An approach for feature extraction and selection from non-trending data for machinery prognosis

James Kuria Kimotho[1], and Walter Sextro[2]

[1,2] *University of Paderborn, Pohlweg 47-49, 33098 Paderborn, Germany*
james.kuria.kimotho@uni-paderborn.de

ABSTRACT

With the paradigm shift towards prognostic and health management (PHM) of machinery, there is need for reliable PHM methodologies with narrow error bounds to allow maintenance engineers take decisive maintenance actions based on the prognostic results. Prognostics is mainly concerned with the estimation of the remaining useful life (RUL) or time to failure (TTF). The accuracy of PHM methods is usually a function of the features extracted from the raw data obtained from sensors. In cases where the extracted features do not display clear degradation trends, for instance highly loaded bearings, the accuracy of the state of the art PHM methods is significantly affected. The data which lacks clear degradation trend is referred to as non-trending data. This study presents a method for extracting degradation trends from non-trending condition monitoring data for RUL estimation. The raw signals are first filtered using a discrete wavelet transform (DWT) denoising filter to remove noise from the acquired signals. Time domain, frequency domain and time-frequency domain features are then extracted from the filtered signals. An autoregressive (AR) model is then applied to the extracted features to identify the degradation trends. Features representing the maximum health information are then selected based on a performance evaluation criteria using extreme learning machine (ELM) algorithm. The selected features can then be used as inputs in a prognostic algorithm. The feasibility of the method is demonstrated using experimental bearing vibration data. The performance of the method is evaluated on the accuracy of RUL estimation and the results show that the method can be used to accurately estimate RUL with a maximum error of 10%.

1. INTRODUCTION

The last one decade has seen focus shifting towards predictive maintenance strategies where maintenance action is taken based on future health state prediction of a component or system. Accurate prediction of the future health or damage propagation of a component provides the maintenance engineers with time to appropriately schedule maintenance without affecting operations. Autonomous systems can also use the predictions to adapt to the prevailing conditions, such that their missions are accomplished. Unlike diagnosis which deals with events that have already occurred, prognosis is much more difficult since it deals with stochastic events that are yet to occur (Kim, Tan, Mathew, & Choi, 2012). Although numerous prognostic methods have been proposed, they are still at the experimental stage (Dragomir, Gouriveau, Minca, & Zerhouni, 2009). This could be attributed to the wide error bounds associated with most algorithms such that maintenance engineers would not have confidence to allow a system to operate once a fault has been identified. Another challenge is the long computational time for both training and prediction, displayed by most methods, rendering them unsuitable for real time prognosis.

Prognosis is a function of the features extracted from the raw data and therefore it is important that the features extracted contain maximum information regarding degradation trend for the predictions to be accurate. Depending on the type of component or system, the observed features may show an increasing or decreasing trend. However, there are situations where the data does not show any observable trend (non-trending data), making long-term prognosis very difficult. Figure 1 (a) shows data with observable degradation trend while (b) shows data with no observable trend until failure.

Various attempts to extract features that represent degradation trends in machinery components have been made. Amongst these attempts is the use of time-frequency methods and autoregressive models. Gu et al., (Gu, Zhao, & Zhang, 2013) introduced a hybrid approach based on autoregressive filter to remove discrete frequencies from a bearing signal and empirical mode decomposition to extract the residual signal which contains the degradation trend. However, the performance of this approach on the ability to provide accurate RUL estimation was not evaluated. Junsheng et al., (Junsheng, De-

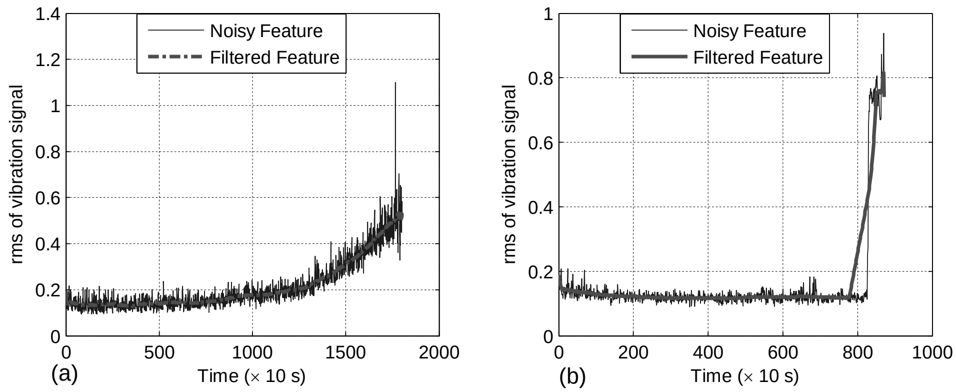

Figure 1. Extracted features from (a) trending data and (b) non-trending data.

jie, & Yu, 2006) presented a fault diagnosis approach based on empirical mode decomposition (EMD) and autoregressive (AR) model which was verified with experimental data on the ability to accurately diagnose bearing faults. Zhang et al., (Y. Zhang, Zuo, & Bai, 2013) proposed a fault diagnosis and performance degradation method based on multiple features, where singular values and autoregressive model parameters are extracted from the results of empirical mode decomposition of the raw signals. Kernel principal component analysis (KPCA) was then employed for feature transformation and reduction. However, the capability of the method to identify performance degradation was not verified. In addition, although PCA has been successfully applied in feature reduction for diagnosis purposes, it may not suitable for prognosis since it orthogonally transforms features which makes them loose important degradation information. Recently, EMD has received a lot of attention as a feature extraction technique for non-stationary signals (Georgoulas, Loutas, Stylios, & Kostopoulos, 2013; Wang, Lu, Wang, Liu, & Fan, 2013; Yu, Dejie, & Junsheng, 2006; Xiong & Yang, 2012). This method is able to enhance impulses in signals, associated with faults and is therefore more suited for fault diagnosis rather than prognosis, where the ability to identify continued degradation is important. Since not all of the extracted features contain important information on degradation trend, it is important to select the features that accurately represent the degradation process.

There has been considerable effort to develop algorithms for automatic feature selection. However, most of these algorithms are focused on feature selection for fault diagnosis or health state based prognosis. In this case, features are selected based on their capability to discriminate between different classes (fault categories or health states). Linear discriminant analysis (LDA) which is based on the assumption that different classes generate data based on Gaussian distributions has been employed in feature selection for fault classification (Bator, Dirks, Monks, & Lohweg, 2012). Distance evaluation technique which computes the largest distance separating data between classes is another feature selection technique that has been employed to select the optimal features that represent the different health states of a degrading component (Kim et al., 2012). Camci et al., (Camci, Medjaher, Zerhouni, & Nectoux, 2012) proposed a feature evaluation method for effective bearing prognostics based on separability value. The features were divided into time segments and the separability of the segments based on 25th and 75th percentile distributions computed. The overall separability value of each feature was then computed as a feature evaluation value. However, the performance of this method for accurate prognosis was not evaluated. The use of separability of features as a method of feature selection can also be found in (Medjaher, Camci, & Zerhouni, 2012). Benkedjouh et al., (Benkedjouh, Medjaher, Zerhouni, & Rechak, 2013) employed isometric feature mapping reduction technique to find a small number of features that represent a large number of observations. The accuracy of the method on ability to improve prognosis was not evaluated. Other methods of feature selection or selection can be found in (Li et al., 2011; Sugumaran, Muralidharan, & Ramachandran, 2007; K. Zhang, Li, Scarf, & Ball, 2011). Saxena and Vachtsevanos, (Saxena & Vachtsevanos, 2007) explored the capabilities of multicore cell processing environment for feature extraction and selection for on-board diagnosis and prognosis. Their effort was concentrated on developing parallel algorithms for Fast Fourier Transforms (FFTs) that could speed up their implementation. Tran and Yang, (Tran & Yang, 2010) presented a method for feature selection based on classification and regression trees. The feature selection was however conducted for classification of faults only and not for prognosis. Ramasso and Gouriveau, (Ramasso & Gouriveau, 2010) proposed a prognostics method involving three modules, observation selection, prediction and classification. A method for feature selection was also presented but found to have high computational requirements. From the literature surveyed, it is evident that there is a need to develop an effective feature selection approach for prognosis based on regression approach.

This paper presents a feature extraction method based on combination of a wavelet denoising filter and autoregressive model with automatic model order selection for feature extraction and the use of kernel based ELM for feature selection based on performance evaluation criteria of the extracted features. This feature extraction approach has the capability of extracting degradation trends from non-trending data. The performance of the method based on ability to provide accurate RUL estimations using ELM is also demonstrated.

2. PROPOSED METHOD

The proposed method involves denoising the raw signals using discrete wavelet transform (DWT) denoising then extracting time, frequency and time-frequency domain features. An AR model is then established for each of the extracted features. The optimum features are then selected using kernel based ELM algorithm. Finally the performance of the method is evaluated using the same kernel based ELM algorithm. Figure 2 shows the workflow of the proposed method.

2.1. Feature Extraction

Feature extraction involves deriving time, frequency and time-frequency domain features from the raw signals which are sampled at suitable frequencies. Signals acquired from some machinery components such as faulty bearing are normally considered non-stationary, that is, frequency varies with time, and hence the extraction of time-frequency features. In this work, wavelet packet decomposition (WPD) is employed for the extraction of the time-frequency features. The denoised signal is decomposed up to 3 levels using bior3.7 wavelet. The detail coefficients from level 1 to 3 and the approximate coefficient for level 3 are then obtained. The wavelet energy is then computed from the wavelet coefficients. Fast Fourier Transform (FFT) is employed to extract the frequency domain features. A total of 19 features, 12 time domain, 3 frequency domain and 4 time-frequency domain from each signal may be extracted from the denoised signals. A summary of these features is presented in Table 1 (Galar, Kumar, & Zhao, 2012; Maio et al., 2012).

2.2. Autoregressive (AR) Model

AR model represents a time series in which the next value in the sequence is predicted based on a certain number of previous values. The AR model parameters may contain important information regarding the condition of a component (Y. Zhang et al., 2013). The following model is established to each of the extracted features f to obtain degradation trend:

$$f_n = \sum_{k=1}^{p} a_k f_{n-k} + e_n, \qquad n = 1, 2...N \tag{1}$$

where a_k are the model parameters, p is the model order, e_n is the residual of the model and N is the number of data points in f. In this work, the model parameters were determined using the Yule-Walker method (Stoica, Friedlander, & Son-

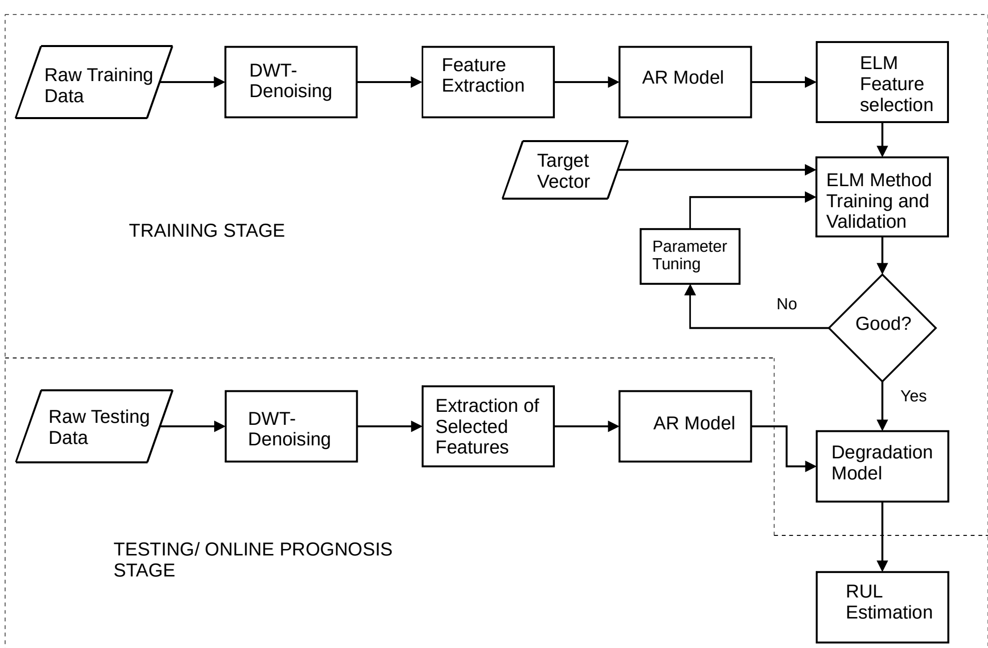

Figure 2. Workflow of the proposed method.

Table 1. Features extracted from denoised signals

Time domain features		Shape factor	$SF = \frac{RMS}{\frac{1}{n}\sum_{i=1}^{n}\lvert x_i\rvert}$	Power spectral density of FFT	$PSD = \sum_{k=-\infty}^{\infty} r_k e^{-i\omega k}$
RMS	$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$	Line integral	$LI = \sum_{i=0}^{n} \lvert x_{i+1} - x_i\rvert$	**Time –frequency domain**	
Variance	$var = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$	Peak to peak value	$PP = \max(x_i) - \min(x_i)$	Energy of WPD detail coefficient one	$D_1 = \sum_{i=1}^{n1} cD1_i$
Peak value	$PvT = \max(\lvert x_i\rvert)$	Shannon entropy	$Ent = -\sum_{i=1}^{n} x_i^2 \log(x_i^2)$	Energy of WPD detail coefficient two	$D_2 = \sum_{i=1}^{n2} cD2_i$
Crest factor	$CF = \frac{PvT}{RMS}$	Skewness	$Sk = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)^3}$	Energy of WPD detail coefficient three	$D_3 = \sum_{i=1}^{n3} cD3_i$
Kurtosis	$Kurt = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{n \times var^2} - 3$	**Frequency domain**		Energy of WPD approximate coefficient three	$A_3 = \sum_{i=1}^{n3} cA3_i$
Clearance factor	$Clf = \frac{PvT}{\left(\frac{1}{n}\sum_{i=1}^{n}\sqrt{\lvert x_i\rvert}\right)^2}$	Peak value of FFT	$PvF = \max(r_k)$, $r_k = \sum_{i=-\infty}^{\infty} x(t)e^{-i\omega t}$		
Impulse factor	$IF = \frac{PvT}{\frac{1}{n}\sum_{i=1}^{n}\lvert x_i\rvert}$	Energy of FFT	$En = \sum_{f=1}^{N} r_k$		

derstrom, 1988). The performance of the AR model depends on the choice of the model order. In this study, the Akaike information criteria, AIC introduced by Akaike was employed (Ayalew, Babu, & Rao, 2012):

$$AIC(p) = log(\hat{\sigma}_p) + \frac{2p}{N}, \tag{2}$$

where,

$$\hat{\sigma}_p = \frac{1}{(N-p)} \sum_{n=p+1}^{N} (f_n - \sum_{k=1}^{p} a_k f_{n-k})^2, \tag{3}$$

The model order is varied from 1 to 100 and the model order yielding the minimum AIC is selected. The feasibility of this approach is demonstrated using the impulse factor IF, extracted from the filtered signal. For each sampled signal with M data points, IF is obtained as follows:

$$IF = \frac{\max(x_K)}{\frac{1}{M}\sum_{K=1}^{M}\lvert x_K\rvert}, \qquad K = 1, 2, ..M \tag{4}$$

Figure 3(a) shows the impulse factor of a bearing vibration signal before application of AR model, in which the degradation trend is not clearly identifiable. Figure 3(b) shows the AR model (f_{IF}) of the feature, which presents a clearer degradation trend or fault evolution trend. The AR model also acts as a filtering method, thus eliminating the noise within the extracted feature.

2.3. Extreme Learning Machine (ELM)

Extreme learning machine is a relatively new simple learning algorithm for single-hidden layer feedforward neural network (SLFN) which was first proposed by Huang in 2005 (Huang, Zhu, & Siew, 2006). Figure 4 shows the structure of a SLFN with radial basis function (RBF) hidden neurons. x_j is the input vector at the input neuron j, a_i is the input weight connecting the hidden neuron i and the input neurons, b_i is the bias of the hidden neuron β_i is the output weight of the hidden neuron i and y is the output (Huang et al., 2006).

In ELM, the input weights and hidden layer biases of SLFN are randomly generated, while the output weights linking the hidden layer to the output layer are determined through simple generalized inverse operation of the hidden layer output matrices (Huang et al., 2006). The ELM learning process is extremely fast compared to other machine learning algorithms such as support vector machines and artificial neural networks with back propagation (Huang et al., 2006). The kernel based ELM has two parameters (regularization parameter C and kernel parameter γ) that tuning. In this work, $C = 7000$ and $\gamma = 2.9$ were employed.

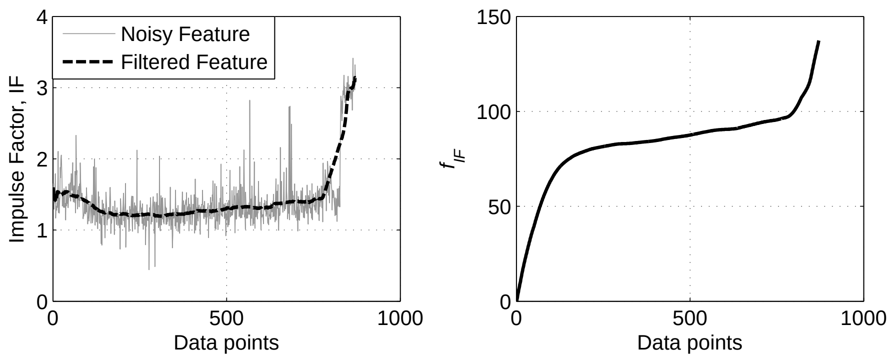

Figure 3. Extracted feature (a) Feature before application of AR model and (b) after applying AR model.

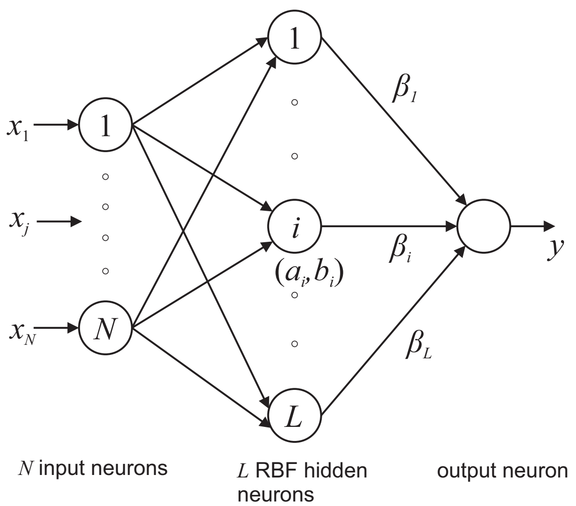

Figure 4. Structure of SLFN with RBF hidden neurons.

2.3.1. ELM Based Feature Selection

Feature selection is important for machinery prognosis in order to reduce computational time and effort, and also to avoid over-fitting of data which results to large prediction errors. In this work, kernel based extreme learning machines was employed for feature selection due to its robust predictions and fast training and prediction times. The AR features are first evaluated individually on their ability to provide accurate prognosis. The input to the ELM method is the AR features while the target vector is the fraction of the remaining useful life. The mean square error computed from the target fractional RUL and estimated fractional RUL of the training data for each individual input feature is obtained and values for all the inputs are normalized between 0 and 1. A performance evaluation criterion, PEC is then defined by:

$$PEC = 1 - \frac{mse}{\max(mse)} \tag{5}$$

where μ is the normalized training mean square error, mse.

Figure 5 shows the workflow of the feature selection algorithm.

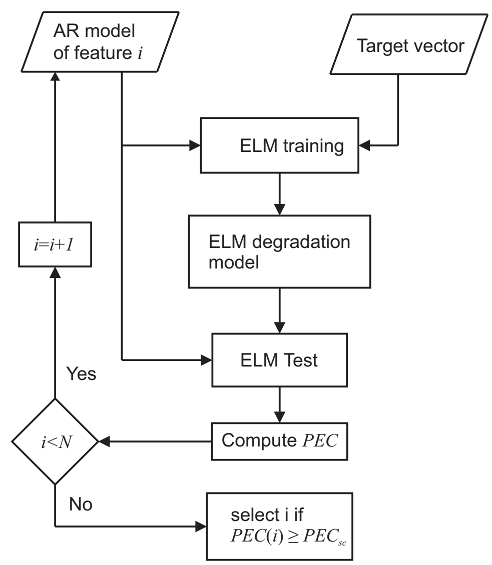

Figure 5. Workflow of the ELM-based feature selection algorithm. i is the feature index and N is the total number of features.

To obtain the selection criteria PEC_{sc}, the PEC is varied from 0 to 1 and the mse of the training data set is obtained. The PEC value that yields the minimum mse is taken as the feature selection criteria.

2.3.2. ELM Based RUL Estimation

During the training stage of the method, the selected features are used as inputs to the PHM algorithm while the fractional remaining useful life is used as the target vector. The fractional RUL is used to take care of the varying lifetimes of machinery components. A degradation model is obtained after training, which is used together with the testing input features to predict the fractional RUL of the test data.

Given the current time, t_c, and the fractional RUL, F_c, the es-

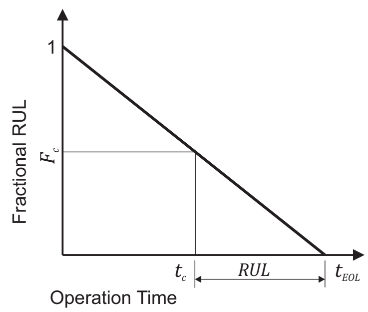

Figure 6. Estimating the RUL from the current time and fractional RUL.

timated remaining useful life RUL can be obtained by similar triangles as follows:

$$RUL = t_{EOL} - t_c = \frac{t_c}{1 - F_c} F_c \tag{6}$$

where t_{EOL} is the time to end of life of the component.

3. APPLICATION EXAMPLE

To demonstrate the applicability of the method, a case study was conducted. Run to failure rolling element bearing data provided for the 2012 PHM data challenge was employed (Nectoux et al., 2012). The data consists of run to failure vibration data recorded by two accelerometers, along the vertical direction and along the horizontal direction, sampled at a frequency of 25.6 kHz with 2560 samples recorded at intervals of 10 seconds. Two complete run to failure data sets are provided for algorithm training and five truncated run to failure data are provided for testing. The challenge is to provide an estimation of the remaining useful life of the test bearings (Nectoux et al., 2012).

The features detailed in section 2.1 were extracted and an AR model applied. The proposed feature selection method described in section 2.3.1 was then applied. Figure 7 shows the mse as a function of PEC.

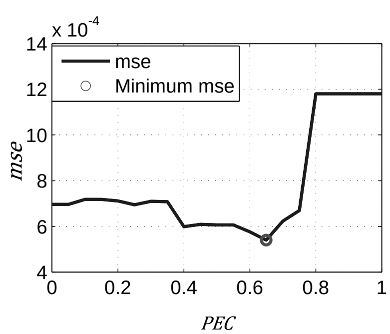

Figure 7. Mean square error, mse as a function of performance evaluation criteria, PEC for the training data set.

From Figure 7, it is evident that features with a performance evaluation criteria value of 0.65 yield the lowest mse. Therefore a selection criterion of $PEC_{sc} = 0.65$ was employed in this study. Based on this selection criterion, 11 out of 38 features were selected. Figure 8 shows the PEC value of each feature. It can be observed that not many features from the vertical acceleration were selected. The vibration signal from the vertical accelerometer was highly impulsive which led to high mean square errors. Although the features extracted from the vertical accelerometer may not be suitable for prognosis, they may provide valuable information about the nature and location of faults within the bearings.

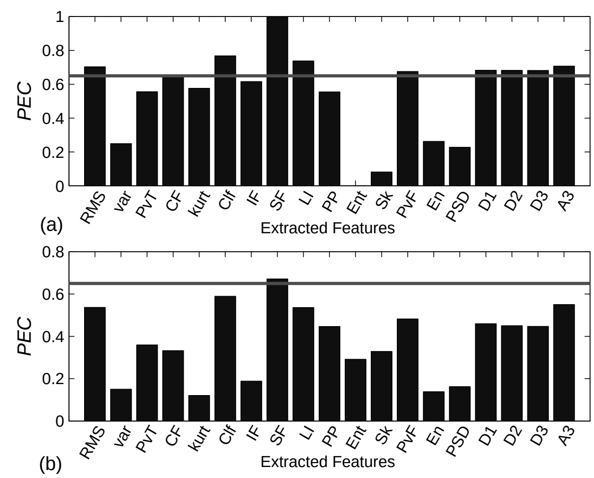

Figure 8. Performance evaluation criteria of the extracted features described in section 2.1, (a) horizontal vibration and (b) vertical vibration.

The selected features were then extracted from the denoised signals of the training and test data. An AR model was applied to the resulting features in order to obtain inputs to the ELM algorithm. The ELM method was then trained with the AR features as the input and fractional lifetime as the target vector. A degradation model consisting of the number of neurons, the input and output weights of the hidden layer was obtained. The AR features from the test data were then used as inputs to the degradation model and the estimated fractional lifetime obtained as the output.

Using Eq. 6, the RUL of the five test bearings were computed from the fractional lifetime obtained as the output from the ELM algorithm. Figure 9 shows curves of the estimated RUL, the actual RUL and predicted RUL of bearing 1_3. RUL_c is the RUL at the current time. The predicted RUL is obtained by fitting a linear curve from the current time to the point where the RUL is zero

Figure 9 shows that the accuracy of the method increases towards the end of life of the component. This is the most critical stage of the prognosis since it signifies that the maintenance engineers should start planning for maintenance.

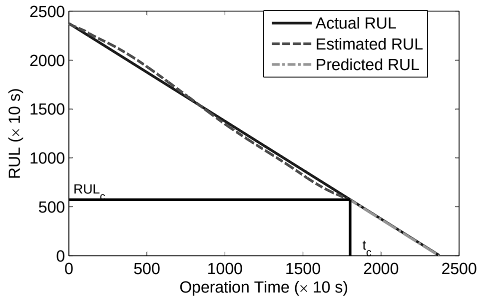

Figure 9. Estimated, actual and predicted RUL for bearing 1_3 versus the operation time.

The performance of the method was evaluated based on performance metrics such as percentage RUL error and percentage accuracy as shown below:

$$\%Error = \frac{ActRUL - estRUL}{ActRUL} \times 100, \quad (7)$$

where $ActRUL$ is the actual RUL of the test data provided for validation of the prognostic methods and $estRUL$ is the estimated RUL. Table 2 shows the performance of the proposed method based on percentage error and percentage accuracy. The estimated RUL is based on the current time.

Table 2. Performance of the proposed method based on prognostic performance metrics.

Test	% Error
Bearing 1_3	0.44
Bearing 1_4	5.31
Bearing 1_5	-4.94
Bearing 1_6	-8.41
Bearing 1_7	3.71

The negative percentage error signifies late prediction or overestimation of RUL, which is usually not desirable in machinery prognosis since the machine may breakdown before the scheduled maintenance, depending on the margin of error of the estimation. Table 1 shows the proposed method yields accuracies within 10% error bounds. This would be a good reference for maintenance.

4. CONCLUSION

The accuracy of any prognostic algorithm is a function of the information contained in the input features which are extracted from the raw condition monitoring data. A method for feature extraction and selection for machinery prognosis based on autoregressive modeling and extreme learning machine is presented. The proposed feature extraction method is able to identify degradation trends in condition monitoring data for effective prognosis. The feasibility of the method is demonstrated using bearing run-to-failure experimental data. The results show that the proposed method is effective in estimating the remaining useful life of machinery components, with error bounds within a 10% bandwidth. The results also show that feature selection improves the accuracy of RUL estimation significantly. Therefore, we conclude that the proposed method of feature extraction and selection can be used as an effective tool for estimating the remaining useful life of machinery components.

5. ACKNOWLEDGEMENT

The authors wish to acknowledge the support of the Chair, Department of Mechatronics and Dynamics, University of Paderborn, German Academic Exchange Service (DAAD) and the Ministry of Higher Education Science and Technology (MoHEST), Kenya.

REFERENCES

Ayalew, S., Babu, M. C., & Rao, L. K. M. (2012). Comparison of new approach criteria for estimating the order of autoregressive process. *IOSR Journall of Mathematics*, *1*(3), 10–20.

Bator, M., Dirks, A., Monks, U., & Lohweg, V. (2012). Feature extraction and reduction applied to sensorless drive diagnosis. *22. Workshop, Computational Intelligence, Dortmund*, 163–178.

Benkedjouh, T., Medjaher, K., Zerhouni, N., & Rechak, S. (2013). Remaining useful life estimation based on nonlinear feature reduction and support vector machines. *Engineering Application of Artificial Intelligence*, *26*, 1751–1760.

Camci, F., Medjaher, K., Zerhouni, N., & Nectoux, P. (2012). Feature evaluation for effective bearing prognostics. *Quality and Reliability Engineering International*.

Dragomir, O. E., Gouriveau, F. D., Minca, E., & Zerhouni, N. (2009). Review of prognostic problem in condition-based maintenance. *European Control Conference, ECC'09*.

Galar, D., Kumar, U., & Zhao, W. (2012). Remaining useful life estimation using time trajectory tracking and support vector machines. In *25th international congress on condition monitoring and diagnostic engineering.*

Georgoulas, G., Loutas, T., Stylios, C. D., & Kostopoulos. (2013). Bearing fault detection based on hybrid ensemble detector and empirical mode decomposition. *Mechanical Systems and Signal Processing*, *41*(1-2), 510–525.

Gu, H., Zhao, J., & Zhang, X. (2013). Hybrid methodology of degradation feature extraction for bearing prognostics. *Mainten ance and Reliability*, *15*(2), 195–201.

Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, *70*, 489–501.

Junsheng, C., Dejie, Y., & Yu, Y. (2006). A fault diagnosis approach for roller bearings based on emd method and ar model. *Mechanical Systems and Signal Processing*, *20*, 350–362.

Kim, H.-E., Tan, A. C. C., Mathew, J., & Choi, B.-K. (2012). Bearing fault prognosis based on health state probability estimation. *Expert Systems with Applications*, *39*, 5200–5213. doi: 10.1016/j.eswa.2011.11.019

Li, B., Zhang, P.-L., Tian, H., Mi, S.-s., Liu, D.-s., & Ren, G.-q. (2011). : A new feature extraction and selection scheme for hybrid fault diagnosis of gearbox. *Expert Systems with Applications*, *38*, 10000–10009.

Maio, F. D., Hu, J., Tse, P., Petch, M., Tsui, K., & Zio, E. (2012). Ensemble-approaches for clustering health status of oil sand pumps. *Expert Systems with Applications*, *39*, 4847-4859.

Medjaher, K., Camci, F., & Zerhouni, N. (2012). Feature extraction and evaluation for health assessment and failure prognostics. *Proceedings of First European Conference of the Prognostics and Health Management Society,*.

Nectoux, P., Gouriveau, R., Medjaher, K., Ramasso, E., Morello, B., Zerhouni, N., & Varnier, C. (2012). Pronostia: An experimental platform for bearing accelerated degradation tests. *IEEE International Conference on Prognostics and Health Management, Denver, CO, USA*.

Ramasso, E., & Gouriveau, R. (2010). Prognostics in switching systems: Evidential markorvian classification of real-time neuro-fuzzy predictions. In *Ieee international conference on prognostics and system health management, macau.*

Saxena, A., & Vachtsevanos, G. (2007). Optimum feature selection and extraction for fault diagnosis and prognosis. In *Proceedings of the 2007 aaai fall symposium on artificial intelligence for prognosis.*

Stoica, P., Friedlander, B., & Sonderstrom, T. (1988). A high-order yule-walker method for estimation of the ar parameters of an arma model. *Systems and Control Letters*, 99–105.

Sugumaran, V., Muralidharan, V., & Ramachandran, K. I. (2007). Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearings. *Mechanical Systems and Signal Processing*, *21*, 930–942.

Tran, V. T., & Yang, B.-S. (2010). Machine fault diagnosis and condition prognosis using classification and regression trees and neuro-fuzzy inference systems. *Control and Cybernetics*, *39*, 22-54.

Wang, Z., Lu, C., Wang, Z., Liu, H., & Fan, H. (2013). Fault diagnosis and health assessment for bearings using the mahalanobis-taguchi system based on EMD-SVD. *Transactions of the Institute of Measurement and Control*, *35*(6), 798–807.

Xiong, X., & Yang, S. (2012). A new procedure for extracting fault feature of multi-frequency signal from rotating machinery. *Mechanical Systems and Signal Processing*, *32*, 306–319.

Yu, Y., Dejie, Y., & Junsheng, C. (2006). A roller bearing fault diagnosis method based on emd energy entropy and ann. *Journal of Sound and Vibration*, *294*, 269–277.

Zhang, K., Li, Y., Scarf, P., & Ball, A. (2011). Feature selection for high-dimensional machinery fault diagnosis data using multiple models and radial basis function networks. *Neurocomputing*, *74*, 2941–2952.

Zhang, Y., Zuo, H., & Bai, F. (2013). Classification of fault location and performance degradation of a roller bearing. *Measurement*, *46*, 1178–1189.

BIOGRAPHIES

James Kuria Kimotho studied mechanical engineering at the Jomo Kenyatta University of Agriculture and Technology, Kenya. Since 2012, he is with the research group Mechatronics and Dynamics at the University of Paderborn. His research focuses on prognostic and health management of mechatronic systems.

Walter Sextro studied mechanical engineering at the Leibniz University of Hanover and at the Imperial College in London. Afterwards, he was development engineer at Baker Hughes Inteq in Celle, Germany and Houston, Texas. Back as research assistant at the University of Hanover he was awarded the academic degree Dr.-Ing. in 1997. Afterward he habilitated in the domain of mechanics under the topic Dynamical contact problems with friction: Models, Methods, Experiments and Applications. From 2004-2009 he was professor for mechanical engineering at the Technical University of Graz, Austria. Since March 2009 he is professor for mechanical engineering and head of the research group Mechatronics and Dynamics at the University of Paderborn.