DOI 10.5162/iCCC2024/P7

Predicting the Remaining Useful Life of Oscillating Bearings via Recurrent and Convolutional Neural Networks Trained on Rotating Bearings

Lukas Mattenklodt[1], Jonathan Diez[1], Antje Dittmer[1], Jens Windelberg[1]

[1] German Aerospace Center (DLR) Institute of Flight Systems, Lilienthalplatz 7, 38108 Brunswick, Germany

{lukas.mattenklodt, jonathan.diez, antje.dittmer, jens.windelberg}@dlr.de

Summary:

A Convolutional Neural Network (CNN) and a Recurrent Neural Network (RNN) are employed to predict the Remaining Useful Life (RUL) of fully rotating bearings using acceleration data. The CNN utilizes frequency domain features, while the RNN incorporates both time and frequency domain features. Initially tested on a public dataset, both models are further applied to new test bench data of oscillating bearings. The study highlights the importance of time information in RUL prediction, evidenced by the RNN's good performance compared to the CNN's poorer results for oscillating bearings.

Keywords: Machine Learning, Remaining Useful Life, Predictive Maintanance, Rolling Bearings, Oscillating Bearings

Introduction

The degradation of rolling bearings has been widely studied, leading to various methods for Remaining Useful Life (RUL) predictions. Publicly available datasets exist, e.g., the NASA Bearing Database from the University of Cincinnati, with run-to-failure test data available on Kaggle [1]. As there are no similar datasets for oscillating bearings, we adopt a two-step approach: First, applying a Convolutional Neural Network (CNN) and a Recurrent Neural Network (RNN), inspired by [2] and [3], to estimate RUL for fully rotating bearings (FRB) from [1]. Secondly, the models are evaluated, with and without re-training, on a new dataset from the German Aerospace Center, featuring oscillating bearings (OBs) [4]. Results show that the RNN, analyzing both time and frequency data, effectively predicts RUL for both types of bearings.

Methodology

Acceleration data from two testbench datasets are used in this survey. The NASA Kaggle dataset contains test runs of several FRB of the type *Rexnord ZA-2115 double row*. The rotation speed for these runs was kept constant at 2000 RPM under the application of a radial load. Acceleration was recorded every 10 minutes for 1 second at a time. The DLR bearing test bench is designed for performance degradation tests for OBs. We operated a motor-driven shaft with two spaced *SKF Explorer 3209 A* bearings, with axial force applied via disc springs adjusted by front cover screws. Axial force and shaft angular position are measured by a load cell and a rotary encoder, respectively. The housing had four acceleration sensors (three screwed, one glued) and two temperature sensors, plus an ambient temperature sensor. Note that in this study only the acceleration measurements are used to investigate the applicability of the Neural Networks (NN) models learned from the NASA dataset. The data used in this study are collected in a 120-hour life cycle assessment. The shaft rotates and applies a reference signal with constant amplitude oscillations ranging from 20° to -20° as a sequence of trapezoidal and sinusoidal trajectories with a period of 1 second under constant axial load. The shaft reference angle position as well as the measured angle of a new and a degraded bearing can be seen in figure 1. Data was recorded every 2 minutes but downsampled to 10-minute intervals to match the NASA dataset.

At the time of publication, a complete test run of a single, newly manufactured bearing under oscillating conditions was available. The moving mean square of the acceleration exceeded $2\,\mathrm{m/s^2}$ after 77 h, reaching an End of Life (EOL) criterion for small bearings, see [4]. But onto inspection of the sensor signals and taking into account that this criterion is defined for FRB, the EOL for this OB test is defined as the decrease of the maximum actual shaft position exceeding 7.5 % after 43 h.

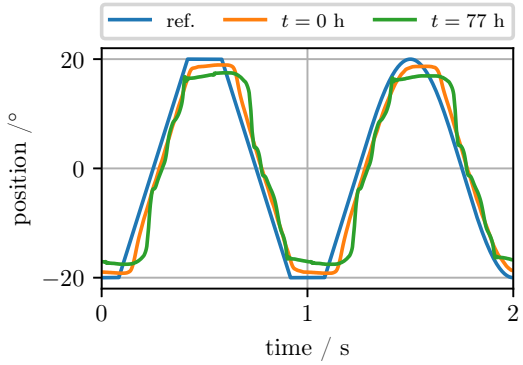

Fig. 1. DLR test bench shaft reference, measured shaft angle of a pristine ($t = 0h$) and degraded ($t = 77h$) bearing.

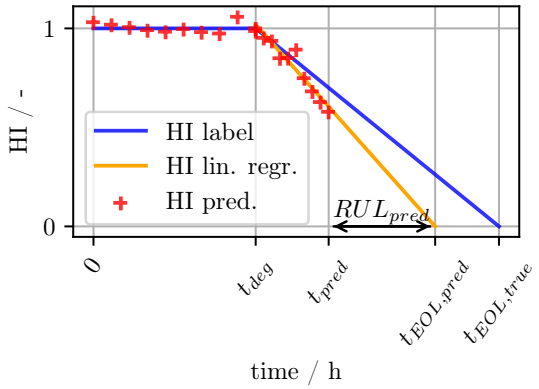

Fig. 2. Visualization of HI label and linear regression based on examplaricly predicted HIs in order to compute $t_{EOL,pred}$

To assess the transferability of AI algorithms from FRB to OB, a CNN, see [2], and an RNN, discussed in [3], are evaluated. The CNN receives a two-dimensional, squared feature map as input. This map holds the maximum amplitude values of 64 bins computed by an FFT along the first axis and the 64 time steps into the past along the second. The RNN on the other hand receives a one-dimensional input, consisting of features from the time, frequency and time-frequency domain as described in [3]. Since no code was provided by either publication, the networks are programmed, trained and validated from scratch.

Both methods aim for estimating the Health Indicator (HI), see figure 2, from the respective accelerometer data features. Following [5], HI is defined as 1 for a healthy bearing and as 0 for a bearing that has reached its EOL. In between t_{deg}, the point in time of first measurable degradation, and $t_{EOL,true}$, the HI is assumed to drop linearly. In short, at an arbitrary point in time $t = t_{pred}$, the HI is defined as

$$\mathrm{HI}(t) = \begin{cases} 1 & \text{if } t \leq t_{deg} \\ 1 - \frac{t - t_{deg}}{t_{EOL,true} - t_{deg}} & \text{if } t_{deg} < t < t_{EOL,true} \end{cases}$$

As described in [5], this definition of HI does not match the true degradation perfectly, which shows exponential behavior, but was utilized for simplicity. Also note that $t_{EOL,true}$ might not be known and be approximated using measurement criteria.Within this definition of HI, the time of EOL t_{EOL} can be conceived geometrically as the intersection of the HI with the x-axis. The actual target variable RUL, defined as the remaining time until the bearing meets the respective EOL criterion, can directly be derived from $t_{EOL,pred}$ with

$$\mathrm{RUL}(t) = t_{EOL,pred} - t$$

Note that the network only predicts the RUL after time t_{deg}, with t_{deg} identified using an Auto-Encoder as described in [6]. The two networks are firstly trained solely on FRB data. To test the transfer capabilities, these trained networks are then evaluated on OB data. After that, the networks got an additional training on OB data to get information about the adaptive potential.

Data Preparation

Features are extracted from the FRB NASA dataset without any pre-processing and NN models are trained based on [2] and [3] respectively. Since the NNs are primarily trained on FRB data it is necessary to adapt the basic structure of the OB data to them. From the 4 available sensors of the DLR test bench, only data from a single one was utilized after visual inspection. The target trapezoidal trajectory implies a discontinuous acceleration profile. Thus, exiting a broad frequency range (see [7] for a detailed explanation of signal analysis), the non-continuity excites the resonance frequencies of the whole structure connected to the motor drive shaft, i.e., both bearing and test bench. To acquire a more comprehensive understanding of the disturbance's influence on the frequency spectrum along a flank, the acceleration data are analyzed at different proximities from the flank's commencement. The obtained frequency spectra are shown in figure 3. Since the system response of the test stand was not captured, one can only guess that the peaks represent the resonance frequencies. The plot reveals that initiating the analysis later on the flank results in the resonant frequencies emerging with diminished intensity, suggesting that these frequencies are indeed excited by the impact at the beginning of the edge and then damped over time. Due to

these distinct peaks, the OB and FRB spectra differ greatly from each other, Unfortunately, this problem also affects the sinusoidal trajectories. In this case, the acceleration of the target trajectory is continuous, however, in reality the test stand does not move completely smooth (see figure 1), introducing the same difference to the spectrum. To tackle this issue, the analysis of the OB data focused on the last quarter of the flank to minimize the influence of the discontinuous acceleration. Additionally, the amplitudes of the resonance frequencies of the test stand are set to 0 in order to align the underlying character of the OB and FRB datasets. Resonance frequencies are defined as frequencies with amplitudes greater than 3 standard deviations of the spectrum. Also notable is that the motor of the test stand was not powerful enough to guarantee a deflection of up to $\pm 20°$ over the entire period. Instead, a drop in the maximum deflection can be seen over the duration of the test, see figure 5. The profile of the maximum amplitude of the system is similar to the profile of the label HI described in the previous chapter. This observation has led to the EOL criterion being based on the deflection, as the function of the assembly can no longer be guaranteed outside a certain deflection tolerance range. For this study, in the absence of a well-defined criteria for EOL of OB, the tolerance limit was set to 7.5% deviation from target trajectory, based on visual inspection of the acceleration and other available sensor data.

Training and Validation Process

The CNN topology performing best on the FRB data consists out of 3 convolutional layers with channel size 4, 8, 16 respectively and a kernel size of 2 in both dimensions followed by 6 dense layers with 200, 100, 50, 30, 8, 1 neurons respectively. All neurons used *relu* activation except the last layer. The most performant RNN consisted of 5 LSTM layers with 50, 100, 200, 100 and 50 neurons respectively. *tanh* activation was used for all layers but the last. The last layer of both networks was activated by the *sigmoid* function to clip the output to the interval (0,1). Before computing $t_{EOL,pred}$, the networks predictions are smoothed by performing a local linear regression with a window size of 50 as suggested in [2]. Both networks performed well on the validation set, see figure 4. In the linear segment of the label, both RNN and CNN achieved a coefficient of determination $R^2 > 0.8$, see table 1, indicating that the predictions correlate with the HI label. The comparison with the original papers ([2] and [3]) is made more difficult by the fact that both

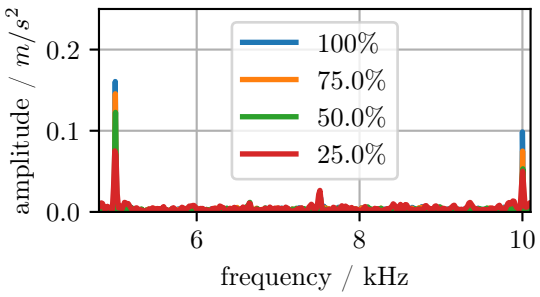

Fig. 3. Frequency spectra of acceleration data gained from the last 25%, 50%, 75% and 100% of the flank.

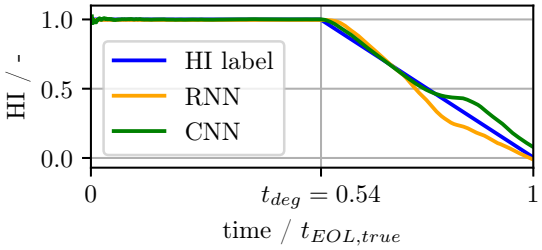

Fig. 4. CNN and RNN predictions on the FRB validation data (test run 2 of NASA data set); $t_{deg} = 89\,h$, $t_{EOL,true} = 164\,h$

papers use different metrics. However, the visual comparison confirms a similar performance.

Results

After the models are fully trained and validated exclusively on the FRB data, they are tested on the OB data. The predictions of both, the CNN and RNN, can be seen in figure 5. The difference in prediction accuracy is distinct. The CNN fails to transfer the prediction capabilities from the FRB data to the OB data and predicts noisy peaks rather than a degradation profile. The amplitudes of the peaks, however, do correlate with the maximum deflection profile, i.e. the label. The RNN on the other hand recognizes t_{deg} and predicts a degradation profile that correlates with the label. The respective R^2 values can be found in table 1. They clearly highlight this difference, confirming a close correlation between the RNN predictions and the label. Note that these predictions are based on trainings done exclusively on FRB data.

It can be seen, however, that the slope of the RNN predictions is less steep compared to the slope of the maximum amplitude. In other words, the RNN successfully identifies the degradation but fails to apply the EOL criterion. Given that the EOL cri-

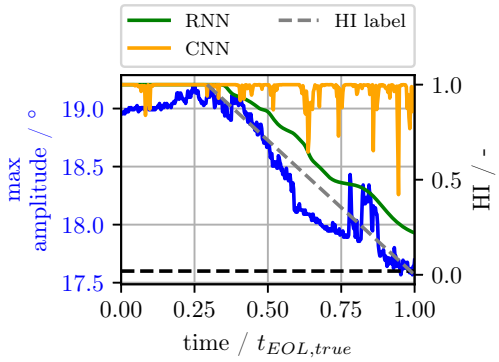

Fig. 5. CNN and RNN predictions on the OB validation data; $t_{deg} = 13\,h$, $t_{EOL,true} = 43\,h$; label marked grey

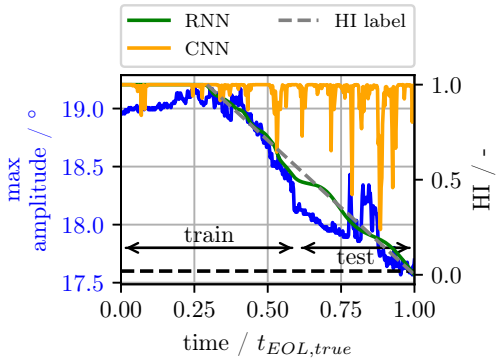

Fig. 6. CNN and RNN predictions on the OB validation data after additional training on OB data

Tab. 1. R^2 values with and without retraining on OB data

Train.: FRB, Test: FRB		**Train.: FRB, Test: OB**		**Train.: FRB&OB, Test: OB**	
R^2_{CNN}	R^2_{RNN}	R^2_{CNN}	R^2_{RNN}	R^2_{CNN}	R^2_{RNN}
0.88	0.89	0.05	0.82	0.06	0.97

terion was selected arbitrarily, this behavior is not unforeseen. In order to test whether the RNN is able to learn the defined EOL of the OB data and whether the CNN is capable of adapting to OB data in general, further trainings are conducted utilizing the OB data. Since only one entire OB test run was available at the time of publication, the first 60% of that data set was defined to be training data and the rest to be validation data. As shown in figure 6 and table 1 the RNN altered its predictions to depict the labels with an even higher accuracy. The additional training significantly enhanced prediction quality, as the RNN effectively integrated capabilities gained from both training sessions on two distinct datasets. The additional training altered the CNN predictions as well, which contains larger peaks after the second training and, in connection with that, a slightly increased R^2 value compared to before the training on OB data. However, the predictions still differ substantially from the labels.

Conclusion

This paper investigated the transferability of RNNs and CNNs in predicting the RUL of OB, utilizing training data from FRB. Both RNN and CNN showed effectiveness on FRB data, aligning with prior research. However, for OB data, the RNN outperformed the CNN, especially in managing the distinct frequency variations between OB and FRB data. Efforts to adjust CNN for these frequency discrepancies were mostly ineffective. Therefore, RNN proved to be a more versatile and transferable model in this scenario. Given the limited OB dataset, these findings are initial insights rather than conclusive assessments. Future research should focus on enhancing CNN transferability with additional inputs and performing a broader quantitative analysis with an expanded database.

References

[1] V. Tyagi, “Bearing dataset.” Kaggle, 2024.

[2] L. Ren, Y. Sun, H. Wang, and L. Zhang, “Prediction of bearing remaining useful life with deep convolution neural network,” *IEEE Access*, vol. PP, pp. 1–1, 02 2018.

[3] L. Guo, N. Li, F. Jia, Y. Lei, and J. Lin, “A recurrent neural network based health indicator for remaining useful life prediction of bearings,” *Neurocomputing*, vol. 240, 02 2017.

[4] M. A. Ismail, *Bearing Fault Detection and Quantification for Flight Control Electro-Mechanical Actuators*. Shaker Verlag, 2020.

[5] Z. Ye, Q. Zhang, S. Shao, T. Niu, and Y. Zhao, “Rolling bearing health indicator extraction and rul prediction based on multi-scale convolutional autoencoder,” *Applied Sciences*, vol. 12, no. 11, 2022.

[6] J. Diez, L. Mattenklodt, A. Dittmer, and J. Windelberg, “Anomaly detection of rotating and oscillating bearings using auto encoder,” in *Proceedings of iCampus Cottbus Conference 2024 (To appear)*, 2024.

[7] L. Ljung, *System Identification*, pp. 163–173. Birkhäuser Boston, 1998.