Rotating Machinery Prognostics and Application of Machine Learning Algorithms: Use of Deep Learning with Similarity Index Measure for Health Status Prediction

Asheber Wagshum Techane[1], Yu-Fu Wang[2], and Bereket Haile Weldegiorgis[3]

[1,2]*Precision Machinery Research and Development Center, Taichung, 40850, Taiwan*
e10509@mail.pmc.org.tw
e10201@mail.pmc.org.tw

[3]*National Taiwan University of Science and Technology, Taipei, 10607, Taiwan*
D10103814@mail.ntust.edu.tw

ABSTRACT

The internet of things (IOT) enabled presence of abundant sensors on smart machineries and the recent advance in deep learning is accelerating the development of predictive maintenance in production systems with less time and fair amount of effort. In this work a Deep learning Neural Networks (DNN) based bearing health monitoring system with index of similarity check is developed and tested for its effectiveness. The assessment procedure followed in here trains a DNN model on a time series data segmented to a vector size equal to number of data points per cycle as training and test data sets. Moreover the model measures the similarity of the test signal to an anchor signal selected from each fault class. The classification performance comparison done proves that DNN with fair depth and large data perform better and can be extended to other problems in intelligent maintenance systems development efforts. The proposed system is intuitive and has minimal complexity with uncompromised fault detection accuracy.

1. INTRODUCTION

The advance in automation and robotics has envisioned the realization of the 4th industrial revolution aka industry 4.0. Moreover the tremendous progress seen in machine learning (ML) in general and deep learning (DL) in particular has doubled the effort and enhanced the path to realization of predictive maintenance goals. High volume of research works are being done on machinery prognostic and health management (PHM) which taps the availability of big data enabled analytics due to abundance and sophisticated sensors incorporated in the machineries as part of growing machine hardware intelligence (E. L. Lee, 2016; J. Lee et al., 2014). Uses of big data analytics to take proactive action with regard to machine performance, however is dependent on the algorithms chosen and the quality of the feature extracted out of the data (Tsui, Chen, Zhou, Hai, & Wang, 2015).

Most fault identification methodologies in rotating machineries, such as induction motor and gearboxes, uses vibration data. The commonly used analysis techniques are based on statistical time domain, frequency domain, and the joint time-frequency domain features (Saucedo-Dorantes, Delgado-Prieto, Ortega-Redondo, Osornio-Rios, & Romero-Troncoso, 2016; Tran, Yang, Gu, & Ball, 2013). One limitation of frequency domain analysis is its inability to handle non-stationary waveform signals, commonly observed during machine faults (Zhang, 2017). Time-frequency domain techniques, such as short time Fourier transform (STFT) and wavelet methods have been applied to rotating machineries health prognostics to overcome this drawback (Q. Sun, Chen, Zhang, & Xi, 2004).

While dozens of methods have been developed to analyze bearing conditions, most of these studies fail to quantify the damage level, mainly focus on classification task. Moreover there is vivid lack of intuitiveness and needs domain expertise to understand and apply to real problem solutions. This research work uses raw time series data in training deep learning neural network (DNN) to monitor health degradation of roller bearing in rotating machineries. To enhance the detection and diagnostic of bearing condition the dominant health indicators of its components through deep layer learning algorithm is extracted. The two basic procedures followed in this work in estimating the bearing components health status are:

i. Segment the raw time series data to one cycle vector in accordance to setup parameters during the collection phase. Use this vectors and form training dataset to build a DNN model.

ii. Train Support Vector Machine (SVM) model to classify the bearing based on manually extracted features.

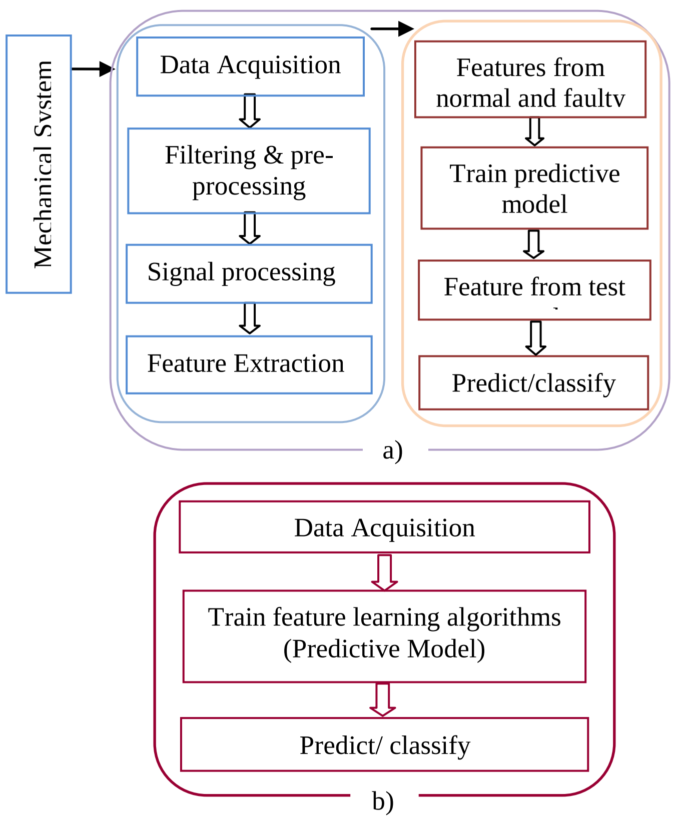

Figure 1. Steps in bearing health assessment: a) Traditional method, b) DNN Method

The traditional approach in bearing performance prediction involves collecting raw vibration data, feature extraction and system modeling (Figure 1a).These models base on hand crafted features, which at times my result in loss of some important information. This work aims at system modeling and health performance prediction for rotating machinery components using cyclic raw vibration data. It mainly focuses on application of deep learning algorithm to bearing health assessment in feature extraction and classification, and present analysis result on performance of the proposed model. In this paper raw time series data collected with enhanced noise filtering is used to train deep neural network (Figure 1b).

The reset of the paper is organized as follows: In section 2 we will discuss data acquisition procedures; while bearing vibration signal feature extraction is explained section 3; in section 4 we describe machine learning training algorithm used. Section 5 discusses experiment and result, and finally conclusion is presented in section 6.

2. DATA ACQUISITION

It is challenging to collect good quality vibration data from rotating machinery in production line due to complex nature of the signal as well as noise and resonance effects of other components. Nonetheless vibration signals obtained from the vicinity of a bearing assembly contain rich information about the bearing condition(Q. Sun et al., 2004).

AC servo motor actuated test rig equipped with single axis accelerometer, (PCB 352C33 Quartz ICPs in combination with NI-9234 signal acquisition module) is prepared as a test set up. The accelerometer is placed on the outer surface of the bearing housing (Figure 2) to measure the vibration signals. Data is collected at sampling rate of 12 KHz with a block size of 40960 data points captured; 64 files of data (blocks of data vectors of size 40960) are taken for each class bearing; normal, inner race, outer race and ball defect bearings (Figure 3).

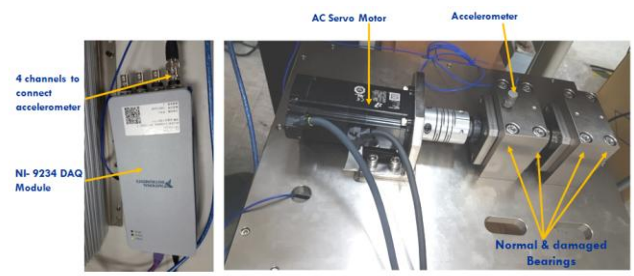

Figure 2. NI-9234 DAQ module and Bearing Test rig

In case of feature based health monitoring system the collected data needs to go through preprocessing steps to extract the selected features. In deep learning neural network, however, the raw time series data, segmented into proper size (number of data points per revolution of the shaft) is directly used in training DNN.

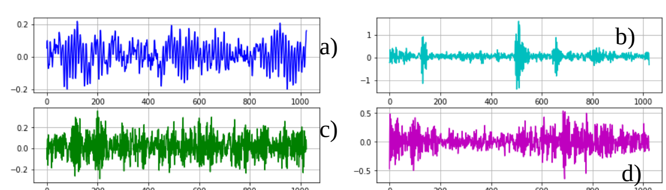

Figure 3. Time series vibration signal: a) normal b) inner c) outer race and d) roller defect bearings.

In fact the biggest concern in deep learning is the availability of sufficient labeled data for training; nonetheless this is less a challenge in the current problem since one can collect enough data within few weeks or months easily using relevant sensors.

3. FEATURE EXTRACTION

The goal of feature extraction is to find out as many as possible unique features to enable accurate recognition of bearing performance status at real time. To develop bearing health monitoring model without a need for manual feature extraction, deep neural network (DNN) is one of the best and versatile candidate, which uses its layers' depth in sifting the data and effectively identifies the characteristics features. However this model depends on availability of big data. In case of poor knowledge about the data, in fact unsupervised machine learning algorithms such as self organizing mapping (SOM) will have a better performance in extracting features inherent to the data set (Lu, Sun, Tao, Liu, & Lu, 2013).

To extract features in the frequency domain as well as the envelope method the peak information at the bearing characteristic frequencies are key indicators for assessing the health of the rolling element bearing. These frequencies are estimated from bearing geometric parameters and mounted shaft speed. The mathematical formula for the characteristic ball pass frequencies (BPF) is reported in numerous literatures (Soualhi, Medjaher, & Zerhouni, 2015). These values for the experimental test rig in this study are reported in Table 1.

Table 1 Bearing Parameters

Inner race rotational speed(rpm) ω_i	3000
Outer race rotational speed(rpm) ω_0	0
Bearing pitch diameter(mm)	31
Roller/ball diameter(mm)	6.35
Contact angle θ	15^0
Number of rolling element	12
@ 3000 RPM constant shaft speed	(RPS)
Outer race ball pass frequency (BPFO)	240.642
Inner race ball pass frequency(BPFI)	287.486
Ball Spin Frequency (BSF)	117.269
Fundamental Train Frequency (FTF)	20.053

Some of the common statistical features extracted from the time domain signal for use in SVM model, procedure ii, training includes; maximum (X_{max}), mean (μ), root mean square (RMS), variance, standard deviation (σ), skewness kurtosis, peak-to-peak and crest factor (Bornn, Farrar, Park, & Farinholt, 2009). In addition to the above time series features BPFO, BPFI, BSF and FTF values are considered in training feature vector of SVM model.

The characteristics of bearing dynamic response due to defect at particular component (roller, cage, inner and outer race) when it passes through the load zone of the bearing, shows there is a correlation between the defect and impulse patterns observed in one cycle of the signal (Randall & Antoni, 2011; You & Meng, 2011). This characteristic is important in selecting the data segment size. The segmented data is centralized, i.e. raw time data is zero mean normalized for use in the DNN algorithms, which will help in removing the DC data and faster convergence in learning phase.

4. MODEL TRAINING

While data may have been acquired with reasonable degree of accuracy and in abundance, the data needs to be analyzed and understood to the extent it gives proper information about the machine performance history. Thus this study proposes to use supervised machine learning algorithms, Deep Neural network. In developing the DNN model, segmented and normalized raw time series data is used to avoid any extra preprocessing and make the task more intuitive. For the purpose of comparison SVM model is also developed on manually extracted features, and the result of the two algorithms is reported for the data used.

The data is partitioned randomly to 90 % training set (2,359,296 data points), 5 % validation set (131,072 data points), and 5% test set (131,072 data points) for all class of bearings. It is divided into eleven training batches and one test batch, each training batch with 214400 data points. The test batch contains only 50 randomly-selected signals segments (vector of size 400) from each class. The training batches contain the signal vectors in random order, but some training batches may contain more from one class than another for CWRU data. In training phase the four class of bearing category is encoded as a (4xm) vector Y with $y[i]=1$ for class label i and zero otherwise, m is number of training examples. For instance $y=[1\,0\,0\,0]^T$ represent normal bearing.

4.1. Deep Learning Neural Networks(DNN)

Given training example (x, y), it uses deep layers of neuron to learn the underlying features to predict the bearing conditions. DNN is mostly applied in image recognition and classification problems (Bengio, 2009; Schmidhuber, 2015; Szegedy, Toshev, & Erhan, 2013). For X is input, W and b weight and bias network parameters respectively is updated during the training phase of the model to predict the output Y, as in typical DNN model in Figure 4.

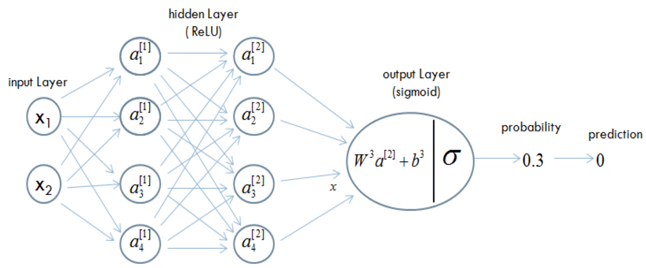

Figure 4. Multi-layer Artificial Neural Network model

For non-stationary non-linear nature of vibration signal neural network is one of the ideal candidate algorithms of choice from pattern recognition and condition monitoring tool (Hinton, Osindero, & Teh, 2006; Jia, Lei, Lin, Zhou, & Lu, 2016; Jing, Zhao, Li, & Xu, 2017). The two major steps in DNN learning are forward and back ward propagation. The forward and backward pass will let to run until convergence is achieved to the level of error criteria. The general steps in building a Neural Network model is (Figure 4):

1. Define the network structure (i.e. number of input, hidden, and output units). Initialize the weight and bias parameters, $W^{[l]}$ and $b^{[l]}$ respectively.
2. Loop:
   - Implement forward propagation, Eq. (1).

$Linear \to ReLu \to Linear \to ReLu \to Linear \to Sigmoid / SoftMax$

$$\begin{aligned} Z_1 &= W^{[1]}X + b_1 \to A_1 = relu(Z_1) \\ Z_2 &= W^{[2]}A_1 + b_2 \to A_2 = relu(Z_2) \\ Z_3 &= W^{[3]}A_2 + b_3 \to \tilde{y} = \sigma(Z_3) \end{aligned} \quad (1)$$

   - Compute cost.
   - Implement backward propagation to get the gradients.
   - Update parameters (gradient descents).

The problem in this work is approached as multi-class classification problem; thus DNN output activation layer uses **Softmax** activation, logistic regression for ***C*** classes. For X , Y training and label input data sets the model learning is implemented as forward and backwards propagation steps followed by gradient decent optimization and parameter updating with the objective of minimizing the loss function, Eq. (2), using cross entropy.

$$J = -\sum_{i=0}^{m} (y^{(i)} \log(\tilde{y}^{(i)} + (1 - y^{(i)}) \log(1 - \tilde{y}^{(i)})) \quad (2)$$

Where $\tilde{y}^{(i)}$ and $y^{(i)}$ are predicted and actual outputs of the i^{th} training steps respectively.

Developing an optimized DNN model is not an easy task; usually a problem of high variance and bias is faced, slow convergence while determining the optimal network and the whole process is highly iterative. To mitigate the problem of over and under fitting various methods have been suggested by Machine Learning researchers' community, such as regularization, data augmentation, early stopping, and dropout techniques.

Due to the nature of physical interactions between bearing components, vibration signal may not exactly identical in a given cycle. To accommodate this variation we adopt similarity check function Eq.(3) whose parameter propagate to compute resemblance of the signal during training, which originally used in (Y. Sun, Wang, & Tang, 2014) face representation and identification task with slight change on constant parameter **d**. This can essentially be considered as a correlation of two discrete signals with trained scaling and shifting parameters.

$$simIndex(x_i, x_j, y_{ij}) = \frac{1}{2}(y_{ij} - \sigma(\theta d + \nu))^2, \quad (3)$$

Where x_i, x_j are segmented signal in the training data set $y_{ij} = 1$ if the signal are same and $y_{ij} = -1$ otherwise. With $d = \frac{x_i \bullet x_j}{\|x_i\| * \|x_j\|}$ is cosine similarity between two signals, $\tau\{\theta, \nu\}$ are learnable parameters.

Thus the DNN model parameters are trained based on optimizing cross-entropy loss function Eq. (2). Similarity check, Eq. (3) is computed during training and cache gradients in every pass to update these parameters after each epoch, Eq. (4) and Eq. (5). Similarity check index tell how far apart are the two signals under consideration. The anchor signal to compare with during testing is selected from a signal in training dataset based on similarity index value; signal with higher similarity within the same class is selected as anchor of the group.

$$\nabla \tau = \lambda \frac{\partial simIndx(x_i, x_j, y_{ij})}{\partial \tau} \quad (4)$$

$$\begin{aligned} W^{[l]} &= W^{[l]} - \alpha * dW^{[l]} \\ b^{[l]} &= b^{[l]} - \alpha * db^{[l]} \\ \tau &= \tau - \alpha * d\tau \end{aligned} \quad (5)$$

Where $W^{[l]}$ and $b^{[l]}$ are layer L weight and biases respectively.

Once training is completed; in testing phase the test data is segmented into one cycle size points, same as training signal size, and run prediction followed by computation of similarity index value of the test case with the anchor signal of each class. Then the final prediction result is verified based on the average of weighted sum of each prediction (Algorithm1):

Algorithm1: Similarity check algorithm

Given: $X\_anch = \{X\_anch\}_0^m$ anchor signal from each class.

Segment test signal into **N** vector of size same as training signal.

$X\_test = \{x_1, x_2, ..., x_N\}$

Run **DNN** model and predict the class of each segment of the test signal.

$y\_test = \{y_1, y_2, ..., y_N\}$

for ***i*** in range**(0,C)**: #C-number of class

for **j** in range**(0,N):**

$sIndx[i, j] = simIndex(x_i, x_j, y_{ij})$

Compute current time prediction as weighted sum of prediction each segment of the test signal as final verification step.

$$Y = \frac{1}{N}\sum_{n=1}^{N} sInd[:, n] * y_n$$

In case of monitoring a bearing health in its life time the index expected to get bigger showing more similarity to the other classes (fault bearings).

5. EXPERIMENT AND ANALYSIS

A Bearing health diagnosis system based on deep learning neural network system is trained. The system is applied to classify a bearing into four categories based on a vibration signal collected and segmented to one cycle data points. Effectiveness of the system developed is also tested on data from test rig we have developed as well as a dataset acquired from online repository bearing vibration database(CWRU) (Western, 2013). The performance of DNN on standardized data set is shown in Table-3.

Table 2. Classification accuracy (%)

Class Label	1 hidden layer	2hidden layer	3 hidden layer	SVM
Normal	86.00	92.00	100.0	87.19
Inner race	64.00	88.33	91.17	77.78
Outer race	52.00	88.33	94.17	79.12
Roller fault	78.00	93.50	99.33	77.32

The accuracy reported in Table 2 is average result on CWRU database. The overall accuracy for best performing model on test dataset in 4-layer network goes up to 98.5%. Figure 5 shows training performance for two layers DNN.

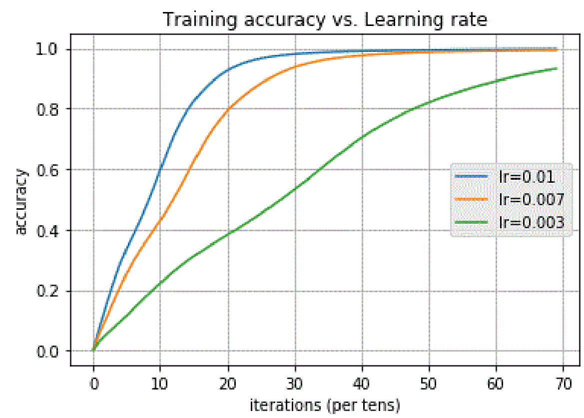

Figure 5. Two Layer model training accuracy

The main worry in using the deep learning approach is the time and computational cost while using batch gradient decent optimization algorithms, Eq. (5). Stochastic gradient descent is fast enough to train in this case; however the parameters oscillate in approaching to the minimum los point, which may leads it to trap in local minima.

The training performance for varying learning rate and optimization algorithms in Figure 6 reveals that adaptive learning methods convergence faster and has better test accuracy. However several hyper parameters need tuning. On the other hand figure 6c shows the batch size relation to the convergence time and accuracy, smaller mini-batch learn faster and have wavy gradient to absolute minima.

The analysis result shows that detection of ball defect and normal bearing is high in all the test cases, mostly near 99 % prediction accuracy (Figure 7). During testing we have also observed that there was high confusion of inner race fault as ball and outer race defect; however none of the faults confused to normal, which is a good characteristic in classifying bearing into normal and failed binary classes. The DNN system developed in this work is implemented using TensorFlow v1.4.

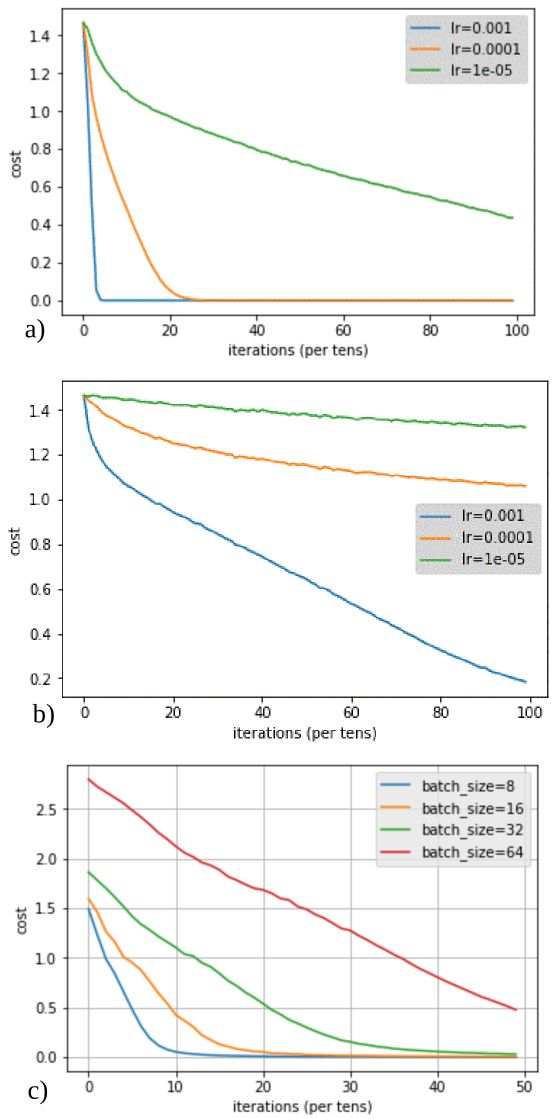

Figure 6. Cost (error) vs. learning methods: a) Adaptive learning b) Gradient descent c) batch size effect on convergence.

One remarkable observation of the analysis result is that detection accuracy varies for loading variation and also for position of outer race fault. The confusion matrix for 200 test samples for 0.014" fault size and outer race fault @6:00 is shown in Table-3.

Table 3. Confusion Matrix (CWRU data)

	Predict			
Actual	Normal	Inner	Outer	Ball
Normal	50	0	0	0
Inner	0	48	0	0
Outer	0	0	50	1
Ball	0	2	0	49

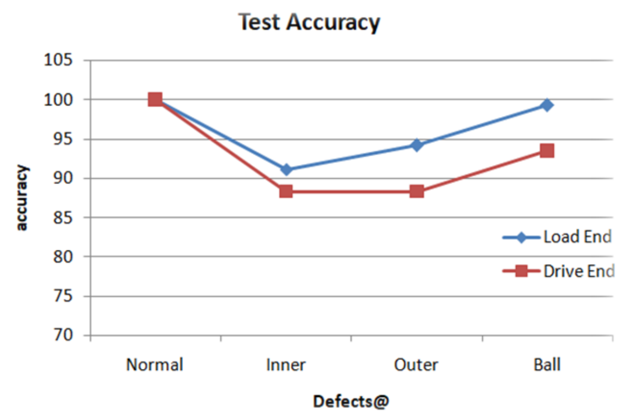

Figure 7 Average Test Accuracy curve

The performance on a data collected from our test setup platform have similar (with up to 8% difference in some cases) characteristics to CWRU data. Nonetheless the test from our data set shows higher confusion rate for some defect, e.g. ball to inner race confusion is higher than CWRU data, up to 5 more cases confused on 50 test cases. The reason for difference in prediction accuracy is partly due to the test setup (closely placed bearing in our test setup) but this needs to be verified in series of experiments.

6. CONCLUSION

In this paper deep learning neural network algorithms based system is developed to model fault detection in roller bearing based on historical data collected from a test rig operated in different conditions moreover the similarity of test signals is measured to check how close predicted class to normal anchor signal. The analysis result has proven the effectiveness of using DNN algorithms in extracting features from raw vibration data and predicting bearing status from cyclic raw vibration time domain signal. From practical point of view the fact the network learns to compute the similarity index in relation to normal bearing will enhance the confidence of prediction as well as have the potential to quantify the relative damage level, at least relative damage. In general the test results show that deep learning from raw time series data segmented to one cycle have robust performance. With good choice of the hyper parameters DNN outperform other algorithms on training and testing datasets.

ACKNOWLEDGEMENT

This project is funded by Taiwan government under grant number 106EC17A251448 of industrial research support.

REFERENCES

Bengio, Y. (2009). Learning Deep Architectures for AI. *Foundations and Trends® in Machine Learning, 2*(1), 1-127. doi: 10.1561/2200000006

Bornn, L., Farrar, C. , Park, G., & Farinholt, K. (2009). Structural Health Monitoring With Autoregressive Support Vector Machines. *Journal of Vibration and Acoustics, 131*(2), 021004-021004-021009. doi: 10.1115/1.3025827

Hinton, G., Osindero, S., & Teh, Yee-Whye. (2006). A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation, 18*(7), 1527-1554. doi: 10.1162/neco.2006.18.7.1527

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, 72-73*(Supplement C),303-315.doi: https://doi.org/10.1016/j.ymssp.2015.10.025

Jing, L., Zhao, M., Li, P., & Xu, X.. (2017). A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. *Measurement, 111*(Supplement C), 1-10. doi: https://doi.org/10.1016/j.measurement.2017.07.017

Lee, E., (2016). *Predictive Manufacturing: The Next Transformation*.

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. doi: https://doi.org/10.1016/j.ymssp.2013.06.004

Lu, C., Sun, Q., Tao, L., Liu, H., & Lu, C., (2013). Bearing Health Assessment Based on Chaotic Characteristics. *Shock and Vibration, 20*(3). doi: 10.3233/sav-130765

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

Saucedo-Dorantes, J., Delgado-Prieto, M., Ortega-Redondo, J., Osornio-Rios, R., & Romero-Troncoso, R., (2016). Multiple-Fault Detection Methodology Based on Vibration and Current Analysis Applied to Bearings in Induction Motors and Gearboxes on the Kinematic Chain. *Shock and Vibration, 2016*, 13. doi: 10.1155/2016/5467643

Schmidhuber, J., (2015). Deep learning in neural networks: An overview. *Neural Networks, 61*(Supplement C), 85-117. doi: https://doi.org/10.1016/j.neunet.2014.09.003

Soualhi, A., Medjaher, K., & Zerhouni, N., (2015). Bearing Health Monitoring Based on Hilbert-Huang Transform, Support Vector Machine, and Regression. *IEEE Transactions on Instrumentation and Measurement, 64*(1), 52-62. doi: 10.1109/TIM.2014.2330494

Sun, Q., Chen, P., Zhang, D., & Xi, F., (2004). Pattern Recognition for Automatic Machinery Fault Diagnosis. *Journal of Vibration and Acoustics, 126*(2), 307-316. doi: 10.1115/1.1687391

Sun, Yi, Wang, X., & Tang, Xu,. (2014). Deep Learning Face Representation by Joint Identification-Verification. *CoRR, abs/1406.4773*.

Szegedy, C., Toshev, A., & Erhan, D., (2013). *Deep neural networks for object detection*. Paper presented at the Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2, Lake Tahoe, Nevada.

Tran, V., Yang, Bo-Suk, Gu, F., & Ball, A., (2013). Thermal image enhancement using bi-dimensional empirical mode decomposition in combination with relevance vector machine for rotating machinery fault diagnosis. *Mechanical Systems and Signal Processing, 38*(2), 601-614. doi: https://doi.org/10.1016/j.ymssp.2013.02.001

Tsui, K., Chen, N., Zhou, Q., Hai, Y., & Wang, W., (2015). Prognostics and Health Management: A Review on Data Driven Approaches. *Mathematical Problems in Engineering, 2015*, 17. doi: 10.1155/2015/793161

Western, Reserve University Case. (2013). Bearing Vibration Data Set *http://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website*.

You, M. Y., & Meng, G. (2011). A generalized similarity measure for similarity-based residual life prediction. *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering, 225*(3), 151-160. doi: 10.1177/0954408911399832

Zhang, R., Deng, W.; Zhu, M., (2017). Using Deep Neural Networks to Automate Large Scale Statistical Analysis for Big Data Applications. *Cornell University Library*.