

# Outlier Modeling in Gear Bearing Using Autoencoder for Remaining Useful Life Prediction

Sunny Singh \*

Department of Electronics and Communication Engineering,  
National Institute of Technology,  
Bhopal, India  
sunnysinghnitb@gmail.com

Praneet Shiv

Department of Electronics and Communication Engineering,  
National Institute of Technology,  
Bhopal, India  
praneettshivv@gmail.com

Atif Ahmed

Department of Electronics and Communication Engineering,  
National Institute of Technology,  
Bhopal, India  
atif.ahmed.1120@gmail.com

**Abstract:** In this paper, we introduce the Prognostics and Health Management of gear bearing system using autoencoder neural networks. Bearings and gears are the most common mechanical components in rotating machines, and their health conditions are of great concern in practice. This study presents an outlier modeling method for forecasting the gear bearing system failure using the health indicators constructed from mechanical signal processing and modeling. Outlier modeling aims to find patterns in data that are significantly different from what is defined as normal. In the unsupervised outlier modeling setting, prior labels about the anomalousness of data points are not available. In such cases, the most common techniques for scoring data points for outlyingness include distance-based methods density-based methods, and linear methods. The conventional outlier modeling methods have been used for a long time to detect anomalous observations in data. However, this paper shows that autoencoders are a very competitive technique compared to other existing methods. The developed method is demonstrated using the IMS bearing data from NASA Acoustics and Vibration Database.

**Keywords:** autoencoder; prognostics and health management; outlier modeling; remaining useful life; gear bearing

## I. Introduction

Gear bearing condition monitoring and diagnostics has received considerable attention for many years because gear bearings are critical to almost all forms of rotating machinery [1], [2] and are among the most common machine elements. Bearing failure is one of the foremost causes of breakdowns in rotating machinery and such failure can be catastrophic, resulting in costly downtime. To prevent unexpected bearing failure, Prognostics and Health Management (PHM) has been used extensively for examining the gear bearing health conditions. One of the key issues in gear bearing prognostics is to detect the defect at its incipient stage and alert the operator before it develops into a catastrophic failure. Therefore, predicting their health condition is necessary to prevent any unexpected accidents caused by gear bearing failures.

PHM [3], [4] is an emerging discipline to scientifically manage the health condition of engineering systems and their critical components, which has attracted much attention from engineers and scholars in recent years [5]–[9]. PHM is mainly concerned with three aspects: construction of health indicators, remaining useful life (RUL) prediction, and health management. Health indicators aim to evaluate the current health condition of an engineering system and its critical components, which is then used to infer their RUL [10], [11]. Based on the first two aspects, the optimal health management schedule is planned to minimize costs and prevent unexpected accidents [12]–[14].

In this paper, we introduce the outlier modeling using autoencoder neural networks to predict the health condition of a gear bearing system. The term “outlier” has come from the field of statistics, wherein outlier modeling has been studied for a long time. Outliers are also referred to as an “anomaly” in the literature. The most quoted definition of outlier comes from Hawkins’ 1980 book: “An outlier is an observation which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism” [15]. Detecting outliers is important because anomalous data often implies negative or even destructive consequences. Detecting and then removing outliers can improve the performance of classification, clustering and regression algorithms because even a single anomalous value can significantly bias these algorithms. For example, Chen recently showed that a single anomalously smooth exemplar will condemn semi-supervised time series classification algorithms to fail [16].

Similarly, an autoencoder is a special type of multi-layer neural network that performs hierarchical and nonlinear dimensionality reduction of the data. Typically, the number of nodes in the output layer is the same as the input layer, and the architecture is layered and symmetric. The goal of an autoencoder is to train the output to reconstruct the input as closely as possible. The nodes in the middle layers are smaller in number, and therefore the only way to reconstruct the input is to learn weights so that the intermediate outputs of the nodes in the middle layers represent reduced representations. Fig. 1 illustrates a fully connected autoencoder.

Since the autoencoder creates a reduced representation of the data, it is a natural approach for discovering outliers. The basic idea here is that outliers are much harder to be accurately represented in this form than the inliers (or normal points). Therefore, on reconstructing an outlier, the error will be large. This provides a natural way to score a data point. Nonlinear dimensionality reduction methods such as spectral transformations [17] have recently been explored in the literature with some success. In this light, it is somewhat surprising that the success with neural networks has been limited. An important issue is that the outliers are often themselves included within the training model. As a result, overfitting becomes increasingly likely. This is one of the reasons that neural networks have not achieved much success in spite of the known success of other dimensionality reduction methods in outlier detection. Autoencoder ensemble learning methods [18] present a natural solution to address this dilemma.

Outlier modeling aims to find patterns in data that are significantly different from what is defined as normal. One of the challenges of outlier modeling is the lack of labeled examples, especially for the anomalous classes. We describe an autoencoder neural network-based approach to detect anomalous instances in the IMS bearing data from NASA Acoustics and Vibration Database [19]. In this work, we train the net to build a model of the normal examples, which is then used to predict the class of previously unseen instances based on the reconstruction error. The input to this network is also the desired output. The results demonstrate that the proposed method is promising for the outlier modeling of the gear bearing system.

The remainder of this paper is organized as follows. We will discuss related works in Section 2. Section 3 discusses our proposed autoencoder method for outlier detection. Section 4 discusses the experimental results, while the conclusions and future works are presented in Section 5.

## 2. Related Works

The problem of outlier modeling has been studied widely in the community [20]. Numerous methods such as distance-based methods [21]–[23], density-based methods [24], linear methods [25],

and spectral methods [26, 17] have been proposed. Recently, ensemble methods have found an increasing interest in the literature [18, 27, 20]. Several ensemble methods such as feature bagging [28], subspace histograms [29], high-contrast subspaces [30], and locally relevant subspaces [31, 32] have been proposed. The spectral methods in [26, 17] can also be viewed as nonlinear dimensionality reduction methods that reduce the data representation in a nonlinear way in order to score data points as outliers.

Outlier modeling has been used for a long time to detect and, where appropriate, remove anomalous observations from data. Outliers arise due to mechanical faults, changes in system behavior, fraudulent behavior, human error, instrument error or simply through natural deviations in populations. The original outlier detection methods were arbitrary but now, principled and systematic techniques are used, drawn from the full gamut of Computer Science and Statistics.

A number of distinguished engineers and scholars have also conducted reviews of RUL prediction. Heng et al. [11] summarised conventional reliability models, condition-based prognostic models, and their hybrid models. Ye and Xie [33] summarised a number of degradation models and comprehensively compared stochastic process models with general path models. Si et al. [34], [35] discussed various prognostic methods based on statistical modeling. Lee et al. [36] clarified the relationship between machine diagnostics and prognostics and then summarised many prognostic methods for predicting the RUL of critical components such as gear bearings. Zhang and Lee [37] reviewed prognostic methods for rechargeable lithium-ion batteries, which are also potentially useful for predicting the RUL of machines, especially gear bearings. The main difference between battery prognostics and gear bearing prognostics is that the health status of rechargeable lithium-ion batteries can be quantified and described by the battery capacity, which is calculated by integrating the battery current over time in the process of discharging. However, for gear bearing prognostics, it is rare to discover a simple and direct health indicator to track the current health condition.

In this paper, we introduce autoencoders for unsupervised outlier modeling. One problem with neural networks is that they are sensitive to noise and often require large data sets to work robustly while increasing data size makes them slow. As a result, there are only a few existing works in the literature on the use of neural networks in outlier modeling. Experimental results comparing the proposed approach with state-of-the-art detectors are presented on the IMS bearing data set showing the robustness of our approach.

### 3. Theory of Autoencoders

An autoencoder is a type of artificial neural network used to learn efficient data codings in an unsupervised manner. The aim of an autoencoder is to learn a representation (encoding) for a set of data, typically for dimensionality reduction. Along with the reduction side, a reconstructing side is learned, where the autoencoder tries to generate from the reduced encoding a representation as close as possible to its original input.

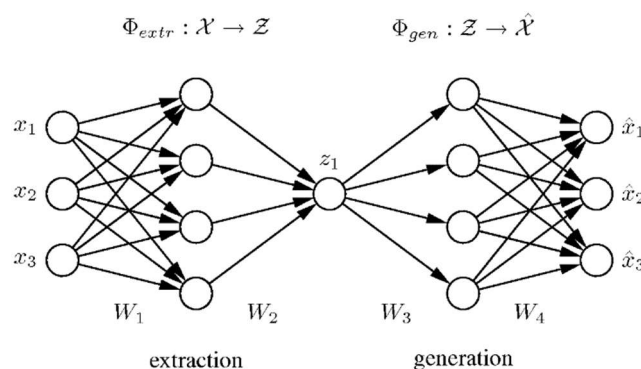


Figure 1. An autoencoder representation.

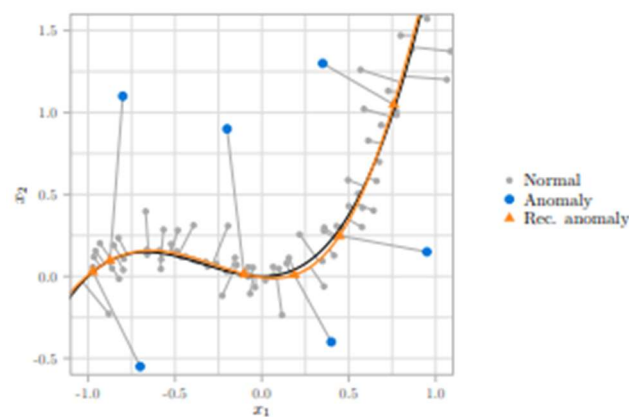
Outliers are data points that differ significantly from the remaining data. In the unsupervised outlier modeling setting, prior labels about the anomalousness of data points are not available. In

such cases, the most common techniques for scoring data points for outlyingness include distance-based methods, density-based methods, and linear methods. An overview of different outlier detection algorithms may be found in [20]. The basic approach in neural networks is to use a multi-layer symmetric neural network to reconstruct (i.e, replicate) the data. The reconstruction error is used as the outlier score.

The motivation for the ability of autoencoders to detect outliers is based on two observations on reconstructions and outliers. First, all reconstructions must lay on the reconstruction manifold and this manifold follows the noise-free relations in the data. Second, outliers are rare and deviate from the general pattern in the data. Using these we can formulate the following main reasons that drive the ability of autoencoders to detect outliers:

1. outliers are not projected orthogonally onto the reconstruction manifold,
2. outliers often have a larger distance to the reconstruction manifold than normal observations

The first reason holds because outliers are in general not well represented in the training data and the autoencoder has therefore not learned to map them to the closest point on the reconstruction manifold. The projection to the closest point corresponds to an orthogonal projection. In general, the projections of outliers on the reconstruction manifold will thus not be orthogonal.



**Figure 2.** Again, we see points generated by the same model but now six anomalous points are also depicted, along with their reconstructions. Reconstructions of normal points have been omitted for clarity.

The second holds for outliers that do not conform to the general pattern in the data in that they deviate more from the noise-free relations between the variables, compared to normal observations. Because the reconstruction manifold does follow these relations, the outliers have a larger distance to the reconstruction manifold than normal observations.

Both reasons lead to a high reconstruction error, which forms the basis of the outlier factor (OF), the metric used to distinguish normal observations from aberrant ones. Following Hawkins et al. (2002b) [38], we define the outlier factor of the  $i^{\text{th}}$  observation to be the average squared reconstruction error over all  $P$  features:

$$OF_i = \frac{1}{P} \sum_{p=1}^P (x_{i,p} - f_p(x_i | \theta))^2$$

In Fig. 2, we see an example of an autoencoder applied for outlier modeling. The six outliers, depicted as blue points, have a larger distance to the reconstruction line than normal points because they show a large deviation from the noise-free relations between  $x_1$  and  $x_2$ . Moreover, the projections onto the manifold are not orthogonal, indicating that the learned identity mapping is not optimal in a least-squares setting for these observations.

The length of the line that connects each point to its reconstruction in Euclidean space is the reconstruction error [38]. These reconstruction errors are larger for outliers in comparison with

normal observations and can thus be used to identify outliers, where large errors indicate that a data point might be corrupted.

In this paper, we employ multi-layer perceptron neural networks with one or more hidden layers, and the same number of output neurons and input neurons, to model the data. In this model, the input variables are also the output variables so that this forms an implicit, compressed model of the data during training. A measure of the outlyingness of individuals is then developed as the reconstruction error of individual data points. This approach has linear analogues in Principal Components Analysis [39]. The insight exploited in this paper is that the trained neural network will reconstruct some small number of individuals poorly and these can be considered as outliers. We measure outlyingness by ranking data according to the magnitude of the reconstruction error. This compares to SmartSifter [40] which similarly builds models to identify outliers but scores the individuals depending on the degree to which they perturb the model.

The selection of the number of hidden neurons has a couple of implications. If it is too large, the system will be over-specified. Conversely, if it is too small, the system can become overgeneralized and therefore poorly infers specific cases. We find that the choice of hidden unit quantity significantly affects the technique accuracy. Interestingly, further experiments on different data sets show that having the number of hidden units equal to the input-output units consistently yields good detection rate even though depending on the data set, it may not be the optimum architecture.

Architecturally, the simplest form of an autoencoder is a feedforward, non-recurrent neural network very similar to the many single layer perceptrons which makes a multilayer perceptron (MLP) – having an input layer, an output layer and one or more hidden layers connecting them – but with the output layer having the same number of nodes as the input layer, and with the purpose of reconstructing its own inputs [39].

In the context of outlier modeling and condition monitoring, the basic idea is to use the autoencoder network to “compress” the sensor readings to a lower-dimensional representation, which captures the correlations and interactions between the various variables.

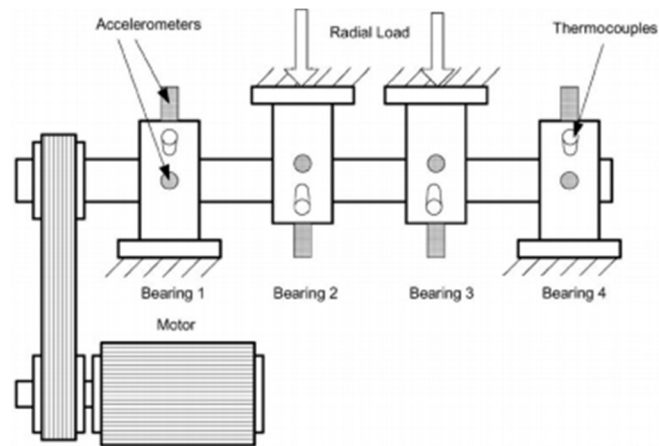
The autoencoder network is then trained on data representing the normal operating state, with the goal of first compressing and then reconstructing the input variables. During the dimensionality reduction, the network learns the interactions between the various variables and should be able to reconstruct them back to the original variables at the output. The main idea is that as the monitored equipment degrades, this should affect the interaction between the variables (e.g. changes in temperatures, pressures, vibrations, etc.). As this happens, one will start to see an increased error in the networks reconstruction of the input variables. By monitoring the reconstruction error, one can thus get an indication of the health of the monitored equipment, as this error will increase as the equipment degrades. We, then, use the probability distribution of the reconstruction error to identify whether a data point is normal or anomalous.

#### 4. Experimental Results

Any machine, whether it is a rotating machine or a non-rotating machine will eventually reach a point of poor health. This signals that there might be a need for some maintenance activity to restore the full operating potential. In simple terms, identifying the health state of the equipment is the domain of condition monitoring [13]. The most common way to perform condition monitoring is to look at the sensor measurements (i.e., data from vibration sensors, temperature sensors, and rotational speed sensors, among others) from the machine and to impose a minimum and maximum value limit on it. If the current value is within the bounds, then the machine is healthy. If the current value is outside the bounds, then the machine is unhealthy and an alarm is sent.

This procedure of imposing hard coded alarm limits is known to send a large number of false alarms, that is, alarms for situations that are actually healthy states for the machine. There are also missing alarms, that is, situations that are problematic but are not alarmed. Hence, the health of a complex piece of equipment cannot be reliably judged based on simple Statistical analysis [20]. That is why we have presented an autoencoder neural network-based approach in this paper.

The vibration signals used in this paper were provided by the Center for Intelligent Maintenance Systems (IMS), University of Cincinnati, USA, in collaboration with the National Aeronautics and Space Administration (NASA). A schematic of the experimental test rig is shown in Fig. 3. Four Rexnord ZA-2115 double row bearings are installed on the shaft. Each bearing contains 16 rollers (for each row), a pitch diameter of 2.815 in., a roller diameter of 0.331 in., and a tapered contact angle of 15.171 [19].



**Figure 3.** Bearing test rig.

Three sets of tests were made. Each set is an experiment of 4 bearings. In this way, 12 bearings are used but only 4 bearings have reached failure with known defects. Each data set describes a run-to-failure experiment. It consists of individual files that are 1-second vibration signal snapshots recorded at specific intervals (every 10 min). Each file consists of 20,480 points with the sampling rate set at 20 kHz. The rotation speed was kept constant at 2000 RPM by an AC motor coupled to the shaft via rub belts. A radial load of 6000 pounds is applied onto the shaft and bearing by a spring mechanism. All bearings are force lubricated. Records (row) in the data ASCII files are data points. Data collection is provided by NI DAQ Card 6062E [41]. In this paper, only the bearing 1 of the second ending with an outer race defect is used as shown in Fig. 4.



**Figure 4.** Outer race defect of bearing 1 of testing 2.

Our goal is to detect gear bearing degradation and give a warning that allows for predictive measures to be taken in order to avoid a gear bearing failure.

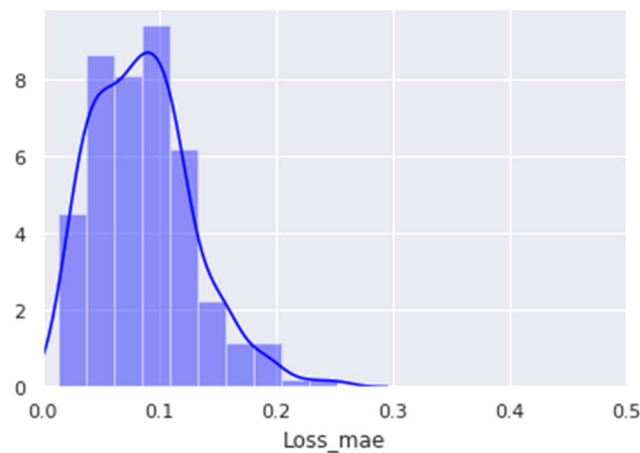
Generating an outlier modeling model involves training a neural network and then finding a suitable threshold. The methodology is:

1. Generate a training set of  $N$  normal examples.
2. Generate a validation set of  $M$  normal examples.
3. Create a feed-forward network with random initial weights. The number of units in the input-output layer is equal to the number of variables in the data set. The number of hidden neurons is determined empirically.

4. Use back-propagation to train the network. Training ceases when the error on the validation set begins to rise.
5. Choose the maximum error in training to be the threshold [39].

Three sets of data each consisting of four bearings were run to failure under constant load and running conditions. The vibration measurement signals are provided for the datasets over the lifetime of the bearings until failure. The failure occurred after 100 million cycles with a crack in the outer race [19]. As the equipment was run until failure, data from the first two days of operation was used as training data to represent normal and healthy equipment. The remaining part of the datasets for the time leading up to the bearing failure was then used as test data, to evaluate whether the different methods could detect the bearing degradation in advance of the failure.

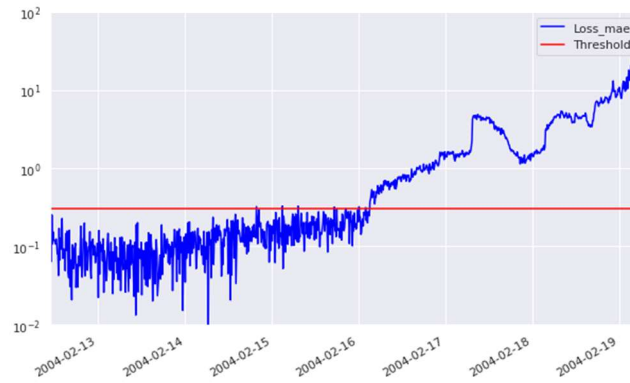
This approach consisted of using an autoencoder neural network to look for outliers (as identified through an increased reconstruction loss from the network). We also here use the distribution of the model output for the training data representing “healthy” equipment to detect outliers. The distribution of reconstruction loss (mean absolute error) for the training data is shown in Fig. 5.



**Figure 5.** Distribution of reconstruction loss for “healthy” equipment.

Using the distribution of the reconstruction loss for healthy equipment, we can define a threshold value for what to consider an outlier. From the distribution above, we can define a loss  $> 0.25$  as an outlier. The evaluation of the method to detect equipment degradation now consists of calculating the reconstruction loss for all data points in the test set and comparing the loss to the defined threshold value for flagging this as an outlier. Using the above approach, we calculate the reconstruction loss for the test data in the time period leading up to the bearing failure, as illustrated in Fig. 6.

In Fig. 6, the blue points correspond to the reconstruction loss, whereas the red line represents the defined threshold value for flagging an outlier. The bearing failure occurs at the end of the dataset. This illustrates that this modeling approach was able to detect the upcoming equipment failure about 3 days ahead of the actual breakdown (where the reconstruction loss crosses the threshold value).



**Figure 6.** Predicting bearing failure.

## 5. Conclusions and Future Works

In this paper, we studied the relationship between health indicators and RUL prediction in the framework of PHM and pointed out that health indicators are the key to RUL prediction. We have shown how autoencoders can be used for outlier modeling. The number of units in the input and output layer corresponds to the number of data attributes. Only normal instances are used for training. We use this model to develop a score for outlyingness where the trained model is applied to the whole data set to give a quantitative measure of the outlyingness based on the reconstruction error. The output of the training process is a predictive model and a corresponding threshold value. Our method not only gives a ranked estimation of the anomalous degree of each instance but also provides an outlier label for direct decision-making. Our approach takes the view of letting the data speak for itself without relying on too many assumptions.

The experiment result demonstrates that using an autoencoder with only one hidden layer is a promising approach for outlier modeling. Even though the optimum number of hidden neurons is dependent on the data dimensionality, we have managed to narrow down to an optimum range. We suggest having this number slightly fewer than, equal to or slightly higher than the number of input-output units are all reasonably good options. When an exhaustive search is impossible, we recommend using the same number of units for all three layers. This finding is somewhat surprising because the intuition is that the number of units in the hidden layer should be smaller than that of the two outer layers to enable data compression and helps the network generalizes unseen examples.

With the reduced cost of capturing data through sensors, as well as the increased connectivity between devices, being able to extract valuable information from data is becoming increasingly important. Finding patterns in large quantities of data is the realm of machine learning and statistics, and there are huge possibilities to harness the information hidden in these data to improve performance within several different domains. Moreover, an outlier modeling system provides you with a real-time interpretation of data activity. Outlier modeling and condition monitoring, as covered in this paper, are just one of many possibilities. The long-term learning potential of these outlier modeling systems puts them in a constant state of evolution. The more experience these tools develop, the more potent they will become. In the future, this will not just result in quicker response times but better insights as well.

Another important future work would be to investigate other alternatives for choosing a threshold value. For large data sets, we plan to investigate whether increasing the number of hidden layers can help in improving performance. We also would like to have the method tested on more data sets of different dimensions and application domains. This paper will be helpful for designing further advanced gear bearing health indicators and provides a basis for predicting the remaining useful life of gear bearings. The proposed technique improves significantly over the traditional methods for outlier modeling. Furthermore, it is also competitive with respect to state-of-the-art methods.



## References

- [1] R. B. Randall, *Vibration-Based Condition Monitoring Industrial, Aerospace and Automotive Applications*. Chichester, U.K.: Wiley, 2011.
- [2] R. X. Gao and R. Yan, *Wavelets: Theory and Applications for Manufacturing*. Cham, Switzerland: Springer, 2010
- [3] M. Pecht, *Prognostics and Health Management of Electronics*, 1st ed. London, U.K.: Wiley, 2008.
- [4] K. Goebel, M. Daigle, S. Abhinav, S. Sankararaman, I. Roychoudhury, and J. Celaya, *Prognostics: The Science of Making Predictions*. New York, NY, USA: CreateSpace, 2017.
- [5] N.-H. Kim, D. An, and J.-H. Choi, *Prognostics and Health Management of Engineering Systems: An Introduction*. Cham, Switzerland: Springer, 2016.
- [6] G. Niu, *Data-driven Technology for Engineering System Health Management: Design Approach, Feature Construction, Fault Diagnosis, Prognostics, Fusion and Decisions*. Cham, Switzerland: Springer, 2016.
- [7] Y. Lei, *Intelligent Fault Diagnosis and Remaining Useful Life Prediction of Rotating Machinery*, 1st ed. Amsterdam, The Netherlands: Elsevier, 2016.
- [8] J. Han, *Machine Learning and Knowledge Discovery for Engineering Systems Health Management*. London, U.K.: Chapman & Hall, 2011.
- [9] Y. Peng, M. Dong, and M. J. Zuo, "Current status of machine prognostics in condition-based maintenance: A review," *Int. J. Adv. Manuf. Technol.*, vol. 50, nos. 1–4, pp. 297–313, 2010.
- [10] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1483–1510, 2006.
- [11] A. Heng, A. C. C. Tan, J. Mathew, N. Montgomery, D. Banjevic, and A. K. S. Jardine, "Intelligent condition-based prediction of machinery reliability," *Mech. Syst. Signal Process.*, vol. 23, no. 5, pp. 1600–1614, 2009.
- [12] V. Makis and A. K. S. Jardine, "A note on optimal replacement policy under general repair," *Eur. J. Oper. Res.*, vol. 69, pp. 75–82, Aug. 1993.
- [13] D. Tang, V. Makis, L. Jafari, and J. Yu, "Optimal maintenance policy and residual life estimation for a slowly degrading system subject to condition monitoring," *Rel. Eng. Syst. Safety*, vol. 134, pp. 198–207, Feb. 2015.
- [14] M. Marseguerra, E. Zio, and L. Podofillini, "Condition-based maintenance optimization by means of genetic algorithms and Monte Carlo simulation," *Rel. Eng. Syst. Safety*, vol. 77, no. 2, pp. 151–165, 2002.
- [15] Hawkins, D.M.: *Identification of outliers*, vol. 11. Springer (1980).
- [16] Chen, Y., Hu, B., Keogh, E., Batista, G.E.: *Dtw-d: time series semi-supervised learning from a single example*. In: *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 383–391. ACM (2013)
- [17] S. Sathe, C. Aggarwal. *LODES: Local Density Meets Spectral Outlier Detection*. *SDM*, 2013.
- [18] C. Aggarwal. *Outlier Ensembles: Position Paper*, *ACM SIGKDD Explorations*, 2013.
- [19] *IMS bearings data set*; 2017. <http://data-acoustics.com/measurements/bearing-faults/bearing-4/>
- [20] C. Aggarwal, S. Sathe. *Outlier Ensembles: An Introduction*. Springer, 2017.
- [21] F. Angiulli, C. Pizzuti. *Fast outlier detection in high dimensional spaces*, *PKDD*, 2002.
- [22] E. Knorr, and R. Ng. *Algorithms for Mining Distance based Outliers in Large Datasets*. *VLDB*, 1998.
- [23] S. Ramaswamy, R. Rastogi, K. Shim. *Efficient Algorithms for Mining Outliers from Large Data Sets*. *SIGMOD*, 2000.
- [24] M. Breunig, H.-P. Kriegel, R. Ng, J. Sander. *LOF: Identifying Density-based Local Outliers*, *SIGMOD*, 2000.

- [25] M. Shyu, S. Chen, K. Sarinnapakorn, L. Chang. A novel anomaly detection scheme based on principal component classifier. ICDMW, 2003.
- [26] X. Dang, B. Misenkova, I. Assent, R. Ng. Outlier detection with space transformation and spectral analysis. SDM, 2013.
- [27] C. Aggarwal, S. Sathe. Theoretical Foundations and Algorithms for Outlier Ensembles. ACM SIGKDD Explorations, 2015.
- [28] A. Lazarevic, V. Kumar. Feature Bagging for Outlier Detection, KDD, 2005.
- [29] S. Sathe, C. Aggarwal. Subspace Outlier Detection in Linear Time with Randomized Hashing. ICDM, 2016.
- [30] F. Keller, E. Muller, K. Bohm. HiCS: High-Contrast Subspaces for Density-based Outlier Ranking. ICDE, 2012.
- [31] E. Muller, M. Schiffer, T. Seidl. Statistical Selection of Relevant Subspace Projections for Outlier Ranking. ICDE, 2011.
- [32] E. Muller, I. Assent, P. Iglesias, Y. Mülle, K. Bohm. Outlier Ranking via Subspace Analysis in Multiple Views of the Data, ICDM, 2012.
- [33] Z.-S. Ye and M. Xie, "Stochastic modelling and analysis of degradation for highly reliable products," *Appl. Stochastic Models Bus. Ind.*, vol. 31, no. 1, pp. 16–32, Jan. 2015.
- [34] X.-S. Si, W. Wang, C.-H. Hu, and D.-H. Zhou, "Remaining useful life estimation—A review on the statistical data driven approaches," *Eur. J. Oper. Res.*, vol. 213, pp. 1–14, Aug. 2011.
- [35] X.-S. Si, Z.-X. Zhang, and C.-H. Hu, *Data-Driven Remaining Useful Life Prognosis Techniques: Stochastic Models, Methods and Applications*. Cham, Switzerland: Springer, 2017
- [36] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel, "Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications," *Mech. Syst. Signal Process.*, vol. 42, nos. 1–2, pp. 314–334, 2014.
- [37] J. Zhang and J. Lee, "A review on prognostics and health monitoring of Li-ion battery," *J. Power Sour.*, vol. 196, no. 15, pp. 6007–6014, 2011.
- [38] Hawkins, Williams, G., Baxter, R., He, Hongxing, S., and Gu, L. (2002a). A comparative study of rnn for outlier detection in data mining. In *Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on*, pages 709712. IEEE.
- [39] R. Hecht-Nielsen. Replicator neural networks for universal optimal source coding. *Science*, 269(1860-1863), 1995.
- [40] K. Yamanishi, J. Takeuchi, G. Williams, and P. Milne. On-line unsupervised outlier detection using finite mixtures with discounting learning algorithm. In *Proceedings of KDD2000*, pages 320–324, 2000.
- [41] J. Ben Ali, N. Fnaiech, L. Saidi, B. Chebel-Morello, F. Fnaiech, "Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals," *Applied Acoustics*, 89, pp 16-27, 2015