A Machine Learning Suite for Machine Components' Health-Monitoring

Ramin M. Hasani[1*]**, Guodong Wang**[1*]**, and Radu Grosu**[1]

[1]Cyber Physical Systems (CPS), Technische Universität Wien (TU Wien), 1040 Vienna, Austria

* Equal Contributions

Abstract

This paper studies an intelligent technique for the health-monitoring and prognostics of common rotary machine components, with regards to bearings in particular. During a run-to-failure experiment, rich unsupervised features from vibration sensory data are extracted by a trained sparse auto-encoder. Then, the correlation of the initial samples (presumably healthy), along with the successive samples, are calculated and passed through a moving-average filter. The normalized output which is referred to as the auto-encoder correlation based (AEC) rate, determines an informative attribute of the system, depicting its health status. AEC automatically identifies the degradation starting point in the machine component. We show that AEC rate well-generalizes in several run-to-failure tests. We demonstrate the superiority of the AEC over many other state-of-the-art approaches for the health monitoring of machine bearings.

1 Introduction

Machine health management plays a key role in almost all industries. A key subsidiary of health management is condition based monitoring (CBM) where one prognoses abnormal status of a machine based on extracted features from a group of implemented sensors and parameters. The CBM procedure, therefore, includes two steps; 1) Feature extraction during a run-to-failure experiment and 2) Data processing for predicting the degradation starting point and monitoring the defect propagation during the test.

Numerous methods have been proposed for the prognostic of key machine components. In many cases, handcrafted time- and frequency-domain features are derived from the sensors mounted on the machine are used directly or post-processed by numerous methods in order to predict the condition of the machine component (Lee et al. 2014). Recently, on the other hand, artificial intelligence (AI) solutions have been vastly utilized in fault classification and condition monitoring (Jia et al. 2016; Sun et al. 2016; Thirukovalluru et al. 2016). AI techniques significantly enhance the quality of the feature extraction and data processing. However, they have not yet realized a fully-automated method without the use of prior knowledge for health monitoring tests.

The main attributes of an ideal health-condition monitoring and prognostic method are described as follows: The method is able to autonomously collect useful unsupervised features from the available sensory data, regardless of its size. Such data-driven procedures should enable users to monitor both online, and offline tests, while providing an intelligent trend on the machine's health, as well as identify its degradation starting point. It is highly desirable that the method automatically combine the steps of the CBM process and be entirely human-labor independent. Moreover, the ideal approach is universal and can be feasibly applied to the prognoses of various key machine components such as bearings, gears and spindles.

In the present study, we propose a novel prognostic method for machine bearings, as a critical machine component which reasonably satisfies the main characteristics of an ideal CBM method. The technique is called autoencoder correlation-based (AEC) prognostic algorithm. We train a sparse autoencoder for extracting unsupervised features from collected sensory data in several test-to-failure experiments and correspondingly, compute the Pearson correlation of the extracted features of the initial samples, with the upcoming samples. The output is then passed through a moving average (MA) filter. AEC algorithm then normalizes the output of the filter and accurately illustrates the health condition of the system. We evaluate the performance of our algorithm over several run-to-failure tests of machine bearings and prove its superiority in finding the degradation starting point compared to the existing methods.

2 Related Works

Comprehensive reviews on useful prognostic methods for industrial machines including bearings have been proposed (Jardine, Lin, and Banjevic 2006; Lee et al. 2014). Traditionally, time domain statistical features such as root mean squared (RMS), Kurtosis, and Spectral Kurtosis have been utilized for monitoring the status of key machine components (Qiu et al. 2006). Although such features can capture regularities in the data, in low signal to noise ratios (SNR) environments with high dimensional prediction properties, they face difficulties concerning predicting the status of the system under test. Unsupervised feature extraction techniques such as principle component analysis (PCA)-based methods (He, Kong, and Yan 2007), followed by a post-processing stage,

such as hidden Markov model (HMM) for the health degradation monitoring, provided a reasonable prediction on the status of the system (Yu 2012). Moreover, the combination of the frequency-domain feature extraction methods such as *wavelet packet decomposition* with HMM are also employed in the calculation of the remaining useful life (RUL) of the key machine components (Tobon-Mejia et al. 2012). Tobon et al. proposed a useful trend on the RUL and successfully tested it on few bearing run-to-failure experiments. The approach is however subjected to modifications to be used in other test-beds (Tobon-Mejia et al. 2012).

Kalman filter (KF) has also been used in condition monitoring (Wang et al. 2016). Reuben et al. proposed a method based on switching KF which can be simultaneously employed in the diagnosis and prognosis of the data (Reuben and Mba 2014). Wang et al. deployed an enhanced KF with expectation-maximization (EM) algorithm for providing accurate estimations on the health condition of the machines (Wang et al. 2016).

Data-driven approaches using AI techniques have brought significant advancements to the *diagnostic* and health condition monitoring of key machine components. Various neural network architectures, specifically autoencoders, successfully classify faults into varying subgroups, through the use of pre-processed statistical time or frequency domain features, and features in general extracted using prior knowledge (Jia et al. 2016; Sun et al. 2016; Thirukovalluru et al. 2016). Although the proposed methods are not fully automated, they are less human-labor dependent. Although data-driven technologies show their success through their accurate diagnoses and fault classifications (Jia et al. 2018; Zhao et al. 2018), such methods have not been yet utilized in *prognoses*. Here, for the first time, we train a sparse autoencoder directly over the vibration data and compute the correlation of the extracted features and therefore, provide a comprehensive *prognostic* and health monitoring method which its performance is superior compared to many of the remarked approaches.

3 Autoencoder Correlation based (AEC) Prognostic Method

In this section, we describe the working mechanism of the automated AEC fault prognostic method, in the run-to-failure settings of machine bearings. Figure 1 symbolically illustrates the structure of the AEC. An autoencoder (AE) network is directly trained over the vibration data-samples. The autoencoder generates rich nonlinear features from the sensory data for each raw input sample. Afterward, the correlation coefficient matrix of the past samples up to the current sample is computed. The correlation rate of the features corresponding to the samples generated at the beginning of the run-to-failure process, with the other available samples, is then normalized and correspondingly filtered through a moving average (MA) filter. The normalization step also linearly broadens the AEC rate spectrum to provide a better distinction for the faulty samples. This feature provides an output rate which predicts the status of the system at each sampling step. We thoroughly sketch our design procedures below.
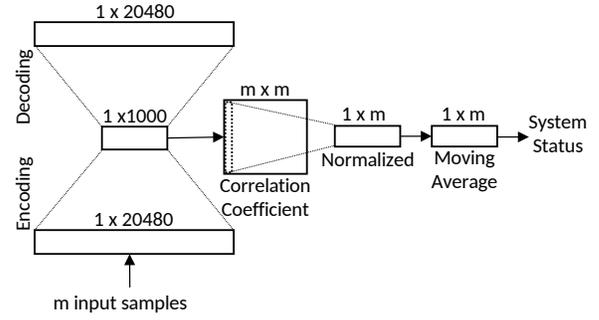

Figure 1: The architecture of the AEC.

Sparse Autoencoder Revisit

An autoencoder (Boureau, Cun, and others 2008), tries to learn an abstract of the identity function, to estimate the same input patterns at its output. One can place constraints on the network by limiting the size of the hidden layer and presumably discover attractive features from the data.

Let us define $x \in \mathbb{R}^{D_x}$, a D_x-dimension input to the autoencoder; the encoder initially maps x to a lower dimension vector $z \in \mathbb{R}^D$, and correspondingly generates an estimation of x, $\hat{x} \in \mathbb{R}^D$:

$$z = f(Wx + b_1); \quad \hat{x} = f(W^T z + b_2), \tag{1}$$

where f, in our design, is a saturating linear transfer function denoted in Equation 2 , $W \in \mathbb{R}^{D_x \times D}$ stands for the weight matrix, and vectors $b_1 \in \mathbb{R}^{\mathbb{D}}$, $b_2 \in \mathbb{R}^{\mathbb{D}_\frown}$ represent the bias values.

$$f(z) = \begin{cases} 0, & \text{if } z \leq 0 \\ z, & \text{if } 0 < z < 1 \\ 1, & \text{if } z \geq 1 \end{cases} \tag{2}$$

We define the following cost function to be optimized subsequently:

$$C = \underbrace{\frac{1}{N} \sum_{n=1}^{N} \sum_{i=1}^{I} (x_{in} - \hat{x}_{in})^2}_{\text{mean squared error}} + \underbrace{\lambda R_{L2}}_{L_2 \text{ regularization}} + \underbrace{\sigma R_{sparse}}_{\text{sparsity regularization}} . \tag{3}$$

The first term in C is the mean squared error. The second term denotes an L_2 regularization. The third term in the cost function equation determines a sparsity regularization, which is formulated by KL divergence, with an effectiveness coefficient, σ.

We then train the network by applying a scaled conjugate gradient (SCG) algorithm. Note that we chose a vanilla sparse-autoencoder topology since its performance over several test environments has been shown to be adequate to output useful features from raw vibration sensory data. Depending on the properties of the test environment, one can deploy many other architectures such as variational AE (Kingma and Welling 2013).

Correlation Analysis, Normalization and Filtering

For each training sample, D nonlinear features are constructed by the autoencoder. We calculate the linear dependencies of the abstract representation of each sample by computing the Pearson correlation coefficient (Fisher 1925). The correlation coefficient matrix for the available samples during the run-to-failure test is then calculated. The first column of the CC matrix depicts the correlation of the first sample data which is recorded at the beginning of the run-to-failure process, with the other available samples. We normalize this vector between zero and one, and define it as the criteria for predicting the degradation starting point and consequently determining the health status of the system.

We finally, smoothen the shape of the output by passing it through a moving average filter. The filter is designed as follows:

$$\hat{m} = \frac{1}{w_{size}}(y(n)+y(n-1)+\cdots+y(n-(w_{size}-1))). \quad (4)$$

For the sample data y, the filter slides a sample-window of length w_{size}, over the data, and calculates the average of the covered data in each window.

4 Experiments with AEC

In this section, we evaluate the performance of our fault prognostic method by employing it in several run-to-failure experiments on bearings. We initially introduce the dataset contents together with the objective of the tests and illustrate the performance of the AEC method in various run-to-failure scenarios. We finally benchmark our results with the state-of-the-art methods in fault prognostic of the bearing machines.

IMS Bearing Dataset from PCoE NASA Datasets

We use the bearing dataset provided by the Center for Intelligent Maintenance Systems (IMS), University of Cincinnati (Lee et al. 2007), collected from the Prognostics Data Repository of NASA (PCoE Accessed 2017). In the setup, a shaft is coupled to an AC motor and is rotating at a speed of 2000 RPM while a 6000 lbs load is installed on it. Four force-lubricated bearings are mounted on the shaft. Accelerometers with high sensitivity are placed for each bearing for recording the vibrations. Three test-to-failure experiments are performed independently. In such tests, failures usually happened at the end of the test (Qiu et al. 2006).

In the first experiment, two accelerometers are utilized for each bearing while in the second and third experiments one accelerometer is used. Datasets contain one-second recordings from the accelerometers with a sampling frequency of $20 KHz$, every 10 min, during the run-to-failure tests (Qiu et al. 2006). Table 1 represents the properties of the collected data in each experiment. For our simulations, we only train the AE network on the faulty bearings; therefore, we have four different experiment settings as 1) Dataset 1 bearing 3 (S1B3), 2)Dataset 1 Bearing 4 (S1B4), 3) Dataset 2 bearing 1 (S2B1) and 4) Dataset 3 Bearing 3 (S3B3).

Table 1: IMS Bearing tests specification

Experiment	# of Samples	sample size	Faulty Bearing	test-to-failure time
S1	2156	4×20480	B3 and B4	35 days
S2	984	4×20480	B1	8 days
S3	4448	4×20480	B3	31 days

Results

AEC method is implemented in MATLAB. The motivation was to make AEC compatible with the common industrial production systems and especially online test-to-failure environmental tools, which are forming a direct interface with MATLAB. The training process is performed on a Microsoft Azure NC-Series virtual machine powered by one NVIDIA Tesla K80 GPU. We demonstrate our method's performance in two general frameworks. In the first setup, we train the autoencoder with all the available data, for each experiment, in order to *monitor the status* of the system. In the second framework, we train the autoencoder with 70% of the data and keep 30% of it for testing in order to investigate the *prediction performance* of the proposed method. The training time varies between 55 to 80 minutes for each test-bed, depending on the number of the available input samples.

Since we only consider the vibration samples of one bearing in each simulation, the input samples' dimension, under 20 kHz sampling frequency, is a 20480-length vector. We choose 1000 hidden units for the AE which enables us to extract 1000 features from each large input vector. The size of the latent space is selected, empirically.

We then calculate the correlation coefficient matrix of the input samples (depending on the framework in which we are working on, for the first case we feed in all the available data while for the second framework we dedicate 70% of the data for training and the network has to output a prediction on the status of the system based on the previously observed samples) and normalize the correlation rate of the initial sample (which is considered to represent the health-status of the system) with the succeeding samples. We then filter out the output and provide a representation of the status of the system.

Framework1 - Health-Condition Monitoring   Figure 2 illustrates the output of the AEC for the four test environments of the three run-to-failure experiments, where we monitor the recorded data. A high AEC rate (Red), corresponds to a healthy behavior of the system while a downward trend depicts the starting point of an abnormal state. High AEC indicates more correlated samples with the initial health-level of the system. AEC clearly displays the beginning of a faulty trend together with its propagation effect. AEC can even develop a reasonably accurate prognosis for the noisy dataset, S3B3, whereas many other approaches cannot.

Figure 2A to 2D, graphically indicates the status of the four experiments S1B3, S1B4, and S2B1 and S3B3, respectively. The Color bar represents the AEC rate from 0 (blue) to 1 (red). We can distinctly observe where a significant change in the AEC rate has occurred, and stop the process accordingly. We call a sample abnormal when its AEC rate is measured 90% below the recorded sample at $t-100$. The AEC del-

Table 2: Detection performance. HMM-DPCA: Hidden Markov model with dynamic PCA, HMM-PCA: Hidden Markov model with PCA (Yu 2012). MAS-Kortusis: Moving average spectral kurtosis (Kim et al. 2016). VRCA: Variable-replacing-based contribution analysis (Yu 2012). – means that the dataset has not been analyzed for the corresponding method

Algorithm	S1B3	S1B4	S2B1	S3B3
	Degradation starting data-point	Relative detection accuracy		
AEC	2027 | 94.5%	**1641 | 99.0%**	547 | 98.1%	**2367 | 99.0%**
HMM-DPCA	2120 | 90.3%	1760 | 94.5%	539 | 98.8%	–
HMM-PCA	–	1780 | 93.5%	**538 | 99.0%**	–
RMS	2094 | 91.5%	1730 | 95.9%	539 | 98.8%	No detection
MAS-Kurtosis	**1910 | 99.0%**	1650 | 98.5%	710 | 82.5%	No detection
VRCA	–	1727 | 96.0%	–	No detection

Table 3: Prediction accuracy of the AEC method in different bearing run-to-failure tests

Experiment	Fault starting point	Prediction Accuracy
S1B3-sensor1	2120	95.68%
S1B3-sensor2	2122	95.59%
S1B4-sensor1	1681	98.14%
S1B4-sensor2	1673	98.51%
S2B1	610	93.60%
S3B3	2435	98.47%

icately indicates the start of a faulty status and represents its gradual propagation. Table 2 summarizes the detection performance of our AEC method. It also provides the performance of some of the existing monitoring approaches, for the IMS dataset. Detection performance is defined as the sample-time (sample number) at which the algorithm notices initiation of the degradation. Therefore, early fault detection results in better performance. The relative detection accuracy is computed as the difference of the detected degradation starting point (DSP) of a method for a particular dataset, to the best DSP amongst all the other compared methods. The corresponding relative accuracy of the DSP detection is provided in Table 2. We can observe that only AEC provides a reasonable DSP for all four experiments. AEC did not provide the best performance in all the monitoring experiments. An earlier detection on the initiation of a fault in the system, for the S1B3 experiment, is produced by the moving average spectral kurtosis (MAS-Kurtosis) method (Kim et al. 2016). Such high performance is achieved because in some special conditions, frequency domain features provide a rich representation on the abnormal fluctuations in the system, though not generalizable to all scenarios, as it is shown in Table 2.

AEC's performance for the S1B4 experiment is superior over the other methods. Degradation of the system starts relatively early in this dataset, and the vibration data on the bearing realizes a nonlinear plateau of fluctuations as it is depicted in Figure 2D. Depending on the sensitivity of the methods, the sample point at which the fault starts to prop-
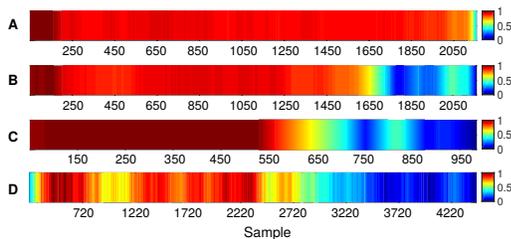

Figure 2: Visualization of the status of four different bearings in four run-to-failure experiments. A) S1B3. B) AEC for S1B4 C) AEC for S2B1. D) AEC for S3B3.

agate is detected. MAS-Kurtosis technique also provides a high-performance prediction on this dataset.

In the experiments of S2B1, the performance of a hidden Markov model equipped with a principle component analysis (HMM-PCA) (Yu 2012), slightly outperforms the other methods. The reason is the existence of linear dependencies in the recorded vibration data and the status of the system. This matter can potentially lead a more non-linear approximation method to an over-fitting region. However, since the performance of other methods is also reasonably high, the over-fitting factor in this test-case is not considerable.

S3B3's test-to-failure process is considered a challenge to monitor, as very few approaches can accurately monitor its state, while none could provide the degradation starting point, whereas AEC provides reasonable detection on it.

Framework2 - Online Prognostic In the second framework, we train the autoencoder over the first 70% of the available sensory data in order to evaluate the prediction performance of the model in an online monitoring setting. Here, we study six cases including S1B3-sensor1, S1B3-sensor2, S1B4-sensor1, S1B4-sensor2, S2B1 and S3B3 where the predicted AEC rate in each experiment, is represented in Figure 3A to 3F, respectively. A sample is defined as the initiation of an abnormal state (equivalent to DSP), where its AEC rate reaches 90% of the rate of the 100 steps earlier sample. The prediction process starts from the samples collected from day five on since before that, the status of the system has been considered to be normal (Qiu et al. 2006). By using this analogy, the degradation starting point is calculated with a high level of accuracy, and the propagation of fault is captured during the simulated run-to-failure test. Table 3 summarizes the predicted degradation starting point together with the prediction accuracy in each experiment. The prediction error is computed as the ratio of the difference between the predicted sample and the monitored DSP, to the total number of samples in each experiment.

The AEC produces a reasonable degree of prediction, as shown in Figures 3A to 3F. The surprising results here is that even the *prediction accuracy* of the AEC in a prognostic setup, for some datasets, outperforms the *detection performance* of the other methods in the monitoring framework. For instance, in S1B4, the 1681th sample is predicted as the DSP, which is an earlier detection than the DSP provided by the HMM-DPCA, HMM-PCA and the VRCA methods. Note that providing a metric for computing the prediction accuracy is only possible in an *online* prognostic and monitoring setting. Other methods discussed in this report have not offered an online measurement of the DSP in their context.

Therefore, our method stands alone in the online monitoring phase, and we are not able to benchmark our approach with the other technologies.

We observed highly correlated activity and a reasonable degree of consistency when comparing the performance of AEC in the monitoring framework (shown in Figure 2) as well as in the prognostic framework (illustrated in Figure 3).

Qualitative Comparison of AEC with the Other Prognostic Methods for the Machine Bearing   In this section, we compare AEC with many existing methods for *status monitoring and prognostic technologies*, over several performance criteria. Such qualifications are determined as follows: *Generalizability:* The ability of the process to detect and predict faulty trends in a system for various machine bearing testbeds. *Status Monitoring:* The ability of the method to provide a useful health-status trend during the run-to-failure experiment of bearings. *Automated:* A fully autonomous fault prognostic method. *Unsupervised:* Capability of the technique to extract information from raw sensory data without any prior knowledge and supervision. *Detection Sensitivity:* Ability of the method to provide a reasonably quick predictive alert on the degradation starting point in the test. *Fault-type Detection (Diagnostics):* Ability of the technique to detect a particular type of defect in the system and classify them into different fault classes.

Table 4 comprehensively illustrates a qualitative comparison among various fault prognostic methods utilized for bearings, based on the mentioned attributes. The assessment on the performance of each method is carefully performed based on the provided results and the detailed specifications of the methods in their corresponding report. Under such evaluations, results suggest the superiority of the AEC algorithm over many existing methods in the *Prognosis* of the machine bearings, where it can precisely capture the initial degradation point and provide a useful trend for the spread of a defect, automatically. AEC, however, lacks the *diagnostic* feature, to specify the type of fault that occurs in the system. A classifier has to be incorporated and trained over supervised data to capture the *diagnostic* attribute as well.

Table 4: Qualitative comparison of the performance of the existing approaches on the bearing prognostic. HMM-DPCA: Hidden Markov model with dynamic PCA, HMM-PCA: Hidden Markov model with PCA (Yu 2012). MAS-Kortusis: Moving average spectral kurtosis (Kim et al. 2016). VRCA: Variable-replacing-based contribution analysis (Yu 2012). EET: Energy Entropy trend (Kim et al. 2016). WPSE-EMD: Wavelet packet sample Entropy (Wang et al. 2011) - Empirical mode decomposition (Lei et al. 2007). Spectral-ANN: Third-order spectral + artificial neural networks (Yang et al. 2002). Fuzzy-BP: Fuzzy logic with back-propagation (Satish and Sarma 2005). SVM: Support Vector Machine (Yang, Zhang, and Zhu 2007). GA-SVR: Genetics algorithm-Support vector regression (Feng et al. 2009). GLR-ARMA: Generalized likelihood ratio - Autoregressive moving average (Galati, Forrester, and Dey 2008). ++: Highly satisfies. +: Satisfies. -: The attribute is not covered -+: The attribute is fairly covered.

Algorithm	Generalizability	status monitoring	Automated	unsupervised	Detection sensitivity	fault-type detection
AEC	+	+	+	+	++	-
HMM-DPCA	-+	+	-	+	+	-
HMM-PCA	-	+	-	+	+	-
RMS	-	+	-	-	-+	-
Kurtosis	-	+	-	-	-+	-
MAS-Kurtosis	-	+	-	-	-+	-
Spectral-ANN	-+	-	-	-	+	+
VRCA	-+	+	-	-	++	-
EET	+	+	-	+	-	-
WPSE+EMD	-	+	-	-	+	-
Fuzzy-BP	-+	-	+	-	+	+
SVM	-	-	-	-	+	++
GA-SVR	-	+	-	-	+	-
GLR-ARMA	-	+	-	-	+	-+

5   Conclusions

We introduced a machine learning suite for fault *prognosis* in machine bearings. We showed that AEC successfully monitors the status of the bearings in various experiments which confirms its generalization ability. AEC also accurately predicts the degradation starting point while providing an informative trend on the defect's propagation. Furthermore, the AEC algorithm generated rich unsupervised features from raw vibration input data, automatically. In the future, we intend to apply AEC to the prognostic of the other key machine components such as gears, cutting tools and spindles.
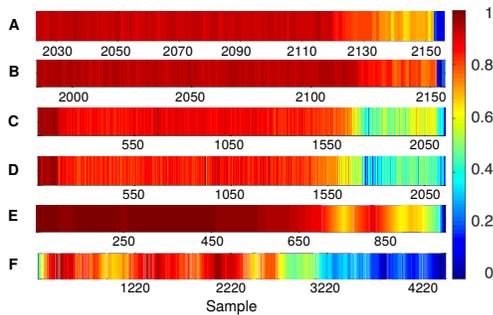
Acknowledgments

Figure 3: Online monitoring of the status of the system in different experiments: A) S1B3 first sensor B) S1B3 second sensor C) S1B4 first sensor D) S1B4 second sensor. E) S2B1 F) S3B3. The color-bar represents the AEC rate.

References

[Boureau, Cun, and others 2008] Boureau, Y.-l.; Cun, Y. L.; et al. 2008. Sparse feature learning for deep belief networks. 1185–1192.

[Feng et al. 2009] Feng, F. Z.; Zhu, D. D.; Jiang, P. C.; and Jiang, H. 2009. Ga-svr based bearing condition degradation prediction. In *Key engineering materials*, volume 413, 431–437. Trans Tech Publ.

[Fisher 1925] Fisher, R. A. 1925. *Statistical methods for research workers*. Genesis Publishing Pvt Ltd.

[Galati, Forrester, and Dey 2008] Galati, F. A.; Forrester, D.; and Dey, S. 2008. Application of the generalised likelihood ratio algorithm to the detection of a bearing fault in a helicopter transmission. *Australian Journal of Mechanical Engineering* 5(2):169–176.

[He, Kong, and Yan 2007] He, Q.; Kong, F.; and Yan, R. 2007. Subspace-based gearbox condition monitoring by kernel principal component analysis. *Mechanical Systems and Signal Processing* 21(4):1755–1772.

[Jardine, Lin, and Banjevic 2006] Jardine, A. K.; Lin, D.; and Banjevic, D. 2006. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing* 20(7):1483–1510.

[Jia et al. 2016] Jia, F.; Lei, Y.; Lin, J.; Zhou, X.; and Lu, N. 2016. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mechanical Systems and Signal Processing* 72:303–315.

[Jia et al. 2018] Jia, X.; Jin, C.; Buzza, M.; Di, Y.; Siegel, D.; and Lee, J. 2018. A deviation based assessment methodology for multiple machine health patterns classification and fault detection. *Mechanical Systems and Signal Processing* 99:244–261.

[Kim et al. 2016] Kim, S.; Park, S.; Kim, J.-W.; Han, J.; An, D.; Kim, N. H.; and Choi, J.-H. 2016. A new prognostics approach for bearing based on entropy decrease and comparison with existing methods. *Annual Conference of the Prognostics and Health management Society*.

[Kingma and Welling 2013] Kingma, D. P., and Welling, M. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

[Lee et al. 2007] Lee, J.; Qiu, H.; Yu, G.; Lin, J.; et al. 2007. Bearing data set. *IMS, University of Cincinnati, NASA Ames Prognostics Data Repository, Rexnord Technical Services*.

[Lee et al. 2014] Lee, J.; Wu, F.; Zhao, W.; Ghaffari, M.; Liao, L.; and Siegel, D. 2014. Prognostics and health management design for rotary machinery systems-reviews, methodology and applications. *Mechanical systems and signal processing* 42(1):314–334.

[Lei et al. 2007] Lei, Y.; He, Z.; Zi, Y.; and Hu, Q. 2007. Fault diagnosis of rotating machinery based on multiple anfis combination with gas. *Mechanical systems and signal processing* 21(5):2280–2294.

[PCoE Accessed 2017] PCoE. Accessed 2017. PCoE, Prognostics Center of Excellence of NASA: The prognostics data repository. `https://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/`.

[Qiu et al. 2006] Qiu, H.; Lee, J.; Lin, J.; and Yu, G. 2006. Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. *Journal of sound and vibration* 289(4):1066–1090.

[Reuben and Mba 2014] Reuben, L. C. K., and Mba, D. 2014. Diagnostics and prognostics using switching kalman filters. *Structural Health Monitoring* 13(3):296–306.

[Satish and Sarma 2005] Satish, B., and Sarma, N. 2005. A fuzzy bp approach for diagnosis and prognosis of bearing faults in induction motors. In *Power Engineering Society General Meeting, 2005. IEEE*, 2291–2294. IEEE.

[Sun et al. 2016] Sun, W.; Shao, S.; Zhao, R.; Yan, R.; Zhang, X.; and Chen, X. 2016. A sparse auto-encoder-based deep neural network approach for induction motor faults classification. *Measurement* 89:171–178.

[Thirukovalluru et al. 2016] Thirukovalluru, R.; Dixit, S.; Sevakula, R. K.; Verma, N. K.; and Salour, A. 2016. Generating feature sets for fault diagnosis using denoising stacked auto-encoder. In *Prognostics and Health Management (ICPHM), 2016 IEEE International Conference on*, 1–7. IEEE.

[Tobon-Mejia et al. 2012] Tobon-Mejia, D. A.; Medjaher, K.; Zerhouni, N.; and Tripot, G. 2012. A data-driven failure prognostics method based on mixture of gaussians hidden markov models. *IEEE Transactions on Reliability* 61(2):491–503.

[Wang et al. 2011] Wang, F.; Zhang, Y.; Zhang, B.; and Su, W. 2011. Application of wavelet packet sample entropy in the forecast of rolling element bearing fault trend. In *Multimedia and Signal Processing (CMSP), 2011 International Conference on*, volume 2, 12–16. IEEE.

[Wang et al. 2016] Wang, Y.; Peng, Y.; Zi, Y.; Jin, X.; and Tsui, K.-L. 2016. A two-stage data-driven-based prognostic approach for bearing degradation problem. *IEEE Transactions on Industrial Informatics* 12(3):924–932.

[Yang et al. 2002] Yang, D.-M.; Stronach, A.; MacConnell, P.; and Penman, J. 2002. Third-order spectral techniques for the diagnosis of motor bearing condition using artificial neural networks. *Mechanical systems and signal processing* 16(2-3):391–411.

[Yang, Zhang, and Zhu 2007] Yang, J.; Zhang, Y.; and Zhu, Y. 2007. Intelligent fault diagnosis of rolling element bearing based on svms and fractal dimension. *Mechanical Systems and Signal Processing* 21(5):2012–2024.

[Yu 2012] Yu, J. 2012. Health condition monitoring of machines based on hidden markov model and contribution analysis. *IEEE Transactions on Instrumentation and Measurement* 61(8):2200–2211.

[Zhao et al. 2018] Zhao, M.; Jia, X.; Lin, J.; Lei, Y.; and Lee, J. 2018. Instantaneous speed jitter detection via encoder signal and its application for the diagnosis of planetary gearbox. *Mechanical Systems and Signal Processing* 98:16–31.