>

$$\ell = \ell_{task} + \sum_{i=1}^t \lambda_i \ell_{Coral} \quad (3)$$

### 5. DATASET

The Smart Maintenance Living Lab (SMLL) is an open test research platform that aims to assist in the adoption of condition monitoring technologies.<sup>2</sup> The platform consists of a fleet of seven identical drive-train setups that perform accelerated lifetime tests on bearings. The fleet offers three advantages: first, it enables faster data collection; second, the drive-train systems can exhibit variability, providing the opportunity to train and evaluate robust models; and finally, the most recent data collected includes recordings of vibration and acoustics, enabling the evaluation of acoustic data as a potential alternative to vibration data.<sup>2,46</sup> This dataset was previously evaluated for the remaining useful life estimation task using a set of engineered features, as well as for transfer learning across different operating conditions.<sup>42</sup> At the time of writing, the SMLL dataset consists of 145 bearing tests; however, data collection is ongoing. Out of these 145 test runs, there are acoustic recordings for 23 tests. For some of the recordings, multiple bearings are tested at the same time, which has an impact on the acoustic recordings. This is done intentionally, as one of the aspects of the SMLL platform is the ability to collect data in parallel. Additionally, this allows the generation of more challenging scenarios for the design and evaluation of PdM solutions. From the 23 tests with acoustic recordings, 7 tests are performed with a secondary test running in parallel.

Fig. 3 shows the testing station. The station is covered by a protective plastic to reduce the environmental noise. Two microphones are oriented towards the bearing test stations, one within the enclosure (*internal*) and one outside (*external*).

#### 5.1. Source and target datasets

This research focuses on the transfer from the vibration modality, which has abundant high-quality data, to the acoustic modality, which has less data and compromised quality due to environmental noise. Certain bearings are omitted either because they were tested under loads or speeds for which only a single bearing is available or due to problems with data quality. It’s worth noting that the full SMLL dataset contains additional bearings for which vibration data is available but not acoustic data. However, this additional data was not considered for this research.

The final dataset selection consists of 17 bearing recordings tested under a fixed load of 2 kN. Out of these tests, 3 were conducted at a fixed speed (2 at 2,000 rpm and one at 1,900 rpm), while the rest followed the saw-tooth speed profile.

# Unsupervised transfer learning for bearing monitoring



Figure 3: SMLL setup

Out of the 17 bearings, 2 were not indented at the beginning of the test and did not fail at the end of the test (which had a duration of 2 hours). Lastly, out of the 14 tests with a saw-tooth profile, 7 are double-tests. Table 1 provides a summary of the dataset and lists the number of bearings per condition.

For each test, only the initial 35% of the samples are used for training, validation, or testing, with a maximum of 800 samples (equivalent to approximately 2 hours and 13 minutes).

Table 1: Summary of the dataset and experiment conditions.

Bearing type	FAG 6205-C-TVH
Initial condition	- Healthy: 2 - Indented 15 ( $400 \pm 25 \mu\text{m}$ )
Stop condition	- Healthy: 2 hrs after stable temperature - Indented: Vibrations exceeding 20 g
Sampling rate	50 kHz
Acquisition frequency	Every second
Operating speed	- Fixed: 3 (2 at 2,000 rpm, 1 at 1,900 rpm) - Sawtooth: 14 (from 1,000 to 2,000 rpm)
Operating load	2 kN

## 5.2. Data collection

All tests use FAG 6205-C-TVH model bearings. Tests are conducted with two different speed profiles and varying initial conditions. Test speed can be either constant or follow a saw-tooth profile. Constant speeds are fixed at either 1,900 rpm or 2,000 rpm. In the saw-tooth profiles the speed varies from 1,000 rpm up to 2,000 rpm. Tests with a saw-tooth profile begin at a speed of 1,000 rpm and increase in increments of 100 rpm. Each speed is kept constant for 60 s, and once the speed reaches 2,000 rpm the speed is set back to 1,000 rpm. The load for the tests is fixed at 9 kN. The initial condition indicates whether the bearing is indented at the start of the test. The indentations of the bearings are meant to accelerate degradation. The indentation diameters are within  $400 \pm 25 \mu\text{m}$ . These indents are small enough for the bearing to be considered healthy at the beginning of the test but significant enough to ensure that degradation begins within a few hours. The stopping condition corresponds to the moment when peak vibrations reach a magnitude of 20 g for indented bearings or 2 hours after temperature stabilizes for not-indented bearings. Fig. 4 shows examples of

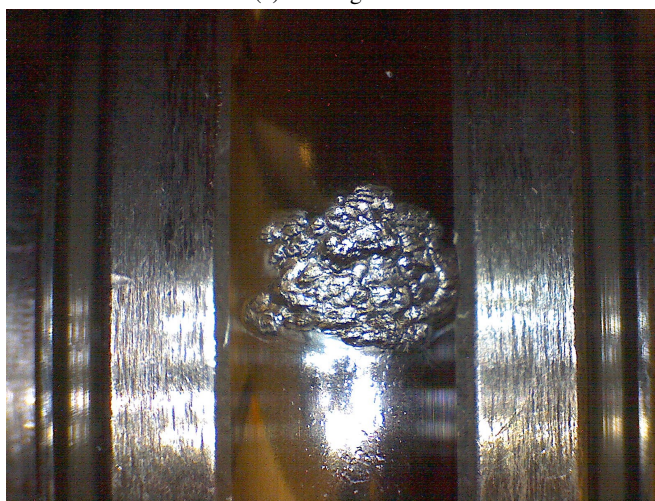
## Unsupervised transfer learning for bearing monitoring

damaged bearings after the accelerated life tests. Finally, some of the tests are conducted with a second healthy bearing tested in the adjacent drive-train. This is done in order to generate additional noise that can interfere with the recording and generate more challenging scenarios.

Tests are measured with vibration sensors, in addition to an array microphone for the most recent tests. The vibration and acoustic signal are sampled at 50 KHz.



(a) Bearing A43.



(b) Bearing A47.

Figure 4: Examples of bearings' damage at the end of the tests.

## 6. METHODOLOGY

This section discusses first the theory behind the spectrogram features and the selected feature representation, followed by the details concerning the architecture selection and parameter tuning, and the steps followed for validation and testing.

# Unsupervised transfer learning for bearing monitoring

## 6.1. Features across domains

In order for UTL techniques to succeed, there needs to be underlying shared information across domains. It can be expected that the spectrogram representation of the data from source and domain share information about the bearing condition. Sounds are mechanical vibrations that travel across a medium. When a sound changes from medium there is no change in its frequency. Eq. (4) shows the relation between frequency ( $f$ ), speed of a sound ( $v$ ) and wavelength ( $\lambda$ ). The frequency of a sound is independent of the propagation medium, while the speed of sound is defined by the medium, hence, the change in the sound corresponds to a change in the wavelength of the sound with respect to the impedance of the medium. As waves travel from the metal frame into the air (less dense medium), the sound speed decreases as well as the wavelength. Therefore, it is to be expected that the characteristic frequencies of the bearing are present in the sound recordings. However, different distortions in the spectrogram are introduced such as: attenuation in some frequency bands, shifts in the phase, introduction of harmonics, and increase in noise. Furthermore, as the microphones are exposed to other machines that are in operation, external noise is also introduced. In addition, the relative position of the microphones to the testbed can also change between tests.

$$\lambda = \frac{v}{f} \quad (4)$$

To summarize, the main distortions are caused by the change in medium impedance, which causes a drop in the transmitted energy and introduces harmonics; and external disturbances, which correspond to events not related to the bearing's operation.

## 6.2. Mel spectrograms

The vibration signals over non-overlapping windows of 10 seconds are transformed into spectrograms using the short-time Fourier transform (STFT), and are then log transformed to obtain the mel spectrogram. This representation is preferred as it reduces the number of features without a noticeable drop in accuracy, which in turn allows reducing the number of parameters for the models. In order to reduce the impact of external noises the mean of the spectrograms over each window of 10 s is used instead of the full spectrogram. This allows filtering sporadic noises, but does not cancel other noises such as the ones introduced in the double tests. Table ?? in the annexes summarizes the data processing steps. Fig. 5 shows the mel spectrogram representation for a window of 10 s. Notice how the high frequency part of the spectrum is compressed (binned). In addition, notice the considerable magnitude difference between the vibration and the acoustic domain (attenuation).

## 6.3. Speed identification

The task that the models need to achieve is the correct identification of the operating speed during healthy operating conditions. The target is defined as the smoothed one-hot encoded vector of the speeds. Due to expected similarity in the frequency content between adjacent speeds, the target vector is softened by a smoothing factor ( $\alpha$ ). Notice

# Unsupervised transfer learning for bearing monitoring

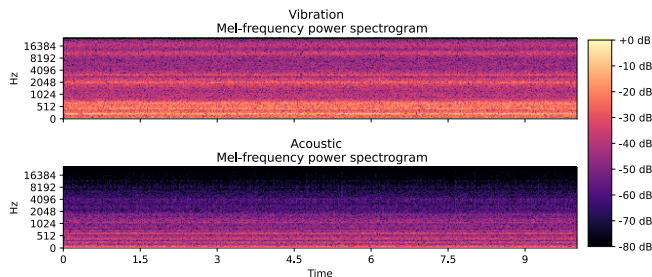


Figure 5: Vibration (top) and acoustic (top) signals representations as mel spectrograms. Notice how the high-frequency band gets compressed over the logarithmic scale, and how a considerable attenuation of the signal occurs when changing the medium.

that the edge cases only have one adjacent neighboring speed. Label smoothing has previously shown to improve the performance via generalizations.<sup>26</sup>

## 6.4. Parameter tuning and validation

The selected architecture is an Artificial Neural Network (ANN). The parameter search is done using 5-times repeated 6-Fold cross-validation over the values shown in Table 2. The repeated fold procedure is selected to obtain better out-of-bag (OOB) error estimations by ignoring variations caused by the network’s weights random initialization. The number of folds was selected in order to test no more than 3 bearings as OOB samples per fold. Consider that the folds are intended to be balanced in the number bearings each contains, but due to changes in the test’s lengths the actual number of samples varies. The dataset partition with the bearing’s ids and the dataset size are shown in Annex Tables 8 and 9. In order to keep a consistent unsupervised learning methodology, the best parameters are selected based on the validation error over the source domain, as selecting it based on the error over the acoustic domain would defeat the purpose of a real UTL. Parameter tuning is only performed for the ReverseGrad and Coral models. For simplicity, the baselines use the same architecture and parameters as the best result found for the ReverseGrad. The fine-tuning baseline reuses the model learned from the baseline and is adapted using the training data of the target domain.

Table 2: Parameter search grid. Details of the feature extractor are detailed in Table 3.

Parameter	Values
2 layers feature extractor	False: Single layer in feature extractor True: Two layers in feature extractor
Embedding dimension	64, 128
Dropout rate	0, 0.1, 0.2

The loss functions for each model are the following:

1. Baseline model. Categorical cross entropy with class weights.
2. Baseline using fine-tuning. Categorical cross entropy with class weights.
3. ReversGrad. Categorical cross entropy with class weights for labels and binary cross entropy for domains with  $\gamma = 10$ .

Table 3: Architecture for each model or branch. Each row is the number of nodes at the corresponding level. Layers with \* have shapes or operating conditions dependent on the parameter tuning shown in Table 2.

		<b>Baseline</b>	
	<b>Model (or branch)</b>	<b>Fine tuning baseline</b>	<b>ReverseGrad</b>
		<b>ReverseGrad (Class head)</b>	<b>(Domain head)</b>
		<b>Coral</b>	
Feature extractor	Input	(256,)	-
	Dense	(128,)	-
	Batch Norm,	(128,)	-
	Optional	(128,) if emb. dim. == 128	-
	Dense*	(64,) if emb. dim. == 64	-
Classifier	Dense*	(emb. dim.,)	(emb. dim.,)
	Dense*	(emb. dim./2,)	(emb. dim./2,)
	Batch Norm.	(emb. dim./2,)	(emb. dim./2,)
	Dropout*	(emb. dim./2,)	(emb. dim./2,)
	Output	(10,)	(2,)

4. Coral. Categorical cross entropy with class weights for labels and coral loss with  $\lambda = 10$ .

For all the tests the validation metric and early stop criterion is the top-k-categorical accuracy with  $k = 2$  with a patience of 10 epochs, after which the best found weights are restored before performing the OOB evaluation. Top-k-categorical accuracy is selected following the same reasoning as the soft-labeling, which reduces the penalty imposed by making errors between adjacent speeds.

## 7. RESULTS AND DISCUSSION

The following section presents the results for the baselines and the UTL techniques. First the individual results of each of the TL methods are presented, namely ReverseGrad in Section 7.1 and Coral in Section 7.2. These sections describe the best found parameters for each approach and their maximum performance. Finally, Section 7.3 presents the results of the baselines, namely the model with adaptation and the model with supervised adaptation, which then are compared against the transfer learning models. This section highlights the advantages found and how each model compares against the others.

### 7.1. ReverseGrad

The results for the parameter search of the ReverseGrad model are summarized in Table 4 and presented in Fig. 6. The most significant parameter is the number of layers for the feature extractor, and to a lesser extent the dropout rate. However, on the target’s domain accuracy, the embedding size seems to also have an effect. The best parameter found for the ReverseGrad model corresponds to a single layer feature extractor, with a dropout rate of 0.2, and an embedding layer of 128 nodes. This model gives a top-2 validation accuracy of 97.65% in the source domain and 74.56% in the validation target.

Table 4: Cross-validation results for the validation error of the ReverseGrad tests. Metric corresponds to Top-2 categorical accuracy. Results have been ordered based on the source accuracy mean.

2 layers feature	Dropout rate	Embedding size	Val. Source accuracy	Standard dev.	Val. Target accuracy	Standard dev.
True	0.1	64	77.23	11.94	48.77	9.69
True	0.0	64	80.32	15.67	51.50	10.85
True	0.2	64	82.16	11.10	49.84	11.69
True	0.0	128	85.09	8.58	53.65	9.26
True	0.2	128	86.73	12.19	59.18	11.16
True	0.1	128	87.41	7.21	57.27	9.04
False	0.0	64	95.69	3.57	62.72	11.08
False	0.2	64	96.48	2.60	66.36	7.63
False	0.1	64	96.52	3.21	67.52	7.65
False	0.0	128	96.61	3.73	70.94	7.62
False	0.1	128	97.23	2.26	68.66	6.92
False	0.2	128	97.65	2.14	74.58	6.76

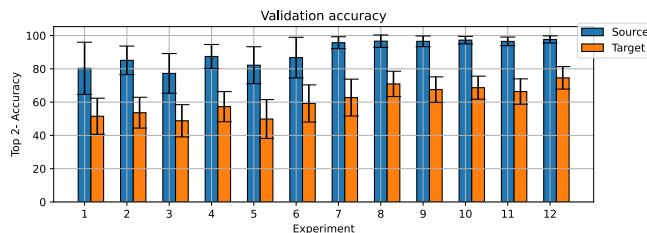


Figure 6: Validation results of the search architecture for the ReverseGrad experiments. Experiments ordered as in Table 4.

## 7.2. Coral

The results for the parameter search of the Coral model are summarized in Table 5 and presented in Fig. 7. The parameter search points out that the only significant parameter for the source’s domain performance is the embedding size, while the other parameters seem to have no effect on the target’s domain performance. The best parameter found corresponds to a single layer feature extractor, with no dropout, and an embedding layer of 128 nodes.

Based on the validation error over the source domain, the Coral method underperforms against ReverseGrad. Meaning that the coral loss has caused the model to perform worse over the source domain. In addition, the performance over the target domain is low over the validation data (55.25%).

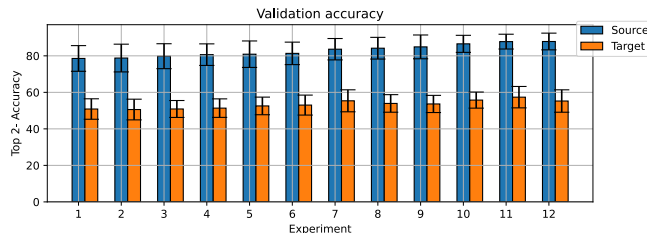


Figure 7: Validation results of the search architecture for the Coral experiments. Experiments ordered as in Table 5

## Unsupervised transfer learning for bearing monitoring

Table 5: Cross-validation results for the validation error of the Coral tests. Metric corresponds to Top-2 categorical accuracy. Results have been ordered based on the source accuracy mean.

2 layers feature	Dropout rate	Embedding size	Val. Source accuracy	Standard dev.	Val. Target accuracy	Standard dev.
True	0.2	64	78.56	7.03	50.87	5.60
True	0.0	64	78.78	7.59	50.58	5.66
True	0.1	64	79.80	6.83	50.89	4.63
False	0.0	64	80.64	5.91	51.35	5.07
False	0.2	64	80.90	7.24	52.56	4.82
False	0.1	64	81.34	6.15	53.02	5.47
True	0.0	128	83.63	5.85	55.36	6.02
True	0.2	128	84.18	5.94	53.92	4.80
True	0.1	128	84.90	6.51	53.64	4.70
False	0.2	128	86.56	4.66	55.75	4.40
False	0.1	128	87.78	4.05	57.36	5.85
False	0.0	128	87.85	4.57	55.25	6.13

### 7.3. Baselines and OOB evaluation

Table 6 summarizes the results for the baselines and UTL techniques over the train, validation and test for each domain. The table compares the best-found model for each of the techniques based on the validation performance over the source domain as presented in the previous section.

To start the analysis, a first reference value for the OOB accuracy is the accuracy obtained from a random guess estimator. Due to the 2-top accuracy metric not being evenly distributed, the expected value is estimated from simulating a 2 million sample distribution with random guess for the evaluated. The expected top-2 accuracy for a random guesser is 18.16% with no significant difference if the class distributions were taken into account. All the models evaluated surpass this value, therefore it is clear that the models have successfully learned meaningful information from the spectrogram representation.

The first baseline, in which no transfer learning was performed, is able to generalize correctly and has a good performance for the source domain. Interestingly, this model also has a good performance in the target domain (accuracy of 85.51% on the validation and 83.18% on the test). This is a clear indicator that the two domains have relatively small differences. It is likely that the Batch Normalization layers are able to compensate for part of the domain shift. The second baseline, the STL model via fine tuning, achieves as expected the best performance on the target domain test (93.25%). This value can be seen as a reference of what would be the best expected accuracy for the UTL models. Notice that the standard deviations in the OOB tests for both source and domain are considerably large, even after having performed repeated folds to get a better estimate. This points to considerable heterogeneity between different tests.

**ReverseGrad** The results for ReverseGrad indicate an accuracy of 91.73% for the test data in the target domain, which aligns with the source accuracy and is notably close to the performance of the STL baseline. It is worth noting that the standard deviations over the test data are relatively large (greater than 10%). An aspect that needs to be further investigated concerns the relatively poor performance over the validation data on the target (74.58%). This discrepancy

## Unsupervised transfer learning for bearing monitoring

Table 6: Mean top-2 accuracy and standard deviation in parenthesis for the train, validation and out-of-bag test. The training metric is not evaluated at the target, as the information is not available during the co-training. Validation is provided for reference but in practice it is neither available. Fine-tuning results at the source correspond to those of the baseline.

Model	Source (Vibration)			Target (Acoustic)		
	Training	Validation	Test	Training	Validation	Test
Baseline	96.26 (0.30)	99.22 (0.58)	92.03 (8.99)	N/A	85.51 (3.99)	83.18 (8.94)
Fine-tuning	- (-)	- (-)	- (-)	N/A	- (-)	93.25 (7.73)
ReverseGrad	95.50 (0.03)	97.65 (2.14)	91.64 (9.11)	N/A	74.58 (6.76)	91.73 (9.07)
Coral	74.87 (2.65)	87.85 (4.57)	88.18 (11.10)	N/A	55.25 (6.13)	87.27 (11.10)

may be attributed to the small dataset size and potential imbalance in the folds. Notice that the partitions were done at random but due to the small number of tests it can be prone to imbalanced folds.

Despite these limitations, the ReverseGrad model successfully transfers acquired knowledge from the source to the target domain. This is evident when comparing the improvements in the source domain’s accuracy to the corresponding increases in the target domain’s accuracy, as shown in Fig. 6.

**Coral** The results for Coral demonstrate an accuracy of 87.27% for the test data in the target domain. This performance is notably lower than that achieved by the ReverseGrad method (91.73%), yet it surpasses the baseline without transfer (83.18%). It is hypothesized that the poor performance may be attributed to a limited parameter search. Furthermore, it is possible that the Batch Normalization layers are interfering with the Coral loss.

In summary, it is essential to emphasize that the substantial variability in the target domain indicates the persistence of significant heterogeneity within that domain. This heterogeneity contributes to the observed large standard deviations in performance estimates.

## 8. FUTURE WORK

This study serves as an illustrative example of applying unsupervised transfer learning to transfer models between different sensor modalities without requiring labeled data. However, in practical applications, machine speed is typically a user input. This raises the question of what practical benefits can be derived from the ability to predict the operating speed.

An early objective of this research was the identification of anomalies in an unsupervised way. The original idea was to use the embedding generated after the feature extraction stage for a downstream task of single class classification. The detection then could be reported as the moment of degradation onset. However, the changes between two speeds generate states which, although not anomalous, differ from the learned embedding and are immediately labeled as anomalies (early false positive). The results of this evaluation were insufficient and are not reported here.

To address these questions and advance this research, two paths for future research are proposed:

## Unsupervised transfer learning for bearing monitoring

- Condition based monitoring based on the deviation between the predicted and the known speed. This output can be seen as a deviation between the expected speed and the actual and can be used in order to diagnose possible errors in the machine such as dose caused by component wear and lubricant quality.
- Anomaly detection by means of a single-class classifier.<sup>47</sup> The goal is to use the embedding layer that represents all the healthy operation for all possible speeds and then use the embedding for the downstream task.
- Evaluating the external microphone. This research used the audio recordings within the enclosure with the aim of reducing the impact of environmental noise, which is still recorded and proved to be a challenge. In order to create realistic solutions for PdM, it would be required to use the data of the external microphone, which is likely to contain more noise.

## 9. CONCLUSION

This work presents the first study of transfer learning across different non-image modalities in the context of PdM. This research opens the possibility of performing monitoring tasks of mechanical components using microphones, which offer advantages over vibrations sensors in terms of sensor costs and ease of deployment. In order to obtain better predictive models for the acoustic domain, the information of the vibration data is used in a transfer learning scenario. The evaluated methodologies show different degrees of performance. The performance of the ReverseGrad method is almost as good as the fine tuning approach (top 2-categorical accuracy of 91.73% against 93.25%). On the other hand, the results of the Coral technique were mixed, having a low performance on validation accuracy and slightly above the baseline without transferral. The poor performance could be caused by a not thorough search in parameters or conflicts between the batch normalization layers and the coral loss function. It is likely that a more extensive parameter search could yield better results for the Coral model.

This research presents a first study of unsupervised transfer learning techniques for non-image modalities with the aim of creating embeddings that can summarize healthy behavior of bearings. This is a first step towards unsupervised transferral for predictive maintenance applications.

## ACKNOWLEDGMENTS

The authors would like to thank Bram Robberechts for performing the tests within the Smart Maintenance Living Lab project. The photos shown in Figure 3 and Figure 4 belong to the SMLL dataset. The authors thank Flanders Make for allowing the photos to be included. This work was supported by the Flemish Government under the “Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen” programme and the VLAIO subsidies for research.

*Conflict of interest* The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## Unsupervised transfer learning for bearing monitoring

Table 7: Pre-processing steps.

Step	Parameters used
Data capturing	Vibration and acoustic data collected at 50 KHz
Filtering	Bandpass Butterworth filter (order 5) with cutoffs at 500 and 12 KHz
Data windowing	Non-overlapping 10 s windows
Spectrogram	Hann windowing with of segments with length of 0.6 s with 50% overlap
Mel transform	256 mel-bands. Linear bands up to 1 KHz, with logarithmic bands for higher frequencies. Bands defined by: $mel = 2595 * \log_{10}(1 + f/700)$

## ANNEX

Table 8: Fold divisions and corresponding bearing ids. Due to small delays between the vibration recordings, there may be small differences in the sample size between domains for the same bearing.

Fold	Train Source	Train Target	Val Source Val Target	Test Source Test target
1	A83 A85 A147 A148 A150 A154 A155 A158 A181	A184 A146 A151 A152 A156	A147 A148 A181	A149 A153 A182
2	A84 A85 A146 A148 A151 A152 A153 A154 A158	A83 A149 A155 A181 A182	A148 A151 A155	A147 A150 A156
3	A83 A146 A150 A151 A153 A154 A158 A181 A182	A85 A147 A148 A149 A156	A147 A148 A151	A84 A152 A155
4	A84 A85 A147 A148 A149 A150 A155 A158 A182	A146 A152 A153 A154 A156	A147 A149 A156	A83 A151 A181
5	A83 A84 A85 A149 A150 A153 A155 A156 A181	A147 A151 A152 A158 A182	A147 A151 A181	A146 A148 A154
6	A146 A147 A148 A149 A150 A151 A152 A153 A156	A83 A84 A154 A155 A181 A182	A149 A153 A182	A85 A158

Table 9: Sample size for each fold

<b>Fold #</b>	<b>Training source</b>	<b>Training target</b>	<b>Val. source Val. target</b>	<b>Test source Test target</b>
1	4935	1997	1155	1451
2	4780	1971	1503	1632
3	4100	2483	1503	1800
4	5068	2186	1800	1129
5	4560	2652	1452	1171
6	4720	2463	1451	1200

### References

- [1] Murphy, C. Choosing the most suitable predictive maintenance sensor. *Analog Devices*. (2020).
- [2] Ooijselaar, T. H., Pichler, K., Di, Y., Devos, S., Volckaert, B., Van Hoecke, S., and Hesch, C. Smart machine maintenance enabled by a condition monitoring living lab. *IFAC-PapersOnLine*. **52**, (15). (2019). doi: <https://doi.org/10.1016/j.ifacol.2019.11.704>.
- [3] Heng, A., Zhang, S., Tan, A. C., and Mathew, J. Rotating machinery prognostics: State of the art, challenges and opportunities. *Mechanical Systems and Signal Processing*. **23**, (3). (2009). doi: <https://doi.org/10.1016/j.ymsp.2008.06.009>.
- [4] Jardine, A. K., Lin, D., and Banjevic, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*. **20**, (7). (2006). doi: <https://doi.org/10.1016/j.ymsp.2005.09.012>.
- [5] Zhang, R., Peng, Z., Wu, L., Yao, B., and Guan, Y. Fault diagnosis from raw sensor data using deep neural networks considering temporal coherence. *Sensors*. **17**, (3). (2017). doi: <https://doi.org/10.3390/s17030549>.
- [6] Gebraeel, N., Lawley, M., Liu, R., and Parmeshwaran, V. Residual life predictions from vibration-based degradation signals: A neural network approach. *IEEE Transactions on Industrial Electronics*. **51**, (3). (2004). doi: <https://doi.org/10.1109/TIE.2004.824875>.
- [7] Huang, R., Xi, L., Li, X., Richard Liu, C., Qiu, H., and Lee, J. Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods. *Mechanical Systems and Signal Processing*. **21**, (1). (2007). doi: <https://doi.org/10.1016/j.ymsp.2005.11.008>.
- [8] Sateesh Babu, G., Zhao, P., and Li, X.-L. Deep convolutional neural network based regression approach for estimation of remaining useful life. *Database Systems for Advanced Applications*. **9642**. (2016). doi: [https://doi.org/10.1007/978-3-319-32025-0\\_14](https://doi.org/10.1007/978-3-319-32025-0_14).
- [9] Yoo, Y. and Baek, J.-G. A novel image feature for the remaining useful lifetime prediction of bearings based on continuous wavelet transform and convolutional neural network. *Applied Sciences*. **8**, (7). (2018). doi: <https://doi.org/10.3390/app8071102>.
- [10] Guo, X., Chen, L., and Shen, C. Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. *Measurement: Journal of the International Measurement Confederation*. **93**. (2016). doi: <https://doi.org/10.1016/j.measurement.2016.07.054>.
- [11] Xia, M., Li, T., Xu, L., Liu, L., and De Silva, C. W. Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks. *IEEE/ASME Transactions on Mechatronics*. **23**, (1). (2018). doi: <https://doi.org/10.1109/TMECH.2017.2728371>.

- [12] Wen, L., Li, X., Gao, L., and Zhang, Y. A new convolutional neural network-based data-driven fault diagnosis method. *IEEE Transactions on Industrial Electronics*. **65**, (7). (2018). doi: <https://doi.org/10.1109/TIE.2017.2774777>.
- [13] Guo, L., Lei, Y., Xing, S., Yan, T., and Li, N. Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data. *IEEE Transactions on Industrial Electronics*. **66**, (9). (2019). doi: <https://doi.org/10.1109/TIE.2018.2877090>.
- [14] Eren, L., Ince, T., and Kiranyaz, S. A generic intelligent bearing fault diagnosis system using compact adaptive 1d cnn classifier. *Journal of Signal Processing Systems*. **91**, (2). (2019). doi: <https://doi.org/10.1007/s11265-018-1378-3>.
- [15] Xu, G., Liu, M., Jiang, Z., Söffker, D., and Shen, W. Bearing fault diagnosis method based on deep convolutional neural network and random forest ensemble learning. *Sensors*. **19**, (5). (2019). doi: <https://doi.org/10.3390/s19051088>.
- [16] Li, S., Liu, G., Tang, X., Lu, J., and Hu, J. An ensemble deep convolutional neural network model with improved d-s evidence fusion for bearing fault diagnosis. *Sensors*. **17**, (8). (2017). doi: <https://doi.org/10.3390/s17081729>.
- [17] Hemmer, M., Van Khang, H., Robbersmyr, K., Waag, T., and Meyer, T. Fault classification of axial and radial roller bearings using transfer learning through a pretrained convolutional neural network. *Designs*. **2**, (4). (2018). doi: <https://doi.org/10.3390/designs2040056>.
- [18] Chen, R., Huang, X., Yang, L., Xu, X., Zhang, X., and Zhang, Y. Intelligent fault diagnosis method of planetary gearboxes based on convolution neural network and discrete wavelet transform. *Computers in Industry*. **106**. (2019). doi: <https://doi.org/10.1016/j.compind.2018.11.003>.
- [19] Zhao, X. and Jia, M. A novel unsupervised deep learning network for intelligent fault diagnosis of rotating machinery. *Structural Health Monitoring*. **19**, (6). (2020). doi: <https://doi.org/10.1177/1475921719897317>.
- [20] Zhang, K., Tang, B., Deng, L., and Liu, X. A hybrid attention improved ResNet based fault diagnosis method of wind turbines gearbox. *Measurement*. **179**. (2021). doi: <https://doi.org/10.1016/j.measurement.2021.109491>.
- [21] Pinedo-Sánchez, L. A., Mercado-Ravell, D. A., and Carballo-Monsivais, C. A. Vibration analysis in bearings for failure prevention using CNN. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*. **42**, (12). (2020). doi: <https://doi.org/10.1007/s40430-020-02711-w>.
- [22] Zhang, W., Peng, G., Li, C., Chen, Y., and Zhang, Z. A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors*. **17**, (2). (2017). doi: <https://doi.org/10.3390/s17020425>.
- [23] Guo, Y., Shi, H., Kumar, A., Grauman, K., Rosing, T., and Feris, R. Spottune: Transfer learning through adaptive fine-tuning. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. (2019). doi: <https://doi.ieeecomputersociety.org/10.1109/CVPR.2019.00494>.

- [24] Yosinski, J., Clune, J., Bengio, Y., and Lipson, H. How transferable are features in deep neural networks? *Proceedings of the 27th International Conference on Neural Information Processing Systems*. **2**. (2014). doi: <https://dl.acm.org/doi/10.5555/2969033.2969197>.
- [25] He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. (2016). doi: <https://doi.org/10.1109/CVPR.2016.90>.
- [26] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, Los Alamitos, CA, USA, Rethinking the inception architecture for computer vision. (2016) 2818–2826. doi: <https://doi.ieeecomputersociety.org/10.1109/CVPR.2016.308>.
- [27] Zhang, K., Tang, B., Deng, L., Tan, Q., and Yu, H. A fault diagnosis method for wind turbines gearbox based on adaptive loss weighted meta-ResNet under noisy labels. *Mechanical Systems and Signal Processing*. **161**. (2021). doi: <https://doi.org/10.1016/j.ymsp.2021.107963>.
- [28] Hong, F., Song, J., Meng, H., Rui, W., Fang, F., and Guangming, Z. A novel framework on intelligent detection for module defects of PV plant combining the visible and infrared images. *Solar Energy*. **236**. (2022). doi: <https://doi.org/10.1016/j.solener.2022.03.018>.
- [29] Marei, M., Zaatari, S. E., and Li, W. Transfer learning enabled convolutional neural networks for estimating health state of cutting tools. *Robotics and Computer-Integrated Manufacturing*. **71**. (2021). doi: <https://doi.org/10.1016/j.rcim.2021.102145>.
- [30] Janssens, O., Van De Walle, R., Loccupier, M., and Van Hoecke, S. Deep learning for infrared thermal image based machine health monitoring. *IEEE/ASME Transactions on Mechatronics*. **23**, (1). (2018). doi: <https://doi.org/10.1109/TMECH.2017.2722479>.
- [31] Lu, W., Liang, B., Cheng, Y., Meng, D., Member, S., and Yang, J. Deep model based domain adaptation for fault diagnosis. *IEEE Transactions on Industrial Electronics*. **64**, (3). (2017). doi: <https://doi.org/10.1109/TIE.2016.2627020>.
- [32] Hu, Q., Si, X., Qin, A., Lv, Y., and Liu, M. Balanced Adaptation Regularization Based Transfer Learning for Unsupervised Cross-Domain Fault Diagnosis. *IEEE Sensors Journal*. **22**, (12). (2022). doi: [10.1109/JSEN.2022.3174396](https://doi.org/10.1109/JSEN.2022.3174396).
- [33] Zhu, Y., Zhu, C., Tan, J., Tan, Y., and Rao, L. Anomaly detection and condition monitoring of wind turbine gearbox based on LSTM-FS and transfer learning. *Renewable Energy*. **189**. (2022). doi: [10.1016/J.RENENE.2022.02.061](https://doi.org/10.1016/J.RENENE.2022.02.061).
- [34] Li, X., Zhang, W., Ma, H., Luo, Z., and Li, X. Partial transfer learning in machinery cross-domain fault diagnostics using class-weighted adversarial networks. *Neural Networks*. **129**. (2020). doi: <https://doi.org/10.1016/j.neunet.2020.06.014>.
- [35] Case Western Reserve University Bearing Data Center. Bearing data center: Seeded fault test data. (2009).

- [36] Li, X., Zhang, W., Ding, Q., and Sun, J. Q. Multi-Layer domain adaptation method for rolling bearing fault diagnosis. *Signal Processing*. **157**. (2019). doi: <https://doi.org/10.1016/j.sigpro.2018.12.005>.
- [37] Deng, Y., Huang, D., Du, S., Li, G., Zhao, C., and Lv, J. A double-layer attention based adversarial network for partial transfer learning in machinery fault diagnosis. *Computers in Industry*. **127**. (2021). doi: <https://doi.org/10.1016/j.compind.2021.103399>.
- [38] Lessmeier, C., Kimotho, J., Zimmer, D., and Sextro, W. 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. *Proceedings of the European Conference of the Prognostics and Health Management Society*. **3**. (2016). doi: <https://doi.org/10.36001/phme.2016.v3i1.1577>.
- [39] Wang, B., Lei, Y., Li, N., and Li, N. A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. *IEEE Transactions on Reliability*. **69**, (1). (2020). doi: <https://doi.org/10.1109/TR.2018.2882682>.
- [40] Nectoux, P., Gouriveau, R., Medjaher, K., Ramasso, E., Chebel-Morello, B., Zerhouni, N., and Varnier, C. PRONOSTIA: An experimental platform for bearings accelerated degradation tests. *Conference on Prognostics and Health Management*. (2012).
- [41] J, L., H, Q., G, Y., J, L., and Rexnord Technical Services. *IMS, University of Cincinnati. "Bearing Data Set"*. NASA Ames Research Center, Moffett Field, CA, 2007.
- [42] Nieves Avendano, D., Vandermoortele, N., Soete, C., Moens, P., Ompusunggu, A. P., Deschrijver, D., and Van Hoecke, S. A semi-supervised approach with monotonic constraints for improved remaining useful life estimation. *Sensors*. **22**, (4). (2022). doi: <https://doi.org/10.3390/s22041590>.
- [43] Blitzer, J., McDonald, R., and Pereira, F. Domain Adaptation with Structural Correspondence Learning. *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*. (2006). doi: <https://doi.org/10.3115/1610075.1610094>.
- [44] Ganin, Y. and Lempitsky, V. Unsupervised domain adaptation by backpropagation. *Proceedings of the 32nd International Conference on International Conference on Machine Learning*. **37**. (2015). doi: <https://dl.acm.org/doi/10.5555/3045118.3045244>.
- [45] Sun, B. and Saenko, K. Computer Vision – ECCV 2016 Workshops, Eds. Hua, G. and Jégou, H. Springer International Publishing, Cham, Deep coral: Correlation alignment for deep domain adaptation. (2016) 443–450. doi: [https://doi.org/10.1007/978-3-319-49409-8\\_35](https://doi.org/10.1007/978-3-319-49409-8_35).
- [46] Moens, P., Bracke, V., Soete, C., Vanden Hautte, S., Nieves Avendano, D., Ooijevaar, T., Devos, S., Volckaert, B., and Van Hoecke, S. Scalable fleet monitoring and visualization for smart machine maintenance and industrial iot applications. *Sensors*. **20**, (15). (2020). doi: <https://doi.org/10.3390/s20154308>.
- [47] Perera, P. and Patel, V. M. Learning deep features for one-class classification. *IEEE Transactions on Image Processing*. **28**, (11). (2019). doi: <https://doi.org/10.1109/TIP.2019.2917862>.