Deep Health Indicator Extraction: A Method based on Auto-encoders and Extreme Learning Machines

Yang Hu[1], Thomas Palmé[2], and Olga Fink[3]

[1,3]*Zurich University of Applied Sciences, Rosenstr. 3, Winterthur, 8401, Switzerland,*
yang.hu@zhaw.ch
olga.fink@zhaw.ch

[2]*General Electric (GE) Switzerland, Brown Boveri Str. 7, Baden, 5401, Switzerland*
thomas.palme@ge.com

ABSTRACT

In this paper, we propose a novel deep learning method for feature extraction in prognostics and health management applications. The proposed method is based on Extreme Learning Machines (ELM) and Auto-Encoders (AE), which have demonstrated very good performance and very short training time compared to other deep learning methods on several applications, including image recognition problems. The proposed approach is applied to vibration condition monitoring data to extract features from normal operation (i.e. fault free conditions) without any additional expert knowledge or prior information on the type of signals and the information content in the datasets. The approach demonstrates a better performance in terms of trendability and monotonicity compared to commonly applied feature extraction methods.

1. INTRODUCTION

The decreasing cost of sensors and data transmission, the improved functionality, and reliability of the sensors have increased the availability of data on the system condition. This development has had a significant impact not only in terms of volume but also in terms of velocity, and the variety of data streams. Prognostics and health management (PHM) attempts to make better use of this information on the system condition to provide a holistic view on the health state of the system and to enable decision support on the optimal preemptive maintenance actions and logistic decisions (Lee et al., 2014).

The increased availability of the condition monitoring data has increased the application and development of data-driven PHM approaches. Particularly, the application of machine learning approaches has been increasing. Different machine learning approaches have been introduced for the classification of the type of fault causing the malfunctioning of engineering systems as well as predicting the Remaining Useful Life (RUL), including K-Nearest-Neighbor Clustering (KNN) (Lei & Zuo, 2009), Artificial Neural Networks (ANN) (Mavromatidis, Acha, & Shah, 2013; Shao, Zhu, Cao, & Shen, 2014), Support Vector Machines (SVM) (Selak, Butala, & Sluga, 2014) and Fuzzy Classifiers (FC) (Lemos, Caminhas, & Gomide, 2013).

In general, machine learning algorithms require as input high-level informative features to achieve a good performance. These features need to provide a good representation of the system health states. However, raw signals obtained by condition monitoring sensors are typically high dimensional and are often also highly redundant, particularly for critical systems that require the sensors for control purposes. High dimensional correlated signals are very difficult to analyse due to the computational burden, curse of dimensionality, and the complexity of the required approaches to extract uncorrelated useful information, which often results in ill-posed problems Therefore, it is not optimal to use the raw condition monitoring signals directly as input for machine learning algorithms.

In such case, several signal processing, and feature extraction and selection approaches are required to extract the relevant information from raw signals. For example, numerous approaches have been proposed to extract useful features from raw vibration signals of bearings. The feature extraction approaches can be classified in 1) Time Domain approaches, such as statistical indicators, autoregressive modelling and empirical mode decomposition (Weizhong Yan, Qiu, & Iyer, 2008); 2) Frequency Domain approaches, such as Spectral Analysis, Envelope Analysis, Higher Order Spectrum; 3) Wavelet analysis such as Continuous/Discrete Wavelet Transform, Morlet Wavelet, Hilbert-Huang Transform (Weizhong Yan et al., 2008). Some recent studies also use more advanced algorithms for feature extraction, like iterative envelope analysis and a low-pass filtering operation (Ming, Zhang, Qin, & Chu, 2016), Empirical Wavelet Transform (EWT) method via data-driven adaptive Fourier spectrum segment (Pan, Chen, Zi, Li, & He, 2016) and sparse representation in wavelet basis (Fan et al., 2015). Even with advanced feature extraction approaches, the extracted features are in some cases still not sufficiently informative, and the feature reduction or selection methods are required for processing the features further and reduce the dimension of input for saving computation burden and improve the robustness of PHM models (Benkedjouh, Medjaher, Zerhouni, & Rechak, 2013; Bolón-Canedo, Sánchez-Maroño, & Alonso-Betanzos, 2015; Emmanouilidis, Hunter, MacIntyre, & Cox, 1999; Gowid, Dixon, & Ghani, 2015; Yang, Liao, Meng, & Lee, 2011).

Generally, the feature extraction methods mentioned above are developed based on domain and engineering knowledge of special fields, which can be called "handcrafted" features (WZ Yan & Yu, 2015). A major limitation of "handcrafted" features is that it is a manual process and an ad-hoc problem, which requires extensive expert involvement and interaction to develop specific solutions for each individual case. In addition, the accuracy of these "handcrafted" features may not be optimal. Thereby, the performance of PHM models highly depends on the experience and the expertise of the experts designing the features. A universally applicable feature extraction approach is, therefore, required, that is able to process raw condition monitoring signals in an automatic and unsupervised way and provides highly informative features that are directly applicable to PHM modeling. It is threfore very valuable to develop a general.approach for feature learning without any dependence on the expert experience and knowledge.

The problem of feature extraction is a type of representation learning problems, which has gained a lot of attention and also importance in the last decade since deep learning has been introduced. Deep learning has become one of the most frequently applied approaches for representation learning and has been applied in many different applications, including image and speech recognition, object tracking and language processing. Bengio, Courville, & Vincent (2013) provide a review of the state of the art representation learning approaches. Deep learning approaches include auto-encoders, restricted Boltzmann machines, semi-supervised embedding, convolutional neural networks, and kernel principal component analysis. Recently, deep learning has also been applied in the field of PHM. In (Gan, Wang, & Zhu, 2016; Jia, Lei, Lin, Zhou, & Lu, 2016; Tamilselvan & Wang, 2013) deep belief networks were applied for fault diagnostic modeling. However, those applications focused developing a "deep diagnostic models". However, the input features of the models remain "handcrafted". In (WZ Yan & Yu, 2015), the authors use a stacked denoising auto-encoder to extract useful features from the signals of gas turbine combustors. The models based on automatically extracted features shows a better detection performance compared to those based on "handcrafted" features.

In this paper, we propose a novel deep learning method for feature extraction in PHM applications. The proposed method is based on Extreme Learning Machines (ELM) and Auto-Encoders (AE), which are very popular machine learning approaches with proven performance on many openly available classification and regression benchmark datasets (G. Bin Huang, 2015; G.-B. Huang, Zhou, Ding, & Zhang, 2012). The ELM-AE approach has been applied to the image recognition problems with very good classification and regression accuracy on benchmark datasets and very short training time compared to other deep learning methods including deep belief networks, stacked AE, deep Boltzmann machine (Cambria et al., 2013). The proposed approach is applied to IEEE PHM 2012 Data Challenge for extracting health indicators of bearings under accelerated testing experiments.

The remainder of this paper is organized in following way: Section 2 and 3 introduce of the methodology of ELM-AE; Section 4 shows its application on the bearing dataset and Section 5 summarizes the results and provides some practical recommendations for industrial users to support the application of the proposed methodology in their applications.

2. EXTREME LEARNING MACHINES

ELM are generalized Single-hidden Layer Feedforward Neural networks (SLFNs) (G.-B. Huang, Zhu, & Siew, 2004). The output $f_L(\mathbf{x}_\mathrm{k})$ of the ELM with L hidden nodes can be written as:

$$f_L\left(\mathbf{x}_k\right)=\sum_{i=1}^{L}\boldsymbol{\beta}_i\cdot g_i\left(\mathbf{a}_i\mathbf{x}_k+b_i\right)=\mathbf{H}\boldsymbol{\beta} \quad (1)$$

where $\mathrm{g_i}$ denotes the activation function, $\boldsymbol{\beta}_i$ the output weight, $\mathbf{a}_i, b_i$ the parameters of the activation function in the hidden layer and $\mathbf{x}_k$ the k-th input pattern with k=1,…, N_T. $\mathbf{H}\boldsymbol{\beta}$ is the matrix form of ELM with H as the hidden layer output matrix:

$$\mathbf{H}=\begin{bmatrix}\mathbf{h}(\mathbf{x}_1)\\ \vdots\\ \mathbf{h}(\mathbf{x}_{N_T})\end{bmatrix}=\begin{bmatrix}g(\mathbf{a}_1,b_1,\mathbf{x}_1) & \cdots & g(\mathbf{a}_L,b_L,\mathbf{x}_1)\\ \vdots & \ddots & \vdots\\ g(\mathbf{a}_1,b_1,\mathbf{x}_{N_T}) & \cdots & g(\mathbf{a}_L,b_L,\mathbf{x}_{N_T})\end{bmatrix}_{N_T\times L}$$

The weights between the input and hidden layer and the bias ($\mathbf{a}_\mathrm{i}, b_\mathrm{i}$) are set randomly sampled from an arbitrary distribution. Only the weights between the hidden layer and the output($\boldsymbol{\beta}_\mathrm{i}$) need to be estimated in the training process using the following formulation:

$$\min_{\boldsymbol{\beta}}\left(\|\boldsymbol{\beta}\|^2 + C\|\mathbf{H}\boldsymbol{\beta}-\mathbf{t}\|^2\right) \tag{2}$$

Equation (2) represents the minimum of the sum of the L_2 norm of $\boldsymbol{\beta}$ and the training error, where $\mathbf{t}$ is the training target, and C is the trade-off parameter between the training error and the generalization ability. Applying ridge regression, the solution of $\boldsymbol{\beta}$ is:

$$\boldsymbol{\beta} = \left(C^{-1}\mathbf{I} + \mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{t} \tag{3}$$

The main advantage of ELM is that they don't require the iterative fine-tuning of network parameters, which is an extensive calculation. The calculation of $\boldsymbol{\beta}$ is only based on matrix multiplication and is very fast compared to methods without an analytical solution. .

3. AUTO-ENCODERS BASED ON ELM

The basic principle of an AE is to encode the input into a hidden layer and decode it back to the original input. If the desired reconstruction accuracy is obtained, the encoded hidden layer can be used as the representation of the original input. The basic structure is shown in Figure **1**:

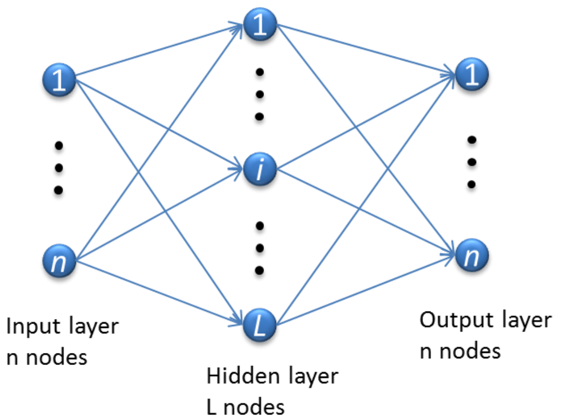

Figure 1. Structure of an Auto-Encoder

AE can be treated as an Auto-Associative SLFN. The network is trained by setting the target output equal to the input. This training process is very similar to equation(2), with the difference that the training target is equal to the input:

$$\begin{gathered}\sum_{i=1}^{L}\boldsymbol{\beta}_i \cdot g_i\left(\mathbf{a}_i\mathbf{x}_k + b_i\right) = \mathbf{x}_k \\ \min_{\boldsymbol{\beta}}\left(\|\boldsymbol{\beta}\|^2 + C\|\mathbf{H}\boldsymbol{\beta}-\mathbf{x}\|^2\right)\end{gathered} \tag{4}$$

$\boldsymbol{\beta}$ can be solved by:

$$\hat{\boldsymbol{\beta}} = \left(C^{-1}\mathbf{I} + \mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x} \tag{5}$$

In the first layer of the stacked auto-encoder, the extracted feature vector $\mathbf{f}$ of original input $\mathbf{x}$ can be written as:

$$\mathbf{f} = \mathbf{x}\hat{\boldsymbol{\beta}} \tag{6}$$

The dimension of $\mathbf{f}$ is equal to the number of neurons in the hidden layer L. By setting L smaller than the dimension of $\mathbf{x}$, a condensed representation of the input is achieved. Typically, a single AE cannot fully extract useful features from x. Therefore, stacked AE can be applied. The structure of stacked AE is shown in Figure 2.

In this stacked ELM-AE, the input of j-th layer ($j > 1$) is the feature vector extracted by the previous layer, and the feature vector extracted by the last layer $\mathbf{f}_N$ is the final learned feature vector. By layer-wise stacking the single AE, information hidden in the original input $\mathbf{x}$ is hierarchically extracted by each layer. The most distinguishing characteristic of stacked ELM-AE is that all layers in the network (each single AE representing a layer) are independent of each other. The parameters of each layer only need to be learned once, and are then fixed after the training. By being trained and stacked independently , this network design overcomes the drawback of back-propagation learning requiring an iterative fine-tuning of the parameters in all the layers. Since the learning process is reduced to matrix multiplication, the learning process becomes computationally very efficient. For more details on ELM-AE, please refer to (Cambria et al., 2013)

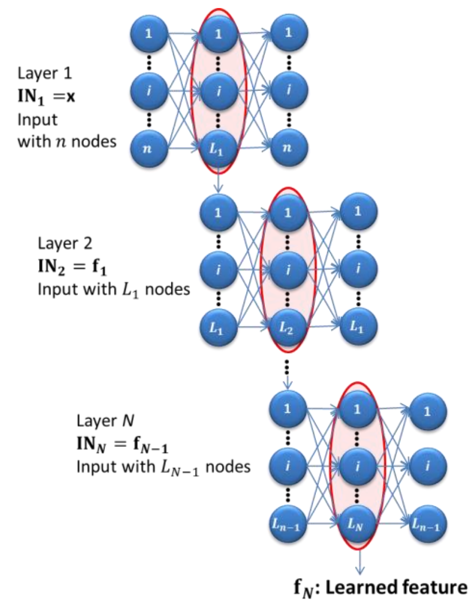

Figure 2. Stacked ELM based AE

For PHM applications, the input to the network, $\mathbf{x}$, can be set as raw monitoring signals. The feature vector $\mathbf{f}_\mathrm{N}$, obtained in the last layer is the extracted condensed and highly informative representation of $\mathbf{x}$, which can be used as health indicators for better understanding the system condition. The extracted health indicator vector, $\mathbf{f}_{\mathrm{N}}$, can also be applied to build detection, diagnostic and prognostic models.

4. CASE STUDY

4.1. Applied dataset

To verify the added value of the ELM-AE algorithm in PHM applications, we apply it to a dataset from IEEE PHM 2012 Data Challenge. This dataset is recorded using an experimental platform called PRONOSTIA, which motivates to provide real data related to the accelerated degradation of bearings performed under constant/variable operating conditions. The details of the experiment design are described in (Nectoux et al., 2012).

This dataset has been used in several research studies for predicting the remaining useful life of bearings. As a first step, relevant features need to be extracted from the raw vibration signals. Different feature extraction approaches can be used, including methods like statistical indicators (root mean square, standard deviation , crest factor, skewness, etc.) (Randall & Antoni, 2011), Wavelet Packet Decomposition (WPD) (Medjaher, Tobon-Mejia, & Zerhouni, 2012), Power Spectrum Density (PSD) (Nectoux et al., 2012) and ISOMAP (Benkedjouh et al., 2013). A more comprehensive overview of the different signal processing approaches for bearing vibration monitoring signals including a review of about 200 papers can be found in (Rai & Upadhyay, 2016).

The features extracted from vibration signals are typically developed based on the domain and engineering knowledge and are applicable for a specific application. They can, therefore, also be referred to as "handcrafted" features.

Contrary to the application of very specialized feature extraction approach, in the following section, we apply stacked ELM-AE for feature extraction of bearing vibration monitoring signals in a completely unsupervised way, and compare the results to some commonly applied "handcrafted" features to evaluate the effectiveness of ELM-AE.

4.2. Results of the case study

Three degraded bearings are selected for performing the experiment, the working condition of them are the same: rotation speed of 1800 rpm and loading with 4000 N. The recorded durations of these three bearings are 27990, 23740 and 14270 seconds. The vibration signals are sampled at 25.6 kHz, and generally, 2560 samples (0.1 seconds) are recorded every 10 seconds. However, during the data acquisition process, there are some time gaps during which the signals were not recorded. The overall dimensions of the raw signals of these three selected bearings are $1629 \times 2560, 2375 \times 2560, 1428 \times 2560$. The first bearing degradation trajectory is used as the training set for training the ELM-AE, and the remaining two bearings are used to test the performance of the algorithm.

For the purpose of prognostics, the features extracted from the vibration signals should have the characteristics of monotonicity and trendability. Monotonicity is an important characteristic of a health indicator since it is generally assumed that industrial components do not undergo self-healing, which would result in non-monotonic indicator trends. Trendability indicates if the degree of the degradation has a regular shape and can be described by a functional form. If the extracted feature provides good monotonicity and trendability, the prognostic model can be easily built and good RUL prediction can be achieved.

In order to extract one feature from 2560 dimension signals, we apply an ELM-AE with four layers, with 1000, 100, 10 and 1 hidden neurons in the respective layers.

To compare the performance of ELM-AE to that of a feature extraction approach with a proven track record, we also extract statistical indicator features (standard deviation) and the Discrete wavelet transform (DWT) using Sym6 basis at level 3. The plots of the raw signals (acceleration) and extracted features of the two tested bearings are shown in figure 3 to 6.

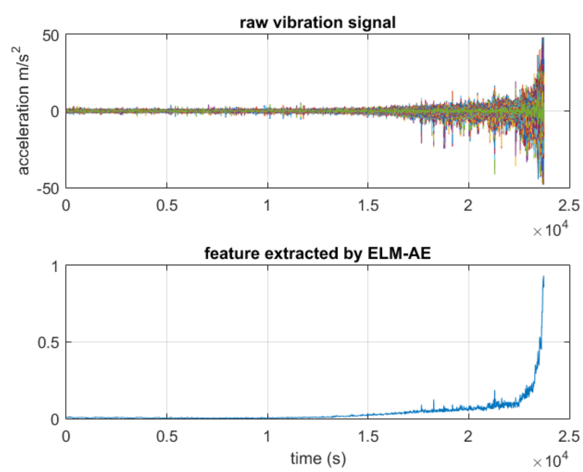

Figure 3. Raw signals and ELM-AE feature of bearing 2

The selected two bearings have different characters. Bearing 2 has a relatively gradual degradation process. However, it contains a lot of "noise spikes", especially towards the end of the test phase. Bearing 3 has a sudden jump of vibration, at the point of time 10,810, and then the degradation develops at a very fast speed. The details interpretations are stated in the next section.

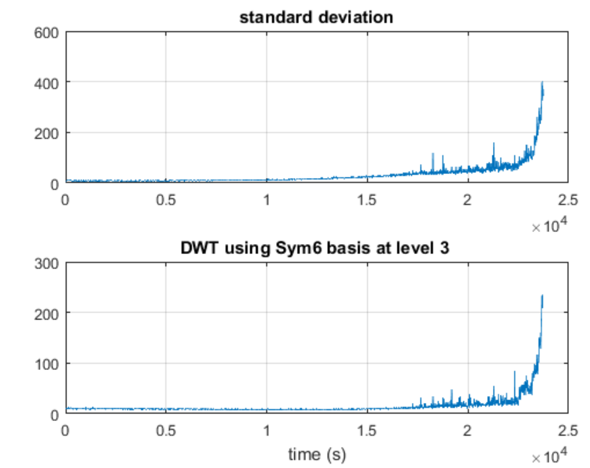

Figure 4. "handcrafted" features of bearing 2

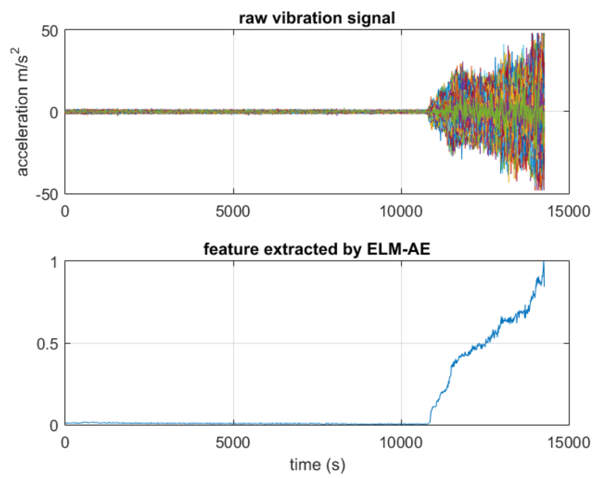

Figure 5. Raw signals and ELM-AE feature of Bearing 3

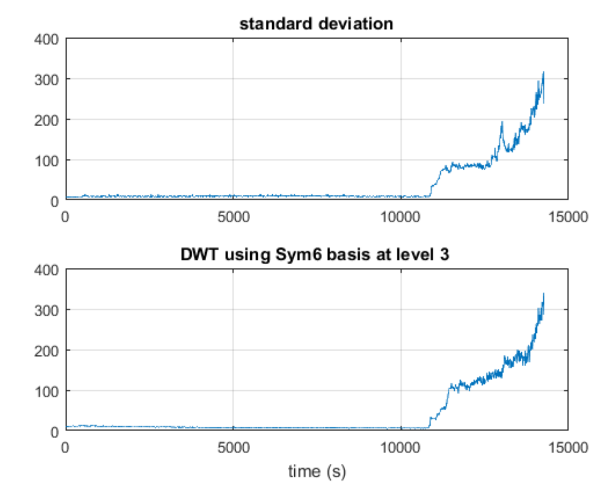

Figure 6. "handcrafted" features of bearing 3

4.3. Interpretation and Discussion of Results

For both of bearings, the degradation is presented by the increase of acceleration, which means the vibration amplitude is large when the bearing is highly degraded. For bearing 2, the results of the "handcrafted" extracted features: standard deviation and DWT are heavily affected by "noise spikes" in the raw signals. This decreases the performance of these two indicators in terms of monotonicity and makes it difficult to build a reliable prognostic model based on these features. On the contrary, the health indicator extracted by ELM-AE is less affected by the "noise spikes" and the intensity of the spikes in the in the ELM-AE indicator is much smaller. This demonstrates the ability of ELM-AE to effectively extract the tendency information hidden in the raw signals and eliminate the effects of noise, even with a high degree of noise in the signal. The ELM-AE are able to extract the relevant information without any additional information on the type of signals and the type of relevant information content in the signals.

The raw signals of bearing 3 are not stable towards the end phase of the test period. The peak values during time 11,500 to 12,000 are in fact larger than those during the time from 12,200 to 12,900. However, it doesn't mean the health status of the bearing has improved. The peak values are not the only indicator of the health condition of bearings. This kind of non-monotonicity in raw signals cannot be filtered by the feature of standard deviation. The health indicators extracted by the "handcrafted" feature extraction approaches are, therefore, not able to extract monotonous health indicators from this dataset and these health indicators are therefore not useful for prognostics models. On the contrary, ELM-AE are able to extract a monotonous health indicator that is able to represent the system health condition and can also be used within prognostics models with good results. Additionally, the health indicator extracted by ELM-AE is smoother than the one extracted by the DWT.

5. CONCLUSIONS

In this paper, we proposed to apply a deep learning approach based on the stacked auto-encoders and extreme learning machines to extract health indicators from bearing vibration monitoring signals. The obtained results were compared to those derived from commonly applied "handcrafted" indicators.

Even though the vibration signals of the two datasets applied to test the algorithm showed very different characteristics compared to those applied to train the algorithm, ELM-AE were able to extract the relevant information in an unsupervised way, without any expert knowledge integrated into the learning process of the algorithm.

The ELM-AE algorithm is able to achieve a better performance compared to the "handcrafted" algorithm.

ELM-AE can be easily applied to any other type of feature extraction problems without integrating additional expert knowledge. The approach provides a general solution to feature extraction problems. This general applicability is in contrast to the typically applied approach in the PHM domain of designing a specific solution for a specific application. It is expected that ELM-AE are applicable to many different types of condition monitoring signals and will be able to effectively extract health indicators from these signals. This will not only enable faster and more reliable detections and predictions of the remaining useful life but will also decrease the costs for algorithm development due to the general applicability of the approach and a fast training time.

ACKNOWLEDGEMENT

This research was funded by the Swiss Commission for Technology and Innovation under grant number 17833.2 PFES-ES.

NOMENCLATURE

PHM	Prognostics and Health Management
RUL	Remaining Useful Life
KNN	K-Nearest-Neighbor Clustering
ANN	Artificial Neural Networks
SVM	Support Vector Machines
FC	Fuzzy Classifiers
EWT	Empirical Wavelet Transform
ELM	Extreme Learning Machines
ELM-AE	Extreme Learning Machines and Auto-Encoder
SLFN	Single-hidden Layer Feedforward Neural networks
PSD	Power Spectrum Density
WPD	Wavelet Packet Decomposition
DWT	Discrete wavelet transform

REFERENCES

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(8), 1798–1828. http://doi.org/10.1109/TPAMI.2013.50

Benkedjouh, T., Medjaher, K., Zerhouni, N., & Rechak, S. (2013). Remaining useful life estimation based on nonlinear feature reduction and support vector regression. *Engineering Applications of Artificial Intelligence*, *26*(7), 1751–1760. http://doi.org/10.1016/j.engappai.2013.02.006

Bolón-Canedo, V., Sánchez-Maroño, N., & Alonso-Betanzos, A. (2015). Recent advances and emerging challenges of feature selection in the context of big data. *Knowledge-Based Systems*, *86*, 33–45. http://doi.org/10.1016/j.knosys.2015.05.014

Cambria, E., Huang, G.-B., Kasun, L. L. C., Zhou, H., Vong, C. M., Lin, J., … Liu, J. (2013). Extreme Learning Machines [Trends & Controversies]. *IEEE Intelligent Systems*, *28*(6), 30–59. http://doi.org/10.1109/MIS.2013.140

Emmanouilidis, C., Hunter, A., MacIntyre, J., & Cox, C. (1999). Selecting features in neurofuzzy modeling by multiobjective genetic algorithms. *Artificial Neural Networks, 1999. ICANN 99. Ninth International Conference on (Conf. Publ. No. 470)*. http://doi.org/10.1049/cp:19991201

Fan, W., Cai, G., Zhu, Z. K., Shen, C., Huang, W., & Shang, L. (2015). Sparse representation of transients in wavelet basis and its application in gearbox fault feature extraction. *Mechanical Systems and Signal Processing*, *56*, 230–245. http://doi.org/10.1016/j.ymssp.2014.10.016

Gan, M., Wang, C., & Zhu, C. (2016). Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings. *Mechanical Systems and Signal Processing*, *72*, 92–104. http://doi.org/10.1016/j.ymssp.2015.11.014

Gowid, S., Dixon, R., & Ghani, S. (2015). A novel robust automated FFT-based segmentation and features selection algorithm for acoustic emission condition based monitoring systems. *Applied Acoustics*, *88*, 66–74. http://doi.org/10.1016/j.apacoust.2014.08.007

Huang, G. Bin. (2015). What are Extreme Learning Machines? Filling the Gap Between Frank Rosenblatt's Dream and John von Neumann's Puzzle. *Cognitive Computation*, *7*(3), 263–278. http://doi.org/10.1007/s12559-015-9333-0

Huang, G.-B., Zhou, H., Ding, X., & Zhang, R. (2012). Extreme Learning Machine for Regression and Multiclass Classification. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, *42*(2), 513–529.

Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2004). Extreme learning machine: a new learning scheme of feedforward neural networks. In *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on* (Vol. 2, pp. 985–990 vol.2).

Jia, F., Lei, Y., Lin, J., Zhou, X., & 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. http://doi.org/10.1016/j.ymssp.2015.10.025

Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., & 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. http://doi.org/10.1016/j.ymssp.2013.06.004

Lei, Y., & Zuo, M. J. (2009). Gear crack level identification based on weighted K nearest neighbor classification algorithm. *Mechanical Systems and Signal Processing*, *23*(5), 1535–1547. http://doi.org/10.1016/j.ymssp.2009.01.009

Lemos, A., Caminhas, W., & Gomide, F. (2013). Adaptive fault detection and diagnosis using an evolving fuzzy classifier. *Information Sciences*, *220*, 64–85. http://doi.org/10.1016/j.ins.2011.08.030

Mavromatidis, G., Acha, S., & Shah, N. (2013). Diagnostic tools of energy performance for supermarkets using Artificial Neural Network algorithms. *Energy and Buildings*, *62*, 304–314. http://doi.org/10.1016/j.enbuild.2013.03.020

Medjaher, K., Tobon-Mejia, D. a, & Zerhouni, N. (2012). Remaining Useful Life Estimation of Critical Components With Application to Bearings. *IEEE Transactions on Reliability*, *61*(2), 292–302. http://doi.org/10.1109/TR.2012.2194175

Ming, A. B., Zhang, W., Qin, Z. Y., & Chu, F. L. (2016). Fault feature extraction and enhancement of rolling element bearing in varying speed condition. *Mechanical Systems and Signal Processing*, *76*, 367–379. http://doi.org/10.1016/j.ymssp.2016.02.021

Nectoux, P., Gouriveau, R., Medjaher, K., Ramasso, E., Chebel-Morello, B., Zerhouni, N., & Varnier, C. (2012). PRONOSTIA : An experimental platform for bearings accelerated degradation tests.

Pan, J., Chen, J., Zi, Y., Li, Y., & He, Z. (2016). Mono-component feature extraction for mechanical fault diagnosis using modified empirical wavelet transform via data-driven adaptive Fourier spectrum segment. *Mechanical Systems and Signal Processing*, *72*, 160–183. http://doi.org/10.1016/j.ymssp.2015.10.017

Rai, A., & Upadhyay, S. H. (2016). A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. *Tribology International*, *96*, 289–306. http://doi.org/10.1016/j.triboint.2015.12.037

Randall, R. B., & Antoni, J. (2011). Rolling element bearing diagnostics-A tutorial. *Mechanical Systems and Signal Processing*, *25*(2), 485–520. http://doi.org/10.1016/j.ymssp.2010.07.017

Selak, L., Butala, P., & Sluga, A. (2014). Condition monitoring and fault diagnostics for hydropower plants. *Computers in Industry*, *65*(6), 924–936. http://doi.org/10.1016/j.compind.2014.02.006

Shao, M., Zhu, X.-J., Cao, H.-F., & Shen, H.-F. (2014). An artificial neural network ensemble method for fault diagnosis of proton exchange membrane fuel cell system. *Energy*, *67*, 268–275. http://doi.org/10.1016/j.energy.2014.01.079

Tamilselvan, P., & Wang, P. (2013). Failure diagnosis using deep belief learning based health state classification. *Reliability Engineering & System Safety*, *115*(0), 124–135. http://doi.org/http://dx.doi.org/10.1016/j.ress.2013.02.022

Yan, W., Qiu, H., & Iyer, N. (2008). *Feature extraction for bearing prognostics and health management (phm)-a survey (preprint)*. DTIC Document.

Yan, W., & Yu, L. (2015). On Accurate and Reliable Anomaly Detection for Gas Turbine Combustors: A Deep Learning Approach. *Proceedings of the Annual Conference of the*.

Yang, Y., Liao, Y., Meng, G., & Lee, J. (2011). A hybrid feature selection scheme for unsupervised learning and its application in bearing fault diagnosis. *Expert Systems with Applications*, *38*(9), 11311–11320. http://doi.org/http://dx.doi.org/10.1016/j.eswa.2011.02.181