



Machinery fault diagnosis based on a modified hybrid deep sparse autoencoder using a raw vibration time-series signal

Syahril Ramadhan Saufi^{1,2} · Muhammad Firdaus Isham^{1,2} · Zair Asrar Ahmad^{1,2} · M. Danial Abu Hasan^{1,2}

Received: 16 February 2022 / Accepted: 20 September 2022 / Published online: 5 October 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Intelligent fault diagnosis (IFD) is an effective system to ensure the safe operation of mechanical components such as bearings, gears, and blades. The main challenge of IFD using traditional methods lies in finding the good features that reflect the machine conditions that need prior knowledge and essential expertise to identify the features. In order to solve this problem, this paper introduces a new IFD technique based on a deep sparse autoencoder (DSAE) using a raw vibration time-domain signal (1D time-series signal). A novel method called modified hybrid DSAE was developed to avoid the need for feature extraction and selection steps. First, a resilient backpropagation learning algorithm was employed on the three hidden layers of the DSAE network. Then, a combination of sigmoid and rectified linear unit (RELU) activation functions was proposed for the DSAE network. Finally, the hyperparameters of the DSAE network were optimized using the grey wolf optimizer technique (GWO). The proposed method was applied to analyze the 1D time-series signal of five machinery datasets, including three bearings, one gearbox, and one turbine blade. The analysis proved that the diagnosis performance achieved by the proposed model is highly reliable and applicable to fault diagnosis of machinery components. The model achieved 100 per cent diagnosis accuracy on four datasets (MFPT, CWRU, gearbox, turbine blade) and 95 per cent diagnosis accuracy on the MaFaulDa dataset. The results from the proposed model show superior diagnostic accuracy compared to other related studies.

Keywords Fault diagnosis · Grey wolf optimiser · Hybrid activation function · Deep sparse autoencoder

1 Introduction

Mechanical transmission systems contain core components such as bearings, gears, and blades. These components have been widely used in motor vehicles, heavy machinery, and wind turbines (Sun and Li 2022). Modern industries use a wide range of rotating machinery systems, which are expected to run continuously. Consequently, an unexpected machinery breakdown due to failure in these vital components can lead to high maintenance costs. Maintenance costs should be reduced as much as possible in a profit-making industry. Hence, recent decades have seen increasingly rapid advances in intelligent machinery fault diagnosis to provide an accurate and robust diagnosis system. Intelligent fault

diagnosis (IFD) systems aim to analyse the collected data and automatically deliver a diagnosis result. Previous work reported that signal processing, feature extraction, feature selection, and classification methods are common procedures in IFD systems that use traditional machine learning models (Wang et al. 2021). Signal processing is mainly used to filter the unwanted signal or noise and retain the important signal reflecting the machinery conditions (Habbouche et al. 2021). This is necessary because the signals obtained from machinery components contain nonlinear and non-stationary characteristics. In addition, extracting and selecting the best features from the processed signal is still a challenging task that needs much prior domain knowledge (Chen et al. 2021). The features from the signal domain are manually extracted based on empirical analysis, which leads to inconsistent and unreliable diagnosis results (Ayas and Ayas 2022). The performance of traditional machine learning models such as artificial neural networks (ANNs) and support vector machines (SVMs) heavily depends on the extracted features. In most diagnoses, some insensitive features do not reflect the machine conditions. Hence, feature selection must be

✉ Syahril Ramadhan Saufi
msyahramadhan41@gmail.com

¹ Faculty of Mechanical Engineering, Universiti Teknologi Malaysia (UTM), 81310 Johor Bahru, Malaysia

² Institute of Noise and Vibration, Universiti Teknologi Malaysia (UTM), 54100 Kuala Lumpur, Malaysia

carried out to remove unwanted features. A literature review reveals that an IFD system that utilizes traditional machine learning highly depends on feature extraction and selection steps to provide an accurate diagnosis result (Yan et al. 2022).

In recent years, the deep learning model has been widely used in many fields due to its capacity for generalizing and learning complex data distribution. Deep stacked sparse autoencoders have been successfully applied in various applications such as speech recognition, image classification, and dimensionality reduction (Xu et al. 2014; Guo et al. 2016; Qi et al. 2017). According to a study, implementing autoencoders in a deep learning model produces greater classification accuracy compared to a series of convolutional neural networks (Gupta and Gupta 2021). The ability of deep sparse autoencoders (DSAEs) to extract high-order correlations and deep features from data using multiple hidden layers is the main reason for their successful application in many fields (Mao et al. 2021). Due to this ability, the DSAE model can use its network to perform the feature extraction and selection process, which is very useful in fault diagnosis analysis. This study uses a DSAE because a DSAE network can learn the input feature using the unsupervised greedy layer-wise pre-training process (Hinton et al. 2006). Bengio et al. elucidated three important steps of the greedy layer-wise pre-training method (Bengio et al. 2007): firstly, pre-train one layer of the deep learning model at a time in a greedy way; secondly, preserve the information from the input by using unsupervised learning at each layer; and finally, fine-tune the whole network to the target of interest. This pre-training step is performed for better generalization and to prevent the network from overfitting. Previous research has described many methods that have been developed to improve current rotating machinery IFD systems. However, none of them is applicable for multiple rotating machinery components. In addition, a problem with a deep learning model such as DSAE is hyperparameter selection, which in current practice is based on manual selection (Domhan et al. 2015). Hence, a modified hybrid DSAE model was proposed to improve the standard autoencoder model and provide a more robust IFD system that can be used in various applications. Resilient backpropagation (Rprop) has been implemented in the DSAE model to improve its training performance. The Rprop model can eliminate the magnitude of the gradient to update the weight of the connection in the DSAE network (Prasad et al. 2013). The activation function hyperparameter is also extremely important for constructing a deep learning network. The sigmoid function has been used for several years since the development of the neural network. In the last few years, the rectified linear unit (RELU) activation function has been proposed to replace the sigmoid function. The sigmoid function has a major issue with vanishing gradients and may

cause the output to become saturated. Hence, RELU was developed to overcome this issue, as the RELU function can reduce the gradient vanishing problem. However, RELU still has problems with the activation “blowing up” or “dying”. The activation function is still in the process of development. In current practice, the activation function is set similarly for all hidden layers. In this research, an activation function combination is proposed in which each hidden layer of the deep network contains a different activation function. Instead of developing a new activation function, both sigmoid and RELU are used to develop the deep network of the sparse autoencoder.

Implementation of grey wolf optimization (GWO) for the DSAE model helps reduce the burden of selecting suitable hyperparameters. Each sparse autoencoder layer has more than four hyperparameters that need to be set (Saufi et al. 2019). Manual hyperparameter tuning is time-intensive and greatly affects the autoencoder’s performance. Therefore, automated hyperparameter selection has been implemented in this study using GWO. In addition, feature extraction is another problem of fault diagnosis analysis because features should be selected according to the machine’s type and condition (Zhang et al. 2020). The proposed model does not involve any feature extraction process; the raw data is directly input into the DSAE network, which can extract and select the important features from the input data (Jing et al. 2017). Most deep learning models use time–frequency images as the input data, requiring advanced signal processing techniques to transform the time-series signal into the time–frequency domain (Xin et al. 2020; Kancharla et al. 2022; Zuo et al. 2022). The proposed model was evaluated with five different machinery datasets: three bearings, one gearbox, and one turbine blade. The rest of this paper is structured as follows: Sect. [Structure of the deep sparse autoencoder](#) introduces the standard architecture of DSAE. Section [Proposed model of machinery fault diagnosis system](#) describes the proposed model using GWO and Rprop. Section [Machinery datasets](#) shows data collected from five different databases, and Sect. [Results and discussion](#) presents the results of the proposed model. Finally, Sect. [Conclusions](#) concludes the paper.

2 Structure of the deep sparse autoencoder

An autoencoder network is developed based on ANN architecture and can learn high-level data representation in an unsupervised manner. The basic structure of an autoencoder contains encoder and decoder functions. The encoder function $h = f(w_1x + b)$ of the autoencoder is used to map the input ($x \in R^n$) into latent space representation ($h \in R^k$) in the form of a compressed feature. This compressed feature is reconstructed using the decoder function $\hat{x} = g(w_2h + b)$ and

produces an output y with similar properties to the input x (Saufi et al. 2018). The network reconstructs the output from the high-dimensional data by minimizing the cost function, E , through the backpropagation algorithm (Praveen et al. 2018). The overall cost function of the sparse autoencoder contains mean square error as the first term and weight regularization as the second term that can be defined as follows:

$$E = \left[\frac{1}{N} \sum_{i=1}^N \left(\frac{1}{2} h_{w,b}(x) - y^2 \right) \right] + \frac{\lambda}{2} \sum_{l=1}^{nl} \sum_i^{sl-1} \sum_j^{sl} \left(W_{ij}^{(l)} \right)^2 \tag{1}$$

where $\frac{1}{N} \sum_{i=1}^N \left(\frac{1}{2} \|h_{w,b}(x) - y\|^2 \right)$ represents mean square error (MSE), λ is the weight decay, nl is the number of a layer in the sparse autoencoder network, sl represents the number of neurons in layer l , and $W_{ji}^{(l)}$ is the weight that connected neuron i in layer $l + 1$ and neuron j in layer l , and $\frac{\lambda}{2} \sum_{l=1}^{nl} \sum_i^{sl-1} \sum_j^{sl} \left(W_{ij}^{(l)} \right)^2$ is the weight regularization terms. The weight regularization term prevents the autoencoder network from overfitting by adjusting the weight and bias magnitudes (Saufi et al. 2019). Meanwhile, the overall cost function of sparse autoencoder, E_{sparse} imposes an additional constraint term onto the hidden units of the autoencoder to increase the sparsity of the hidden unit in the autoencoder network (Wang et al. 2018), which can be expressed as follows;

$$E_{sparse} = E + \beta \sum_{j=1}^n KL(\rho \parallel \hat{\rho}_j) \tag{2}$$

where $KL(\rho \parallel \hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1-\rho}{1-\hat{\rho}_j}$ represents the Kulback-Leibler term, β is a weight of sparsity penalty term, ρ is the sparsity parameter and $\hat{\rho}_j$ is the average activation of the hidden unit.

The sparse autoencoder is usually stacked in several layers with a softmax classifier to make a DSAE. The compressed feature is input into a softmax classifier for multi-fault diagnosis purposes. The softmax function is expressed as follows:

$$h_{\theta}(x^i) = \frac{e^{x^i}}{\sum_{i=1}^k e^{x_k}} \tag{3}$$

3 Proposed model of machinery fault diagnosis system

The primary focus of this research is to develop a sparse autoencoder that can deal with any kind of machinery dataset using a raw vibration time-series signal (1D time-series signal). The proposed method used a three-layer sparse autoencoder with the softmax classifier. It is important

to note that during the analysis, the proposed model was directly fed with segmented raw data from the vibration sensor. Hence, the sparse autoencoder model was improved by implementing the Rprop algorithm in this section. In addition, the activation function used in the three layers of the sparse autoencoder combined two different activation functions: sigmoid and RELU. The hyperparameters of the sparse autoencoder (e.g., some hidden nodes, weight decay parameter λ , sparsity parameter ρ , and weight of sparsity penalty term β) were optimized using the GWO method. According to our previous study, three layers of sparse autoencoder can work well for fault diagnosis analysis (Saufi et al. 2018). Significant contributions to this research were using the 1D time-series signal, Rprop, GWO, and the combination of activation functions in the sparse autoencoder model. The flowchart of the proposed model is shown in Fig. 1.

3.1 Grey wolf optimization (GWO)

To solve the hyperparameter selection problem discussed in the preceding section, GWO was utilised for automated hyperparameter selection. The GWO algorithm produced very competitive performance compared to the other well-known optimization algorithms in engineering problems (Mirjalili et al.

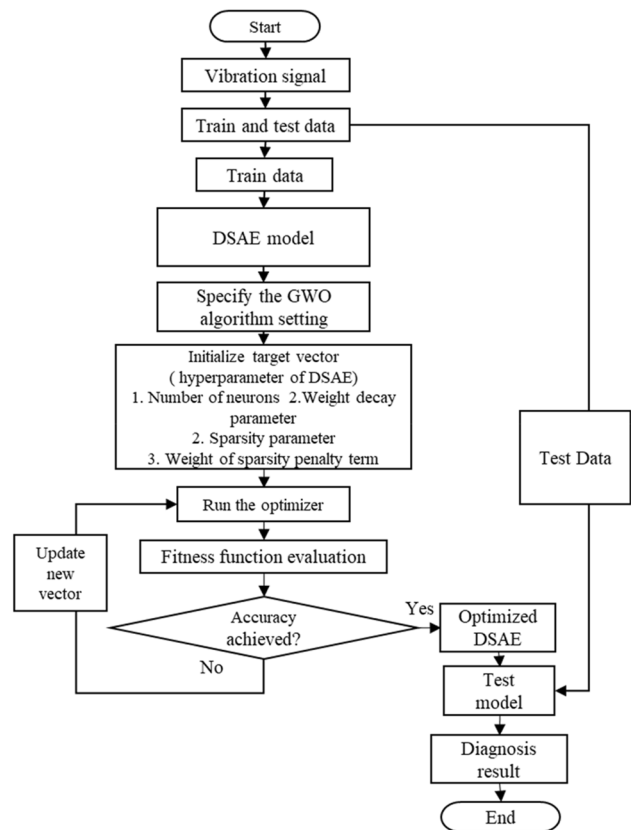


Fig. 1 Flowchart of proposed diagnosis system

2014). Moreover, the GWO method requires less parameter-setting than other optimization algorithms, such as differential evolution, genetic algorithm, and particle swarm optimization. The GWO algorithm mimics the leadership hierarchy of grey wolves and their hunting behaviour. There are four levels of social order in the grey wolf population: alpha α , beta β , delta δ , and omega ω . The wolves' leader belongs to group alpha α ; they are considered the dominant wolf, and the group members must obey their decisions. The other three levels consist of beta β , delta δ , and omega ω wolves. The beta β wolf is the second tier in the wolf hierarchy and assists the alpha in decision making. The delta δ is the third tier, and these wolves act as scouts, sentinels, elders, hunters, and caretakers. The lowest level is omega ω , wherein the wolves serve as scapegoats. As stated above, the GWO algorithm is based on grey wolves' hunting behaviour. When hunting, grey wolves will first encircle their prey, and the prey distance is defined as follows:

$$X_i^d(t+1) = X_i^d(t) - A_i^d \left| C_i^d X_p^d(t) - X_i^d(t) \right| \quad (4)$$

where t is the current iteration number, X_i is the grey wolf's position vector, and X_p is the prey's position vector. The parameters A and C are the coefficient vectors. The A and C values are calculated using the following equations:

$$A = 2a \cdot r_1 - a \quad (5)$$

$$C = 2 \cdot r_2 \quad (6)$$

where vector components linearly decrease from 2 to 0 over the entirety of the iterations, r_1 and r_2 are random vectors between 0 and 1. The grey wolves update their positions based on prey location. The alpha, beta, and delta wolves are assumed to be better able to determine the potential prey's position and to be in a better position to attack the target. The first three best solutions remain while the omegas reposition themselves based on the best search agents. The position update is described in Eqs. (7) and (8).

$$\begin{cases} X_{i,\alpha}^d(t+1) = X_{i,\alpha}^d(t) - A_{i,1}^d \left| C_{i,1}^d X_{i,\alpha}^d(t) - X_i^d(t) \right| \\ X_{i,\beta}^d(t+1) = X_{i,\beta}^d(t) - A_{i,2}^d \left| C_{i,2}^d X_{i,\beta}^d(t) - X_i^d(t) \right| \\ X_{i,\delta}^d(t+1) = X_{i,\delta}^d(t) - A_{i,3}^d \left| C_{i,3}^d X_{i,\delta}^d(t) - X_i^d(t) \right| \end{cases} \quad (7)$$

$$X_i^d(t+1) = \frac{X_{i,\alpha}^d(t+1) + X_{i,\beta}^d(t+1) + X_{i,\delta}^d(t+1)}{3} \quad (8)$$

4 Machinery datasets

In this study, five datasets were used to evaluate the proposed model. Three datasets were obtained through the data-acoustics.com database, including the MFPT bearing dataset, high-speed gearbox dataset, and UNSW turbine blade dataset. The other two bearing datasets were obtained from the CWRU bearing database and MaFaulDa. The bearing datasets from CWRU and MaFaulDa each contained three types of fault conditions and one normal condition at an operating speed of 1800 rpm. Both datasets were divided into four classes for fault diagnosis using the proposed model during the analysis. The bearing dataset from MFPT contains two fault conditions and one normal condition at an operating speed of 1500 rpm, and the data is distributed to three classes. The MFPT dataset contains different sampling rates for normal and fault samples. The gearbox dataset from a high-speed turbine comprises two types: normal and fault conditions. During the data acquisition process, the 3-MW wind turbine gearbox was driven at 1800 rpm. Further, the turbine blade dataset from UNSW contains three conditions: a normal blade, a normal blade with an air jet, and a blade fault. The experimental turbine rig has 19 blade arrangements. Detailed descriptions of the five datasets are shown in Table 1, and Fig. 2 presents the vibration signals of the five datasets. All time-series signals were divided into one-cycle segments, and the segmented signals were sampled for training and testing, as shown in Table 1.

4.1 Preliminary analysis of hybrid activation functions

This section conducted a preliminary analysis of the activation function combinations. The CWRU dataset was used for the analysis. Since there are three layers of sparse autoencoder, several combinations of the sigmoid (Sig) and RELU functions can be obtained, as shown in Figs. 3 and 4. Figure 3 represents the proposed model's diagnosis performance (percentage error), and Fig. 4 illustrates the time required for the model to complete the training process. Seven combinations of sigmoid and RELU can be developed on three layers of the sparse autoencoder. As illustrated in Fig. 3, no significant difference was observed in the diagnosis performance when all layers of the sparse autoencoder were set with only Sig or RELU functions. The model produced 29% and 26.1% diagnosis errors with Sig and RELU, respectively. According to the results, three combinations had low diagnosis errors: RELU-Sig-Sig, RELU-RELU-Sig, and RELU-Sig-RELU. RELU-RELU-Sig produced just a 1.7% error but required 13.714 s to

Table 1 Details of machinery datasets

Dataset	Description	Number of Class	Details	Data distribution
Bearing	Experimental rig from CWRU	Class 1: Normal, Class 2: Outer race, Class 3: rolling element Class 4: inner race	Sampling rate: 48 kHz Record length: 10 s Operating speed: 1800 rpm	Train data: 250 data samples for each fault condition Test data: 100 data samples for each fault condition
	MFPT Experimental rig	Class 1: Normal Class 2: Outer race Class 3: inner race	Sampling rate: Baseline data: 97 kHz Record length: 6 s Fault data: 48 kHz Record length: 3 s Operating speed: 1500 rpm	
	Machinery Fault Database (MaFaulDa)	Class 1: Normal, Class 2: Outer race, Class 3: rolling element Class 4: Cage	Sampling rate: 51.2 kHz Record length: 5 s Operating speed: 1800 rpm	
Gearbox dataset	High-speed turbine	Class 1: Normal Class 2: Fault	Sampling rate: 97 kHz Record length: 6 s Operating speed: 1800 rpm	
Turbine blade dataset	UNSW turbine test rig	Class 1: Normal Class 2: Normal with air jet Class 3: Blade fault	Sampling rate: 63 kHz Record length: 10 s Operating speed: 2000 rpm	

complete the training process, as shown in Fig. 4. The RELU-Sig-RELU combination had a 2.7% error with 6.794 s of training time. There was no significant difference in the diagnosis error between RELU-RELU-Sig and RELU-Sig-RELU, but there was a substantial difference in their training times. For further analysis, the RELU-Sig-RELU combination was used on five machinery datasets. During the analysis, the diagnosis error of the proposed model could be further reduced when the model was optimized using the GWO method. The results demonstrated that the combination of activation functions greatly influences diagnosis performance and training time.

5 Results and discussion

Five datasets from three different mechanical components were used to verify the proposed model. Based on the 1D time series in Fig. 1, it is difficult to distinguish the fault condition by observing the time-series signals. Therefore, the 1D time-series data was input into the proposed model without additional signal processing, manual feature extraction, or feature selection steps. The analysis started with the pre-selection of the number of hidden layers. The hidden layers of the DSAE network were predetermined using several datasets, and according to the analysis, three was a suitable number.

The GWO algorithm optimized 12 hyperparameters of the proposed model's network, with four hyperparameters for each sparse autoencoder layer. Any metaheuristic optimization model requires a set of ranges for the hyperparameter

of interest. The range for hidden node size was set in decrement order from layer to layer of the proposed model's network to reduce the feature dimension from the raw signals. In addition, the ranges of the weight decay parameter, sparsity parameter, and weight of sparsity penalty term are determined by referring to ref (Lee 2017). Figure 5 illustrates the result of the GWO optimization process during the proposed model's hyperparameter selection. The GWO algorithm required fewer than 30 fitness evaluations for the MFPT bearing, UNSW turbine blade, and gearbox datasets to reach the lowest objective value for diagnosis error, which is 0%. The GWO models achieved 0% error on the CWRU dataset after 47 fitness evaluations but required 103 fitness evaluations to reach the 0% objective value on the MaFaulDa dataset. This analysis indicates that the GWO algorithm can optimize the DSAE's hyperparameters with approximately 100 fitness evaluations. The training performance of the proposed model when the networks achieved a correct configuration of hyperparameters is shown in Fig. 6, which contains the training plots for the five datasets. Based on the training plot, the convergence speed of the proposed model is the fastest when the model is trained with a turbine blade dataset, as the model requires less than 50 epochs to achieve 100% training accuracy. By contrast, the proposed model required more than 150 epochs to reach 100% training accuracy when using the MaFaulDa dataset. The training process took less than 70 epochs when the proposed model was trained with the MFPT, CWRU, and gearbox datasets. The overall analysis shows that the proposed model converges to 100% training accuracy in less than 200 epochs. The network configuration must be optimized to prevent the

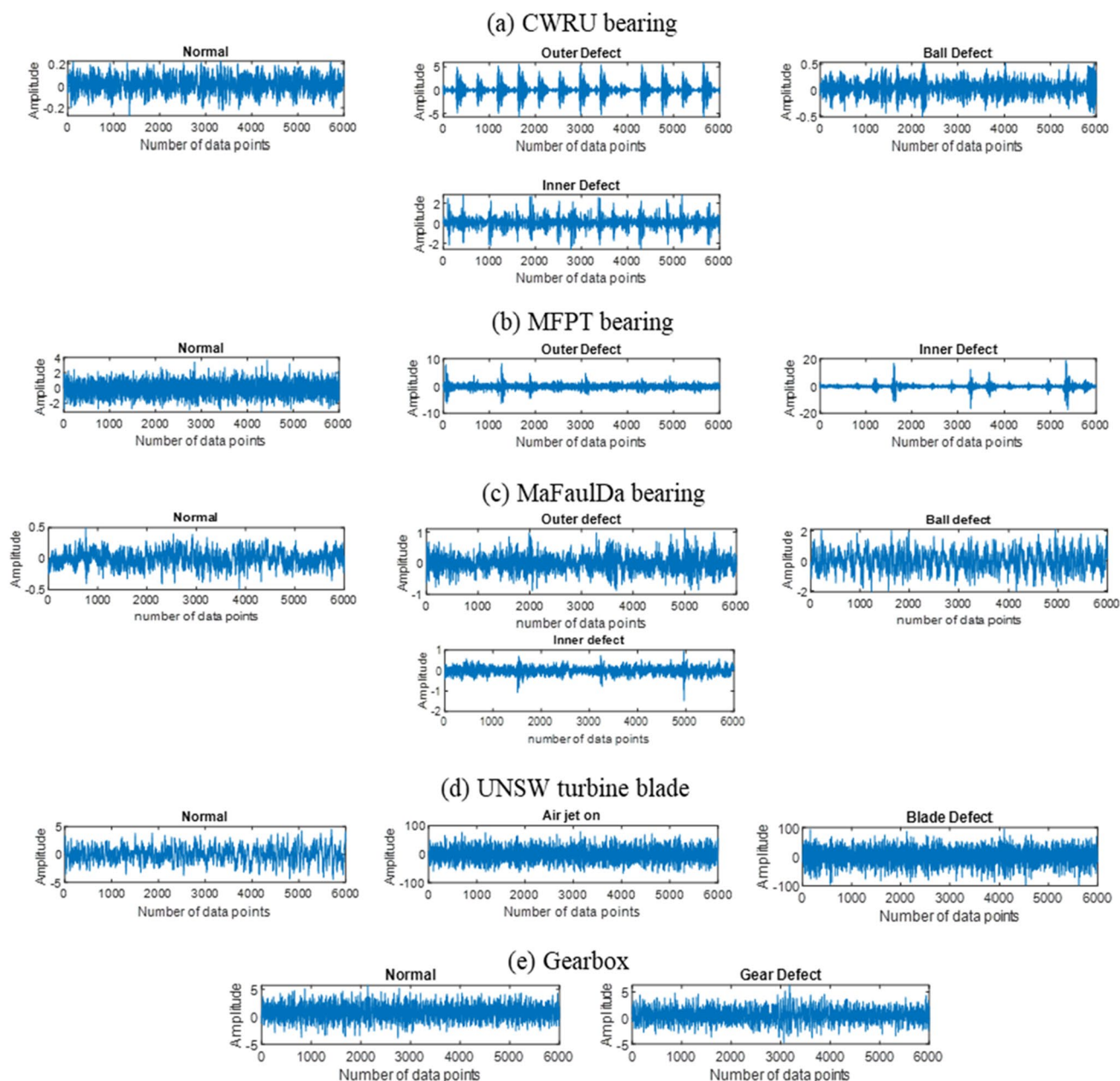


Fig. 2 Vibration time-series signals: **a** CWRU bearing, **b** MaFaulDa bearing, **c** MFPT bearing, **d** UNSW turbine blade, and **e** gearbox

proposed model from overfitting, a common problem with the deep learning model. In addition, the high-dimensional feature learned by the proposed model was visualized via t-distributed stochastic neighbour embedding (t-SNE). The visualization further validates how well the model clusters the input features. The visualization of the t-SNE scatter plot is shown in Fig. 7. It is worth noticing that the features extracted by the proposed model from the 1D time-series signal can be clearly distinguished from one class to another. The results in Fig. 7 demonstrate that the proposed model effectively clustered the training data for all datasets. The

t-SNE visualization indicated that the proposed model was able to mine the fault characteristics and discover discriminative information. The nonlinear function utilized in the proposed model can encode the input data in each sparse autoencoder layer, which helps the softmax classifier to classify the machinery conditions accurately.

Once the proposed model's network achieved an optimal configuration during the training process, the network was tested with another dataset that contained 100 data samples for each fault condition; the test result is illustrated in Fig. 8. The result reflects the effectiveness of the model, which is

Fig. 3 Diagnosis error on a combination of Sig and RELU activation functions on five trials with average error

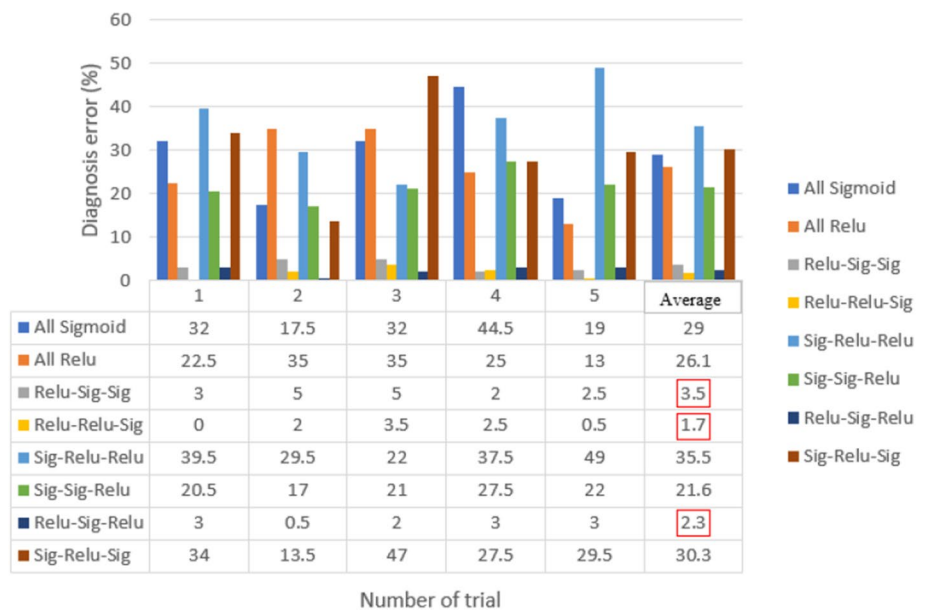
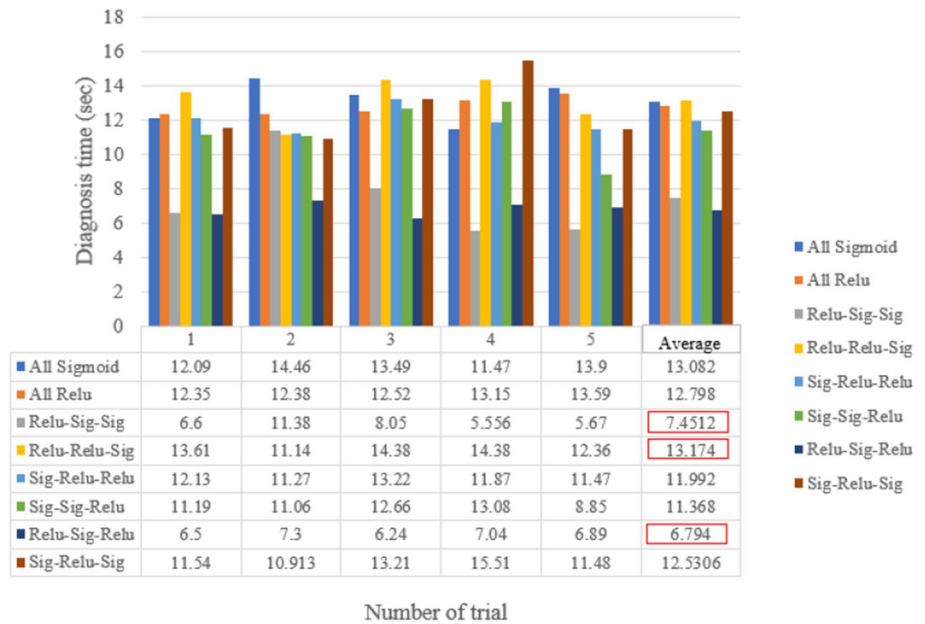


Fig. 4 Time required for the network to complete training on a combination of Sig and RELU activation functions on five trials with average time



equal to the average percentage of samples correctly classified in each class. The proposed model achieved 100% classification accuracy on the CWRU bearing, MFPT bearing, gearbox, and UNSW turbine blade datasets and 95% classification accuracy on the test dataset from MaFaulDa. The proposed model accuracy on the MaFaulDa dataset reached 82% in class 2 (outer race defect) and 98% in class 3 (ball defect). The class 2 error rate, in particular, contributed to a sizeable decline in the overall classification performance of the proposed model on the MaFaulDa dataset. The long training process required to analyse the MaFaulDa data is more evidence of the difficulty of analysing this dataset.

The t-SNE visualization in Fig. 7(c) proved that the high-dimensional features data were so close to each other that the proposed model had difficulty achieving 100% classification accuracy. In addition, GWO algorithms required the highest fitness evaluation to search the hyperparameters of the proposed model when training with the MaFaulDa dataset, as shown in Fig. 6. The MaFaulDa dataset presented a difficulty for the proposed model to classify the time-domain signal based on its fault condition. Nevertheless, the overall classification performance of the proposed model on all datasets achieved a satisfactory result, proving that the model can effectively learn the features from the raw signal

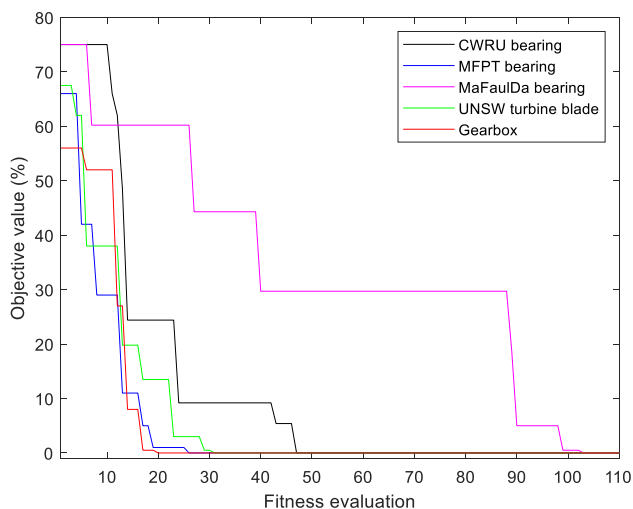


Fig. 5 Performance of grey wolf optimizer on DSAE hyperparameter optimization

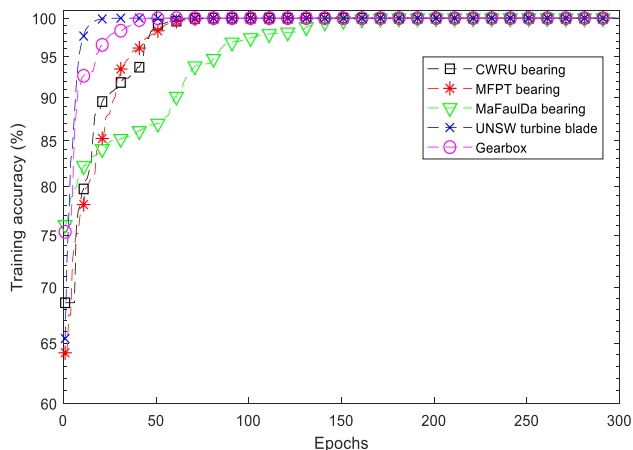


Fig. 6 Training process of the proposed model

even without signal processing, manual feature extraction, or feature selection processes in the early stages. However, the signal processing step is still needed to pre-process the signal if the condition of the signal is too noisy.

The proposed model provided a robust and accurate method of diagnosing bearing, gearbox, and blade faults and was shown to solve the problems of the standard autoencoder model. To further demonstrate the superiority of the proposed model, three deep learning models were employed for comparison: standard autoencoder (AE), deep neural network (DNN), and deep belief network (DBN). All models were tested using the data sample shown in Table 1. The selected deep learning models are unable to perform well when the hyperparameter values are kept the same for all types of datasets. Hence, the configurations for standard AE, DNN, and DBN models were

manually selected until the models achieved satisfactory classification accuracy on the datasets. Table 2 presents the four models' classification accuracy on the testing dataset. The proposed model achieved more than 95% classification accuracy for all datasets and thus performed better than standard AE, DNN, and DBN. On the other hand, the standard AE model achieved the worst classification accuracy among the selected deep learning models. The DBN model's classification accuracy on the CWRU dataset was 100%, which was better than its performance on the other four datasets. The DBN model achieved its lowest classification accuracy, 82.5%, on the MaFaulDa dataset. This shows that the DBN model is capable of processing the CWRU dataset effectively compared to the other four datasets. In addition, DBN's overall performance was slightly better than the DNN model, as the DBN model outperformed DNN on the CWRU, MFPT, and MaFaulDa datasets. The DNN model reached 96% classification accuracy for the gearbox dataset, and its lowest classification accuracy was on the MaFaulDa dataset. Based on the comparative analysis, all models performed poorly when analysing the MaFaulDa dataset due to its low signal-to-noise ratio. We believed that the DNN and DBN models could provide performance competitive with the proposed model in terms of classification accuracy if their structures were well-tuned by experts.

The proposed IFD system was compared with an IFD system that uses traditional machine learning. The statistical feature was manually extracted from the 1D time-series signal to six time-domain features, which were then used to train and test the traditional machine learning models such as SVM, fuzzy classifier, k nearest neighbour (KNN), decision tree, discriminant analysis, and the combination of principal component analysis with SVM. The common time-series features are kurtosis, skewness, crest, shape, impulse, and margin. Table 3 presents the classification accuracy of all models. The classification accuracy on the CWRU, MFPT, gearbox, and UNSW turbine blade datasets when using the selected traditional machine models ranged from 79.0 to 93.5%, much lower than the accuracy achieved by deep learning models. However, all models achieved less than 80% classification accuracy on the MaFaulDa dataset. Based on the analysis, the results produced by using traditional classifiers vary greatly, and the classifier performance depends on the extracted feature. For example, when the six-time domain was directly input into the SVM, the model produced results that ranged from 77 to 89.5% classification accuracy. The performance of SVM increased from 3 to 10% with the implementation of principal component analysis (PCA) on six time-domain features. The PCA-SVM combination produced higher classification accuracy than the other five traditional machine learning models. It is worth

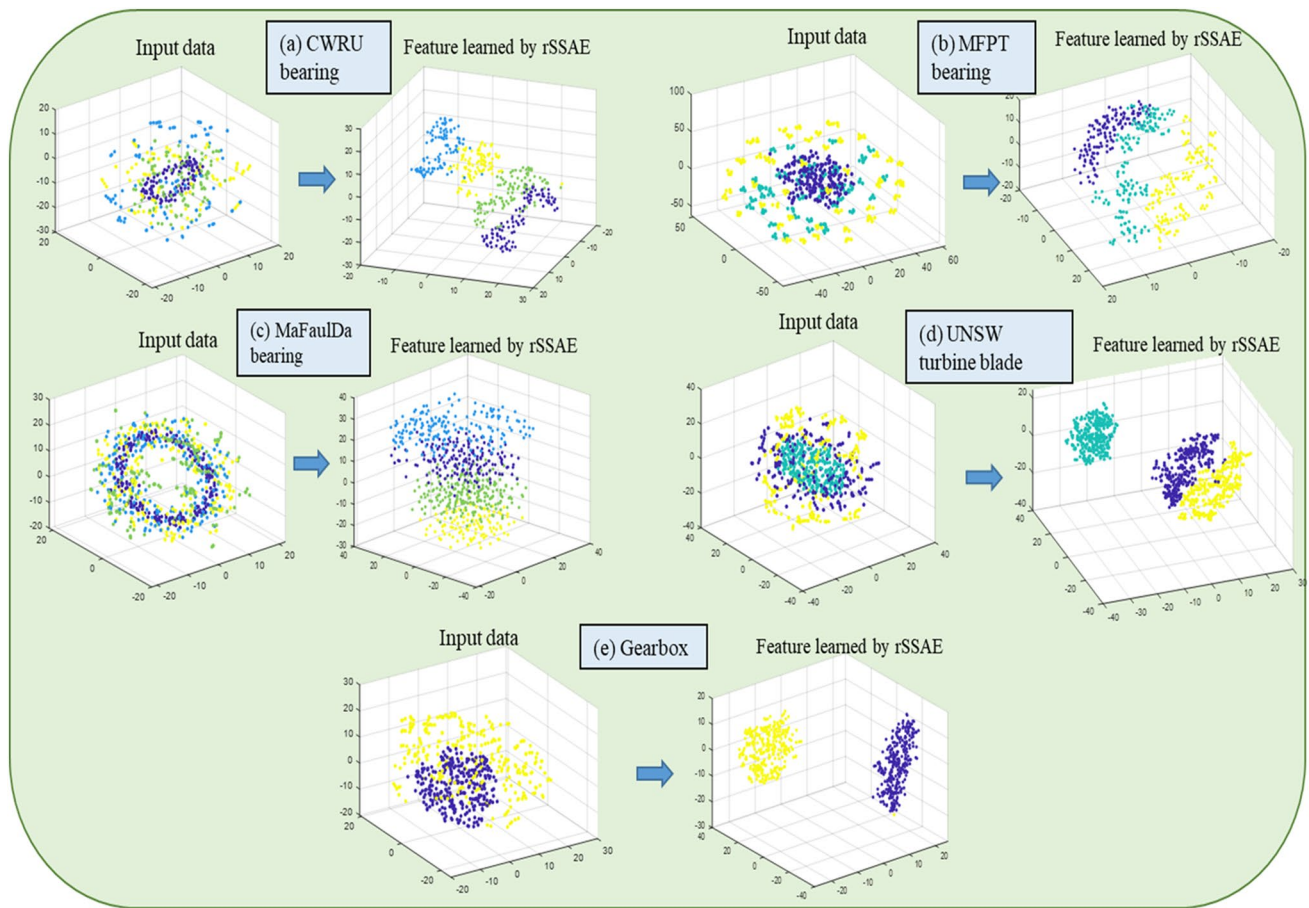


Fig. 7 t-SNE visualization of feature learned by proposed model for every type of signal: **a** CWRU bearing, **b** MFPT bearing, **c** MaFaulDa, **d** UNSW turbine blade, and **e** gearbox

Fig. 8 Classification accuracy of the proposed model using the test dataset

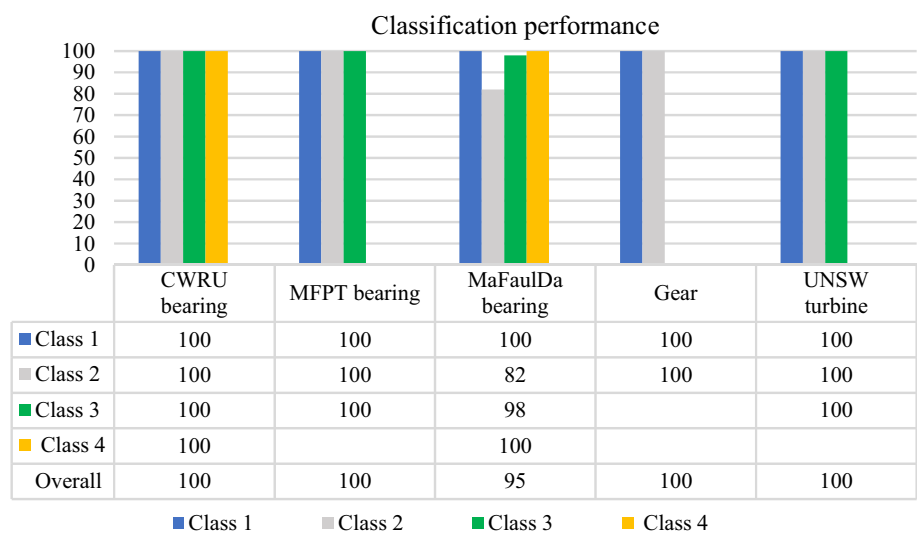


Table 2 Accuracy classification rates of different deep learning models

Methods	Proposed model (%)	Standard AE (%)	DNN (%)	DBN (%)
Classification accuracy (%)				
CWRU	100.0	82.5	95.0	100.0
MFPT	100.0	79.6	93.5	93.8
MaFaulDa	95.0	73.5	77.3	82.5
Gearbox data	100.0	87.0	96.0	90.0
Turbine blade	100.0	74.0	91.8	87.7

Table 3 Classification accuracy using traditional fault diagnosis

	CWRU (%)	MFPT (%)	MaFaulda (%)	Gear-box (%)	UNSW Turbine Blade (%)
SVM	89.17	88.7	77.0	89.5	83.0
Fuzzy classifier	79.8	79.3	61.0	92.0	81.3
KNN	79.5	86.0	70.3	87.5	79.0
Decision Tree	86.94	87.3	69.3	89.5	80.0
Discriminant Analysis	82.0	86.0	71.5	90.0	74.0
PCA-SVM	92.10	90.7	78.3	93.5	93.0

Table 4 List of related studies that used the same datasets as in this manuscript

Authors	Method	Dataset	Diagnosis accuracy
(Wang et al. 2022)	Convolutional Neural Networks–Granular Computing (CNN-GC) and extracted features	CWRU	99.8%
(Kancharla et al. 2022)	Fast Fourier Transform, Multi-Layer Convolutional Auto-Encoder (MLCAE), and K-Nearest Neighbors (KNN) classifier	CWRU	100.0%
(Zuo et al. 2022)	Probabilistic spiking response model (PSRM), spiking neural network (SNN), and Local mean decomposition (LMD)	CWRU MFPT	99.38% 99.72%
(Verstraete et al. 2017)	Deep convolutional neural network with time–frequency images	MFPT	98.1%
(Saufi et al. 2018)	Modified Autoencoder with wavelet time–frequency image	Mafulda	99.0%
(Isham et al. 2018)	Improved variational mode decomposition with the neural network using selected statistical features	Gear	95.8%
None	No classification diagnosis analysis on this data	UNSW	None

reporting that traditional machine learning could deliver a competitive performance with deep learning models in terms of training time if the feature is appropriately designed.

5.1 Comparative study with related literature analysis

The performance of the proposed model is compared with related studies as shown in Table 4. All of the types of machinery data that are used in this manuscript are available online. However, most of the machinery fault diagnosis studies use the CWRU dataset. According to the literature, limited research has been conducted that used the MFPT, MaFaulDa, and high-speed turbine gearbox datasets. No study has been done on the UNSW dataset. The model

proposed by Wang et al. (2022), Kancharla et al. (2022), and Zuo et al. (2022) was able to attain high diagnosis accuracy on the CWRU dataset; however, the method they used required the integration of two deep learning models with signal processing. In contrast, the proposed model requires proper tuning of hyperparameters to achieve an accurate diagnosis. For the MFPT and MaFaulDa datasets, the performance of the proposed model is competitive with a model in ref (Verstraete et al. 2017). However, this study used the time–frequency image as input data to the deep learning model. It is important to note that it is more difficult to classify a raw vibration signal than a time–frequency image, as the raw signal contains unwanted noise that may bury the valuable information in the signal. Meanwhile, the analysis of the gearbox dataset uses a shallow learning model with modified variational mode decomposition (Isham et al.

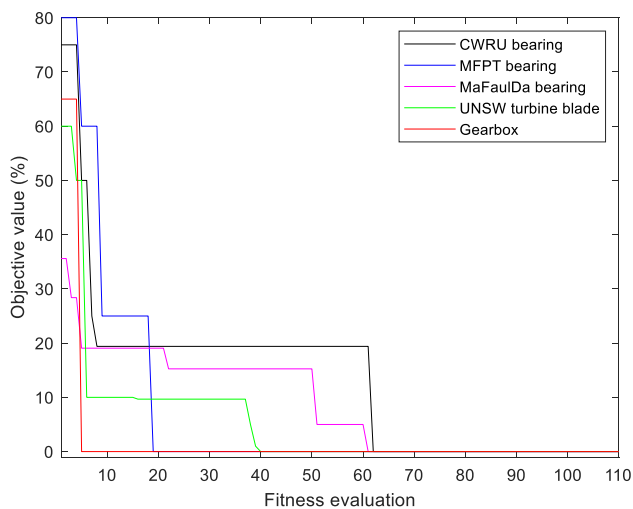


Fig. 9 Performance of GTO on DSAE hyperparameter optimization

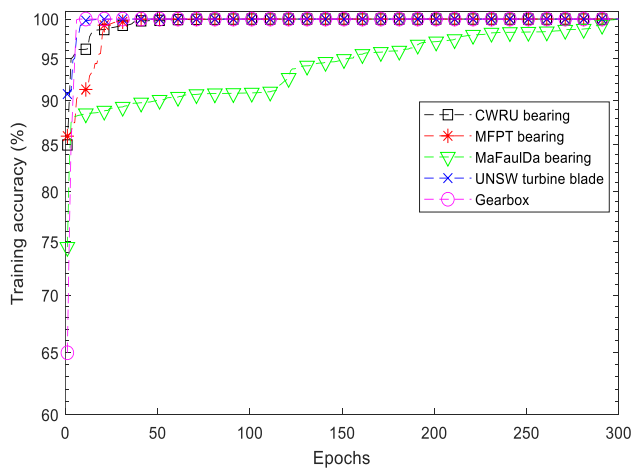


Fig. 10 Training process of the proposed model with hyperparameters optimized by GTO

2018). Finally, the analysis of the UNSW dataset is still at the signal processing stage; hence no comparative study can be done. In short, the proposed model worked well on five types of machinery datasets, including three bearings, one gearbox, and one turbine blade. Even though the proposed model performs similarly to some related study models, it can diagnose different machinery components without changing the main architecture. This proves that hyperparameter tuning is essential in the deep learning model.

5.2 Comparative study with artificial gorilla troops optimizer

Artificial gorilla troops optimizer (GTO) is a new optimization algorithm (Abdollahzadeh et al. 2021). In this paper, the

GTO algorithm was used to optimize the proposed model, and the optimization result is shown in Fig. 9. It was demonstrated that the GTO algorithm could optimize the DSAE's hyperparameters, as the DSAE model reached the 0% objective value for all datasets with different fitness evaluations. The algorithm required more than 60 fitness evaluations for the CWRU dataset and fewer than ten for the gearbox dataset. As illustrated in the training plot in Fig. 10, the DSAE model can reach 100% training accuracy for all datasets. However, DSAE required more than 250 epochs to achieve 100% training accuracy on the MaFaulDa dataset, showing that the hyperparameters optimized by GWO are better than GTO. The training performance of the DSAE model on the other four datasets showed similar performance when optimized with GWO as with GTO.

6 Conclusions

This paper proposed an improved sparse autoencoder model by implementing a Rprop learning algorithm on the three layers of the DSAE network and optimizing the network hyperparameters using the GWO algorithm. This paper proposed a unique flexible autoencoder to classify the 1D time-series signal from three machinery components. The proposed model achieved 100 per cent diagnosis accuracy on four datasets (MFPT, MaFaulDa, gearbox, turbine blade) and above 95% accuracy on the CWRU dataset. The proposed model outperformed three deep learning models and five traditional machine learning models on all datasets. The proposed model shows superior accuracy and offers several benefits: (i) it does not require any signal processing method; (ii) it is capable of analysing raw vibration time-domain signals, and thus no feature extraction and feature selection steps are required; (iii) the structure is unique and user-friendly because all the hyperparameters are optimized with the GTO, and the network is trained using the Rprop algorithm and iv) the mix of activation functions is capable of increasing diagnostic accuracy and reducing training time. Thus, our future work will extend the proposed model's use to other applications.

Acknowledgements This research was supported by Universiti Teknologi Malaysia through UTM fundamental research (UTMFR) grant Q.J130000.3851.22H06.

References

- Abdollahzadeh B, Soleimani Gharehchopogh F, Mirjalili S (2021) Artificial gorilla troops optimizer: a new nature-inspired

- metaheuristic algorithm for global optimization problems. *Int J Intell Syst* 36:5887–5958. <https://doi.org/10.1002/int.22535>
- Ayas S, Ayas MS (2022) A novel bearing fault diagnosis method using deep residual learning network. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-021-11617-1>
- Bengio Y, Lamblin P, Popovici D, Larochelle H (2007) Greedy layer-wise training of deep networks. *Adv Neural Inf Process Syst* 19:153
- Chen CC, Liu Z, Yang G et al (2021) An improved fault diagnosis using 1d-convolutional neural network model. *Electron* 10:1–19. <https://doi.org/10.3390/electronics10010059>
- Domhan T, Springenberg JT, Hutter F (2015) Speeding Up Automatic Hyperparameter Optimization of Deep Neural Networks by Extrapolation of Learning Curves. In: *International Joint Conference on Artificial Intelligence (IJCAI 2015)*. pp 3460–3468
- Guo X, Shen C, Chen L (2016) Deep Fault Recognizer: an integrated model to denoise and extract features for fault diagnosis in rotating machinery. *Appl Sci* 7:41. <https://doi.org/10.3390/app7010041>
- Gupta S, Gupta MK (2021) A comparative analysis of deep learning approaches for predicting breast cancer survivability. *Arch Comput Methods Eng*. <https://doi.org/10.1007/s11831-021-09679-3>
- Habbouche H, Amirat Y, Benkedjough T, Benbouzid M (2021) Bearing fault event-triggered diagnosis using a variational mode decomposition-based machine learning approach. *IEEE Trans Energy Convers* 8969:1–9. <https://doi.org/10.1109/TEC.2021.3085909>
- Hinton GE, Osindero S, Teh Y-W (2006) A fast learning algorithm for deep belief nets. *Neural Comput* 18:1527–1554. <https://doi.org/10.1162/neco.2006.18.7.1527>
- Isham MF, Leong MS, Lim MH, Ahmad ZA (2018) Variational mode decomposition for rotating machinery condition monitoring using vibration signals. *Trans Nanjing Univ Aero Astro*. 35:38–50
- 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* 111:1–10. <https://doi.org/10.1016/j.measurement.2017.07.017>
- Kancharla CR, Vankeirsbilck J, Vanoost D et al (2022) Latent dimensions of auto-encoder as robust features for inter-conditional bearing fault diagnosis. *Appl Sci*. <https://doi.org/10.3390/app12030965>
- Lee J (2017) A Performance Comparison of Auto-Encoder and Its Variants for Classification. 207–211
- Mao W, Feng W, Liu Y et al (2021) A new deep auto-encoder method with fusing discriminant information for bearing fault diagnosis. *Mech Syst Signal Process* 150:107233. <https://doi.org/10.1016/j.ymssp.2020.107233>
- Mirjalili S, Mirjalili SM, Lewis A (2014) Grey Wolf Optimizer. *Adv Eng Softw* 69:46–61. <https://doi.org/10.1016/j.advengsoft.2013.12.007>
- Prasad N, Singh R, Lal SP (2013) Comparison of back propagation and resilient propagation algorithm for spam classification. *Proc Int Conf Comput Intell Model Simul*. <https://doi.org/10.1109/CIMSim.2013.14>
- Praveen GB, Agrawal A, Sundaram P, Sardesai S (2018) Ischemic stroke lesion segmentation using stacked sparse autoencoder. *Comput Biol Med* 99:38–52. <https://doi.org/10.1016/j.compbiomed.2018.05.027>
- Qi Y, Shen C, Wang D et al (2017) Stacked sparse autoencoder-based deep network for fault diagnosis of rotating machinery. *IEEE Access* 5:15066–15079. <https://doi.org/10.1109/ACCESS.2017.2728010>
- Saufi SR, Ahmad ZA, Leong MS, Lim MH (2018) Differential evolution optimization for resilient stacked sparse autoencoder and its applications on bearing fault diagnosis. *Meas Sci Technol* 29:125002. <https://doi.org/10.1088/1361-6501/aae5b2>
- Saufi SR, Ahmad ZA, Leong MS, Lim MH (2019) Challenges and opportunities of deep learning models for machinery fault detection and diagnosis : a review. *IEEE Access* 7:122644–122662. <https://doi.org/10.1109/ACCESS.2019.2938227>
- Sun Y, Li S (2022) Bearing fault diagnosis based on optimal convolution neural network. *Meas J Int Meas Confed* 190:110702. <https://doi.org/10.1016/j.measurement.2022.110702>
- Verstraete D, Ferrada A, Droguett EL et al (2017) Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings. *Hindawi Shock Vib* 2017:1–29. <https://doi.org/10.1155/2017/5067651>
- Wang Y, Liu M, Bao Z, Zhang S (2018) Stacked sparse autoencoder with PCA and SVM for data-based line trip fault diagnosis in power systems. *Neural Comput Appl* 5:1–13. <https://doi.org/10.1007/s00521-018-3490-5>
- Wang X, Mao D, Li X (2021) Bearing fault diagnosis based on vibro-acoustic data fusion and 1D-CNN network. *Meas J Int Meas Confed* 173:108518. <https://doi.org/10.1016/j.measurement.2020.108518>
- Wang X, Yang J, Lu W (2022) Bearing fault diagnosis algorithm based on granular computing. *Granul Comput*. <https://doi.org/10.1007/s41066-022-00328-z>
- Xin Y, Li S, Wang J et al (2020) Intelligent fault diagnosis method for rotating machinery based on vibration signal analysis and hybrid multi-object deep CNN. *IET Sci Meas Technol* 14:407–415. <https://doi.org/10.1049/iet-smt.2018.5672>
- Xu J, Xiang L, Hang R, Wu J (2014) Stacked Sparse Autoencoder (SSAE) Based Framework for Nuclei Patch Classification on Breast Cancer Histopathology. 999–1002
- Yan J, Kan J, Luo H (2022) Rolling bearing fault diagnosis based on Markov transition field and residual network. *Sensors* 22:3936. <https://doi.org/10.3390/s22103936>
- Zhang J, Sun Y, Guo L et al (2020) A new bearing fault diagnosis method based on modified convolutional neural networks. *Chin J Aeronaut* 33:439–447. <https://doi.org/10.1016/j.cja.2019.07.011>
- Zuo L, Xu F, Zhang C et al (2022) A multi-layer spiking neural network-based approach to bearing fault diagnosis. *Reliab Eng Syst Saf* 225:108561. <https://doi.org/10.1016/j.ress.2022.108561>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.