

Damage Detection using Machine Learning for PHM in Gearbox Applications

Lisa Binanzer, Tobias Schmid, Lukas Merkle and Martin Dazer

Institute of Machine Components, University of Stuttgart, 70569 Stuttgart, Germany

lisa.binanzer@ima.uni-stuttgart.de

st157630@stud.uni-stuttgart.de

lukas.merkle@ima.uni-stuttgart.de

martin.dazer@ima.uni-stuttgart.de

ABSTRACT

Early damage detection in gearbox applications enables the implementation of Prognostics and Health Management (PHM). On the one hand, the earliest possible damage detection provides a precise in-sight into the state of health of a gearbox. In addition, early damage detection offers the possibility to slow down the damage progress and extend the remaining useful life (RUL) by intervening in the operating state at an early damage stage. The main contribution of this work is that existing Machine Learning tools are applied to the challenge of very early damage detection in gearboxes. Thus, the need for complex physically based data evaluation is avoided. The aim of this investigation is a comparison of two different machine learning approaches. To investigate the detection possibilities test bench experiments were conducted with a single stage spur gearbox. For a comprehensive investigation, i.e. to detect damage under different operating conditions, the test runs are carried out at different damage sizes, speeds and torques. Based on the recorded vibration data, the damage detection is examined. Two machine learning approaches of anomaly detection are considered: An encoding approach and a loss approach. The same sparse auto-encoder is developed for both approaches Both machine learning approaches are able to detect even the smallest damage of about 0.5 % in most operating states. The loss approach allows the different damage stages to be recognized much more clearly than the encoding approach. The comparison of the different approaches provides valuable insights for the further development of robust damage detection algorithms.

1. INTRODUCTION

In many mobile and stationary applications, gearboxes are essential for adjusting speed and torque. The greater the power

that needs to be transmitted, the larger and more expensive the corresponding gear units are. In gearboxes one of the most common types of damage on a tooth flank is pitting. As soon as a pitting exceeds a size of 4 % in relation to the size of the tooth flank, the gear is considered as failed according to the 2016 International Organization for Standardization [ISO] report. Damaged tooth flanks are one of the leading reasons of downtime and each failure can be associated with high repair costs and time-consuming repair work. This particularly applies to large gearboxes and applications in remote locations, such as offshore wind power drives. For this reason, gearboxes in critical industrial applications are often equipped with condition monitoring systems (CMS) based on vibration sensors. They continuously monitor the current state of health of the gearbox. If the CMS detects damage, depending on the damage extent, a complete shutdown or a load reduction can be initiated. Expensive subsequent damage can be prevented and, in case of a reduced load, the remaining useful life (RUL) of the gearbox can be extended until it is repaired or replaced. However, the CMS's are only developing their full potential if damage can be detected at a very early stage.

The earliest possible damage detection in gearboxes enables comprehensive Prognostics and Health Management (PHM) to be implemented in gearbox applications. According to Goebel, Celaya, Sankararaman, Roychoudhury, Daigle and Abhinav (2017), a PHM approach consists of the 5 sub-areas of the system: data, diagnosis, prognosis, optimization and the system itself. The earliest possible damage detection affects all of these areas.

First of all, the health of a system, which according to the 2017 Institute of Electrical and Electronics Engineers [IEEE] committee standards include all information regarding the functionality of a system, can be diagnosed much more precisely using the data of the system. The time gained by early damage detection can be used to acquire more data for a possible RUL prediction. Finally, health management of the system can be realized. Health management describes the control of damage according to the aim of the PHM solution

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

(Bertsche & Dazer, 2023). A potential goal is the optimal utilization of the RUL without unexpected failure until scheduled maintenance. For instance, this can be achieved by avoiding particularly damaging operating points. Another option is to implement an adaptive operating strategy that enables the extension of the RUL without any loss of performance (Gretzinger, Lucan, Stoll & Bertsche, 2020). Due to the application of the operating strategy, the load on the pre-damaged tooth is significantly reduced. This results in a slow-down of the damage progress. The other teeth on the circumference of the gear, which can still withstand the designated load, are being slightly overloaded. Thus, the load reduction is compensated without any overall power reduction. Control of the plant is carried out by a corresponding optimization algorithm.

Overall, comprehensive PHM in gearbox applications offers numerous benefits. However, damage detection at a very early stage is a prerequisite. The aim of this study is to investigate the earliest possible damage detection in gearboxes with the help of machine learning. Experiments were conducted on a test gearbox, vibration data was recorded and evaluated using an Autoencoder (AE).

2. EXPERIMENTS AND MACHINE LEARNING APPROACHES

Following, the test bench experiments for the earliest possible damage detection are described first. Subsequently, the developed autoencoder is presented. Finally, the two machine learning approaches (encoding and loss), which are based on the developed autoencoder, are discussed in more detail.

2.1. Test Bench Experiments

The test gearbox is designed as a single stage spur gear unit. Figure 1 illustrates the design of the test gearbox. The gear ratio is $i = 25/36 = 0.69$. More information on the test gears in (Binanzer, Merkle, Dazer & Nicola, 2023).

The test bench is set up as an inline concept (2 motor concept). The electric drive motor loads the transmission input side and the second electric motor loads the transmission output side. Torque measuring shafts and incremental encoders for speed and angular position measurement are mounted between the electric motors and the test gearbox. Further information on the test bench setup can be found in (Binanzer et al., 2023). The mounted test gearbox on the test bench is shown in figure 2.

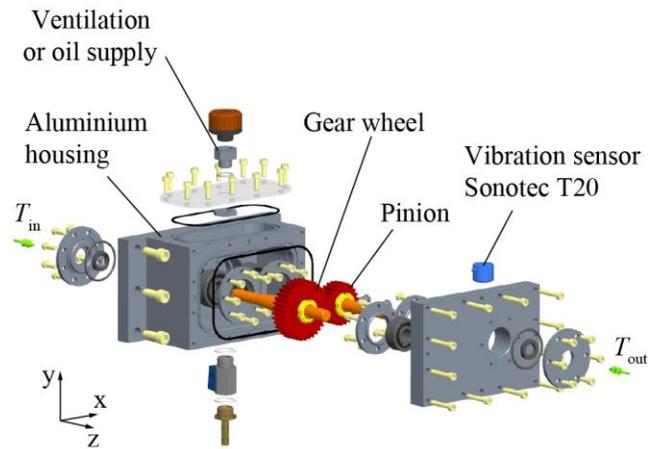


Figure 1. Design of the test gearbox.

For lubricating the tooth contact, FVA reference oil no. 3 (see the 1985 Research Association for Drive Technology [FVA] report) is used. This is a mineral oil without additives with a viscosity corresponding to ISO 3448 (see the report (ISO, 2010)). To ensure constant test conditions, the oil is preconditioned in an external fluid tempering device to 29.54 °C. A gear pump supplies the oil to the tooth contact. A Pt100 temperature sensor in the oil supply measures the oil temperature during the test runs.

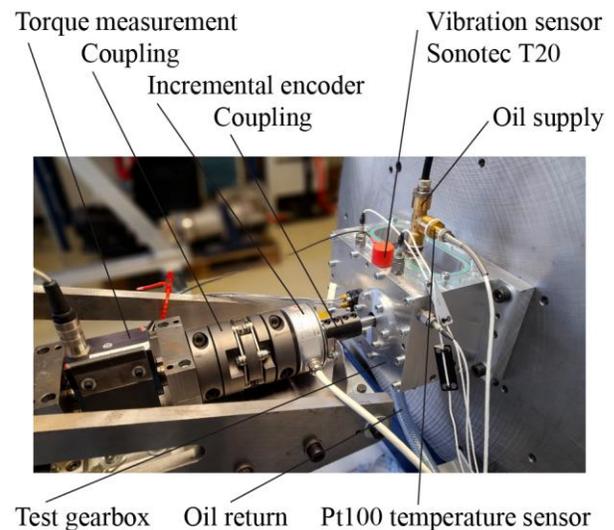


Figure 2. Test gearbox mounted.

The test gearbox is equipped with a Sonotec T20 sensor. This ultrasonic accelerometer is located between the bearings and measures in the y-direction. The maximum measuring frequency of the Sonotec T20 sensor is 100 kHz. However, the sampling rate of the Sonotec T20 sensor is limited to 96 kHz due to the maximum sampling rate of the data acquisition system that is used (PAK MK2). Thus, according to the Nyquist-

Shannon sampling theorem, maximum frequencies of 48 kHz can be measured with the system.

Since the earliest possible damage detection using machine learning algorithms is to be investigated as part of this work, damage well below the 4 % criterion is examined. Artificially generated pitting serves as representative gear damage. Following a test series without damage, a total of three damage sizes are tested - small (S), medium (M) and large (L). The tests without damage serve as a reference and thus as a training data set for the machine learning algorithms. The pitting damage is applied using a numerically controlled milling machine. This ensures that the pitting can be easily and reproducibly manufactured on the tooth flanks. A suitable radius milling cutter with a head diameter of 2 mm is used. Due to the higher number of load cycles, pitting damage usually occurs on the pinion. Consequently, the artificial damage is applied on the pinion. The gear wheel remains undamaged. The pitting is located in the center of a tooth flank below the pitch circle. Figure 3 shows the pinion with the manufactured pitting damage.

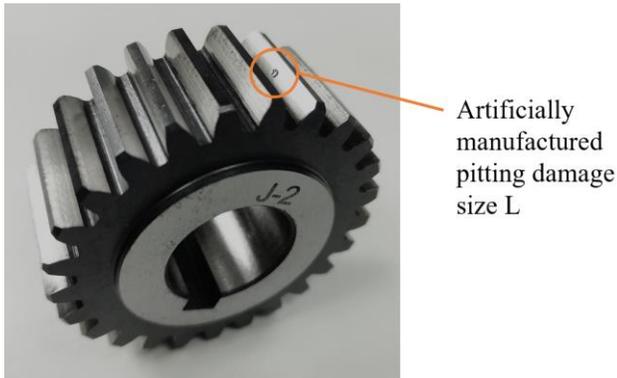


Figure 3. Pinion of second gear pair with manufactured pitting damage size L (1.72 %).

For a comprehensive investigation, i.e. to detect damage under different operating conditions, the test runs are carried out not only at different damage sizes, but also at different speeds and torques. In each of the four test series (no damage, S, M, L), six operating conditions are tested. These six operating states result from a combination of two speed levels (72 rpm, 636 rpm) and three torque levels (18 Nm, 24 Nm, 30 Nm). Within a test series, the six operating states are varied randomly in their sequence. The measurement duration for each test run is 100 s. Between the test series, the damage on the pinion is then artificially applied and milled larger. In this way, increasing damage on the pinion can be tested and the different damage sizes can be directly compared with each other. In order to achieve a more meaningful result, the tests are carried out with a total of three pairs of gears. The pinion and gear are always tested in the same pairs. This ensures that the algorithms do not detect any anomalies caused by manufacturing tolerances or material deviations of different pinion - gear combinations. The only difference between the test

series of a gear pair is the increasing damage on the pinion. The surface area of each pitting is measured after the milling process using a digital microscope (see table 1).

Table 1. Pitting surface areas.

Gear pair	Pitting level	Pitting surface in mm ²	Relative surface area in %
1	S	0.48	0.61
1	M	0.97	1.23
1	L	1.40	1.77
2	S	0.40	0.51
2	M	0.92	1.16
2	L	1.36	1.72
3	S	0.62	0.78
3	M	1.01	1.28
3	L	1.40	1.77

2.2. Autoencoder

A test run duration of 100 s and a sample rate of 96 kHz result in 9,600,000 data points per test (acceleration over time). These data points are then converted into the frequency spectrum. For this purpose, Fast Fourier Transforms (FFT) of 10,000 data points each are performed, thus 960 FFT's per test. Each FFT results in 5,001 frequency points. The two rotational speeds of the tests result in the following: At a speed of 72 rpm, approximately 11.6 FFTs are conducted for each revolution of the pinion. At a speed of 636 rpm, approximately 1.3 FFTs are generated for each revolution of the pinion.

The AE developed in this study emerged from literature research and empirical hyperparameter tuning. It is a multilayer AE, which consists of three joined AEs, each with a hidden layer, see figure 4. The number of units in layer l is defined as s_l . The input of the first AE used in this study contains $s_1 = 5,001$ units corresponding to the number of frequency points per FFT. Accordingly, layer 3 and 5 have $s_3 = s_5 = 5,001$ units. The first hidden layer (layer 2) has $s_2 = 560$ units, the second hidden layer (layer 4) has $s_4 = 200$ units. Finally, the third hidden layer (layer 6) learns a compressed representation of the frequency spectrum with $s_6 = 50$ features. The output after the last decoding process again has $s_7 = 5,001$ frequency points. The aim of the AE is to ensure that the output \hat{x} is the most accurate possible reproduction of the input x . For this purpose, the AE has to learn a function $h_{W,b}(x)$ with the parameters W and b , for which the following is valid:

$$h_{W,b}(x) \approx x \tag{1}$$

Since the AE in this paper has a total of 7 layers, layer 1 corresponds to the input and layer 7 to the output:

$$x = x^{(1)} \quad (2)$$

$$\hat{x} = h_{W,b}(x) = x^{(7)} \quad (3)$$

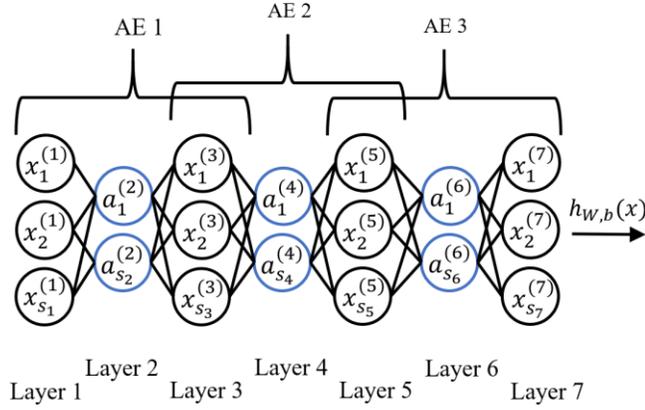


Figure 4. Structure of the Autoencoder.

For parameter W , the notation $W_{i,j}^{(l)}$ is used and this is associated with the weighting of the connection between unit j in layer l and unit i in layer $l + 1$ (Ng, 2011). Parameter $b_i^{(l)}$ is the bias associated with unit i in layer $l + 1$ (Ng, 2011). Eq. (4) and (5) are valid to the layers of the AE. For this, $a_{s_l}^{(l)}$ corresponds to the output of the hidden unit s_l in layer $l = 2, 4, 6$ and $x_{s_l}^{(l)}$ corresponds to the output of unit s_l in layer $l = 3, 5, 7$.

$$a_{s_l}^{(l)} = f \left(\sum_{i=1}^{s_l-1} W_{s_l,i}^{(l-1)} x_i^{(l-1)} + b_{s_l}^{(l-1)} \right) \quad (4)$$

$$x_{s_l}^{(l)} = f \left(\sum_{i=1}^{s_l-1} W_{s_l,i}^{(l-1)} a_i^{(l-1)} + b_{s_l}^{(l-1)} \right) \quad (5)$$

The function $f(z)$ with $f: \mathbb{R} \rightarrow \mathbb{R}$ is called activation function. The sigmoid function (Eq. (6)) is chosen as the activation function in this study. This function can assume values between 0 and 1, see figure 5.

$$f(z) = \frac{1}{1 + \exp(-z)} \quad (6)$$

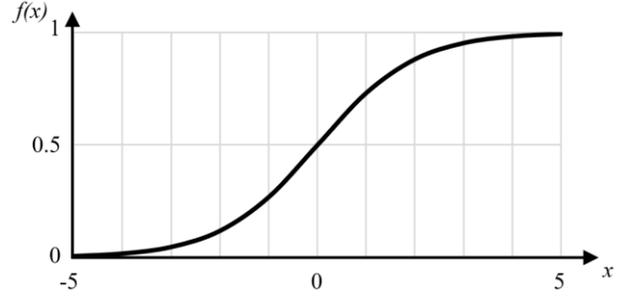


Figure 5. Sigmoid function.

The AE requires training with a training data set with m training examples. The data from the tests without damage serves as the training data set. Thus, the training data set consists of 960 FFTs each ($m = 960$). If training data set number m is used as input, then $x^{(l),(m)}$ results in layer l . The three AEs are trained one after the other. The corresponding loss function $J_{AE}(W, b)$ with $AE = 1, 2, 3$ is defined for training the AE for both approaches (encoding and loss) and consists of three terms:

$$\begin{aligned} J_{AE}(W, b) &= \left[\frac{1}{m} \sum_{i=1}^m \left(\|x^{(l+2),(i)} - x^{(l),(i)}\|^2 \right) \right] \\ &+ \lambda \sum_{l=1}^{l+1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(l)})^2 \\ &+ \beta \sum_{j=1}^{s_l} KL(\rho \| \hat{\rho}_j) \end{aligned} \quad (7)$$

with

$$l = 1 \text{ for } AE = 1;$$

$$l = 3 \text{ for } AE = 2;$$

$$l = 5 \text{ for } AE = 3$$

The first term of the loss function is the average sum-of-squares error term (Ng, 2011). It describes the deviation of \hat{x} from x and can therefore also be described as a reconstruction error. The second term is the regularization term, also known as the weight decay term, which tends to reduce the size of the weights W and helps to prevent overfitting (Ng, 2011). It therefore ensures that the network does not simply memorize the data, but learns the underlying structure. The weight decay parameter λ determines the relative importance of the second term. $\lambda = 0.001$ is selected.

Additionally, a sparsity constraint is imposed on the hidden units of layer 2, 4 and 6. Therefore a sparse AE is obtained and the third term of the loss function is the sparsity penalty term. With $\beta = 1$, the weighting of the term corresponds to a single weighting compared to the first term. A sparsity constraint permits the AE to learn the underlying structure in the data even with a large number of hidden units. For this

purpose, the average activation of unit j in the hidden layer l is calculated:

$$\hat{\rho}_j = \frac{1}{m} \sum_{i=1}^m [a_j^{(l)}(x^{(i)})] \quad (8)$$

This average activation of the neuron should match the selected sparsity constraint ρ :

$$\hat{\rho}_j = \rho = 0.3 \quad (9)$$

The sparsity penalty term penalized if $\hat{\rho}_j$ deviates significantly from ρ . The penalty term is based on the Kullback-Leibler (KL) divergence:

$$\begin{aligned} & \sum_{j=1}^{s_l} KL(\rho \parallel \hat{\rho}_j) \\ &= \sum_{j=1}^{s_l} \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j} \end{aligned} \quad (10)$$

The aim of training the AE is to minimize the function $J(W, b)$ as much as possible. $J(W, b)$ becomes as small as possible when an optimal combination of the parameters W and b is found. First, these parameters are randomly initialized, then backpropagation is applied. During backpropagation, the connections of the AE's units are either strengthened or weakened via the weightings to further minimize the loss between input and output. Each AE is trained for 100 iterations.

2.3. First evaluation approach: Encoding

After training the AE on the basis of the vibration data of a test without damage, the trained AE is used to evaluate all data of an individual operating condition of this gear pair. For this purpose, the recorded vibration data of the test without damage and the three tests with damage sizes S, M and L are appended to each other and an FFT is generated from 10,000 data points each. This results in a total of 3,840 FFTs. These are each encoded to 50 features using the trained AE. A Principal Component Analysis (PCA) is then used to determine a one-dimensional real number from these 50 features. According to the number of FFTs, 3,840 real numbers are obtained. The first 960 one-dimensional representations of the frequency spectrum are those of the test without damage. The data points are normalized between 1 and 2.

The data points of the test without damage are transformed using a Box Cox transformation. The aim is to modify the data in such a way that it is closer to a normal distribution. This provides a standardized baseline for the comparison with the untransformed data from the tests with damage. For the comparison, the arithmetic mean μ and the standard deviation σ are determined from the transformed data without damage. Three intervals are defined based on the standard deviation:

$$1^{\text{st}} \text{ interval: } \mu \pm \sigma \quad (11)$$

$$2^{\text{nd}} \text{ interval: } \mu \pm 2\sigma \quad (12)$$

$$3^{\text{rd}} \text{ interval: } \mu \pm 3\sigma \quad (13)$$

In case of an optimal normal distribution, 68.27 % of the data are in the first interval, 95.45 % in the second interval and 99.73 % in the third interval, see figure 6.

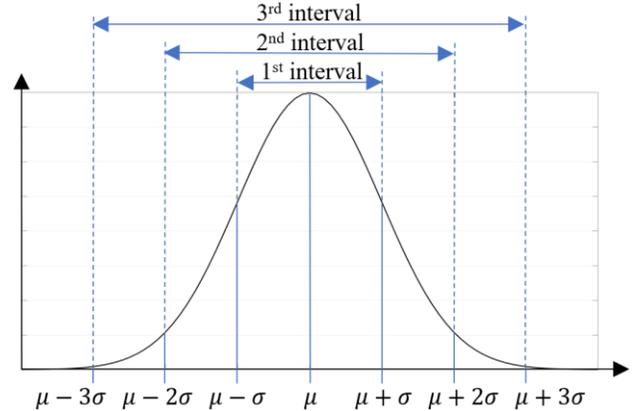


Figure 6. Normal distribution.

Based on the determined intervals, it is calculated how many of the 960 data points each of the tests with damage size S, M and L are outside the intervals. Since the limits of the intervals are further apart as the interval increases, the proportion of data points of the tests with damage that are outside the limits decreases. The first interval will therefore always show a greater deviation between tests without and with damage than intervals 2 and 3.

However, the different machine learning approaches (encoding and loss) can be compared with each other on the basis of the interval method, as it is used in both approaches. If a deviation between the test without and with damage can still be detected with the second or third interval in one approach, the damage is more clearly detectable and the approach is therefore more suitable. The approaches can therefore also be compared in terms of how much the detectability decreases with increasing interval. The less influence the used interval of an approach has on the detectability, the more suitable the approach is.

In the context of this paper, no fixed threshold for the data proportion outside the limits is to be defined with which damage can be detected. Instead, the two approaches (encoding and loss) are to be evaluated and compared with each other based on the predefined intervals.

2.4. Second evaluation approach: Loss

In the second evaluation approach, the AE is also trained on the basis of the vibration data from a test without damage. All vibration data from the test without damage and the three damage variables S, M and L of the individual operating state

are then appended to each other. After 3840 FFTs have been generated from 10,000 data points each, these are encoded and decoded using the trained AE. The FFTs produced by the AE are then compared to the original FFTs by calculating the mean squared error (first term of the loss function, see Eq. 7). The loss approach, such as the encoding approach, provides 960 real numbers per test without damage, damage size S, M and L. Based on this, a Box Cox transformation can be performed again on the data without damage and the three intervals can be determined (see Eq. (11), (12) and (13)). Subsequently, the proportion of data points outside the intervals is calculated for each of the tests with damage.

3. RESULTS

The results of the encoding and the loss approach are presented below.

3.1. Encoding approach

Figures 7 to 12 show the results of the encoding approach of the first gear pair for each operating condition. The lower and upper limits of the first interval are determined on the data without damage (see Eq. (11)). The upper and lower limits are marked with a red line. All blue data points are within this interval, all orange ones outside. For the damage sizes S, M and L, the proportion of data points inside and outside the interval is calculated.

For all operating conditions, it can be seen that the plotted data points per damage size scatter significantly. Overall, this applies even more to the higher speed level of 636 rpm than to the lower speed level of 72 rpm. The most difficult pitting to detect using the encoding approach for the first gear pair is pitting M at 72 rpm and 24 Nm (see figure 8). Here, only 20.2 % of the data points lie outside the first interval. The difference is therefore not as significant as with the other pitting sizes or operating conditions, for which at least 33.5 % of the data points always lie outside the interval.

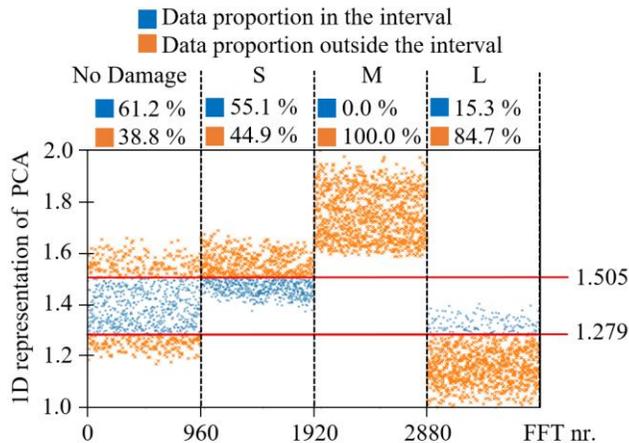


Figure 7. Encoding approach, 1st gear pair, 72 rpm, 18 Nm, 1st interval.

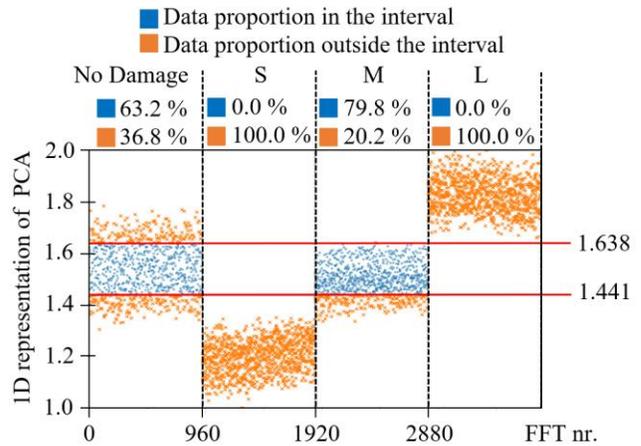


Figure 8. Encoding approach, 1st gear pair, 72 rpm, 24 Nm, 1st interval.

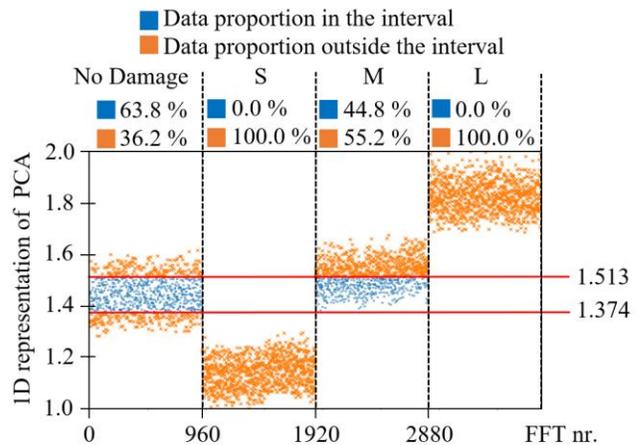


Figure 9. Encoding approach, 1st gear pair, 72 rpm, 30 Nm, 1st interval.

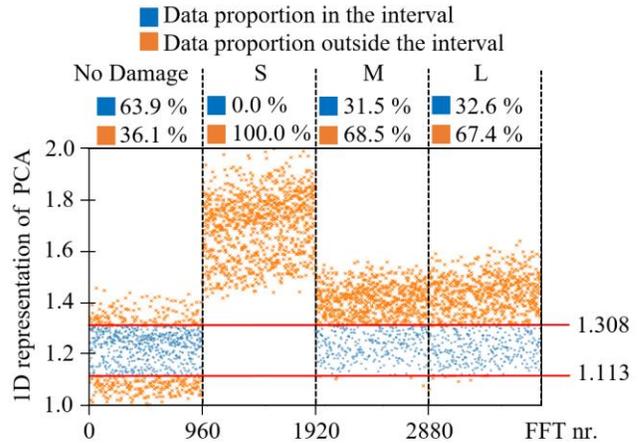


Figure 10. Encoding approach, 1st gear pair, 636 rpm, 18 Nm, 1st interval.

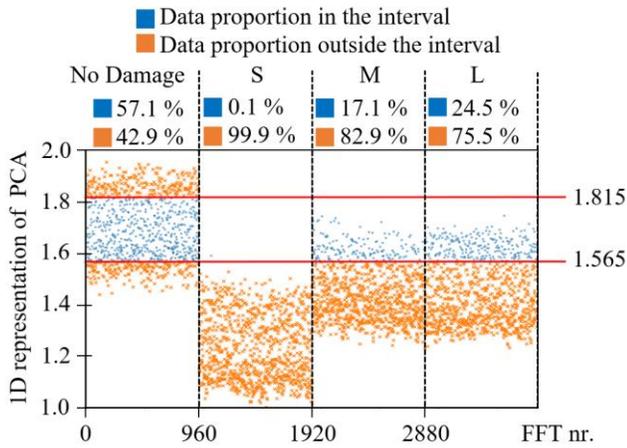


Figure 11. Encoding approach, 1st gear pair, 636 rpm, 24 Nm, 1st interval.

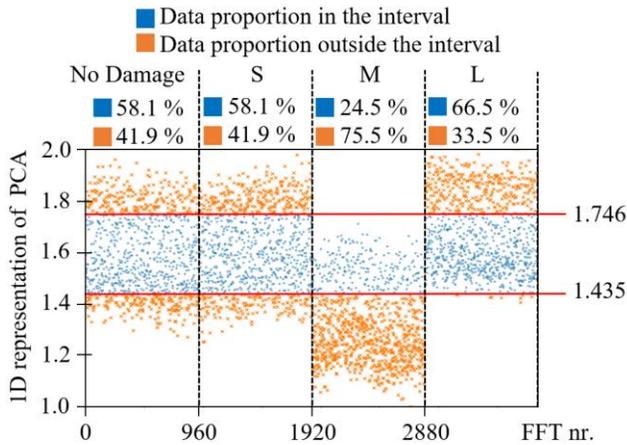


Figure 12. Encoding approach, 1st gear pair, 636 rpm, 30 Nm, 1st interval.

In addition to the encoding evaluation of the first gear pair using the first interval, the second and third intervals are also determined (see Eq. (12) and (13)). As mentioned, this is only used for comparison with the loss approach, as the detectability decreases as the interval increases. However, the aim is to evaluate how much the detectability decreases. For the operating condition 72 rpm and 18 Nm, this results in figure 13 (second interval) and figure 14 (third interval).

While for the first interval only 61.2 % of the data points without damage are within the limits (see figure 7), in the second interval 99.1 % (see figure 13) and in the third interval all data points (see figure 14) are within the limits. As the limits therefore have a greater distance, it is more difficult to detect a difference to the data points of the experiments with damage. Even in the evaluation with the second interval, only 1.7 % of the data points for pitting size S are outside the interval (see figure 13). If the third interval is used for the

evaluation, no difference is recognizable, as all data points of pitting S are within the interval (see figure 14). Pitting sizes M and L are also more difficult to detect as the interval increases.

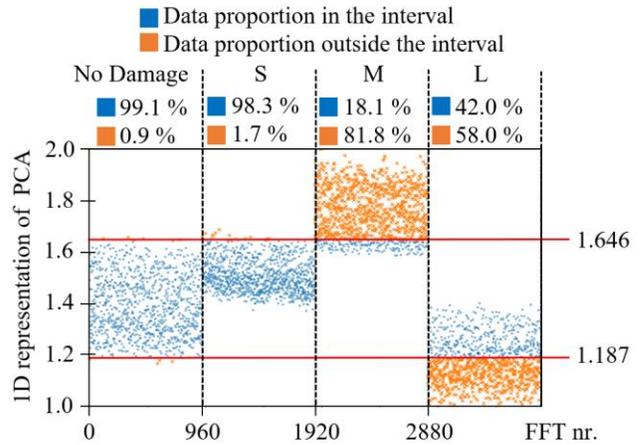


Figure 13. Encoding approach, 1st gear pair, 72 rpm, 18 Nm, 2nd interval.

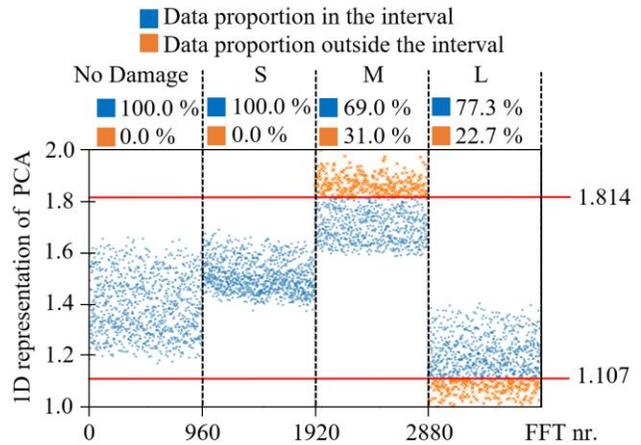


Figure 14. Encoding approach, 1st gear pair, 72 rpm, 18 Nm, 3rd interval.

Figure 15 presents the evaluation of the encoding approach for all operating conditions of the first gear pair. The proportion of data points for damage sizes S, M and L that are outside the respective interval is shown.

When using the second interval, some pitting can no longer be detected (pitting S at 72 rpm and 18 Nm, pitting M at 72 rpm and 24 Nm, pitting S and L at 636 rpm and 30 Nm) or less clearly (13.4 % for pitting M at 72 rpm and 30 Nm). For all other pitting and operating conditions, at least 40.8 % of the data points are always outside the limits of the second interval. If the third interval is used for detection, the detectability of the pitting decreases significantly.

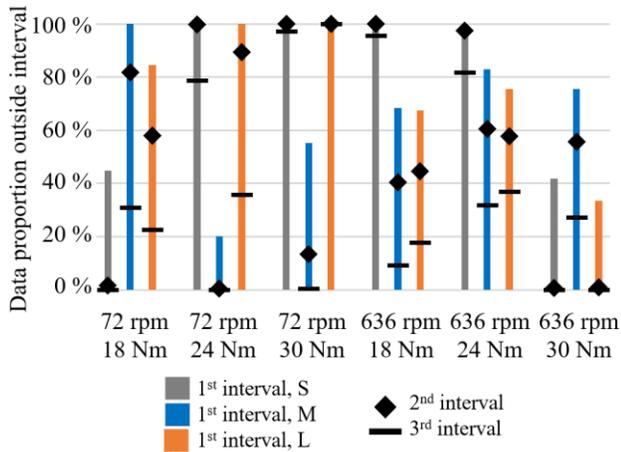


Figure 15. Encoding approach, 1st gear pair.

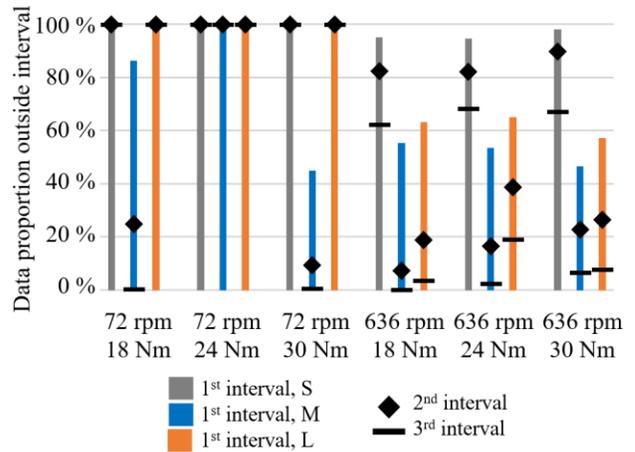


Figure 17. Encoding approach, 3rd gear pair.

Figures 16 and 17 present the results of the encoding approach for the second and third gear pair in all operating conditions.

Considering the second and third gear pair, it is noticeable that the global tendency of the encoding approach corresponds to that of the first gear pair. When using the first interval, at least 39.8 % of the data for the second gear pair is always outside the limits, with one exception (28.6 % at 72 rpm and 30 Nm). For the third gear pair, a minimum of 44.9 % of the data is always outside the limits when using the first interval. If the second and third intervals are considered, the proportion of data points outside the limits decreases significantly for certain operating conditions. Especially with the third interval, some damage can no longer be detected.

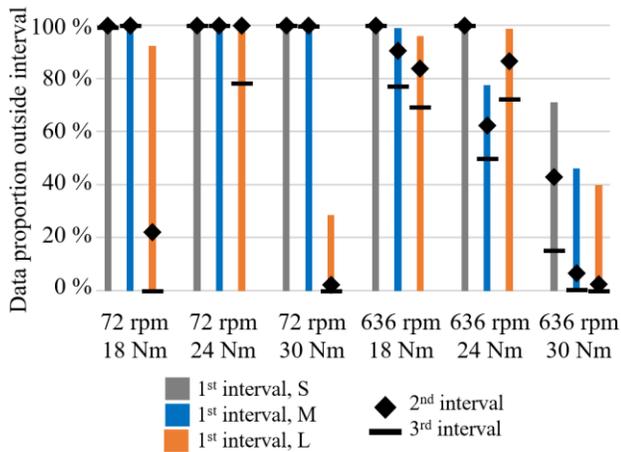


Figure 16. Encoding approach, 2nd gear pair.

3.2. Loss approach

The results of the loss approach of the first gear pair for each operating condition are given in figures 18 to 23. Again, the lower and upper limits of the first interval are determined using the data without damage (see Eq. (11)) and marked with a red line. The proportion of data points within and outside this interval is identified.

Overall, the results of the loss approach at the low speed level of 72 rpm have a low scatter of the data points. At the higher speed level of 636 rpm, larger scatter is recognizable. In all operating conditions, the data points of all pitting sizes are at least 96.8 % outside the limits – except for pitting L at operating condition 636 rpm and 18 Nm (77.0 %, see figure 21) and operating condition 636 rpm and 30 Nm (52.7 %, see figure 23). Overall, all pitting of the first gear pair can therefore be detected using the first interval of the loss approach.

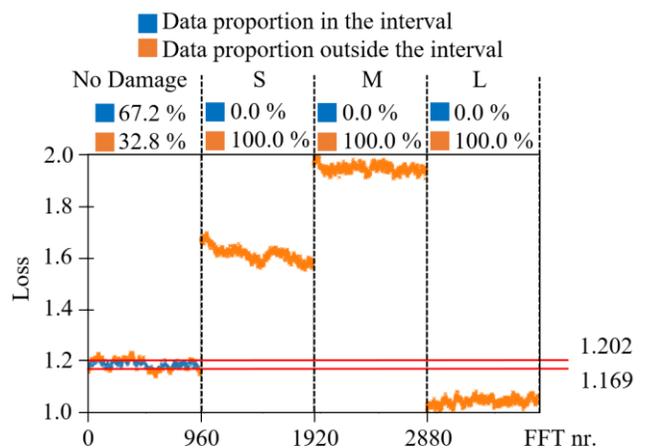


Figure 18. Loss approach, 1st gear pair, 72 rpm, 18 Nm, 1st interval.

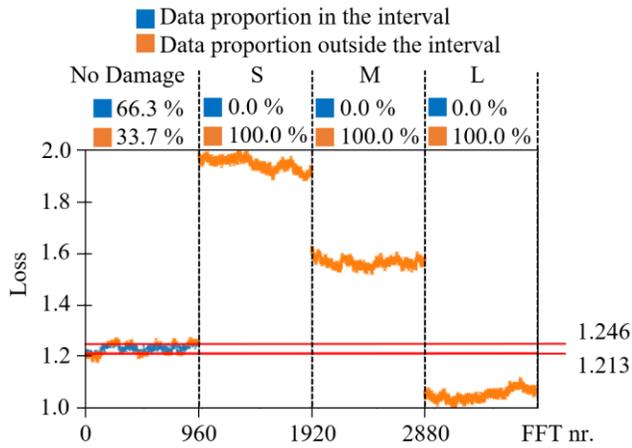


Figure 19. Loss approach, 1st gear pair, 72 rpm, 24 Nm, 1st interval.

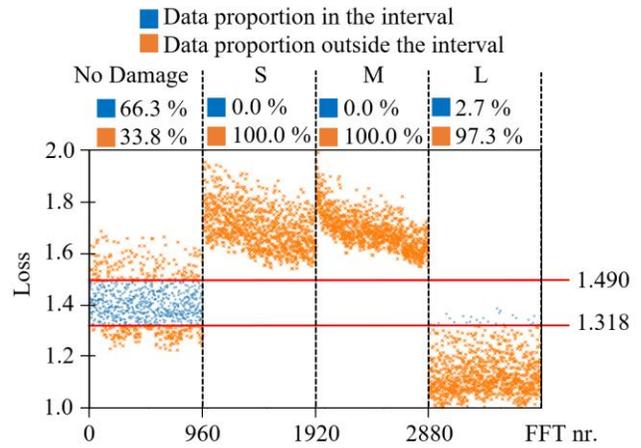


Figure 22. Loss approach, 1st gear pair, 636 rpm, 24 Nm, 1st interval.

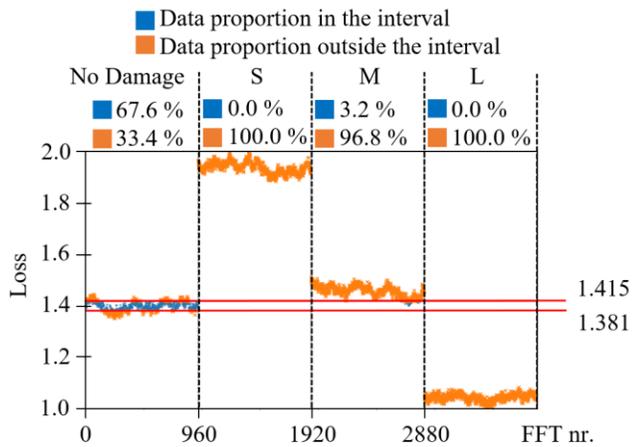


Figure 20. Loss approach, 1st gear pair, 72 rpm, 30 Nm, 1st interval.

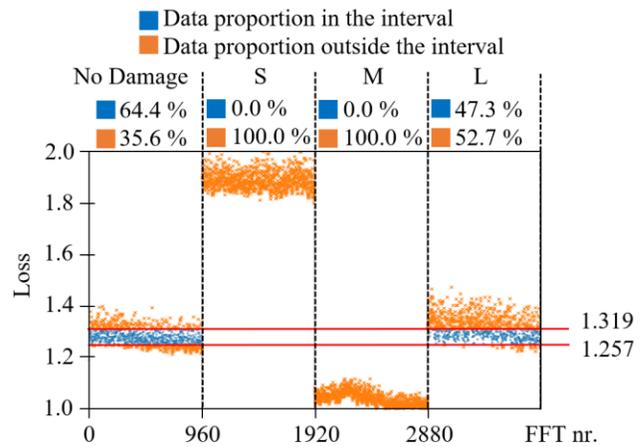


Figure 23. Loss approach, 1st gear pair, 636 rpm, 30 Nm, 1st interval.

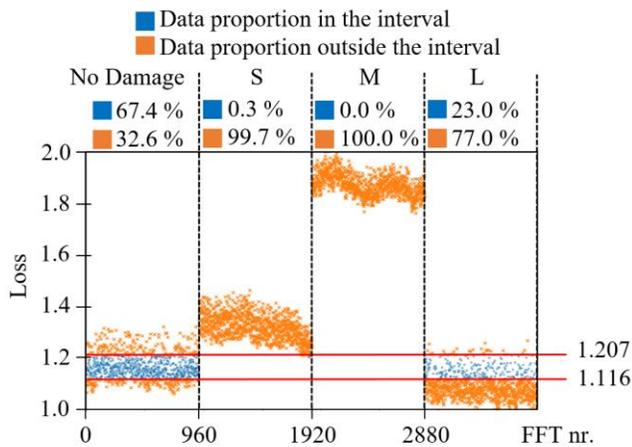


Figure 21. Loss approach, 1st gear pair, 636 rpm, 18 Nm, 1st interval.

In addition to the first interval, the second and third intervals are also determined for the loss approach (see Eq. (12) and (13)). The evaluation of all operating conditions of the first gear pair is illustrated in figure 24. The proportion of data points for damage sizes S, M and L that are outside the respective interval is shown.

When using the second interval, pitting L at operating condition 636 rpm and 30 Nm is the worst detectable pitting with only 16.4 % of the data outside the interval. Otherwise, at least 51,3 % of the data is always outside the limits. Using the third interval, pitting L is not detectable at operating condition 636 rpm and 30 Nm. Here, only 0.3 % of the data is outside the limits. Overall, pitting is more difficult to detect when using the third interval, especially at the higher speed level of 636 rpm.

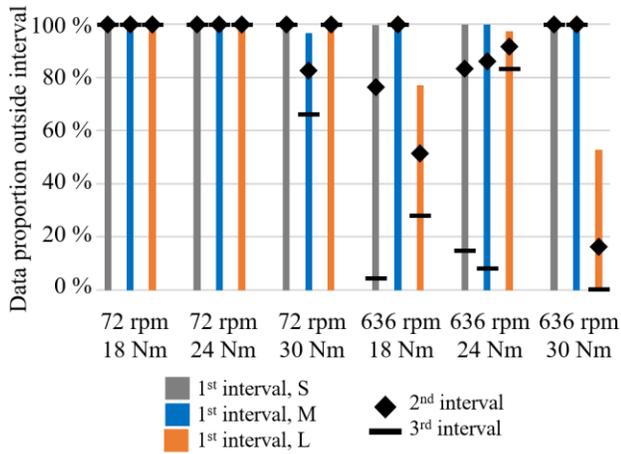


Figure 24. Loss approach, 1st gear pair.

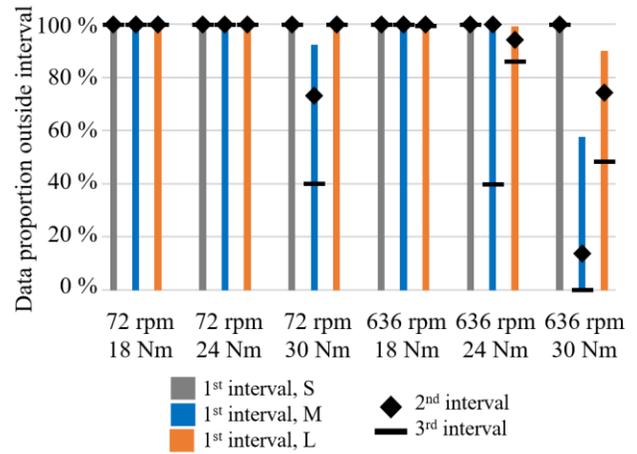


Figure 26. Loss approach, 3rd gear pair.

Figures 25 and 26 present the results of the loss approach of the second and third gear pairs for all operating conditions.

When using the first interval, at least 89.7 % of the data for the second gear pair is always outside the limits. For the third gear pair, a minimum of 57.7 % of the data is always outside the limits when using the first interval. If the second interval is calculated, 52.2 % of the data for the second gear pair is always outside the limits and 13.7 % for the third gear pair. When using the third interval, a decrease in the proportion of data points outside the limits for individual operating states and pitting sizes can be seen, similar to the first gear pair.

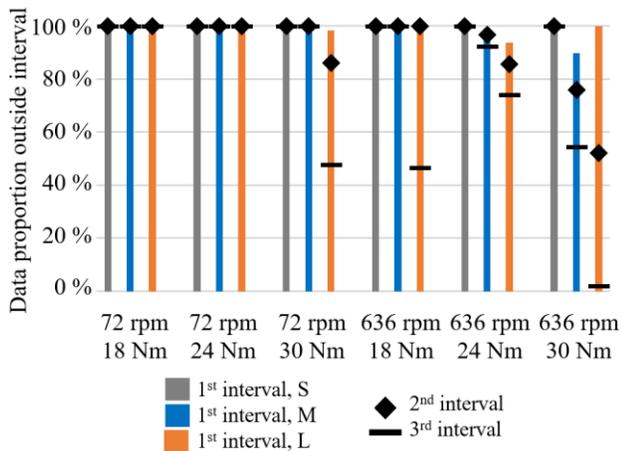


Figure 25. Loss approach, 2nd gear pair.

4. DISCUSSION

Overall, the comparison of the encoding and loss approach shows that the data points scatter considerably more in the encoding approach. As a result, the upper and lower limits of the first interval are significantly further apart in the encoding approach than in the loss approach. When using the second and third intervals, this results in greater differences with the encoding approach because the limits are then frequently so far apart that pitting detection is no longer possible.

However, if the first interval is used for pitting detection with the encoding approach, a minimum of 20.2 % of the data points are always outside the limits for the first gear pair, 39.8 % for the second gear pair and 44.9 % for the third gear pair.

In total, the loss approach shows significantly better pitting detection than the encoding approach. When using the first interval, a minimum of 52.7 % of the data points are always outside the limits for the first gear pair, 89.7 % for the second gear pair and 57.7 % for the third gear pair.

A total of 18 cases are examined with the two approaches (3 pitting sizes in 6 operating states). In order to be able to compare the approaches even better, it is considered in how many of the 18 cases there is a significant deviation - i.e. at least 50 % of the data points outside the limits. The result can be found in figure 27.

For the first interval of the encoding approach, between 14 and 16 cases have a deviation greater than 50 %, depending on the gear pair. With the first interval of the loss approach, all damage sizes in all operating states of all gear pairs have a minimum deviation of 50 %. With the second interval of the encoding approach, the 50 % criterion only applies to 10 to 13 cases, depending on the gear pair. With the loss approach, it still applies to a minimum of 17 cases. When using the third interval, the number of cases in which at least 50 %

of the data points lie outside the interval is reduced for the encoding approach to between 5 and 12 cases. Large differences within the gear pairs can therefore also be seen here. In contrast, the loss approach shows 13 to 15 cases.

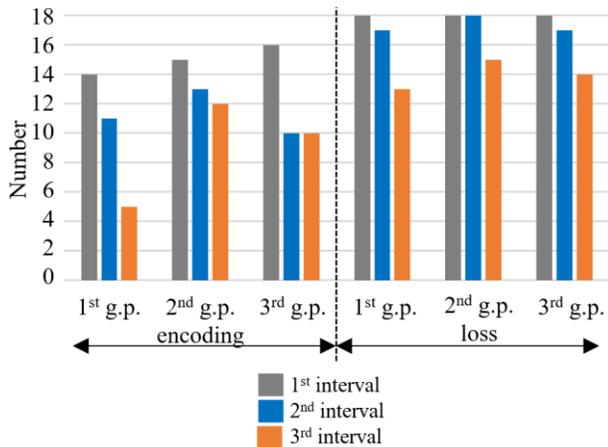


Figure 27. Number of cases with at least 50 % of the data outside the interval, per approach and gear pair (g.p.).

Overall, the loss approach is much better suitable for recognizing a clear difference between the tests without and with damage. In the second interval of the loss approach, a minimum of 50 % of the data points are still outside the limits in at least 17 cases. The encoding approach does not even achieve this for the first interval. In addition, the loss approach is not only less sensitive to different damage sizes and operating conditions, but also to different gear pairs. However, this paper does not consider how well false positives can be excluded with the two approaches.

Regardless of the comparison of the two approaches, even the smallest investigated pitting with a size of 0.61 % (1st gear pair), 0.51 % (2nd gear pair) and 0.78 % (3rd gear pair) could be detected in the context of this study. In contrast to the approach presented in (Binanzer et al., 2023), in which an AE was combined with a Long Short Term Memory (LSTM) network, detection is also possible with a purely unsupervised algorithm. This offers the advantage for the application that no labeled training data is required. Only data from a test without damage is required for training.

The detectable pitting sizes in the scope of this work are a significant improvement on other investigations. There are various approaches for pitting detection in gearboxes using vibration sensors. The approaches differ on the one hand in the investigated pitting size and in the methods of sensor data evaluation.

Qu, M. He, Deutsch and D. He (2017) investigated one row of pitting damage along the tooth width of one tooth. A stacked autoencoder network was used to perform the dictionary learning in sparse coding and automatically extract features from the raw vibration data. With these features a

backpropagation neural network was trained to identify the damage.

Fan, Zhou, Wu and Guo (2017) developed a gear damage detection and localization approach by analyzing the vibration signal of an individual tooth and Support Vector Machines (SVM). The dispersion degree and vibration accelerations of the waveform of an individual gear tooth were studied to investigate the characteristics of gear tooth under normal, small failure (< 5 % damaged tooth area) and serious failure (> 5 % damaged tooth area) conditions.

An unsupervised feature extraction method called disentangled tone mining was presented by Qu, Zhang, M. He, D. He, Jiao and Zhou (2019). This method was able to identify the fault level directly from the frequency spectrum of the measured vibration data. Pitting sizes between 4.33 % and 24.91 % were investigated in a single stage spur gearbox.

Medina, Cerrada, Cabrera, Sanchez, Li and Oliveira (2019) used a LSTM network for classifying nine levels of pitting. The smallest investigated pitting had a size of 4.16 %.

Pitting sizes of less than 1 % were detected by Grzeszkowski, Nowoisky, S., Scholzen, Kappmeyer, Gühmann, Brimmers and Brecher (2020) using a SVM classifier. A disadvantage of the SVM classifier is that it is a supervised algorithm and therefore requires labeled training data.

Damage detection with purely physically based data evaluation, with pitting sizes between 6.3 % and 41.7 %, was presented by Sowana und Chandrasekaran (2020). In each case, the root mean square (RMS) value of the structure-borne noise data in the time domain of the undamaged and damaged gear was compared.

Sarvestani, Rezaeizadeh, Jomehzadeh and Bigani (2020) also examined the detection of naturally occurring pitting damages with a size of 30 %, 60 % and 90 % using purely physically based methods. The frequency spectrum of the structure-borne noise data was divided into six ranges. The damage was best detected in the second gear mesh harmonic range.

Häderle, Merkle and Dazer (2024) presented another physically based data analysis approach. It is shown that the greatest percentage difference between undamaged and damaged gears can be determined for the harmonics of the gear mesh frequency (GMF) and the sidebands between 24,000 Hz and 40,300 Hz. Thus, it was possible to detect very small pitting sizes between 0.42 % and 1.83 %.

5. CONCLUSION

In order to increase the service life of gearboxes, avoid unexpected failures and thus reduce overall operating, maintenance and labor costs, comprehensive PHM has to be implemented in gearbox applications. Adaptive operating strategies can even extend the RUL without any loss of performance. In order for the PHM of gearboxes to achieve its full

potential, damage detection at the earliest possible stage is essential.

In this study, two unsupervised machine learning approaches (encoding and loss approach) were developed and the detection of artificially manufactured damage on the tooth flank of a test gearbox was investigated.

In particular, the loss approach is more capable of identifying a difference between no damage and damage than the coding approach, regardless of the size of the pits and operating conditions. The loss approach is also less sensitive to different gear pairs, which have slightly different properties due to material and manufacturing tolerances.

Overall, it can be stated that the main contribution of this work is that existing Machine Learning tools have been applied to the challenge of a very early damage detection in gearboxes. Without the need of complex physically based evaluation methods of the vibration data, the smallest pitting of about 0.5 % could be detected regardless of the operating condition. The use of the sparse AE was described in detail and two evaluation methods were compared.

REFERENCES

- Bertsche, B. & Dazer, M. (2022). *Zuverlässigkeit im Fahrzeug- und Maschinenbau: Ermittlung von Bauteil- und System-Zuverlässigkeiten*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Binazer, L., Merkle, L., Dazer, M. & Nicola, A. (2023). Pitting Detection for Prognostics and Health Management in Gearbox Applications. *International Conference on Gears 2023 (VDI-Berichte*, vol. 2422, pp. 97-108), September 13-15, Munich, Germany. doi:10.51202/9783181024225
- Fan, Q., Zhou, Q., Wu, C. & Guo, M. (2017). Gear tooth surface damage diagnosis based on analyzing the vibration signal of an individual gear tooth. *Advances in Mechanical Engineering (AIME)*, vol. 9 (no. 6), pp. 1-14. doi: 10.1177/1687814017704356
- German Institute for Standardization (DIN) (2010). Industrial liquid lubricants - ISO viscosity classification (ISO 3448:1992). In DIN, *DIN ISO 3448:2010-02*. Berlin, Germany: Beuth Verlag GmbH. doi: 10.31030/1562009
- Goebel, K., Celaya, J., Sankararaman, S., Roychoudhury, I., Daigle, M. & Abhinav, S. (2017) *Prognostics: The Science of Making Predictions*. CreateSpace Independent Publishing Platform.
- Gretzinger, Y., Lucan, K., Stoll, C., & Bertsche, B. (2020). Lifetime Extension of Gear Wheels using an Adaptive Operating Strategy. *Proceedings of 7th International Conference Integrity-Reliability-Failure* (pp. 703-710), September 6-10, Funchal, Portugal. https://fe.up.pt/irf/Proceedings_IRF2020/
- Grzeszkowski, M., Nowoisky, S., Scholzen, P., Kappmeyer, G., Gühmann, C., Brimmers, J. & Brecher, C. (2020). Classification of Gear Pitting Severity Levels using Vibration Measurements. *In tm - Technisches Messen*, vol. 87 (no. s1), pp. s56-s61. doi: 10.1515/teme-2020-0026
- Häderle, P., Merkle, L. & Dazer, M. (2024). Vibration Analysis for Early Pitting Detection During Operation. *Forschung im Ingenieurwesen*. vol. 88 (article nr. 15). doi: 10.1007/s10010-024-00743-5
- Institute of Electrical and Electronics Engineers (IEEE) (2017). IEEE Standard Framework for Prognostics and Health Management of Electronic Systems. In IEEE, *IEEE Std 1856-2017* (pp. 1-31). doi: 10.1109/IEEE-ESTD.2017.8227036
- International Organization for Standardization (ISO) (2016). Calculation of load capacity of spur and helical gears – Part 5: Strength and quality of materials. In ISO, *ISO6336-5:2016(E)*, (p. 5). Genève, Switzerland: International Standards Organization.
- Medina, R., Cerrada, M., Cabrera, D., Sanchez, R.-V., Li, C. & Oliveira, J. V. D. (2019). Deep Learning-Based Gear Pitting Severity Assessment Using Acoustic Emission, Vibration and Currents Signals. *Proceedings of 2019 Prognostics and System Health Management Conference (PHM-Paris 2019)* (pp. 210-216), May 2-5, Paris, France. doi: 10.1109/PHM-Paris.2019.00042
- Ng, A. (2011). Sparse autoencoder. CS294A Lecture notes, 72, pp. 1-19.
- Qu, Y., He, M., Deutsch, J. & He, D. (2017). Detection of Pitting in Gears Using a Deep Sparse Autoencoder. *Applied Science. Special Issue Deep Learning Based Machine Fault Diagnosis and Prognosis*, vol. 7 (no. 5), pp. 515-529. doi: 10.3390/app7050515
- Qu, Y., Zhang, Y., He, M., He, D., Jiao, C. & Zhou, Z. (2019). Gear pitting fault diagnosis using disentangled features from unsupervised deep learning. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, vol. 233 (no. 5), pp. 719-730. doi: 10.1177/1748006X18822447
- Research Association for Drive Technology (FVA) (1985). FVA-Heft 180 Referenzöle - Datensammlung für Mineralöle.
- Sarvestani, E. S., Rezaeizadeh, M., Jomehzadeh, E. & Bigani, A. (2020). Early Detection of Industrial-Scale Gear Tooth Surface Pitting Using Vibration Analysis. *Journal of Failure Analysis and Prevention*, vol. 20, pp. 768-788. doi: 10.1007/s11668-020-00874-1
- Sonawane, P. R. & Chandrasekaran, M. (2020). Investigation of gear pitting defect using vibration characteristics in a single-stage gearbox. *The International Journal of Electrical Engineering & Education*, vol. 57 (no. 3). doi: 10.1177/0020720918813837