Gearbox Fault Diagnosis Using a Deep Learning Model With Limited Data Sample

Syahril Ramadhan Saufi, Zair Asrar Bin Ahmad, Mohd Salman Leong, and Meng Hee Lim

***Abstract*—Massive volumes of data are needed for deep learning (DL) models to provide accurate diagnosis results. Numerous studies of fault diagnosis systems have demonstrated the effectiveness of DL models over shallow machine learning (SL) in terms of feature extraction, feature dimensional reduction and diagnosis performance. Occasionally, during data acquisition, a problem with a sensor renders some of the data potentially unsuitable for further analysis, leaving only a small data sample. To compensate for this deficiency, a DL model based on a stacked sparse autoencoder (SSAE) model is designed to deal with limited sample data. In this article, the fault diagnosis system is developed based on time-frequency image pattern recognition. Therefore, two gearbox datasets are used to evaluate the proposed diagnosis system. The results from the experiments prove that the proposed system is capable of achieving high diagnostic accuracy even with limited sample data. The proposed fault diagnosis system achieved 100% and 99% diagnosis performance on experimental gearbox and wind turbine gearbox datasets, respectively. The proposed diagnosis system increased diagnosis performance between 10% and 20% over the standard SSAE model. In addition, the proposed model achieved higher diagnosis performance compared to deep neural network and convolutional neural networks models.**

***Index Terms*—Fault diagnosis, gearbox, image recognition, limited data sample, stacked sparse autoencoder (SSAE).**

I. Introduction

FAULT detection and diagnosis, condition monitoring, and system health management are the primary descriptors used to describe the maintenance scheme of machinery components. Most failures in rotating machinery are related to the mechanical transmission system, which contributes 30% of the machine's total maintenance cost [1]. The gearbox is a vital transmission system composed of gears, bearings, and drive shafts that form a complex system [2]. Gears contribute to 60% of gearbox failures, as reported by Yang *et al*. [3]. The breakdown of a gearbox causes a halt in the operation of the machine, followed by a lengthy maintenance process to identify and resolve the problem.

Traditionally, fault diagnosis systems based on shallow learning (SL) models involve five main steps: data/signal acquisition, data/signal preprocessing, feature extraction, feature reduction/selection, and fault diagnosis [4]. The vibration method has been proven to be effective in collecting signals from machinery under high-speed operating conditions. However, the vibration signal is often masked by unwanted signals from unrelated components [5]. Signal preprocessing steps are, therefore, required to filter these unwanted signals. In recent years, deep learning (DL) models have reduced the fault diagnosis process to three main steps: signal acquisition, signal preprocessing, and fault diagnosis [6]. Because DL models have multiple hidden layers, the feature extraction and selection tasks are automated. DL has shown its effectiveness in many applications such as medicine [7], natural language processing, speech recognition, and computer vision [8]. DL models such as convolutional neural networks (CNNs), stacked sparse autoencoders (SSAEs), and deep belief networks (DBNs) have been extensively used in fault diagnosis systems, as reviewed by Zhao *et al*. [9]. The details of DL model variants have been discussed by Guo *et al*. [10].

Even though DL has shown its superior diagnosis performance over SL, DL models require large datasets for the learning process to be effective [11]. For example, Shao *et al.* fed a machinery dataset consisting of 6000 to 9000 samples into a convolutional neural network (CNN) model to produce an accurate fault diagnosis result [12]. Guo *et al.* fed their proposed DL model with 32 256 samples for the training process [13]. Jian *et al.* diagnosed motor bearing conditions using 5,000 data samples in order to obtain a satisfactory diagnosis result [14]. Zhong *et al*. utilised 3000 samples for each fault condition for their generative adversarial network [15]. The great advantage of DL models over SL models is that DL model performance increases with an increase in the size of the training dataset. Increased size of datasets does not affect the performance of SL models once they reach a certain diagnosis performance. Yang *et al*. and Chen *et al*. discussed the difficulty of obtaining enough data samples in industrial environments to train DL models [3][16]. One factor that can cause a shortage of data

Manuscript received June 3, 2019; revised August 18, 2019 and December 23, 2019; accepted January 11, 2020. Date of publication January 20, 2020; date of current version June 22, 2020. This work was supported in part by the Higher Institution Centre of Excellence under Grant Scheme R.K130000.7809.4J225, Grant R.K130000.7809.4J226, Grant R.K130000.7843.4J227, and Grant R. J130000.7824.4J234; and in part by Q. K130000.2543.11H36, Q.J130000.2424.03G86, and Q.J130000.3551.07G54. Paper no. TII-19-2350. *(Corresponding author: Syahril Ramadhan Saufi.)*

S. R. Saufi, Z. A. B. Ahmad, and M. S. Leong are with the School of Mechanical Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, Johor Bahru 81310, Malaysia, and also with the Malaysia and Institute of Noise and Vibration, Universiti Teknologi Malaysia, Kuala Lumpur 54100, Malaysia (e-mail: msramadhan93@gmail.com; zair@mail.fkm.utm.my; salman.leong@gmail.com).

M. H. Lim is with the Institute of Noise and Vibration, Universiti Teknologi Malaysia, Kuala Lumpur 54100, Malaysia (e-mail: limmenghee@gmail.com).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TII.2020.2967822

TABLE I
AUTOENCODER MODEL FOR GEAR FAULT DIAGNOSIS

Additional method	Hyperparameter setting	Data size train/test	Ref
SAE with Gauss-Binary restricted Boltzmann machine	Manual selection	11,200/ 2,800	[28]
Data augmentation	Manual selection	50,000/-	[29]
Ensemble EMD, autoregressive model	Manual selection	300/300	[30]
Fast Fourier transform	Manual selection	665/100	[31]
Multimodal information with frequency transform	Hidden neurons are optimised using grid search. Others are manually selected.	8,000/ 2,000	[32]

samples is sensor problems. Zhang *et al*. emphasized that sensor problems result in poor data quality [17]. DL models are prone to overfitting problems when the data sample size is small [18], [19], Han *et al*. mentioned that a proper architecture settings could prevent the DL model from overfitting and they diagnosed gearbox component using a small dataset sample with 1000 to 5000 samples size [19].

The suitable configuration of DL networks for certain applications remains an open problem. DL models require many hyperparameters to be set manually, which is time-intensive. Number of hidden layers, learning rate, number of hidden nodes, number of epochs, and activation function are among the general hyperparameters of DL models. Each specific type of DL model has its own hyperparameters. For example, CNN models have pooling function, padding, kernel size, number of kernel and kernel stride [20]. The details of CNN hyperparameters are discussed in Sadoughi and Hu [21]. Meanwhile, SSAE models have weight regulariser and sparsity regulariser that need to be set before the training process [22]. Among DL models, the sparse autoencoder has a simple architecture that contains a unique characteristic: the model can be pretrained via greedy layer-wise unsupervised techniques to initialise the parameters (weight and bias) of the network. This process eliminates the burden on the network of tuning the model's parameters during the actual training process. This process has been proposed by Hinton *et al*. [23]. In addition, the SSAE model contains fewer hyperparameter values compared to CNN model. Thus, optimizing the SSAE's hyperparameter is much easier compared to the CNN model. The use of an autoencoder model for gear fault diagnosis is given in Table I. It can be seen that most of the analyses used manual selection of the autoencoder's hyperparameters. In addition, the authors integrated the autoencoder with another method for accurate fault diagnosis.

The proposed model is developed based on a SSAE network, which has several features that help avoid overfit and underfit problems. The SSAE model has several advantages such as unsupervised pretraining, simple architecture and a fast training process [24]. The novelty of this article is to develop the SSAE network with metaheuristics algorithm so that it can be used for diagnosis analysis on limited data sample. The first is a proper hyperparameter selection on SSAE network using a particle swarm optimisation algorithm. The important hyperparameter of SSAE network is the regularizer which should be perfectly tuned as the regularizer is able to reduce the overfitting problem. This is due to the limited data sample that may cause the network to easily overfit the training data. The optimal configuration of hyperparameters varies according to the application, and manual hand tuning or the grid search method are commonly used for hyperparameter selection [25]. Second, instead of being manually selected, the hidden neuron number is simultaneously optimised with other SSAE hyperparameters. Improper hidden neuron size will lead to underfit or overfit problems [26]. For instance, a machine learning model will experience an underfit problem when the number of hidden neurons is lower than the complexity of the input data [146]. Third, the use of low-dimension input data reduces the training load of an SSAE model, and the hyperparameter optimisation process can thus be performed thoroughly. An overfitting problem may occur when the data has high dimensionality [27]. The proposed fault diagnosis model is as follows.

1) Throughout the analysis, the proposed model is fed with a low sample size dataset in which only a single data sensor is used, in order to study the ability of the proposed model to diagnose gearbox conditions with limited data.
2) The proposed model is compared with a standard SSAE model to determine the importance of hyperparameter optimisation. Moreover, the proposed model is compared with other DL models such as CNNs and deep neural networks (DNNs) on the basis of diagnosis performance.
3) The proposed model is compared with SL models to identify the advantages of the proposed model on automated feature extraction.

The rest of the article is organised as follows. Section II discusses the background of the proposed method. Section III describes the experimental procedure. In Section IV, the result produced by the proposed model is discussed. Finally, Section V presents the conclusion and future work.

II. METHODS

A. Proposed Diagnosis System

The proposed diagnosis system is based on the combination of SSAE, particle swarm optimisation (PSO) and fast kurtogram images, as shown in Fig. 1. In this article, a kurtogram that is usually used for bearing components has been implemented in the gearbox system [33]. The analysis involved low-dimensional input data, in the form of a 28×28 grayscale image. A 28×28 image size was adequate for fault diagnosis analysis, as mentioned in [34]. In order to verify the training performance of the proposed diagnosis system, t-stochastic neighbour embedding (t-SNE) is used to visualise the features learned by SSAE's hidden layers in 2-D space. The t-SNE model was proposed by Maaten and Hinton [35]. Two learning algorithms, scaled conjugate and the resilient backpropagation algorithm, were also used throughout the analysis. Moreover, the PSO algorithm optimised four hyperparameters for each sparse autoencoder including number of hidden nodes, weight decay, sparsity,

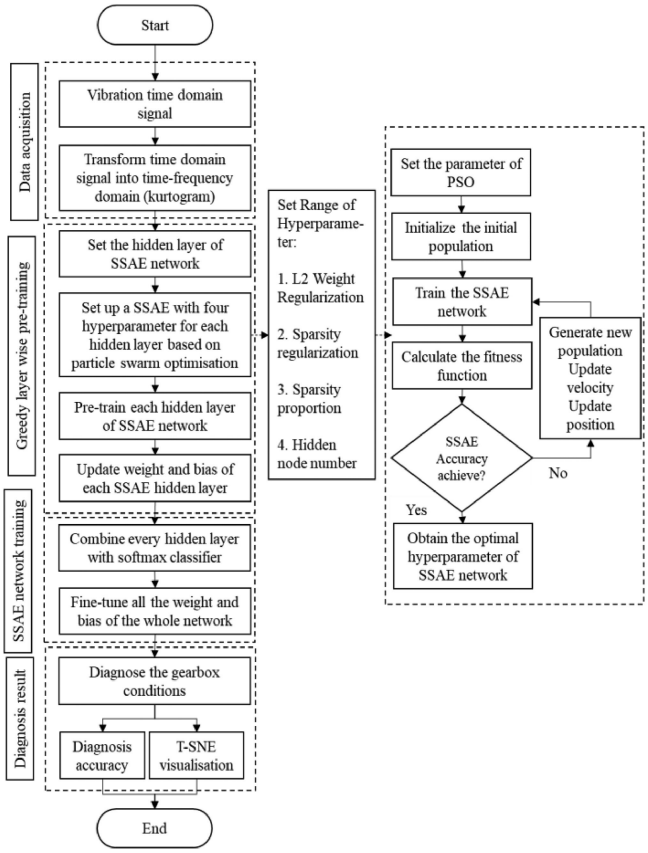

Fig. 1. Proposed diagnosis system.

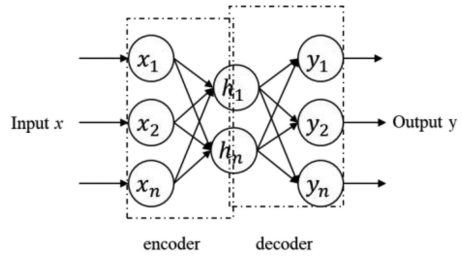

Fig. 2. Simplified structure of sparse autoencoder.

and weight of sparsity penalty term. The objective function is developed based on the classification error of the SSAE model, as shown

$$\text{Test\_error} = \frac{\text{Actual\_value} - \text{predicted\_value}}{\text{total\_number\_of\_prediction}} \tag{1}$$

B. SSAE Architecture

The stacked sparse autoencoder is developed by stacking multiple sparse autoencoders. A sparse autoencoder contains encoder, hidden layer, and decoder functions, as shown in Fig. 2. The encoder maps the input data using the function $h = f(w_1 x + b)$ into a hidden representation $(h \in R^k)$. The hidden representation is reconstructed using the function $y = g(w_2 h + b)$. The sparse autoencoder (SAE) enforces a constraint

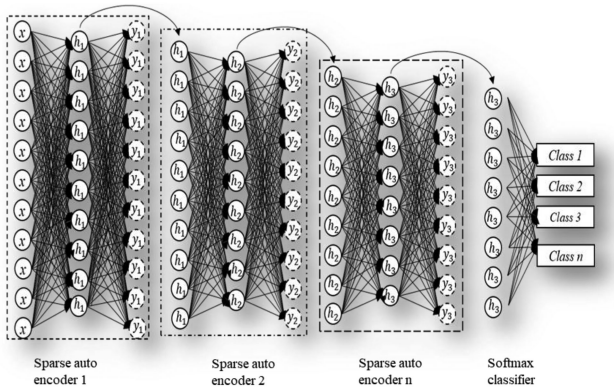

Fig. 3. Structure of SSAE.

on the hidden units of the autoencoder that causes the activation of inactive hidden units [36]. The reconstruction error of a sparse autoencoder is shown

$$E(w,b) = \frac{1}{2}\|h_{w,b}(x) - y\|^2 + \beta \sum_{j=1}^{n} KL(\rho \,\|\tilde{\rho}) + \frac{\lambda}{2}\sum_{l=1}^{nl}\sum_{i}^{sl-1}\sum_{j}^{sl}\left(\omega_{ij}^{(l)}\right)^2 \tag{2}$$

where $KL(\rho\|\tilde{\rho}) = \rho \log \frac{\rho}{\tilde{\rho}_j} + (1-\rho)\log \frac{1-\rho}{1-\tilde{\rho}_j}$ represents the Kullberg–Leibler divergence, β is a weight of sparsity penalty term, ρ is the sparsity parameter and $\widehat{\rho_j}$ is the average activation of the hidden unit.

The final layer of an SSAE network is a softmax classifier that aims to classify the features that were processed by the SAE. The softmax equation is defined as

$$h_\theta(x^i) = \begin{bmatrix} p(y^i = 1 \,|\, x^i;\theta) \\ p(y^i = 2 \,|\, x^i;\theta) \\ \vdots \\ p(y^i = k \,|\, x^i;\theta) \end{bmatrix} = \frac{1}{\sum_{l=1}^{k} e^{\theta_j^T x^i}} \begin{bmatrix} e^{\theta_1^T x^i} \\ e^{\theta_2^T x^i} \\ \vdots \\ e^{\theta_k^T x^i} \end{bmatrix} \tag{3}$$

where $\theta_1, \theta_2, \ldots, \theta_k \in \Re^{n+1}$ are the model parameters and $1/\sum_{i=1}^{k} e^{\theta_j^T x^i}$ normalizes the distribution to ensure that the sum is equal to one. The overall structure of an SSAE network is shown in Fig. 3.

C. Particle Swarm Optimization (PSO)

The development of artificial intelligence technology has allowed techniques, such as the genetic algorithm, PSO, and differential evolution (DE) to be applied to optimise the hyper-parameters of machine learning. Optimisation techniques were comprehensively reviewed in [37], and the authors' analysis showed that PSO methods is comparable to the DE algorithm.

The PSO algorithm consists of three important parameters: a personal experience (Pbest), overall experience (Gbest), and the present movement of the particles to determine their next positions in the search space.

Each particle updates its position and velocity based on

$$v_{id}^{k+1} = w * v_{id}^{k} + c_1 * \left(p\text{best}_{id}^{k} - p_{id}^{k}\right) + c_2 * \text{rand}_2 * \left(g\text{best}_{id}^{k} - p_{id}^{k}\right) \quad (4)$$

and

$$p_{id}^{k+1} = p_{id}^{k} + v_{id}^{k+1} \quad (5)$$

where d is the dimension of a particle in the kth iteration $(1 \leq d \leq n)$ and v is the velocity of the ith particle, which is restricted to the range $[-v_{\max}, v_{\max}]$. Rand*1*() and rand*2*() are random numbers in the range of [0, 1], w is the weight of inertia, and positive constants c_1 and c_2 represent personal and social learning factors, respectively. The capabilities of global exploration and local exploration can be balanced using

$$w_k = w_{\max} - \frac{w_{\max} - w_{\min}}{t_{\max}} * t \quad (6)$$

where $w_{\min}$ represents the minimum inertia weight, $w_{\max}$ represents the maximum inertia weight and t represents the current iteration. During the analysis, the swarm size is set to 4 with 25 iterations.

D. Time-Frequency Transformation

Time-frequency analysis is useful to reveal the energy contained in the machinery signal in time and frequency representations. Several types of time-frequency transformations such as spectral kurtosis diagram (kurtogram), empirical mode decomposition (EMD), short-time Fourier transform (STFT), and wavelet transform (WT) have been used for machinery fault diagnosis. However, all of these methods have flaws. It is difficult to select the mother wavelet function for WT [38], while EMD methods suffer from mode mixing problems [39]. In addition, in STFT models, the time and frequency resolution are linked together [40]. The fast kurtogram method provides some advantages over other methods as the current generation of fast kurtogram does not require any parameters to be set, making the model easy to use. Therefore, the fast kurtogram was chosen for the proposed diagnosis system to reduce the number of parameters that needed to be set.

1) Fast Kurtogram: The kurtogram is the fourth order of spectral analysis, which is known as spectral kurtosis. Antoni imposed four constraints on kurtogram analysis to increase the versatility of the process, which makes the model more sensitive to nonstationary signals [41]. The details of the fast kurtogram algorithm can be found in [41]. The kurtogram can be defined using

$$k_x f(x) = \frac{\langle H^4(t,f)\rangle}{\langle H^2(t,f)\rangle^2} - 2 \quad (7)$$

where $\langle . \rangle$ represents the operator of time averaging and $H(t, f)$ represents the enveloping process of signal $x(t)$ [42]. $H(t, f)$ is estimated using

$$H(t,f) = \sum_{n=t}^{t+N_n-1} w(n-t)x(n)e^{-j2\pi fn} \quad (8)$$

where $w(t)$ represents the length of the window.

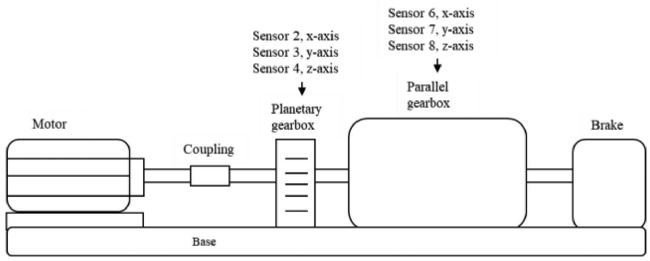

Fig. 4. Experimental rig of the gearbox.

TABLE II
DATA DISTRIBUTION

Gear condition	Training data	Testing data	Class
Normal	100	50	1
Crack on gear feet	100	50	2
Missing one gear foot	100	50	3
Crack on gear root	100	50	4
Wear on gear surface	100	50	5
Total data	500	250	

III. EXPERIMENTAL PROCEDURE

The experiments involved one experimental gearbox and one wind turbine gearbox. The data were divided into training and testing sets, using 100 data points per gear condition for training and 50 data points per gear condition for testing. Since there are no published studies that describe the minimum size of a dataset that can be used for training a DL model. Hence, we used less dataset compared to size of Han *et al*. dataset [19]. The experimental rig is shown in Fig. 4, and the details of the experimental setup can be found in [12]. In this article, only gear faults were used for the analysis. Five types of gear signals were obtained during data acquisition: normal, crack on gear feet, missing one gear foot, crack on gear root, and wear on gear surface. Each gear conditions was assigned a class for the fault diagnosis analysis, as given in Table II. The gearbox was operated at 20 and 30 Hz with eight sensors located on the test rig. Sensors 1 and 5 were a motor and torque sensor, respectively, sensors 2, 3, and 4 were located on the planetary gearbox and sensors 6, 7, and 8 were located on the parallel gearbox. In this study, only sensors 2, 3, 4, 6, 7, and 8 were used to evaluate the proposed diagnosis system. The sampling frequency was set at 2 000 Hz. With sampling frequency information, the time domain signal of the gearbox can be transformed into time-frequency domain using kurtogram analysis, and example images are shown in Fig. 5.

IV. RESULT AND DISCUSSION

Instead of a multisensor fusion system, this article evaluated the proposed model on a single sensor. A multisensor fusion system might produce an accurate diagnosis result; however, in real applications, not all of the sensors necessarily provide a quality signal for the analysis. Thus, the main objective is to

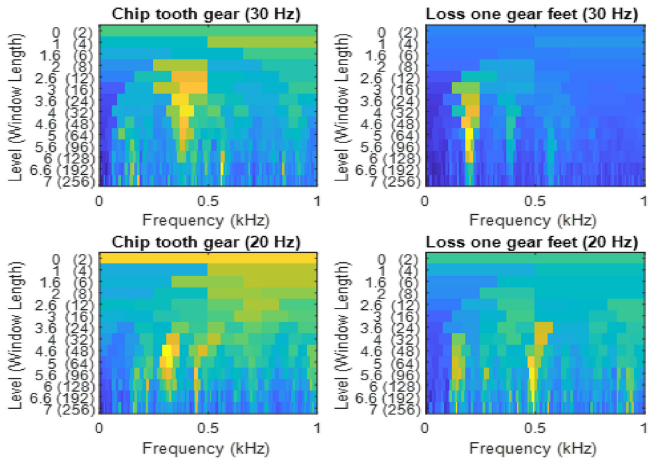

Fig. 5. Kurtogram images of gear signals for chipped gear tooth and loss of one gear foot.

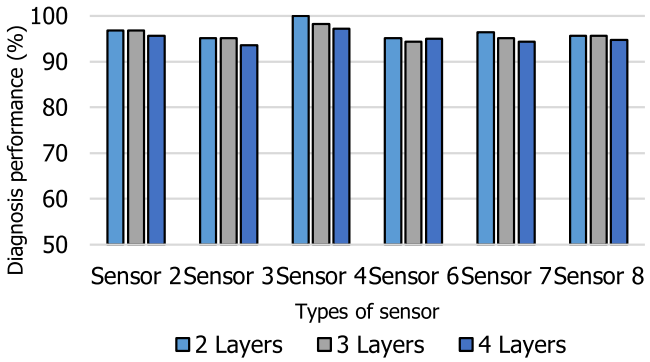

Fig. 6. Diagnosis performance of the proposed model (30 Hz).

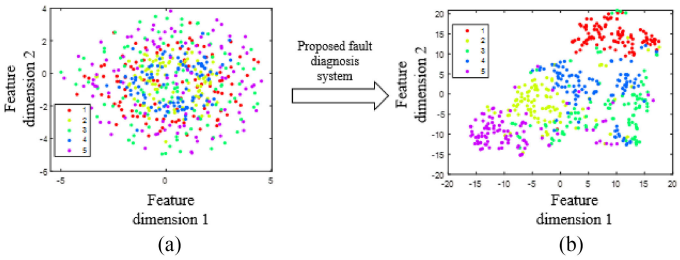

Fig. 7. T-SNE visualisation of SSAE model with two layers with sensor 4 data (30 Hz).

perform accurate fault diagnosis on limited data samples. In this article, the effect of the number of hidden layers in the network on diagnosis performance is evaluated, as hidden layer size is one of the factors that can cause the architecture to overfit [43]. The diagnosis performance of the proposed system is shown in Fig. 6 and 8 for 30 and 20 Hz, respectively.

It can be seen that the performance of the proposed model varies according to the type of sensor. Effective sensor selection is a crucial process in fault diagnosis analysis, as too much information from multiple sensors results in increased data processing needs and system costs [44]. In this case, sensor 4

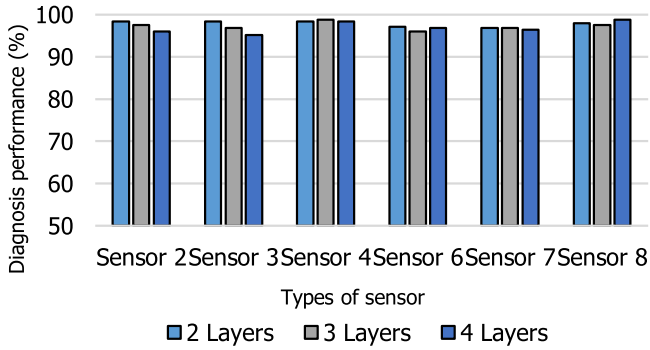

Fig. 8. Diagnosis performance of the proposed model (20 Hz).

signal contains more information of fault characteristics compared to the other sensors, as the proposed model achieved highest diagnosis performance. The result is shown in Fig. 6. No significant difference was observed in the performance of SSAEs with different hidden layer sizes. Therefore, two layers are more suitable for diagnosis, as this size requires less training time for parameter and hyperparameter optimisation.

The visualisation analysis was conducted on the two-layer SSAE model using the t-SNE method as the model achieved 100% diagnosis performance. As Fig. 7(a) illustrates, the t-SNE model was unable to visualise the data obtained directly from the sensor. In Fig. 7(b), the t-SNE was capable of visualising the features processed by the proposed system, and a specific class of each feature can be observed. Even though some of the classes overlap with each other, the proposed model achieved a 100% diagnosis performance once the feature was fed into the softmax layer. It is worth mentioning that the quality of hidden layer feature visualisation depends on the t-SNE model configuration, which is based on the default setup. The combination of SSAE hidden layer features and t-SNE could provide a highly effective fault visualisation system in the future. The diagnosis can be done based on the visualisation of the t-SNE, where the number of faults can be determined using the t-SNE mapping. Therefore, the dependency of the SSAE model on the target value can be reduced. The current limitation of the machine learning model requires an exact target during the training process in order to perform well. However, the target data are usually difficult to obtain in real industries [45].

Subsequently, the proposed diagnosis system was evaluated on the 20 Hz dataset, and the result is shown in Fig. 8. The system achieved the highest performance accuracy at 98.8% for sensors 4 and 8. Three hidden layers were required to analyse the data from sensor 4, and four hidden layers for sensor 8. The overall performance of the proposed system was above 95%, which is slightly higher compared to performance on the 30 Hz dataset. The feature learned by the proposed system was visualised via the t-SNE method as shown in Fig. 9 and 10. Fig. 9(a) and 10(a) represent the t-SNE visualisation from the vibration signal before being processed by the proposed diagnosis system. When the vibration signal is analysed using the proposed system, the t-SNE model was capable of clustering the features from the gearbox dataset, as shown in Fig. 9(b) and 10(b).

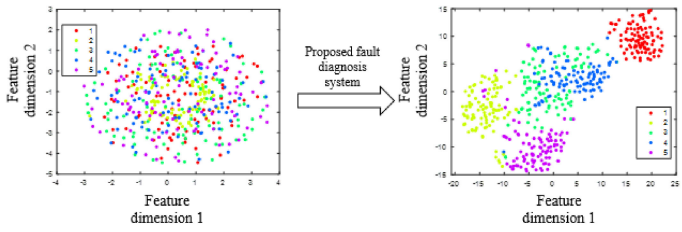

Fig. 9. T-SNE visualisation of SSAE model with three layers with sensor 4 data (20 Hz).

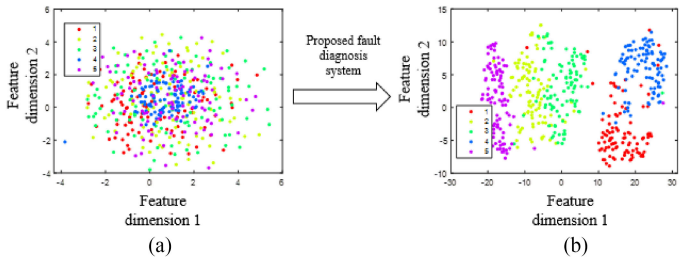

Fig. 10. T-SNE visualisation of SSAE model with four layers with sensor 8 data (20 Hz).

TABLE III
DIAGNOSIS PERFORMANCE OF DL MODELS (30 HZ)

Sensor location	Standard SSAE	DNN	CNN
Sensor 2	86.4%	77.6%	80.4%
Sensor 3	85.6%	67.2%	75.6%
Sensor 4	80.4%	80.8%	81.6%
Sensor 6	84.0%	76.0%	82.0%
Sensor 7	70.0%	63.6%	79.6%
Sensor 8	74.0%	65.2%	83.2%

TABLE IV
DIAGNOSIS PERFORMANCE OF DL MODELS (20 HZ)

Sensor location	Standard SSAE	DNN	CNN
Sensor 2	86.0%	80.4%	87.6%
Sensor 3	86.0%	86.8%	81.2%
Sensor 4	82.4%	84.8%	83.6%
Sensor 6	85.6%	80.0%	76.8%
Sensor 7	81.2%	75.6%	81.6%
Sensor 8	85.6%	79.6%	84.4%

A. Comparative Study of DL Models

Several DL models including standard SSAE, DNN, and CNN were selected for comparative study. The models were evaluated using six types of data collected from sensors at different locations. The diagnosis performance of each DL model is given in Tables III and IV for 30 and 20 Hz datasets, respectively. Only a slight difference was observed between the diagnosis performance of standard SSAE and the CNN models. DNN is unsuitable for these types of datasets, as the model achieved more than 80% diagnosis performance only on sensor 4. The standard SSAE model achieved approximately 70 to 86.4% diagnosis accuracy, while the proposed diagnosis system based on the SSAE model achieved 94 to 100% accuracy, an increase of 10 to 20%. Analysis of these SSAE models revealed that they suffered from inadequate hyperparameter tuning. Therefore, it is proven that hyperparameter selection for the SSAE model is important for accurate diagnosis. In addition, Lu *et al*. and Albelwi *et al*. emphasized the important of hyperparameter selection in the DNN and CNN, respectively [46]. Albelwi and Mahmood stated that hyperparameter selection requires a prior knowledge to find the optimal configuration for a CNN model [47].

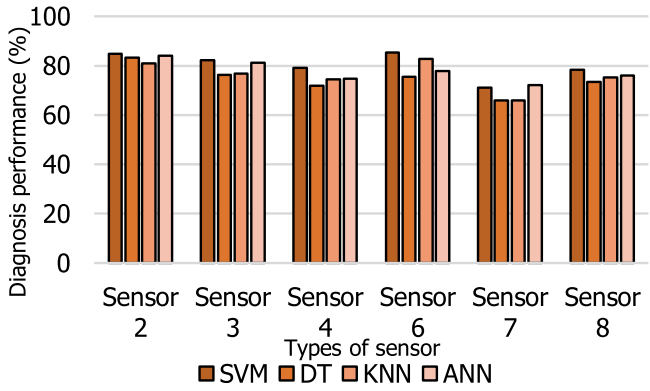

Fig. 11. Diagnosis performance of the SL model (30 Hz).

B. Traditional Fault Diagnosis

Ten statistical features were extracted from the vibration signal: root mean square, standard deviation, peak amplitude, energy, kurtosis, skewness, margin factor, impulse factor, crest factor, and shape factor. The analysis involved the use of SL models, such as an artificial neural network (ANN), support vector machine (SVM), decision tree (DT), and k-nearest neighbour (KNN). The ANN hyperparameters were set to a single hidden layer with 500 hidden neurons and a gradient descent algorithm. Meanwhile, a radial basis function kernel was used for the SVM model. For the DT model, the split criterion is the Gini diversity index and Euclidean distance is the hyperparameter used for KNN model. The result is shown in Fig. 11 and 12 for the 30 Hz and 20 Hz datasets, respectively. As can be seen in Fig. 11, there was no significant difference in the diagnosis performance of the shallow models. The accuracy of the models was around 60% to 85%, with the highest diagnosis performance achieved by the SVM model using the sensor 6 signal for the 30 Hz dataset and the sensor 2 signal for 20 Hz. In comparison with the proposed model system, the SL model produced inaccurate diagnosis results.

C. Wind Turbine Gearbox Analysis

The proposed model was evaluated on the real structure of a wind turbine gearbox, as pictured in Fig. 13. The dataset contained two types of vibration signals from normal and fault conditions. The wind turbine gearbox was driven at 30 Hz during

TABLE V
DIAGNOSIS PERFORMANCE ON WIND TURBINE GEARBOX

Types of model	Proposed model			Standard SSAE	CNN	DNN	SVM	KNN	DT
	2 layers	3 layers	4 layers						
Diagnosis performance (%)	99.0	99.0	99.0	84.0	90.0	82.0	86.8	85.9	82.6

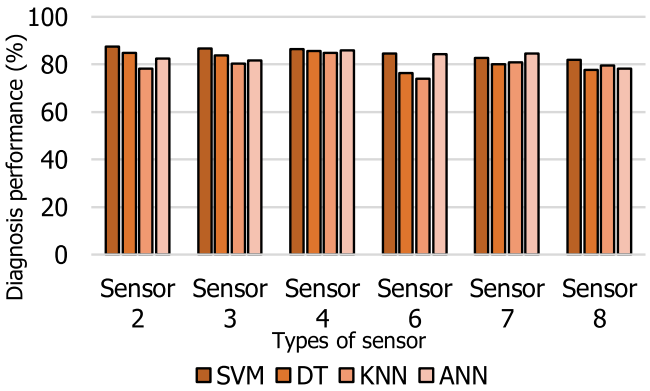

Fig. 12. Diagnosis performance of the SL model (20 Hz).

Fig. 13. Wind turbine gearbox [48].

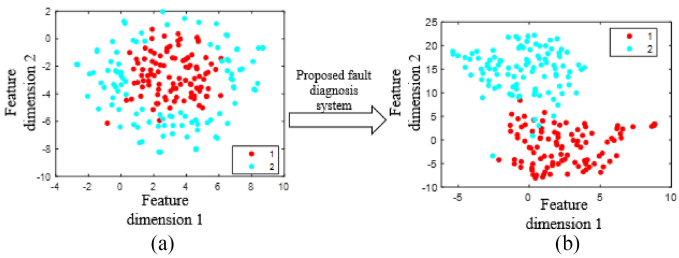

Fig. 14. T-SNE visualisation of SSAE model with two layers.

the data acquisition process. The diagnosis performance is given in Table V. The number of layers did not affect performance when using the wind turbine gearbox dataset; the proposed model achieved 99% diagnosis performance using two, three, or four layers. It was proved that accurate diagnosis performance can be achieved using a three-step fault diagnosis system without need of advanced signal processing, feature extraction and feature selection. The t-SNE analysis shown in Fig. 14(b) provides strong evidence of the results obtained by the proposed diagnosis system. In general, the data obtained from the real structure was heavily masked by unwanted noise, which made it difficult to process, but the proposed model was nonetheless able to achieve satisfactory diagnosis performance. This proves that the proposed system is robust.

V. CONCLUSION

In this article, a fault diagnosis system based on the optimised SSAE model was proposed. In the proposed system, the standard SSAE was modified so that the hyperparameters of each hidden layer were optimised using particle swarm optimisation. In addition, the proposed diagnosis system implements the image recognition-based approach without need for manual feature extraction. Moreover, the system was built to be suited to limited sample size.

The proposed diagnosis system was evaluated on one experimental gearbox dataset and a wind turbine gearbox dataset. The results proved that the proposed model can adaptively extract features from the image of time-frequency analysis and can achieve satisfactory performance using data from a single sensor. The combination of the proposed model and t-SNE was the potential to be used for unsupervised fault visualisation method, which can provide rapid fault detection and diagnosis on machinery components.

REFERENCES

[1] F. Elasha, M. Greaves, and D. Mba, "Planetary bearing defect detection in a commercial helicopter main gearbox with vibration and acoustic emission," *Struct. Heal. Monit.*, vol. 17, no. 5, pp. 1192–1212, 2018.
[2] Y. Wang, S. Yang, and R. V. Sanchez, "Gearbox fault diagnosis based on a novel hybrid feature reduction method," *IEEE Access*, vol. 6, no. 1, pp. 75813–75823, 2018.
[3] D. Yang, Y. Liu, S. Li, X. Li, and L. Ma, "Gear fault diagnosis based on support vector machine optimized by artificial bee colony algorithm," *Mech. Mach. Theory*, vol. 90, pp. 219–229, 2015.
[4] S. Khan and T. Yairi, "A review on the application of deep learning in system health management," *Mech. Syst. Signal Process.*, vol. 107, pp. 241–265, 2018.
[5] T. Wang, Q. Han, F. Chu, and Z. Feng, "Vibration based condition monitoring and fault diagnosis of wind turbine planetary gearbox: A review," *Mech. Syst. Signal Process.*, vol. 126, pp. 662–685, 2019.
[6] S. R. Saufi, Z. A. bin Ahmad, M. S. Leong, and M. H. Lim, "Low-Speed bearing fault diagnosis based on ArSSAE model using acoustic emission and vibration signals," *IEEE Access*, vol. 7, no. 1, pp. 46885–46897, 2019.
[7] J. Xu, L. Xiang, R. Hang, and J. Wu, "Stacked sparse autoencoder (SSAE) based framework for nuclei patch classification on breast cancer histopathology," in *Proc. IEEE 11th Int. Symp. Biomed. Imag.*, 2014, pp. 999–1002.
[8] L. Deng, G. Hinton, and B. Kingsbury, "New types of deep neural network learning for speech recognition and related applications: an overview," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2013, pp. 8599–8603.
[9] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mech. Syst. Signal Process.*, vol. 115, pp. 213–237, 2019.
[10] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27–48, 2016.
[11] L. Zhang, Z. He, and Y. Liu, "Deep object recognition across domains based on adaptive extreme learning machine," *Neurocomputing*, vol. 239, pp. 194–203, 2017.

[12] S. Shao, S. McAleer, R. Yan, and P. Baldi, "Highly accurate machine fault diagnosis using deep transfer learning," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2446–2455, Apr. 2019.
[13] S. Guo, T. Yang, W. Gao, C. Zhang, and Y. Zhang, "An intelligent fault diagnosis method for bearings with variable rotating speed based on pythagorean spatial pyramid pooling CNN," *Sensors (Switzerland)*, vol. 18, no. 11, p. 3857, 2018.
[14] X. Jian, W. Li, X. Guo, and R. Wang, "Fault diagnosis of motor bearings based on a one-dimensional fusion neural network," *Sensors (Switzerland)*, vol. 19, no. 1, p. 122, 2019.
[15] C. Zhong, K. Yan, Y. Dai, N. Jin, and B. Lou, "Energy efficiency solutions for buildings: Automated fault diagnosis of air handling units using generative adversarial networks," *Energies*, vol. 12, no. 3, p. 527, 2019.
[16] Y. Chen, G. Peng, C. Xie, W. Zhang, C. Li, and S. Liu, "ACDIN: Bridging the gap between artificial and real bearing damages for bearing fault diagnosis," *Neurocomputing*, vol. 294, pp. 61–71, 2018.
[17] Q. Zhang, L. T. Yang, Z. Chen, and P. Li, "A survey on deep learning for big data," *Inf. Fusion*, vol. 42, pp. 146–157, 2018.
[18] M. Xia, T. Li, L. Xu, L. Liu, and C. W. de Silva,"Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 1, pp. 101–110, Feb. 2018..
[19] T. Han, C. Liu, W. Yang, and D. Jiang, "A novel adversarial learning framework in deep convolutional neural network for intelligent diagnosis of mechanical faults," *Knowl., Based Syst*., vol. 165, pp. 474–487, 2019.
[20] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," *Mech. Syst. Signal Process.*, vol. 100, pp. 439–453, 2018.
[21] M. Sadoughi and C. Hu, "Physics-based convolutional neural network for fault diagnosis of rolling element bearings," *IEEE Sens. J.*, vol. 19, no. 11, pp. 4181–4192, Jun.1,2019.
[22] C. Lu, Z. Y. Wang, W. L. Qin, and J. Ma, "Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification," *Signal Process.*, vol. 130, pp. 377–388, 2017.
[23] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput*., vol. 18, no. 7, pp. 1527–1554, 2006.
[24] H. O. A. Ahmed, M. L. D. Wong, and A. K. Nandi, "Effects of deep neural network parameters on classification of bearing faults," in *Proc. 42nd Annu. Conf. IEEE Ind. Electron. Soc.*, 2016, pp. 6329–6334.
[25] J. Bergstra, B. Komer, C. Eliasmith, D. Yamins, and D. D. Cox, "Hyperopt: A Python library for model selection and hyperparameter optimization," *Comput. Sci. Discovery*, vol. 8, no. 1, p. 014008, 2015.
[26] K. G. Sheela and S. N. Deepa, "Review on methods to Fix number of hidden neurons in neural networks," *Math. Probl. Eng.*, vol. 2013, 2013. [Online]. Available: https://doi.org/10.1155/2013/425740
[27] A. Widodo and B.-S. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mech. Syst. Signal Process.*, vol. 21, no. 6, pp. 2560–2574, 2007.
[28] J. Li, X. Li, D. He, and Y. Qu, "A novel method for early gear pitting fault diagnosis using stacked SAE and GBRBM," *Sensors (Switzerland)*, vol. 19, no. 4, pp. 1–17, 2019.
[29] X. Li, Z. Liu, Y. Qu, and D. He, "Unsupervised gear fault diagnosis using raw vibration signal based on deep learning," in *Proc. Prognostics Syst. Health Manage. Conf.*, 2019, pp. 1025–1030.
[30] Y. Qi, C. Shen, D. Wang, J. Shi, X. Jiang, and Z. Zhu, "Stacked sparse autoencoder-based deep network for fault diagnosis of rotating machinery," *IEEE Access*, vol. 5, pp. 15066–15079, 2017.
[31] Y. Ren Wang, Q. Jin, G. Dong Sun, and C. Fei Sun, "Planetary gearbox fault feature learning using conditional variational neural networks under noise environment," *Knowl., Based Syst*., vol. 163, pp. 438–449, 2019.
[32] M. Ma, C. Sun, and X. Chen, "Deep coupling autoencoder for fault diagnosis with multimodal sensory data," *IEEE Trans. Ind. Informat.*, vol. 14, no. 3, pp. 1137–1145, Mar. 2018.
[33] P. H. Nguyen and J. M. Kim, "Multifault diagnosis of rolling element bearings using a wavelet Kurtogram and vector median-based feature analysis," *Shock Vib.*, vol. 2015, 2015.
[34] D. Verstraete, M. Engineering, and M. Engineering, "Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings," *Hindawi Shock Vib.*, vol. 2017, pp. 1–29, 2017.
[35] L. Van Der Maaten and G. Hinton, "Visualizing high-dimensional data using t-sne," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, 2008.
[36] Y. Wang, M. Liu, Z. Bao, and S. Zhang, "Stacked sparse autoencoder with PCA and SVM for data-based line trip fault diagnosis in power systems," *Neural Comput. Appl.*, vol. 5, pp. 1–13, 2018.
[37] M. N. Ab Wahab, S. Nefti-Meziani, and A. Atyabi, "A comprehensive review of swarm optimization algorithms," *PLoS One*, vol. 10, no. 5, pp. 1–36, 2015.
[38] M. Zhao, B. Tang, and Q. Tan, "Bearing remaining useful life estimation based on time-frequency representation and supervised dimensionality reduction," *Meas. J. Int. Meas. Confederation*, vol. 86, pp. 41–55, 2016.
[39] Y. Wei, Y. Li, M. Xu, and W. Huang, "A review of early fault diagnosis approaches and their applications in rotating machinery," *Entropy*, vol. 21, no. 4, pp. 1–26, 2019.
[40] P. K. Wong, J. Zhong, Z. Yang, and C. M. Vong, "Sparse Bayesian extreme learning committee machine for engine simultaneous fault diagnosis," *Neurocomputing*, vol. 174, pp. 331–343, 2016.
[41] J. Antoni, "Fast computation of the kurtogram for the detection of transient faults," *Mech. Syst. Signal Process.*, vol. 21, no. 1, pp. 108–124, 2007.
[42] Y. Wang, Z. He, and Y. Zi, "Enhancement of signal denoising and multiple fault signatures detecting in rotating machinery using dual-tree complex wavelet transform," *Mech. Syst. Signal Process.*, vol. 24, no. 1, pp. 119–137, 2010.
[43] C. Liu, G. Cheng, X. Chen, and Y. Pang, "Planetary gears feature extraction and fault diagnosis method based on VMD and CNN," *Sensors (Switzerland)*, vol. 18, no. 5, 2018.
[44] L. Liu, D. Liu, Y. Zhang, and Y. Peng, "Effective sensor selection and data anomaly detection for condition monitoring of aircraft engines," *Sensors (Switzerland)*, vol. 16, no. 5, 2016.
[45] X. Li, W. Zhang, Q. Ding, and J. Q. Sun, "Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation," *J. Intell. Manuf.*, pp. 1–20, 2018. [Online]. Available: https://doi.org/10.1007/s10845-018-1456-1
[46] W. Lu, B. Liang, Y. Cheng, D. Meng, S. Member, and J. Yang, "Deep model based domain adaption for fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 64, no. 3, pp. 2296–2305, Mar. 2017.
[47] S. Albelwi and A. Mahmood, "A framework for designing the architectures of deep convolutional neural networks," *Entropy*, vol. 19, no. 6, 2017.
[48] E. Bechhoefer, "Acoustics and vibration database," 2014. [Online]. Available: http://data-acoustics.com/measurements/gear-faults/gear-1/, Accessed on: Jun. 25, 2018.

Syahril Ramadhan Saufi is working toward the Ph.D. degree in mechanical engineering from Noise and Vibration Institute, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia.

His current research interests include machinery fault diagnosis and artificial intelligence.

Zair Asrar bin Ahmad received the Ph.D. degree in mechanical engineering from Otto-von-Guericke-University Magdeburg, Magdeburg, Germany, in 2011.

He is currently a Senior Lecturer with the Faculty of Engineering, Universiti Teknologi Malaysia, Skudai, Malaysia. His current research interests include civil engineering, acoustic engineering and structural engineering.

Mohd Salman Leong received the B.Sc. degree in mechanical engineering and the Ph.D. degree in rotor dynamics from Heriot-Watt University, in 1983.

He has more than 35 years professional engineering consulting experience, and is acknowledged by the industry and government agencies as the leading authority in acoustics, noise and vibration in the country. He has been involved in many of the mega-projects and high impact consulting and investigation projects in oil and gas, power generation, infrastructure and construction industries. He was a Professor and a Principal consultant, and the Founding Director with the Institute of Noise and Vibration, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia.

Meng Hee Lim received the Ph.D. degree in vibration from the Universiti Teknologi Malaysia, in 2010.

His research interests in blade fault diagnosis in turbomachinery. He has more than 13 years professional consulting experience in acoustics and vibration, with specialization in machinery diagnostics. He has extensive experience in power generation gas and steam turbines, blade failures diagnostics and experimental modal analysis. He is currently a Senior Lecturer and a Consultant with the Noise and Vibration Institute, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia.