

Simulation-to-real unsupervised domain adaptation gear diagnostics based on entropy

Henrique Duarte Vieira de Sousa^{1,2}, Rui Zhu^{1,2}, Toby Verwimp^{1,2}, Alexandre Maurício^{1,2}, Hao Wen^{1,2}, Seyed Ali Hosseinli^{1,2} and Konstantinos Gryllias^{1,2}

¹ *Department of Mechanical Engineering, KU Leuven*

henriqueduarte.vieiradesousa@kuleuven.be

rui.zhu@kuleuven.be

toby.verwimp@kuleuven.be

alex.ricardomauricio@kuleuven.be

hao.wen@kuleuven.be

ali.hosseinli@kuleuven.be

konstantinos.gryllias@kuleuven.be

² *Flanders Make @ KU Leuven*

Celestijnenlaan 300, Box 2420, 3001 Leuven, Belgium

ABSTRACT

Gearbox failures are a major contributor to the levelized cost of electricity of wind turbines, owing to their long repair time and large share of total turbine downtime. Condition monitoring enables early fault detection, reducing unplanned downtime and the associated economic impact, while digital twins provide virtual representations of physical assets that support advanced diagnostics. This work couples the two: a phenomenological gear vibration model generates a synthetic dataset spanning healthy operation and several gear pitting severities, and healthy measurements from the real asset are used to estimate a transfer function, through cepstrum liftering, that is applied to the synthetic signals so that they better reproduce the experimental response. By stochastically varying the model parameters within prescribed ranges, a labeled synthetic dataset is produced and used to train a deep learning unsupervised domain adaptation (UDA) network that couples adversarial training with entropy-based sample weighting and an information maximization loss, reducing the mismatch between simulated and measured data and improving diagnostic robustness on real measurements. The framework is validated on vibration data from an in-house back-to-back, single-stage helical gear test rig, in which pitting was artificially introduced on the tooth flank. Against state-of-the-art (SOTA) adversarial methods and a source-only baseline, the proposed approach more accurately distinguishes healthy op-

eration from multiple pitting severity levels, supporting its potential for diagnosis on real applications.

1. INTRODUCTION

Gearboxes transmit power by adapting the available torque and speed, and are used across many sectors, including automotive, aviation, and renewable energy. In wind energy in particular, gearbox failures are among the most impactful failure modes (Santelo, de Oliveira, Maciel, & Monteiro, 2022), contributing substantially to the levelized cost of electricity.

Consequently, condition monitoring of high-risk components is essential to maximize availability and production. It supports the detection and diagnosis of faults and the estimation of the remaining useful life (RUL) of the gears. Although these tasks can be addressed with signal processing alone, the rise of deep learning (DL), together with the growing volume of available data, has moved machine learning to the forefront of Prognostics and Health Management (PHM). The use of DL, however, raises two main challenges: the scarcity of high-quality labeled data, and the distribution shift between training and testing data, both of which limit performance (W. Li et al., 2022).

Physics-based and phenomenological models have increasingly been used to address these challenges, initially focusing on rolling-element bearings owing to their comparatively simpler dynamics. In (Gryllias & Antoniadis, 2012), a support vector machine (SVM) was trained on data from a bearing phenomenological model to distinguish healthy, inner-race, and outer-race conditions. In (Sobie, Freitas, & Nico-

Henrique Duarte Vieira de Sousa et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

lai, 2018), a bearing dynamic model derived from bond graph theory was used to train several machine learning (ML) models, which outperformed statistical classifiers. Neither work, however, addressed domain shift. Transfer learning (TL) was subsequently combined with synthetic data to tackle both challenges at once. In (Peng et al., 2025), signals from a phenomenological model were transformed into a cyclostationary map representation, which is well suited to the cyclostationary behavior of bearings and gears (Antoni, 2009), and were combined with adversarial training for bearing diagnostics; in (Hosseini, Ooijevaar, & Gryllias, 2024), a phenomenological model was paired with an adversarial domain adaptation (DA) method that encodes speed and time information to estimate the RUL.

For gears, although several models have been proposed (McFadden, 1986; D’Elia, Cocconcelli, Rubini, & Dalpiaz, 2017; Lei et al., 2018; Y. Li, Ding, He, & Lin, 2016; Koutsoupakakis, Seventekidis, & Giagopoulos, 2023; Jiang et al., 2024; Lou, Kumar, & Xiang, 2022), only a few have been integrated into ML-based diagnostics. In (Koutsoupakakis et al., 2023), data from a multibody dynamic model were combined with a convolutional neural network (CNN) to discriminate between healthy, missing-tooth, and surface-fault conditions, but without considering the domain shift. In (Jiang et al., 2024), a phenomenological model was used together with a domain mapping applied to both synthetic and real data before feeding UDA methods; the approach was evaluated against its no-mapping counterpart on several fault types (broken tooth, tooth crack, and shaft misalignment) under different operating conditions. Finally, in (Lou et al., 2022), a finite element (FE) model of gears and bearings generated synthetic data that were aligned to real data through a generative adversarial network (GAN), after which a CNN classifier was trained on several bearing and gear faults.

Overall, synthetic data approaches have focused more on bearings than on gears, and the gear studies either rely on FE models or on earlier DA methods. A phenomenological gear model is attractive precisely because its simplicity enables fast data generation and parameter tuning, yet such models have not been combined with more recent unsupervised DA methods. Motivated by this gap, this paper combines a modified version of the model of (D’Elia et al., 2017) with a novel adversarial UDA network, EM-DAAN (Entropy-measured Dynamic Adversarial Adaptation Network), and demonstrates accurate diagnosis of different gear pitting severities. The methodology overview is presented in Figure 1. The phenomenological simulator produces a labeled simulated dataset spanning the target fault classes; the real dataset is split into an unlabeled training portion and a labeled test portion held out for evaluation. The transfer function, estimated from the healthy real measurements, feeds back into the simulator as the rig-specific coloring filter, anchoring the simulation to the target rig. The

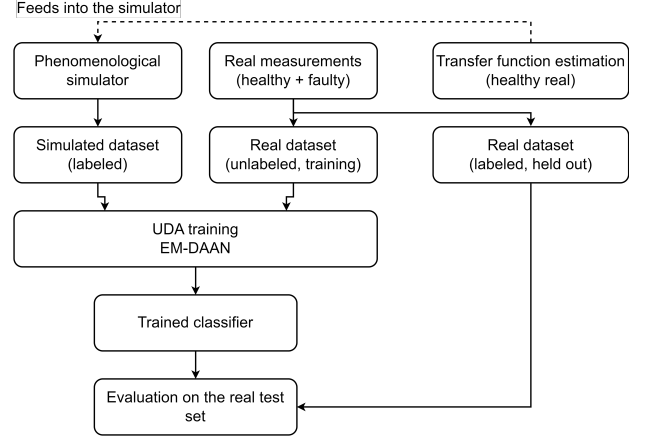


Figure 1. End-to-end methodology overview.

two datasets jointly drive an unsupervised domain-adaptation training procedure, EM-DAAN, that produces a fault classifier, which is then evaluated on the held-out real test set. The remainder of the paper is organized as follows. Section 2 reviews the state of the art for both the model and DA; Section 3 presents the contributions of this work; Section 4 reports a case study comparing the method against existing alternatives; and Section 5 concludes.

2. THEORETICAL BACKGROUND

2.1. Phenomenological Model

The phenomenological model of gear vibration aims to reproduce the meshing response of a gear pair in both healthy and faulty states. The first systematic description was given by (Randall, 1982), who decomposed the gear vibration into distinct effects: load variations arising from load sharing between teeth, tooth deflection, machining errors (such as ghost components and random profile deviations), uniform wear, and amplitude and phase modulations whose behavior differs for local and distributed faults.

Building on this, (McFadden, 1986) represented the meshing vibration as a harmonic series of the tooth-meshing frequency and introduced amplitude and phase modulation functions, $a_m(t)$ and $b_m(t)$, to describe the effect of local faults such as cracks. For a gear with T teeth rotating at shaft frequency f_{rs} , the vibration signal $g(t)$ is

$$g(t) = \sum_{k=1}^K X_k [1 + a(t)] \cos(2\pi k T f_{rs} t + \phi_k + b(t)), \quad (1)$$

where X_k and ϕ_k are the amplitude and phase of the k -th meshing harmonic, and the amplitude and phase modulation functions are

$$a(t) = \sum_{l=1}^L A_l \cos(2\pi l f_{rs} t + \alpha_l), \quad (2)$$

$$b(t) = \sum_{l=1}^L B_l \cos(2\pi l f_{rs} t + \beta_l). \quad (3)$$

(D’Elia et al., 2017) reformulated this methodology in the angular domain and added an explicit model for the impulses produced by local faults, with this contribution instead being formulated in the time domain. The impulse train caused, for example, by pitting on the tooth flank is

$$d(t) = \sum_{j=1}^J d_0 \exp\left(-\frac{(t - jT_{rs})^2}{4\sigma^2}\right) \cos(2\pi f_n (t - jT_{rs})), \quad (4)$$

where f_n is the structural resonance (carrier) frequency and $T_{rs} = 1/f_{rs}$ is the nominal shaft rotation period. The impulse is a Gaussian window of standard deviation σ and amplitude d_0 , whose extent reflects the number of affected teeth. The fault also perturbs the existing amplitude and phase modulations through a local change of $+d_0$ and $-d_0$, respectively. Under the speed fluctuation, the impulse instants are obtained from the speed profile through the angle-time relation rather than from a constant period. The load variation and distributed faults are modeled with Eqs. 2 and 3. Finally, a small speed fluctuation is superimposed on the otherwise constant operating condition to better emulate reality:

$$f_{rs}(\theta_s) = f_c + f_d \cos(f_m \theta_s), \quad (5)$$

where f_c is the carrier (mean shaft) frequency, f_d the frequency deviation, and f_m the modulation frequency. Every step except the addition of the time-domain impulses is carried out in the angular domain.

2.2. Domain Adaptation

Using simulated signals removes the dependence on scarce labeled measurements from the asset under test. On its own, however, a model cannot match real data, because the two signals differ in nature and the measurements contain behavior the model does not capture. This distribution mismatch motivates DA (Pan & Yang, 2010), which transfers knowledge from a labeled source domain $\mathcal{D}_S$ (here, the synthetic dataset) to an unlabeled target domain $\mathcal{D}_T$ (here, the real dataset), reducing the gap between their distributions and improving target performance.

This work adopts a closed-set setting, in which the two domains share the same label space ($\mathcal{Y}_S = \mathcal{Y}_T$), while both their marginal and class-conditional feature distributions differ because of the discrepancies between the phenomenological model and the real setup.

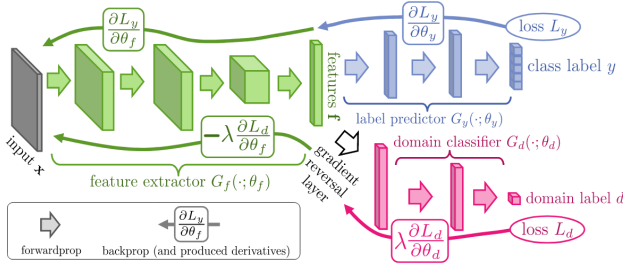
Within this setting, several categories can be distinguished, the most common being: instance-based methods, which rely on the selection or re-weighting of samples; model-based methods, which rely on sharing a common structure and parameters; and feature-based methods, which rely on the align-

ment of feature distributions (W. Li et al., 2022). The latter category is further divided into several subcategories, including discrepancy-based methods, such as MK-MMD (Long, Cao, Wang, & Jordan, 2015) and Deep CORAL (Sun & Saenko, 2016); reconstruction-based methods, relying on autoencoders, as adopted by (Wen, Gao, & Li, 2019); and adversarial-based methods, such as DANN (Ganin et al., 2016; Ganin & Lempitsky, 2015).

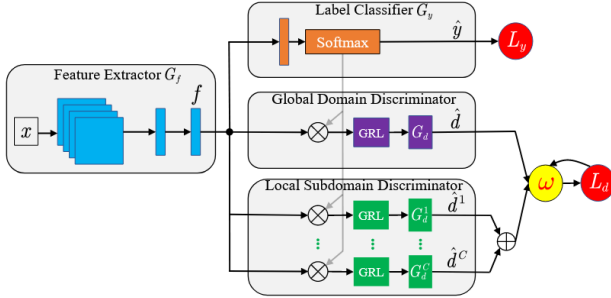
More recently, additional alternatives have also been explored. These include contrastive learning, a self-supervised method that pairs similar samples while pushing apart dissimilar ones (Ozyurt, Feuerriegel, & Zhang, 2023), and new feature extractor architectures such as transformers (Liu, Cui, Wang, & Cheng, 2025). Prototype-based DA instead uses prototypes derived from intermediate representations to allow for local alignment (Fang, Chen, et al., 2024). A departure from traditional approaches is source-free DA, in which a model is pre-trained on the source domain and subsequently adapted to the target domain without access to the source data (Fang, Yap, Lin, Zhu, & Liu, 2024). That constraint does not apply here, since the source data are synthetic and can be regenerated on demand. Finally, generative models, in particular diffusion-based approaches, have been applied to the simulation-to-real setting to synthesize target-like data or to learn a mapping between domains (Xu, Tang, & Pang, 2026). In the present framework, that role is already filled by the phenomenological model.

In this work, the focus will be on adversarial-based methods. Following the introduction of GANs (Goodfellow et al., 2014), adversarial DA emerged as a family of closed-set methods. It began with the Domain-Adversarial Neural Network (DANN) (Ganin et al., 2016; Ganin & Lempitsky, 2015) (Figure 2a) and was later extended by the Dynamic Adversarial Adaptation Network (DAAN) (Yu, Wang, Chen, & Huang, 2019) (Figure 2b).

Both comprise a feature extractor G_f , a label classifier G_y , and a domain discriminator G_d . The feature extractor, typically a deep network of convolutional and activation layers, learns a high-level latent representation. The label classifier operates on this representation and, through supervised classification of the source data, learns the decision boundaries between classes. The domain discriminator instead predicts whether a sample originates from the source or the target domain. During backpropagation, a gradient reversal layer (GRL) negates the gradient of the domain-classification loss as it flows into the feature extractor. As a result, the label classifier and the domain discriminator each minimize their own loss, while the feature extractor simultaneously minimizes the classification loss and maximizes domain confusion. This adversarial interaction yields a domain-invariant feature space in which both source and target samples can be classified.



(a) DANN (Ganin et al., 2016).



(b) DAAN (Yu et al., 2019).

Figure 2. Adversarial domain adaptation methods.

For DANN, the objective is

$$\mathcal{L}(\theta_f, \theta_y, \theta_d) = \frac{1}{N_S} \sum_{i=1}^{N_S} \mathcal{L}_y^i(\theta_f, \theta_y) - \lambda \left(\frac{1}{N_S} \sum_{i=1}^{N_S} \mathcal{L}_d^i(\theta_f, \theta_d) + \frac{1}{N_T} \sum_{i=1}^{N_T} \mathcal{L}_d^i(\theta_f, \theta_d) \right). \quad (6)$$

DAAN performs both global and class-conditional alignment, adding C local domain discriminators, one per class. Each sample is routed to the local discriminators through its classifer pseudo-probabilities, for both source and target,

$$z_c^i = p_c^i z^i. \quad (7)$$

The two alignment terms are balanced by a dynamic adversarial factor computed during training from a proxy $\mathcal{A}$ -distance (Ben-David, Blitzer, Crammer, & Pereira, 2006), evaluated with the discriminator losses:

$$d_{A,G}(\mathcal{D}_S, \mathcal{D}_T) = 2(1 - 2\mathcal{L}_{d,G}), \quad (8)$$

$$d_{A,L}(\mathcal{D}_S^c, \mathcal{D}_T^c) = 2(1 - 2\mathcal{L}_{d,L}^c), \quad (9)$$

where $\mathcal{L}_{d,G}$ is the global discriminator loss and $\mathcal{L}_{d,L}^c$ the loss of the local discriminator of class c . These yield the dynamic adversarial factor

$$\omega = \frac{d_{A,G}(\mathcal{D}_S, \mathcal{D}_T)}{d_{A,G}(\mathcal{D}_S, \mathcal{D}_T) + \frac{1}{C} \sum_{c=1}^C d_{A,L}(\mathcal{D}_S^c, \mathcal{D}_T^c)}, \quad (10)$$

and the DAAN objective

$$\mathcal{L}(\theta_f, \theta_y, \theta_d) = \frac{1}{N_S} \sum_{i=1}^{N_S} \mathcal{L}_y^i(\theta_f, \theta_y) - \lambda \left(\omega \mathcal{L}_{d,G}(\theta_f, \theta_d) + (1 - \omega) \sum_{c=1}^C \mathcal{L}_{d,L}^c(\theta_f, \theta_d^c) \right). \quad (11)$$

Equation 11 reduces to the DANN objective of Eq. 6 when $\omega = 1$, i.e., when only global alignment is active, even though the original formulation states this behavior for $\omega = 0$.

3. PROPOSED METHODOLOGY

This section builds on the methods of Section 2, extending both the synthetic data generation and the domain adaptation algorithm.

3.1. Improved Phenomenological Model

The phenomenological model is built upon that of (D'Elia et al., 2017), extending the angular-domain gearbox model in three directions:

1. A measurement chain model anchored on the target rig through an estimated transfer function and a layered noise architecture (a broadband background applied before the transfer function, a structural mode bank, and a residual sensor noise term).
2. Explicit modeling of physical mechanisms omitted by (D'Elia et al., 2017) but present in practice and noted by (Randall, 1982), namely shaft-rotation harmonics and per-tooth manufacturing variability.
3. A refinement of the fault signature following (McFadden, 1986), decoupling the per-tooth amplitude and phase modulation depths and extending the model to multi-tooth faults with per-revolution correlated jitter.

3.1.1. Transfer Function Estimation and Structural Bank

The transfer function estimation follows the background-based framework of (Matania, Klein, & Bortman, 2022; Bachar et al., 2023), in which the smoothed spectral background of a healthy measurement is interpreted as the magnitude of the transmission path filter. Whereas (Bachar et al., 2023) extract this background with the adaptive clutter separation (ACS) algorithm, (Matania et al., 2022) show that cepstrum liftering yields an equivalent estimate; the latter is adopted here.

The transfer function between the gear excitation and the accelerometer is estimated from healthy measured signals, under the assumption that the machine response in stationary operation is dominated by the structural transmission path. For each healthy record, a Welch power spectral density is

computed with a Hann window of fixed length and overlap, converted to a magnitude spectrum, and averaged across records in the log-magnitude domain. The real cepstrum of the averaged log-spectrum is liftered with a half-Hann low-pass window at a chosen cutoff frequency, leaving a broadband spectrum dominated by structural resonances (Randall, Peeters, Antoni, & Manzato, 2012) and free of the dynamic components associated with the shafts, bearings, and gears. The liftered log-magnitude is then mean-subtracted, so that its shape is retained while extreme gains are avoided, and interpolated (piecewise cubic Hermite interpolating polynomial) to the target sampling grid where required. A minimum-phase reconstruction from the real cepstrum (Pei & Lin, 2006) finally recovers the complex transfer function $H(f)$. Minimum phase is assumed because a measured phase response was not available; it is a standard and accessible choice for mechanical systems, whose limitations are acknowledged in the literature (Lu, Borghesani, Randall, Smith, & Peng, 2023).

To reproduce the layered noise of real measurements, two components are added to the phenomenological signal before applying $H(f)$. First, a broadband Gaussian background is added in the time domain at a fixed fraction of the signal RMS, so that the subsequent multiplication by $H(f)$ imparts the same spectral coloring to the noise floor as to the deterministic components. Second, a structural mode bank is injected as a sum of discrete sinusoids at the dominant resonance frequencies of the machine, compensating for the narrowband peaks smoothed out by the cepstrum lifter.

The structural mode bank is derived directly from $H(f)$ rather than from an independent measurement, because the cepstrum liftering smooths sharp narrowband features out of the spectral envelope, partially attenuating the local resonances present in the real measurement. The magnitude of $H(f)$ in dB is first smoothed with a wide Gaussian kernel to obtain a slowly varying baseline. The residual, i.e., the difference between $H(f)$ in dB and this baseline, isolates the local maxima that stand out above the slow envelope. Peaks are extracted from the residual using a prominence threshold and a minimum frequency spacing constraint, and each becomes a sinusoid with a fixed base amplitude A and an initial phase. Because the bank is added before the transfer-function multiplication, its components are colored by $H(f)$ together with the deterministic signal and the background layer.

Finally, a Gaussian term is added at a controlled signal-to-noise ratio (SNR) to represent residual sensor and acquisition noise that is not shaped by the structural path, and the signal is mean-centered, as is standard in condition monitoring. The result is a simulated signal that more closely resembles those acquired on the actual gearbox.

3.1.2. Addition of Physical Behaviors

Two further extensions enrich the phenomenological signal. First, shaft-rotation harmonics are added as sinusoidal components at integer multiples of the shaft rotation frequency; these represent content the base model neglects and would enable the simulation of shaft-related faults such as unbalance and misalignment. Second, to capture manufacturing tolerances, small zero-mean perturbations are drawn once per gear, assigning an amplitude and a phase offset to each tooth. Because this pattern repeats with the shaft period, it produces shaft-related sidebands around the gear meshing harmonics, reproducing the tooth-to-tooth variability observed in real measurements.

3.1.3. Modification of the Local Fault Modeling

The local fault model is generalized to an arbitrary number of simultaneously faulty teeth, enabling multi-pit scenarios. Figure 3 compares one, two, and three pitted teeth at different positions across one revolution, showing the time-domain contribution that is added to the angle-generated signal. For visualization, a small $\sigma = 0.0005$ is used to narrow the impulse window; the operative value is set from the real asset. The extension applies consistently to both the time-domain impulsive contribution and the local fault modulation of the meshing signal. To capture cycle-to-cycle variability within a single recording, a per-revolution stochastic jitter is applied to the fault parameters: in Figure 3, although all signals share the same nominal fault amplitude, different seeds yield different impulse amplitudes. The same jitter realization is shared between the impulsive contribution and the local fault modulation, enforcing the physical correlation between them. Finally, the local change applied to the modulation function is decoupled into independent amplitude and phase depths, supporting fault scenarios such as the crack signature described by (McFadden, 1986), though not utilized in this paper, as pitting shares the same value for both types of modulation (D’Elia et al., 2017).

3.2. Dataset Generation

Data generation relies on a configuration file holding every parameter required to synthesize a signal, together with template classes that define a base configuration from which signals are derived. Diversity is introduced by assigning a range to each relevant parameter and drawing a value from it with a random generator, while storing the seed for reproducibility. The procedure is summarized in Figure 4.

3.3. Entropy-Measured DAAN (EM-DAAN)

The proposed EM-DAAN extends DAAN by transferring the entropy weighting mechanism previously used with DANN to the global and local discriminators, adding an information

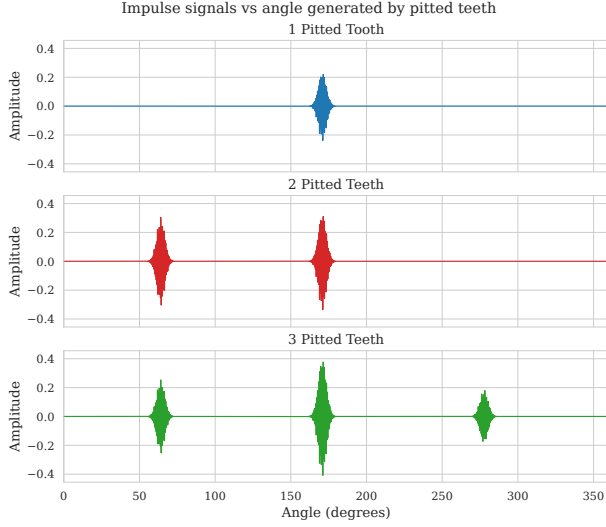


Figure 3. Comparison of three fault scenarios (1 vs. 2 vs. 3 pitted teeth) for a 13/50 gear pair with the pinion at 2400 rpm, over one revolution of the 50-tooth wheel.

maximization loss, and introducing stability improvements to the dynamic adversarial factor.

3.3.1. Entropy Weighting

Recent DA methods use prediction entropy as a dynamic per-sample weight, emphasizing samples differently according to the optimization goal. Two complementary routes have been proposed.

Let $H^i = -\sum_{c=1}^C p_c^i \log p_c^i$ denote the Shannon entropy of sample x_i , where p_c^i is the softmax probability of class c . In the first route, (Wang & Zhang, 2020) relax the adversarial alignment on samples the classifier already aligns well and focus it on the poorly aligned ones, weighting each sample in the DANN loss by its entropy:

$$\mathcal{L}_{adv} = -\frac{1}{N_S} \sum_{i=1}^{N_S} (1 + H_p^i) \mathcal{L}_d^i(\theta_f, \theta_d) - \frac{1}{N_T} \sum_{i=1}^{N_T} (1 + H_p^i) \mathcal{L}_d^i(\theta_f, \theta_d), \quad (12)$$

where $H_p^i = H^i/C$ is the entropy normalized by the number of classes, following (Wang & Zhang, 2020). A similar scheme was used by (Meng, Xuan, Xu, & Liu, 2022) with the raw entropy H^i , i.e., without the $1/C$ factor.

(Wang & Zhang, 2020) further add an entropy term on the target predictions to the total loss, to discourage misclassified target samples:

$$\mathcal{L}_{ent} = -\frac{1}{N_T} \sum_{i=1}^{N_T} p_i \log p_i. \quad (13)$$

A second route, the CDAN+E method of (Long, Cao, Wang, & Jordan, 2018), instead makes the discriminator focus on easy-to-transfer (low entropy) samples by assigning them larger weights, using $w^i = 1 + \exp(-H^i)$.

This work adopts the first route. The methods above weight a single discriminator, whereas DAAN has a global discriminator and one local discriminator per class, which calls for two distinct weighting rules. The global branch should focus on uncertain samples, giving more weight to high-entropy ones. The local branches should instead favor confident samples, since high-entropy samples carry unreliable pseudo-labels and weighting them strongly would corrupt the class-conditional alignment. Using the entropy normalized by $\log C$, $\bar{H}^i = H^i / \log C \in [0, 1]$, so that a confident sample has $\bar{H}^i = 0$ and a uniform prediction has $\bar{H}^i = 1$, the global, w_G^i , and local, w_L^i , weights are

$$w_G^i = 1 + \rho \bar{H}^i, \quad (14)$$

$$w_L^i = 1 + \rho(1 - \bar{H}^i). \quad (15)$$

For stability, the weights are normalized jointly across source and target for both the global and local branches by their mean, which preserves the relative differences between samples; these can be amplified through the hyperparameter ρ . Finally, the weights are detached from the computational graph, so they scale the adversarial loss without propagating gradients back into the classifier.

3.3.2. Information Maximization Loss

To further promote alignment, an additional term is added to the total loss, which already comprises the source classification and adversarial losses. Following (Wang & Zhang, 2020), an entropy-based loss is used; however, the plain target entropy loss of (Wang & Zhang, 2020) can cause model collapse (Niu et al., 2023), since entropy minimization makes the trivial solution of assigning all samples to one class the steepest descent direction. To prevent this, the Information Maximization (IM) loss (Gomes, Krause, & Perona, 2010; Liang, Hu, & Feng, 2020) is adopted, comprising two terms and a constant offset:

$$\begin{aligned} \mathcal{L}_{IM} &= \mathcal{L}_{ent} - \mathcal{L}_{div} \\ &= -\frac{1}{N_T} \sum_{i=1}^{N_T} \sum_{c=1}^C p_c^i \log p_c^i + \sum_{c=1}^C \bar{p}_c \log \bar{p}_c + \log C, \end{aligned} \quad (16)$$

where $\bar{p}_c = \frac{1}{N_T} \sum_{i=1}^{N_T} p_c^i$ is the marginal class probability over the target batch. The diversity term $\mathcal{L}_{div}$ (Liang et al., 2020) penalizes assigning all samples to one class: a collapsed prediction yields $\bar{p}_c \rightarrow 1$ for a single class and $\mathcal{L}_{div} \rightarrow 0$, raising the total loss. The $\log C$ offset shifts $\mathcal{L}_{IM}$ to non-negative values.

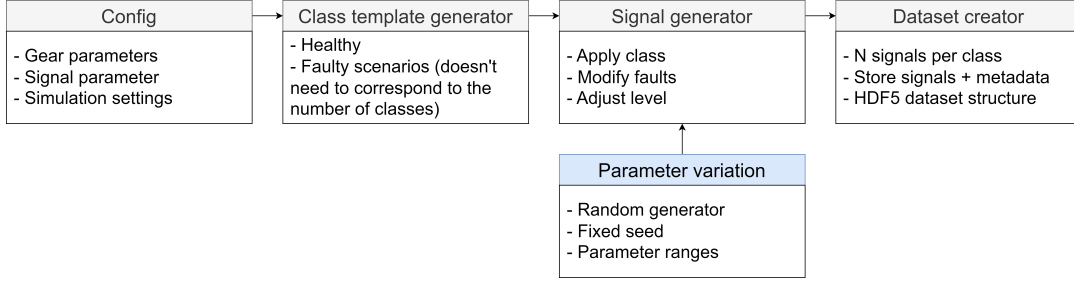


Figure 4. Workflow of the dataset generation.

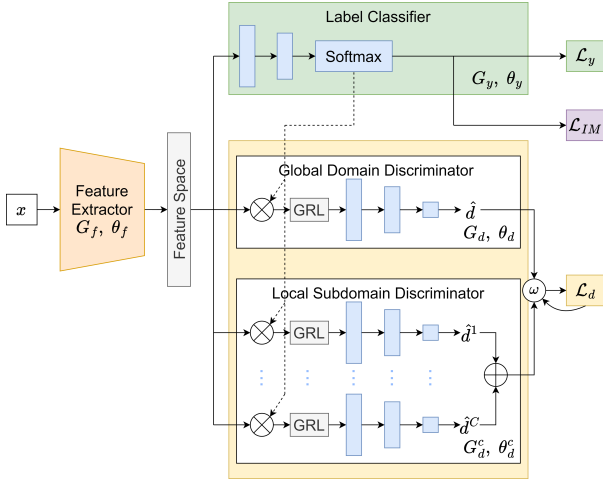


Figure 5. EM-DAAN diagram.

3.4. Learning Method

The total loss combines three components:

$$\mathcal{L} = \mathcal{L}_{class} + \lambda_{adv}(\tilde{\omega} \mathcal{L}_{G,adv}^H + (1 - \tilde{\omega}) \mathcal{L}_{L,adv}^H) + \lambda_{IM} \mathcal{L}_{IM}, \quad (17)$$

where $\mathcal{L}_{G,adv}^H$ is the weighted version of the global loss, $\mathcal{L}_{L,adv}^H$ is the weighted version of the local loss, and λ_{adv} and λ_{IM} are trade-off parameters.

Compared with the original DAAN, the dynamic adversarial factor is modified and denoted $\tilde{\omega}$. Exploiting the equivalence of the original ω to the proxy $\mathcal{A}$ -distance, the global and local $\mathcal{A}$ -distances of Eqs. 8 and 9 are evaluated from the domain classification error rather than from the raw discriminator losses, where the error of each discriminator is taken as $\min(\text{acc}, 1 - \text{acc})$ so that it remains in $[0, 1/2]$, as required by the proxy $\mathcal{A}$ -distance. The factor is then agnostic to magnitude shifts introduced by the entropy weighting and, like the proxy $\mathcal{A}$ -distance, is bounded, which the raw losses are not. This yields a more stable $\tilde{\omega}$ that does not saturate at the extremes of its range. The overall method is summarized in Figure 5.

4. METHODOLOGY APPLICATION

4.1. Experimental Dataset

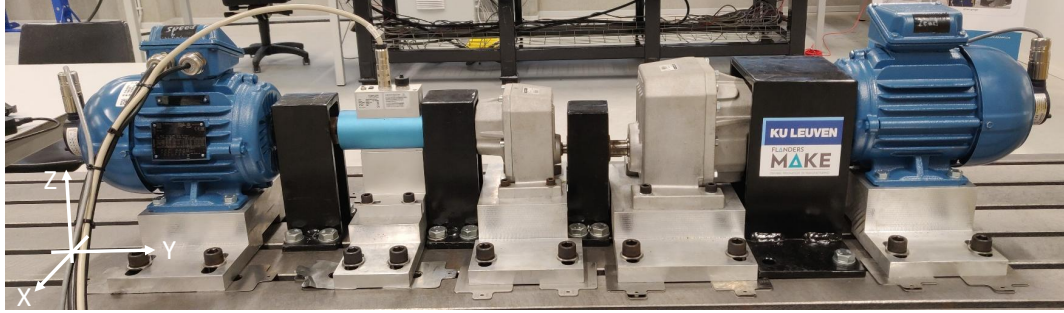
For this study, the experimental dataset first presented in (Zhu, Verwimp, Mauricio, & Gryllias, 2026) is used as the case study. The measurements were conducted on a back-to-back gearbox rig at KU Leuven, shown in Figure 6a. The rig comprises two electrical motors, A and B, a torque transducer, a test gearbox, and a drive gearbox. Motor A rotates the setup while Motor B applies the load, with an encoder and the torque transducer closing the control loops for speed and load, respectively. The test gearbox is single-stage, with 13 teeth on the input shaft and 50 on the output shaft; the drive gearbox is also single-stage, with 71 and 18 teeth. Vibration was acquired with piezoelectric triaxial accelerometers mounted on the test gearbox housing (the one used is indicated in Figure 6b), sampled at 102.4 kHz by a data acquisition system.

Faults were progressively seeded on the output gear of the test gearbox and the vibration response was acquired in healthy and each faulty condition. The faults (Figure 7), were artificially introduced on the teeth in increasing severity to emulate progressive pitting. The fault combinations, which define the classes, are listed in Table 1. The dataset, as presented in (Zhu et al., 2026), spans several combinations of speed and load imposed on the gearbox through the motors, with speed levels of 600, 1200, 1800, and 2400 rpm and load levels of 2.0, 2.6, and 3.3 Nm. Because the operating condition profiles were continuously monitored, the segment corresponding to the 2400 rpm and 3.3 Nm condition was selected as it offered the highest SNR. The segment boundaries were identified by low-pass filtering the torque transducer and motor encoder signals to remove high-frequency content, and comparing the filtered signals against the nominal values to obtain the corresponding timestamps. The resulting operating conditions are shown in Figure 8, with the selected one shaded in gray.

4.2. Method Preparation

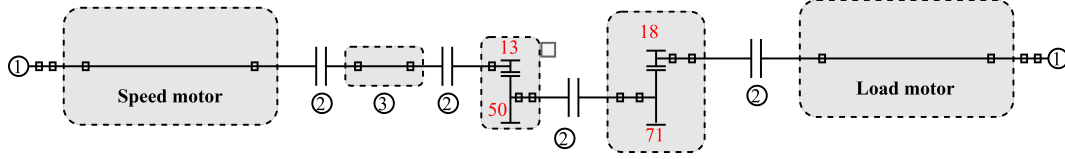
4.2.1. Network Configuration

The network architecture is described in Tables 2 and 3. In the feature extractor, group normalization (Wu & He,



Motor A Torque transducer Test Gearbox Drive Gearbox Motor B

(a) Gearbox setup.



□ Bearing ① Rotary encoder ② Flexible coupling ③ Torque sensor □ Accelerometer Z Tooth number

(b) Schematic of the gearbox setup.

Figure 6. Overview of the experimental setup at KU Leuven, adapted from (Zhu et al., 2026).

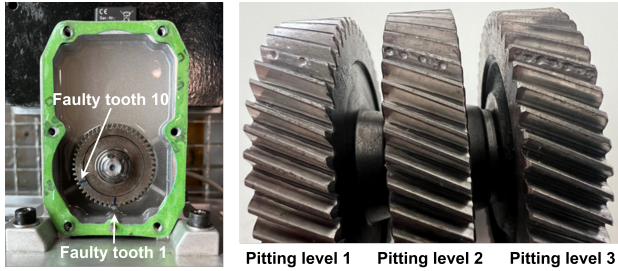


Figure 7. Different levels of damage used for the test runs, adapted from (Zhu et al., 2026).

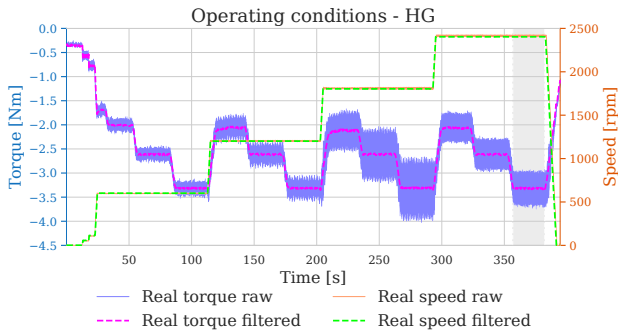


Figure 8. Different operating conditions tested on the gearbox rig for the healthy condition (HG is healthy gear).

Table 1. List of runs for the different configurations, which correspond to the selected classes.

Test	Location of pitted tooth	Tooth condition
1	—	Healthy
2	Tooth 1	Pitting level 1
3	Tooth 1	Pitting level 2
4	Tooth 1	Pitting level 3
5	Teeth 1 and 10	Pitting levels 3 and 2

2018) was preferred over the more common batch normalization, as it remained stable throughout the runs, and GeLU activations (Hendrycks & Gimpel, 2016) were chosen over ReLU for their smoother gradient. The classifier outputs 5 logits, and it is trained through cross-entropy loss, while the domain discriminator outputs a single logit and is trained with a binary cross-entropy loss. For these two networks, dropout (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014) was used after the activation layers to reduce overfitting.

All experiments share the same training configuration. The learning rate follows the schedule proposed in (Ganin et al., 2016):

$$\mu_p = \frac{\mu_0}{(1 + \alpha p)^\beta} \quad (18)$$

where p denotes the training progress, varying linearly from 0 to 1, and the remaining constants are set to $\mu_0 = 1 \times 10^{-4}$, $\alpha = 5$, and $\beta = 0.75$. To prevent the feature extractor from

Table 2. Network parameters of the feature extractor.

Layer	Type	Activation
1	1D CNN (1 → 16, $k = 15$, $s = 1$, $p = 7$, $d = 1$)	GeLU
2	1D CNN (16 → 32, $k = 9$, $s = 1$, $p = 4$, $d = 1$)	GeLU
3	1D CNN (32 → 48, $k = 7$, $s = 2$, $p = 3$, $d = 1$)	GeLU
4	1D CNN (48 → 64, $k = 5$, $s = 2$, $p = 2$, $d = 1$)	GeLU
5	1D CNN (64 → 64, $k = 3$, $s = 2$, $p = 1$, $d = 1$)	GeLU
6	1D CNN (64 → 64, $k = 3$, $s = 1$, $p = 2$, $d = 2$)	GeLU
7	Adaptive Average Pooling (output = 4)	-

Table 3. Network parameters of the classifier and discriminator.

Layer	Type	Activation
1	Linear (256 → 64)	GeLU
2	Dropout (0.5)	—
3	Linear (64 → 32)	GeLU
4	Dropout (0.5)	—
5	Linear (32 → 5; Classifier) (32 → 1; Discriminator)	-

overfitting to source-domain features before sufficient adversarial pressure builds up, its learning rate is scaled by a factor of 0.1, while the adversarial layers are assigned a multiplier of 0.5.

The GRL parameter follows the schedule from the same reference:

$$\lambda_p = \frac{2}{1 + \exp(-\gamma p)} - 1 \quad (19)$$

with $\gamma = 5$. For the entropy term, ρ was set to 3. All loss trade-off coefficients were fixed at 1 to ensure consistency across methods and avoid method-specific tuning. This is a reference setting rather than an optimal one, and it is expected that performance could be improved by tuning these parameters for each method.

The optimizer was Adam (Kingma & Ba, 2015) with a weight decay of 1×10^{-5} , and all input signals were normalized using the mean and standard deviation of the training set. Training ran for 300 epochs with a batch size of 64 and a 75/25 train-validation split; the target training windows are cycled so that every minibatch contains equal numbers of source and target samples.

All hyperparameter choices were made empirically rather than through a formal ablation study.

4.2.2. Evaluation Metrics

Based on the definitions of true positive (TP), true negative (TN), false positive (FP), and false negative (FN), the accuracy, F1-score, and Matthews correlation coefficient (MCC), as in (Peng et al., 2025), are used to compare performance across methods:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \quad (20)$$

$$\text{F1-score} = \frac{2 \text{TP}}{2 \text{TP} + \text{FP} + \text{FN}}, \quad (21)$$

$$\text{MCC} = \frac{(\text{TP} \cdot \text{TN}) - (\text{FP} \cdot \text{FN})}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}}. \quad (22)$$

The definitions above are given for the binary case; for the five-class problem, the F1-score is macro-averaged over classes and the MCC is computed in its multiclass form.

Model selection was based on the F1-score evaluated on the source validation set. Alternative criteria such as the transfer score of (Yang, Qian, Xu, Wang, & Xie, 2024) were considered but could not be used here: the comparison includes a source-only baseline, for which no target information is available, so a criterion common to all methods is required. However, in UDA, no metric evaluated on the source domain can guarantee good performance on the target domain.

Furthermore, t-SNE (van der Maaten & Hinton, 2008) is used to project the high-dimensional feature space onto a two-dimensional map, enabling a qualitative inspection of how source and target features cluster per class. Its main limitation is that a two-dimensional embedding cannot preserve all the structure of the feature space: distances and cluster sizes are not directly meaningful, the result depends on initialization, and some class-discriminative information may not be visible in the plot.

Lastly, the proxy $\mathcal{A}$ -distance (PAD), an estimate of transferability (Ben-David et al., 2006), is computed by averaging the linear-SVM domain classification error over $k = 5$ stratified random splits, with the SVM regularization parameter C_{SVM} selected by cross-validation over a log-spaced grid on the training half of each split. The PAD is calculated taking into account that the source and target datasets are imbalanced, by subsampling the larger domain to match the size of the smaller one, trying to mitigate one of the limitations shown in (Kunte, Cornelis, Colangeli, & Gryllias, 2023).

Together with the statistical metrics, these tools help to assess whether the learned features are domain-invariant, i.e., whether the domain adaptation was successful.

As a UDA method, the approach is transductive: the unlabeled target data are available during adaptation, and performance is reported on that same target distribution, as is standard in this setting. Generalization to independent fault re-

alizations is therefore not assessed here and is left to future work.

4.2.3. Data Comparison

Because experimental data were scarce, with each measurement, one per class, providing only about 25 s of signal, the target recordings were split before segmentation to avoid data leakage. The first 75 % of each recording was assigned to training, and the last 25 % to testing. However, to reduce correlation, a guard of 51 200 samples (0.5 s) was discarded on each side of the split, removing 1 s in total between training and test regions. 1 s segments with 90 % overlap were then extracted within each region separately. This yielded 177 segments per class for training, and 49 for testing. Each 1 s segment was downsampled from 102.4 kHz to 25.6 kHz, giving 25 600 points per segment. Lastly, due to the nature of the dataset, Gaussian noise was added to the target training set to expose the adaptation to a noisier target distribution and to discourage memorization of the small target set, giving a 20 dB SNR.

The synthetic dataset was built from three templates (healthy, one pitted tooth, and two pitted teeth), from which the five severity classes of Figure 7 were derived. Each class comprises 500 one-second signals at 25.6 kHz, generated in 349 s in total. Tooth variability, load modulation, and shaft harmonics were enabled for every class. For the faulty classes, progressively higher meshing harmonics were used together with the local fault contribution; the amplitude and phase modulation depths, although decoupled in the model, were set equal for this study, recovering the coupled behavior of the original model of (D’Elia et al., 2017). The distributed amplitude and phase modulations were also enabled to introduce sidebands around the meshing harmonics that accompany the onset of pitting. The carrier frequency of the local-fault impulses was approximated by inspecting the spectra of all runs and locating candidate resonance lines, yielding 6810 Hz, with σ about 0.005 s. Were the gearbox available for an offline test, modal analysis (Jiang et al., 2024) could instead identify the structural bands, of either the gear or the housing, that govern impact propagation. The transfer function was estimated from the full healthy signal at the original sampling rate using 5 s Welch segments, and is shown in Figure 9.

These choices allow a direct comparison between the synthetic and real datasets. Figure 10a overlays the time-domain signals for every class, and Figure 10b compares their spectra. In both figures, a synthetic signal of each class is shown against the first 1 s portion of the real recording.

Despite some residual discrepancies, most visibly an elevated meshing-frequency amplitude in the synthetic signal, the combination of the estimated transfer function, the Gaussian background, and the structural bank reproduces the overall spectral envelope of the real signal, including the struc-

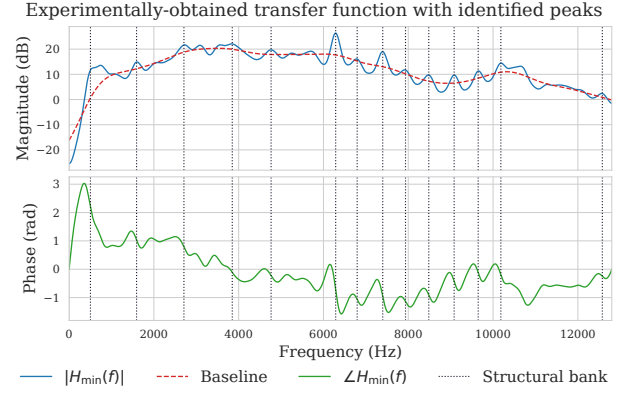


Figure 9. Experimentally obtained transfer function with the identified structural-bank components, based on the healthy run.

Table 4. Summary of the experimental results (accuracy, F1-score, and MCC) for the proposed method, compared with SOTA methods, shown as mean $\pm$ standard deviation for 5 runs. The best values are in bold.

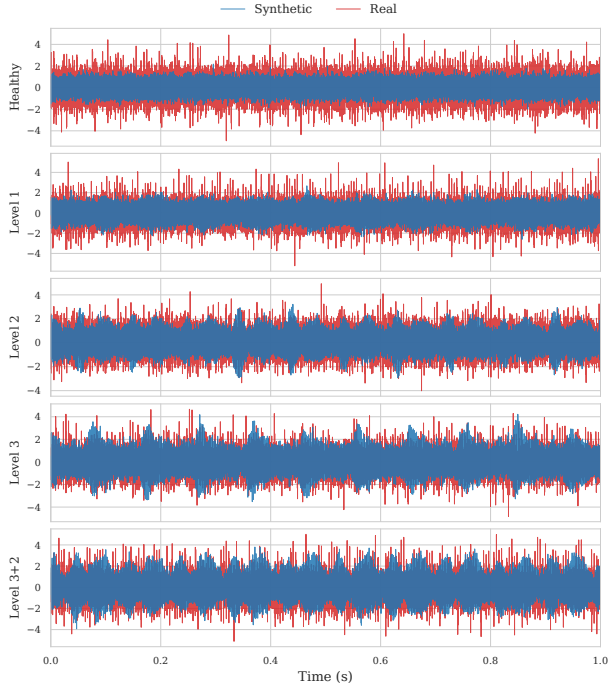
	Baseline	DANN	DAAN	Proposed method
Accuracy (%)	20.00 $\pm$ 0.00	51.84 $\pm$ 15.67	29.80 $\pm$ 6.99	92.49 $\pm$ 15.02
F1-score (%)	9.10 $\pm$ 3.00	46.33 $\pm$ 17.80	19.25 $\pm$ 5.25	92.46 $\pm$ 15.08
MCC (%)	0.00 $\pm$ 0.00	41.61 $\pm$ 19.60	14.40 $\pm$ 10.25	90.62 $\pm$ 18.76

tural coloring. One family of features not captured by the phenomenological model is the set of sidebands in the mid-frequency range; because these appear in all classes of the real signal, they should not carry class-discriminative information, and the UDA procedure is expected to filter them out as a class-invariant nuisance. A second source of discrepancy is the total signal energy: the phenomenological model concentrates its contribution at the selected meshing harmonics, whereas the real signal carries additional broadband content from sources not represented in the model. This effect is most pronounced for the healthy condition, where no local fault impulses are present to contribute broadband energy to the simulated spectrum.

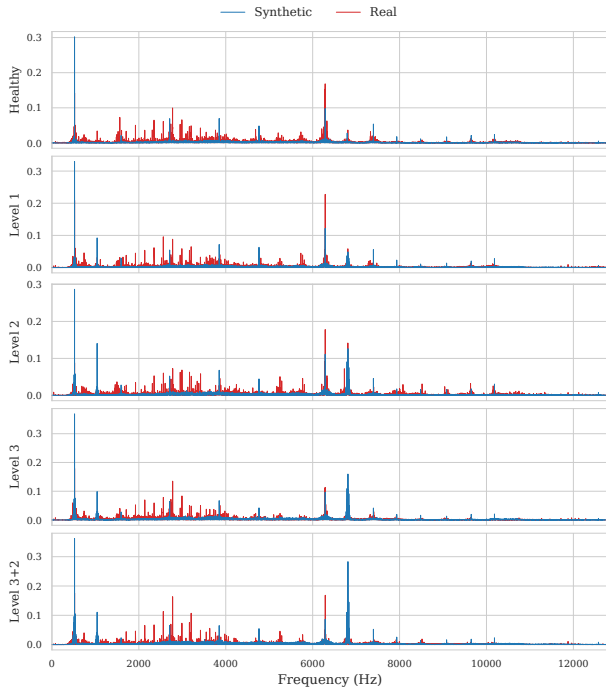
4.3. Results and Discussion

Five runs were performed per method, varying the seed to change the weight initialization, the source train/validation data splits, and the remaining stochastic processes of the algorithm. EM-DAAN was compared against a source-only trained classifier, DANN (Ganin et al., 2016), and DAAN (Yu et al., 2019). All networks were trained in PyTorch 2.6.0 on an RTX A1000 mobile GPU, requiring on average 5.38 s per epoch (training and validation) for the baseline, 9.50 s for DANN, 9.39 s for DAAN, and 10.07 s for EM-DAAN.

The averaged confusion matrices are shown in Figure 11, and a summary of the metrics introduced in Section 4.2.2 is given in Table 4. EM-DAAN outperforms both SOTA ap-



(a) Time-domain.



(b) Frequency-domain.

Figure 10. Comparison between synthetically generated and real signals.

proaches and the source-only baseline by a wide margin. Although a formal component-wise ablation is left for future work, the chosen comparison already provides a coarse form of it, since DAAN and DANN correspond to the proposed architecture with the introduced mechanisms progressively disabled. For this setup, its entropy-based mechanisms appear to largely close the residual gap (Figure 11d), whereas DANN and DAAN correctly classify some of the classes, and the baseline completely collapses (Figures 11a–11c). Their poor performance can possibly be attributed to the standard trade-off values of 1, kept uniform to avoid method-specific tuning (Section 4.2.1), as their optimum might lie at other values. The relatively high standard deviation stems from one under-performing seed in the proposed approach, and one over-performing seed in DANN, while the baseline and DAAN remain rather stable. Furthermore, it is important to highlight the struggle of DAAN, which was unexpected, but could also be related to the chosen metric for best epoch. Lastly, these results were obtained in a limited-data regime, where 90% overlap and a single recording per class were used for the target dataset. Therefore, with a larger and more diverse data, the SOTA methods that lack entropy-based mechanisms would most likely improve, although this remains to be demonstrated.

A t-SNE visualization of the classifier logits is also produced for a single representative seed, common to all four networks. For the proposed approach (Figure 12d), the faulty samples form well-separated, class-pure structures, consistent with the strong performance reported in all the metrics. The remaining three show substantial inter-class overlap. DANN and DAAN (Figures 12b and 12c) clearly confuse the pitting classes, as already noted in the confusion matrices, and the baseline (Figure 12a) separates them even less.

Finally, the average proxy $\mathcal{A}$ -distance was computed for each method, with results being reported in Table 5. An apparent tension emerges between the classification metrics and the PAD: the proposed method achieves the highest metrics by a wide margin but the worst PAD of the three UDA methods, while DANN and DAAN show the opposite behavior. This dissociation is consistent with the entropy and IM mechanisms, which drive confident, class-correct predictions on the target without explicitly aligning the source and target feature distributions; the features therefore remain partially domain-separable, hence the higher PAD, even though the classifier resolves the target classes correctly. PAD, which measures feature-distribution overlap, is thus an unreliable proxy for transfer quality when the adaptation acts on the predictions rather than on the features. This compounds the PAD limitations noted in Section 4.2.2.

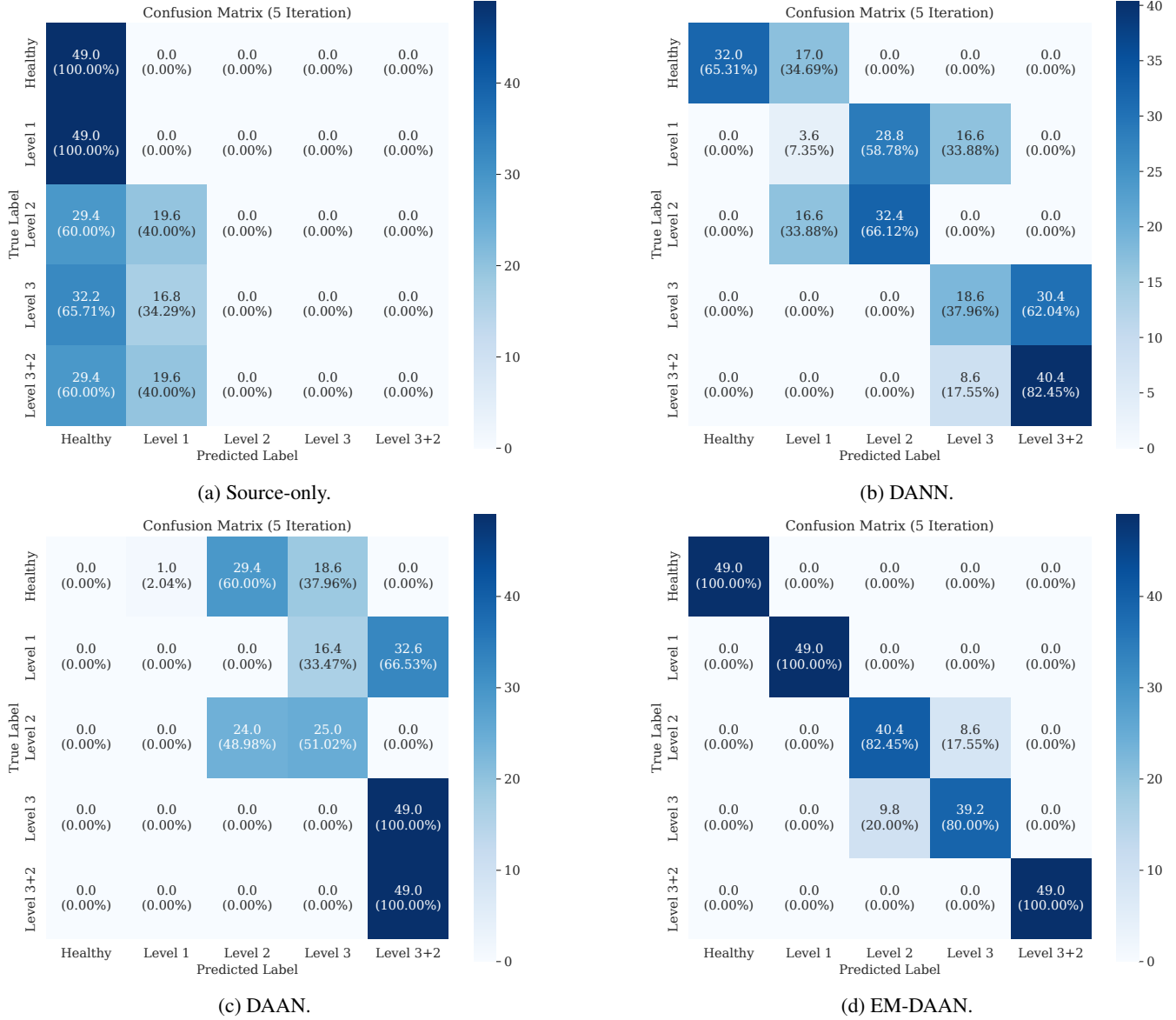


Figure 11. Confusion matrix results for all the applied methods.

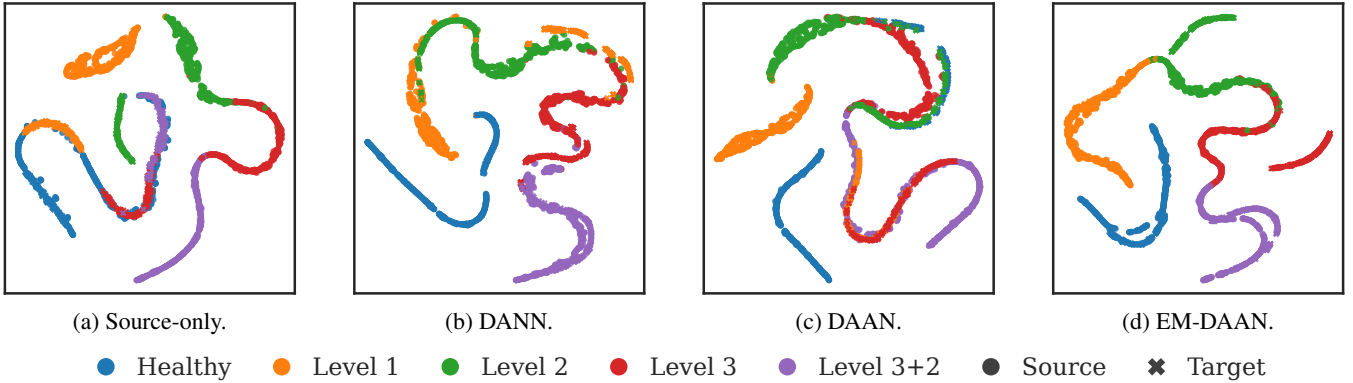


Figure 12. Visualization of the logit distributions via t-SNE for all the applied methods, with source samples shown as circles and target samples as crosses, and the different classes being color-coded.

Table 5. Proxy $\mathcal{A}$ -distance (PAD) for the proposed method, compared with SOTA methods, shown as mean $\pm$ standard deviation for 5 runs. The best values (lower is better) are in bold.

	Baseline	DANN	DAAN	Proposed method
PAD	1.79 $\pm$ 0.17	0.97 $\pm$ 0.44	1.30 $\pm$ 0.30	1.47 $\pm$ 0.21

5. CONCLUSION

In this work, an improved phenomenological gearbox simulator, extended to support multi-tooth faults, shaft-rotation harmonics, and a measurement-chain model anchored by an estimated transfer function, was combined with a deep UDA procedure that aligns simulated and real distributions and largely closes the sim-to-real gap on the target rig. The entropy-based mechanism, in particular, demonstrates strong capability for guiding the model toward confident and accurate predictions on the tested dataset, although the limited amount of target data is acknowledged as a constraint on the generality of the conclusions. The performance was evaluated against a source-only baseline and two SOTA methods, DANN and DAAN, with the proposed method outperforming both by a wide margin, showing a high standard deviation due to an outlier run. The t-SNE visualization of the classifier logits shows that the proposed method produces well-separated clusters for each class, while the other methods show substantial inter-class overlap. The proxy $\mathcal{A}$ -distance, a measure of feature-distribution overlap, is shown to be an unreliable proxy for transfer quality when the adaptation acts on the predictions rather than on the features.

Future work will focus on evaluating the same approach in an extended version of the applied target dataset, covering as well the additional speed and load combinations available in the test campaign, allowing more detailed conclusions to be drawn on the generality of the approach through a significance analysis study. A formal component-wise ablation will also be performed to isolate the individual contribution of each introduced mechanism, complementing the coarse progression from DANN to DAAN to EM-DAAN reported here, together with a sensitivity study of the introduced hyperparameters and a convergence analysis, both beyond the scope of the present contribution. Regarding robustness, the noise injected into the target training data hardens the training protocol but does not constitute a noise-robustness study, so a systematic sweep over test-time SNR levels is planned. Finally, the framework will be extended to independent test rigs and to real-to-real adaptation, as well as to other fault types such as bearing faults, while the carrier frequency will be identified through modal analysis rather than spectral inspection.

ACKNOWLEDGMENT

The authors acknowledge the financial support of the Clean Energy Transition Partnership, funded from the European Union's Horizon Europe research and innovation program under grant agreement no. GA 101069750, and VLAIO, via the Hybrid Wind project. Additionally, the authors would like to gratefully acknowledge the support of the Fonds Wetenschappelijk Onderzoek Vlaanderen (FWO) under the research grant no. 1SE0125N.

REFERENCES

- Antoni, J. (2009, May). Cyclostationarity by examples. *Mechanical Systems and Signal Processing*, 23(4), 987–1036. doi: 10.1016/j.ymssp.2008.10.010
- Bachar, L., Matania, O., Cohen, R., Klein, R., Lipsett, M. G., & Bortman, J. (2023, December). A novel hybrid physical AI-based strategy for fault severity estimation in spur gears with zero-shot learning. *Mechanical Systems and Signal Processing*, 204, 110748. doi: 10.1016/j.ymssp.2023.110748
- Ben-David, S., Blitzer, J., Crammer, K., & Pereira, F. (2006). Analysis of Representations for Domain Adaptation. In *Advances in neural information processing systems* (Vol. 19, pp. 137–144). MIT Press.
- D'Elia, G., Cocconcelli, M., Rubini, R., & Dalpiaz, G. (2017, May). Step-by-step algorithm for the simulation of a faulted gearbox. In *Proceedings of Surveillance 9 2017*.
- Fang, Y., Chen, C., Zhang, W., Wu, J., Zhang, Z., & Xie, S. (2024, November). Prototype learning for adversarial domain adaptation. *Pattern Recognition*, 155, 110653. doi: 10.1016/j.patcog.2024.110653
- Fang, Y., Yap, P.-T., Lin, W., Zhu, H., & Liu, M. (2024, June). Source-free unsupervised domain adaptation: A survey. *Neural Networks*, 174, 106230. doi: 10.1016/j.neunet.2024.106230
- Ganin, Y., & Lempitsky, V. (2015, June). Unsupervised Domain Adaptation by Backpropagation. In *Proceedings of the 32nd international conference on machine learning* (Vol. 37, pp. 1180–1189). PMLR.
- Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., ... Lempitsky, V. (2016). Domain-Adversarial Training of Neural Networks. *Journal of Machine Learning Research*, 17(59), 1–35.
- Gomes, R., Krause, A., & Perona, P. (2010). Discriminative Clustering by Regularized Information Maximization. In *Advances in neural information processing systems* (Vol. 23, pp. 775–783). Curran Associates, Inc.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... Bengio, Y. (2014). Generative Adversarial Nets. In *Advances in neural information processing systems* (Vol. 27, pp. 2672–2680). Curran

- Associates, Inc.
- Gryllias, K. C., & Antoniadis, I. A. (2012, March). A Support Vector Machine approach based on physical model training for rolling element bearing fault detection in industrial environments. *Engineering Applications of Artificial Intelligence*, 25(2), 326–344. doi: 10.1016/j.engappai.2011.09.010
- Hendrycks, D., & Gimpel, K. (2016). *Gaussian Error Linear Units (GELUs)*. (arXiv:1606.08415 [cs.LG]) doi: 10.48550/arXiv.1606.08415
- Hosseini, S. A., Ooijevaar, T., & Gryllias, K. (2024, June). Simulation-based remaining useful life prediction of rolling element bearings under varying operating conditions. In *Proceedings of the European Conference of the PHM Society 2024*. doi: 10.36001/phme.2024.v8i1.4044
- Jiang, F., Lin, W., Wu, Z., Zhang, S., Chen, Z., & Li, W. (2024, August). Fault diagnosis of gearbox driven by vibration response mechanism and enhanced unsupervised domain adaptation. *Advanced Engineering Informatics*, 61, 102460. doi: 10.1016/j.aei.2024.102460
- Kingma, D. P., & Ba, J. (2015, May). Adam: A Method for Stochastic Optimization. In *Proceedings of the 3rd international conference on learning representations (iclr 2015)*.
- Koutsoupakis, J., Seventekidis, P., & Giagopoulos, D. (2023, May). Machine learning based condition monitoring for gear transmission systems using data generated by optimal multibody dynamics models. *Mechanical Systems and Signal Processing*, 190, 110130. doi: 10.1016/j.ymssp.2023.110130
- Kunte, D., Cornelis, B., Colangeli, C., & Gryllias, K. (2023, July). Investigating the usage of Proxy-A-Distance as a measure of dataset shift detection and quantification in an automotive booming noise classification setting. In *Proceedings of Surveillance, Vibrations, Shock and Noise*. Toulouse, France.
- Lei, Y., Liu, Z., Wang, D., Yang, X., Liu, H., & Lin, J. (2018, June). A probability distribution model of tooth pits for evaluating time-varying mesh stiffness of pitting gears. *Mechanical Systems and Signal Processing*, 106, 355–366. doi: 10.1016/j.ymssp.2018.01.005
- Li, W., Huang, R., Li, J., Liao, Y., Chen, Z., He, G., ... Gryllias, K. (2022, March). A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: Theories, applications and challenges. *Mechanical Systems and Signal Processing*, 167, 108487. doi: 10.1016/j.ymssp.2021.108487
- Li, Y., Ding, K., He, G., & Lin, H. (2016, December). Vibration mechanisms of spur gear pair in healthy and fault states. *Mechanical Systems and Signal Processing*, 81, 183–201. doi: 10.1016/j.ymssp.2016.03.014
- Liang, J., Hu, D., & Feng, J. (2020, July). Do We Really Need to Access the Source Data? Source Hypothesis Transfer for Unsupervised Domain Adaptation. In *Proceedings of the 37th international conference on machine learning* (Vol. 119, pp. 6083–6094). PMLR.
- Liu, D., Cui, L., Wang, G., & Cheng, W. (2025, March). Interpretable domain adaptation transformer: A transfer learning method for fault diagnosis of rotating machinery. *Structural Health Monitoring*, 24(2), 1187–1200. doi: 10.1177/14759217241249656
- Long, M., Cao, Y., Wang, J., & Jordan, M. I. (2015). Learning Transferable Features with Deep Adaptation Networks. In *Proceedings of the 32nd international conference on machine learning* (Vol. 37, pp. 97–105). Lille, France: PMLR.
- Long, M., Cao, Z., Wang, J., & Jordan, M. I. (2018). Conditional Adversarial Domain Adaptation. In *Advances in neural information processing systems* (Vol. 31, pp. 1640–1650). Curran Associates, Inc.
- Lou, Y., Kumar, A., & Xiang, J. (2022). Machinery Fault Diagnosis Based on Domain Adaptation to Bridge the Gap Between Simulation and Measured Signals. *IEEE Transactions on Instrumentation and Measurement*, 71, 1–9. doi: 10.1109/TIM.2022.3180416
- Lu, R., Borghesani, P., Randall, R. B., Smith, W. A., & Peng, Z. (2023, February). Removal of transfer function effects from gear vibration signals under constant and variable speed conditions. *Mechanical Systems and Signal Processing*, 184, 109714. doi: 10.1016/j.ymssp.2022.109714
- Matania, O., Klein, R., & Bortman, J. (2022, March). Novel approaches for the estimation of the spectrum background for stationary and quasi-stationary signals. *Mechanical Systems and Signal Processing*, 167, 108503. doi: 10.1016/j.ymssp.2021.108503
- McFadden, P. D. (1986, April). Detecting Fatigue Cracks in Gears by Amplitude and Phase Demodulation of the Meshing Vibration. *Journal of Vibration and Acoustics*, 108(2), 165–170. doi: 10.1115/1.3269317
- Meng, Y., Xuan, J., Xu, L., & Liu, J. (2022, March). Dynamic Reweighted Domain Adaption for Cross-Domain Bearing Fault Diagnosis. *Machines*, 10(4), 245. doi: 10.3390/machines10040245
- Niu, S., Wu, J., Zhang, Y., Wen, Z., Chen, Y., Zhao, P., & Tan, M. (2023). Towards stable Test-Time adaptation in dynamic wild world. In *Proceedings of the 11th international conference on learning representations (iclr 2023)*.
- Ozyurt, Y., Feuerriegel, S., & Zhang, C. (2023). Contrastive Learning for Unsupervised Domain Adaptation of Time Series. In *Proceedings of the 11th international conference on learning representations (iclr 2023)*.
- Pan, S. J., & Yang, Q. (2010, October). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345–1359. doi: 10.1109/TKDE

- .2009.191
- Pei, S.-C., & Lin, H.-S. (2006, October). Minimum-Phase FIR Filter Design Using Real Cepstrum. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 53(10), 1113–1117. doi: 10.1109/TCSII.2006.882193
- Peng, D., Yazdaniyanasr, M., Mauricio, A., Verwimp, T., Desmet, W., & Gryllias, K. (2025, April). Physics-driven cross domain digital twin framework for bearing fault diagnosis in non-stationary conditions. *Mechanical Systems and Signal Processing*, 228. doi: 10.1016/j.ymssp.2024.112266
- Randall, R. B. (1982, April). A New Method of Modeling Gear Faults. *Journal of Mechanical Design*, 104(2), 259–267. doi: 10.1115/1.3256334
- Randall, R. B., Peeters, B., Antoni, J., & Manzato, S. (2012, September). New cepstral methods of signal pre-processing for operational modal analysis. In *Proceedings of the International Conference on Noise and Vibration Engineering (ISMA2012)* (pp. 755–764). Leuven, Belgium.
- Santelo, T. N., de Oliveira, C. M. R., Maciel, C. D., & Monteiro, J. R. B. d. A. (2022, April). Wind Turbine Failures Review and Trends. *Journal of Control, Automation and Electrical Systems*, 33(2), 505–521. doi: 10.1007/s40313-021-00789-8
- Sobie, C., Freitas, C., & Nicolai, M. (2018, January). Simulation-driven machine learning: Bearing fault classification. *Mechanical Systems and Signal Processing*, 99, 403–419. doi: 10.1016/j.ymssp.2017.06.025
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15(56), 1929–1958.
- Sun, B., & Saenko, K. (2016). Deep CORAL: Correlation Alignment for Deep Domain Adaptation. In *Computer vision – European Conference on Computer Vision 2016 workshops* (Vol. 9915, pp. 443–450). Cham: Springer. doi: 10.1007/978-3-319-49409-8_35
- van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9, 2579–2605.
- Wang, S., & Zhang, L. (2020, July). Self-adaptive Reweighted Adversarial Domain Adaptation. In *Proceedings of the twenty-ninth international joint conference on artificial intelligence* (pp. 2547–2553). IJCAI. doi: 10.24963/ijcai.2020/440
- Wen, L., Gao, L., & Li, X. (2019, January). A new deep transfer learning based on sparse auto-encoder for fault diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 49(1), 136–144. doi: 10.1109/TSMC.2017.2754287
- Wu, Y., & He, K. (2018, September). Group Normalization. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 3–19). doi: 10.1007/978-3-030-01261-8_1
- Xu, Z., Tang, G., & Pang, B. (2026, January). Simulation driven speed aware conditional diffusion model and its application to domain generalization fault diagnosis of bearings. *Measurement*, 258, 119418. doi: 10.1016/j.measurement.2025.119418
- Yang, J., Qian, H., Xu, Y., Wang, K., & Xie, L. (2024, January). Can we evaluate domain adaptation models without target-domain labels? In *Proceedings of the 12th international conference on learning representations (iclr 2024)*. (arXiv: 2305.18712)
- Yu, C., Wang, J., Chen, Y., & Huang, M. (2019, November). Transfer Learning with Dynamic Adversarial Adaptation Network. In *2019 IEEE International Conference on Data Mining (ICDM)* (pp. 778–786). Beijing, China: IEEE. doi: 10.1109/ICDM.2019.00088
- Zhu, R., Verwimp, T., Mauricio, A., & Gryllias, K. (2026, January). An Improved Instantaneous Wavelet Bispectrum Feature for gear fault diagnostics using a Wireless MEMS sensor. *Mechanical Systems and Signal Processing*, 244, 113729. doi: 10.1016/j.ymssp.2025.113729

BIOGRAPHIES

Henrique Duarte Vieira de Sousa received his Bachelor and Master of Mechanical Engineering from University of Porto, Portugal in 2022 and 2024, respectively. In 2024, he joined the LMSD (Mecha(tro)nic System Dynamics) Division of the Department of Mechanical Engineering at KU Leuven as a Ph.D. researcher. His research interests lie in the areas of condition monitoring, transfer learning, and digital twins.

Rui Zhu received his M.Sc. in Mechanical Engineering from Southeast University, China, in 2019. Since 2020, he has been a PhD researcher in the LMSD (Mecha(tro)nic System Dynamics) Division of the Department of Mechanical Engineering at KU Leuven. His research interests focus on condition monitoring of rotating machinery with advanced signal processing methodologies.

Toby Verwimp received his M.Sc. in mechanical engineering from KU Leuven, Leuven, Belgium in 2021. He joined the LMSD (Mecha(tro)nic System Dynamics) Division in the Department of Mechanical Engineering at KU Leuven as a PhD researcher in 2021. His research interests include condition monitoring of rotating machinery using, among others, rotary encoders and cameras.

Alexandre Maurício received his M.Sc. degree in Mechanical Engineering from the Faculty of Engineering of the University of Porto (FEUP) in 2017. Then he obtained his Ph.D. degree at the LMSD (Mecha(tro)nic System Dynamics) Division of the Department of Mechanical Engineering at KU Leuven. His main research interests focus on the diagnostics of rotating machinery operating under time varying conditions.

Hao Wen received his Bachelor of Science degree from the China University of Geosciences, China and his Master of Science degree in Mechanical Engineering from the University of Manchester, UK. He joined the LMSD (Mecha(tro)nic

System Dynamics) Division in the Department of Mechanical Engineering at KU Leuven as a PhD researcher in 2022. His research interests lie in the areas of fault detection utilizing unsupervised learning and generative modeling.

Seyed Ali Hosseinli received his B.Sc. and M.Sc. degrees in Railway Engineering from the Iranian University of Science and Technology (IUST) and in Mechanical Engineering from the Sharif University of Technology (SUT), respectively. He is currently a Ph.D. researcher at LMSD (Mecha(tro)nic System Dynamics) Division in the Department of Mechanical Engineering at KU Leuven. His research interests focus on machine learning techniques in prognostics and fault diagnosis of rotary equipment.

Konstantinos Gryllias received the Diploma and Ph.D. degrees in Mechanical Engineering from the National Technical University of Athens, Greece, in 2004 and 2010, respectively. He is currently a Professor of Vibro-Acoustics of Machines and Transportation Systems with the Department of Mechanical Engineering, KU Leuven, Belgium. He also serves as the Manager of the University Core Lab Flanders Make@KU Leuven Motion Products, Belgium. His research interests include condition monitoring, signal processing, prognostics, and health management of mechanical and mechatronic systems.