Context-aware machine learning for estimating the remaining useful life of bearings under varying speed operating conditions

Seyed Ali Hosseinli[1, 2, 3], Ted Ooijevaar[4] and Konstantinos Gryllias[1, 2, 3]

[1] *Department of Mechanical Engineering, KU Leuven*

[2] *Flanders Make @ KU Leuven*

Celestijnenlaan 300, BOX 2420, 3001 Leuven, Belgium

[3] *Leuven.AI - KU Leuven Institute for AI, B-3000 Leuven, Belgium*

[4] *Flanders Make vzw, CoreLab MotionS, 3001 Leuven, Belgium*

konstantinos.gryllias@kuleuven.be

ABSTRACT

Remaining useful life estimation is a crucial and complicated task in predictive maintenance in order to reduce downtime and avoid catastrophic breakdowns in industrial plants. Thanks to the recent advances in our machine learning era, deep learning models can effectively deal with modeling complex phenomena such as the bearing degradation process, specifically under varying operating conditions. However, obtaining large labeled datasets for training the data-dependent deep learning models is challenging and expensive. To overcome this limitation, a phenomenological model has been used in this study as an effective approach to creating synthetic run-to-failure datasets under varying operating conditions. The suggested methodology is able to adjust synthetic run-to-failure datasets to the different periodic speed profiles, including the speed ranges that pass the resonance frequency of the structure. A Context-aware Domain Adversarial Neural Network is proposed to remove the domain shift between the simulated signals and the real ones and to enable the deep learning model to understand the varying speed operating conditions and the sequential order of the measurements. The simulated signals are used as the source domain and a limited number of the real signals are used as the unlabeled samples for the domain adaptation task. Speed and time information are encoded as one-dimensional vectors and fed to the Domain Adversarial Neural Network, leading to an improvement in the model performance and its generalization ability. A dataset captured in a bearing test rig is adopted to verify the proposed method. Results show that context awareness can result in better performance and also more robust predictions against major speed changes in varying speed scenarios compared to the non-context-aware models.

1. INTRODUCTION

Rolling element bearings are one of the industrial components that are mostly used in rotary equipment, and they are prone to failures, which can lead to machine breakdown (Buzzoni et al., 2020; Randall & Antoni, 2011).

Remaining useful life (RUL) estimation plays an important role in reducing the risk of long downtimes and catastrophic breakdowns. It can be determined by the remaining time left before a health indicator crosses a certain threshold such that the component is no longer able to operate in the desired way (Lei et al., 2018). There are many approaches to achieving this aim. However, Artificial Intelligence (AI), specifically Deep Learning (DL), recently has shown brilliant results due to its ability to model processes with high complexity (Lei et al., 2018).

In order to have an effective DL model, the availability of a large amount of labeled data is crucially important. However, collecting such a dataset is usually laborious and of high cost (Chen et al., 2018; C. Liu & Gryllias, 2022). Utilizing simulated datasets for training the DL models has emerged as a viable approach to alleviate the impact of limited availability of real labeled datasets, thus enhancing the overall performance of the models.

Gryllias and Antoniadis (2012) generated artificial signals for different types of faults in bearings and classified the real samples using a Support Vector Machine (SVM) model. Chen et al. (2018) proposed a DL model to learn the fault patterns from artificial bearing datasets and applied the model to real bearing datasets. Farhat et al. (2021) developed a

model of a rotor-ball bearing system in order to build a digital twin (DT) to generate simulated vibration signals for different fault severity in bearings and classified them by machine learning algorithms.

However, the simulated datasets cannot or at least cannot yet cover all the aspects of the faulty bearings or the degradation process, and there is always a distribution mismatch between real signals and simulated ones, leading to generalization problems while deploying the trained models on real industrial datasets (C. Liu & Gryllias, 2022). In order to address this issue, domain adaptation or transductive transfer learning methods in machine learning have been used to improve the generalization ability by transferring the knowledge learned from the source domain $\mathcal{D}_S$ to the target domain $\mathcal{D}_T$. In this approach, the feature space between the domains is the same but the marginal probability distributions of them are different (Pan & Yang, 2010).

C. Liu and Gryllias (2022) proposed a simulation-based transfer learning method based on domain adaptation that leveraged simulated signals for training in order to classify bearing faults in real samples. They showed that domain adaptation will improve the accuracy when it is applied to simulation-based neural networks. J. Liu et al. (2023) proposed a method that used a dynamic model of rolling bearings to generate vibration signals for different types of faults. Then, a non-adversarial domain adaptation network was used to provide knowledge for other working conditions. Domain adaptation showed an improvement in the accuracy of the classifiers. C. Liu et al. (2020) used a phenomenological model to generate synthetic signals that simulated a degradation process in bearings, aiming to create a synthetic run-to-failure dataset for training a transfer learning model based on domain adversarial neural networks (DANN). They showed that domain adaptation significantly enhanced the accuracy of the predictions.

Moreover, varying speed condition poses an additional challenge that impacts the performance of deep models developed based on the steady operating condition assumption (Chi et al., 2022). Varying speed operating conditions can be seen in industrial robots, wind turbines, servo motors, etc. (Cocconcelli et al., 2011; Xia et al., 2022; Xue et al., 2017). In some cases, in condition monitoring strategies, the tachometer pulses are also available that are commonly used in signal processing methods to avoid misinterpretations when dealing with varying speed operating conditions. Likewise, tachometer data or speed signals can also be used in the AI-based methods as a context that helps the network improve its accuracy. “Context” can be defined as any information about working conditions such as load, temperature, and speed that has a significant effect on the equipment’s behavior (Leturiondo et al., 2017).

Ciani et al. (2019) used the temperature and humidity information as contexts or covariates to improve the reliability estimation of wind turbines. Rezamand et al. (2021) introduced a method to utilize the wind speed and ambient temperature of the assets as well as vibration features to estimate the RUL of wind turbines. They showed that speed and temperature information will improve the accuracy under varying operating conditions compared to the model that did not use this information. Chi et al. (2022) employed sparse auto-encoders to encode the speed information, intending to combine them with deep vibration features obtained by transformers. This fusion aimed to enhance the accuracy of bearing fault classifications under variable speed operating conditions.

Driven by the observation that the RUL prediction under varying speed operating conditions is still an open research question, this study aims to present a chain of methods for RUL prediction in bearings under varying speed operating conditions. The whole chain can be summed up as follows:

- Generate synthetic signals under varying speed operating conditions using a phenomenological model and simulate the degradation process in bearings to create a synthetic run-to-failure dataset.
- Propose a context-aware neural network architecture based on the DANN model and train it using the synthetic labeled samples as the source domain and the real unlabeled samples as the target domain for the domain adaptation task.

The rest of this article is organized as follows. Section 2 explains the fundamental theory that has been used in the proposed methods. Section 3 describes the proposed methodologies to predict the RUL. Section 4 introduces a run-to-failure dataset under varying speed operating conditions and shows the results of the proposed model applied to this dataset. Finally, Section 5 provides the conclusion of the paper.

2. THEORETICAL PART

2.1. Phenomenological model

The phenomenological simulation of the bearing vibration signals tries to emulate the real vibration signals coming from a real bearing by comparing the whole bearing structure to a single-degree-of-freedom (SDOF) vibration system subjected to a repetitive excitation force due to the localized faults presented in the bearing. The initial idea was proposed by McFadden and Smith (1984) and then it was improved by Antoni (2007) in order to have a more realistic spectral analysis. The simulated vibration signal can be generated by the following formula (Antoni, 2007):

$$x(t) = \sum_{i=-\infty}^{+\infty} h(t - iT - \tau_i) q(iT) A_i + n(t) \quad (1)$$

where $h(t)$ is the impulse response of the equivalent SDOF system. T is the time period between two consecutive

impacts. i is the index of the i-th impact due to the fault, $n(t)$ accounts for the possible noise presented in the signals, and q is the amplitude modulating function due to the load distribution. A and τ are the parameters in order to take into account the randomness of the impact intensities and the moments that the impacts occur, respectively. According to (Antoni, 2007):

$$E\{\tau_i \tau_j\} = \delta_{ij}\sigma_\tau^2$$
$$E\{A_i^2\} = 1 + \delta_{ij}\sigma_A^2 \quad (2)$$

where σ_τ and σ_A are the standard deviations, and δ_{ij} is the Kronecker symbol.

The time period between two consecutive impacts surely depends on the rotational speed of the inner race of the bearing, and the mean value of the time interval ΔT is expressed by:

$$E\{\Delta T\} = \frac{E\{\Delta\theta\}}{2\pi f_r} \quad (3)$$

where f_r is the inner race rotational speed and $\Delta\theta$ is the angular distance between two consecutive impacts which its mean value is expressed by:

$$E\{\Delta\theta\} = \frac{2\pi}{O_{imp}} \quad (4)$$

where O_{imp} is the characteristic fault order, and it is defined as follows for different types of faults (Buzzoni et al., 2020):

$$\begin{aligned} &\text{Outer race} && \frac{n}{2}\left(1 - \frac{d}{D}\cos(\beta)\right) \\ &\text{Inner race} && \frac{n}{2}\left(1 + \frac{d}{D}\cos(\beta)\right) \\ &\text{Rolling element} && \frac{D}{2d}\left(1 - \left(\frac{d}{D}\cos(\beta)\right)^2\right) \\ &\text{Cage} && \frac{1}{2}\left(1 - \frac{d}{D}\cos(\beta)\right) \end{aligned} \quad (5)$$

where n is the number of rolling elements in the bearing, D is the pitch circle diameter, d is the bearing roller diameter and β represents the contact angle.

2.2. Domain Adversarial Neural Network (DANN)

DANN is an architecture in machine learning which is used to remove the domain shift between the deep features extracted from the source domain dataset and the target domain dataset by capturing the underlying shared information across domains. In this way, the extracted features from both domains are domain-invariant which is one of the important properties in order to increase the generalization ability of the neural network. As illustrated in Figure 1, the network consists of three different parts: feature extractor, domain classifier, and label predictor or regressor.

The feature extractor G_f is typically a deep neural network that learns to extract high-level representations from the input data. It transforms the input samples into a latent representation that encodes useful features for the following layers. The domain classifier G_d is another neural network component that tries to predict the domain of the input samples based on the extracted features. Its purpose is to learn to differentiate between the source and target domains.

The objective of the domain classifier is to maximize its accuracy in predicting the domain while the feature extractor tries to minimize the accuracy of the domain classifier. This is the result of the gradient reversal between these two parts of the network. Therefore, the training process is done in an adversarial manner and at the end of the training process, the domain classifier is no longer able to distinguish between the features coming from different domains. In other words, the feature extractor layer is able to extract domain-invariant features (Farahani et al., 2021). Label predictor or regressor layer G_r are also trying to use the domain-invariant features to estimate the output.

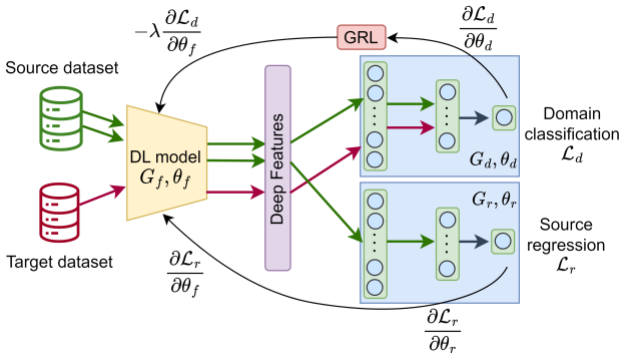

Figure 1. DANN architecture for regression task

Two different losses are defined corresponding to the regressor and domain classifier as follows (Ganin et al., 2016):

$$\mathcal{L}_r^i(\theta_f, \theta_r) = \mathcal{L}_r\left(G_r\left(G_f(x_i;\theta_f);\theta_r\right), y_i\right)$$
$$\mathcal{L}_d^i(\theta_f, \theta_d) = \mathcal{L}_d\left(G_d\left(G_f(x_i;\theta_f);\theta_d\right), d_i\right) \quad (6)$$

where θ_f, θ_r, and θ_d represent the parameters of the G_f, G_r, and G_f respectively.

The goal of the DANN is to optimize the loss function:

$$\mathcal{L}(\theta_f, \theta_r, \theta_d) = \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}_r^i(\theta_f, \theta_r) - \lambda\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}_d^i(\theta_f, \theta_d) + \frac{1}{n'}\sum_{i=n+1}^{N}\mathcal{L}_d^i(\theta_f, \theta_d)\right) \quad (7)$$

where n and n' are the number of samples presented in the source domain and the target domain datasets respectively, and λ is a hyperparameter that controls the trade-off between

the classification loss and the domain adversarial loss during training (Ganin et al., 2016). The optimization algorithms like Adam, SGD, or RMSProp can lead equation 7 to reach a saddle point, which allows for the labeling of the unlabeled data in the target domain (C. Liu et al., 2020).

2.3. Context Awareness

In the area of natural language processing (NLP), positional encodings are used to make the transformers aware of the relative or absolute order of the words inside a sentence (Vaswani et al., 2017). In order to encode the positional information, sine and cosine functions with different frequencies can be used (Vaswani et al., 2017):

$$PE_{(pos,2i)} = sin\left(\frac{pos}{10000^{2i/d_{model}}}\right)$$
$$PE_{(pos,2i+1)} = cos\left(\frac{pos}{10000^{2i/d_{model}}}\right) \quad (8)$$

where pos is the position of the word, d_{model} is the dimension of the word embeddings, and i represents the dimension of the positional encoding. The use of sine and cosine functions with varying frequencies ensures that the positional encodings are unique for each position, providing a way for the model to distinguish between different positions in the input sequence. And more importantly, each positional encoding with offset k, PE_{pos+k} can be described as a linear function of the positional encodings PE_{pos}, allowing the model easily learn the relative dependencies (Vaswani et al., 2017).

The fascinating part of the story is that all the concepts explained above can be transferred to the prognosis research area because the sequential order of the vibration signals measured from a bearing is of high importance in prognosis and it can be an indicator of how the damage is progressing through time. To put it simply, imagine that there are three signals v_1, v_2, and v_3 measured from a bearing at time t_1, t_2, and t_3. The interpretation of these data is different from when the order of the signals changes to v_2, v_3, and v_1. In other words, the same vibration signals in different contexts can be interpreted in different ways.

Additionally, equation 8 can be used to encode the speed information. Although this method of encoding does not have physical meaning, it can make the neural network aware of the differences between the vibration signals coming from different operating conditions. Each operating condition should have a unique encoding that is assigned to the vibration signals.

The method for specifying the value of d_{model} will be described in section 4.1, and the value of pos is an integer number that starts from 1, indicating the sequential order of each measurement. The same way is followed to encode the speed information. For example, $pos = 1$ is used for the lowest speed when the speed profile is stepwise. For each pos the value of i starts from 0 and ends in $\frac{d_{model}}{2}$ to form a vector of length d_{model}. For each i there are two values, one from $sine$ function and the other from $cosine$ function. For example, the encoding for $pos = 1$ is $[PE_{(1,0)}, PE_{(1,1)}, \dots, PE_{(1,d_{model})}]$ which is a one-dimensional vector.

Context-aware machine learning in prognosis is a way to make the neural network aware of the true relationship between the measurements as well as the operating conditions.

3. PROPOSED METHODOLOGY

As discussed in Section 2, the first step is to create a synthetic dataset using the phenomenological model in order to train the context-aware DANN (CA-DANN) model. Generated signals should be adapted according to the real signals from different aspects. Firstly, the frequency content and the fault characteristic frequencies should be the same in real and synthetic signals. Then, the effect of speed on the amplitude of real signals before the anomaly should be investigated and the same effect will be applied to the synthetic signals. More importantly, the health indicator (HI) of the synthetic signals should be in a way to follow a degradation trajectory after the anomaly so that they can replicate a run-to-failure process. In this paper, peak to peak (PP) of the vibration signal is used as the health indicator. Figure 2 shows the whole algorithm for estimating the RUL of the bearings.

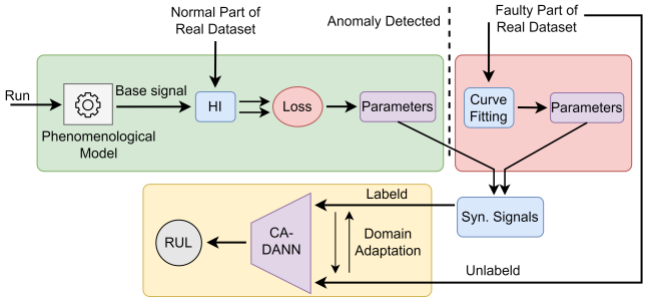

Figure 2. The architecture of the proposed algorithm

3.1. Synthetic run-to-failure dataset

The first step is to replicate the real signals using equation 1. Equation 5 helps to find the time distance between the impacts due to the fault, and equation 9 determines the impulse response $h(t)$ in equation 1:

$$h(t) = e^{-\zeta\omega_n t}\, sin\left(\sqrt{1-\zeta^2}\omega_n t\right)$$
$$\omega_n = \sqrt{k/m} \quad (9)$$
$$\zeta = \frac{c}{2m\omega_n}$$

where c and k are the damping coefficient and stiffness respectively, and m is the equivalent mass. A simple Fast

Fourier method can be applied to the real measurements to determine the dominant natural frequency. Additionally, the logarithmic decrement can be used to see at which rate the amplitude of the impact responses in real measurements is decreasing in order to determine ζ.

Subsequently, a simple assumption should be made in order to easily capture the effect of speed and degradation severity on the amplitude of the signals. Every synthetic signal can be seen as a multiplication of a base signal generated by the phenomenological model and two modifier functions:

$$\boldsymbol{S}_{s,i} = \boldsymbol{S}_{b,i}.(\boldsymbol{R}_i \times D_i) \tag{10}$$

$$D_i = \begin{cases} 1 & t_i < t_{anomaly} \\ e^{a(t_i - t_{anomaly})} & t_i \geq t_{anomaly} \end{cases} \tag{11}$$

where $\boldsymbol{S}_s$ and $\boldsymbol{S}_b$ are the synthetic adapted signal and the synthetic base signal generated by the phenomenological model respectively. $\boldsymbol{R}$ is a linear function of the speed signal from the real measurements, D is a piecewise function that describes the degradation process, and t_i represents the time at the acquisition number i. Function D is assumed to be an exponential function that can describe the dynamic of the degradation in bearings. Its parameter, a, should be found by curve fitting on the health indicator of the real data that comes after the anomaly point.

As defined in equation 11, the effect of degradation severity is stable before the anomaly, therefore, the real signals before the anomaly can be used to recognize the influence of speed on the amplitude of the signals. Figure 3 (a) shows a speed profile and its effect on the peak to peak of vibration. The important point here is to consider the possibility of the structure resonance when speed is varying. In other words, increasing speed does not necessarily result in an increasing amplitude. Figure 3 (b) and (c) show two possible behaviors that can be seen in speed varying scenarios. Increasing amplitude with increasing speed can be a sign of passing no resonance frequency in that specific speed range (Salunkhe & Desavale, 2021).

Function $\boldsymbol{R}_i$ in equation 10 is expressed as follows:

$$\boldsymbol{R}_i = c_j \,.\, \boldsymbol{rpm}_i \tag{12}$$

where c_j is a constant parameter specific to a certain speed j, and $\boldsymbol{rpm}_i$ is the speed signal in the i-th measurement. c_j can be determined using the available real signals before the anomaly. The amplitude of the synthetic signals should be adapted by equation 12 in a way that both real and synthetic signals have the same peak to peak. For this purpose, a mean square error (MSE) function will be formed to measure the error between the peak to peak of the real and synthetic signals that come from the same operating speed, intending to find the c_j that results in the minimum error between peak to peak of the real and synthetic signals at speed j. This method can capture the resonance effect as well, since the parameter c_j is independently determined for each speed.

Figure 4 (b) and (c) show two samples of the MSE loss function for two different speeds to find the optimum constant parameter $c_{j=m,n}$.

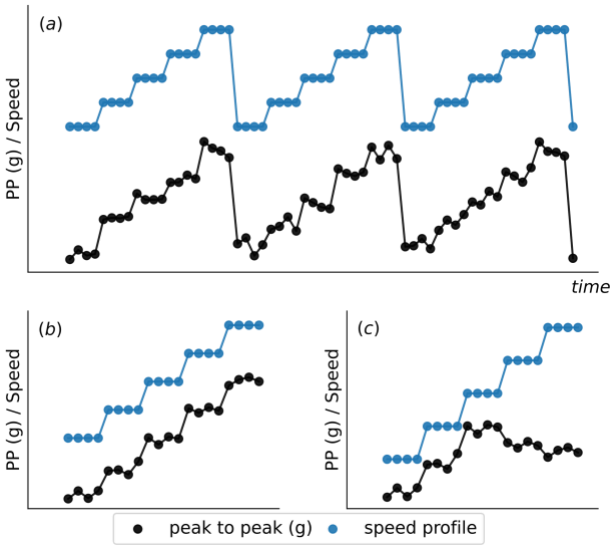

Figure 3. (a) Varying speed effect on the peak to peak amplitude of the signals, (b) in case of no resonance, (c) in case of passing the resonance frequency

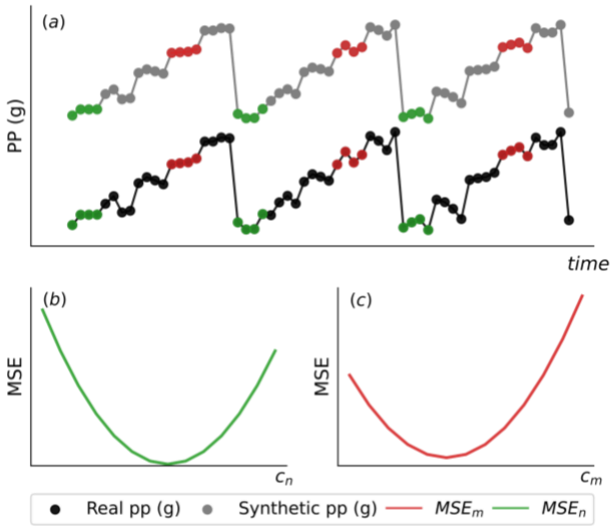

Figure 4. (a) Peak to peak of synthetic and real signals before anomaly (red: speed *m*, green: speed *n*), (b) and (c) MSE of peak to peak of real and synthetic signals for two different speeds

Figure 5 (a) shows the peak to peak of synthetic signals that replicate the real ones before the anomaly. By knowing function $\boldsymbol{R}_i$, the effect of speed can be removed from the health indicator by equation 13, meaning that any changes in the health indicator that is not associated with speed can manifest itself more clearly. This method will be used to find anomalies.

$$HI_{n,i} = HI\left(\boldsymbol{S}_{r,i}/\boldsymbol{R}_i\right) \quad (13)$$

HI_n represents the normalized health indicator (peak to peak), and $\boldsymbol{S}_r$ is the real vibration signal. Investigating the effect of speed on vibrations will be done in a specific time interval, for example from t_0 to t_1 in Figure 5 (a). Every real peak to peak after t_1 will be normalized using equation 13 with its corresponding c_j and $\boldsymbol{rpm}_i$ (c_j has been determined in the previous steps and $\boldsymbol{rpm}_i$ comes with each new measurement). Figure 5 (b) shows the normalized peak to peak whose mean and standard deviation in the time interval $[t_0, t_1]$ can be used as the threshold for anomaly detection.

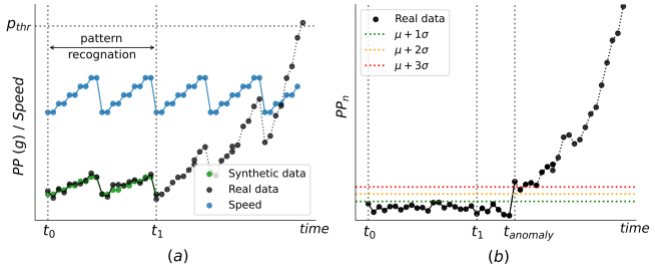

Figure 5. (a) Peak to peak of real and synthetic signals, (b) anomaly detection by the normalized peak to peak

After the anomaly, a limited number of signals will be used for curve fitting in order to find parameter a in equation 11. As Figure 6 (a) shows, curve fitting is done based on the normalized peak to peak because the effect of speed has been removed, and it is only the degradation severity that plays a role. By introducing slight variations in the parameter a in function D, different degradation trajectories can be built in order to have a big synthetic run-to-failure dataset. Figure 6 (b) shows the typical degradation trajectories of the peak to peak of the synthetic dataset.

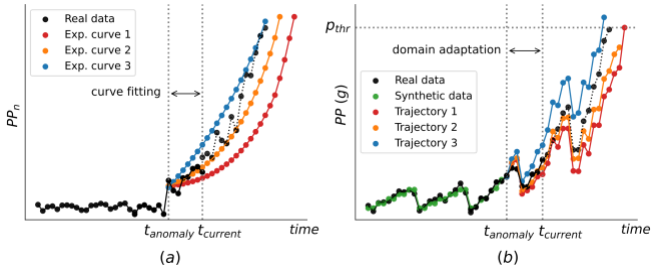

Figure 6. (a) curve fitting according to the normalized peak to peak, (b) peak to peak of the synthetic dataset

3.2. Context-aware DANN

As discussed in Section 2.3, two contexts, speed, and order of the measurements can be seen as two inputs in addition to the raw vibration signal. As illustrated in Figure 7, a Convolutional Neural Network (CNN) is used to extract the local information and deep features automatically from the raw vibration signals without any expert knowledge. The extracted features will be concatenated by two inputs (contexts) to form a bigger 1-D vector which is followed by two parallel Fully Connected (FC) layers, a domain discriminator and a source regressor. The loss function of the regressor part is the mean squared error and the loss function of the discriminator part is the binary cross entropy which is expressed as follows:

$$\mathcal{L}_d = -y.log(\bar{y}) - (1-y).log(1-\bar{y}) \quad (14)$$

where $y \in \{0, 1\}$ is the domain label and $\bar{y}$ is the predicted value between 0 and 1. Table 1 and Table 2 show the network parameters in detail.

Table 1. Network parameters of the feature extractor

Layer	Type	Filter/ Kernel/Stride	Activation function
1	1D CNN	4/128/16	ReLU
2	Max Pooling	-/8/8	-
3	1D CNN	16/16/8	ReLU
4	Max Pooling	-/8/8	-

Table 2. FC parameters in the regressor and the discriminator

Regressor part		Discriminator part	
Layer	Units/Activation function	Layer	Units/Activation function
1	64/softplus	1	64/softplus
2	32/softplus	2	32/softplus
3	1/softplus	3	1/sigmoid

As depicted in Figure 7, the gradient reversal layer (GRL) with the trade-off parameter $\lambda = 0.1$ is also added as the first layer of the discriminator part to reverse the gradient in the backpropagation process. It should be noted that the length of the input signal should be long enough to include the sufficient number of impacts due to the fault in bearing. The number of the impacts due to the ball defect in one revolution of shaft is lower than other type of faults. On the other hand, at the lower shaft speeds, more datapoints are needed to cover a complete revolution of shaft. Using the equation 14, the number of datapoints needed to cover at least 1 impact due to the ball defect can be calculated. This number will cover more than one impact if a different type of defect is present at any speed.

$$L_c = \frac{F_S}{BSF} \quad (14)$$

where L_c is the critical length of signal, BSF is the ball spin frequency at the lowest shaft speed, and F_S is the sampling frequency.

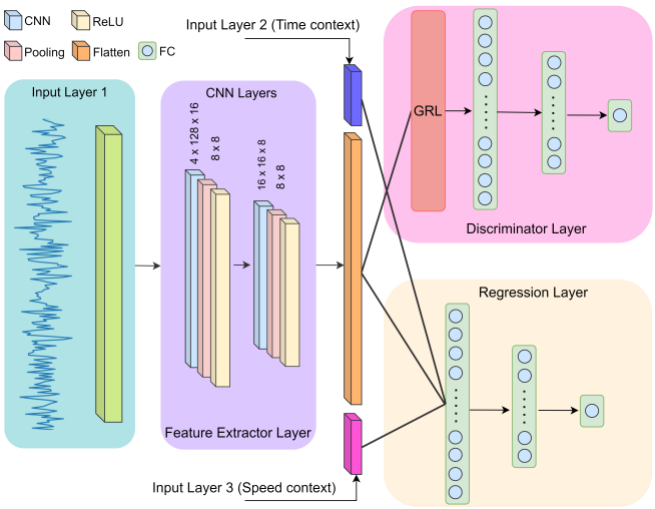

Figure 7. The architecture of the proposed CA-DANN

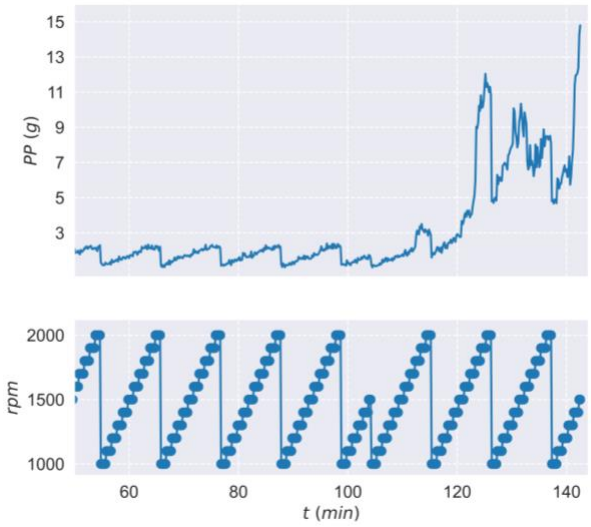

Figure 8. Peak to peak and its corresponding speed profile of bearing A148

4. APPLICATION OF THE METHODOLOGY AND RESULTS

4.1. Smart Maintenance dataset

Smart Maintenance (SM) dataset provided by Flanders Make (Ooijevaar et al., 2019) consists of accelerated life tests where indentations were deliberately created on the inner races (IR) of bearings using a Rockwell-C indenter with a force of 100 *kg* before the tests started to run. The diameter of the indentations is carefully regulated to be within the range of 400 ± 25 μm. The radial load is 9 *kN* and the rotational speed follows a periodic stepwise profile starting from 1000 rpm to 2000 rpm, each step is 100 rpm and is maintained for 60 seconds. The type of test bearings is 6205-C-TVH from FAG. The sampling rate frequency is 50 kHz, and the stopping criterion for the experiments was based on the peak to peak value reaching 20g. However, some experiments were not ended at 20g. Therefore, in order to have a consistent dataset, a new threshold at 15g is considered for the end of life. Figure 8 shows the peak to peak and speed profile of one of the experiments. Table 3 shows the specifications of all the bearings that are used in this study.

Table 3. Bearing information in the SM dataset

Bearing	Test duration	Anomaly detected	Fault
A148	143.3 min.	at 112.6 min.	IR
A150	197.5 min.	at 169.1 min.	IR
A153	229.3 min.	at 207.8 min.	IR
A154	126.0 min.	at 98.8 min.	IR
A155	369.3 min.	at 348.6 min.	IR
A156	251.5 min.	at 224.0 min.	IR

Table 4. Length of the inputs of the proposed architecture

Input No.	Length
Input 1 (raw signal)	25000
Input 2 (time context)	24
Input 3 (speed context)	24

Referring to the equation 14, a signal of 25000 datapoints is set as the input to make sure that at least 20 impacts will be covered in the critical scenario for the Smart Maintenance dataset. The length of the other inputs, d_{model} in equation 8, should be lower than the length of the deep features after Flatten layer in order to make sure that the following layers will learn the features effectively. For the architecture described in Table 1, the length of the deep features for the input length of 25000 is 112.

4.2. Results and discussions

The phenomenological model is used to generate synthetic signals. Figure 10 shows one of the examples of the generated signals for bearing A148.

The results of the proposed method for anomaly detection are mentioned in Table 3. After the anomaly, the DANN model needs some real signals to perform the domain adaptation task. In this study, 10 minutes of the real dataset after anomaly is used for this purpose. It should be noted that these samples are unlabeled, therefore, they will be used in the inference stage as well. Despite the passage of time, labeling the past samples is still valuable, since it indicates what were the predictions from a few moments ago which can be used for decision-making.

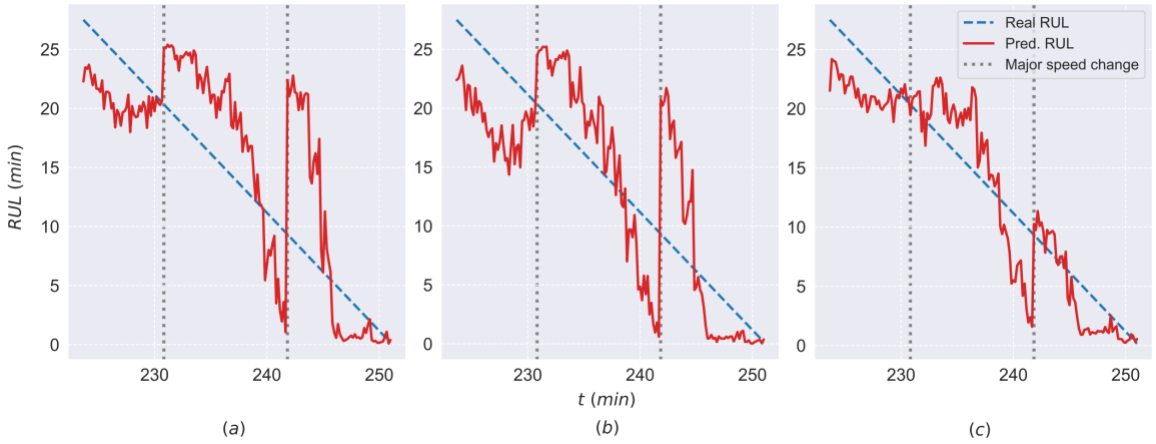

Figure 9. Predicted RUL of the bearing A156, (a) CNN, (b) DANN, (c) CA-DANN

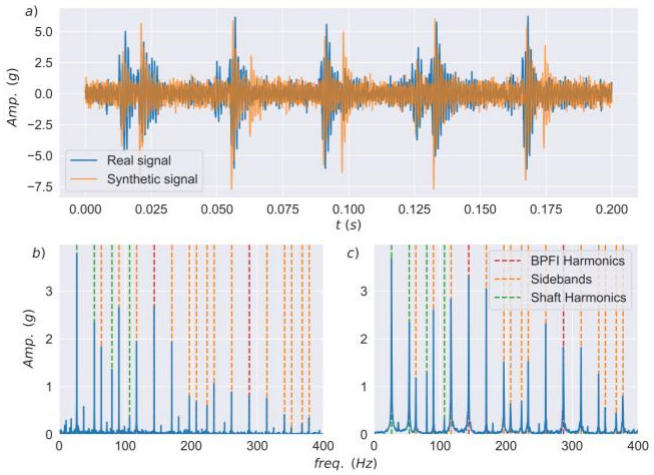

Figure 10. (a) Comparison between the real and generated signals, (b) envelope spectrum of the real signal and (c) generated signal

The synthetic run-to-failure dataset generated by the steps described in Section 3.1 is used to train 4 models. A simple CNN model that neither includes the discriminator part of the architecture nor contexts, a DANN model without contexts, the proposed CA-DANN model with time context, and the CA-DANN model with both time and speed contexts. Table 4 shows the length of the inputs of the CA-DANN model.

Table 5 shows the superior performance of the CA-DANN model which in all cases can improve the mean squared error (MSE) and the mean absolute error (MAE) of the RUL predictions. However, including the time information only as the context is not sufficient in varying speed operating conditions. As discussed before, the important point of the context awareness models is to make the model aware of the operating conditions and any other information that influences the physical behavior of the assets. This fact is perfectly shown in Figure 11 where by using the t-distributed stochastic neighbor embedding (t-SNE) technique the feature distribution of the second to the last fully connected layer in the regressor part of the CA-DANN model is visualized. This layer outputs a 32-dimensional feature space that t-SNE can reduce the dimension to a lower one such as a 2-dimensional feature space which is easier to visualize.

Table 5. MSE and MAE of the predicted RUL of the SM bearings in minutes

Bearing	Error	CNN	DANN	CA-DANN (time context)	CA-DANN
A148	MSE	6.22	6.76	6.68	**6.02**
	MAE	4.68	4.91	5.14	**4.51**
A150	MSE	7.46	6.45	7.04	**6.03**
	MAE	6.12	5.42	5.75	**4.71**
A153	MSE	3.80	3.44	3.17	**2.13**
	MAE	3.19	2.65	2.83	**1.77**
A154	MSE	6.69	6.81	6.39	**4.72**
	MAE	5.57	5.60	4.73	**3.81**
A155	MSE	6.97	5.94	6.78	**4.20**
	MAE	5.44	4.79	5.19	**3.58**
A156	MSE	5.44	5.49	4.51	**3.12**
	MAE	4.40	4.79	3.89	**2.58**

Figure 11 shows how contextual information helps the CA-DANN model distinguish between different speeds, resulting in better predictions. More importantly, context awareness makes the model more robust against the major changes in the speed profile. For example, bearing A156 underwent two major changes in operating speed after the anomaly. As depicted in Figure 9, major speed changes from 2000 rpm to 1000 rpm in a short time interval led the CNN and DANN models to have a higher error in the predictions. However, the CA-DANN model effectively addresses these abrupt changes

by understanding that the high fluctuations of the amplitude of the signals at these moments are not due to the degradation severity but because of the operating speed.

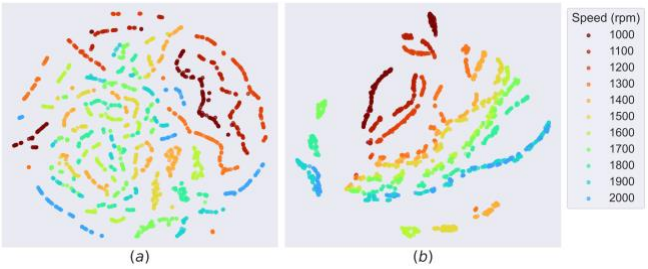

Figure 11. t-SNE visualization of the second to the last layer of the regressor part, bearing A156, (a) DANN, (b) CA-DANN

5. CONCLUSION

The objective of this study is to utilize a phenomenological model to create synthetic run-to-failure datasets under varying speed operating conditions for training the machine learning models. Additionally, the proposed CA-DANN model facilitates the incorporation of contextual information in a deep-learning model for prognosis. The suggested approach demonstrates that integrating contextual information into the architecture of the DANN model enables the model to gain knowledge about both the operating conditions and the sequential order of measurements. Although the method for encoding the contextual information does not convey any physical meaning, it can help the network differentiate between the signals coming from different operating conditions and find the relative relationships between them. The experimental results on the SM dataset prove that the CA-DANN model is able to digest the speed changes to some degree and offers more reliable predictions at the moment of major speed changes. Thanks to the t-SNE technique, the fact that CA-DANN can discriminate between different operating conditions has been shown. On the other hand, leveraging the phenomenological model and simulated signals to create a synthetic run-to-failure dataset enhances the cost-effectiveness of the proposed approach by minimizing the reliance on historical run-to-failure datasets. The flexibility of the proposed method in recognizing the speed influences on the amplitude of the signals makes it applicable to the various periodic speed profiles, and also different speed ranges, including those that may or may not pass the resonance frequency of the structure. Experimental results have shown the capability of the proposed method compared to the models that are not informed of the contextual information.

ACKNOWLEDGMENT

The authors would like to acknowledge the support of Flanders Make, the strategic research center for the manufacturing industry in the context of the DGTwin Prediction project, the Flemish Government under the "Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen" program, and the Flemish Supercomputer Center (VSC) for the computation resource.

REFERENCES

Antoni, J. (2007). Cyclic spectral analysis of rolling-element bearing signals: Facts and fictions. *Journal of Sound and Vibration*, *304*(3–5), 497–529. https://doi.org/10.1016/j.jsv.2007.02.029

Buzzoni, M., D'Elia, G., & Cocconcelli, M. (2020). A tool for validating and benchmarking signal processing techniques applied to machine diagnosis. *Mechanical Systems and Signal Processing*, *139*, 106618. https://doi.org/10.1016/j.ymssp.2020.106618

Chen, Y., Peng, G., Xie, C., Zhang, W., Li, C., & Liu, S. (2018). ACDIN: Bridging the gap between artificial and real bearing damages for bearing fault diagnosis. *Neurocomputing*, *294*, 61–71. https://doi.org/10.1016/j.neucom.2018.03.014

Chi, F., Yang, X., Shao, S., & Zhang, Q. (2022). Bearing Fault Diagnosis for Time-Varying System Using Vibration–Speed Fusion Network Based on Self-Attention and Sparse Feature Extraction. *Machines*, *10*(10), 948. https://doi.org/10.3390/machines10100948

Ciani, L., Galar, D., & Patrizi, G. (2019). Improving context awareness reliability estimation for a wind turbine using an RBD model. *2019 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 1–6. https://doi.org/10.1109/I2MTC.2019.8827041

Cocconcelli, M., Rubini, R., Zimroz, R., & Bartelmus, W. (2011). Diagnostics of ball bearings in varying-speed motors by means of Artificial Neural Networks. *8th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies 2011, CM 2011/MFPT 2011*, *2*, 760 – 771. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84905727322&partnerID=40&md5=9502865e8a2f88027b67d77bf5ec422b

Farahani, H. S., Fatehi, A., Nadali, A., & Shoorehdeli, M. A. (2021). Domain Adversarial Neural Network Regression to design transferable soft sensor in a power plant. *Computers in Industry*, *132*, 103489. https://doi.org/https://doi.org/10.1016/j.compind.2021.103489

Farhat, M. H., Chiementin, X., Chaari, F., Bolaers, F., & Haddar, M. (2021). Digital twin-driven machine learning: ball bearings fault severity classification. *Measurement Science and Technology*, *32*(4), 044006. https://doi.org/10.1088/1361-6501/abd280

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., & Lempitsky, V.

(2016). *Domain-Adversarial Training of Neural Networks*.

Gryllias, K. C., & Antoniadis, I. A. (2012). 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. https://doi.org/10.1016/j.engappai.2011.09.010

Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical Systems and Signal Processing*, *104*, 799–834. https://doi.org/10.1016/j.ymssp.2017.11.016

Leturiondo, U., Salgado, O., Ciani, L., Galar, D., & Catelani, M. (2017). Architecture for hybrid modelling and its application to diagnosis and prognosis with missing data. *Measurement*, *108*, 152–162. https://doi.org/10.1016/j.measurement.2017.02.003

Liu, C., & Gryllias, K. (2022). Simulation-Driven Domain Adaptation for Rolling Element Bearing Fault Diagnosis. *IEEE Transactions on Industrial Informatics*, *18*(9), 5760–5770. https://doi.org/10.1109/TII.2021.3103412

Liu, C., Mauricio, A., Qi, J., Peng, D., & Gryllias, K. (2020). *Domain Adaptation Digital Twin for Rolling Element Bearing Prognostics*.

Liu, J., Cao, H., Su, S., & Chen, X. (2023). Simulation-Driven Subdomain Adaptation Network for bearing fault diagnosis with missing samples. *Engineering Applications of Artificial Intelligence*, *123*. https://doi.org/10.1016/j.engappai.2023.106201

McFadden, P. D., & Smith, J. D. (1984). Model for the vibration produced by a single point defect in a rolling element bearing. *Journal of Sound and Vibration*, *96*(1), 69–82. https://doi.org/10.1016/0022-460X(84)90595-9

Ooijevaar, T. H., Pichler, K., Di, Y., Devos, S., Volckaert, B., Hoecke, S. Van, & Hesch, C. (2019). Smart Machine Maintenance Enabled by a Condition Monitoring Living Lab. *IFAC-PapersOnLine*, *52*(15), 376–381. https://doi.org/10.1016/j.ifacol.2019.11.704

Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191

Randall, R. B., & Antoni, J. (2011). Rolling element bearing diagnostics—A tutorial. *Mechanical Systems and Signal Processing*, *25*(2), 485–520. https://doi.org/10.1016/j.ymssp.2010.07.017

Rezamand, M., Kordestani, M., Orchard, M. E., Carriveau, R., Ting, D. S.-K., & Saif, M. (2021). Improved Remaining Useful Life Estimation of Wind Turbine Drivetrain Bearings Under Varying Operating Conditions. *IEEE Transactions on Industrial Informatics*, *17*(3), 1742–1752. https://doi.org/10.1109/TII.2020.2993074

Salunkhe, V. G., & Desavale, R. G. (2021). An Intelligent Prediction for Detecting Bearing Vibration Characteristics Using a Machine Learning Model. *Journal of Nondestructive Evaluation, Diagnostics and Prognostics of Engineering Systems*, *4*(3). https://doi.org/10.1115/1.4049938

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All You Need*.

Xia, B., Wang, K., Xu, A., Zeng, P., Yang, N., & Li, B. (2022). Intelligent Fault Diagnosis for Bearings of Industrial Robot Joints Under Varying Working Conditions Based on Deep Adversarial Domain Adaptation. *IEEE Transactions on Instrumentation and Measurement*, *71*, 1–13. https://doi.org/10.1109/TIM.2022.3158996

Xue, L., Li, N., Lei, Y., & Li, N. (2017). Incipient Fault Detection for Rolling Element Bearings under Varying Speed Conditions. *Materials*, *10*(6), 675. https://doi.org/10.3390/ma10060675

BIOGRAPHIES

Seyed Ali Hosseinli received his BSc and MSc 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 the Department of Mechanical Engineering of KU Leuven, Belgium. His research interests focus on machine learning techniques in prognostics and fault diagnosis of rotary equipment.

Ted Ooijevaar is a Monitoring Technology Domain Lead and Senior Research Engineer at Flanders Make since 2015. In this role he defines, leads and performs industry-driven applied research and development to support companies in the manufacturing industry. Ted received a PhD degree in field of Mechanical Engineering from the University of Twente and has special expertise in the field of sensing and monitoring, signal processing and data analytics, dynamics and mechanics, modeling and experimental testing.

Konstantinos Gryllias holds a 5 years engineering diploma degree and a Ph.D. degree in Mechanical Engineering from the National Technical University of Athens, Greece. He holds an associate professor position on vibro-acoustics of machines and transportation systems at the Department of Mechanical Engineering of KU Leuven, Belgium. He is also the manager of the University Core Lab Flanders Make@KU Leuven - MPro of Flanders Make, Belgium. His research interests lie in the fields of condition monitoring, signal processing, prognostics, and health management of mech. & mechatronic systems