Simulation-based remaining useful life prediction of rolling element bearings under varying operating conditions

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

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

[2] *Flanders Make @ KU Leuven*

Celestijnenlaan 300, BOX 2420, 3001 Leuven, Belgium

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

konstantinos.gryllias@kuleuven.be

ABSTRACT

Remaining useful life (RUL) prediction of rolling element bearings is a complex task in the frame of condition monitoring which brings cost benefits to the industry by reducing unexpected downtimes and failures. Data-driven approaches based on deep learning have demonstrated exceptional performance in estimating RUL effectively. Nevertheless, challenges such as data scarcity for model training and varying operating conditions add more complexity to prognostic tasks using these methods. This study proposes a methodology for simulating the vibration signals during the degradation process of bearings in order to mitigate the need for historical data for training the models. Simulations are realized using a phenomenological model whose free parameters are adapted based on real measurements so that the simulated run-to-failure datasets are under the same influence of speed as the real dataset with almost the same degradation rate. The simulated dataset is used for model training. Moreover, the proposed methodology is able to react to the shaft speed and be flexible at the predictions when the speed of the bearing varies. The proposed model can take extra information regarding the operating speed and the sequential ordering of the measurements to be aware of the working conditions and the dynamics of the damage progression. The positive effect of the extra information is shown in the results. Model training is based on an unsupervised domain adaptation approach to reduce domain discrepancy between the simulated and real feature space. The effectiveness of the proposed method is examined according to bearing run-to-failure tests under varying operating conditions.

1. INTRODUCTION

Improving the accessibility of industrial assets is a crucial factor in boosting productivity and efficiency, leading to cost benefits for industries. This is achieved by exploiting the full life of components and avoiding premature replacements. Rolling element bearings, being the key component of rotary equipment in the industry, are prone to failures due to their frequent operation in harsh and demanding conditions, including high temperatures, heavy loads, and contaminated surroundings, which increase the possibility of unexpected failures. Failures could spread to the entire machine and lead to unplanned downtimes (Buzzoni et al., 2020; Tajiani & Vatn, 2023).

RUL estimation techniques can be employed to determine the remaining time until the failure occurs. Failure is defined as a state in which a health indicator crosses a predefined threshold such that the component is no longer able to operate in the desired way (Lei et al., 2018). Numerous methods can be employed to achieve this objective. However, data-driven approaches, particularly those based on deep learning, have recently demonstrated outstanding results due to their capability of modeling processes with high complexity like degradation process in bearings (Fink et al., 2020). The main bottleneck of using deep learning methods is the necessity of having a large amount of labeled pre-recorded run-to-failure data for training the models. This process is time-consuming, labor-intensive, and crucially, in industrial settings, failures may be infrequent, and maintenance is usually performed before failures (Arias Chao et al., 2022), which makes it hard to have a perfect dataset for training the models, since for model training, datasets should cover the whole life of bearings until the point of failure, and the degradation process is normally a long process that could take months or even years (Chen et al., 2023). Additionally, labeled data from

different bearings or machines operating under different conditions may be insufficient due to different deterioration trajectories and conditional probabilities. As a solution, leveraging simulated datasets for training deep learning models has surfaced as a practical approach to mitigate the constraints imposed by the limited availability of real labeled datasets. This approach enhances the overall performance of the models by providing a broader and more diverse set of training data (Hosseinli et al., 2023). Gryllias and Antoniadis (2012) generated artificial signals by a phenomenological model for different types of localized faults in bearings and then trained a Support Vector Machine (SVM) model using them. The real samples were then classified using the trained SVM model. Cui et al. (2020) proposed a method based on a 5-DOF dynamic model of bearings coupled with surface topography excitation to create a dictionary of many different degradation processes. Then, based on the similarity of the tested bearing and the simulated ones, the RUL of the tested bearing can be estimated based on the life label of the most matched sample. Deng et al. (2023) developed a 5-DOF dynamic model of bearings and generated a large amount of samples. Then, a particle filter-based dynamic calibration method was used to calibrate the parameters of the model based on observations. The simulated dataset was further used to train a deep learning model and estimate the RUL of real samples. Ai et al. (2023) utilized a phenomenological model to create a dataset for three types of fault: ball, inner race, and outer race for fault diagnosis. A deep learning model based on the transfer learning approach was then trained to remove the gap between the distributions of the real and simulated signals for fault classification of the real dataset.

Moreover, varying operating conditions, which can be seen in industrial cases such as wind turbines, servo motors, compressors, etc., pose another challenge for estimating the RUL of bearings. Developing a RUL prediction method that can respond to the operating conditions is of high importance since the developed models based on the assumption of steady operating conditions could not have satisfying performance under varying operating conditions (Chi et al., 2022; Liao & Tian, 2013). (Wang et al., 2021) developed a model-based method for RUL estimation by considering the joint dependency of degradation rate and time-varying operating conditions. The parameters of a system state function and an observation function were then estimated to model the degradation process of the system and predict the RUL of bearings. (Li et al., 2019) developed a state-space model for systems working under varying operating conditions. The model considered two effects of the varying operating conditions: changes in degradation rates and jumps in degradation signals. By estimating the underlying system state and the remaining time until it reaches a failure predefined threshold, the RUL of tested bearings was estimated. Zhang et al. (2022) proposed a normalization method that recalibrates the upward and downward abrupt jumps of sensor readings at the operational conditions change points. Then, the normalized sensor features and operating condition features were fed to a gated recurrent unit (GRU) to estimate the RUL of the aircraft turbofan engine dataset provided by NASA.

Motivated by the observation that the literature lacks a comprehensive exploration of RUL estimation under varying operating conditions using deep learning, in this paper, the proposed methodology consists of different steps including data simulation as a way of mitigating the influence of data scarcity and then a deep learning model based on a domain adaptation approach which gets the raw vibration signals as input as well as supplementary information on working conditions in which the signals are acquired in order to make the model aware of varying operating conditions. The rest of the paper can be summarized as follows:

1. Utilize a phenomenological model that simulates the general vibration signals of the bearings under different fault modes: ball, inner race, and outer race defects.
2. Adapt the phenomenological model based on the healthy real signals to tune the dynamic parameters of the model and also identify the effect of varying speed conditions on the amplitude of vibration signals.
3. Separate the effect of speed from the peak-to-peak health indicator so that it only indicates the degradation process which is void of the influence of speed and realizing anomaly detection based on this new health indicator (normalized peak-to-peak).
4. Realize curve-fitting on the normalized peak-to-peak after the anomaly to find out how fast the damage is progressing and then generate many synthetic run-to-failure data under the same influence of speed and damage progression as the real data to create a big training dataset for training a deep learning model.
5. Train a deep learning model according to the domain adversarial method to decrease the discrepancy between the unlabeled real data and the labeled simulated data. The deep learning model takes two additional inputs: speed information and sequence information in order to better understand the working conditions and the sequential relationship between each measurement.
6. Estimate the RUL of the real measurements using the trained deep-learning model.

In other words, the proposed methodology, as shown in Figure 1, is a digital twin (DT) that requires no historical data and is able to adapt itself to different rotating speeds, fault modes, and degradation rates of bearings.

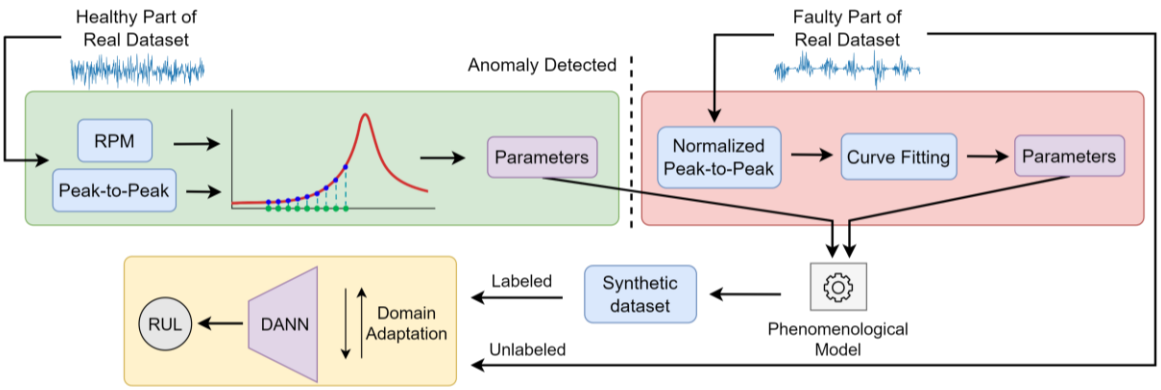

Figure 1. Pipeline of the proposed methodology

Peak-to-peak values are used here as a reference health indicator to tune the digital twin, since the EoL criterion is assumed to be defined on the peak-to-peak, and the synthetic dataset needs to be created under the same definition as the real dataset's EoL to be able to mimic historical datasets for training. Moreover, a few unlabeled real data that comes after the anomaly detection are used for unsupervised domain adaptation. They are unlabeled because their corresponding RULs are not known at this stage.

The rest of the paper is organized as follows. First, the fundamental theoretical background used in the proposed methodology is shortly introduced in Section 2. Moreover, the proposed approach to adapt the DT and predict the RUL is introduced in Section 3. Furthermore, a run-to-failure dataset captured under varying speed operating conditions is presented in Section 4, the methodology is applied, and the effectiveness of the proposed approach is demonstrated. Finally, Section 5 provides the conclusion of the paper.

2. THEORETICAL BACKGROUND

2.1. Phenomenological model

The phenomenological simulation of bearing vibration signals involves replicating the actual vibration signals of a real bearing. The simulation method achieves this by comparing the entire bearing and its supporting structure to a single-degree-of-freedom (SDOF) vibration system and introducing consecutive impulses to excite the structure, mirroring the effects of localized faults within the bearing. This approach allows for a representation that emulates the characteristics of real-world bearing vibration signals. 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:

$$x_k(t) = S(k)\,.\,D(k)\sum_{i=-\infty}^{+\infty} h(t - iT - \tau_i)q(iT)A_i + n(t) \qquad (1)$$

where $S(k)$ and $D(k)$ are the amplitude modifiers regarding the speed and damage influence on the amplitude of the signals, respectively, for the k-th simulated signal in a degradation process. $h(t)$ is the impulse response of the equivalent SDOF system. T is the time 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):

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

where σ_τ and σ_A are the standard deviations, and δ_{ij} is the Kronecker symbol. The time period between two consecutive impacts depends on the rotational speed of the inner race of the bearing, and the mean value of the time interval ΔT is expressed by:

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

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:

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

where O_{imp} is the characteristic fault order, and it is defined as follows for different types of faults:

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

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 adaptation

Acknowledging that simulated signals are derived from a simplistic model, incapable of capturing all aspects of faulty bearings or the degradation process, a distribution mismatch between real and simulated signals arises. This mismatch poses challenges in generalization when deploying trained models on real datasets. To tackle this issue, a domain adaptation method is employed to enhance generalization by transferring knowledge acquired from the source domain $\mathcal{D}_S$, where simulated signals originate, to the target domain $\mathcal{D}_T$, representing real-world datasets (Pan & Yang, 2010). This facilitates improved performance and adaptability of trained models. In this case, the marginal probability distributions of source and target domains, $P(x_S)$ and $P(x_T)$, are assumed to be different due to the simplicity of the simulations, but their conditional probability distributions, $P(y_S|x_S)$ and $P(y_T|x_T)$, are assumed to be the same due to the fact that in the preprocessing steps, the digital twin is tuned based on the real data in terms of the type of fault, the influence of speed, and the dynamic characteristics of the bearing.

This paper employs a Domain Adversarial Neural Network (DANN) to tackle the abovementioned domain shift problem. The network, illustrated in Figure 1, comprises three key components: a feature extractor G_f a domain classifier G_d, and a label predictor or regressor G_r. The feature extractor G_f is typically a deep neural network responsible for learning high-level representations from input data. It transforms input samples into a latent representation encoding valuable features for subsequent layers. The domain classifier G_d is another neural network component that predicts the domain of input samples based on extracted features, aiming to differentiate between the source and target domains. During training, the domain classifier seeks to maximize its accuracy, while the feature extractor aims to minimize this accuracy by gradient reversal. This adversarial training process results in the domain classifier being unable to distinguish features from different domains, indicating that the feature extractor can extract domain-invariant features. Additionally, the label predictor or regressor layer G_r utilizes these domain-invariant features to estimate the output, contributing to the overall goal of addressing the domain shift problem (Ganin et al., 2016). The objective function of the model is:

$$\mathcal{L}(\theta_f, \theta_r, \theta_d) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}_r^i(\theta_f, \theta_r) - \lambda\left(\frac{1}{n}\sum_{i=1}^{n} \mathcal{L}_d^i(\theta_f, \theta_d) + \frac{1}{n'}\sum_{i=n+1}^{N} \mathcal{L}_d^i(\theta_f, \theta_d)\right) \tag{6}$$

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 regression loss and the domain adversarial loss during training. $\mathcal{L}_r$ and $\mathcal{L}_d$ are defined as:

$$\mathcal{L}_r^i(\theta_f, \theta_r) = \mathcal{L}_r\big(G_r\big(G_f(x_i;\theta_f);\theta_r\big), y_i\big)$$
$$\mathcal{L}_d^i(\theta_f, \theta_d) = \mathcal{L}_d\big(G_d\big(G_f(x_i;\theta_f);\theta_d\big), d_i\big) \tag{7}$$

where θ_f, θ_r, and θ_d are the trainable parameters of the G_f, G_r, and G_d respectively.

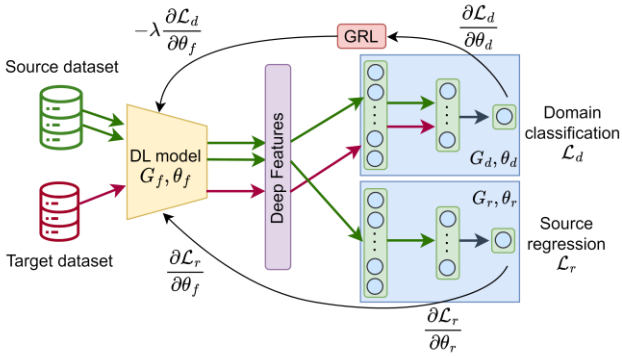

Figure 2. DANN architecture for regression task

3. Proposed Approach

To create a synthetic run-to-failure dataset by the digital twin, at first the dynamic characteristics of the phenomenological model should be tuned based on the real data. Then, the modifier functions $S(k)$ and $D(k)$, introduced in Section 2, are determined.

3.1. Dynamic characteristics

Given the fact that rolling element bearings vibrate even in healthy conditions due to the waviness of the surfaces and other sources of imperfections (Harsha et al., 2003; Jawad & Jaber, 2022), the resonance frequency of the structure can still be seen in the frequency content of the vibration signals (Ghafari et al., 2010). Therefore, by considering the Fast Fourier Transform (FFT) of the healthy signals, the dominant natural frequency of the structure, ω_n, can be found and then used in the phenomenological model as an equivalent SDOF system. Moreover, the logarithmic decrement can be used to see at which rate the amplitude of the impact responses, x, in real measurements is decreasing in order to determine ζ.

$$h(t) = e^{-\zeta\omega_n t}\, sin\left(\sqrt{1-\zeta^2}\,\omega_n t\right)$$
$$\zeta = \frac{\delta}{\sqrt{4\pi^2+\delta^2}} \tag{8}$$
$$\delta = ln\left|\frac{x_1}{x_2}\right|$$

3.2. Influence of speed, $S(k)$

The real signals before the detection of the anomaly can be used to recognize the influence of speed on the amplitude of the vibration signals. Figure 3 (a) shows a speed profile and its effect on the peak-to-peak amplitude of the 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. Figures 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 crossing no resonance frequency in that specific speed range (Salunkhe & Desavale, 2021).

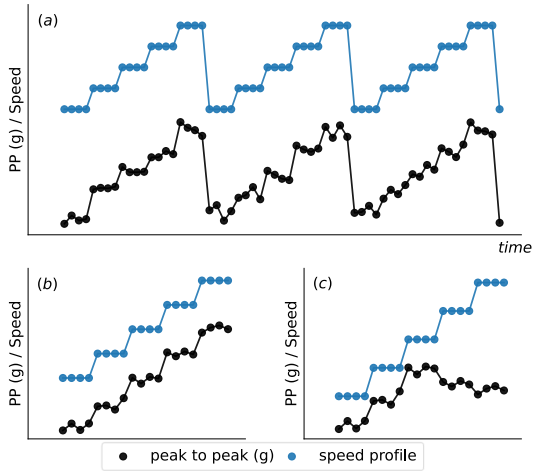

Figure 3. (a) Varying speed effect on the peak-to-peak amplitude of the signals, (b) in case of no resonance crossing, (c) in case of crossing a resonance frequency

Corresponding to each operating speed, a constant parameter c can be found to create a link between the rotating speed and the vibration amplitude or the peak-to-peak:

$$P_j = c_j \,.\, rpm_j \tag{9}$$

where P is the peak-to-peak of real signals, c is a constant parameter, rpm is the operating speed, and j is the index of measurements. In this way, any non-linearity between speed and vibration due to the frequency response of the structure can be captured (Figure 4).

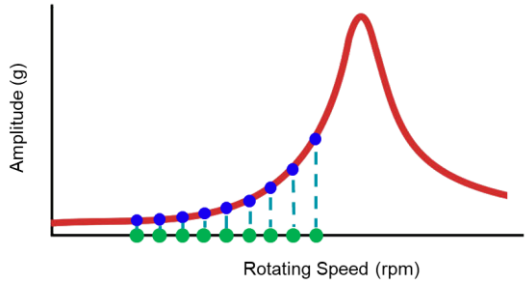

Figure 4. Capturing the complex relationship between speed and vibration amplitude in real measurements

It should be noted that these analyses must be done based on a few measurements at the beginning of the operation of the machine to be far from the influence of bearing faults. After the detection of the anomaly, when the synthetic run-to-failure dataset should be created, a speed profile should also be provided so that the digital twin can generate signals accordingly. Since a random speed profile might be desired at this stage, an interpolation would be needed to find the correct value of c while dealing with unseen speed values. Then, the modifier function $S(k)$ in equation 1 will be:

$$S(k) = c_k \,.\, rpm_k \tag{10}$$

3.3. Normalized health indicator

By knowing the relation between speed and vibration, the effect of speed can be removed from the peak-to-peak amplitude of the vibration signals by equation 11, meaning that any changes in the health indicator that are not associated with speed can manifest itself more clearly. This method will be used to find anomalies. The Normalized peak-to-peak amplitude, P_N, is constructed as follows:

$$P_{N,j} = \frac{P_j}{c_j \,.\, rpm_j} \tag{11}$$

Obviously, for the healthy samples before the detection of the anomaly, $P_N \sim 1$. Figure 5 (b) shows the normalized peak-to-peak amplitude whose mean and standard deviation in the time interval $[t_0, t_1]$, which is at the beginning of the measurements, can be used as the threshold for anomaly detection. $K_{anomaly}$ is used to refer to the k-th signal in the degradation process where the anomaly occurs. As shown in Figure 5 (b), the fluctuations caused by the varying speed profile no longer exist in the normalized peak-to-peak amplitude.

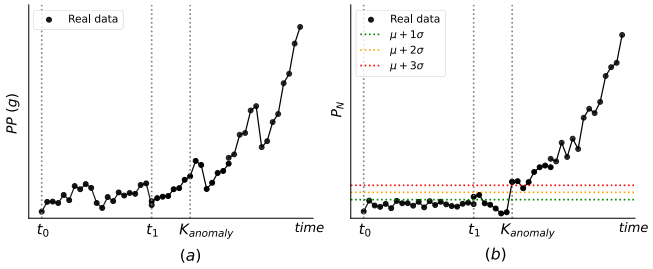

Figure 5. (a) Peak-to-peak amplitude of real data, (b) anomaly detection by the normalized peak-to-peak

3.4. Influence of damage, $D(k)$

After the anomaly, a limited number of signals will be used for curve fitting in order to estimate the degradation rate of the real bearing so that the digital twin will be able to generate a synthetic run-to-failure dataset with almost the same degradation rate as the real bearing. As shown in Figure 6 (a), curve fitting is done based on the normalized peak-to-peak

amplitude because the effect of speed has been removed, and it is only the degradation process that plays a role. The modifier function $D(k)$ in equation 1 can be modeled by an exponential function to approximate the degradation trajectory of the normalized peak-to-peak amplitude:

$$D(k) = e^{a(k-K_{anomaly})} \tag{12}$$

where a is a constant parameter that defines the degradation rate. By introducing slight variations in the parameter a in function $D(k)$, various degradation trajectories can be built in order to have a big synthetic run-to-failure dataset. The variations are such that the time difference between the synthetic EoLs is limited to the *Simulation range* as shown in Figure 6 (b). This figure shows the typical degradation trajectories of the peak-to-peak amplitude of the synthetic dataset generated by equation 1. Domain adaptation is also done using a few real unlabeled available measurements after the detection of the anomaly.

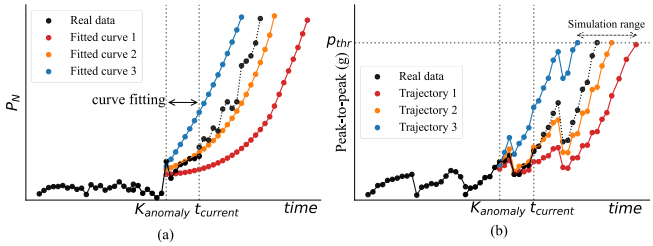

Figure 6. (a) Curve fitting according to the normalized peak-to-peak amplitude, (b) peak-to-peak amplitude of the synthetic dataset

3.5. Encoding

In order to encode the speed and sequence label of each measurement, which will be fed into the deep learning model, this paper adopted one of the well-known methods of information encoding from the natural language processing (NLP) research domain. 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). To encode the positional information, *sine* and *cosine* functions with different frequencies can be used:

$$PE_{(pos,2i)} = sin\left(\frac{pos}{10000^{2i/d_{model}}}\right)$$
$$PE_{(pos,2i+1)} = cos\left(\frac{pos}{10000^{2i/d_{model}}}\right) \tag{13}$$

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. Unique encoding for each position is achieved by using *sine* and *cosine* functions with varying frequencies, making the model able to distinguish the sequential order of measurements. It is important to highlight that each positional encoding with offset k, PE_{pos+k} can be described as a linear function of the positional encodings PE_{pos}, this characteristic enables the model to readily learn the relative dependencies between different positions, contributing to the model's ability to capture sequential information and relationships effectively (Vaswani et al., 2017).

The concept of positional encoding can be transferred to the prognosis research domain. The sequential order of vibration signals obtained from a bearing holds significant importance in prognosis, serving as an indicator of how the damage progresses over time. Moreover, equation 13 can be utilized for encoding speed information. While this encoding method might lack a direct physical interpretation, it serves the purpose of making the neural network aware of distinctions among vibration signals working in different conditions. Each operating condition should have a unique encoding by which the raw vibration signals are accompanied while feeding to the model. d_{model} is a hyperparameter that will be determined in section 4.1, and the value of pos is an integer number that starts from 1, representing 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 rotational speed. For each pos, the value of i starts from 0 and ends in $\frac{d_{model}}{2}$, forming 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 $[PE_{(1,0)}, PE_{(1,1)}, \dots, PE_{(1,d_{model}-1)}]$ which is a one-dimensional vector.

3.6. RUL curve for varying speed scenario

One of the most important outcomes of the proposed methodology is to see how the speed is influencing the RUL. Obviously, for higher speeds, lower RUL is expected, and vice versa. To the best knowledge of the authors, no paper has taken into account the effect of speed on the RUL curve.

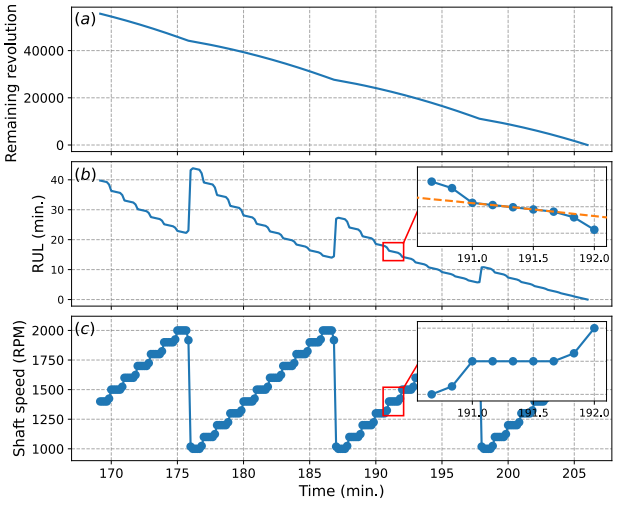

Figure 7. (a), (b) Remaining revolutions and remaining time after the detection of anomaly, (c) the corresponding speed profile

The proposed idea is to estimate the remaining revolutions until the end-of-life, and in the post-processing step, the number of revolutions can be transferred to the remaining time by simply dividing it by the operating speed. Figure 7 shows the remaining revolutions and the RUL for a measurement campaign which has been done under varying speed operating conditions. Equation 14 is used to calculate the total number of revolutions after the detection of anomaly which are then used as the labels in the model training.

$$Rev_{Total} = \sum_{i=K_{anomaly}}^{N} rpm_i . \Delta t$$

$$Rev_n = Rev_{Total} - \sum_{i=1}^{n} rpm_{i+K_{anomaly}} . \Delta t \qquad (14)$$

$$RUL_n = \frac{Rev_n}{rpm_{n+K_{anomaly}}}$$

where Rev_{Total} is the total number of revolutions after the detection of anomaly, N is the number of samples in the run-to-failure experiment, Δt is the length of each measurement. Rev_n and RUL_n are the remaining revolutions and the remaining time until the end-of-life for the n-th sample, respectively. Rev_n is used as the labels for model training.

The important point is that the slope of the RUL curve is -1 as long as the speed is constant, as shown in Figure 7 (b) by the orange line, preserving the most useful property of the RUL curve.

3.7. Deep-learning model

As mentioned in Section 2.2, a deep learning model based on the DANN model is used to estimate the RUL of the real bearings. Two supplementary information, speed and sequential ordering of the measurements, have been encoded and will be fed to the model as extra inputs in addition to the raw vibration signal. As shown in Figure 8, Convolutional Neural Network (CNN) is used to extract the local information and deep features automatically from the raw vibration signals. The extracted features are concatenated by two encoded inputs 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:

$$\mathcal{L}_d = -y . log(\bar{y}) - (1 - y) . log(1 - \bar{y}) \qquad (15)$$

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. As depicted in Figure 8, 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.

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/ ReLU	1	64/ ReLU
2	32/ ReLU	2	32/ ReLU
3	1/ ReLU	3	1/sigmoid

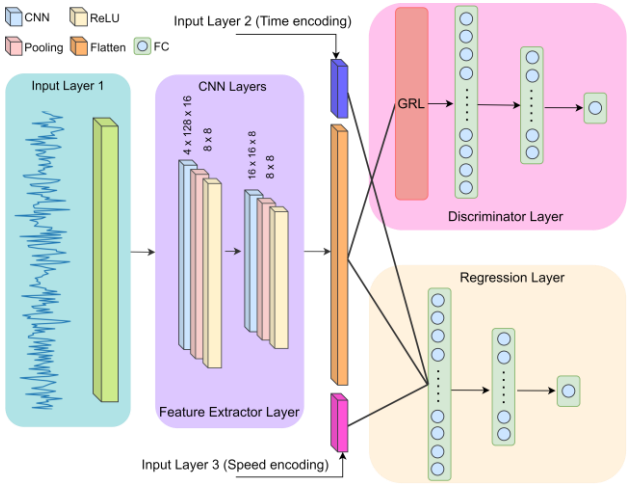

Figure 8. The architecture of the proposed model

It is important to emphasize that the length of the input signal must be sufficiently long to encompass an adequate number of impacts resulting from faults in the bearing. Notably, the number of impacts due to a ball defect in one revolution of the shaft is lower compared to other types of faults. Using equation 16, the number of data points 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 (Hosseinli et al., 2023).

$$L_c = \frac{F_S}{BSF} \qquad (16)$$

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

4. Application of the Methodology and Results

4.1. Case study

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 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 a peak-to-peak amplitude of 15g is considered the end-of-life criterion in this study. Figure 9 shows the peak-to-peak amplitude and the speed profile of one of the measurement campaigns. Table 3 shows the specifications of all the bearings used in this study.

Table 3. Bearing information in the SM dataset

Bearing	Test duration	Anomaly detected at	Fault
A148	142.5 min.	112.6 min.	IR
A150	197.5 min.	169.1 min.	IR
A153	229.3 min.	207.8 min.	IR
A154	126.0 min.	98.8 min.	IR
A155	369.3 min.	348.6 min.	IR
A156	251.3 min.	224.0 min.	IR

Referring to equation 16, a signal of 25000 data points 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 encodings, d_{model} in equation 13, should be kept lower than the length of the deep features after the Flatten layer. This ensures that subsequent layers can effectively learn the deep features by maintaining a lower dimensionality for these encodings compared to the deep features, the model can efficiently process and integrate additional information without overwhelming the learning process or introducing unnecessary complexity. For the architecture described in Table 1, the length of the deep features for the input length of 25000 is 112.

Table 4. Length of the inputs of the proposed architecture

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

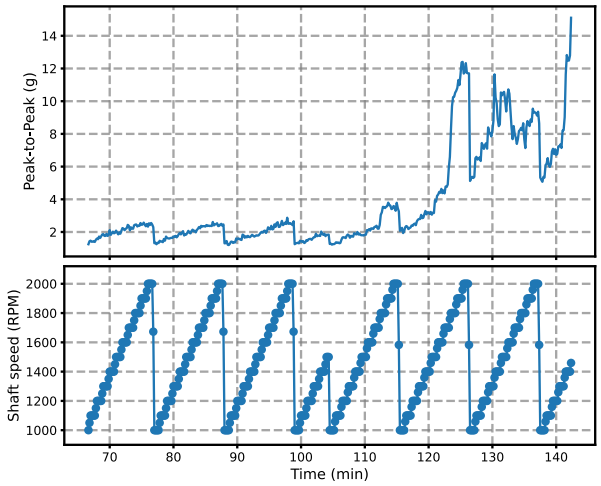

Figure 9. Bearing A148, peak-to-peak amplitude and speed profile

4.2. Results and discussions

Before the anomaly occurs, the relation between speed and vibration amplitude is analyzed. After the anomaly, curve fitting is done based on the available data, as shown in Figure 12. 10 minutes of measurements are used at this stage. This few unlabeled available real data is also used for domain adaptation while model training. It should be highlighted that the unlabeled past samples will be labeled in the inference stage. 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 10 shows the result of anomaly detection and the corresponding speed profile for bearing A148.

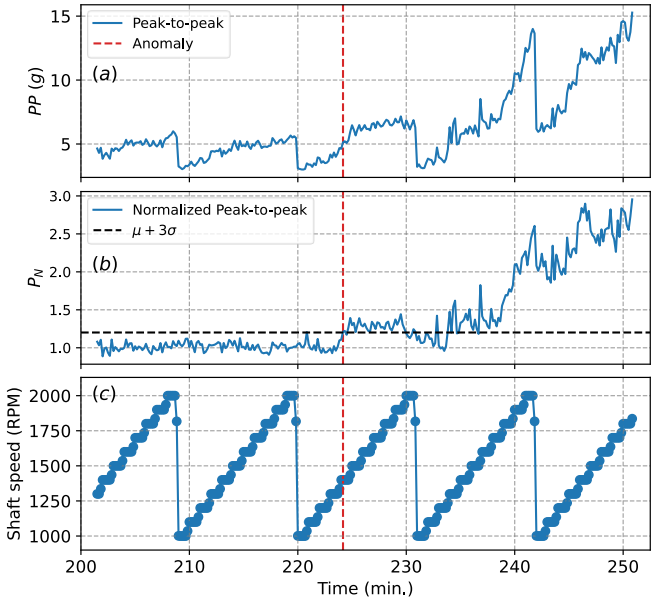

Figure 10. Bearing A156, (a) Peak-to-peak amplitude, (b) Normalized peak-to-peak amplitude, (c) Corresponding speed profile

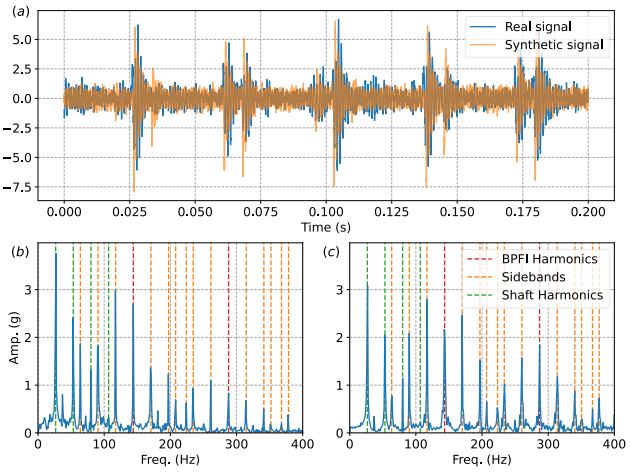

Figure 11. Bearing A148; (a) Comparison between the real and generated signals, (b) envelope spectrum of the real signal, (c) generated signal

Figure 11 shows one of the examples of the generated signals for bearing A148 after the anomaly. The results of the proposed method for anomaly detection are mentioned in Table 3.

After adapting the digital twin in terms of the speed influence and degradation rate, 6 trajectories are generated as shown in Figure 12. Their corresponding raw vibration signals will be the input of the deep learning model for training. The *Simulation range* is chosen to be 40 minutes. The synthetic run-to-failure dataset is used to train 5 models. A simple CNN model that neither includes the discriminator part of the architecture nor encodings, a DANN model without encodings, proposed model 1 with CNN and only speed encodings, proposed model 2 with CNN and only sequence encodings, and proposed model 3 with DANN and both speed and sequence encodings. Table 5 shows the superior performance of the proposed model 3 which in all cases can improve the root mean squared error, RMSE, of the RUL predictions compared to the DANN and CNN model. As discussed before, the important point of feeding the operating condition and the sequential information to the models is to make the model aware of the working environment and any other information that influences the physical behavior of the assets. This fact is perfectly shown in Figure 13 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 proposed 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. Figure 13 shows how the extra information fed to the model helps to distinguish between different speeds, resulting in better predictions. More importantly, supplementary information fed to the model 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 detection of the anomaly. As depicted in Figure 14, abrupt 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. Proving that these models have less control over the predictions when speed plays an impactful role. The proposed model makes satisfying predictions at the moment of abrupt speed changes and the predicted RUL is not too far from the ground truth, showing that the proposed model understands the relationship between rotating speed, vibration, and degradation severity thanks to the encodings. Most importantly, the estimated RUL is reactive to the operating speed. Higher speeds lead to lower RUL and vice versa. This property of the proposed method makes it applicable to real industrial cases where a varying speed profile is used.

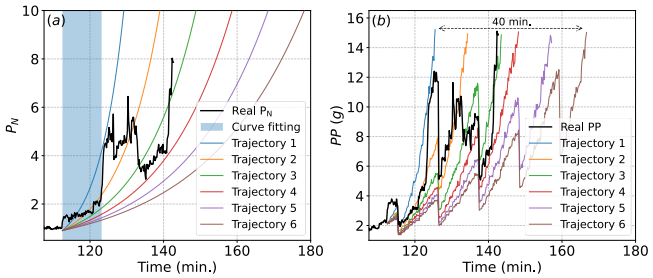

Figure 12. Bearing A148; (a) several trajectories by curve fitting on the real normalized peak-to-peak, (b) peak-to-peak of the generated signals by digital twin

Table 5. RMSE of the predicted RUL of the SM bearings in minutes

Bearing	CNN	DA NN	Proposed model 1	Proposed model 2	Proposed model 3
A148	6.6	7.8	7.3	**5.7**	5.8
A150	7.6	7.9	6.5	**5.7**	**5.7**
A153	4.2	3.8	2.8	2.5	**2.4**
A154	6.2	6.0	5.2	**2.7**	**2.7**
A155	5.5	5.3	4.7	3.7	**2.9**
A156	7.4	7.0	4.5	3.6	**2.7**

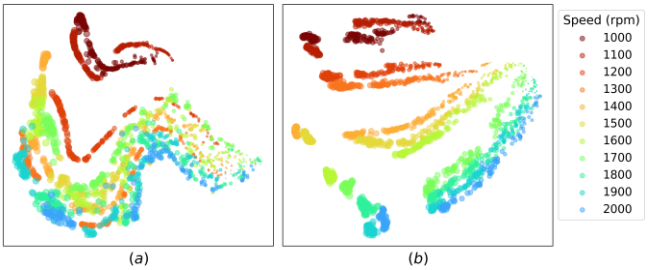

Figure 13. t-SNE visualization of the second to the last layer of the regressor part, the size of circles is proportional to the RUL, bearing A156, (a) CNN, (b) Proposed model 3

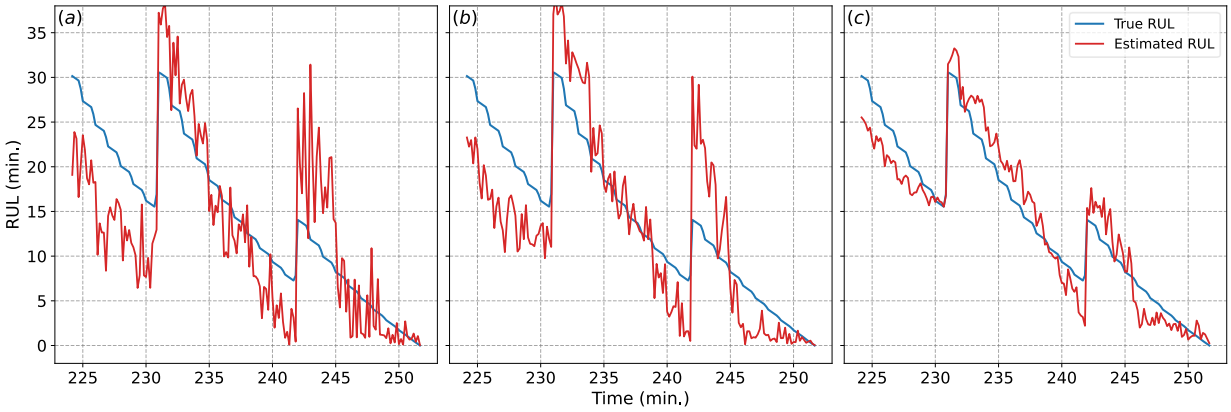

Figure 14. Predicted RUL of the bearing A156, (a) CNN, (b) DANN, (c) Proposed model 3

5. CONCLUSION

The objective of this study is to create a digital twin based on the simple physical knowledge of the fault progression phenomena. Utilizing a phenomenological model to generate synthetic signals helps to have a big synthetic run-to-failure dataset under varying speed operating conditions for training the machine learning models. On the other hand, leveraging the phenomenological model and simulated signals enhances the cost-effectiveness of the proposed approach by minimizing the reliance on historical run-to-failure datasets. The proposed methodology shows how the synthetic dataset should be adapted while facing varying speed scenarios. Additionally, the proposed model facilitates the integration of supplementary information regarding the working conditions and sequential ordering of measurements in a deep-learning model for prognosis and demonstrates that using extra information in the architecture of the DANN model enables the model to gain knowledge about both the operating conditions and the dynamics of damage progression. Moreover, a few unlabeled measurements from the real dataset after anomaly are used for domain adaptation in an adversarial way to reduce the gap between the feature distribution of the real and simulated dataset. Encoding the extra information, despite the lack of physical meaning, can aid the network in distinguishing signals from different operating conditions and identifying their relative relationships. Experimental results on the SM dataset demonstrate that the proposed model achieves improved RUL estimation accuracy, particularly in scenarios involving abrupt speed changes, and delivers more reliable predictions. The estimated RUL can also react to the operating speed which is a must in prognosis and decision making. Thanks to the t-SNE technique, the model's ability to discriminate between different operating conditions has been validated. The flexibility of the proposed method in recognizing the speed influences on the amplitude of the signals makes it applicable to the various speed profiles including random profiles, and also different speed ranges, whether or not they cross the resonance frequency of the structure. Experimental results have shown the capability of the proposed method compared to the models that do not utilize encodings.

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

Ai, T., Liu, Z., Zhang, J., Liu, H., Jin, Y., & Zuo, M. (2023). Fully Simulated-Data-Driven Transfer-Learning Method for Rolling-Bearing-Fault Diagnosis. *IEEE Transactions on Instrumentation and Measurement*, *72*. https://doi.org/10.1109/TIM.2023.3301901

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

Arias Chao, M., Kulkarni, C., Goebel, K., & Fink, O. (2022). Fusing physics-based and deep learning models for prognostics. *Reliability Engineering and System Safety*, *217*. https://doi.org/10.1016/j.ress.2021.107961

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, J., Huang, R., Chen, Z., Mao, W., & Li, W. (2023). Transfer learning algorithms for bearing remaining useful life prediction: A comprehensive review from an industrial application perspective. *Mechanical Systems and Signal Processing*, *193*, 110239. https://doi.org/10.1016/j.ymssp.2023.110239

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

Cui, L., Wang, X., Wang, H., & Jiang, H. (2020). Remaining useful life prediction of rolling element bearings based on simulated performance degradation dictionary. *Mechanism and Machine Theory*, *153*, 103967. https://doi.org/10.1016/J.MECHMACHTHEORY.2020.103967

Deng, Y., Du, S., Wang, D., Shao, Y., & Huang, D. (2023). A Calibration-Based Hybrid Transfer Learning Framework for RUL Prediction of Rolling Bearing Across Different Machines. *IEEE Transactions on Instrumentation and Measurement*, *72*, 1–15. https://doi.org/10.1109/TIM.2023.3260283

Fink, O., Wang, Q., Svensén, M., Dersin, P., Lee, W. J., & Ducoffe, M. (2020). Potential, challenges and future directions for deep learning in prognostics and health management applications. *Engineering Applications of Artificial Intelligence*, *92*. https://doi.org/10.1016/j.engappai.2020.103678

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., & Lempitsky, V. (2016). *Domain-Adversarial Training of Neural Networks*.

Ghafari, S. H., Abdel-Rahman, E. M., Golnaraghi, F., & Ismail, F. (2010). Vibrations of balanced fault-free ball bearings. *Journal of Sound and Vibration*, *329*(9), 1332–1347. https://doi.org/10.1016/j.jsv.2009.11.003

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

Harsha, S. P., Sandeep, K., & Prakash, R. (2003). The effect of speed of balanced rotor on nonlinear vibrations associated with ball bearings. *International Journal of Mechanical Sciences*, *45*(4), 725–740. https://doi.org/10.1016/S0020-7403(03)00064-X

Hosseinli, S. A., Ooijevaar, T., & Gryllias, K. (2023). Context-aware machine learning for estimating the remaining useful life of bearings under varying speed operating conditions. *Annual Conference of the PHM Society*, *15*(1). https://doi.org/10.36001/phmconf.2023.v15i1.3571

Jawad, S., & Jaber, A. (2022). Bearings Health Monitoring Based on Frequency-Domain Vibration Signals Analysis. *Engineering and Technology Journal*, *41*(1), 86–95. https://doi.org/10.30684/etj.2022.131581.1043

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

Li, N., Gebraeel, N., Lei, Y., Bian, L., & Si, X. (2019). Remaining useful life prediction of machinery under time-varying operating conditions based on a two-factor state-space model. *Reliability Engineering and System Safety*, *186*, 88–100. https://doi.org/10.1016/j.ress.2019.02.017

Liao, H., & Tian, Z. (2013). A framework for predicting the remaining useful life of a single unit under time-varying operating conditions. *IIE Transactions (Institute of Industrial Engineers)*, *45*(9), 964–980. https://doi.org/10.1080/0740817X.2012.705451

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

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

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

Tajiani, B., & Vatn, J. (2023). Adaptive remaining useful life prediction framework with stochastic failure threshold for experimental bearings with different lifetimes under contaminated condition. *International Journal of System Assurance Engineering and Management*, *14*(5), 1756–1777. https://doi.org/10.1007/s13198-023-01979-0

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

Wang, H., Liao, H., & Ma, X. (2021). Remaining Useful Life Prediction Considering Joint Dependency of Degradation Rate and Variation on Time-Varying Operating Conditions. *IEEE Transactions on Reliability*, *70*(2), 761–774. https://doi.org/10.1109/TR.2020.3002262

Zhang, Z., Chen, X., & Zio, E. (2022). A framework for predicting the remaining useful life of machinery working under time-varying operational conditions[Formula presented]. *Applied Soft Computing*, *126*. https://doi.org/10.1016/j.asoc.2022.109164

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 Ph.D. degree in the 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 system.