

Bearing Fault Diagnosis Using Deep Sparse Autoencoder

S R Saufi^{1*}, Z A B Ahmad¹, M S Leong¹ and L M Hee¹

¹Institute of Noise and Vibration, Faculty of Mechanical Engineering, Universiti Teknologi Malaysia (UTM), Jalan Sultan Yahya Petra 54100 Kuala Lumpur, Malaysia

*Corresponding email: msramadhan93@yahoo.com

Abstract. Rolling element bearing is an important component in various machinery. Faulty on bearing cause severe equipment damage that lead to high maintenance cost. The development of deep learning has been paid a considerable amount of attention to fault diagnosis on rolling element bearing. Traditional machine learning such as Artificial Neural Network and Support Vector Machine have problems of lacking expression capacity, existing the curse of dimensionality, require manual feature extraction and require an additional feature selection. Deep learning model has the ability to effectively mine the high dimensional features and accurately recognize the health condition. In consequence, deep learning model has turned into an innovative and promising research in bearing fault diagnosis field. Thus, this paper tends to proposed Deep Sparse Autoencoder (DSAE) with Teager Kaiser Energy Operator (TKEO) to diagnose the bearing condition. DSAE is one of deep learning model which uses the architecture of neural network. During the analysis, the hyperparameter of DSAE model was optimized by Ant Lion Optimization. The analysis results show that the proposed TKEO-DSAE achieved 99.5% accuracy of the fault diagnosis. The comparative study between proposed model and ANN proved that deep learning model outperform traditional machine learning model on bearing fault diagnosis.

Keywords: bearing; fault diagnosis; teager kaiser energy operator; deep sparse autoencoder; ant lion optimization

1. Introduction

Rolling element bearing (REB) is broadly used in domestic and industrial applications. The bearing provides a physical support to a component and allows that components to rotate with less friction. Working condition of this component depends on the smooth and quiet running of the bearings. Bearing is extensively used in industry field such as induction machines [1] wind turbines [2], helicopters [3], and automotive [4]. Bearing is considered as a critical component in a majority of machines due to exposure on extreme conditions (dirty conditions, high temperature, overstress) continuously. Extreme conditions cause the generation of fault on bearing components such as outer race, inner race and rolling element. Therefore, Fault diagnosis is introduced in providing an effective maintenance strategy to diagnose the fault growth in bearing component. Various monitoring method have been introduced for bearing fault diagnosis such as vibration analysis, acoustic analysis, thermal analysis and motor current analysis. However, vibration analysis has been proven to be an effective monitoring method for bearing fault diagnosis [5- 8].

Presently, the integration of machine learning on bearing fault diagnosis has been paid a lot of attention by researchers as machine learning could produce more accurate outcomes with a consistent result. Artificial neural network (ANN) and Support Vector Machine (SVM) are among the popular



machine learning model which widely used on bearing fault diagnosis. However, this machine learning needs a manual feature extraction from the bearing vibration signal. The extracted features require a selection process in order to choose the best feature that reflect the bearing conditions. The process of feature extraction and feature selection is time-intensive. In 2006, Hinton and Salakhutdinov develop a modification version of the neural network by increasing the number of hidden layer [9] which later called as deep learning. Deep learning model capable of extracting high dimensional feature without the needs of a human assistant. Convolutional Neural Network (CNN), Deep Belief Network (DBN), and Deep Sparse Autoencoder (DSAE) are the examples of deep learning model which is widely used in many applications [10- 12]. Among the deep learning model, DSAE model has the capability to be trained based on unsupervised manner [13]. The DSAE model is the integration of multiple neural network architecture. The DSAE aims to produce the output that contains similar characteristics to its input. DSAE is trained with the same mathematical model as feedforward neural network that utilized gradient descent algorithms for the back-propagation process.

Due to the complexity of the bearing signal, a lot of consideration is required to acquire the important feature from the bearing signal. This is due to the prediction performance of deep learning model is influenced by the quality of the signal. It is known that during data acquisition from the sensor, the captures signal contains the signal of interest and noise. Noise is a type of disturbance that bury the useful important feature in the signal. Therefore, a suitable signal processing method is required to filter the noise so that useful feature can be extracted from the signal. There is a lot of signal processing method utilized in bearing fault diagnosis field such as wavelet transform, empirical mode decomposition, variational mode decomposition, etc. In this study, the Teager Kaiser Energy Operator (TKEO) is used instead of another signal processing method due to its ability to pre-process the signal by filtering the low-frequency background signals [14]. TKEO has been paid less attention toward bearing fault diagnosis.

This paper aims to solve bearing fault diagnosis using Teager Kaiser Energy Operator (TKEO) and deep sparse autoencoder (DSAE). The DSAE model is optimized via Ant Lion Optimization. The performance of (DSAE) model is compared with Artificial Neural Network (ANN).

2. Teager Kaiser Energy Operator Theory

The Teager–Kaiser energy operator was introduced by Teager and Kaiser to measure the energy of a time domain signal that was generated by the mechanical process [15]. This method is proposed for AM-FM demodulation and to estimate the envelope of the amplitude of AM signals [16]. The TKEO process on continuous time signals $x(t)$ is defined in the following Equation (1);

$$x(t) = \left(\frac{dx}{dt}\right)^2 - x(t)\left(\frac{d^2x}{dt^2}\right) \quad (1)$$

Meanwhile, the Equation (2) defined the discrete model of TKEO;

$$x(n) = x^2n - x(n-1)x(n+1) \quad (2)$$

Several studies used TKEO for signal processing. For example, Rodriguez extracted important feature from the Amplitude Modulated (AM) signal using TKEO. The result shows that features extracted after the signal has been processed by TKEO outperform the diagnosis results of the extracted features from the raw time vibration signal [17]. Meanwhile, Kwak et al. used TKEO in their research to enhance the signal peak induced by the defect located on bearing component [14]. Tran et al. used to TKEO to reveal the fault pattern from the reciprocating compressor valves signals [18].

3. Deep Sparse Autoencoder

The autoencoders aim to produce the output that contains similar characteristics to its input. Sparse autoencoder used the architecture of neural network which is suitable for the dimensional reduction of the features. The hidden layer representation of sparse autoencoder network is illustrated in the following Equation (3);

$$h(x) = f(w_1x_1 + b_1) \quad (3)$$

where $f(z)$ is the nonlinear activation function to maps the input. Generally, the logistic sigmoid function $f(z) = 1/1 + e^{-z}$ is used as activation function which maps the input from zero to one. The network output maps the hidden representation h back to a reconstruction $\tilde{x} \in R^n$ is illustrated in Equation (4);

$$\tilde{x} = f(w_2h(x) + b_2) \quad (4)$$

In general, autoencoder is used for feature dimensional reduction and do not contain any sparsity term on its model. The hidden node number of autoencoder should be less than the input size. By adding the sparsity term on the autoencoder cost function, the network of autoencoder can be more versatile since the hidden node number can be any number either lower, equal or higher than input node. Therefore, the cost function of sparse autoencoder is developed based on three terms which are defined in Equation (5);

$$E = \frac{1}{N} \sum_{n=1}^N \sum_{k=1}^K (x_{kn} - \hat{x}_{kn})^2 + \lambda * \Omega_{weights} + \beta + \Omega_{sparsity} \quad (5)$$

The integration of multiple sparse autoencoders is called deep sparse autoencoder (DSAE).

4. Ant Lion Optimization

The Ant Lion Optimization (ALO) is the optimization algorithms in finding the unknown parameters based on the objective function and fitness values. The ALO algorithms are proposed by mimicking the nature of Antlion hunting mechanism. This algorithm was proposed by Mirjalili in 2015 [19]. The antlion hunting mechanism is based on four following process; digging the trap, hide underneath the bottom cone and attack the prey once the prey falls into the trap. Finally, the antlion pulled the prey under the soil and consumed. The leftovers are thrown outside the trap and the antlions prepare the trap for the next hunt. The details regarding the algorithms may refer to Mirjalili works [19].

The ALO algorithms were used to optimized six hyperparameters of DSAE such as weight regularization, sparsity regularization and sparsity proportion that are associated with each sparse autoencoder. In this study, two components of sparse autoencoder are used to construct the DSAE network. Manual selection of six DSAE hyperparameter is time-intensive. The overall step of bearing fault diagnosis is shown in the following Figure 1.

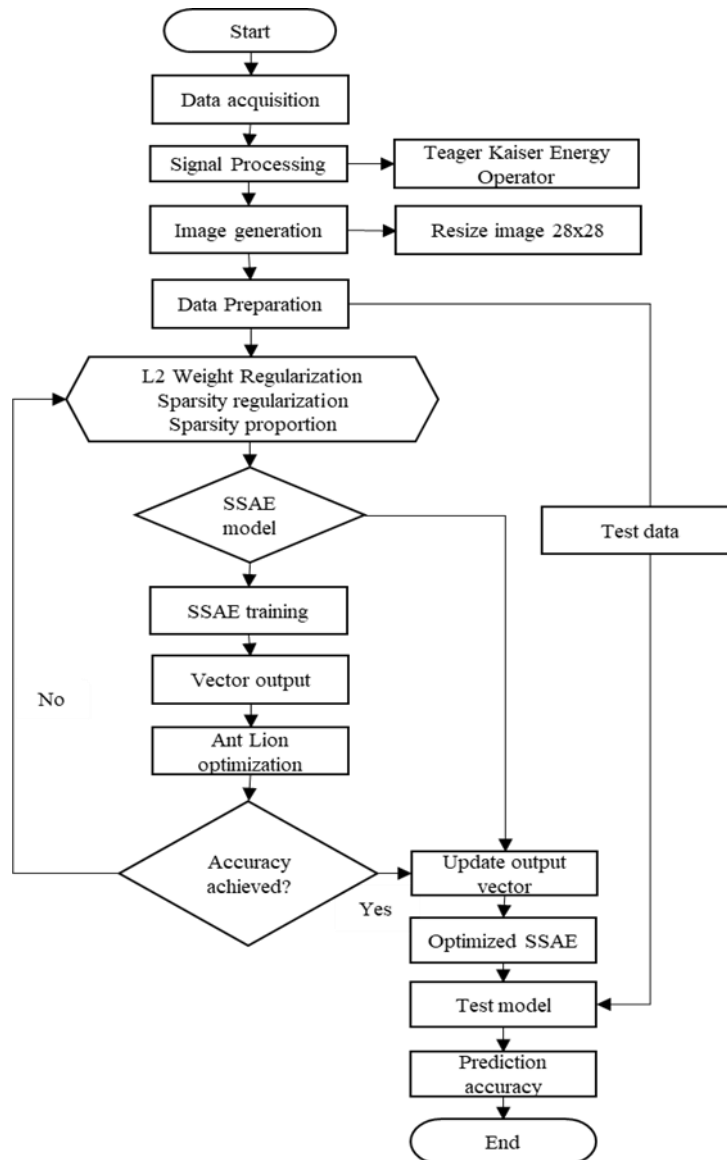


Figure 1. Bearing fault diagnosis of Proposed Method

5. Experimental setup

In this section, the details about experimentation from Case Western Reserve University is discussed. The data were collected with 48kHz from the experimental set up as shown in Figure 2. Every single fault is artificially made on each bearing components such as an outer raceway, inner raceway and rolling element. In this study, the fault diagnosis focus on the 1800rpm speed conditions and 0.007-inch fault size. Based on the data selection, we believed that if the proposed model could diagnose the 0.007-inch fault effectively, then there is no a problem for the proposed model to diagnose 0.014-inch and 0.021-inch fault size. The selected vibration signal is then processed by Teager Keiser Energy Operator (TKEO). The 2D-images of TKEO signal is resized to 28x28 image patches. In addition, the statistical feature from the filtered signal by TKEO is extracted for further analysis.

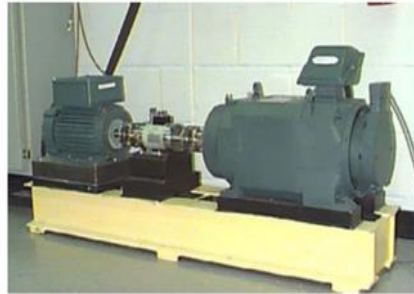


Figure 2. Experimental Rig for bearing fault simulation

There are four types of signal that represent bearing conditions as illustrated in table 1 which are normal, outer race fault, rolling element fault and inner race fault. The vibration signals are segmented to one rotation based on equation 6 and the segmented process is illustrated in Figure 3. The segmented signals are sampled as shown in Table 1 which means the signal is cut into 100 samples on each bearing conditions. The 100 samples are divided into a training dataset and testing dataset. Each segmented vibration signal is processed via TKEO as shown in Figure 4-7. These images from Figure 4-7 are used as input to our proposed method. The proposed model is based on image classification. These image patches are used as an input to DSAE model and ANN model for bearing fault classification analysis. Instead of TKEO images, the statistical features from the filtered signal has been extracted for a comparative study between the proposed method and ANN model with statistical features. There are nine statistical features that were extracted from the filtered signal such as amplitude, mean, variances, standard deviation, RMS, kurtosis, skewness, crest factor and clearance factor.

$$\text{Segmented Signal for one rotation} = \frac{\text{number of sample data in one second}}{\text{speed of motor (RPM)}/60} \quad (6)$$

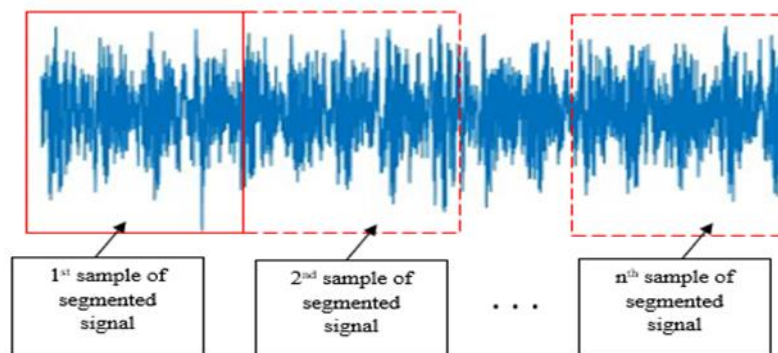


Figure 3. The segmented signal of one rotation

Table 1. Data distribution

Bearing Conditions	Training Data	Testing Data
Normal	50	50
Outer race fault	50	50
Rolling element fault	50	50
Inner race fault	50	50

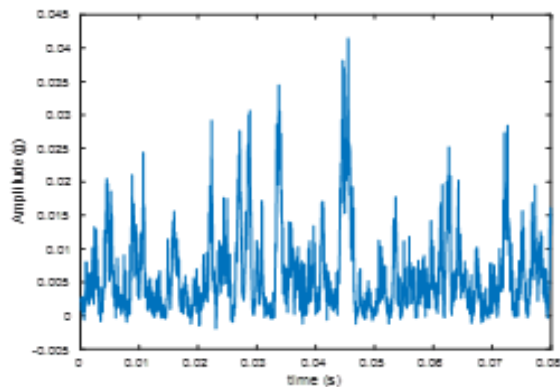


Figure 4. Rolling Element Fault

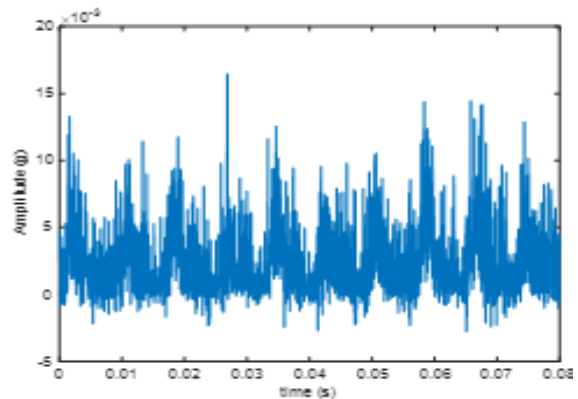


Figure 5. Normal state

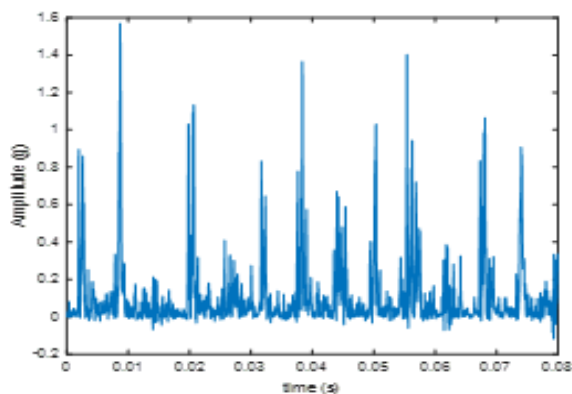


Figure 6. Inner Race Fault

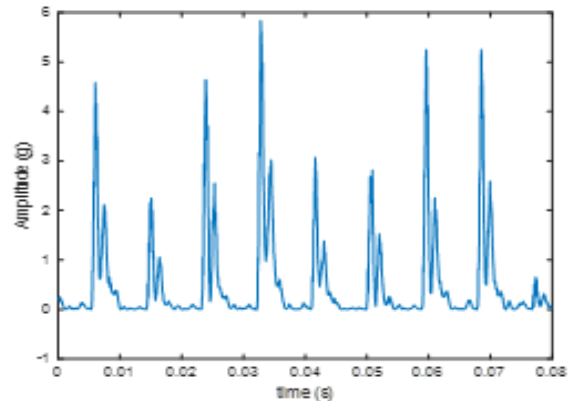


Figure 7. Outer Race Fault

The performance accuracy of DSAE training and supervised fine-tuning stages is affected by many hyperparameters. The hyperparameters of DSAE classifier are the number of the hidden nodes, number of layers, maximum iteration (epoch), sparsity proportion etc. In general, there is no standard method to determine the optimal hyperparameters of DSAE model. As mentioned earlier, three important hyperparameters of DSAE is optimized by ALO algorithms. The range of three hyperparameters that were optimized using ALO is tabulated in Table 2. In addition, the rest hyperparameter of DSAE model is tabulated in Table 2.

Table 2. Setting of parameters for DSAE

Sparse Autoencoder 1	Hidden node number: 300
	Epoch: 300
	L2 Weight Regularization: (0-0.004)
	Sparsity Regularization: (0-4)
	Sparsity Proportion: (0-1)
Encoder and Decoder Transfer function: Logistic Sigmoid	
Sparse Autoencoder 2	Hidden node number: 150
	Epoch: 300
	L2 Weight Regularization: (0-0.004)
	Sparsity Regularization: (0-4)
	Sparsity Proportion: (0-1)
Encoder and Decoder Transfer function: Logistic Sigmoid	

Softmax Layer	Loss function: Cross-entropy Epoch: 300
---------------	--

The comparative study between DSAE and ANN model is conducted. ANN is a traditional machine learning model which has been broadly used in bearing fault diagnosis. The ANN model is trained using two different datasets such as statistical features and image features. The prediction of ANN for both features are presented in the next section. The parameter setting regarding ANN model is illustrated in Table 3.

Table 3. The setting of the ANN hyperparameter

Feed Forward Neural Network	
Activation function hidden layer	Logistic Sigmoid
Number of Hidden Neuron	10-50
Loss function	Cross-entropy

6. Results and Discussion

In this section, the experimental results of TKEO-DSAE, TKEO-ANN, ANN with statistical features are presented. Matlab R2017a installed on a conventional computer with a 2.4GHZ CPU and 8GB memory is used to conduct the experiments. In order to compare the machine learning model, Receiver Operating Characteristics (ROC) curve is used to make the comparison more meaningful. ROC based on the measurement of specificity and sensitivity of all cut off point. The y-axis of ROC curves represents a true positive rate (TPR) while x-axis of ROC curves represents a false positive rate (FPR). TPR denote as the sample ratio that is classified correctly while FPR denotes the sample ratio that is incorrectly classified. The performance of the model is higher if the located on the top and left edges of the plot.

From figure 8, it is shown that DSAE model curve is located closer to the top and edge of the plot. Meanwhile, ANN with statistical features has higher performance compared to ANN model with image features since ANN with statistical features has steeper curve compare to ANN with image features. In order to present more accurate result regarding classification accuracy, the confusion matrix is tabulated in table 4-6. From Table 4 -6, it is clear that SSAE model outperforms ANN methods in term of overall classification accuracy using image patches of TKEO. TKEO-DSAE produced 99.5%. DSAE misclassify one rolling element fault image to normal bearing conditions. However, most of the image sample is correctly classified. Meanwhile, ANN produced 77.2% for image classification. The data used to train and test the ANN model is similar to DSAE dataset. From the confusion matrix in Table 5, ANN model misclassified 41 images which rolling element fault contribute to the highest misclassification rate. Images contain a large size of features which in this study, the size of features is 784. We believe that ANN unable to process the large size of features that make the ANN model unable to produce a satisfactory performance.

Therefore, further study is conducted by extracting nine statistical features from the filtered signal. The analysis shows that ANN model produced 89.0% prediction accuracy of bearing fault classification. The details of the classification are tabulated in Table 6. The ANN model suffers from the same problems which rolling element fault contribute to the highest misclassification among the four classes of bearing conditions. Based on the result, DSAE model outperforms ANN model for bearing fault classification.

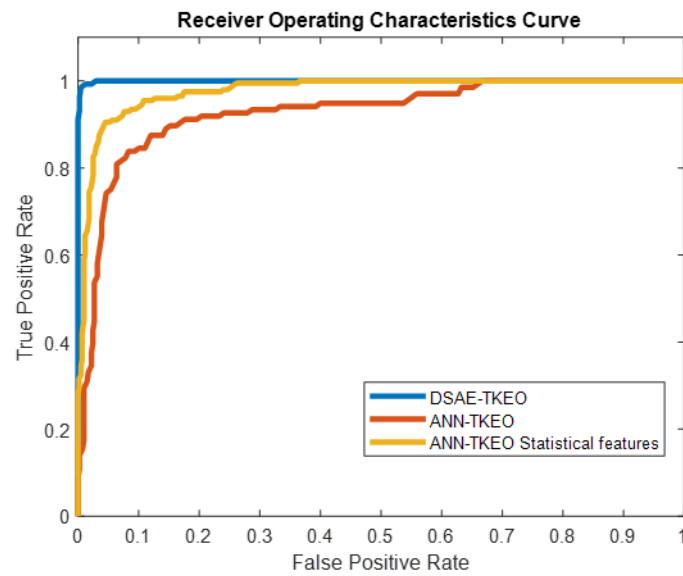


Figure 8. Comparison of ROC curve between three models

Table 4. Confusion matrix of DSAE prediction

Class		Actual			
		Normal	Outer Race Fault	Rolling element Fault	Inner Race Fault
Predicted	Normal	50	0	1	0
	Outer Race Fault	0	50	0	0
	Rolling element Fault	0	0	49	0
	Inner Race Fault	0	0	0	50
Sensitivity (%)		100	100	98	100
Accuracy (%)		99.5			

Table 5. Confusion matrix of ANN prediction

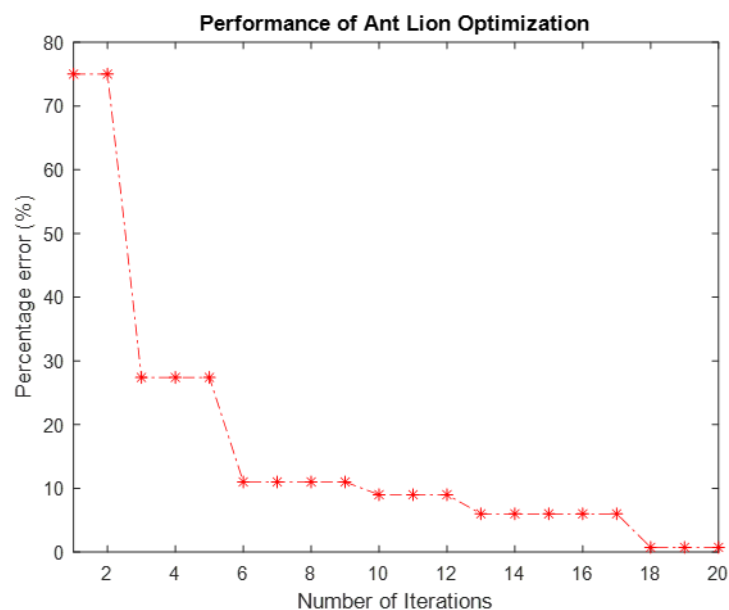
Class		Actual			
		Normal	Outer Race Fault	Rolling element Fault	Inner Race Fault
Predicted	Normal	47	0	10	0
	Outer Race Fault	0	50	0	5
	Rolling element Fault	3	0	25	13

Inner Race Fault	0	0	15	32
Sensitivity (%)	94	100	50	64
Accuracy (%)	77.2			

Table 6. Confusion matrix of ANN prediction using statistical features

Class	Actual			
	Normal	Outer Race Fault	Rolling element Fault	Inner Race Fault
Normal	40	0	8	0
Outer Race Fault	0	50	0	0
Rolling element Fault	10	0	39	1
Inner Race Fault	0	0	3	49
Sensitivity (%)	80	100	78	98
Accuracy (%)	89.0			

The performance of ALO is shown in the following Figure 9. The ALO model is used to optimize the 6 hyperparameters of the DSAE model. The result indicated ALO is able to search the optimal value of hyperparameters as shown in Fig. 9. The ALO require 20 iterations to reach the lowest percentage error of DSAE which is 99.5%. Based on the analysis, we found that ALO requires less parameter setting in order to perform the optimization process of DSAE model.

**Figure 9.** Performance of ALO on DSAE

7. Conclusion

In this paper, the method of applying TKEO and stacked autoencoder DSAE to bearing fault diagnosis problem shows the excellent classification performance compared to TKEO and ANN. During the analysis, the DSAE is optimized using ALO for automated DSAE hyperparameter selection. The proposed model capable of achieving 99.5% prediction accuracy on bearing fault classification. In addition, this DSAE model has good capability to extract a feature from the image compare to ANN model.

Acknowledgements

The authors would like to extend their greatest gratitude to the Institute of Noise and Vibration UTM for funding the study under the Higher Institution Centre of Excellence (HiCoE) Grant Scheme R.K130000.7843.4J227. Additional funding for this research also comes from the UTM Research University Grant (Q.K130000.2543.11H36), and Fundamental Research Grant Scheme (R.K130000.7840.4F653) by The Ministry of Higher Education Malaysia.

References

- [1] Bouzida A, Touhami O, Ibtouen R, Belouchrani A, Fadel M and Rezzoug 2011 *IEEE Trans. Ind. Electron.*, vol. **58**, no. 9, pp. 4385–4395
- [2] Liu W Y, Tang B P, Han J G, Lu X N, Hu N N, and He Z Z 2015 *Renew. Sustain. Energy Rev.*, vol. **44**, pp. 466–472
- [3] Qi X, Theilliol D, Qi J, Zhang Y and Han J 2013 *2013 Int. Conf. Unmanned Aircr. Syst. ICUAS 2013 - Conf. Proc.*, pp. 1114–1118
- [4] Zhao M, Lin J, Miao Y and Xu X 2016 *Meas. J. Int. Meas. Confed.*, vol. **91**, pp. 421–439
- [5] Zhang W, Peng G, Li C, Chen Y and Zhang Z 2017 no. January, pp. 1–21
- [6] Li W, Qiu M, Zhu Z, Wu B and Zhou G 2016 *Meas. Sci. Technol.*, vol. **27**, no. 3
- [7] Dolenc B, Boškoski P and Juričić Đ 2016 *Mech. Syst. Signal Process.*, vol. **66**, pp. 521–532
- [8] Grasso M, Chatterton S, Pennacchi P and Colosimo B M 2016 *Mech. Syst. Signal Process.*, vol. **81**, pp. 126–147, 2016.
- [9] Salakhutdinov R and Hinton G 2009 *Aistats*, vol. **1**, no. 3, pp. 448–455.
- [10] Shao H, Jiang H, Zhao H and Wang F 2017 *Mech. Syst. Signal Process.*, vol. **95**, pp. 187–204
- [11] Shao H, Jiang H, Zhang X and Niu M 2015 *Meas. Sci. Technol.*, vol. 26, no. 11, p. 115002
- [12] Zhang W, Li C, Peng G, Chen Y and Zhang Z 2018 *Mech. Syst. Signal Process.*, vol. **100**, pp. 439–453
- [13] Qu Y, He M, Deutsch J and He D 2017
- [14] Kwak D H, Lee D H, Ahn J H and Koh B H 2013 *Sensors (Basel)*, vol. **14**, no. 1, pp. 283–98
- [15] Kaiser J F 1993 *IEEE Int. Conf. Acoust. Speech Signal Process.*, pp. 149–152 vol.3, 1993.
- [16] Kumar J and Jena P 2017 *Arab. J. Sci. Eng.*, vol. **42**, no. 12, pp. 5003–5013, 2017.
- [17] Henríquez Rodríguez P, Alonso J B, Ferrer M A and Travieso C M 2013 *ISA Trans.*, vol. **52**, no. 2, pp. 278–284
- [18] Tran V T, Althobiani F and Ball A 2014 *Expert Syst. Appl.*, vol. **41**, no. 9, pp. 4113–4122, 2014.
- [19] Mirjalili S 2015 *Adv. Eng. Softw.*, vol. **83**, pp. 80–98