A Gear Health Indicator Based on f-AnoGAN

Hao Wen[1,2], Djordy Van Maele[3,4], Jean Carlos Poletto[3,4,5], Patrick De Baets[3,4] and Konstantinos Gryllias[1,2]

[1] *KU Leuven, Department of Mechanical Engineering, Celestijnenlaan 300, 3001, Leuven, Belgium*
hao.wen@kuleuven.be
konstantinos.gryllias@kuleuven.be

[2] *Flanders Make @ KU Leuven, Leuven, Belgium*

[3] *Ghent University, Soete Laboratory, Technologiepark Zwijnaarde 46, 9052 Zwijnaarde, Belgium*

[4] *Flanders Make @ UGent - Core Lab MIRO, Ghent, Belgium*

[5] *Federal University of Rio Grande do Sul, Laboratory of Tribology, Osvaldo Aranha, 99, 90035-190, Porte Alegre, Brazil*

Abstract

The development of high-quality health indicators based on Artificial Intelligence (AI) for condition monitoring, reflecting the degradation process and trend, remains a key area of research. Unsupervised deep learning methods, such as deep autoencoders and variational autoencoders, are often employed to establish health indicators for rotating machinery. However, commonly used methods frequently face challenges in controlling and evaluating the quality of learned features that represent this distribution, which subsequently impacts the accuracy of the test data analysis and anomaly detection. Additionally, the empirical nature of threshold setting adds an element of uncertainty to detections.

The research propose a novel approach for constructing gear health indicators and performing anomaly detection using Generative Adversarial Networks (GAN), with a particular emphasis on the f-AnoGAN structure. The research focuses on modeling the distribution of vibration signals acquired from healthy systems using adversarial learning. By comparing test samples against this modeled distribution, the degree of similarity or dissimilarity acts as an indicator of anomalies. Owing to the generative process of the GAN architecture (creating data from randomly sampled low-dimensional noise), GAN-based modeling overcomes the limitation of autoencoders by aiming to reconstruct the continuous distribution of systems in healthy conditions from a limited set of healthy (training) samples. In this way, it offers more generalizability than traditional model learning. Moreover, this study proposes a new method for establishing thresholds based on distribution fitting by the anomaly score of healthy data. The proposed f-AnoGAN-based model and thresholding technique is applied, tested and evaluated in a gear-pitting degradation dataset and result in more accurate and timely fault detection, markedly enhancing the ability to identify subtle faults in systems over traditional methods.

1. Introduction

Gears are an indispensable element of rotating machinery, widely employed across industry, including aerospace, rail transport, and industrial sectors (Chen, Jiang, Ding, & Huang, 2022; Salameh, Cauet, Etien, Sakout, & Rambault, 2018). The malfunctioning of gears constitutes a prevalent reason for the failure of machine systems, which can result in substantial economic losses and may even pose risks to human safety (Lee et al., 2014). Consequently, monitoring gear conditions and accurately predicting component failure and fault progression are crucial.

The employment of vibration based condition monitoring at both system and component levels represents a universally endorsed technique within the realm of health monitoring for rotating machinery (Elasha et al., 2014; Teng, Wang, Zhang, Liu, & Ding, 2014; Öztürk, Sabuncu, & Yesilyurt, 2008). The meticulous measurement and subsequent analysis of vibration signals are instrumental in the precise identification of incipient faults, thereby enabling the implementation of preventative and predictive maintenance prior to deterioration and corresponding issues. This proactive approach significantly contributes to the sustenance of system reliability and safety. Moreover, vibration analysis serves as an invaluable source of insight regarding the mechanical condition, since deviations in pivotal rotating elements, such as gears, manifest within the vibration signals (Zhu, Mousmoulis, & Gryllias, 2023; Hendriks, Dumond, & Knox, 2022). The utilization of signal processing tools for the examination of vibratory data aids in the extraction of critical information and indicators spanning both frequency and time-frequency domains. Nevertheless, it is imperative to acknowledge that the interpretations derived from these signal processing outcomes frequently require the expertise of seasoned operators.

With the advancement of artificial intelligence, its application in gear fault detection has gained increasing attention. Artificial intelligence, especially machine learning and deep learning methods, can process and analyze vast amounts of data, uncovering complex patterns and relationships that may be elusive to human experts. This reduces reliance on deep expert knowledge while enhancing the efficiency and accuracy of fault detection, enabling even non-experts to effectively diagnose faults. Among the various techniques, Convolutional Neural Networks (CNN) have demonstrated their versatility in state monitoring applications, including the detection and diagnosis of gear pitting faults (Zhang, Liu, Wang, & Gu, 2022; Xiang, Yang, Hu, Su, & Wang, 2022; Shi et al., 2022; Kim, Na, & Youn, 2022).

Viewing fault detection as a classification problem is a widely adopted strategy. However, obtaining clean, ample, and balanced healthy and especially faulty data, is challenging. Thus, various unsupervised one-class classification methods have been introduced. Unsupervised training methods, which infer based solely on information from healthy data, are limited by their output being the probability of a sample being normal. Thus for detection in a continuous progress, such as a degradation, this type of methods lacks of ability to represent trend. These methods primarily involve two steps: firstly, through the neural network's learning, mastering the distribution of healthy data and gauging the deviation of test data from this baseline; secondly, establishing reasonable thresholds for anomaly detection. A popular method is Deep Support Vector Data Description (DSVDD) (Ruff et al., 2018; Liu & Gryllias, 2020; Peng, Liu, Desmet, & Gryllias, 2023), which uses the Euclidean distance between hidden layer feature representations to characterize the extent of faults, allowing for trend assessment. However, DSVDD faces limitations in feature space representation capability and a lack of control over hidden layers/features.

An alternative unsupervised learning approach involves self-supervised schemes like Autoencoder (AE). By encoding and decoding complete data through neural networks, these models learn the intrinsic structure of the data (C. Zhou & Paffenroth, 2017; Ren, Sun, Cui, & Zhang, 2018; Mao, Feng, Liu, Zhang, & Liang, 2021). Yet, the characteristic of data compression in autoencoders limits their generalization ability, showing a significant dependency on the training data.

Recent years have seen generative models emerge as a new research focus. From the perspective of mechanical fault detection, Generative Adversarial Networks (GANs) (Goodfellow et al., 2014) have been applied primarily in data augmentation tasks (He, Tian, & Zuo, 2022; K. Zhou, Diehl, & Tang, 2023; Qin, Wang, & Xi, 2022; Wang et al., 2019). (Ding, Ma, Ma, Wang, & Lu, 2019) proposed a GAN-based anomaly detection method for bearing fault diagnosis, where the discriminator is used as an anomaly detector. (Dai, Wang, Huang, Shi, & Zhu, 2020) introduces an adversarial learning strategy to optimise the training of autoencoder(the method is also known as adversarial autoencoder) for the establishment of rotating machinery health indicators.

However, the essence of GANs lies in their use of adversarial learning to better fit the distribution of training data, allowing the direct generation of new data from this distribution. This aligns with the upstream task of various anomaly detection algorithms, which is to simulate the distribution of training data.

This study explores the potential of Generative Adversarial Networks in the task of anomaly detection for rotating machinery, based on vibration signals. It proposes a scheme for constructing a gear health indicator using GANs, along with a corresponding threshold setting and an anomaly detection system, aimed at detecting pitting initiation as early as possible. The methodology is validated on a dataset from a gear-accelerated degradation test.

The rest of the paper is organised as follows. In Section 2, the proposed anomaly detection methodology including model training, construction of health indices, and the threshold setting scheme is presented in detail. Then in Section 3 the experimental set up is described, the proposed methodology is applied on the experimental dataset and its effectiveness is analysed. The paper closes at the final section with some conclusions and the potentials of the proposed method in the field of rotating machinery health monitoring.

2. Proposed Method

The proposed detection scheme can be divided into three independent steps:

1. **Offline Distribution Learning**, by generative adversarial learning. In this step, the model is trained only by a limited number of partitioned healthy signals. The generator uses low-dimensional random noise as input and upscales it to the same dimension as the actual samples.

The objective of the generator is to produce signals from the random noise (i.e., feature space) that are as realistic as possible. This process can be considered as the model's grasp of the intrinsic structure of the training signals.

2. **Health Indicator Formation**, by the fast AnoGAN (f-AnoGAN) structure. The well-trained generator from step 1 can upscale any arbitrary low-dimensional feature set to obtain sufficiently realistic signals. This result can be interpreted as having obtained a continuous, infinite set of training samples. Therefore, for a test sample, its state related with health can be determined by whether an identical sample (or, as similar as possible) can be found within this continuous healthy set. The process of finding the corresponding sample, according to the f-AnoGAN structure, is assisted by an independently trained encoder working alongside the generator. After obtaining the corresponding sample for the test signal, the Euclidean distance is measured between signals to gauge the test sample.
3. **Fault Detection**, by a thresholding method. The discrepancy measured as outlined in step 2 is compared against a pre-determined threshold. Samples exceeding this threshold are flagged as potential anomalies, indicating a departure from the healthy signal distribution and hence, identifying possible faults.

2.1. Generative Adversarial Network (GAN) Training

2.1.1. Training Strategy

The training of GAN, depicted in Figure 1, alternates between updating the discriminator (also referred to as the Critic in the context of WGANs) and the generator. The discriminator (Critic model)'s task is to evaluate the realism of both real and generated samples, while the generator aims to produce data that are indistinguishable from real data. The key innovation of WGAN-GP (Gulrajani, Ahmed, Arjovsky, Dumoulin, & Courville, 2017) lies in the gradient penalty term, which enforces a soft version of the Lipschitz constraint by penalizing the gradient norm of the Critic's output with respect to its input.

2.1.2. Loss Composition

Generator Loss: The generator's objective is to minimize the negative average score of the generated samples evaluated by the discriminator:

$$L_G(\theta_G) = -\mathbb{E}_{\tilde{x}\sim P_g}[C(\tilde{x})] \tag{1}$$

where θ_G represents the generator's parameters. The generator is trained to produce samples $\tilde{x}$ that maximize the discriminator's (critic's) score $C(\tilde{x})$, pushing it towards generating more realistic samples.

Discriminator Loss: It includes two components - the average score for the real samples and the average score for the generated (fake) samples.

The objective of the GAN's training can be expressed as:

$$\min_{\theta_G} \max_{\theta_C \in \mathcal{C}} \mathbb{E}_{x\sim P_r}[C(x)] - \mathbb{E}_{\tilde{x}\sim P_g}[C(\tilde{x})] \tag{2}$$

where θ_C represents the critic's parameters. The goal is to train the critic to assign higher scores to real samples $x \sim P_r$ and lower scores to generated samples $\tilde{x} \sim P_g$.

However, the optimizing objective (2) is still not effective enough in the practice of GAN training, and researchers are often plagued by pattern collapse, which has spawned more related studies. Among them, the study of (Gulrajani et al., 2017) has attracted attention by introducing the gradient penalty:

Gradient Penalty (GP): Is calculated by first interpolating between real and fake samples, and then computing the gradient of the critic's scores with respect to these interpolated samples. The penalty is the squared deviation of the gradient norm from 1, averaged across the batch. The final loss function is as follows:

$$\begin{aligned} L(\theta_C) = &\mathbb{E}_{x\sim P_r}[C(x)] - \mathbb{E}_{\tilde{x}\sim P_g}[C(\tilde{x})] + \\ &\lambda \mathbb{E}_{\hat{x}\sim P_{\hat{x}}}\left[(\|\nabla_{\hat{x}} C(\hat{x})\|_2 - 1)^2\right] \end{aligned} \tag{3}$$

where $\hat{x}$ is sampled uniformly along straight lines between pairs of real and generated samples, and λ is a hyperparameter that controls the strength of the penalty, which is set as default value 10 to ensure Critic's gradient comply with the Lipschitz constraint.

This strategy encourages the generator to produce samples that are realistic enough to receive high scores from the discriminator, while the discriminator is penalized for having a gradient norm far from 1, ensuring that it behaves like a smooth function (Gulrajani et al., 2017) that provides useful gradients to the generator throughout the training process.

2.2. Indicator Formation

As previously mentioned, following the training of the GAN, the generator can now represent the complete and continuous distribution of healthy samples found within the training set. Subsequently, the difference between the test signal and the healthy distribution need to be measured to quantify the degree of anomaly in a new signal.

However, this learned distribution is implicit, which means that it is impossible to explicitly write out the mathematical form of this learned data distribution. In the final implementation of AnoGAN (Schlegl, Seeböck, Waldstein, Schmidt-Erfurth, & Langs, 2017), this process is simplified to whether a similar signal can be sampled from the distribution of the

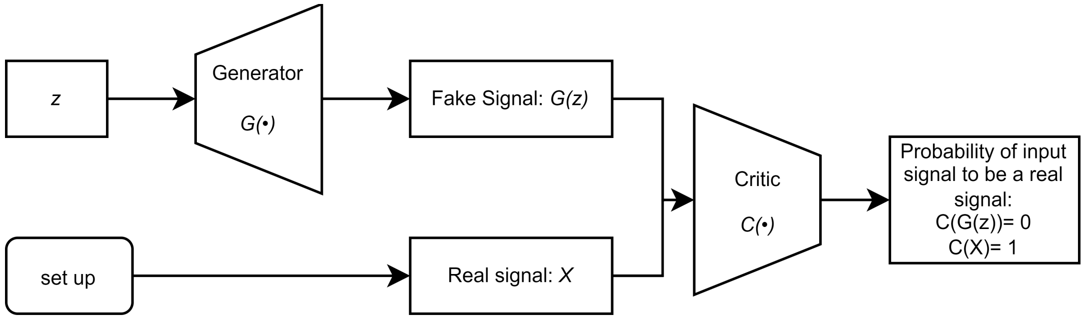


Figure 1. WGAN-GP Training Strategy

generated signal (D_{gen}), that is, finding z in the noise space (the distribution of Z, D_Z, is normally defined as a Gaussian distribution). This process is further reduced iteratively using back-propagation for a substantial number of iterations, such as 10,000 iterations, after which the final sampling result is considered the closest match to the test sample in D_{gen}.

However, the drawbacks of this process are evident; iterating multiple times for a single sample is computationally inefficient, especially when considering practical downstream applications. Moreover, using gradient descent optimization in isolation carries a significant risk of the sampling being trapped in local minima, which can adversely affect the quality of signal sampling in D_{gen}.

To enhance efficiency, the f-AnoGAN introduces an independent encoder for the sampling process. In GAN models, reliable mapping from D_z to D_{gen} is established. The aim of the independent encoder in f-AnoGAN is to facilitate the reverse process: mapping from the complex data distribution D_{gen} back to the simple feature space D_z. This process aids in quickly searching feature vectors z that match the new test sample best, enhancing both accuracy and efficiency in anomaly detection tasks. The obtained vector z is used to generate the corresponding health data $x_{gen} = G(z)$. The generated signal(x_{gen}) is then considered as the generated health signal closest to the tested signal to complete the corresponding indicator calculation.

To train the model, the encoder takes the generated signal X_{gen} as input and Z as output to train the parameters. The formation of the encoder can be depicted in Figure 2.

The loss of the training process of the Encoder is defined as:

$$\begin{aligned}\text{Loss} &= \text{Loss}_{\text{szs}} + \text{Loss}_f \\ &= \text{MSE}(X_{\text{gen}} - G(E(X_{\text{gen}}))) \\ &\quad + \text{MSE}(C(X_{\text{gen}}) - C(G(E(X_{\text{gen}})))) \end{aligned} \tag{4}$$

As mentioned earlier, the detection relies on the discrepancy between the test data and the generated healthy data. The discrepancies in this work are defined as two independent parts:

1. the Euclidean distance in the signal space
2. the Euclidean distance in the feature space, defined by the Critic

The Health Indicator (HI): the Anomaly Score (AS) is defined as the weighted sum of these two distances. In this research, this weighting parameter is not discussed emphatically and both distances are considered equally important, thus, for a tested data x, AS can be expressed as follows:

$$AS = \|x - G(E(x))\| + \|C(x) - C(G(E(x)))\| \tag{5}$$

2.3. Detection Part - Thresholding

The described method evaluates any signal to obtain a unique quantified indicator AS. For anomaly detection tasks, it is necessary to set a threshold based on the AS collection of the given healthy samples. The aforementioned method is applied to evaluate the healthy signals in the validation set to obtain the AS. Then, for the resulting Set_{AS}, the maximum likelihood estimation is used to estimate the parameters according to the assumed distribution type. In this step, research first establishes a distribution bank that includes all common and interesting distributions. Afterward, for Set_{AS}, different distribution estimates can be obtained, Dis_1, Dis_2, etc., along with the estimated distribution parameters. The AIC is taken as the matrix to compare and evaluate different distributions, and to select the optimal distribution based on this comparison as the parametric expression of the AS collection. The resulting dis is the distribution expression of the healthy signals. This distribution is inferred based on actual vibration measurements, and given the potential instability of operating conditions in the actual experimental process, and various interferences in signal measurements (such as the electrical environment), there are outliers in both the AS collection and the estimated distribution. This is also why similar studies do not use the maximum value of the validation's AS as the threshold for judging anomalies. In this method, the threshold setting is based on the estimation of the threshold according to the Distribution's Cumulative distribution function (CDF). In this paper, the AS corresponding to $CDF(\text{AS}) = 0.99$ is

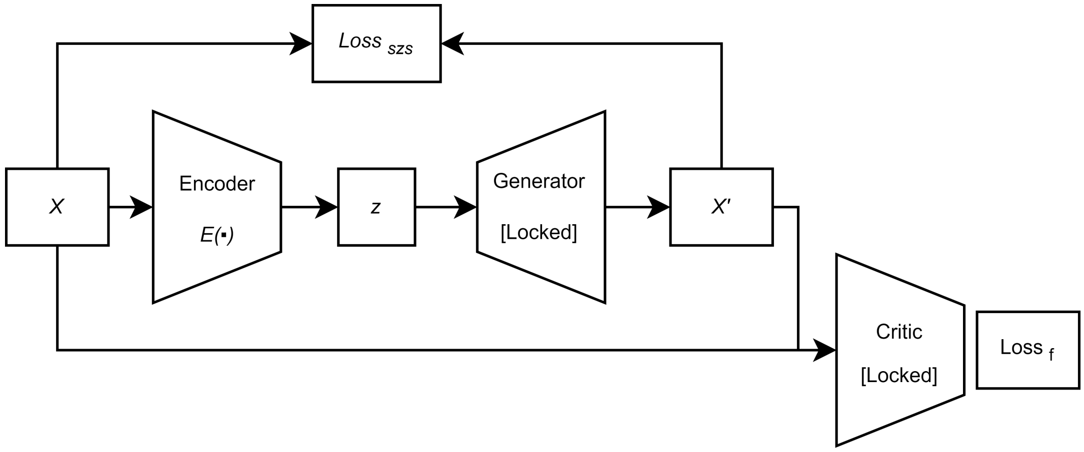


Figure 2. Encoder formation in f-AnoGAN

used as the threshold for judgment of anomaly.

3. Application of the methodology and Results

3.1. Description of the data

The data used for validating the anomaly detection approach in this research were derived from a comprehensive gearbox degradation test (Van Maele et al., 2023).The measurements were conducted on an FZG multi-stage gearbox test rig (Figure 3), where the input and output of two gearboxes were mechanically interconnected, thus forming a mechanical closed loop. The load was provided by a friction disk coupling mechanism situated between the gearboxes, which applies torque to the gear meshing system through angular displacement between two discs at either end. Throughout the test, the torque, applied manually, was maintained at 60-90Nm, and the gear under test was set to a speed of 2560 rpm.

Figure 3. Photo of the multistage FZG test rig

Within the gearbox under investigation, the transmission system consists of three pairs of meshing gears, with their specific locations indicated in Figure 4. The test employed two pairs of helical gears made of 20MnCr5 steel, featuring 41 (monitored) and 38 teeth, respectively. Unlike standard, industrial gears, the gears under observation were not hardened (250HV) to ensure pitting would occur on the monitored surfaces within a reasonable time frame (Van Maele et al., 2023). A camera was used during the operation to periodically record the visual information of the gear surfaces at fixed intervals for the study and quantification of surface pitting. Specifically, the system was slowed down to 1rpm for image capture every 30 minutes during operation. The camera system took five shots of each meshing surface during the collection process, and the sharpest image was algorithmically selected to serve as the basis for quantifying the pitting area. Thus, after the test concluded, a quantitative description of the process of surface pitting area evolution over time was obtained, providing a metric for the degradation process.

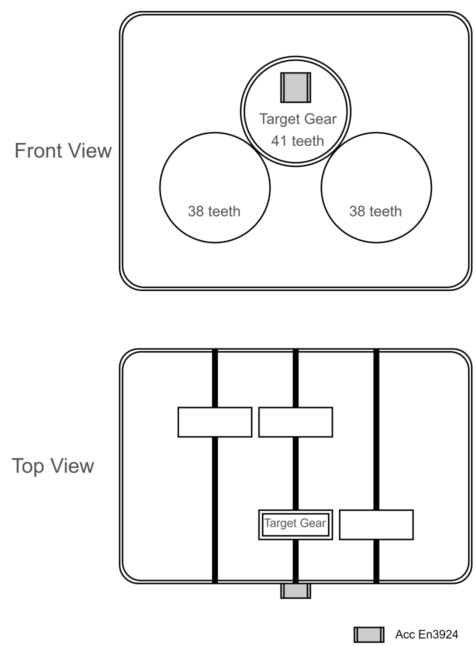


Figure 4. Sketch of the test gearbox setup

In addition to the visual information, during the test also torque, speed, and most importantly, vibration signals were measured. The locations of the vibration sensors are indicated in the accompanying diagram. Vibration signals were sampled at 10-minute intervals, with each sample collected at a sampling rate of 25600Hz over a duration of 10 seconds. In total, this accelerated degradation test lasted approximately 205 hours.

Overall, the test yielded 999 valid vibration data entries (indicating failure after 999 cycles) and 300 sets of synchronized gear surface information. According to tribologists' analysis, significant and observable pitting occurred on all gear surfaces after the 33 cycle.Figure 5 is a depict of the observed degradation process of pitting.

In this study, the vibration signal in the X-direction from the vibration sensor EN-3924, which lies closest to the tested gear, was utilized. A health indicator built on a generative model was used to track the degradation of gears, aiming for earlier detection of pitting formation, which would, in turn, guide maintenance activities.

3.2. Data preprocessing

In this research, the training data were prepared with the following preprocessing steps to use more informative samples. It is postulated that the degradation features are primarily concentrated in the gear mesh frequency (GMF), its harmonics, and the features and structures of the sidebands. In this test, with the input shaft rotating at 2560 rpm, the velocity of the intermediate and target gears was calculated as $v_2 = v_1 \times \text{gear\_ratio} = 39.5\text{Hz}$, and the gear mesh frequency was $\text{GMF} = \text{Teeth\_num} \times v_2 = 1622\text{Hz}$. Table 1 contains the characteristic frequencies of the test rig and the test.

Table 1. Characteristic frequencies of the test gearbox

Speed (Input Shaft)	2560 rpm
Speed (Driver Gear)	42.7 Hz
Speed (Target Gear)	39.5 Hz
GMF (Target Gear)	1622.6 Hz
2*GMF (Target Gear)	3245.2 Hz
3*GMF (Target Gear)	4867.8 Hz

Accordingly, the following preprocessing was applied to the data: initially, vibration signals were passed through a filter targeting the 1500-5000Hz frequency band, which includes the harmonics from the 1st to the 3rd order of the gear mesh frequency, along with the related band components. Afterward, the Discrete Fourier Transform was applied to isolate the informative frequency band, and then the 1500-5000Hz range was extracted to form the training, validation and test set samples.

The complete experimental dataset consists of approximately 999 independent measurements covering the full lifecycle. For this study, the early-life gear signals are selected as training samples to ensure that the training set comprised entirely healthy data to support model building and parameter optimization. A portion of the healthy dataset was also reserved as a validation set due to the encoder architecture of f-AnoGAN. In f-AnoGAN, unlike the original AnoGAN structure that relies solely on random sampling and gradient descent for sampling in D_{gen}, the model treats the training data as input to build the latent feature z via the encoder. Hence, to establish a threshold for anomaly detection, the new, unseen healthy data is still required as a reference.

Figure 6 delineates the division of the dataset in the test. Due to the run-in and gear bedding-in phases, which led to an unstable operational state of the experimental system, the initial two signals were discarded. The complete training set is composed of 19 independent signals, each sliced into time series of 51200 points with a 50% overlap during segmentation. All models in the study, including the generator, discriminator, and encoder, were trained exclusively with the aforementioned samples as per the described method. Furthermore, following the aforementioned method, 9 independent measurements from the healthy system were retained as a validation set, with the data division and sample generation being identical to the training set. All remaining data, encompassing both healthy and anomalous readings, were used as the final test set. It is important to note that the exclusion of data, as well as the delineation of the training, validation, and test sets, was conducted in chronological order following the accelerated degradation life course. In other words, the training and validation sets represent the early service life of the gears, while the test set includes the entire progression from a healthy state through the onset and development of pitting. Figure 7 illustrates the data pre-processing process.

3.3. Results

To verify the methodology's effectiveness, this study sets up comparative experiments and discusses the performance of the proposed model and the Autoencoder (AE). For the f-AnoGAN architecture, all models are set to be based on fully connected networks. To ensure fairness in the comparative experiments, the main model's structure and the number of parameters are kept as consistent and comparable as possible. This implies that both the generator of the proposed approach and the decoder of the AE, which serves as the benchmark method, undertake analogous functions by upscaling low-dimensional variables in the feature space to the target dimension. Consequently, to guarantee comparability between the two models, their parameters and network structures are configured to be identical. Both G and AE are completed by fully connected neural networks, transforming dimensions from $1000 \to 2500 \to 5000 \to 7001$, finally obtaining the spectrum from 1500-5000Hz (with a resolution of 0.5Hz). Based on these two models, model construction in

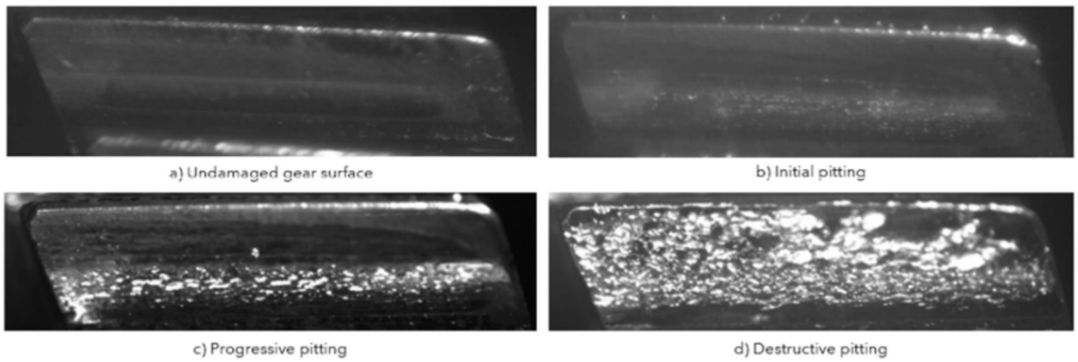


Figure 5. Observed degradation process of pitting

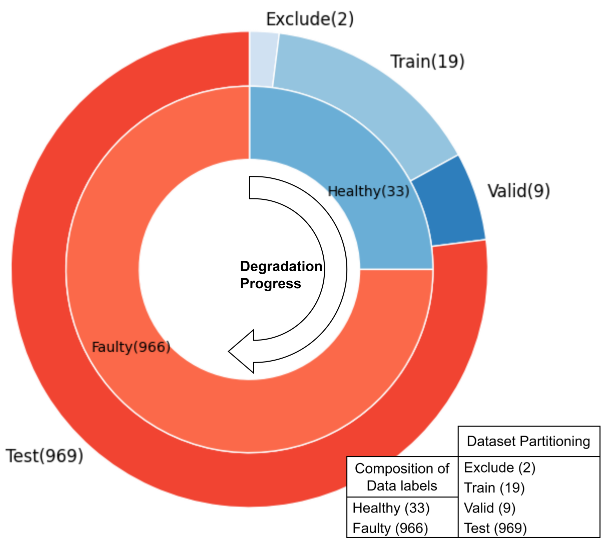


Figure 6. Segmentation of the training, validation, and test sets

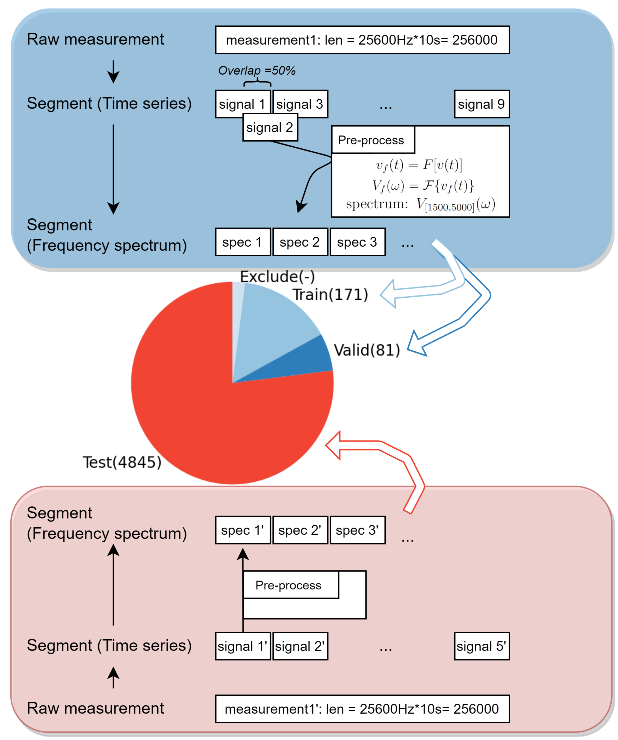


Figure 7. Dataset formation & pre-processing of signals, where $v(t)$ is the time-domain vibration signal, F is the filter, the Discrete Fourier Transform (DFT) of the filtered signal $v_f(t)$ is represented as $V_f(\omega)$.

each method is respectively completed, i.e., completing the Critic in GAN and implementing the encoder in f-AnoGAN. For the AE, the encoder is set according to the structure of the decoder. The AE uses the Mean Squared reconstruction Error (MSE) as its training loss and evaluates the health indicator during an assessment based on the MS reconstruction error.

For the proposed method, the training process generally follows the WGAN-GP (Wasserstein GAN with Gradient Penalty) scheme, setting the model to be trained for 2000 epochs (learning rate = 0.0001) to allow the model parameters to converge. After complete training, the generator can produce specified frequency bands based on any given set of features z. Figure 8 shows examples of training data and Figure 9 shows the generated results after 1990 epochs based on four randomly sampled z values. It is observed that the generated frequency bands closely mimic the features of the training data. From this, it can be inferred that the generator model has grasped the internal structure of the training data, that is, the representation of the service vibration condition of the given gear in the experimental system in the frequency domain. Specifically, the generated samples accurately replicate the gear mesh frequency and its higher-order harmonics, as well as the surrounding sideband performance.

According to the f-AnoGAN architecture, the construction of the Anomaly Score (AS) for any signal is completed with the help of the encoder. Figure 10 illustrates the result of generator sampling based on the aforementioned method and calculating the Euclidean distance in two spaces, with example

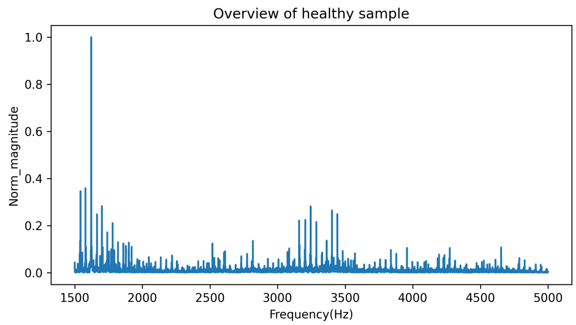


Figure 8. Sample overview in the training set

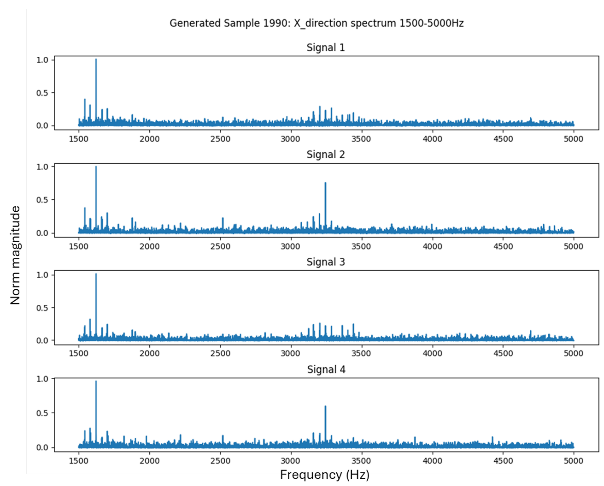


Figure 9. Generated spectrum at the end of the training period

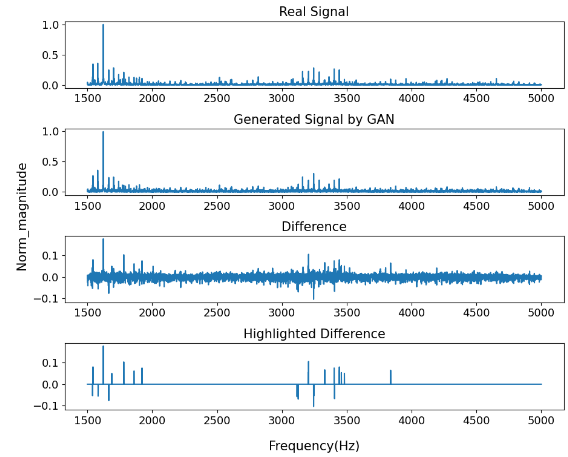


Figure 10. Comparison of the generated spectrum and the original spectrum

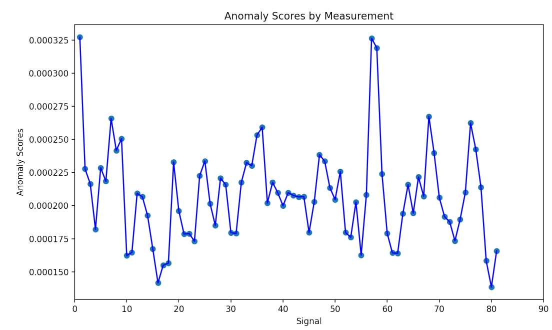


Figure 11. Anomaly score of the validation set (Healthy samples)

signals from the validation set.

Based on all the validation set data, a reference benchmark for anomaly detection, namely the Anomaly Score (AS) collection of healthy data, will be established. Figure 11 shows the AS for 81 samples (from 9 vibration signals) in the validation set 11. The study constructed a distribution bank using some common distributions. Following the aforementioned thresholding method, the fitting of the obtained distribution is as shown in Figure 12. The legend lists the distribution types in the figure, which are sorted in ascending order of the Akaike Information Criterion (AIC). Theoretically, a distribution with a smaller AIC value is closer to the true distribution. According to this criterion, the distribution among tested that best represents the AS of the healthy samples is the lognorm distribution (Figure 12). Additionally its numerical solution for various parameters is obtained, allowing to derive the CDF. Based on the aforementioned method where $CDF(AS_{\text{threshold}}) = 0.99$, the threshold is then determined for determining the anomaly (Figure 13). The threshold is then applied to the test set to evaluate the performance of the proposed method in anomaly detection.

Having completed all the components for anomaly detection as described previously, the Anomaly Score (AS) for each data in the test set is obtained following the aforementioned method. The average AS from the same vibration signal is taken as the AS for that vibration signal. Figure 14 shows the variation of AS across the entire test set in chronological order, along with the threshold performance. The results indicate that, according to the aforementioned method, the onset of failure occurs at the 33rd cycle (Figure 15), which aligns with the onset time of pitting derived from tribologists and visual information.

As a comparison, the AE model was also trained on the training set for 2000 epochs (learning rate = 0.0001), with an early-stopping at patience of 100. Figure 16 illustrates the reconstruction effect and schematic after completing the training.

As mentioned the reconstruction error derived from the AE model served as health indicators. In the comparative experiments, the threshold was established as the mean plus three times the standard deviation of the derived HI. AE-

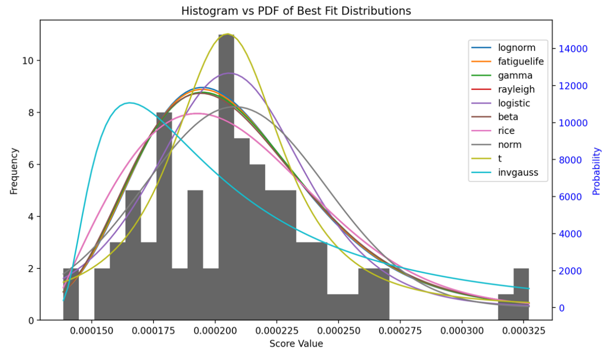


Figure 12. Distribution fit of the anomaly score in the Validation set

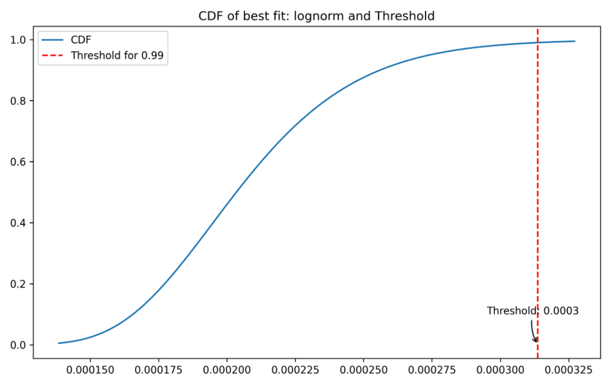


Figure 13. CDF according to the fitted distribution of the Anomaly Score of the healthy spectrum

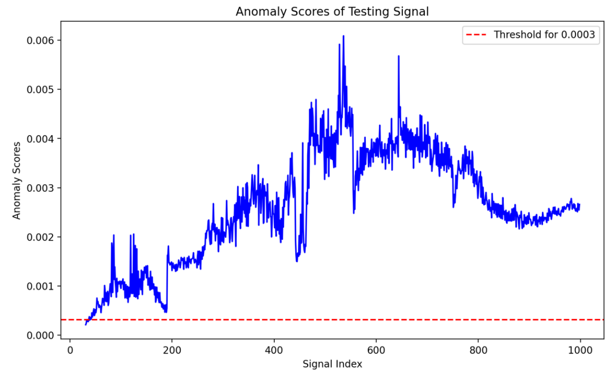


Figure 14. Detection result of f-AnoGAN-based anomaly detection method

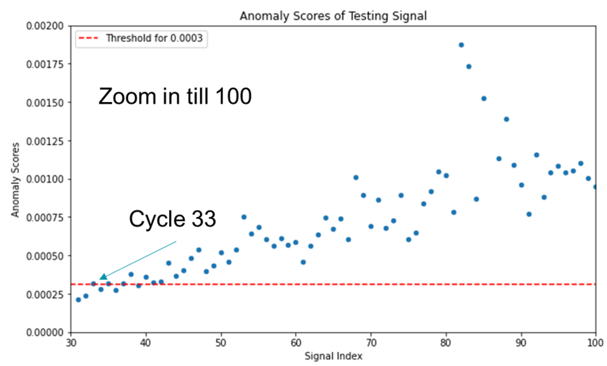


Figure 15. Anomaly Score in the first 100 cycles

based anomaly detection method identified the 38th cycle as the first cycle exceeding the threshold. However, this indicator was less stable, with a significant number of cycles between 38 and 200 falling below the threshold (Figure 18), indicating mis-detections if analogous to a classification problem.

3.4. Discussion

The results indicate that the proposed health indicator scheme and the thresholding method based on the GAN accurately detected the onset of the gear failure. Compared to the traditional unsupervised anomaly detection AE, the GAN-based detection advanced the detection time by 5 cycles. Given that the experiment conducted was an accelerated degradation test, and the gears underwent softening, this gap would be even more significant in actual industrial components.

Furthermore, it is also observed that the AE-reconstruction error, used as a HI, was highly unstable. One reason for this is the instability in the application of torque during the measurement campaign, which gradually diminishes during operation, necessitating the experiment to be halted and torque to be manually reapplied once excessive torque loss occurs. In the trend of HI obtained from AE, each sharp decrease in HI corresponds to the moments when the experiment is stopped and torque is reapplied. The same phenomenon is also observed in GAN-based HI. In GAN-based HI, even though HI is still established based on Euclidean distance, the powerful representation learning capability of the generator model allows it to construct more diverse samples that are more adaptable to certain instabilities in torque interference. Consequently, the differences reflected by HI are more attributable to degradation, with less impact from torque variations. This explains why GAN-based HI demonstrates better trend performance.

4. Conclusion

This work proposes an anomaly detection scheme for the condition monitoring of rotating machinery, focusing on gear fault detection using vibration signals. This method employs Generative Adversarial Networks (GANs) to learn the intrinsic structure and features of the training data's spectrum, particularly aiming to generate non-existent, highly realistic counterfeit samples. Based on the f-AnoGAN architecture, a health indicator is constructed utilizing the quantified Euclidean distance in two independent spaces. The study also employs a distribution fitting-based threshold method to assist in detection. The methodology is validated in a comprehensive

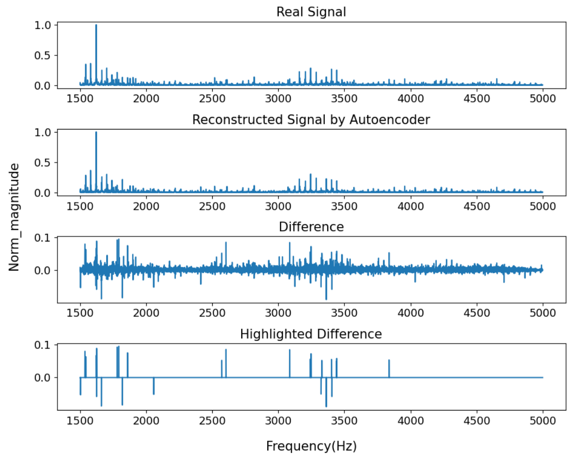


Figure 16. Comparison of the original spectrum and the reconstructed spectrum by an Autoencoder

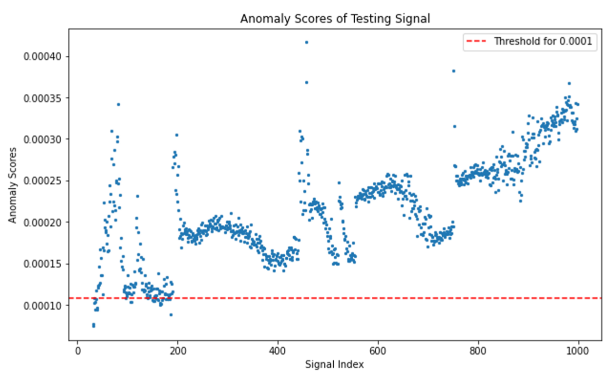


Figure 17. Detection result of Autoencoder-based anomaly detection method

gear accelerated degradation measurement campaign, which includes synchronized visual information collection, thus allowing for precise determination of the initial onset of pitting — the target of anomaly detection based on vibration signals in this research. In comparative experiments, the GAN-based method surpassed traditional unsupervised autoencoders and demonstrated better adaptability to changes in operating conditions, highlighting the performance of generative models with adversarial learning in the field of anomaly detection. Exploring how to better and more controllably utilize its adaptability under changing operating conditions will be the focus of future research.

Acknowledgment

This work was supported by Flanders Make, the strategic research center for the manufacturing industry, in the context of the QED project. Additionally K. Gryllias and Hao Wen would like to acknowledge the support of the FWO Fonds Wetenschappelijk Onderzoek – Vlaanderen in the frames of

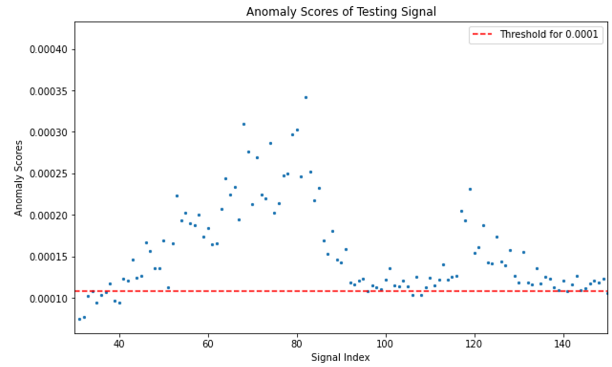


Figure 18. Autoencoder-based reconstruction error in first 200 cycles, first detection in cycle 38 and lots of misdetection in later cycles

GOA3123N project. The computing resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation - Flanders (FWO) and the Flemish Government.

References

Chen, H., Jiang, B., Ding, S. X., & Huang, B. (2022, March). Data-driven fault diagnosis for traction systems in high-speed trains: A survey, challenges, and perspectives. *IEEE Transactions on Intelligent Transportation Systems*, *23*(3), 1700–1716. doi: 10.1109/TITS.2020.3029946

Dai, J., Wang, J., Huang, W., Shi, J., & Zhu, Z. (2020, October). Machinery Health Monitoring Based on Unsupervised Feature Learning via Generative Adversarial Networks. *IEEE/ASME Transactions on Mechatronics*, *25*(5), 2252–2263. doi: 10.1109/TMECH.2020.3012179

Ding, Y., Ma, L., Ma, J., Wang, C., & Lu, C. (2019). A Generative Adversarial Network-Based Intelligent Fault Diagnosis Method for Rotating Machinery Under Small Sample Size Conditions. *IEEE Access*, *7*, 149736–149749. doi: 10.1109/ACCESS.2019.2947194

Elasha, F., Ruiz-Cárcel, C., Mba, D., Kiat, G., Nze, I., & Yebra, G. (2014, July). Pitting detection in worm gearboxes with vibration analysis. *Engineering Failure Analysis*, *42*, 366–376. doi: 10.1016/j.engfailanal.2014.04.028

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., … Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (Vol. 27). Curran Associates, Inc.

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. (2017, December). *Improved training of wasserstein gans* (No. arXiv:1704.00028). arXiv. doi: 10.48550/arXiv.1704.00028

He, R., Tian, Z., & Zuo, M. J. (2022, April). A semi-supervised gan method for rul prediction using failure and suspension histories. *Mechanical Systems and Signal Processing*, *168*, 108657. doi: 10.1016/j.ymssp.2021.108657

Hendriks, J., Dumond, P., & Knox, D. (2022, April). Towards better benchmarking using the CWRU bearing fault dataset. *Mechanical Systems and Signal Processing*, *169*, 108732. doi: 10.1016/j.ymssp.2021.108732

Kim, Y., Na, K., & Youn, B. D. (2022, March). A health-adaptive time-scale representation (htsr) embedded convolutional neural network for gearbox fault diagnostics. *Mechanical Systems and Signal Processing*, *167*, 108575. doi: 10.1016/j.ymssp.2021.108575

Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., & Siegel, D. (2014, January). Prognostics and health management design for rotary machinery systems—reviews, methodology and applications. *Mechanical Systems and Signal Processing*, *42*, 314–334. doi: 10.1016/j.ymssp.2013.06.004

Liu, C., & Gryllias, K. (2020, January). A semi-supervised support vector data description-based fault detection method for rolling element bearings based on cyclic spectral analysis. *Mechanical Systems and Signal Processing*, *140*. doi: 10.1016/j.ymssp.2020.106682

Mao, W., Feng, W., Liu, Y., Zhang, D., & Liang, X. (2021, March). A new deep auto-encoder method with fusing discriminant information for bearing fault diagnosis. *Mechanical Systems and Signal Processing*, *150*, 107233. doi: 10.1016/j.ymssp.2020.107233

Öztürk, H., Sabuncu, M., & Yesilyurt, I. (2008, April). Early Detection of Pitting Damage in Gears using Mean Frequency of Scalogram. *Journal of Vibration and Control*, *14*(4), 469–484. doi: 10.1177/1077546307080026

Peng, D., Liu, C., Desmet, W., & Gryllias, K. (2023, July). Condition monitoring of wind turbines based on anomaly detection using deep support vector data description. *Journal of Engineering for Gas Turbines and Power*, *145*(091009). doi: 10.1115/1.4062768

Qin, Y., Wang, Z., & Xi, D. (2022, January). Tree cyclegan with maximum diversity loss for image augmentation and its application into gear pitting detection. *Applied Soft Computing*, *114*, 108130. doi: 10.1016/j.asoc.2021.108130

Ren, L., Sun, Y., Cui, J., & Zhang, L. (2018, July). Bearing remaining useful life prediction based on deep autoencoder and deep neural networks. *Journal of Manufacturing Systems*, *48*, 71–77. doi: 10.1016/j.jmsy.2018.04.008

Ruff, L., Vandermeulen, R., Goernitz, N., Deecke, L., Siddiqui, S. A., Binder, A., … Kloft, M. (2018, July). Deep one-class classification. In *Proceedings of the 35th international conference on machine learning* (pp. 4393–4402). PMLR.

Salameh, J. P., Cauet, S., Etien, E., Sakout, A., & Rambault, L. (2018, October). Gearbox condition monitoring in wind turbines: A review. *Mechanical Systems and Signal Processing*, *111*, 251–264. doi: 10.1016/j.ymssp.2018.03.052

Schlegl, T., Seeböck, P., Waldstein, S. M., Schmidt-Erfurth, U., & Langs, G. (2017). Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In M. Niethammer et al. (Eds.), *Information processing in medical imaging* (pp. 146–157). Cham: Springer International Publishing.

Shi, J., Peng, D., Peng, Z., Zhang, Z., Goebel, K., & Wu, D. (2022, January). Planetary gearbox fault diagnosis using bidirectional-convolutional lstm networks. *Mechanical Systems and Signal Processing*, *162*, 107996. doi: 10.1016/j.ymssp.2021.107996

Teng, W., Wang, F., Zhang, K., Liu, Y., & Ding, X. (2014, January). Pitting Fault Detection of a Wind Turbine Gearbox Using Empirical Mode Decomposition. *Strojniški vestnik – Journal of Mechanical Engineering*, *60*(1), 12–20. doi: 10.5545/sv-jme.2013.1295

Van Maele, D., Poletto, J. C., Neis, P. D., Ferreira, N. F., Fauconnier, D., & De Baets, P. (2023). Online vision-assisted condition monitoring of gearboxes. In *8th european conference and exhibition on lubrication, maintenance and tribotechnology (lubmat 2023), proceedings.*

Wang, J., Li, S., Han, B., An, Z., Bao, H., & Ji, S. (2019). Generalization of Deep Neural Networks for Imbalanced Fault Classification of Machinery Using Generative Adversarial Networks. *IEEE Access*, *7*, 111168–111180. doi: 10.1109/ACCESS.2019.2924003

Xiang, L., Yang, X., Hu, A., Su, H., & Wang, P. (2022, January). Condition monitoring and anomaly detection of wind turbine based on cascaded and bidirectional deep learning networks. *Applied Energy*, *305*, 117925. doi: 10.1016/j.apenergy.2021.117925

Zhang, Y., Liu, W., Wang, X., & Gu, H. (2022, July). A novel wind turbine fault diagnosis method based on compressed sensing and dtl-cnn. *Renewable Energy*, *194*, 249–258. doi: 10.1016/j.renene.2022.05.085

Zhou, C., & Paffenroth, R. C. (2017, August). Anomaly detection with robust deep autoencoders. In *Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining* (pp. 665–674). Halifax NS Canada: ACM. doi: 10.1145/3097983.3098052

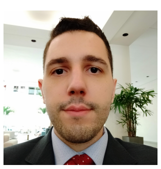
Zhou, K., Diehl, E., & Tang, J. (2023, February). Deep convolutional generative adversarial network with semi-supervised learning enabled physics elucidation for extended gear fault diagnosis under data limitations. *Mechanical Systems and Signal Processing*, *185*, 109772. doi: 10.1016/j.ymssp.2022.109772

Zhu, R., Mousmoulis, G., & Gryllias, K. (2023, July). Wavelet-based high order spectrum for local damage diagnosis of gears under different operating conditions. In *Surveillance, Vibrations, Shock and Noise.* Toulouse, France: Institut Supérieur de l'Aéronautique et de l'Espace [ISAE-SUPAERO].

BIOGRAPHIES

Hao Wen received his Bachelor of Science degree from the China University of Geosciences, China and his Master of Science degree in Mechanical Engineering from the University of Manchester, UK. He joined the Noise and Vibration Research Group in the Department of Mechanical Engineering at KU Leuven, Belgium, as a PhD researcher in 2022. His research interests lie in the areas of fault detection utilizing unsupervised learning and generative modeling.

Djordy Van Maele Received his M. Sc. in electromechanical engineering technologies from Ghent University, Gent, Belgium, in 2021. In 2021 he started working on his doctoral degree in electromechanical engineering, in the field of tribology, at Ghent University until current date.



Jean Carlos Poletto obtained his M.Sc. degree in Mechanical Engineering from the Federal University of Rio Grande do Sul (UFRGS), Brazil, in 2018. He has been developing research on experimental tribology, with active contributions to the field since 2015. Currently, he is working on a joint PhD program between UFRGS, Brazil, and Ghent University, Belgium.

Patrick De Baets received a master's degree in Electromechanical Engineering (1989) from Ghent University. His research, addressing the use of Thin Layer Activation to assess the fretting wear of machine components, resulted in a Ph.D. degree in 1995. He is now a full professor of machine elements and tribology. His research focusses on the tribological response (friction, friction stability and wear) in dry or marginally lubricated conditions of advanced materials such as fibre reinforced composites, high temperature ceramics and various self-lubricating materials. For his research, he has constructed different experimental tribological set-ups, ranging from the N to MN force range. The results of his research have been published in about 300 peer reviewed journal contributions and conference papers and numerous technical reports. Since more than 20 years Patrick De Baets is teaching specialised courses on Machine Design, Machine Elements and Tribology to Bachelor and Master students in Mechanical Engineering. Besides that, he has gained 15 years' experience in teaching general course on mechanical engineering to non-technical audiences, such as e.g. commercial engineering students.

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