

Image-Based Damage Quantification and Vibration-Based Estimation of Gear Tooth Degradation

Paul Stuiwer¹

¹Royal Netherlands Aerospace Centre, Amsterdam, The Netherlands

Paul.Stuiwer@nlr.nl

1 Introduction

Estimating gear damage remains a complex task due to significant variability in gearbox lifespans. The ability to accurately predict gear damage is crucial in various industries such as aerospace and automotive, where gearbox failure can have catastrophic consequences. While conventional data acquisition methods rely on accelerometer data to monitor gearbox health, this data is indirect and can be affected by noise and lead to inaccurate gear damage predictions. Optical images of the tooth surface provide a reference for the gear spall damage but image acquisition is expensive as it disrupts the operation.

The 2026 PHM Data Challenge aims to develop models that can predict gear damage using accelerometer data and only use imaging data during training.

A high-level detail of the methodology is depicted in Figure 1. In this work, an image-based damage indicator (DI) is constructed to quantify gear degradation where candidate image features are evaluated against the Coble metrics [1]. Subsequently, the DIs are used as a reference for developing partial least squares (PLS) regression model using vibration-based condition indicators for gear degradation estimation [2,3]. The method is assessed using leave-one-out cross validation (LOOCV) and applied to unseen experiments to assess the potential.

2 Dataset

The PHM 2026 Data Challenge dataset consists of seven accelerated life experiments (A-G) on a spur-gear gearbox of 28 teeth with a one-to-one gear ratio. Two accelerometers recorded vibration signals throughout the experiments. For only the four training experiments, after every run (at roughly 6-hour intervals), images were taken of each tooth's surface to capture surface degradation serving as ground-truth data. As lifespans vary from 30 to 90 hours, the training degradation observations remain limited, which posed a challenge for model development and generalization.

Paul Stuiwer. 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.

3 Methodology

3.1 Image-based Damage Indicator

The image-based damage indicator is constructed by transforming the tooth-surface images into quantitative features aiming to capture the degradation over the course of the experiment (see Figure 2). The image processing pipeline aims to ensure spatial consistency between the images by aligning them to the reference break-in image using enhanced correlation coefficient (ECC) alignment. Background illumination correction is applied to remove different lighting across images. Two regions of interest were defined: the tooth contact region and a region directly below it where spalling was observed during later runs. For both regions, difference images relative to the test start are generated to capture changes in the tooth surface over time. Then, image features are extracted of the processed images and evaluated for their ability to track degradation according to the Coble metrics that give a score for monotonicity, trendability and prognosability. The scores are combined into a fitness score, that is

$$f = w_m M + w_t T + w_p P \quad (1)$$

where f is the fitness score, M , T and P the monotonicity, trendability and prognosability scores respectively between 0 and 1 and weights $w_m = 0.4$, $w_t = 0.4$ and $w_p = 0.2$. Features are ranked according to their Coble fitness score, and the highest-ranked ones are retained for a principal component analysis (PCA) to obtain a per-tooth level damage indicator. As severe spall damage occurred for only a few teeth, the mean of the top 3 per-tooth damage indicators is taken to get a single gearbox-level damage indicator. Isotonic regression is applied as a post-processing step to enforce a non-decreasing damage trajectory, reflecting the expected progressive nature of gear degradation.

3.2 Sensor-to-Damage mapping

The sensor-to-damage model is a mapping between the sensor data and the constructed image-based damage indicators. It covers a typical data-driven workflow in which data is filtered, physics-based features are derived and selected to minimize the error between the input and output (see Figure 4). In this work, only the context variables and the given condition indicators (CIs) were used as input for the model.

For each test, the speed and torque signals in

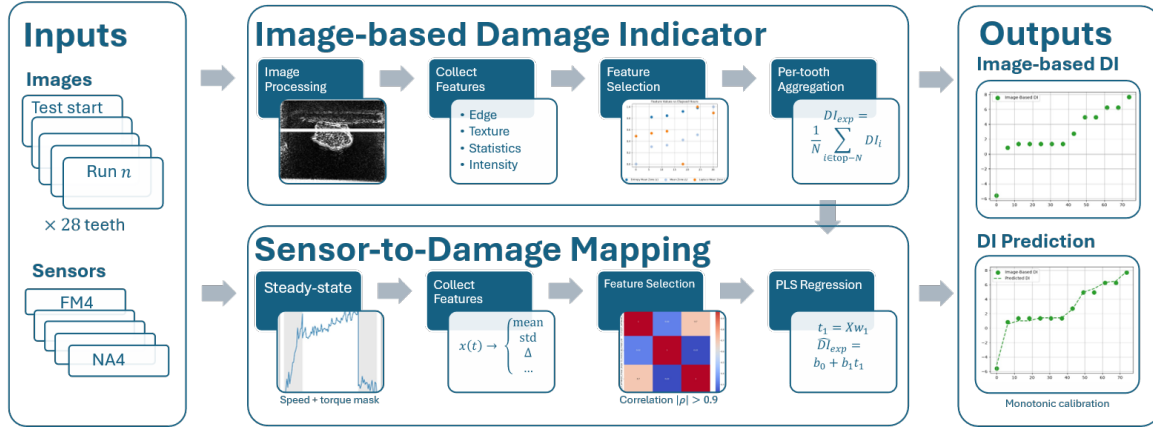


Figure 1: Flowchart of the development of the image-based damage indicators and vibration-based model for gear degradation estimation.

the condition indicators showed a warm-up phase where the speed and torque signals increase until a steady value is reached. At the end, a cool-down phase is recognized where the speed and torque signals decrease. Since most degradation-related features rely on time-synchronous averages and subtraction of gear mesh frequencies, the operating condition where the speed and torque are above 95% of their respective median values, is extracted and applied to all signals.

In each test, for each signal, features are calculated aiming to find features that are related to the gearbox damage condition. These included simple statistical moments and quantiles and energy-related features. Based on the observation that some signals, for example $FLEET_{WTMADE}$, showed a gradual increase over the full operating life, the feature collection also includes the difference features, that is

$$\Delta f = f(x_{\text{last 10 min}}) - f(x_{\text{first 10 min}}) \quad (2)$$

where f represents the feature operator and Δf the difference between the feature values of the signal in the first and last 10 minutes ($x_{\text{first 10 min}}$ and $x_{\text{last 10 min}}$ respectively). In addition, the relative feature values with respect to the feature value based on the first test are included. To minimize redundancy and limit model complexity, vibration features were removed that shared an absolute Pearson correlation coefficient above 0.9. Subsequently, the remaining vibration features were ranked according to their absolute Pearson correlation coefficients with the image-derived damage indicator.

The limited number of independent experiments increase the risk of overfitting the model. For that reason, a relatively simple Partial Least Squares (PLS) regression was chosen to create a mapping between the best top-ranked vibration features and the image-derived damage indicators. By iterating over the top-ranked features, PLS models are fitted with an experiment-level leave-one-out cross validation (LOOCV) approach. The feature set is selected based on the average mean squared error (MSE) across the out-of-fold predictions. This way,

the best-performing PLS model is aimed to generalize well to unseen experiments. With the best-performing PLS model, the damage prediction trajectory of each experiment is post-processed using isotonic regression to enforce the expected monotonically increasing damage trajectory.

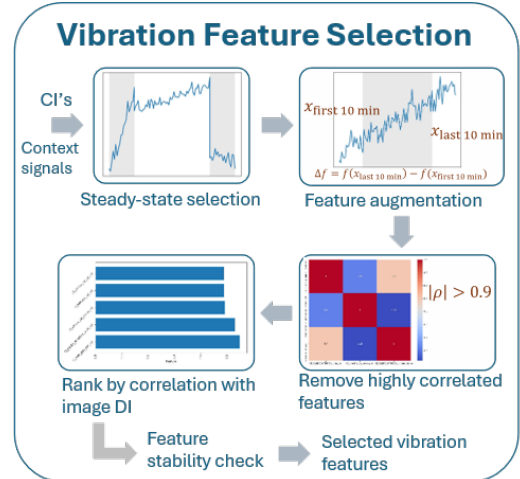


Figure 4: Vibration feature selection pipeline.

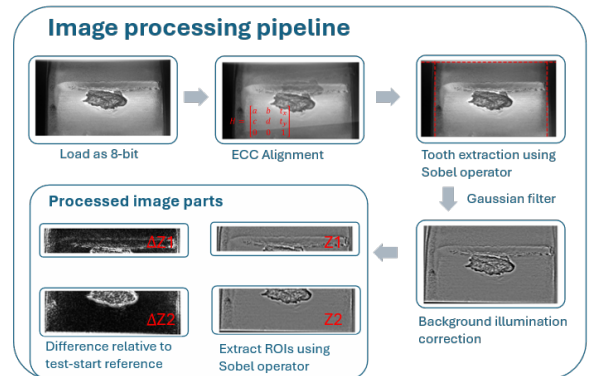


Figure 2: Image processing pipeline to construct the image-derived damage indicator. Z1 refers to the tooth contact region and Z2 refers to the spalling region. The symbol Δ refers to the difference relative to the test start.

4 Results and Discussion

Figure 3 shows the processed difference images over the course of the experiment A for tooth 16. The trajectory shows an increasing difference in pixel intensity and rough texture over the course of the experiment. Spalling is visible in the last run and introduced slight errors in the ECC alignment that can lead to unreliable feature values as unintended high intensity pixels occur at the border of the image. In addition, the poor separation of the contact region (Z1) and spalling region (Z2) depicted by the red dashed line in run 5 can introduce unreliable feature values.

Figure 5a shows the top 5 image features with the highest fitness score. The selected features correspond to changes in intensity and surface texture as the mean relative difference feature in zone 1 captures the overall magnitude of the surface change, while the Canny-based features and entropy capture changes in edge structure and spatial variability as the surface becomes more irregular. The canny variance in the lower region of interest is relevant for the later stages of degradation, where spalling occurs. The results therefore show that the selected features capture the observable changes in the tooth surface, providing a suitable basis for constructing the image-derived damage indicator. The top vibration features that scored highest in correlation with the image-based damage indicator values are depicted in figure 5b. The 3 features selected

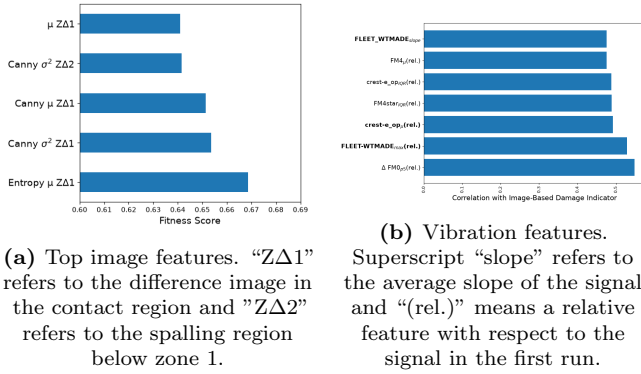


Figure 5: Top image and vibration features.

for the PLS model are highlighted in bold and are based on the lowest mean MSE during leave-one-experiment-out cross validation in combination with plausible test and validation degradation trajectories and lowest test data score. Although PLS was selected to accommodate for correlated vibration indicators, the final model required only three features, suggesting that a simple linear relationship may have been sufficient.

The selected features have a plausible physical relationship with gear degradation as $\text{crest-e}_{\text{op}}$ is related to impulsive changes in the vibration response and $\text{FLEET}_{\text{WTMADE}}$ is related to overall changes in vibration response, but further analysis is required to directly relate them to specific damage of the gearbox. Figure 6 shows the training results of the image-based degradation trajectory per experiment and the PLS predictions. The PLS model reproduces the general degradation trajectory but raw predictions contain non-monotonic variations. In Figure 7, the test and validation predictions are depicted and show the model is able to produce degradation trajectories.

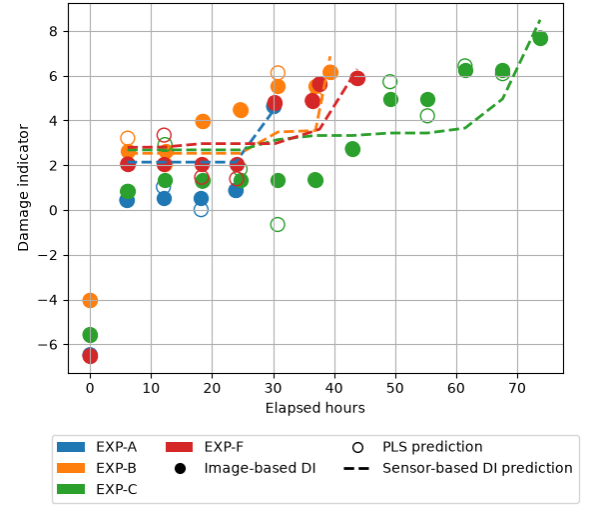


Figure 6: Sensor-based damage estimation using image-derived damage indicators on the training data.

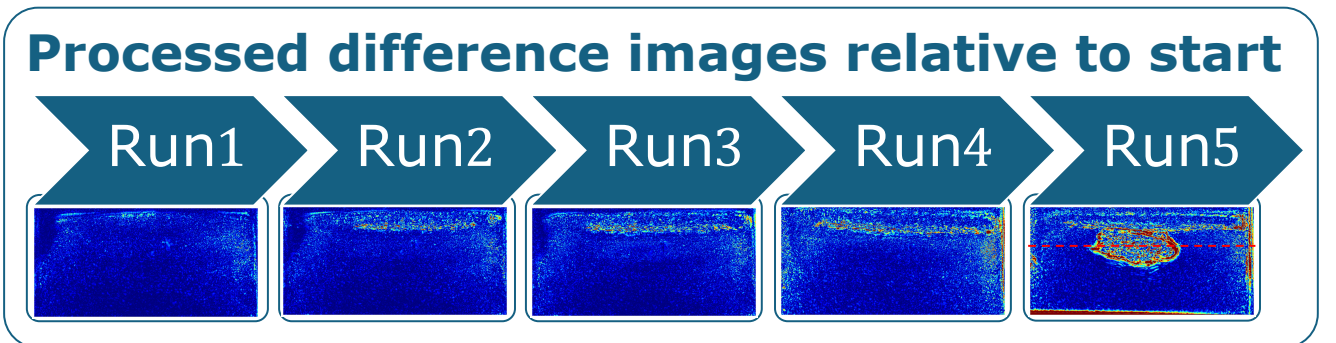


Figure 3: Result of the processed difference images over the course of experiment A for tooth 16. The red dashed line indicates the constructed boundary between the contact region (Z1) and spalling region (Z2).

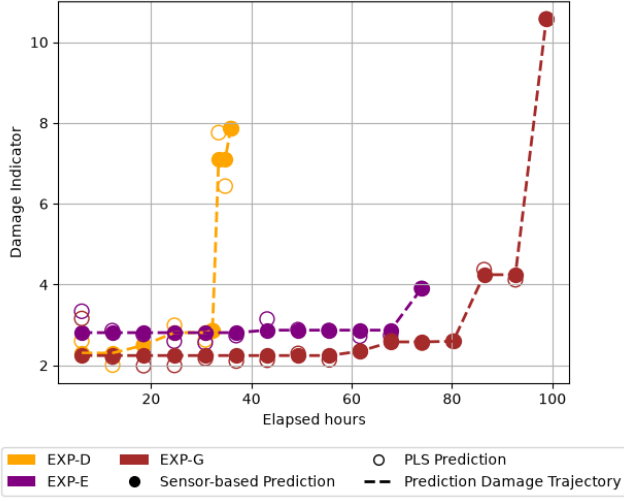


Figure 7: Sensor-based damage predictions on the test and validation data.

5 Conclusion

A vibration-based degradation model for a gearbox was developed using image-based damage indicators from tooth surface images as a reference. The image features captured the progressive damage observed and a PLS model using 3 vibration features reproduced overall damage progression. Despite the limited number of independent experiments and instability in the vibration features, the results show a potential of vibration measurements to estimate image-derived gear degradation.

References

- [1] J. Coble and J. W. Hines, “Identifying Optimal Prognostic Parameters from Data: A Genetic Algorithms Approach,” in *Proceedings of the Annual Conference of the PHM Society*. PHM Society, Mar. 2021.
- [2] V. Sharma and A. Parey, “A Review of Gear Fault Diagnosis Using Various Condition Indicators,” *Procedia Engineering*, vol. 144, pp. 253–263, 2016. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1877705816303484>
- [3] S. Wold, M. Sjöström, and L. Eriksson, “PLS-regression: a basic tool of chemometrics,” vol. 58, no. 2, pp. 109–130. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0169743901001551>