New hybrid method for fault detection in rolling element bearings

Yonatan Nissim [1], Renata Klein [2], Jacob Bortman [1] , Jonathan D. Rosenblatt

[1] *PHM Laboratory, Department of Mechanical Engineering, Ben-Gurion University of the Negev, P.O.B. 653, Beer-Sheva 8410501, Israel*
[2] *R.K. Diagnostics, P.O.B. 101, Gilon, D.N. Misgav 20103, Israel*

Abstract

Rolling Element Bearings (REBs) are key components in rotary machines, tend to suffer from faults causing serious damage to the whole system. Therefore, many methods have been developed, to detect faults and evaluate their degradation using vibration monitoring. However, many of them showed noisiness and lack of monotonicity over time. In this work, we present a hybrid method for fault detection in REBs. To compare and evaluate our results, we also implemented different other methods on data from an endurance test with a propagating fault in the outer race. The results show that the presented hybrid method, which including physical-aware signal processing techniques and fault-related feature extraction combined with deep learning models, can increase the reliability and interpretability of the methods. The health indicator received from the hybrid method showed better trendiness indicating the severity of the fault and improved the health track of the degrading system.

Keywords: anomaly detection, hybrid modelling, fault detection, bearing fault diagnosis ,fault severity, deep neural networks

Introduction

Rolling Element Bearings (REBs) which are key components in rotary machines are prone to failure due to high loads and rough environments. Evaluating machine health can allow the implementation of Condition Based Maintenance (CBM) to optimize maintenance schedules and prevent catastrophic events. Condition Monitoring (CM) can be achieved by analyzing measured vibration signals [1,2]. There are two main approaches for achieving CM, a physics-based and a data-driven. The main advantage of the physics-based approach is its interpretability by its explainable outputs. However, most of the methods are limited to their assumptions, and when the systems become very complex, these methods are less efficient. Data-driven methods are usually developed using historical data without any physical knowledge. A common way to develop fault detection algorithms is by using one-class classification when healthy data is served as the single class. Due to their lack of assumptions, these methods are relatively easy to develop. However, to get satisfactory results, they typically require a lot of labeled data. These issues make a hybrid method an ideal solution for exploiting both advantages and for bridging the gap between the two approaches. Although many methods have developed over the past years using physics-based [3-6], and data-driven[7-9] approaches, many of the Health Indicators (His) trend they present are lack of Monotonicity and Trendability over time. This work is focused on the HI construction process from data acquisition to HI calculation, using a hybrid method to combine the approaches, with the goal of finding a HI that reflects the physical state of the system and perform a Monotone and smooth trend.

Review of existing methods

This section is dedicated to illustrate and examine some known propositions for HI construction in REBs. The first two methods we present here involve some physics-based techniques and the other two are data-driven, which require the use of DL architectures. The first method, K-Nearest-Neighbors (KNN) of physical fault features, was introduced in [3]. The method uses faulty bearing simulations to find the most informative frequency bands in the measured signals. After extracting Spectral Kurtosis (SK), from each selected frequency band, KNN algorithm is then used to measure the similarity between the measured signal and the healthy data. The second method, Norm of Condition Indicators(CIs), is presented in [4]. It uses the bearing tones energy (e.g, BPFO, BPFI, FTF, BSF) as CIs and uses their norm as an HI. The norm is calculated after applying whitening process for the extracted data. The third method uses a Deep Learning (DL) architecture called Auto-Encoder. AE architecture includes an encoder and a decoder. The encoder gets as an input vector in the frequency domain and passes it through a bottleneck of hidden layers to the decoder, which tries to reconstruct the original signal into the output. After the network has been trained on healthy data, a faulty signal that will pass through the AE will get a larger value of Reconstruction Error (RE). This RE is serve as the HI of the signal. The last method uses Generative Adversarial Network (GAN). GAN architecture is composed of two different subnetworks, a Generator (G) and a Discriminator (D) unit. During the training process, these units are acting against each other. The purpose of G is to generate from a random noise a fake signal in the frequency domain which needs to be similar to a real signal. Meanwhile, D aims to distinguish between real signals and fake signals generated from G. At the end of the training process, D can serve as a detector, and therefore the HI in this method is defined as HI = 1-D(X).

Experimental setup

The data used in this work was obtained from our own endurance test experiment. During the experiment, a spall has developed in the outer race of the tested bearing. The experimental conditions were a radial load of 3.3 kN and a rotation speed of 35 Hz. The vibration was collected every 20 minutes for a period of 1 minute in a 50 kHz sampling rate. Speed rotation and environmental temperature were also collected during the experiment. The total number of records was 2404. Records 50-150 were used as a training set while the first 50 records were used as a validation set. The other records were left for examination of the trend of the HI. For each method, the threshold for alert was decided to be the 99th percentile of the HI in the training data.

Existing methods results

Figure 1 shows the results of the four implemented methods. In this plot, we normalized the output of methods between 0-1. As can be observed, at the end of the experiment all the methods clearly detect the developed fault. On the other hand, all the HI trends of the methods behave undoubtedly differently, although they intend to solve the same problem. Among all the methods, the GAN has shown the most different trend with extreme overfitting.

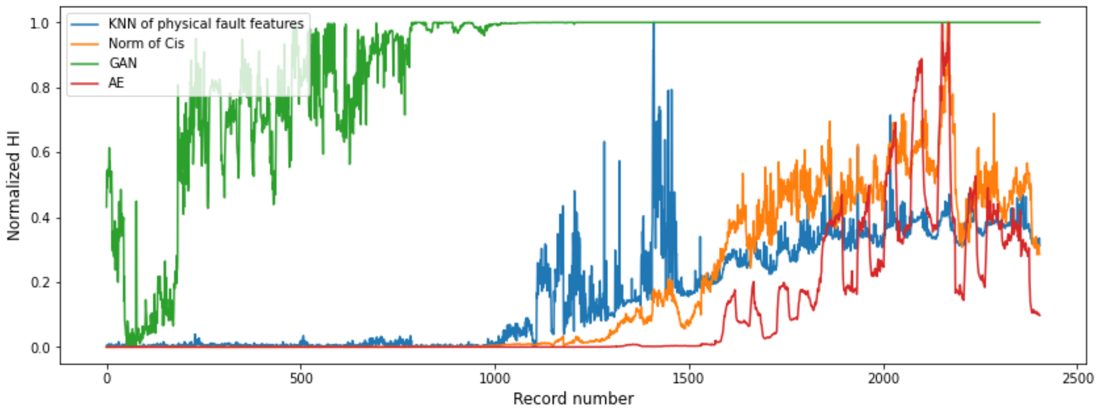

Fig. 1: HI trends over the whole experiment of the four implemented methods.

Furthermore, we saw some problems with these methods. Some of the methods showed a decreasing trend over time which contradicts the physical basic assumption that the system's health cannot be improved over time. Some have shown a noisy trend which may leads to an inaccurate prognostic estimation and decrease the reliability of the degradation assessment. The data-driven methods also show instability in their solution convergence which is expressed in different trends at different runs. These mentioned issues require the use of a reliable method that creates a monotone and smoothed HI that will eventually result in a better prognostic estimation

A new hybrid method

Figure 2 illustrates our proposed hybrid method. In contrary to the naïve data-driven approach, the hybrid method consists physics-based data pre-processing stage and data extraction stage, before applying a data-driven algorithm. The proposed method exploits both the physical knowledge regarding the system and the capabilities of a data-driven algorithm to learn a complex mechanism behavior such as vibration signals.

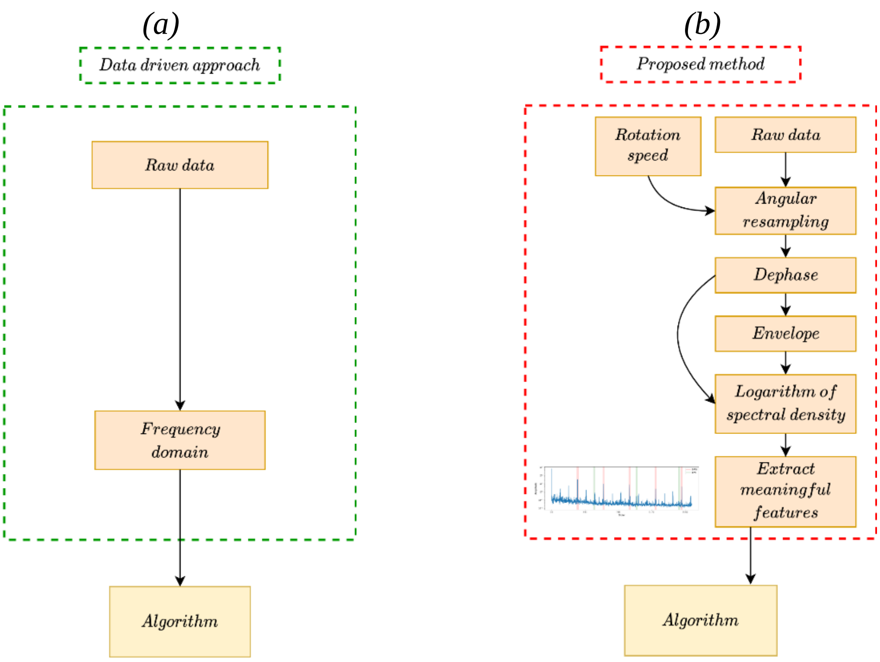

Fig. 2: (a) Naïve data-driven approach scheme; (b) The new hybrid method scheme.

In terms of fault detection for mechanical systems, physics-based pre-processing techniques can improve signal-to-noise in order to accentuate the physical phenomena that arise in the data when a fault occurs in the system. Here we suggest some signal processing techniques that may help to emphasize these phenomena. The angular-resempling technique aims to reduce the effect of the fluctuations in rotation speed during the operation of the machine. Dephasing technique, clean the data from the synchronous rotating components in the system and allowing to monitor more clearly the asynchronous components i.e bearings. The envelope technique emphasizes the physical cyclic impulses that emerge when a fault is appear in the bearing. Finally, the spectrum domain of the data is been calculated for the feature extraction phase.
Bearing tones are the fault-related frequencies that should arise when a fault exists in the bearing. Thus, for the feature extraction phase, we suggest taking the bearing tones and their corresponding harmonics as meaningful features. Moreover, due to noise and resolution errors, it may be more informative to take these specific frequencies with their surrounding bands. In our method, we used 10 harmonics for each fault type and $\pm$ 0.08 [Order] as the surrounding band for each selected frequency. This clean and informative data is then inserted as an input for the AE and the GAN, which will be denoted as Hybrid-AE and Hybrid-GAN respectively.

Hybrid-AE and Hybrid-GAN results

Figure 3 shows the trends derived from the Hybrid-AE and Hybrid-GAN. In the figure, the HIs are normalized between 0 to 1. In comparison to the original AE and GAN, the trends are looking smoother over time and with fewer significant declines. As expected, the hybrid method has improved the shape of the trend and its trendability for both algorithms.

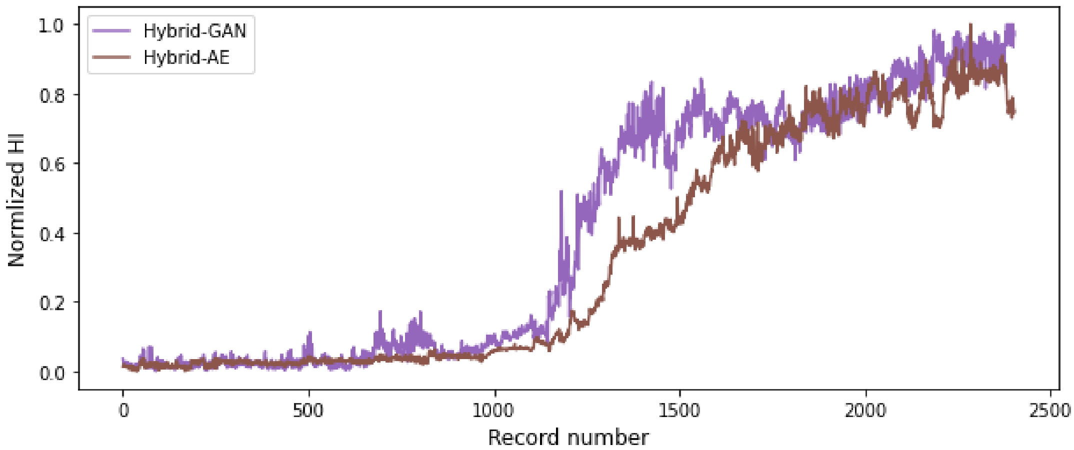

Fig. 3: HI trends over the whole experiment of the GAN and AE, after applying physical pre-processing and fault-related feature extraction.

To evaluate the performance of the methods we conducted a quantitative comparison. Table 1 summarizes the results from all the methods.

Table 1: Quantitative comparison between the methods. Noisiness - MSE with rolling mean; Monotonicity - Spearman correlation with time; FPR - rate of false positives.

Method	Noisiness	Monotonicity	FPR (%)
KNN of physical fault features (Tian et al., 2015)	0.0023	0.882	7.0
Norm of Cis (Bechhoefer and Schlanbusch, 2018)	0.0011	0.941	0.3
AE	0.0027	0.954	0
GAN	0.0030	0.958	83.3
Hybrid-AE	0.0004	0.980	1.7
Hybrid-GAN	0.0010	0.960	0

The noisiness of the trend is measured by the Mean Squared Error (MSE) of the normalized HI with the rolling mean of it. As much as the deviations from the rolling mean are small (i.e smoothed trend), MSE score is smaller. The Monotonicity is measured by Spearman's rank correlation coefficient of the HI with the time vector. To estimate the FPR we need to assume the period in which the system is healthy. Considering that none of the methods except for GAN showed significant changes in the first 400 records, we took records 0-50 and 150-400 as healthy records to evaluate the FPR. Because the threshold is set to be the 99th quantile of the training data, a result of around 1% is expected. Table 1 shows that the physics-based methods showed smoother trends but less monotone than the data-driven. Furthermore, Hybrid-GAN and Hybrid-AE improved all the measures except for the FPR of the Hybrid-AE, which still got a reasonable result of 1.7%.

Conclusions & summary

In both qualitative and quantitative comparisons, the hybrid approach has shown improvements indicating that hybrid thinking may lead to more accurate diagnosis and prognosis. These improvements enable a better understanding of the trend propagation and may result in a better remaining useful life estimation, which is one of the most important goals in prognostics. The physical preprocessing and the extraction of the meaningful areas in the spectrum as features for the algorithm can improve the signal-to-noise ratio of the data, by focusing on the meaningful areas (i.e. bearing tones) but reducing the dimensionality of the data by omitting areas in the spectrum that do not carry information related to the bearing. To conclude, the physics knowledge and data-driven advanced techniques being effective individually, but they strengthen each other and make a strong and promising combination in the future.

References

1. Malla, C., & Panigrahi, I. (2019). Review of condition monitoring of rolling element bearing using vibration analysis and other techniques. *Journal of Vibration Engineering & Technologies*, *7*(4), 407-414.

2. Randall, R. B. (2021). *Vibration-based condition monitoring: industrial, automotive and aerospace applications*. John Wiley & Sons.

3. Tian, J., Morillo, C., Azarian, M. H., & Pecht, M. (2015). Motor bearing fault detection using spectral kurtosis-based feature extraction coupled with K-nearest neighbor distance analysis. IEEE Transactions on Industrial Electronics, 63(3), 1793-1803.

4. Bechhoefer, E., & Schlanbusch, R. (2018, September). Calculating remaining useful life in an embedded system. In Annual Conference of the PHM Society (Vol. 10, No. 1).

5. Gazizulin, D., Kogan, G., Klein, R., & Bortman, J. (2015, March). Towards a physics based prognostic model for bearing-spall initiation and propagation. In *2015 IEEE Aerospace Conference* (pp. 1-10). IEEE.

6. Zhang, B., Sconyers, C., Byington, C., Patrick, R., Orchard, M. E., & Vachtsevanos, G. (2010). A probabilistic fault detection approach: Application to bearing fault detection. *IEEE Transactions on Industrial Electronics*, *58*(5), 2011-2018.

7. Xu, F., & Tse, Y. L. (2018). Roller bearing fault diagnosis using stacked denoising autoencoder in deep learning and Gath–Geva clustering algorithm without principal component analysis and data label. *Applied Soft Computing*, *73*, 898-913.

8. Vakharia, V., Gupta, V. K., & Kankar, P. K. (2015). Ball bearing fault diagnosis using supervised and unsupervised machine learning methods. *Int J Acoust Vib*, *20*(4), 244-250.

9. Sutrisno, E., Oh, H., Vasan, A. S. S., & Pecht, M. (2012, June). Estimation of remaining useful life of ball bearings using data driven methodologies. In *2012 ieee conference on prognostics and health management* (pp. 1-7). IEEE.