Predicting pitting severity in gearboxes under unseen operating conditions and fault severities using convolutional neural networks with power spectral density inputs

Rik Vaerenberg[1,2], Douw Marx[1,2], Seyed Ali Hosseinli[1,2], Fabrizio De Fabritiis[1,2], Hao Wen[1,2], Rui Zhu[1,2] and Konstantinos Gryllias[1,2,3]

[1] *Department of Mechanical Engineering, Division LMSD, KU Leuven, Celestijnenlaan 300, Box 2420, 3001 Leuven, Belgium*
rik.vaerenberg@kuleuven.be
konstantinos.gryllias@kuleuven.be

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

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

Abstract

The PHM North America 2023 Data Challenge tasked participants to diagnose the pitting fault severity of a gearbox from a three-channel vibration signal. This work summarizes the authors' proposed diagnostics solution which consists of a convolutional neural network with an ordinal loss criterion, trained on the power spectral density of the signal. This method is selected based on a rigorous evaluation using three dedicated validation sets, designed to evaluate the model's ability to generalize to unseen operation conditions and fault severities. Ultimately, the proposed approach achieved a competition validation score of 282.2 and a test score of 213.3.

1. Internal validation sets to evaluate model generalisation on unseen data

Three internal validation sets are split from the available training data. Each validation set is used to mimic a part of the true test condition where the model is required to generalize to unseen inputs. Validation set I is a conventional validation set used to evaluate the model on unseen, in-distribution data. This involves setting 20% of the training data aside for validation through random selection. Validation set II contains two unseen speed conditions. Finally, validation set III evaluates the model on two previously unseen fault severity levels. All further modelling decisions in this work were then based on model performance on these internal validation sets.

2. Data pre-processing and augmentation

Welch's estimate for the Power Spectral Density (PSD) (Welch, 1967) proved to be superior to other pre-processing candidates including cyclo-stationary analysis (Antoni, 2009), time-synchronous averaging (Mcfadden & Toozhy, 2000) and frequency whitening methods (Borghesani, Pennacchi, Randall, Sawalhi, & Ricci, 2013). Interestingly, methods that did not rely on angular re-sampling (Fyfe & Munck, 1997) performed best on in distribution data.

Before pre-processing, a rudimentary data augmentation was performed. This involved using signal segments with different time offsets, thereby introducing variations in the temporal alignment of the data for the purpose of improved model generalization. After data augmentation, the PSD was estimated from a signal with 40960 data points using an FFT window length of 8000. This ensured that at least one full revolution (and corresponding fault event) was present in each sample for any of the rotation speeds tested. Finally, the PSD data was re-scaled before model training, with a log transform performing best amongst the implemented scaling methods (Box-Cox (Box & Cox, 1964), z-normalization over samples or frequencies). Interestingly, we found that the model's generalisation capability on unseen speeds was especially sensitive to the re-scaling scheme used.

3. A Convolutional Neural Network (CNN) for ordinal regression

The pre-processed data is used to train a three-layer CNN with an ordinal regression loss function (Rosenthal & Ratna, 2022). In this way, the relative relationship between the ordinal health states is modelled instead of treating each health state as a distinct class, thereby improving model generalizability. CNN convolutions are performed over the frequency axis of the PSD, with the convolutional weights identical for each channel. This ensures that the model captures informative features in each frequency channel while using fewer weights for a more parsimonious model. Further measures for improving model generalizability and combating the risk of over-fitting include dropout, weight decay, and batch normalization.

4. Out of distribution detection for prediction confidence estimates

Although internal validation tests indicated good generalisation of the pitting severity model to unseen operating conditions, generalisation to unseen health states was challenging. Therefore, an out-of-distribution detection model was added, which detects unseen health states based on the final latent layers of the trained CNN. Two anomaly detection schemes were tested for out-of-distribution detection, with the One-Class SVM model (Alam, Sonbhadra, Agarwal, & Nagabhushan, 2020) generally outperforming the Local Outlier Factor model (Breunig, Kriegel, Ng, & Sander, 2000) in identifying data from unseen states. Samples flagged by this model are assigned pitting level predictions with low confidence since predictions on out-of-distribution data are expected to be inaccurate. Validation tests further indicated that the out-of-distribution detection model could detect unseen health states well under known operating conditions but was less successful under unseen operating conditions. Thus, the out-of-distribution model was used only under previously seen operating conditions.

5. Results on test and validation sets

The data competition made use of a scoring system where predictions close to the true fault severity are rewarded and predictions far away from the true fault severity are heavily penalized. Ultimately, the proposed approach achieved a competition validation score of 282.2 and a testing score of 213.3 out of a possible maximum score of 800.

6. Conclusions and Future work

The machine learning solution proposed here incorporates careful validation set design, data augmentation, preprocessing, a sophisticated CNN architecture, and uncertainty quantification through the detection of out-of-distribution samples. Ultimately, the proposed pipeline yields a robust predictive model for detecting pitting faults in gears using vibration data.

In future work, the unexpectedly superior performance of basic signal processing input features as opposed to more sophisticated approaches that rely on angular re-sampling can be investigated. Furthermore, generalization to unseen health states could be improved with improved input features that are sensitive only to fault severity. Finally, the outlier detection model could be improved to perform better on unseen operating conditions.

Acknowledgment

The authors gratefully acknowledge the support of the European Commission under the Marie Sklodowska-Curie program through the ETN MOIRA project (GA 955681). This work was supported by Flanders Make, the strategic research center for the manufacturing industry, in the context of DGTwin Prediction project and the Flemish government under the “AI Flanders” programme.

References

Alam, S., Sonbhadra, S. K., Agarwal, S., & Nagabhushan, P. (2020, May). One-class support vector classifiers: A survey. *Knowledge-Based Systems*, *196*, 105754. doi: 10.1016/j.knosys.2020.105754

Antoni, J. (2009, May). Cyclostationarity by examples. *Mechanical Systems and Signal Processing*, *23*(4), 987–1036. doi: 10.1016/j.ymssp.2008.10.010

Borghesani, P., Pennacchi, P., Randall, R., Sawalhi, N., & Ricci, R. (2013, April). Application of cepstrum pre-whitening for the diagnosis of bearing faults under variable speed conditions. *Mechanical Systems and Signal Processing*, *36*(2), 370–384. doi: 10.1016/j.ymssp.2012.11.001

Box, G. E. P., & Cox, D. R. (1964, July). An Analysis of Transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, *26*(2), 211–243. doi: 10.1111/j.2517-6161.1964.tb00553.x

Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000, May). LOF: Identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data* (pp. 93–104). New York, NY, USA: Association for Computing Machinery. doi: 10.1145/342009.335388

Fyfe, K., & Munck, E. (1997, March). Analysis of computed order tracking. *Mechanical Systems and Signal Processing*, *11*(2), 187–205. doi: 10.1006/mssp.1996.0056

Mcfadden, P., & Toozhy, M. (2000, November). Application of synchronous averaging to vibration monitoring of rolling element bearings. *Mechanical Systems and Signal Processing*, *14*(6), 891–906. doi: 10.1006/mssp.2000.1290

Rosenthal, E., & Ratna, S. (2022, March). *Spacecutter.* https://github.com/EthanRosenthal/spacecutter, https://www.ethanrosenthal.com/2018/12/06/spacecutter-ordinal regression/.

Welch, P. (1967). The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics*, *15*(2), 70–73.