Developing a Hybrid Expert/Data-Driven Health Index for Railway Axleboxes Using Auto-encoder Neural Networks

Alexandre Trilla
Services – R&D
ALSTOM
Barcelona, Spain
alexandre.trilla@alstomgroup.com

Fahd Janjua
Services
ALSTOM
London, United Kingdom
fahd.janjua@alstomgroup.com

Sergi Bermejo
Services – R&D
ALSTOM
Barcelona, Spain
sergi.bermejo@alstomgroup.com

***Abstract*—The railway axlebox is a critical safety component in rolling stock. It bears the weight of the train, and minimises the friction with the rotating axle. Its failure in service might cause derailment. Therefore, its maintenance is decisive. This article presents a methodology to create a standard Health Index that blends the hard threshold-based expert criteria developed at the workshop, with a record of actual vibration data, in order to increase the flexibility of its diagnostic and provide a soft granular feedback. The proposed method makes use of an auto-encoder neural network to accommodate two different approaches: one based on the principal component of a set of vibration features, and the other based on an anomaly detection scheme operating directly at the vibration signal level. The results of the analysis show that the former strategy provides a smooth uncertain index, while the latter yields a sharp sensitive index. Such different diagnosis designs are regarded as an opportunity to better fit the vision of the maintenance course of action.**

***Index Terms*—railway axlebox, health index, auto-encoder, neural networks, principal component, anomaly detection**

I. Introduction

The ability to diagnose component faults in their infancy is currently limited due to sensitivity to signal noise, dependence on environmental and operating conditions, lack of fault detection, and uncertainties in maintenance schedules. Consequently, most maintenance actions are reactive. Additionally, generic preventive diagnostics trigger alarms when some key performance indicator degrades below an acceptable threshold. However, such preventive maintenance is estimated to be applied unnecessarily up to 50% of the time [1].

What is common among the former maintenance approaches is the lack of understanding of the actual asset degradation, and the use of conservative hard thresholds that disguise this knowledge deficiency. The adoption of such a statistical process control strictly used as a quality check, misses important aspects of the degradation process because the real world is not discrete. Therefore, in order to tackle the need for a more refined diagnostic, the predictive maintenance approach suggests using a Health Index (HI) as a finer abstract representation of an asset's degradation.

The HI is a real-valued figure that encodes the condition of an asset, implicitly aggregating the physical sensor data and the failure modes that lead to the numerical diagnosis result. It displays a monotonic decreasing evolution from 10 (brand new condition) to 0 (scrap condition). Standard approaches to Prognostics and Health Management (PHM) like ISO 13374 [2] contemplate the estimation of the HI within the Health Assessment module. Other similar standard approaches like ISO 13373 [3] are specific to vibration condition monitoring, and suggest using 4 hazard zones. And ISO 10816 [4] even defines the alarm thresholds regarding the type of machine following a statistical analysis of a worldwide industry survey.

The context of the present work is framed in the maintenance of axleboxes for the Northern Line metro in London, where a condition-based maintenance approach with vibration inspection has been conducted since 2004 with excellent train availability results. The maintenance team that is responsible for this project has traditionally approached axlebox diagnostics with 3 condition states driven by the magnitude of the vibration, which is a straight indicator of the failure severity [5]. This strategy has led to a positive return of experience after having detected a few incipient failures, and not having experienced any in-service failure since the introduction of the vibration monitoring programme. However, the high variability of the vibration magnitude may lead to unsteady results, which increases the rate of false alarms, and the lack of ability to closely track the evolution of the degradation hinders the possibility to plan for optimum maintenance resources at the depot.

The goal of this article is to develop a standard HI adapted to the Northern Line scenario, leveraging the knowledge gained with experience, and exploiting its vibration monitoring data and maintenance record. To do so, two different strategies for creating the HI are compared and discussed: a traditional one with a feature-based principal component, and a more innovative one with a signal-based anomaly detection technique. Additionally, both approaches share the auto-encoder neural network framework, following the success of this technique that appears in the recent state of the art for bearing diagnostics [6]–[11]. This manuscript is organised as follows: Section II describes the analysis procedure that has been pursued, including the description of the data, the processing techniques, and the method to create the HI. Section III conducts the experiments and reports the outcomes. Section IV discusses the overall results and the limitations of the approach, and Section V concludes the manuscript and reflects on its impact to the current maintenance process.

II. Method

This section describes the creation of a custom HI driven by a record of vibration maintenance data.

A. Axlebox Monitoring Experience

The axlebox bearing is a heavy-duty safety-critical railway component. Its replacement is envisaged to occur every 11 years according to the maintenance plan (during the overhaul action). Also, it is regreased every 4 years, taking 2 years to process the whole fleet, which has 106 trains. In addition, its condition is regularly monitored with vibration inspection tests in order to ensure it operates successfully.

In order to accomplish this monitoring, a sample of the fleet consisting of 28 units (3-car trainsets that equip 24 axleboxes each) is inspected every 6 weeks. The vibration signature of each asset is acquired, and its condition is assessed by the magnitude of its peaks, yielding three severity-ordered states: good G, regular R, and bad B. Their specific thresholds were designed by the maintenance team following their expertise. Given the time that it takes to cycle through all the planned maintenance actions, the diversity in the asset condition is expected to be present in every acquisition test.

The vibration monitoring is conducted on a dedicated and calibrated test track at the depot with a constant speed of 5mph. This test scenario maintains the wheel-track contact noise at a minimum.

Finally, in order to explore the evolution of the degradation and the life of the axleboxes for the Northern Line scenario, the maintenance team set one train for such carefully controlled experimental purposes. After 7 to 8 years of service following the overhaul (i.e., fitting brand new axleboxes), the condition of those assets correspond to the "regular" state. Therefore, it can be estimated that the yearly wear rate in HI units should be given by (1).

$$\frac{R-G}{Service} = \frac{5-10}{7.5} = -0.6667 \ {}^{HI}/_{Yr} \tag{1}$$

Note that (1) gives a weight of 5 to the regular condition. This is somewhat arbitrary, but given that it's the average of brand new and scrap conditions, it may be taken for a reliable indication.

B. Vibration Data Collection

In order to acquire the vibration signature of the axleboxes, a network of intelligent wireless sensors called The Motes is used [12], [13], see Fig. 1. These sensors have been configured to acquire 4 seconds of the vertical vibration axis (normal to the ground), with a sampling frequency of 3.2kHz. This setting ensures that enough wheel rotation is captured, and that the vibration pattern is reliably represented.

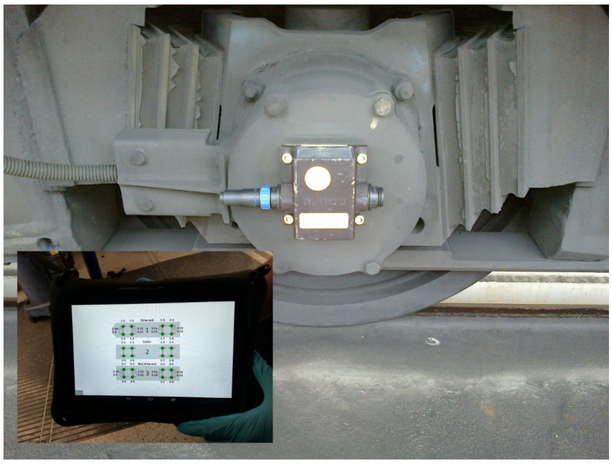

Fig. 1. The Motes in use with an axlebox at the Morden depot in London. The small window at the bottom-left corner also shows a tablet, which is used on board to operate the network of sensors.

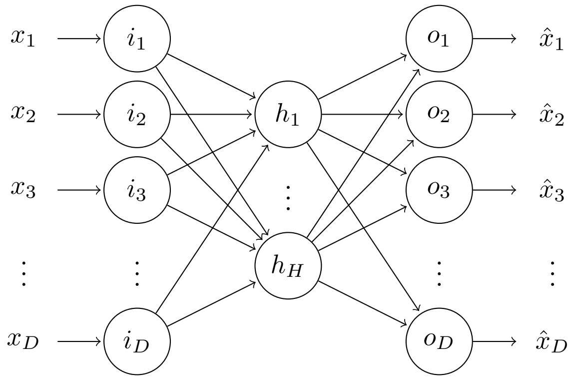

Fig. 2. Auto-encoder architecture. D is the data dimensionality, and H is the size of the hidden layer, which defines the representational capacity of the network. The encoding character of the model is ensured as long as $H < D$.

The dataset of use in this work contains a maintenance record of 2.5 years, with 6200 instances (i.e., the vibration signature of each asset for a particular test).

C. Auto-encoder Neural Network

The auto-encoder neural network is a connectionist machine learning technique that learns to replicate data. The model of use in this work, which exploits auto-association [14], [15], shows a convergent structure from its input dimensionality D into H at half of its length, and then it diverges back to D toward its output, showing a diabolo-like architecture, see Fig. 2. Without loss of generality, the auto-encoders considered in this work have one single hidden "bottleneck" layer, which may have a variable number of hidden units H that define the expressiveness of the network.

An auto-encoder is trained to encode the input into some lower-dimensional representation so that it may thereafter be reconstructed. Hence, the target output of the auto-encoder is the auto-encoder input itself. As a result, the network learns a compressed distributed representation of the data that captures its main factors of variation [16]. The training makes use of gradient descent with the mean-squared error (MSE) of the reconstruction as the cost function. Once the learning is complete (this is an offline procedure), the auto-encoder can then quickly process new data through a set of matrix multiplications [17], which is also very advantageous for industrialisation purposes.

D. Feature-based Principal Component Auto-encoder

The classical approach to designing an automatic pattern recognition system starts with a set of features extracted from the raw input data. Thus, the list of vibration features of use is shown as follows (most of them are described in [18], [19]):

0) Root mean square (RMS): energy content indicator. It operates on the raw acceleration waveform. It shows a positive correlation with fault severity.
1) Variance: signal dispersion indicator. It's the second central moment of its distribution.
2) Skewness: signal asymmetry indicator. It's the third central moment of its distribution.
3) Kurtosis: signal peakdness indicator. It's the fourth central moment of its distribution. A value of 3 is well-recognised to belong to a good condition.
4) Shape factor: signal shape indicator related to its distribution.
5) Crest factor: signal spikiness indicator. Typically used as and impact detector.
6) Entropy: signal uncertainty indicator. Implemented as a magnitude sign randomness indicator.
7) Velocity: velocity signal energy indicator [4]. The raw acceleration signal needs to be numerically integrated prior to computing its RMS.
8) Peak: expert fault severity indicator. It corresponds to the maximum acceleration waveform value.
9) Clearance factor: fault indicator. A peak-based calculation.
10) Impulse factor: another fault indicator. Also, a peak-based calculation.
11) High frequency noise: frequency-based fault indicator. Faults develop noise in the high-frequency range [20].

In order to condense the amount of extracted information for the metro environment at hand, the ranges of the aforementioned features first need to be normalised to similar scale values, and then their amount of correlation has to be reduced. To do so, the Pearson correlation coefficient matrix is used, and the features that show the least amount of mutual relation with the others are selected. This reduces the amount of verbosity at the feature level.

Finally, the Principal Component (PC) is extracted as an aggregation of the resulting explanatory features. The PC is the best one-dimensional linear projection of the feature data that is optimal in a sum-squared error sense [21]. It can be obtained using a feature-set auto-encoder with one single linear hidden unit [17]. However, in this case, the more basic approach of the Karhunen-Loève transform can also be used for computational efficiency [22], generating an equivalent result: the PC is the eigenvector of the feature covariance matrix that corresponds to the largest eigenvalue.

Note that the extraction of the PC is an unsupervised process driven by the maximum variance of the projected dimension. A greater variance is assumed to contain more information, but it must be ensured that the resulting PC follows the direction of the signal, which is indicative of the health condition, rather than the unrelated noise in the vibration data [21].

E. Signal-based Anomaly Detection Auto-encoder

An anomaly detector is a system that only models good healthy data, and then uses this prototype to assess the similarity or distance to the more degraded conditions. Therefore, the machine health is summarised as one statistic: the amount of deviation from healthy data [17].

The proposed approach represents the healthy raw vibration signal directly (instead of the calculated vibration features) with an auto-encoder, inspired by the recent advances in deep learning. The detection of the anomaly occurs when there happens to be a discrepancy between the input and the output of the model: healthy samples show a small difference because the auto-encoder can reliably reconstruct them, but degraded samples get poorly reconstructed and thus display a larger error.

This auto-encoder needs to be expressive enough to represent the complicated interactions at the raw vibration signal level. In this sense, the hidden units also need to be non-linear to help capture the multi-modal aspects of the input distribution [16]. Thus, the hyperbolic tangent is used as the activation function. This approach can be regarded as a non-linear generalisation of the former component analysis [21].

F. Lagging Condition-weighted Average

The construction of a hybrid HI between expert and data-driven criteria entails reaching an agreement between these two complementary approaches to diagnosis. To do so, the distribution of the numerical results obtained with the actual field data, either with the principal component or with the anomaly detection, are further reevaluated and labelled according to the expert diagnostic that follows the magnitude of the vibration. Then, with the resulting condition-driven histogram, a rolling window the size of a bin is swiped along the abscissa, from "good" condition to "bad", and a weighted average is computed, lagging the last computation to ensure a monotonically decreasing evolution.

The two approaches being evaluated need to be forwarded the same data in order to conduct a fair comparison. To do so, the vibration acquisitions are segmented into chunks of 500 vibration samples (a 160ms signal), with an 80% overlap. This is supported by the belief that the redundancy in the data provides knowledge [23]. In addition, the first complete segment is stripped and held out for validation to assess an unbiased effectiveness result.

Finally, in order to characterise the representations of the resulting condition-driven distributions, the Bhattacharyya distance (BD) is computed. The BD quantifies the amount of

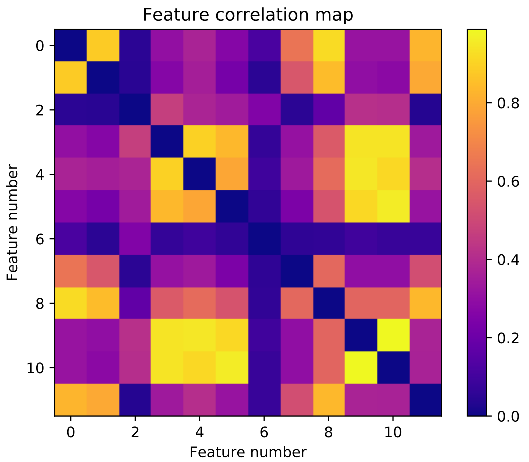

Fig. 3. Pearson correlation coefficient matrix of the vibration features. The feature names are matched to their numbers in Section II-D.

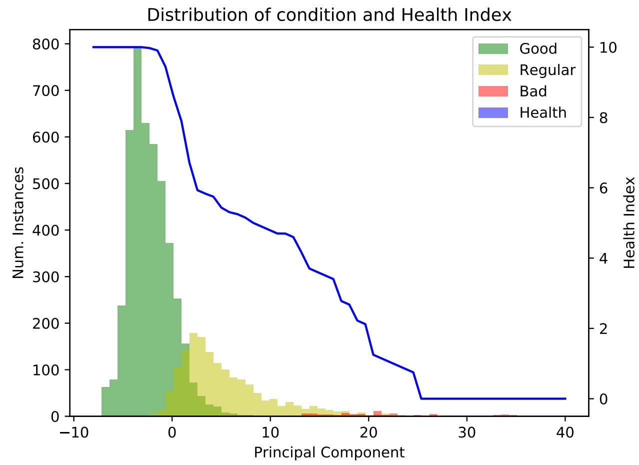

Fig. 4. Health Index based on the feature-based principal component auto-encoder. The ordinate axis for the HI is the one on the right.

overlap between two distributions. It is better at classification than other measures, such as the Mahalanobis or the Kullback-Leibler distances, due to the fact that it incorporates covariances as well as means [24]. The closed-form formula of use (2) assumes Gaussian distributions in the data.

$$BD = \frac{1}{8}(\mu_1 - \mu_2)^T \Sigma^{-1}(\mu_1 - \mu_2) + \frac{1}{2}\ln\frac{|\Sigma|}{\sqrt{|\Sigma_1||\Sigma_2|}} \quad (2)$$

In (2), Σ refers to the average covariance matrix between the two distributions, i.e., the two health conditions being evaluated. It is to note that this formula may still be used even if the underlying distributions are not Gaussian [21].

III. Results

The traditional approach with the feature-based principal component yields the Pearson correlation coefficient matrix, see Fig. 3. It shows that the most uncorrelated vibration features (with a value smaller than 0.8) are: RMS, skewness, entropy, velocity, impulse factor, and high frequency noise. It is to note that the peak value that is used by the expert diagnosis approach is highly correlated with several other features. Therefore, it is excluded for the extraction of the PC. The subsequent calculation of the HI is shown in Fig. 4. It can be seen that the PC effectively captures the condition information of the signal: the progression from the "good" condition to the "bad" condition happens through the "regular" condition, which is consistent with the expected degradation evolution.

For the direct approach using the signal-based anomaly detection auto-encoder, Fig. 5 shows the adjustment of the network's expressiveness. There is an inflection point (i.e., a sudden change of slope) around 40 hidden units that starts to bring the training and test errors closer. For this configuration, the reconstruction fidelity of a vibration sample is shown in Fig. 6. It can be seen how the auto-encoder learns the

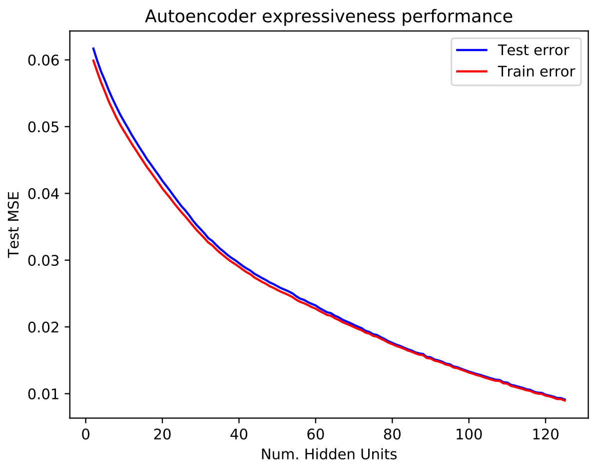

Fig. 5. Auto-encoder expressiveness adjustment. The representational capacity increases with the number of hidden units because the mean-squared error (MSE) of the reconstruction decreases.

regularities of the data, and behaves like a low-pass filter on the original signal, as if it was a de-noising system.

Finally, Fig. 7 shows the HI based on the reconstruction error of the anomaly detection auto-encoder. This approach displays a much more initial abrupt transition as the axlebox condition degrades.

In order to provide further insight into the resulting functions for the HI, Table I shows the Bhattacharyya distances among the different conditions, for the two approaches. It is to note that the anomaly detection auto-encoder provides a solution with more separation among the health conditions.

IV. Discussion

The approach to constructing a HI based on the feature-based principal component shows smooth transitions along the continuum of degradation. The inherent variance-based

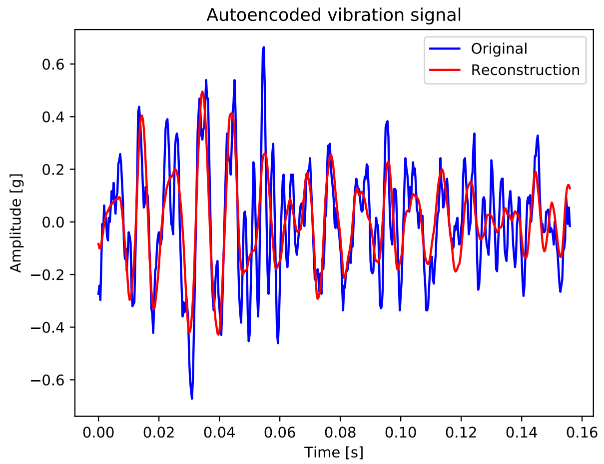

Fig. 6. Reconstruction of a vibration signal using an auto-encoder with 40 hidden units.

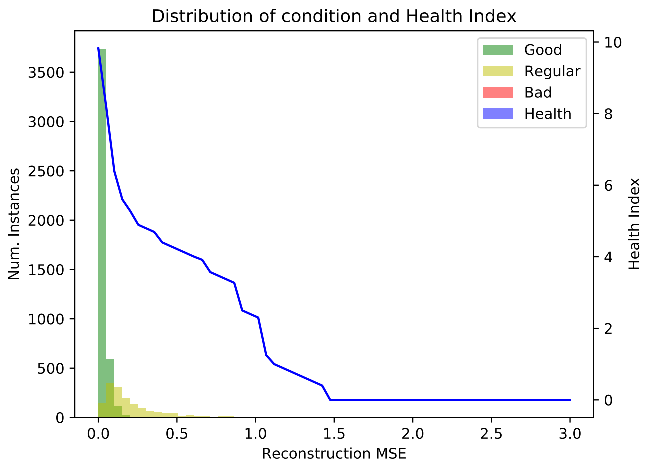

Fig. 7. Health Index based on the signal-based anomaly detection auto-encoder. The ordinate axis for the HI is the one on the right.

information criterion is therefore effective to identify the different condition states, which spread evenly along the range of the independent variable. However, this is partly due to the wide uncertainty that exists within each state. In contrast, the HI based on the anomaly detection approach shows a narrow uncertainty around the good condition, which results in a sharp transition toward the remaining degraded space. Depending on the focus of the maintenance, one approach might be more appropriate than the other. It must be stated that the potential sources of uncertainty that have not been accounted for in this work are mainly driven by the accumulated bogie mileage, the wheel tread conditions, and the car-level design differences: driven, non-driven, and trailer.

TABLE I
BHATTACHARYYA DISTANCES AMONG THE CONDITIONS FOR THE DIFFERENT HI APPROACHES.

HI Approach	G-R	G-B	R-B
Principal Component	1.0999	3.7712	1.2709
Anomaly Detector	1.7149	4.3060	1.4899

It is interesting to note how the high-frequency noise rejection ability provided by the signal-based auto-encoder affects its performance to detect anomalies. The appearance of high-frequency noise is precisely what indicates the presence of a failure [20]. If the auto-encoder is adjusted so that it may not accurately replicate this noise, which is present in degraded data, then it becomes successful in recognising the anomalies.

It is also shown that regardless of the HI approach, its evolution throughout the life of the axlebox is not linear with the vibration indicator, contrary to what might be expected from gradual mechanical wear. Perhaps this progressive evolution is only present within each condition state, and thus the collected dataset is too short to exhibit this extended change. In any case, if the derivative of the HI is taken with respect to the time difference of the acquisitions, seeking a wear rate estimate, the resulting figures get much higher than the expected rate provided by the experimental train. This leads to question the condition weights used to construct the health indexes, which stretch to the complete health range, from 10 to 0. Therefore, the actual vibration dataset of use must only show a segment of the entire HI span, probably somewhere between 7 and 4. Only then may the wear rates converge to similar values. Besides, there are no brand-new nor collapsed bearings in the fleet today, definitely, so it makes perfect sense to narrow the observed health range. However, further prognosis-related details are out of the scope of this study.

V. CONCLUSIONS

At present, the maintenance of axleboxes for the Northern Line fleet is supervised with vibration inspection tests, and the coarse diagnostic results provided by a small set of experience-driven rules is just sufficient to identify potential developing failures when the critical alarms are set. This article presents a more refined approach to diagnostics with a standard Health Index, that adds value to this expert focus with data from the field, and enhances its finesse with additional criteria. Two methods are compared for extracting this health function. On the one hand, the principal component of the vibration features provides a smooth, uncertain index. On the other hand, the anomaly detector based on the vibration waveform provides more certainty, especially around the healthy condition.

Finally, the auto-encoder has proven to be a very effective and versatile technique for PHM. The future work that is currently envisaged may delve into the intermediate representations that can be obtained with the hidden units of the encoding layer. At present, they are totally random, which might make them inconsistent with the expected evolution of the degradation. Maybe the topologically-preserving criteria of other neural networks like the self-organising maps could be of help to better understand them, and also provide new representations able to adapt to other maintenance environments.

REFERENCES

[1] G. W. Vogl, B. A. Weiss, and M. Helu, "A review of diagnostic and prognostic capabilities and best practices for manufacturing," *Journal of Intelligent Manufacturing*, pp. 1–17, 2016. [Online]. Available: https://doi.org/10.1007/s10845-016-1228-8

[2] ISO, "Condition monitoring and diagnostics of machine systems: Data processing, communication and presentation," International Organization for Standardization, Tech. Rep. 13374-1:2003, 2003.

[3] ——, "Condition monitoring and diagnostics of machines - Vibration condition monitoring," International Organization for Standardization, Tech. Rep. 13373-1:2002, 2002.

[4] ——, "Mechanical vibration – Evaluation of machine vibration by measurements on non-rotating parts," International Organization for Standardization, Tech. Rep. 10816-3:2009, 2009.

[5] S. J. Lacey, "An Overview of Bearing Vibration Analysis," *Maintenance and Asset Management*, vol. 23, no. 6, pp. 32–42, Nov./Dec. 2008.

[6] M. Sohaib, and J.-M. Kim, "Reliable Fault Diagnosis of Rotary Machine Bearings Using a Stacked Sparse Autoencoder-Based Deep Neural Network," *Shock and Vibration (Hindawi)*, vol. 2018, no. 2919637, pp. 1–11, 2018.

[7] S. Haidong, J. Hongkai, L. Xingqiu, and W. Shuaipeng, "Intelligent fault diagnosis of rolling bearing using deep wavelet auto-encoder with extreme learning machine," *Knowledge-Based Systems (Elsevier)*, vol. 140, pp. 1–14, Oct. 2018.

[8] J. Di, and L. Wang, "Application of Improved Deep Auto-Encoder Network in Rolling Bearing Fault Diagnosis," *Journal of Computer and Communications (Scientific Research Publishing)*, vol. 6, pp. 41–53, Jul. 2018.

[9] H. O. A. Ahmed, M. L. D. Wong, and A. K. Nandi, "Intelligent condition monitoring method for bearing faults from highly compressed measurements using spare over-complete features," *Mechanical Systems and Signal Processing*, vol. 99, pp. 459–477, Jul. 2018.

[10] D.-T. Hoang, and H.-J. Kang, "A Bearing Fault Diagnosis Method on Autoencoder and Particle Swarm Optimization - Support Vector Machine," *Intelligent Computing Theories and Application. ICIC 2018. Lecture Notes in Computer Science (Springer)*, vol. 10954, 2018.

[11] C. Li, W. Zhang, G. Peng, and S. Liu, "Bearing Fault Diagnosis Using Fully-Connected Winner-Take-All Autoencoder," *Special Edition on Complex System Health Management Based on Condition Monitoring and Test Data*, vol. 6, pp. 6103–6115, Jun. 2017.

[12] A. Trilla and P. Gratacòs, "Maintenance of bogie components through vibration inspection with intelligent wireless sensors: a case-study of axle-boxes and wheel-sets using the Empirical Mode Decomposition technique," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit (ISSN: 2041-3017)*, no. 230, pp. 1408–1414, 2016.

[13] ——, "Condition based maintenance on board," *Chemical Engineering Transactions Journal (ISSN: 1974-9791)*, no. 33, pp. 733–738, 2013.

[14] M. A. Kramer, "Autoassociative Neural Networks," *Computers and Chemical Engineering*, vol. 16, no. 4, pp. 313–328, Apr. 1992.

[15] V. M. Stone, "The auto-associative neural network - a network architecture worth considering," *Proc. of the 2008 World Automation Congress (ISSN: 2154-4824)*, pp. 1–4, Sep. 2008.

[16] Y. Bengio, "Learning Deep Architectures for AI," *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–71, Jan. 2009.

[17] P. Goldthorpe and A. Desmet, "Denoising autoencoder anomaly detection for correlated data," *Proc. of the Fourth European Conference of the Prognostics and Health Management Society (ISSN: 2325-016X).*

[18] J. K. Kimotho, and W. Sextro, "An approach for feature extraction and selection from non-trending data for machinery prognosis," *Proc. of the Second European Conference of the Prognostics and Health Management Society (ISSN: 2325-016X).*

[19] W. Caesarendra, and T. Tjahjowidodo, "A Review of Feature Extraction Methods in Vibration-Based Condition Monitoring and Its Application for Degradation Trend Estimation of Low-Speed Slew Bearing," *Machines (ISSN: 2075-1702)*, vol. 5, no. 4, 2017.

[20] D. Howieson, "Vibration Monitoring: Envelope Signal Processing," *@ptitudeXchange - SKF Reliability Systems*, pp. 1–13, 2003.

[21] R. O. Duda, P. E. Hart, and D. G. Stork, Eds., *Pattern Classification*. Wiley-Interscience, 2001.

[22] J. Almotiri and K. Elleithy and A. Elleithy, "Comparison of autoencoder and Principal Component Analysis followed by neural network for e-learning using handwritten recognition," *Proc. of 2017 IEEE Long Island Systems, Applications and Technology Conference (LISAT)*, pp. 1–5, May. 2017.

[23] J. Hertz, A. Krogh, and R. G. Palmer, Eds., *Introduction to the Theory of Neural Computation*. Addison-Wesley Longman Publishing Co., Inc., 1991.

[24] S. A. Janse, "Inference using Bhattacharyya distance to model interaction effects when the number of predictors far exceeds the sample size," Ph.D. dissertation, University of Kentucky, Lexington, KY, USA, Oct. 2017.