Bearing fault diagnosis based on cross-machine statistical features generalization and improved extreme learning machine

Muhammad Harith Mohd Kamal[1*] , Muhammad Firdaus Bin Isham[1] ,
Amirulaminnur Raheimi[2] , Mohd Syahril Ramadhan Mohd Saufi[1],
Wan Aliff Abdul Saad[1]

[1] Faculty of Mechanical Engineering, Universiti Teknologi Malaysia, Johor, Malaysia
[2] Institute of Noise and Vibration, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia
* Corresponding author's e-mail: mdharithkamal@gmail.com

ABSTRACT

Asset reliability is among the primary objectives in technological advancements and effective maintenance is essential to guarantee optimal performance of machineries while upholding safety requirements. Intelligent models based on machine learning and deep learning techniques have been extensively suggested for advanced maintenance procedures. In recent times, there has been a trend in fault diagnosis studies towards cross-machine diagnosis which involves multiple machines. Therefore, this paper proposes a cross-machine bearing fault diagnosis trained without faulty data of target machine; based on selected generalized statistical vibration features and improved extreme learning machine. This work utilized an online bearing dataset from a source machine and experimental datasets from a target machine. The statistical vibration features were derived from both datasets (online and experimental) and subsequently chosen based on distinctive characteristics in features. Next, specific characteristics will be input into the improved extreme learning machine (ELM) technique for the purpose of fault categorization. The suggested model demonstrated substantial cross-machine classification ability, with an accuracy rate of up to 98.9%.

Keywords: fault diagnosis, bearing, vibration, cross-machine, eel and grouper optimizer, extreme learning machine.

INTRODUCTION

The reliability and safety of machinery are important to ensure smooth operation in any industries. In various industries, rotating machinery equipment always subjected to operate in harsh working environment, such as high temperatures, unsuitable humidity, poor lubrication, and beyond its designed limits [1]. This will cause a performance deterioration and defects propagates in rotating machinery components such as gear, bearing, and shaft [2, 3]. As reported by Zheng et al., bearing component exhibited the highest failure rate, ranging from 45% to 90% as compared with other components in specific rotating machinery equipment [4]. Condition monitoring and fault diagnosis (CMFD) is mainly used for accessing rotating machinery operation to ensure safe operation and avoiding failure or losses.

In recent literature, there are several CMFD approach had been developed which are vibration-based analysis [5–7], temperature-based analysis [8], pressure-based analysis [9], and oil-based analysis [10]. Vibration-Based Analysis effectively detects changes and early faults in rotating machines by capturing both linear and nonlinear traits. Its adaptability to various operating conditions enhances its monitoring capabilities. In contrast, Temperature-Based Analysis is straightforward but may miss early faults due to reliance on significant temperature variations. Its effectiveness diminishes in environments with stable temperatures. Pressure-Based Analysis provides direct insights into system health, particularly in pressure-sensitive systems, but is limited to specific applications and may fail to detect subtle faults. Oil-Based Analysis assesses wear through

lubricant condition, signalling impurities, yet requires frequent sampling and may miss faults not affecting the lubricant. Vibration-based analysis is preferred and popular tool used for CMFD of rotating machinery equipment. This is because of its simple, high sensitivity, low implementation costs, and capability to deliver the important information about the equipment or machine.

Recently, there are many CMFD approaches had been developed for diagnosing rotating machinery equipment using vibration signals. For an example, Saha et al. introduced a method for intelligent diagnostics of bearings utilizing support vector machines (SVM) and particle swarm optimization (PSO) with the aid of vibration signals [11], Youcef et al. proposed bearing fault diagnosis using vibration spectrum and convolutional neural network (CNN) [12], Shao et al. also proposed bearing fault diagnosis using vibration signal optimization using Harris-hawks optimization and SVM [13], etc.

However, current CMFD approach mainly developed for single machine diagnosis which requires new diagnosis approach to be established when replacing the rotating machinery equipment or machine. This is due to domain shift causing reduced accuracy of the prior CMFD method when applied to the modified equipment. Hence, a demand from the industry for an efficient and reliable fault diagnosis especially on different machine diagnosis, a cross-machine diagnosis studied had been gain a lot of attentions recently [14–16]. Jia et al. and Li et al. proposed cross-machine diagnosis for bearing components (DBDP-Net) [14, 16]. He and Shen proposed cross-machine diagnosis mode for spindle motors [15]. Wang et al. proposed cross-machine diagnosis for bearing component based on semi-supervised adversarial domain adaptation [17]. Lv et al. proposed cross-machine diagnosis using deep learning approach [18]. Zhang et al. also proposed cross-machine diagnosis based on pseudo-label transitive domain adaptation networks [19]. Yuan et al. proposed cross-machine using variational autoencoder and deep learning approach [20]. The studies on cross-machine diagnosis have been continuously conducted by many scholars by proposing a new diagnosis approach that uses diagnosis knowledge from an older machine as a source domain and apply to another machine as a target domain. However, the problem on cross-machine diagnosis study remains open to be explored where the domain difference is due to changed working

condition and mechanical structure characteristic, as agreed by [21].

Therefore, this paper proposes novel cross-machine bearing fault diagnosis based on selected statistical features generalization (SFG) and improved extreme learning machines based on eel and grouper optimization (EGO) method, so called SFG-IELM method. In 2006, Huang et al. pioneered the extreme learning machine (ELM) algorithm [22]. The SFG will be selected from time-domain and frequency-domain statistical features. Then, the generalized factor will be determined based on the relationship between source domain features and target domain features for each operational condition. The ELM method has a capability to provide fast learning speed and excellent generalization performance as compared to the conventional back-propagation algorithm and support vector machine [23–25]. The ELM method has been proposed to solve various tasks in many industries; for instance, biomedical image classification [26], automatic spoken language identification [27], forecasting air quality index [28], and traffic flow forecasting [29]. Due to its limitation, the ELM method had been improved recently using a meta-heuristic algorithm to select an optimized number of its parameters such as number of hidden neurons, input weight, bias, etc [30, 31]. Recently, Ali and Mirjalili had proposed a new meta-heuristic algorithm known as eel and grouper optimizer (EGO) method [32]. The EGO method will be used to improve the ELM method in this paper to select an optimized number of hidden neurons, input weight and bias.

The structure of this paper is as follows: Section 2 provides a concise overview of the ELM method, EGO method, and the proposed method. Section 3 discuss and present the datasets and experimental configuration. Section 4 provides an exposition of the findings and further analysis. The final conclusions are formulated in Section 5.

THEORETICAL BACKGROUND

Extreme learning machine

The ELM algorithm is a highly efficient and simple technique for training single-hidden layer feedforward neural networks (SLFNs) [22]. This method relies on the random initialization of connection weights between the input and hidden layers, as well as the biases for the neurons

in the hidden layer, without any modifications throughout the training phase. Furthermore, it attains an optimal solution by varying the number of hidden layer neurons as the only parameter that needs to be set and only output weights are calculated within the algorithm training. Thus, achieves quicker learning rate and enhanced generalization performance compared to traditional feedforward network training methods [22]. The ELM algorithm addresses the difficulties encountered by gradient-based methods in SLFNs, such as instability and divergence with high learning rates, slow convergence with low learning rates, and the risk of converging to local minima that are far from the global minima, which can result in overfitting and prolonged training times [22].

The ELM algorithm consists of three layers: an input layer, a hidden layer, and an output layer. The input layer is comprised of neurons that correspond to the number of input features, while the hidden layer contains a number of neurons that requires adjustment but normally set equal to the number of input features. The output layer consists of neuron number that represents the output classes number, as illustrated in the topology of the ELM algorithm shown in Figure 1. From the input layer, various features (x_i), input weights (w_{ij}), and biases (b_j) are integrated, where $i \in 1, 2, 3 \dots n$ and n is number of features; $j \in 1, 2, 3 \dots \eta$ and η is number of hidden layer neurons; and ij denotes the number of connections between input neuron i and hidden layer neuron j. These features are then processed in the hidden neuron layer incorporating an activation function, f as according to Equation 1, leading to the formation of a matrix H as shown in Equation 2.

$$h_j = (w_{ij}x_i + b_j) \times f \tag{1}$$

$$H = \begin{bmatrix} (w_{11}x_1 + b_1) \times f & \cdots & (w_{n1}x_n + b_1) \times f \\ \vdots & \ddots & \vdots \\ (w_{1\eta}x_1 + b_\eta) \times f & \cdots & (w_{n\eta}x_n + b_\eta) \times f \end{bmatrix} \tag{2}$$

The output matrix can subsequently be represented as shown in Equation 3 and Equation 4, including an output layer (y_k) and an output weight (β_{jk}), where $k \in 1, 2, 3 \dots N$ and N is number of output classes; while jk denotes the number of connections between hidden layer neuron j to output neuron k.

$$y = H\beta \tag{3}$$

$$y_k = \sum \beta_{jk} f[w_{ij}x_i + b_k] \tag{4}$$

Eel and grouper optimizer

EGO, a novel meta-heuristic optimization method, draws inspiration from the collaborative hunting behaviour of moray eels and grouper fish. Groupers hunt in the open waters around coral reefs, while moray eels pursue their prey within the reefs at night. When these two species work together, the chances of their prey escaping from both the grouper in open water and the moray eel in the reefs are significantly reduced [33].

The mutualistic relationship observed during hunting typically initiates when the groupers' attempts to hunt in open water fail, prompting their prey to seek refuge within the reefs. In response, the groupers swim towards nearby eels, signalling and guiding them to the location of the concealed prey. The eels then navigate the reef's crevices to hunt for the prey while the groupers remain near
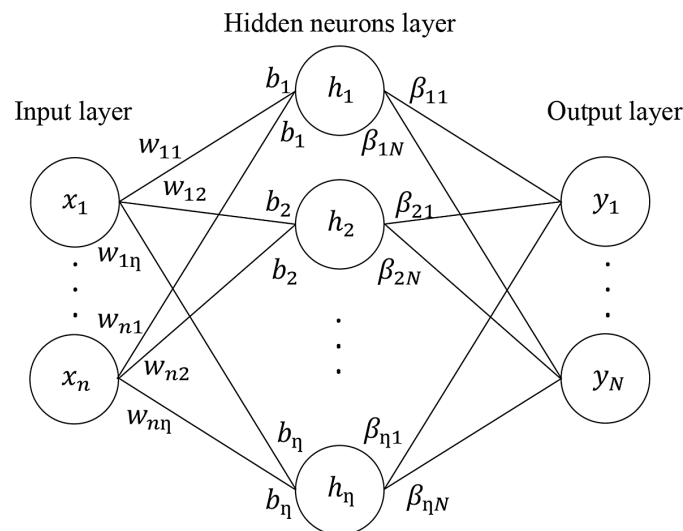

Figure 1. ELM topology

the reef in case the prey escapes back into open water. Should one of the predatory species successfully consume the prey, the other species does not retaliate, suggesting that this non-aggression is part of their cooperative arrangement.

In phase 1, a random search agent is employed to adjust the location of a search agent during the exploration phase. This mechanism, along with specific coefficient values, improves exploration and allows the EGO algorithm to conduct a global search. The mathematical model is outlined as follows in Equation 5 and 6 respectively, where t is the current iteration in i^{th} dimension, X_{rand} represents the random position vector, XG is the best solution from previous iteration, C_1 and C_2 represent the coefficient vector with controlled random values which is more than 1 and less than -1 to ensure exploration, and X_i is a current location vector. Moreover, the fitness is evaluated based on objective function that will be described in methodology section.

$$\vec{X}_i^{t+1} = \vec{X}_{rand} + \vec{C}_1 \cdot |\vec{X}_i^t - \vec{C}_2 \cdot \vec{X}_{rand}| \quad (5)$$

$$\overrightarrow{XG}_i^{t+1} = \vec{X}_i^{t+1}, if\ fitness\ (\vec{X}_i^{t+1}) > \overrightarrow{XG}_i^t \quad (6)$$

XG should changes in each iteration representing the grouper position hunting movement. As iteration increases, C_1 and C_2 is updated using Equation 7, 8 and 9, where r_1 and r_2 are random number between 0 and 1. Equation 9 and 10 ensures shrinking encircling behaviour as α value decreases as iteration increases.

$$\vec{C}_1 = 2a \times r_1 - a \quad (7)$$

$$\vec{C}_2 = 2r_2 \quad (8)$$

$$a = 2 - 2 \times (t/Max\_iteration) \quad (9)$$

$$r_3 = (a - 2) \times r_2 + 2 \quad (10)$$

$$r_4 = 100 \times rand \quad (11)$$

Grouper then signals nearby eel based on starvation rate that increases with iteration as described in Equation 12. Thus, initializes eel position using Equation 13 that will then join the hunt signifying the start of phase 2 of EGO with exploitation algorithm depending on relationship between Equation 11 and 12.

$$Starvation_{rate} = \\ = 100 \times (t/Max\_iteration) \quad (12)$$

$$\overline{XE}_i^t = C_2 \cdot \overrightarrow{XG}_i^t, if\ r_4 \leq starvation\_rate \quad (13)$$

Initially, XP is initialized based on best particle according to fitness. Then, it updates according to cooperative exploitation movement of eel (X_1) and exploration movement of grouper

(X_2) as simulated in Equation 14 and 15 respectively. Equation 16 shows the influence of X_1 and X_2 according to randomized value of p (between 0 to 1); thus, ensuring the algorithm to balance between exploration and exploitation. Then, performance of each search agent is evaluated and XP is updated if better solution is present. The operation continues until the maximum iteration set. The position of prey is the solution obtained by EGO algorithm.

$$X_1 = e^{br_3} \cdot \sin(2\pi r_3) \cdot \vec{C}_1 |\overrightarrow{XE}_i^t - \overline{XP}_i^t| + \overrightarrow{XE}_i^t \quad (14)$$

$$X_2 = \overrightarrow{XG}_i^t + \vec{C}_1 |\overrightarrow{XG}_i^t - \overline{XP}_i^t| \quad (15)$$

$$\vec{X}_i^{t+1} = \left(\frac{0.8X_1 + 0.2X_2}{2}\right), if\ p < 0.5 \\ \vec{X}_i^{t+1} = \left(\frac{0.2X_1 + 0.8X_2}{2}\right), if\ p \geq 0.5 \quad (16)$$

In summary, the EGO algorithm effectively enhances both optimal and diverse solutions while avoiding local optima. Additionally, EGO mitigates the risks associated with local optima due to its population-based approach and adaptable transitions between exploration and exploitation phases. A thorough explanation and benchmark performance results of the optimizer can be found in [33] where EGO has been tested with various benchmark functions, spanning unimodal, multimodal, and composite varieties. The study shows that EGO outperforms many prominent algorithms like Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Differential Evolution (DE) when it comes to achieving global optima with remarkable precision and swift convergence. The mathematical model presented is based on single dimension or single variable; thus, the proposed method requires 3 variables for optimization resulting to a combination of 3 dimensions. Figure 2 illustrates the computational model flowchart of the joint hunting activity performed by these two predatory fish.

METHODOLOGY

As presented in Figure 3, the method is comprised of 2 phases. The first phase process flow is as depicted in the left side of Figure 3. The purpose of the first phase of this method is to determine which features are suitable to classify faults and to obtain the generalization factor. Initially, labelled source domain dataset consisting of 3 different bearing conditions is divided into samples according to cycles of revolution as demonstrated
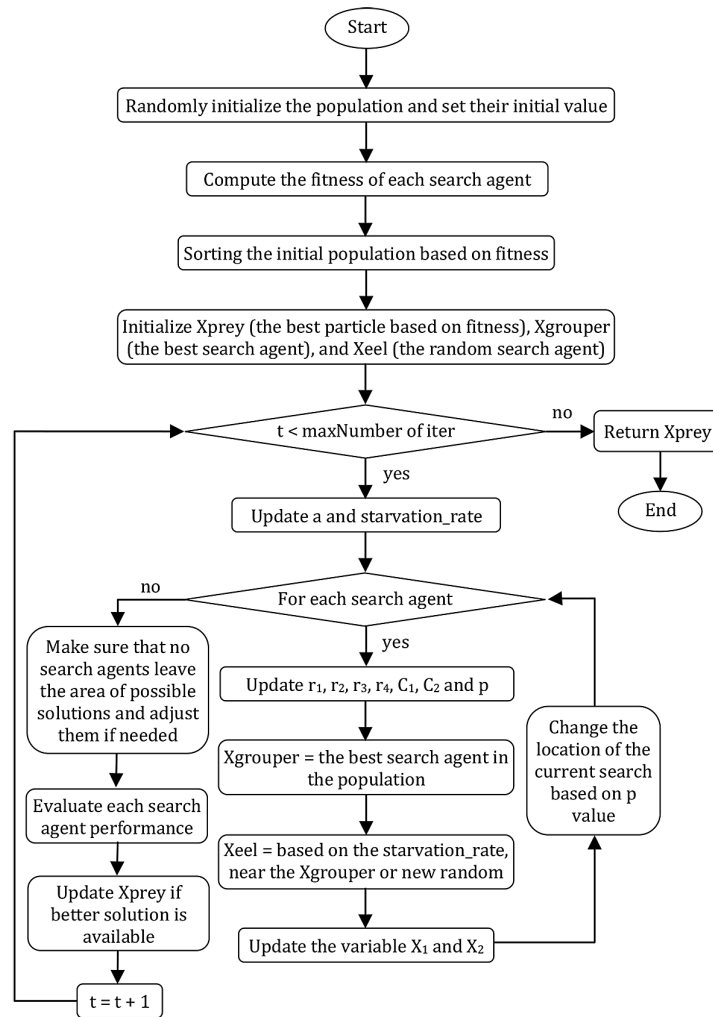
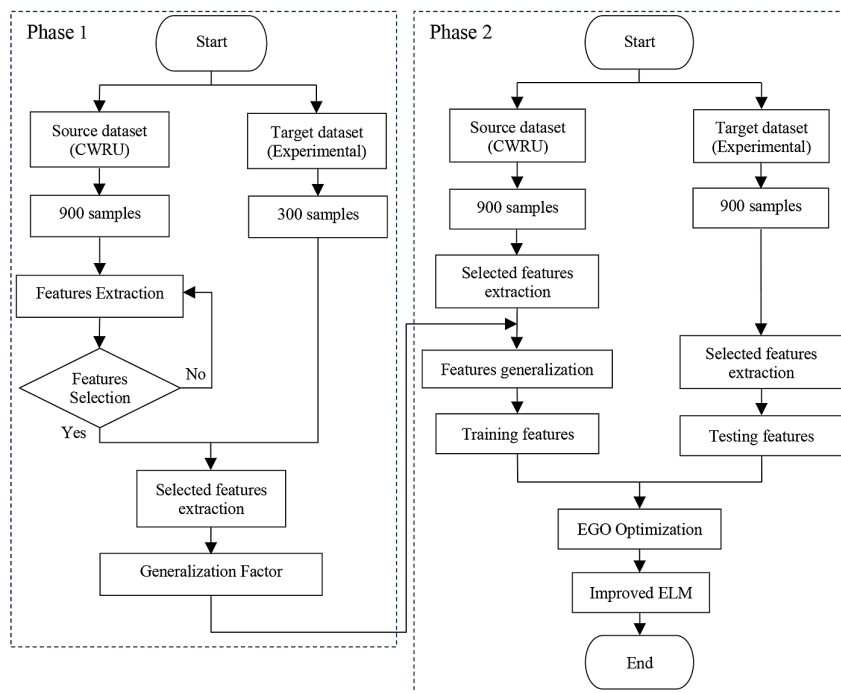**Figure 2**. EGO algorithm flowchart

Figure 3. Proposed method flowchart

in Figure 4. 900 signal samples from 3 bearing conditions are taken to be used for feature extraction. A total of 14 statistical features, 8 features from time domain and 6 features from frequency domain are then extracted from the samples.

Within time domain, the features to be observed are RMS, range, skewness, kurtosis, crest factor, shape factor, impulse factor and margin factor; obtained using equations in Table 1. While frequency domain features to be considered are RMS, skewness, kurtosis, maximum, mean and standard deviation; acquired through equations in Table 2. The features are then selected according to human observation of the features data distribution clustering pattern that displays distinctive borders between bearing conditions. The selected feature will then be utilised for the remaining feature extraction processes of the research.

Afterwards, samples from healthy bearing of target domain dataset; annotated with 300 samples in Figure 3 will be subjected to selected feature extraction. This process is within the assumption that only healthy bearing of target domain dataset is available. Thus, by leveraging source domain healthy bearing extracted features previously procured; generalization factor (*GF*) for each selected feature is then calculated using Equation 17. The purpose of GF is to find the domain shift between source domain and target domain; which will be utilized to in second phase.

$$GF = \frac{mean\ of\ target\ domain\ feature}{mean\ of\ source\ domain\ feature} \quad (17)$$

The econd phase of the methodology, diagnosis model development is as presented in the right side of Figure 4 where 900 samples from all bearing conditions of both domains' datasets are prepared similarly to previous sampling parameter. Then, the selected features are extracted from both domain datasets; features extracted from target domain are assigned as testing data while features extracted from source domain are generalized utilizing generalization factor obtained and then assigned as training data. Generalized training data

Table 1. Vibration time-domain statistical features

Statistical feature	Equation	Statistical feature	Equation
Range	$\max(x_i) - \min(x_i)$	Crest factor	$\dfrac{max\|x_i\|}{\sqrt{\frac{1}{N}\sum_{i=1}^{I} x_i^2}}$
RMS	$\sqrt{\dfrac{\sum_{i=1}^{I} x_i^2}{N}}$	Shape factor	$\dfrac{\sqrt{\frac{1}{N}\sum_{i=1}^{I} x_i^2}}{\frac{1}{N}\sum_{i=1}^{I}\|x_i\|}$
Skewness	$\dfrac{\frac{1}{N}\sum_{i=1}^{I}(x_i - \bar{x})^3}{\left(\sqrt{\frac{1}{N}\sum_{i=1}^{I}(x_i - \bar{x})^2}\right)^3}$	Impulse factor	$\dfrac{max\|x_i\|}{\frac{1}{N}\sum_{i=1}^{I}\|x_i\|}$
Kurtosis	$\dfrac{\frac{1}{N}\sum_{i=1}^{I}(x_i - \bar{x})^4}{\left(\sqrt{\frac{1}{N}\sum_{i=1}^{I}(x_i - \bar{x})^2}\right)^4}$	Margin factor	$\dfrac{max\|x_i\|}{\left(\frac{1}{N}\sum_{i=1}^{I}\sqrt{\|x_i\|}\right)^2}$

Table 2. Frequency-domain statistical features

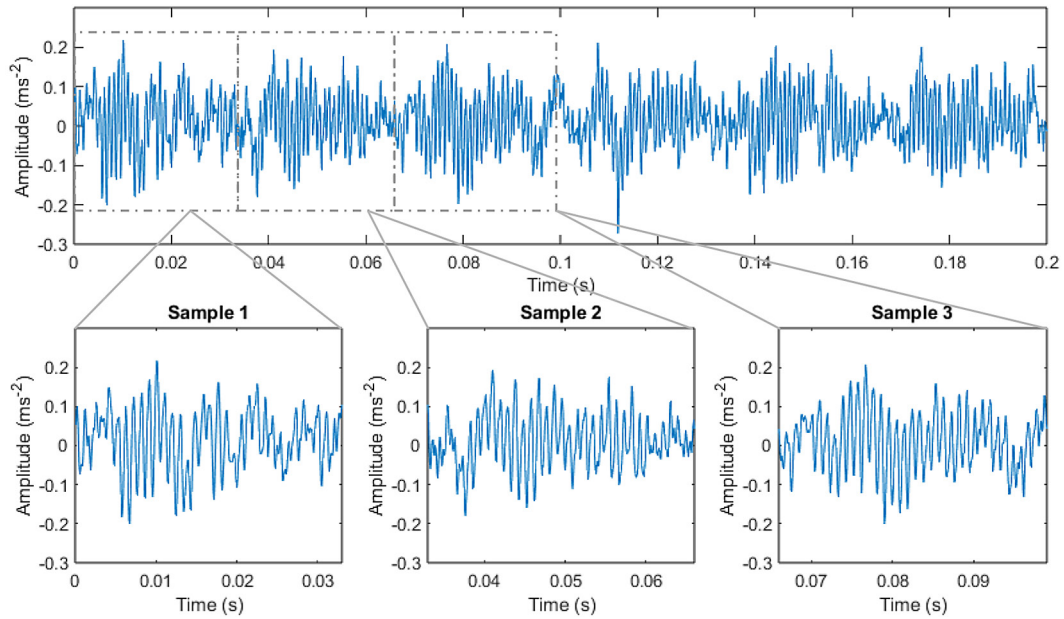| Statistical feature	Equation	Statistical feature	Equation
RMS	$\sqrt{\dfrac{\sum_{k=1}^{K} s_k^2}{N}}$	Mean, $\bar{s}$	$\dfrac{\sum_{k=1}^{K} s_k}{N}$
Standard deviation	$\sqrt{\dfrac{\sum_{k=1}^{K}(s_k - \bar{s})^2}{N}}$	Kurtosis	$\dfrac{\frac{1}{N}\sum_{k=1}^{K}(s_k - \bar{s})^4}{\left(\sqrt{\frac{1}{N}\sum_{k=1}^{K}(s_k - \bar{s})^2}\right)^4}$
Skewness	$\dfrac{\frac{1}{N}\sum_{k=1}^{K}(s_k - \bar{s})^3}{\left(\sqrt{\frac{1}{N}\sum_{k=1}^{K}(s_k - \bar{s})^2}\right)^3}$	Maximum	$\max(s)$

Figure 4. Sampling process procedure

is expected to be aligned with testing data; thus, reducing the domain shift between source domain and target domain as demonstrated in Figure 5.

The training data are labelled and then fed into EGO optimizer to obtain the ideal ELM parameters; number of hidden layer neurons, weight values and bias values for improved ELM method. The importance of this procedure is because the parameter significantly affects ELM accuracy as demonstrated in Figure 6 where ELM is run for 200 times with increasing neuron number ranging from 1 to 200. Highest diagnostic accuracy is observed to be obtained using 9 hidden layer neurons.

Afterwards, the training data is fed into improved ELM to develop the proposed bearing FD model. After the training process, unlabelled testing data are fed into the model for bearing condition prediction. The output is then compared with experimental data label to determine the fault classification accuracy. Equation 18 details the objective function for the EGO.

$$Obj.Function =$$
$$= 100 - \left[\frac{Training\ Acc.\,(\%) + Testing\ Acc.\,(\%)}{2}\right] (18)$$

The datasets utilized in this study were obtained from CWRU and an experimental testing rig. The CWRU dataset was employed for the formulation of the ELM model, while the experimental dataset was used to assess the model's performance. The CWRU dataset, acted as the source domain, was primarily utilized to train the proposed ELM model and was sourced from the CWRU bearing data centre website.
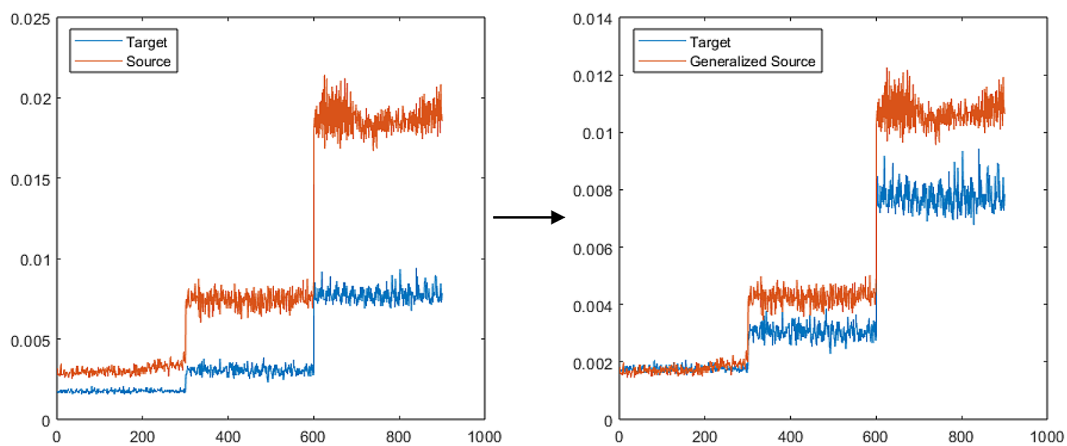

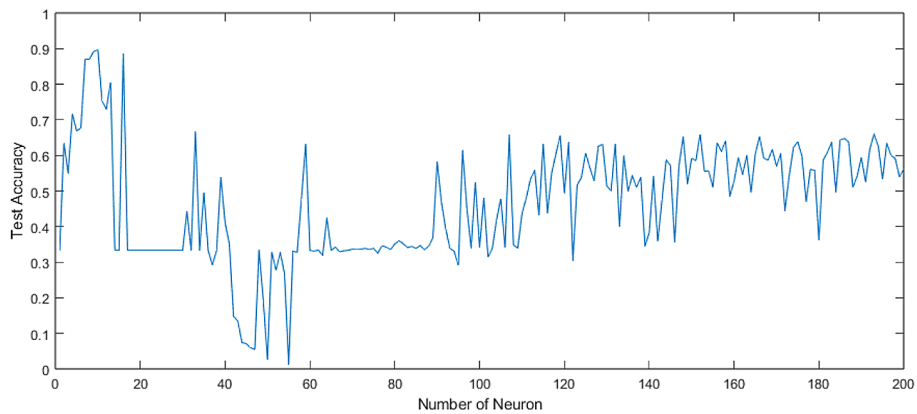**Figure 5.** Generalization effect on frequency-domain mean feature

Figure 6. Number of neuron effect on ELM algorithm accuracy

The online dataset comprises three sets of vibration signals with distinct bearing conditions: healthy, ball fault, and inner race fault. The signals were captured in the laboratory using bearings with artificially induced faults. Each induced fault measures 0.007 inches in diameter running at 1797 RPM with 12 kHz sampling frequency. The test rig configuration was depicted in Figure 7 (a), sourced from CWRU. More details regarding the configuration were available on the website.

The experimental dataset was obtained from the experimental test rig. The test rig simulates a real rotating machine with three bearing conditions: healthy, ball fault, and inner race fault. The test rig is the Machinery Fault and Rotor Dynamics Simulator (MFS-RDS) made by Spectra Quest, as illustrated in Figure 7b operating at 1800 RPM speed with sampling rate 25.6 kHz. The labelled dataset produced by this test rig served as the target domain for this study. Raw vibration signals from both online and experimental datasets were shown in Figure 8.

RESULTS

The result will be presented similarly to the methodology which consists of two phases. Phase 1 presents the results and discussion of features selection and generalization. While phase 2 focuses towards FD model development process.
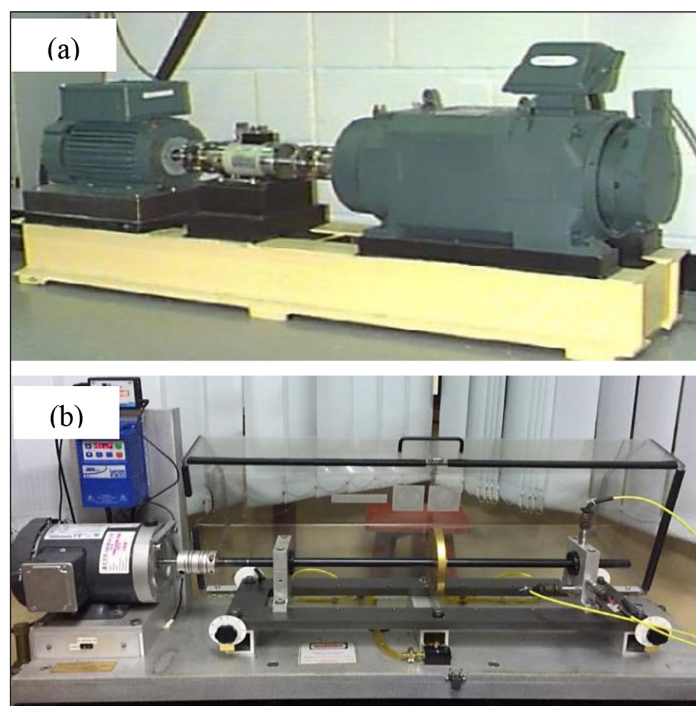

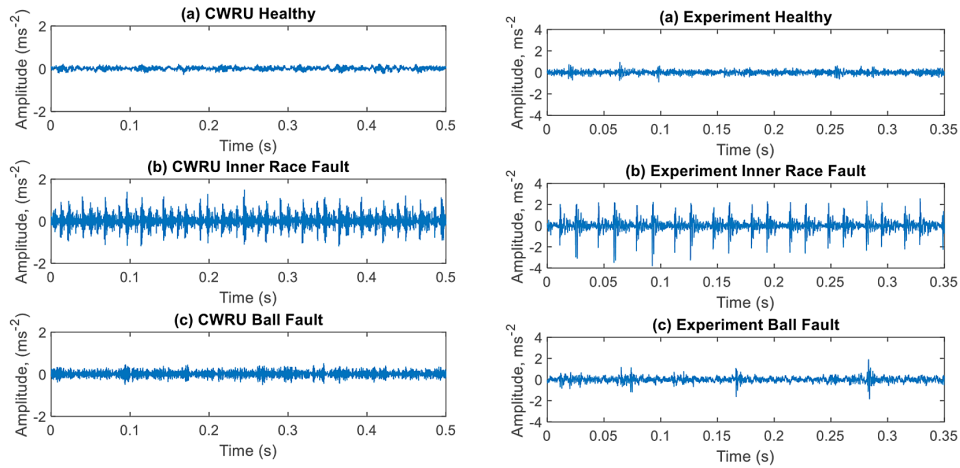**Figure 7.** (a) Online test rig and (b) experimental test rig

Figure 8. Raw vibration signals of CWRU and experimental datasets

Features selection and generalization

Initially, both datasets; CWRU and experimental are subjected to sampling process as accordance to their sampling rate. Details of sampling parameter of phase 1 are as described in Table 3.

Subsequently, A total of 14 statistical features were extracted from the signal samples and specific features were selected according to features distribution clustering. These selections were based on the visualizations as demonstrated in Figure 9 and Figure 10. For this visualization

Table 3. Datasets configuration for features selection and generalization

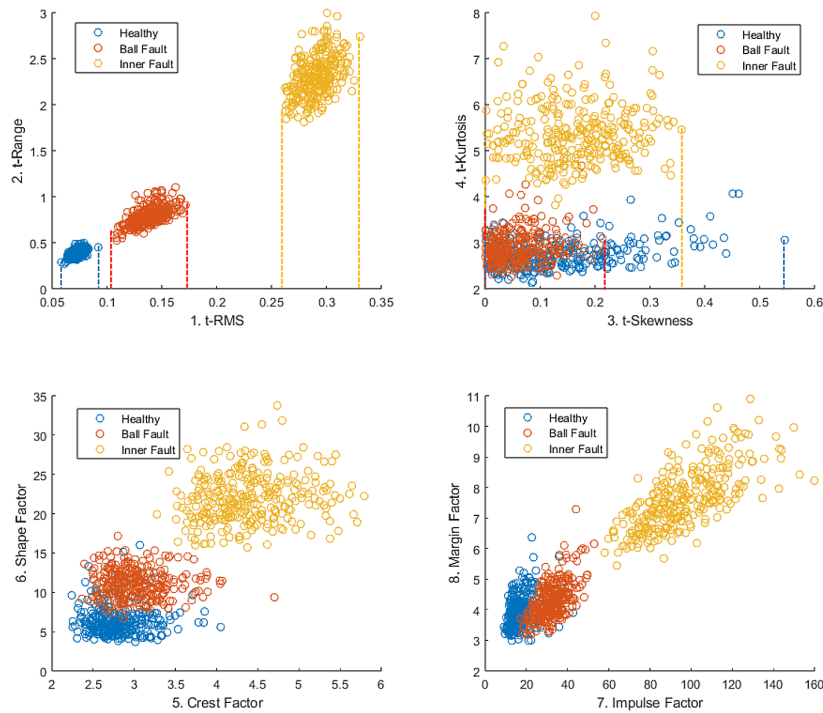| Datasets	Condition	Signal samples	Data per samples
Case western reserve university (CWRU)	Healthy	300	400
	Ball fault	300	400
	Inner fault	300	400
Experimental	Healthy	300	856

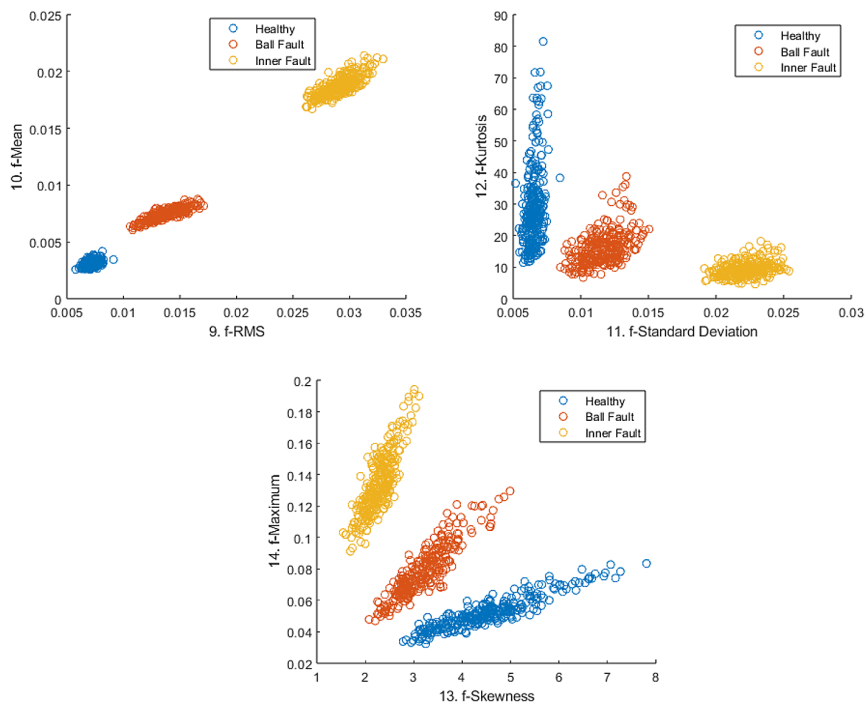**Figure 9.** Clustering of CWRU time domain statistical features

Figure 10. Clustering of CWRU frequency domain statistical features

and selection, only features from CWRU will be used, as it is considered the source domain for this study. Figure 9 shows the clustering for each statistical feature from the time-domain. Based on the clustering, RMS, range, shape factor and impulse factor show 3 clusters as accordance to bearing conditions, which indicates the most sensitive features for these datasets. The selection is according to 1-dimensional cluster; for instance, x-axis on top left chart in Figure 9 labelled with '1. t-RMS' is selected as 3 1-dimensional clusters of data points can be observed. Based on the same selection criteria, that is also the reason for '3. t-Skewness' in top right chart in the same Figure 9 is not selected because no clear clusters can be observed from the feature axis perspective. Figure 10 presented the frequency-domain clustering for each statistical feature. Based on the clustering, the standard deviation and mean are selected with the same reason as previous selections. Frequency domain RMS is not selected as the distribution is similar to time domain RMS.

Hence, a total of six features are selected and used for feature generalization as depicted in Figure 11. For feature generalization, only healthy condition from both CWRU and experimental datasets are used. Then, selected features are extracted from the healthy signal samples of both datasets. Mean from each

feature of both domains are obtained and GF value are calculated as Equation 17. The same approach was implemented all 6 selected features accordingly, and the GF value for each is described in Table 4.

Cross-machine fault classification

For cross-machine fault classification, 300 signal samples had been extracted from each dataset, as described in Table 5. Selected statistical features were then extracted from each signal sample and CWRU features are generalized by multiplying to GF value as according to type of feature as stated in Table 4. The same GF value was used for each condition of source dataset, which are healthy, ball fault, and inner race fault.

The classification was based on the ELM algorithm enhanced with parameters; number of neurons, weight values and bias values optimized by EGO method. The EGO method was employed with input parameters according to Table 6 to find the best parameters for the study. Variables are set to 3 as the 3 parameters for optimized ELM algorithm and boundary set to control range of the values for the IELM parameters. As observed in Figure 12, the convergence rate of the optimizer managed to obtain the optimum values of IELM parameters after 15 iterations.
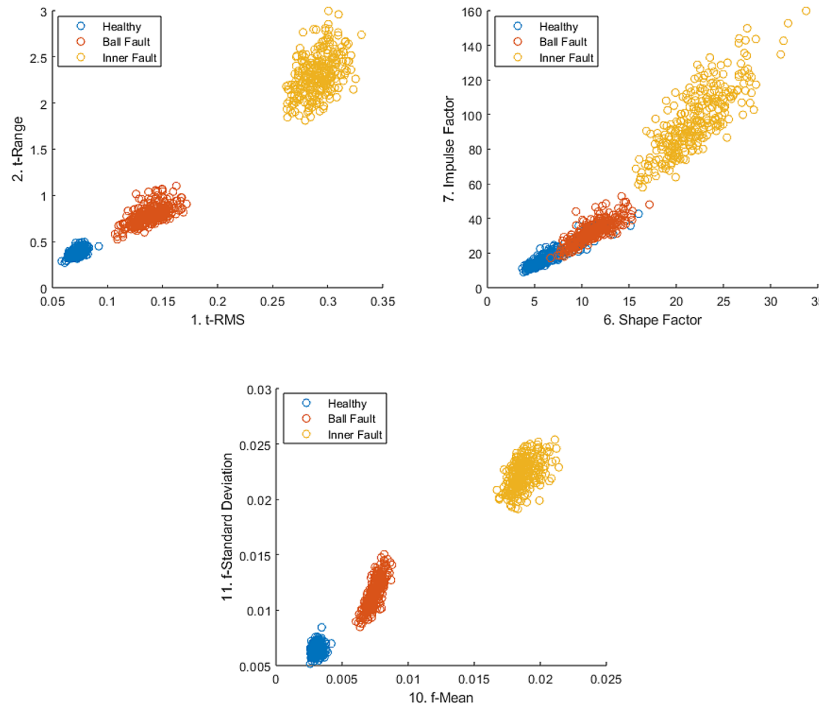
Figure 11. Selected features from time domain and frequency domain

The optimum values of IELM parameters obtained by EGO; hidden neuron number of 5, input weight coefficient of 0.0015, and bias coefficient of 0.0047 are used to train the proposed SFG-IELM. To measure the effect of generalization method employed in this study, 2 different transfer learning FD models based on conventional ELM were trained using non-generalized CWRU selected features and generalized CWRU selected features respectively. To ensure fairness, training and testing dataset configuration

of these 2 models are kept consistent with proposed method. Besides, all models were run for 30 times; where performance is recorded after each fully trained run to obtain the average accuracy of each model for better performance representation and tabulated in Table 7.

The average and overall accuracy are presented in Figure 13 where generalization procedure has improved testing accuracy of conventional ELM model from 70.5% to 85.3% as observed in comparison between Non Gen

Table 4. GF value for each selected feature

Statistical feature	GF value
Time-domain RMS	0.9109
Time-domain range	1.0014
Time-domain shape factor	0.9070
Time-domain impulse factor	0.9584
Frequency-domain mean	0.5726
Frequency-domain standard deviation	0.6327

Table 6. EGO parameters

EGO parameter	Value
Population size	200
Iteration	30
Variable	3
Lower bound	[0 0 1]
Upper bound	[1 1 700]
Others	Default

Table 5. Datasets configuration for classification

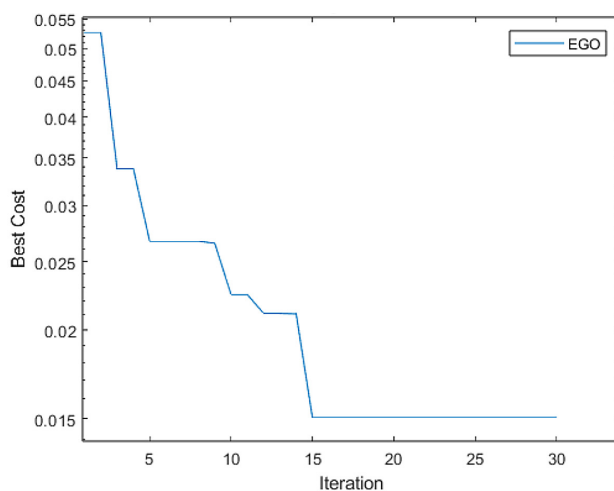| Purpose	Datasets	Condition	Number of samples
Training	Generalized CWRU	Healthy	300
		Inner race fault	300
		Ball fault	300
Testing	Experimental	Healthy	300
		Inner race fault	300
		Ball fault	300

Figure 12. EGO convergence curve

Table 7. Detailed accuracy of 30 runs of non-generalized ELM, generalized ELM and SFG-IELM

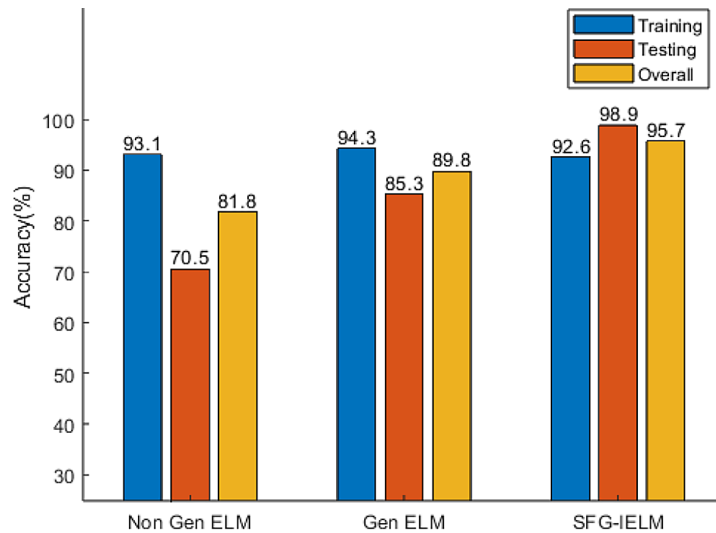| Run	Accuracy (%)					
	Non-generalized ELM		Generalized ELM		SFG-IELM	
	Training	Testing	Training	Testing	Training	Testing
1	96.0	73.4	95.0	88.9	92.8	99.3
2	93.0	53.7	93.8	91.9	92.6	98.6
3	95.6	68.0	95.4	87.3	92.3	99.3
4	92.0	69.4	97.4	99.4	92.8	99.3
5	93.6	72.0	90.1	71.3	92.4	99.3
6	94.4	54.7	95.7	99.0	92.4	98.0
7	93.3	70.8	92.2	88.6	93.2	99.7
8	96.0	68.2	94.7	91.4	92.7	99.3
9	89.4	67.6	95.3	87.4	92.2	97.8
10	92.6	77.0	95.6	93.8	92.7	98.6
11	91.3	76.3	97.3	99.4	92.7	99.3
12	91.2	69.6	96.2	81.8	92.4	99.2
13	97.8	95.9	94.2	74.3	92.6	97.9
14	91.0	67.8	89.7	69.9	92.4	97.9
15	91.8	73.9	94.6	85.8	92.9	98.9
16	96.2	88.7	96.6	96.1	92.3	97.4
17	97.0	75.8	94.8	91.3	92.2	97.8
18	96.6	62.4	90.3	71.7	92.7	99.0
19	90.6	67.0	98.2	97.9	93.1	99.6
20	94.2	73.9	91.1	77.3	93.6	99.7
21	90.8	69.0	96.9	94.2	92.6	98.6
22	93.6	75.2	93.3	76.9	92.3	99.2
23	95.1	81.0	92.0	80.3	92.4	98.3
24	90.7	66.3	90.1	67.9	92.7	99.6
25	90.0	67.3	91.8	77.2	92.0	99.1
26	92.3	69.7	96.9	91.6	92.0	99.4
27	91.2	73.9	95.6	73.4	92.1	99.2
28	92.1	55.9	90.3	73.1	92.6	98.9
29	92.7	62.2	96.6	96.3	92.6	99.0
30	90.7	68.9	97.3	82.2	92.8	98.9
Average	93.1	70.5	94.3	85.3	92.6	98.9
Overall	81.8		89.8		95.7	

Figure 13. Average accuracy comparison

ELM and Gen ELM. Furthermore, the proposed EGO optimization has further improved testing accuracy to 98.9% as shown in Figure 13 labelled with SFG-IELM. Figure 14 represents the stability of the 3 models when trained for 30 times. SFG-IELM is discerned to be more stable when compared to 2 other models; thus, proving that controlling weight and bias values using EGO optimization managed to stabilize the model with optimum performance.

Within transfer learning FD models, testing accuracy is prioritized as it represents the purpose of transfer learning where performance of the model in target domain is pursued. Hence, the testing performance of the proposed method is compared to other recent cross-machine bearing FD models; dictionary domain adaptation transformer (DDAT) [34] and cross-domain manifold structure preservation (CD-MSP) [35] as presented in Figure 15 bar chart. Both of these methods also deployed CWRU bearing dataset as source domain. SFG-IELM accuracy in this study is slightly better than other 2 cross-machine FD models; therefore, indicating the performance to be on par with recent literatures.
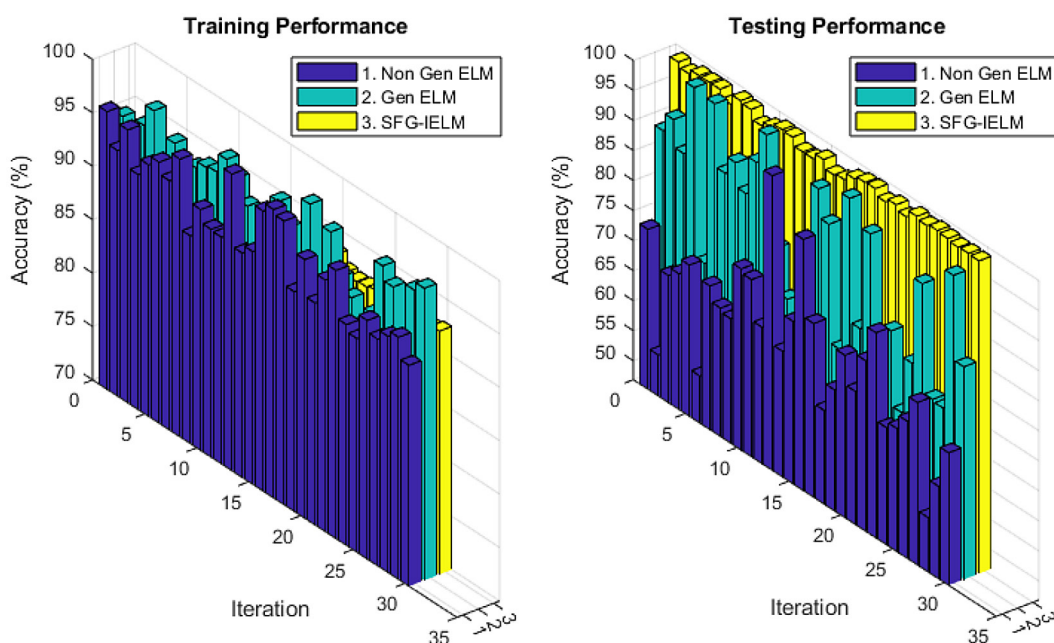

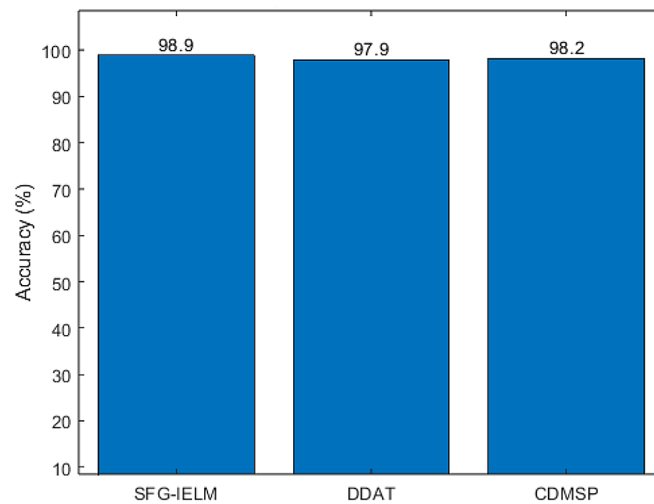**Figure 14.** Detailed performance of 30 runs of non-generalized ELM, generalized ELM and SFG-IELM

Figure 15. Testing performance comparison between SFG-IELM, DDAT and CDMSP

CONCLUSIONS

This study proposed an alternative approach for cross-machine bearing diagnosis based on statistical feature generalization and IELM. Selected statistical features were also presented, with six out of fourteen statistical features from the time-domain and frequency-domain selected. The generalized factor (GF) value was also presented for each selected statistical parameter feature, which will be used for generalizing source-domain to target-domain. A comparison result presented in this paper suggests that the generalization and optimization improved the cross-machine diagnosis performance from 70.5% to 98.9%, which is very significant with 28.4% differences. Moreover, SFG-IELM also exhibits a similar level of performance with recent cross-machine bearing FD literatures. As per future remarks, the proposed cross-machine diagnosis should also be tested with other machine components such as gear, shaft, and blade.

Acknowledgment

REFERENCES

1. Yu K, Fu Q, Ma H, Lin TR, Li X. Simulation data driven weakly supervised adversarial domain adaptation approach for intelligent cross-machine fault diagnosis. Struct Health Monit, 2021; 20: 2182–2198.

2. Saufi SR, Ahmad ZAB, Leong MS, Lim MH Challenges and Opportunities of Deep Learning Models for Machinery Fault Detection and Diagnosis: A Review. IEEE Access 2019; 7: 122644–122662.

3. Li Z, Jiang Y, Hu C, Peng Z. Recent progress on decoupling diagnosis of hybrid failures in gear transmission systems using vibration sensor signal: A review. Measurement 2016; 90: 4–19.

4. Zheng H, Wang R, Yang Y, Yin J, Li Y, Li Y, Xu M. Cross-domain fault diagnosis using knowledge transfer strategy: A review. IEEE Access 2019; 7: 129260–129290.

5. Isham MF, Leong MS, Lim MH, Ahmad ZA. Variational mode decomposition for rotating machinery condition monitoring using vibration signals. Trans Nanjing Univ Aero Astro 2018; 35: 38–50.

6. Isham MF, Leong MS, Lim MH, Bin Ahmad ZA. Intelligent wind turbine gearbox diagnosis using VMDEA and ELM. Wind Energy 2019. https://doi.org/10.1002/we.2323

7. Isham MF, Leong MS, Lim MH, Zakaria MK. A Review on Variational Mode Decomposition for Rotating Machinery Diagnosis. MATEC Web Conf. 255. 2019.

8. Ren Y, Ye Q, Xu X, Huang Q, Fan Z, Li C, Chang W. An anomaly pattern detection for bridge structural response considering time-varying temperature coefficients. Structures 2022; 46: 285–298.

9. Liu SX, Lü M. Fault diagnosis of the blocking diesel particulate filter based on spectral analysis. Processes. 2019. https://doi.org/10.3390/pr7120943

10. Yao H, Zhang X, Guo Q, Miao Y, Guan S. Fault diagnosis method for oil-immersed transformers integrated digital twin model. Sci Rep. 2024. https://doi.org/10.1038/s41598-024-71107-w

11. Saha DK, Hoque ME, Badihi H. Development of

intelligent fault diagnosis technique of rotary machine element bearing: A machine learning approach. Sensors. 2022. https://doi.org/10.3390/s22031073

12. Youcef Khodja A, Guersi N, Saadi MN, Boutasseta N. Rolling element bearing fault diagnosis for rotating machinery using vibration spectrum imaging and convolutional neural networks. The International Journal of Advanced Manufacturing Technology 2020; 106: 1737–1751

13. Shao K, Fu W, Tan J, Wang K. Coordinated approach fusing time-shift multiscale dispersion entropy and vibrational Harris hawks optimization-based SVM for fault diagnosis of rolling bearing. Measurement 2021; 173: 108580.

14. Jia LS, Chow TWS, Wang Y, Ma JH. Dynamic Balanced Dual Prototypical Domain Generalization for Cross-Machine Fault Diagnosis. IEEE Trans Instrum Meas. 2024. https://doi.org/10.1109/TIM.2024.3381292

15. He YM, Shen WM. MSiT: A Cross-Machine Fault Diagnosis Model for Machine-Level CNC Spindle Motors. IEEE Trans Reliab 2024; 73: 792–802.

16. Li C, Wang GB, Zhao SB, Zhong ZX, Lv Y. Cross-domain manifold structure preservation for transferable and cross-machine fault diagnosis. Journal of Vibroengineering. 2024; 26: 1367–1384.

17. Wang XD, Liu F, Zhao DD. Cross-machine fault diagnosis with semi-supervised discriminative adversarial domain adaptation. Sensors. 2020. https://doi.org/10.3390/s20133753

18. Lv MZ, Liu SX, Su XM, Chen CZ. Deep transfer network with multi-kernel dynamic distribution adaptation for cross-machine fault diagnosis. IEEE ACCESS, 2021; 9: 16392–16409.

19. Zhang K, Ding K, Zheng Q, Zou YS, Ding GF. A novel cross-bearing fault diagnosis method based on pseudo-label transitive domain adaptation networks. Journal of Vibration and Control. 2023. https://doi.org/10.1177/10775463231202550

20. Yuan SZ, Liu ZH, Wei HL, Chen L, Lv MY, Li XH. A Variational Auto-Encoder-Based Multisource Deep Domain Adaptation Model Using Optimal Transport for Cross-Machine Fault Diagnosis of Rotating Machinery. IEEE Trans Instrum Meas. 2024. https://doi.org/10.1109/TIM.2023.3331436

21. Feng Y, Chen J, He S, Pan T, Zhou Z. Globally localized multisource domain adaptation for cross-domain fault diagnosis with category shift. IEEE Trans Neural Netw Learn Syst 2023; 34: 3082–3096.

22. Huang G Bin, Zhu QY, Siew CK. Extreme learning machine: Theory and applications. Neurocomputing 2006; 70: 489–501.

23. Huang G-B, Zhou H, Ding X, Zhang R Extreme learning machine for regression and multiclass classification. IEEE transactions on systems, man, and cybernetics Part B, Cybernetics. 2012; 42: 513–29.

24. Huérfano-Maldonado Y, Mora M, Vilches K, Hernández-García R, Gutiérrez R, Vera M A comprehensive review of extreme learning machine on medical imaging. Neurocomputing. 2023. https://doi.org/10.1016/j.neucom.2023.126618

25. Isham MF, Saufi MSR, Waziralilah NF, Talib MHAb, Hasan MDA, Saad WAA Optimized-ELM Based on Geometric Mean Optimizer for Bearing Fault Diagnosis. In: Mohd. Isa WH, Khairuddin IMohd, Mohd. Razman MohdA, Saruchi S 'Atifah, Teh S-H, Liu P (eds) Intelligent Manufacturing and Mechatronics. Springer Nature Singapore, Singapore, 2024; 125–139.

26. Mercaldo F, Brunese L, Martinelli F, Santone A, Cesarelli M. Experimenting with extreme learning machine for biomedical image classification. Applied Sciences (Switzerland). 2023. https://doi.org/10.3390/app13148558

27. Albadr MAA, Tiun S, Ayob M, Nazri MZA, AL-Dhief FT. Grey wolf optimization-extreme learning machine for automatic spoken language identification. Multimed Tools Appl 2023; 82: 27165–27191.

28. Liu C, Pan G, Song D, Wei H. Air quality index forecasting via genetic algorithm-based improved extreme learning machine. IEEE Access 2023; 11: 67086–67097.

29. Wu K, Xu C, Yan J, Wang F, Lin Z, Zhou T. Error-distribution-free kernel extreme learning machine for traffic flow forecasting. Eng Appl Artif Intell. 2023. https://doi.org/10.1016/j.engappai.2023.106411

30. Isham MF, Saufi MSR, Hasan MDA, Saad WAA, Leong MS, Lim MH, Ahmad ZAB. Bearing Fault Diagnosis Using Extreme Learning Machine Based on Artificial Gorilla Troops Optimizer BT - Advances in Intelligent Manufacturing and Mechatronics. In: Abdullah MA, Khairuddin IMohd, Ab. Nasir AF, Mohd. Isa WH, Mohd. Razman Mohd A, Rasid Mohd AH, Zainal SMHF, Bentley B, Liu P (eds). Springer Nature Singapore, Singapore, 2023; 87–103.

31. Isham MF, Leong MS, Lim MH, Ahmad ZAB. Optimized ELM based on Whale Optimization Algorithm for gearbox diagnosis. MATEC Web Conf. 255: 2019.

32. Mohammadzadeh A, Mirjalili S. Eel and grouper optimizer: a nature-inspired optimization algorithm. Cluster Comput 2024; 27: 12745–12786.

33. Mohammadzadeh A, Mirjalili S. Eel and grouper optimizer: a nature-inspired optimization algorithm. Cluster Comput. 2024. https://doi.org/10.1007/s10586-024-04545-w

34. Cui L, Wang G, Liu D, Pan X. Dictionary domain adaptation transformer for cross-machine fault diagnosis of rolling bearings. Eng Appl Artif Intell. 2024. https://doi.org/10.1016/j.engappai.2024.109261

35. Li C, Wang G, Zhao S, Zhong Z, Lv Y. Cross-domain manifold structure preservation for transferable and cross-machine fault diagnosis. Journal of Vibroengineering. 2024. https://doi.org/10.21595/jve.2024.24067