# Using Deep Learning Based Approaches for Bearing Remaining Useful Life Prediction

Jason Deutsch and David He

*Department of Mechanical and Industrial Engineering, University of Illinois at Chicago, Chicago, Illinois, 60607, USA*

*jdeuts4@uic.edu*
*davidhe@uic.edu*

## ABSTRACT

Traditional data driven prognostics requires establishing explicit model equations and much prior knowledge about signal processing techniques and prognostic expertise, and therefore is limited in the age of big data. This paper presents a deep learning based approach for bearing remaining useful life (RUL) prediction with big data. This approach has the ability to automatically extract important features that can be used for RUL predictions. The presented approach is tested and validated using data collected from bearing run-to-failure tests and compared with existing PHM methods. The test results show the promising bearing RUL prediction performance of the deep learning based approach.

## 1. INTRODUCTION

In the age of Internet of Things and Industrial 4.0, the prognostic and health management (PHM) systems are used to collect massive real-time data from the mechanical equipment. Mechanical big data has the characteristics of large-volume, diversity and high-velocity. Effectively mining features from such data and accurately predicting the remaining useful life (RUL) of the equipment in use with new advanced methods become new issues in PHM. Traditionally, data driven prognostics is largely dependent on signal processing and feature extraction techniques. Over the past years, many prognostic methods that require explicit model equations have been developed (Vachtsevanos *et al*., 2006). For example, recurrent neural networks (Malhi *et al*. 2011, Heimes 2008), Kalman filter (Lim *et al*. 2014, Bechhoefer *et al*. 2010), and particle filter (Daigle and Goebel 2013, Baraldi *et al*. 2013, Chen *et al*. 2011,). This critical process of establishing explicit model equations requires much prior knowledge about signal processing techniques and prognostic expertise.



Since its introduction by Hinton *et al*. (2006), deep learning method has become a popular approach for big data process and analysis. Deep learning has the ability to yield useful and important features from data that can ultimately be useful for improving predictive power (Bengio *et al*. 2013). It has also the capability of processing big data and mining hidden information due to its multiple layer structure. The recent success of AlphaGo by Google Deepmind has demonstrated the power of deep learning for big data processing and feature learning. There have been great successes in building deep neural network architectures in various domains such as image recognition, automatic speech recognition, and natural language processing (LeCun *et al.* 2015), and many more. It has also recently shown promising results for machine fault diagnostics on extraction of raw vibration signals (Chen *et al*. 2016) as well as time-domain features (Shao *et al*. 2015). Although much success in deep learning has been focused on classification problems, deep learning has also proven to be successful in solving prediction problems. These domains include predicting car traffic (Lv *et al.* 2015), weather (Hossain *et al.* 2015), wind speed (Tao *et al.* 2014), and internet traffic (Oliveira *et al*. 2014). There are many types of deep learning algorithms present including auto encoders, restricted Boltzman machines, deep belief networks, convolutional neural networks, and more that can also be used for prediction problems. Deep learning represents an attractive option to process mechanical big data for RUL prediction as deep learning has the ability to automatically select features that otherwise require much skill, time, and experience.

In this paper, a deep learning based approach for bearing remaining useful life prediction using vibration sensors is presented. The presented approach is tested and validated using data collected from bearing run-to-failure tests and compared with existing PHM methods. The test results show the promising bearing RUL prediction performance of the presented method.

## 2. THE METHODOLOGY

A Restricted Boltzman Machine (RBM) is used in this paper to model vibration data in order to make predictions $L$ steps ahead in the future. Next, the RBM and RBM based bearing RUL prediction are explained.

### 2.1. Deep Learning Method – The Restricted Boltzman Machine

A Restricted Boltzman Machine is a type of unsupervised machine learning algorithm. It is a generative stochastic artificial neural network that learns a probability distribution over the set of its inputs. A RBM is a bipartite graph, which contains undirected edges from its two layers: a visible and a hidden layer. Each layer contains a collection of neurons/nodes. The visible layer consists of the data's input, where each node/neuron represents a feature of the data. The hidden layer represents the latent variables. As shown in Figure 1, an RBM is "Restricted" because there are no connections between each neuron/node within either the visible or hidden layers. An RBM contains a matrix of weights $W_{ij}$ representing the connection to visible node $v_i$ and hidden node $h_j$. We will let $a_i$ represent the bias term for the visible layer and $b_j$ for the hidden layer.

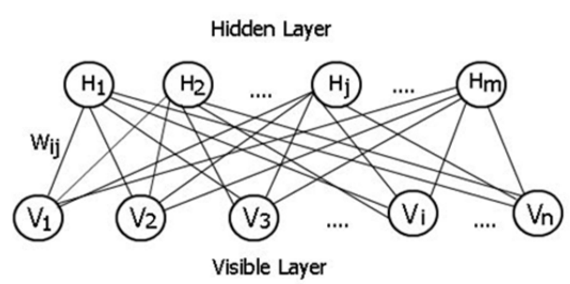


Figure 1. A Restricted Boltzmann Machine

Although not shown in the figure, the bias terms can just be thought of as an extra node for each layer with a fixed value of 1, where the weight/edge simply defines the value of the bias term. The weights and biases are computed by maximizing $P(\boldsymbol{v})$, which is the probability that the network assigns to a visible vector $\boldsymbol{v}$:

$$P(v) = \frac{1}{Z}\sum_{\boldsymbol{h}} e^{-E(\boldsymbol{v},\boldsymbol{h})} \quad (1)$$

where $Z$ is the normalization constant which can be found by summing over all the possible pairs of visible and hidden vectors:

$$Z = \sum_{\boldsymbol{v}}\sum_{\boldsymbol{h}} e^{-E(\boldsymbol{v},\boldsymbol{h})} \quad (2)$$

and the energy function of the joint configuration $(\boldsymbol{v},\boldsymbol{h})$ is given by:

$$E(\boldsymbol{v},\boldsymbol{h}) = -\boldsymbol{a}^T\boldsymbol{v} - \boldsymbol{b}^T\boldsymbol{h} - \boldsymbol{v}^T\boldsymbol{W}\boldsymbol{h} \quad (3)$$

The max of (1) can in theory, be determined by taking its partial log derivative with respect to its parameters $\boldsymbol{W}$, $\boldsymbol{a}$, and $\boldsymbol{b}$:

$$\frac{\partial \log(P(\boldsymbol{v}))}{\partial \boldsymbol{W},\boldsymbol{a},\boldsymbol{b}} = \sum_{v,h} E(v,h) - \sum_{h} E(v,h) \quad (4)$$

Typically (4) is also written as:

$$\frac{\partial \log(P(\boldsymbol{v}))}{\partial \boldsymbol{W},\boldsymbol{a},\boldsymbol{b}} = < v_i h_j >_{data} - < v_i h_j >_{model} \quad (5)$$

where <.> denotes the expectation. However, the expectation $< v_i h_j >_{model}$ in the maximum log likelihood function cannot be easily computed and is thus estimated using Contrastive Divergence (Hinton 2006), which leads to the following weight update equation:

$$W_{ij}{}^k = W_{ij}{}^k + \mathrm{Ŋ}(< v_i{}^k h_j >_{data} - < v_i{}^k h_j >_T) \quad (6)$$

where $T$ represents a full step of Gibbs sampling, Ŋ represents the learning rate and $k$ represents the $k-step$ of contrastive divergence. The biases are also computed by the same process. The neuron activation probabilities are given by the following equations:

$$P(h_j = 1 |\, \boldsymbol{v}) = \sigma(b_j + \sum_{i=1}^{n} W_{ij} v_i) \quad (7)$$

$$P(v_i = 1 |\, \boldsymbol{h}) = \sigma(a_i + \sum_{j=1}^{m} W_{ij} h_j) \quad (8)$$

where σ denotes the logistic function defined as:

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad (9)$$

$n$ represents the number of visible units and $m$ represents the number of hidden units.

### 2.2. Deep Learning Based Bearing Prognostics

In order to use a RBM as a discriminative model, we can first learn the weights and biases in the unsupervised stage of learning illustrated in the previous section. Once the optimal parameters have been determined the output of the last layer (hidden features learned) can be used as an input to a supervised learning algorithm; in this paper a linear regression layer was used as the last layer to make the $L$-step ahead predictions.

The root mean square (RMS) values are used as the fault feature to determine the bearing's degradation over time. This feature serves as the input into the RBM. The RMS at each time interval (denoted as $X_t$) can be calculated as follows:

$$X_t = \sqrt{\frac{1}{n}\sum_{i=1}^{n} f_{ti}^2} \tag{10}$$

where $f_{ti}$ represents the $ith$ raw vibration data point at time interval $t$ and $n$ is the length of the signal.

The time series data of the RMS then must be reconstructed, into a matrix, where each feature (column) represents a lagged order of the time series, the output is the $L$-step ahead (future) RMS value, and each row represents an index in time. Formally, the input can be denoted as:

$$[X_t, X_{t-1}, \dots, X_{t-d+1}], \ \epsilon\, \mathbb{R}^d \tag{11}$$

and the output as:

$$[X_{t+L}, X_{t+L+1}, \dots, X_n] \tag{12}$$

where $d$ represents the embedding dimension and determines the size of the visible layer in the RBM. Once the data has been constructed the RBM can essentially perform automatic feature engineering in order to better capture the dependency of the lagged RMS values onto the future RMS values and thus avoiding the use of some more complex manual feature extractions of the data.

The predicted RUL can then be computed by using the predicted RMS values and the time of the bearing's failure. One can estimate the predicted RUL by the following equation:

$$\widehat{RUL_t} = t_{life} - \phi(F_t) \tag{13}$$

where $\widehat{RUL_t}$ is the predicted remaining useful life at some time $t$, $t_{t_{life}}$ is the total time of the bearing's life, and $\phi(F_t)$ is a function that maps $F_t$, the predicted RMS value, to an estimated point in time of the bearing's life.

### 3. EXPERIMENTAL SETUP

To validate the deep learning based bearing RUL prediction method, vibration data collected from hybrid ceramic bearing run-to-failure tests was used. Figure 1 shows the customized bearing run-to-failure test rig in the laboratory. The key features of the test rig include: (1) It is driven by a 3-HP AC motor with a maximum speed up to 3600 rpm and variable speed controller. (2) It is equipped with a hydraulic dynamic loading system with a maximum radial load up to 4400 lbs or 19.64 kN. (3) An integrated loading and bearing housing that can be used for testing both ball and tapered roller bearings.

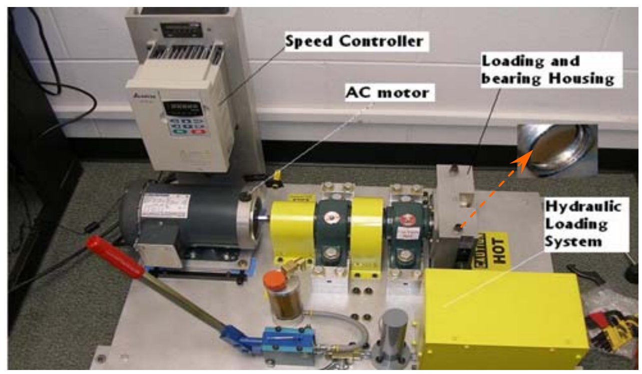


Figure 2. The bearing run-to-failure test rig

An automatic data acquisition system based on National Instrument CI 4462 board and NI LabVIEW software was constructed for data collection purpose. The automatic data acquisition system has the following features: (1) Maximum sampling rate up to 102.4 kHz. (2) 4 Input simultaneous anti-aliasing filters. (3) Software-configurable AC/DC coupling and IEPE conditioning. (4) Vibration analysis functions such as envelope analysis, cepstrum analysis, and so on for computing necessary condition indicators.

Two hybrid ceramic bearings used in the test were ball bearings with stainless steel inner and outer races and ceramic balls. The bearings were mounted on the test rig. Two accelerometers were stunt mounted on the bearing housing in the direction perpendicular to the shaft. The test bearing was mounted on the test rig and the rig was run at a speed of 1800 rpm (30 Hz) and was subjected to a radial load of 600 psi. A sampling rate of 102.4 kHz was used for 2 seconds of data collection at each sampling point. The data was collected every 5 minutes during the test. At the end of the test, the test bearing was disassembled, checked, and photographed. For the first bearing, a total of 173 data files with a length of 14.42 hours were used. For the second bearing, there a total of 804 data files with a length of 67 hours were used. Table 1 provides the run-to-failure test setting for the two bearings and Table 2 the specification of the tested bearings.

Table 1. The run-to-failure test setting

| **Name** | **Type** | **Load (psi)** | **Input Shaft Speed (Hz)** |
|---|---|---|---|
| B1 | Hybrid Ceramic Bearing | 600 | 30 |
| B2 | Hybrid Ceramic Bearing | 600 | 30 |

Table 2. Hybrid ceramic bearing specifications

| **Parameter** | **Specification** |
|---|---|
| Bearing Material | Stainless Steel 440c |
| Ball Material | Ceramic SI3N4 |
| Inner Diameter (d) | 25 m |
| Outer Diameter (D) | 52 m |
| Width (B1) | 15 m |
| Enclosure | Two Shields |
| Enclosure Material | Stainless Steel |
| Enclosure type | Removable (S) |
| Retainer Material | Stainless Steel |
| ABEC/ISO Rating | ABEC #3 / ISOP6 |
| Radial Play | C3 |
| Lube | Klubber L55 Grease |
| RPM Grease (x 1000 rpm) | 19 |
| RPM Oil (x 1000): | 22 |
| Dynamic Load (Kgf) | 1429 |
| Basic Load (Kgf) | 804 |
| Working Temperature Deg C | 121 |
| Weight (g) | 110.32 |

## 4. THE RESULTS

A total of 849 RMS values were recorded at an interval $t$ of 5 minutes based on 2 seconds of data collection at a sampling rate of 102.4 kHz for bearing B2 and a total of 255 RMS values were recorded for bearing B1. A RBM with a linear regression layer as the last layer of the network was developed to model the RMS values. Since the RBM assumes a binary input or a real valued input between [0, 1] the input values were scaled to be between [0,1]. An embedding dimension $d$ = 50 and a training size of 250 examples was empirically found to yield good results in the model for bearing B2 and just 33 examples for bearing B1. Vibration data obtained after the bearings failed were removed from the dataset before training. The hyperparamters were found by using a grid search (exhaustive search). The root mean squared error (RMSE) was used as metrics to determine the most appropriate model. The mean absolute percentage error (MAPE) was also recorded for each model. The MAPE and RMSE are defined by the following equations:

$$MAPE = \left( \frac{1}{n} \sum_{t=1}^{n-L} \frac{A_t - F_t}{A_t} \right) * 100\% \quad (14)$$

$$RMSE = \frac{1}{n} \sum_{t=1}^{n-L} (A_t - F_t)^2 \quad (15)$$

In (14) and (15), $A_t$ = actual value and $F_t$ = predicted value from the Restricted Boltzman Machine and two step values of $L$ = 1 and $L = 10$ were used to predict 5 minutes and 50 minutes respectively into the future for both bearings. Figure 3 and Figure 4 show the plots of RBM predicted RMS values vs. actual RMS values for bearing B2 with $L$ = 1 and $L$ = 10, respectively.

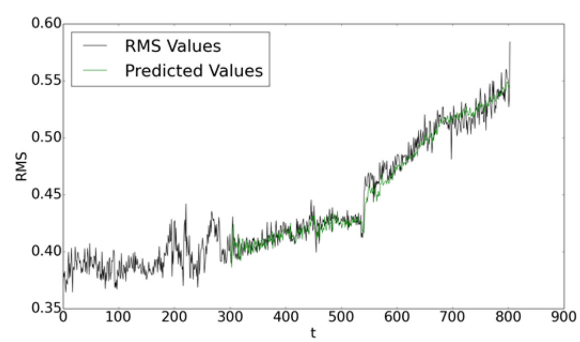


Figure 3. Plot of RBM predicted RMS values vs. actual RMS for bearing B2 with $\boldsymbol{L}$ = 1

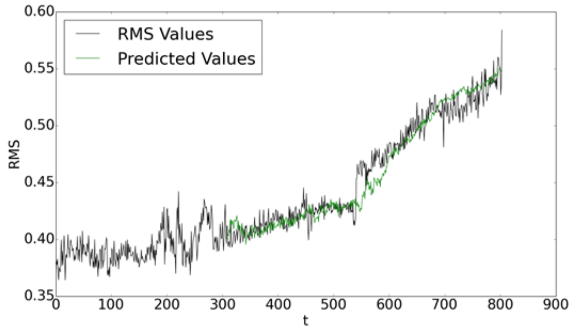


Figure 4. Plot of RBM predicted RMS values vs. actual RMS for bearing B2 with $\boldsymbol{L}$ = 10

Figure 5 and Figure 6 show the plots of RBM predicted RMS values vs. actual RMS values for bearing B1 with $L$ = 1 and $L$ = 10, respectively.

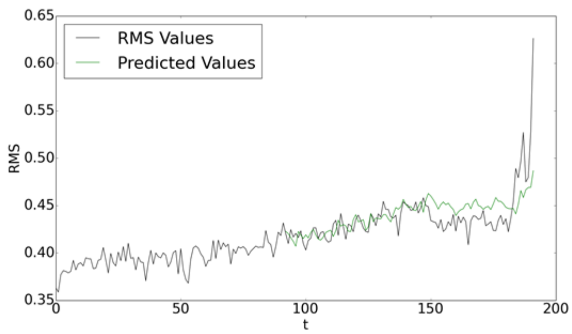


Figure 5. Plot of RBM predicted RMS values vs. actual RMS for bearing B1 with $\boldsymbol{L}$ = 1

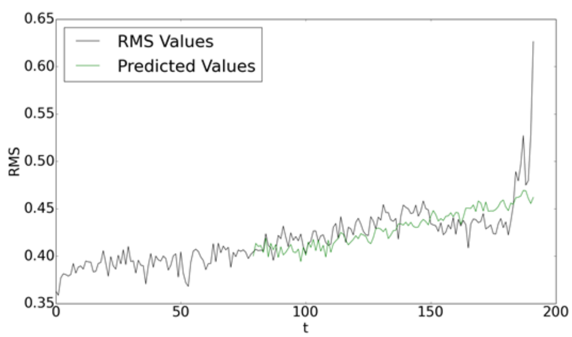


Figure 6. Plot of RBM predicted RMS values vs. actual RMS for bearing B1 with $\boldsymbol{L}$ = 10

From the above figures, it can be seen that the RBM model is able to capture much of the dynamics of the vibration data well throughout the predictions and stay within most of the noise in the data and is able to capture the overall trend of the data.

To evaluate the RUL prediction performance of the RBM model, the last 100 testing points over a time period of 500 minutes (about 8 hours) were used to estimate the bearing RUL. Figure 7 and Figure 8 show the plots of estimated RUL $\widehat{RUL}_t$ vs. true RUL for bearing B2 with $L$ = 1 and $L$ = 10, respectively.

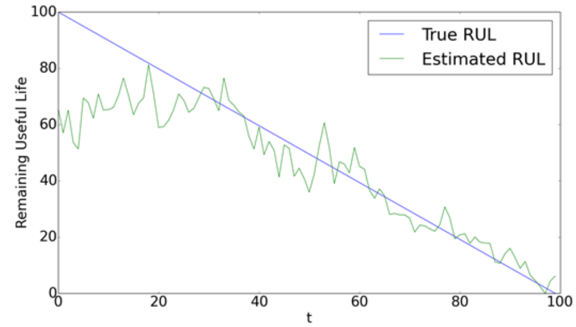


Figure 7. Plot of $\boldsymbol{\widehat{RUL}_t}$ values of bearing B2 with $\boldsymbol{L}$ = 1

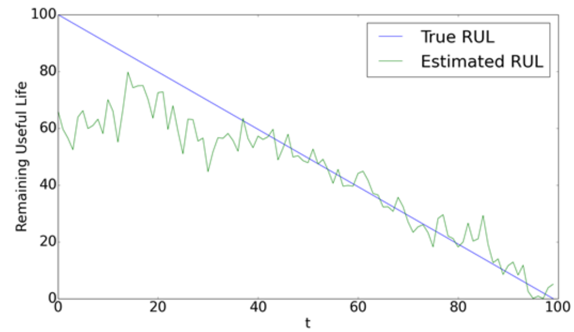


Figure 8. Plot of $\boldsymbol{\widehat{RUL}_t}$ values of bearing B2 with $\boldsymbol{L}$ = 10

Figure 9 and Figure 10 show the plots of estimated RUL $\widehat{RUL}_t$ vs. true RUL for bearing B1 with $L$ = 1 and $L$ = 10, respectively.

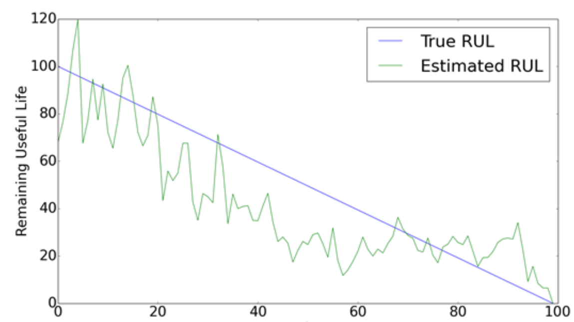


Figure 9. Plot of $\boldsymbol{\widehat{RUL}_t}$ values of bearing B1 with $\boldsymbol{L}$ = 1

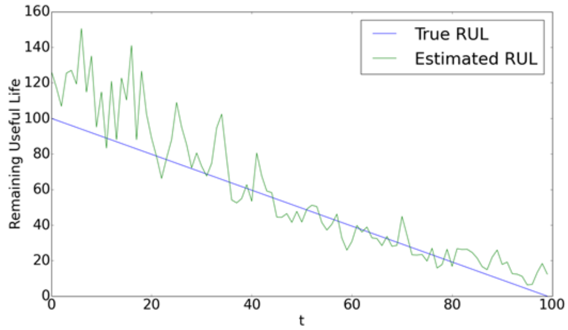


Figure 10. Plot of $\widehat{RUL_t}$ values of bearing B1 with $\boldsymbol{L}$ = 10

From the above figures, it can be seen that the RBM deep learning based approach gives good RUL prediction especially when it approaches to the end of the bearing life. The figures also show that as more data points were fed into the deep learning model, the more accurate RUL prediction was generated by the deep learning model. In addition, RMSE and MAPE were computed for the RUL predictions obtained by the deep learning model and compared with those obtained by particle filter based approach (Li *et al.* 2010). Table 3 and Table 4 show the RMSE and MAPE values of the predictions obtained by deep learning based and particle filter based approaches for bearing B2 and B1, respectively.

Table 3. RMSE and MAPE results of bearing B2

| **Deep learning based approach** | | | | |
|---|---|---|---|---|
| ***L*** | MAPE | RMSE | Learning Rate | Hidden Layer Size |
| 1 | 21.62% | 12.85 | 0.120 | 50 |
| 10 | 23.24% | 13.68 | 0.130 | 50 |
| **Particle filter based approach** | | | | |
| ***L*** | MAPE | RMSE | | |
| 1 | 7.47% | 2.53 | | |
| 10 | 8.73% | 3.65 | | |

Table 4. RMSE and MAPE results of bearing B1

| **Deep learning based approach** | | | | |
|---|---|---|---|---|
| ***L*** | MAPE | RMSE | Learning Rate | Hidden Layer Size |
| 1 | 34.99% | 15.86 | 0.001 | 11 |
| 10 | 43.65% | 20.79 | 0.195 | 91 |
| **Particle filter based approach** | | | | |
| ***L*** | MAPE | RMSE | | |
| 1 | 10.56% | 5.87 | | |
| 10 | 12.42% | 7.21 | | |

From Table 3 and Table 4, it can be seen that the deep learning based approach achieved lower accuracy than the particle filter based approach. However, given that the deep learning based approach doesn't require explicit model equations like particle filter based approach and is scalable for big data applications, the RUL prediction performance achieved by the deep learning based approach has shown its potential for bearing RUL prediction with big data.

## 5. CONCLUSIONS

In the age of Internet of Things and Industrial 4.0, the PHM systems are used to collect massive real-time data from mechanical equipment. PHM big data has the characteristics of large-volume, diversity and high-velocity. Effectively mining features from such data and accurately predicting the remaining useful life of the equipment in use with new advanced methods become new issues in PHM. Traditional data driven prognostics requires establishing explicit model equations and much prior knowledge about signal processing techniques and prognostic expertise, and therefore is limited in the age of big data.

In this paper, a deep learning based approach for bearing remaining useful life prediction using vibration sensors was presented. The presented approach was developed based a RBM. The RBM and RBM based bearing RUL prediction using vibration data were discussed and explained in the paper. To validate the presented deep learning based approach, vibration data collected from two hybrid ceramic bearing run-to-failure tests were used to predict the bearing RUL using the RBM model. The bearing RUL prediction performance of the deep learning based approach was compared with that of a particle filter based approach. The results have shown that the deep learning based approach achieved lower accuracy than the particle filter based approach. However, given that the deep learning based approach doesn't require explicit model equations like particle filter based approach and is scalable for big data applications, the RUL prediction performance achieved by the deep learning based approach has shown its promising capability for bearing RUL prediction with big data. One possible way to improve the accuracy of the presented method is to use stacked RBM structure like that of deep brief network.

## REFERENCES

Baraldi, P., Compare, M., Sauco, S., & Zio, E. (2013). Ensemble neural network-based particle filtering for prognostics. *Mechanical Systems and Signal Processing*, vol. 41, no. 1, pp. 288–300.

Baraldi, P., Mangili, F., & Zio, E. (2012). A kalman filter-based ensemble approach with application to turbine creep prognostics. *IEEE Transactions Reliability*, vol. 61, pp. 966 – 977.

Bechhoefer, E., Clark, S., & He, D. (2010). A state space model for vibration based prognostics. *Proceedings of the 2010 Annual Conference of the Prognostics and Health Management Society*, Portland, OR, October 10 – 16.

Bengio, Y., Courville, A. & Vincent, P. (2013). Representation learning: A review and new perspectives, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 35, no. 8, pp. 1798-1828.

Chen, C., Zhang, B., Vachtsevanos, G., & Orchard, M. (2011). Machine condition prediction based on adaptive neuro–fuzzy and high-order particle filtering. *IEEE Transactions on Industrial Electronics*, vol. 58, no. 9, pp. 4353–4364.

Chen, Z., Zeng, X., Li, W., & Liao, G (2016). Machine fault classification using deep belief network, *2016 IEEE International Instrumentation and Measurement Technology Conference Proceedings*, Taipei, Taiwan, pp. 1-6.

Daigle, M. J. and Goebel, K. (2013). Model-based prognostics with concurrent damage progression processes. *IEEE Transactions on Systems, Man, Cybernetics*, vol. 43, no. 3, pp. 535–546.

Heimes, F. (2008). Recurrent neural networks for remaining useful life estimation. *Proceedings of the 2008 IEEE International Conference of Prognostics and Health Management*, Denver, CO, pp. 1–6.

Hinton, G. E, Osindero, S., & The, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Computing,* vol. 18, no. 7, pp.1527-1554.

Hossain, M., Rekabdar, B., Louis, S. J., & Dascalu, S. (2015). Forecasting the weather of Nevada: A deep learning approach. *International Joint Conference on Neural Networks (IJCNN),* vol., no., pp.1-6.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature.*

Li, R., Ma, J., Panyala, A., & He, D. (2010). Hybrid ceramic bearing prognostics using particle filtering. *Proceedings of the 2010 Conference of the Society for Machinery Failure Prevention Technology*, Huntsville, AL, April 13 – 15, pp. 57 – 69.

Lim, P., Goh, C. K., Tan, K. C., & Dutta, P. (2014). Estimation of remaining useful life based on switching kalman filter neural network ensemble. *Proceedings of the 2014 Annual Conference of the Prognostics and Health Management Society*, Fort Worth, TX, pp. 2–9.

Lv, Y., Duan, Y., Kang, W. Li, Z., Fei-Yue Wang, F.-Y. (2015). Traffic flow prediction with big data: A deep learning approach. *IEEE Transactions on Intelligent Transportation Systems*, vol.16, no.2, pp.865-873.

Malhi, A., Yan, R., & Gao, R. X. (2011). Prognosis of defect propagation based on recurrent neural networks. *IEEE Transactions on Instrument and Measurement*, vol. 60, no. 3, pp. 703–711.

Oliveira, T. P., Barbar, J. S., Soares, A. S. (2014). Multilayer perceptron and stacked autoencoder for Internet traffic prediction. *Network and Parallel Computing*, vol. 8707 of the series Lecture Notes in Computer Science, pp. 61-71.

Shao, H., Jiang, H., Zhang, X., & Niu, M. (2015). Rolling bearing fault diagnosis using an optimization deep belief network. *Measurement Science and Technology,* Volume 26, Number 11.

Tao, Y., Chen, H., & Qiu, C. (2014). Wind power prediction and pattern feature based on deep learning method. *Power and Energy Engineering Conference (APPEEC), 2014 IEEE PES Asia-Pacific* , vol., no., pp.1-4.

Vachtsevanos, G., Lewis, F. L., Roemer, M., Hess, A., & Wu, B. (2006). *Intelligent fault diagnosis and prognosis for engineering system*. Hoboken, NJ: John Wiley & Sons, Inc.