

Stacked Auto-encoder Based Feature Transfer Learning and Optimized LSSVM-PSO Classifier in Bearing Fault Diagnosis

VietHung Nguyen¹, JunSheng Cheng², VanTrong Thai¹

¹*Department of Mechanical Engineering, Hanoi University of Industry, MinhKhai street, 298., 100000, Hanoi, VietNam, nguyenvietchung@hau.edu.vn*

²*College of Mechanical and Vehicle Engineering, Hunan University, Changsha 410082, Hunan, China.*

Abstract: This paper proposes a new diagnosis technique for predicting the big data of roller bearing multi-level fault, which uses the deep learning method for the feature representation of the vibration signal and an optimized machine learning model. First, vibration feature extraction by stacked auto-encoders (VFE-SAE) with two layers in roller bearing fault signals is proposed. The unsupervised learning algorithm in VFE-SAE is used to reveal significant properties in the vibration data, such as nonlinear and non-stationary properties. The extracted features can provide good discriminability for fault diagnosis tasks. Second, a classifier model is optimized based on least squares support vector machine classification and particle swarm optimization (LSSVM-PSO). This model is used to perform supervised fine-tuning and classification; it is trained with the labelled features to identify the target data. Especially, using transfer learning, the performance of the bearing fault diagnosis technique can be fine-tuned. In other words, the features of the target vibration signal can be extracted by the learning of feature representation, which is dependent on the weight matrix of hidden layers of the VFE-SAE method. The experimental results (by analyzing the roller bearing vibration signals with multi-status fault) demonstrate that VFE-SAE based feature extraction in conjunction with the LSSVM-PSO classification is more accurate than other popular classifier models. The proposed VFE-SAE – LSSVM-PSO method can effectively diagnose bearing faults with 97.76 % accuracy, even when using 80 % of the target data.

Keywords: Multi-level fault, SAE transfer learning, vibration feature, deep learning; LSSVM-PSO classifier model.

1. INTRODUCTION

Roller bearings play a crucial role in rotating machinery, and their working condition can seriously influence equipment performance. Accordingly, it is necessary to monitor the bearing conditions and identify faults to avoid fatal system breakdowns and thus prevent economic losses and injury [1]. Therefore, a reliable automatic system that monitors roller bearing health conditions is very beneficial to industry applications.

Feature extraction is an important stage in the classification of vibration data, and suitably extracted features can increase the ability to distinguish between similar classes, thereby improving classification performance. Generally, signal analysis methods have been successfully applied for feature extraction of vibration data. These methods aim to distinguish unusual data characteristics, e.g., the time-domain parameters of kurtosis, skewness, root mean square [2], [3]; and the frequency-domain analysis methods of the fast Fourier transform (FFT) [4]; and the time-frequency domain analysis methods of the Hilbert-Huang transform, local characteristic-scale decomposition (LCD), local mean decomposition (LMD), and empirical mode decomposition (EMD) [5]-[9]. However, feature extraction based on signal analysis is required to obtain deep knowledge of vibration analysis and identify roller bearing fault mechanism.

The learning of feature representation can reveal significant data features. Previous linear methods are normally used for this purpose, e.g., principal component analysis (PCA) and linear discriminant analysis (LDA) [10]-[12], but these methods fail to model nonlinear data structures. Manifold learning methods, which are proposed for nonlinear feature extraction, can characterize the nonlinear relationships between data points [13], [14]. However, they can only process a limited number of data points due to their high computational complexity. Recently, deep learning methods have been the subject of attention for many researchers, as they can process large scale and high-dimensional datasets [1], [15]-[17] and learn the nonlinear feature representation. Therefore, in this work, we propose a deep learning stacked auto-encoder (SAE) network for vibration feature extraction (VFE; VFE-SAE) in bearing fault diagnosis, which can obtain high-level features and improve the separation capability of data. In VFE-SAE, the vibration features are extracted from the original vibration signal as an activation function after unsupervised layer intelligent training by SAE. Due to the learning of feature representation based on transfer learning [18], the target data features are extracted automatically. In the previous layer intelligent training of SAE, the weight matrix at the hidden layers is used to extract the features as the transfer learning. The extracted

features are then used to identify the actual work status by the trained classifier model.

Aiming at the objective of identification of target vibration data, a classifier model is necessary which is constructed to identify the extracted features. Popular classification methods, e.g., k -nearest-neighbor (k-NN) [19], artificial neural networks (ANNs) [20]-[22], or support vector machines (SVM) [20], [23], have been successfully applied to vibration data. However, k-NN and ANN have a number of problems, e.g., restricted by training sample size, the over-fitting problem, and parameter setting. SVM method, which has great generalization capability for the problem of training sample size can solve the problem of over-fitting based on structural risk minimization principle and local optimal solution.

The least squares support vector machine (LSSVM) was proposed by Suykens in 1999 [24] and has been successfully employed to diagnose faults in machinery. In fact, a number of experiments have proven that LSSVM achieves a good compromise between accuracy and generalization performance. For example, Ref. [25]-[27] applied LSSVM to identify the fault statuses in roller bearings; in [28] LSSVM is used to diagnose the faults of centrifugal pump; Ref. [29] used LSSVM in multi-fault diagnosis for rotating machinery. LSSVM can be applied not only to the classification problem but also to the case of regression. LSSVM possess features that characterize the maximum margin algorithm, in which, a non-linear kernel function is introduced by linear learning machine mapping the low-dimensional inseparable feature space into high-dimensional separable feature space. Therefore, in this study, we finally choose LSSVM as a classification model for bearing fault diagnosis. We also used a particle swarm optimization (PSO) algorithm to optimize the parameter pair (γ, σ) of LSSVM (LSSVM-PSO). The LSSVM-PSO model is advantageous because it combines the regression analysis and generalization performance of the basic LSSVM model with the PSO parameter pair (γ, σ) , which is trained depending on the labelled training feature set. The trained LSSVM-PSO model is then used to identify the vibration features of the target data based on VFE-SAE, which is hereafter expressed as the diagnosis technique in this study (VFE-SAE - LSSVM-PSO).

The rest of the paper is organized as follows: Section 2 presents the singular auto-encoder (AE) architecture; Section 3 presents the proposed VFE-SAE method for feature extraction of training data and target data; Section 4 presents the proposed VFE-SAE - LSSVM-PSO diagnosis technique, the optimal LSSVM-PSO model is constructed to classify the high-level vibration feature; Section 5 analyzes and discusses the experimental results of the proposed technique based on roller bearing vibration data with multi-level faults; and Section 6 outlines the conclusions.

2. SINGULAR AUTO-ENCODER ARCHITECTURE

An auto-encoder (AE) is a particular neural network architecture, which works as an unsupervised learning algorithm. Fig.1. shows a three-layer auto-encoder, i.e., an input layer, hidden layer, and output layer, which has two stages: encoder and decoder. In the auto-encoder, the input layer $f = \{f_1, f_2, \dots, f_n\}$, hidden

layer $h = \{h_1, h_2, \dots, h_m\}$, $m \ll n$ and output layer $\tilde{f} = \{\tilde{f}_1, \tilde{f}_2, \dots, \tilde{f}_n\}$ are sequentially connected. The auto-encoder is trained to replicate its input at its output.

The encoder stage accomplishes feature representation from the high-dimensional input $f = \{f_1, f_2, \dots, f_n\}$ to low-dimensionality data in the hidden layer, $h = \{h_1, h_2, \dots, h_m\}$ ($m \ll n$). Similar to the mapping, the input and hidden layers are connected by the feed-forward activation function $h = \mathbf{Sigmoid}(W^{(1)} \cdot f + b^{(1)})$ in which $W^{(1)}$ is weight matrix and $b^{(1)}$ is a bias vector. In other words, each input of vector f is transformed into hidden representation h which can compactly express the features of the input. In contrast, the decoder stage is implemented to reconstruct the input f . At this stage, the input data h back map the output data \tilde{f} with high-level feature representation. The activation function $\tilde{f} = \mathbf{Sigmoid}(W^{(2)} \cdot h + b^{(2)})$ is used to connect the h data with \tilde{f} data. The weight matrix $W^{(2)} = (W^{(1)})^T$ is referred to tied weights and the $b^{(2)}$ is the bias vector in the decoder stage.

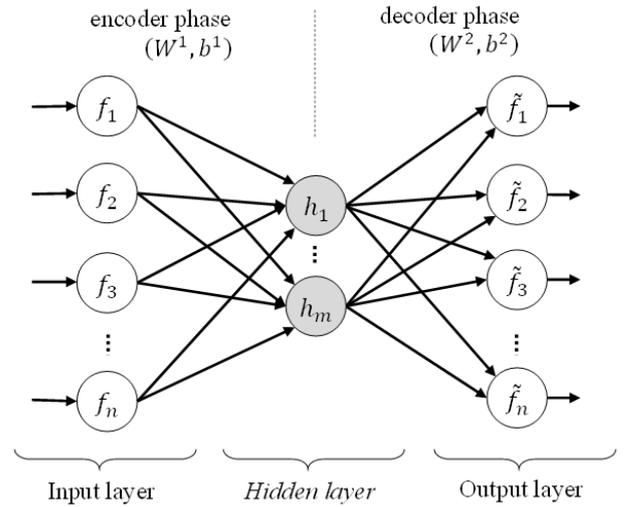


Fig.1. Illustration of deep auto-encoding.

The auto-encoder is optimized by constructing it with the $(W^{(1)}, W^{(2)}, b^{(1)}, b^{(2)})$ parameter set, which aims to minimize the reconstruction error at the output. The cost function can be expressed as follows:

$$C = \frac{1}{N} \sum_n \sum_i (f_i^{(n)} - \tilde{f}_i^{(n)})^2 + \lambda * \Omega_W + \beta * \Omega_S, \quad (1)$$

where i is the number of variables in input data, N is the number of training samples, denoting the L_2 regularization term defined by (2), Ω_S is the sparsity regularization term defined by (3), λ is the coefficient for the Ω_W , β is the coefficient for the Ω_S .

Main ideas, theory and mathematical formulations, include the data on the measuring method and instruments as well as experimental results. This part should be accompanied by exact references.

$$\Omega_W = \sum_k \sum_i \sum_j (W_{ij}^{(k)})^2, \quad (2)$$

$$\Omega_S = \sum_k \sum_j KL(\rho \parallel p_j^{(k)}), \quad (3)$$

where $p_j^{(k)}$ is the mean activation for unit j in layer k , ρ is the desired mean activation. KL is the Kullback-Leibler divergence, which is defined as follows.

$$KL(\rho \parallel p_j^{(k)}) = \rho \log \frac{\rho}{p_j^{(k)}} + (1 - \rho) \log \frac{1 - \rho}{1 - p_j^{(k)}}. \quad (4)$$

Indeed, each auto-encoder is trained independently and the feature data (which contain most of the important input information) is extracted from the auto-encoder in the hidden layer's nodes. In doing so, the data is successfully mined for information. The obtained feature data serves as input for the next auto-encoder, in which higher-level feature representation is generated.

3. PROPOSED VFE-SAE METHOD

A stacked auto-encoder (SAE) is a deep learning network, which consists of multiple layers of AEs. This deep learning network architecture represents the input data features based on unsupervised layer-wise training. SAE is applied to bearing vibration data as vibration feature extraction.

A. Stacked auto-encoders (SAE)

Auto-encoders can be stacked to build deep learning network which has more than one hidden layer [30]. Fig.2. shows a typical example of SAE structure, which includes two encoding layers and two decoding layers. At the encoding stage, the output of the first encoding layer acts as the input data of the second encoding layer. Supposing there are N hidden layers in the encoding part, we have the activation function of the k -th encoding layer:

$$h^{(k+1)} = f_e(W^{(k+1)}y^{(k)} + b^{(k+1)}) \quad (5)$$

$$k = 0, 1, \dots, N - 1,$$

where the input $y^{(0)}$ is the original data x ; the output $h^{(k)}$ of the last encoding layer are the high-level features extracted

by the SAE network. At the decoding stage, the output of the first decoding layer is regarded as the input of the second decoding layer. The decoding function of the k -th decoding layer is.

$$z^{(k+1)} = f_d\left(\left(W^{(N-k)}\right)^T z^{(k)} + b^{(k+1)}\right) \quad (6)$$

$$k = 0, 1, \dots, N - 1,$$

where the input $z^{(0)}$ of the first decoding layer is the output $y(N)$ of the last encoding layer; the output $z(N)$ of the last decoding layer is the reconstruction of the original data x ; f_e, f_d are the transfer functions of encoding and decoding, respectively. f_e, f_d are defined by (7).

$$f(x) = \frac{1}{1 + e^{-x}}. \quad (7)$$

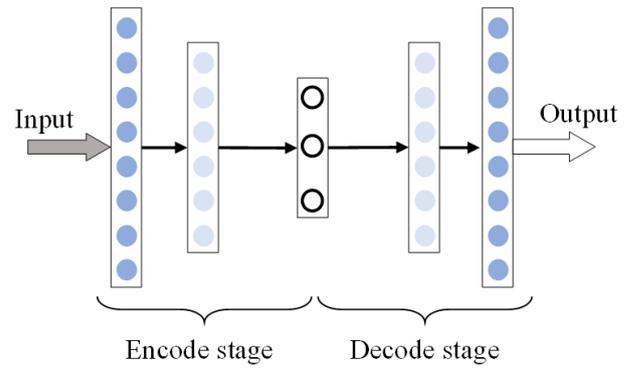


Fig.2. Deep learning SAE network based on two AE.

B. Proposed VFE-SAE based feature extraction

Vibration signal pre-processing is required by using the FFT method which can reveal the transient events or shocks defined as the mechanical disturbances of fault. Based on the transformed data, VFE-SAE will extract the significant features of vibration signal which can generate the best results of diagnosis accuracy.

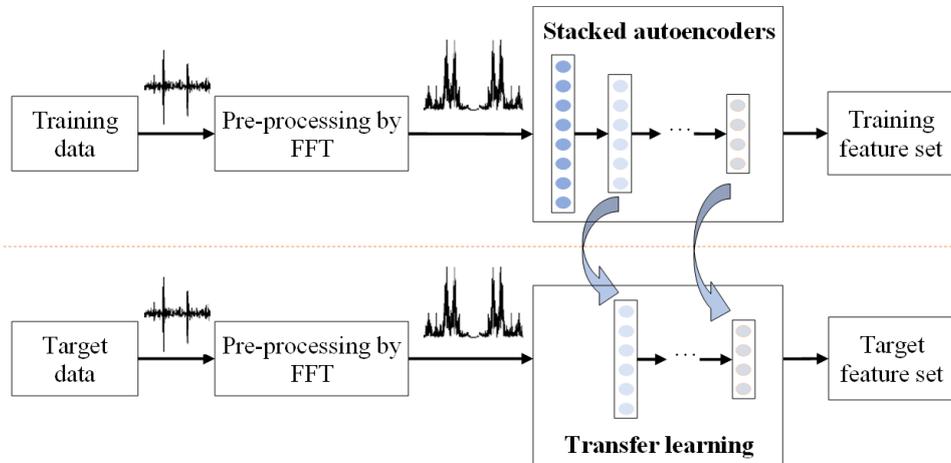


Fig.3. Proposed VFE-SAE method-based feature extraction.

Fig. 3. shows the proposed VFE-SAE method for extracting the features. At the stage of feature presentation learning, SAE is trained as the wise layers with a collection of vibration signals corresponding to bearing conditions without label. The vibration data is pre-processed into frequency domain data and is used in SAE to train the unlabeled data. The output of the SAE is the best of features. While the labelled data is used in fine-tuning of the classification model. At the feature extraction stage of target data, the optimized weighs at the hidden layer are transformed for extracting the target features as a learning process. The extracted feature data is then fed to the well-trained classifier model to identify the target data.

The feature extraction process based on VFE-SAE is provided as follows:

Step 1: Prepare data for training SAE. This data is collected from the history vibration data.

Step 2: Train the first auto-encoder which includes the first encoding layer and the last decoding layer. The network weights $W^{(1)}$, $b^{(1)}$ and the features $h^{(1)}$ which are the output of the first encoding layer can be obtained.

Step 3: Use $h^{(k)}$ as the input data to train the $(k + 1) - th$ encoding layer. Thus, W , $b^{(k+1)}$ and the features $h^{(k+1)}$ can be obtained, where $k = 1, 2, \dots, N - 1$ and N is the number of hidden layers in the deep learning network.

After the layer-wise training is repeated in the number of initializations of auto-encoder, the SAE training process is completed. In the last auto-encoder, the encoded output features are extracted at the hidden layer which represent the highest-level characteristic designed by the initial procedure.

4. PROPOSED FAULT DIAGNOSIS TECHNIQUE BASED ON VFE-SAE AND LSSVM-PSO

The VFE-SAE method-based feature extraction is combined with an optimal classifier model to exploit the effectiveness of a vibration data diagnosis technique, in which the basic LSSVM model is used to build the classifier with the parameter pair optimized by the PSO algorithm, i.e., LSSVM-PSO. The flowchart of the proposed diagnosis technique is depicted in Fig.5.

A. LSSVM-PSO classification model

A diagnosis model is constructed to identify the target vibration data by the proposed feature representation. The LSSVM-PSO model takes the advantages of the basic LSSVM, i.e., regression analysis and generalization with the parameter pair (γ, σ) optimized by the PSO algorithm, which is trained depending on the labeled training feature set as fine-tuning. This trained LSSVM-PSO model is then used to identify the vibration features of the target data as the classifying pattern.

A.1. LSSVM classifier

LSSVM introduces a least squares linear system into the SVM, which was initially put forward by Suykens in 1999 [24]. LSSVM has solved the convex optimization problems outlined below (which attempt to identify an optimal separating hyperplane) with good generalization.

Suppose that $\{(x_i, y_i) | i = 1, 2, \dots, l\}$ is the training set of samples l . The sample of $x_i, i = 1, 2, \dots, l$ corresponds to the

category of $y_i \in (-1, 1)$; the objective function and constraint condition are thus shown as follows:

$$\begin{cases} \min(J_{LS}) \text{ with } J_{LS}(w, e) = \frac{1}{2}w^T w + \frac{1}{2}\gamma \sum_{i=1}^l e_i^2 \\ \text{s. t. } y_i = w^T \varphi(x_i) + b + e_i, \quad i = 1, 2, \dots, l \end{cases}, \quad (8)$$

where e_i is the slack variable and $\gamma \geq 0$ is the penalty factor or regularization parameter.

The γ value will influence the training result of the LSSVM model: a low value reflects a high number of training errors, whereas a high value does not permit any slack variables and consequently increases model complicity. Therefore, it is critical to find a proper value for γ , and it is a vital LSSVM tuning parameter that should be adjusted conscientiously.

Defining the Lagrange function:

$$L(w, b, e, \alpha) = \frac{1}{2}w^T w + \frac{1}{2}\gamma \sum_{i=1}^l e_i^2 - \sum_{i=1}^l \alpha_i (w^T \varphi(x_i) + b + e_i - y_i), \quad (9)$$

where α_i is the Lagrange multiplier, which can be positive or negative. The optimum condition is as follows:

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^l \alpha_i y_i \varphi(x_i) \\ \frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^l \alpha_i y_i = 0 \\ \frac{\partial L}{\partial e_i} = 0 \Rightarrow \alpha_i = \gamma e_i \\ \frac{\partial L}{\partial \alpha_i} = 0 \Rightarrow w^T \varphi(x_i) + b + e_i - y_i = 0 \end{cases}. \quad (10)$$

The matrix equation can be easily obtained as follows:

$$\begin{bmatrix} 0 & Y^T \\ Y & \Omega + \frac{I}{\gamma} \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \bar{I} \end{bmatrix}, \quad (11)$$

where: $Y^T = [y_1, y_2, \dots, y_l]$; I is the unit matrix $\bar{I} = [1, 2, \dots, l]^T$, $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_l]^T$, $\Omega = y_i y_j \varphi^T(x_i) \varphi(x_j) = y_i y_j K(x_i, x_j)$, and $K(x_i, x_j)$ is the SVM Kernel function.

The LSSVM classifying function can be finally obtained as follows:

$$f(x) = \text{sgn}(\sum_{i=1}^l \alpha_i y_i K(x, x_i) + b). \quad (12)$$

Table 1. Common Kernel functions.

Kernels	Expressions	Equation
Linear	$K(x, y) = x^T \cdot y$	(13)
Polynomial	$K(x, y) = (ax^T \cdot y + b)^c$	(14)
Gauss RBF	$K(x, y) = \exp\left(-\frac{\ x - y\ ^2}{2\sigma^2}\right)$	(15)
Sigmoid	$K(x, y) = \tan(ax^T \cdot y + b)$	(16)

A.2. PSO algorithm

PSO is an algorithm for locating the optimum value of continuous nonlinear function introduced by Eberhart and Kennedy in (1995) [31]. PSO implementation is based on position and velocity which are two unique characteristics of

each particle. The particle with no volume is distributed in search space and seeks for the optimum function location. All particles except global best want to accelerate toward the global best and their local best. Therefore, a particle's new velocity can be obtained by summing its last velocity and local and global best trajectory vectors. The basic PSO mathematical expressions are as follows:

$$v_{sd}^{n+1} = wv_{sd}^n + c_1r_1(p_{sd}^n - x_{sd}^n) + c_2r_2(p_{gd}^n - x_{sd}^n), \quad (17)$$

$$x_{sd}^{n+1} = x_{sd}^n + v_{sd}^{n+1}, \quad (18)$$

where it is supposed that species consisting of S particles are marked as $x = (x_1, x_2, \dots, x_S)$ and fly at a certain speed $v_s = [v_{s1}, v_{s2}, \dots, v_{sd}]^T$, $s = 1, 2, \dots, S$ in the solution space. The search space dimensions are denoted by d , and the speed can be adjusted dynamically according to historical behavior. The optimal search position is $p_s = [p_{s1}, p_{s2}, \dots, p_{sd}]^T$ and p_{sd} is the optimum solution searched by particle S in d -dimensional space, w is the dynamic inertia weight value, which is introduced to control the local and global optimization performance; n is iteration number; random variables of r_1 and r_2 obey uniform distribution of the interval $\{0,1\}$; c_1 and c_2 are acceleration constants.

As mentioned in the upper sub-section, the dependency of the mean square error of predicted targets on LSSVM tuning parameters is not direct and linear. In addition, in the radial basis function (RBF) kernel, it is difficult to find proper values for γ and σ through continuous intervals. Due to these constraints, it is necessary to combine PSO with LSSVM to produce better results.

A.3. Optimized LSSVM-PSO classifier model

The parameter pair (γ, σ) of LSSVM plays a crucial role in model construction, it can be obtained using the PSO algorithm. This algorithm is used to explore the search space of the given LSSVM classification problem to find the parameter pair (γ, σ) , which is required to maximize the particular objective of accuracy. The principled training phase of the optimal LSSVM-PSO classifier model includes several main steps as follows:

Step 1: Prepare the training feature set, which is extracted by VFE-SAE after pre-processing the original vibration data.

Step 2: This is an initialization step, in which the initial γ and σ parameters are randomly set for LSSVM. Here, the maximum iteration number t_{max} is set as well as the iterative variable $t = 0$ to perform the training process for the next steps. For this optimization algorithm, the maximum iteration is $t_{max} = 20$.

Step 3: Increase iteration variable by $t = t + 1$.

Step 4: Deterioration evaluation. The deterioration function is employed to evaluate the quality of every element. Equation (19) shows the LSSVM classification accuracy.

$$Deterioration(\%) = \frac{y_{actual} - y_{classified}}{N} 100\%, \quad (19)$$

where y_{actual} is the actual sample, $y_{classified}$ is the classified sample, and N is the total number of samples in the testing process. For a high classification accuracy, the absolute value should be small.

Step 5: Stop criteria checking. If the deterioration function is satisfied by (19) or iteration is maximal, go to Step 7. If not, go to Step 6.

Step 6: Update the new γ and σ parameters and return to Step 3.

Step 7: End the training procedure, the fitting parameters are the optimal output values.

The efficient search capability of PSO algorithm incorporated with the generalization capability of LSSVM can bring out the higher classification accuracy. The LSSVM-PSO architecture is presented in Fig.4. Each reactant represents the candidate solution for the model, which includes the penalty parameter γ and the kernel function parameter σ .

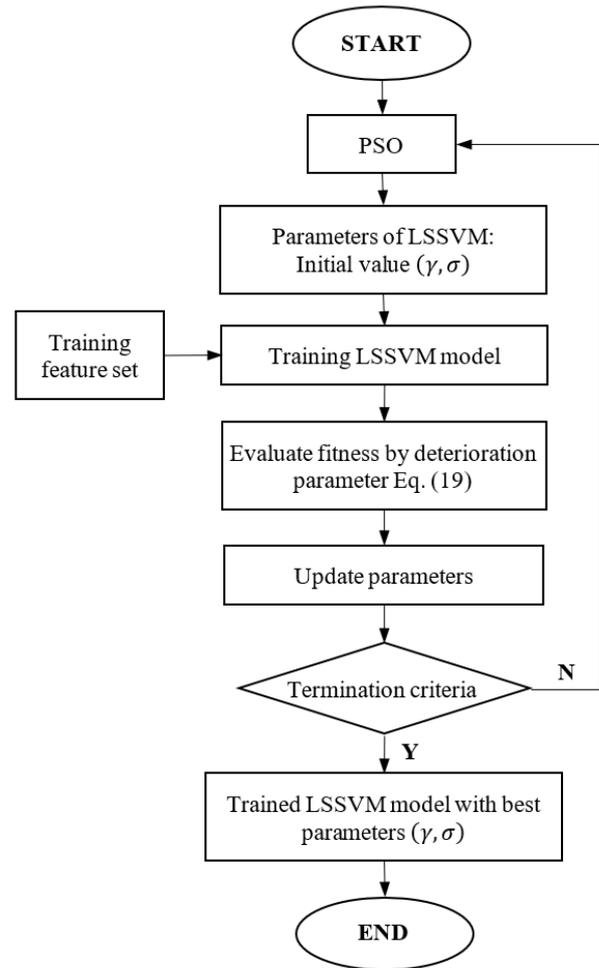


Fig.4. Constructing the optimal LSSVM-PSO classifier model.

B. Proposed VFE-SAE - LSSVM-PSO diagnosis technique

This section presents a complete roller bearing fault diagnosis technique, i.e., the proposed VFE-SAE - LSSVM-PSO. The VFE-SAE feature extraction method effectively combines with the optimized LSSVM-PSO classifier model. In particular, a deep learning SAE network with a strong ability for high-dimensionality data is constructed to extract vibration data features, and, based on layer-wise training, transfer learning is used to increase efficiency by introducing the LSSVM-PSO classifier model. Fig.5. shows the VFE-

SAE - LSSVM-PSO implementation, which is also explained below.

Step 1: Collect vibration data.

Step 2: Conduct FFT-based pre-processing.

Step 3: Conduct VFE-SAE-based feature extraction, the process of which is outlined below.

1. Pre-processed training data is used to train the deep learning SAE network, where the features are extracted from the hidden layers of deep learning SAE network, which is an unsupervised learning method in feature representation.
2. The feature of pre-processed target data is extracted by the transfer learning method. In doing so, the weight

matrices at hidden layer of the SAE network are transformed, and, due to the learning, the target features are extracted.

Step 4: Construct an optimal LSSVM-PSO classifier model to classify the actual vibration.

The extracted feature set of the training data is used to train the optimal classifier model, i.e., the LSSVM classifier and PSO algorithm. In which, the supervised learning algorithm is applied to train the model. Accordingly, the target data can then be obtained by the trained LSSVM-PSO classifier model for identifying the target data.

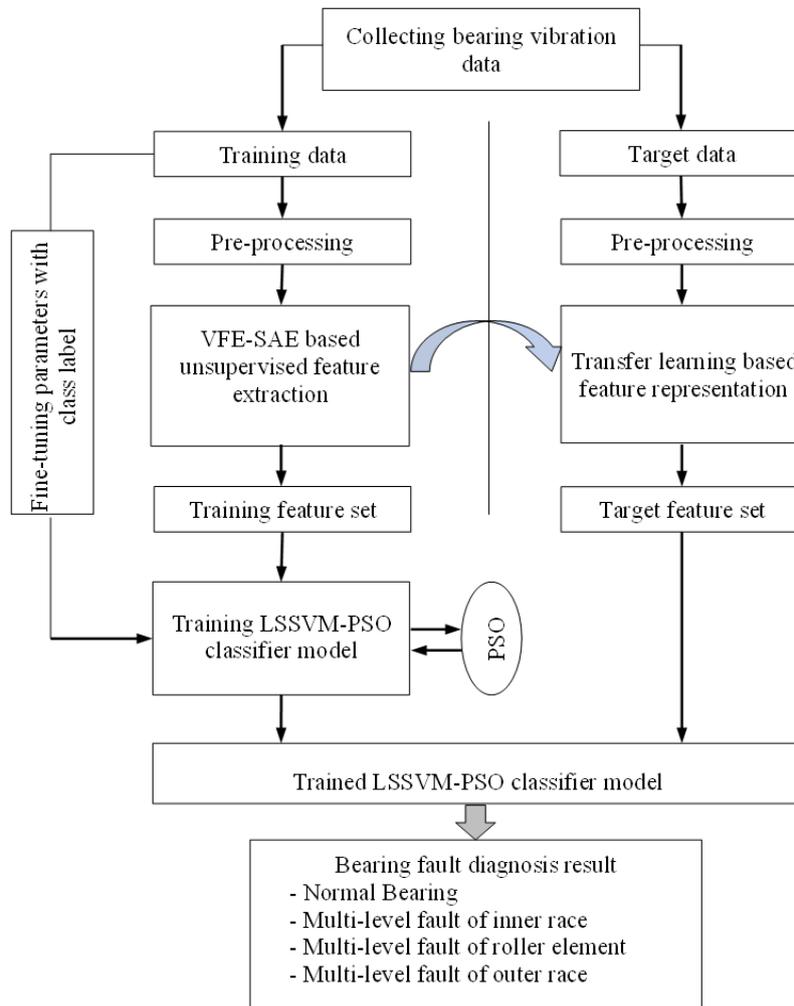


Fig.5. VFE-SAE - LSSVM-PSO diagnosis technique.

5. EXPERIMENTAL RESULTS AND ANALYSIS

A. Data acquisition

The experimental dataset was collected in the Bearing Data Center at Case Western Reserve University (Loparo, 2013) with the experimental setup model shown in Fig.6., which consists of a 2 HP reliance electric motor, a torque transducer, and a dynamometer. A placed DE roller bearing was used as the test bearing with a shaft speed of 1797 rpm and a sampling

frequency of 12 kHz. Four test conditions were examined: healthy bearing (HB), inner-race defect (IRD), outer-race defect (ORD) and roller element defect (RED), which are detailed in twelve classes. The collected time-domain vibration data for these bearing conditions is detailed in Table 2., and a sample of waveforms is depicted in Fig.7. for bearing statuses. As a result, 180 samples in total are acquired in the group with each bearing condition of 15 samples in length of 1024 sample points.

Table 2. Collection of vibration signal samples.

Bearing status	Number of samples	Defect size (inch)	ID class
Normal ($x_1 - x_{15}$)	15	-	1
Inner race fault 1 ($x_{16} - x_{30}$)	15	0.007	2
Inner race fault 2 ($x_{31} - x_{45}$)	15	0.014	3
Inner race fault 3 ($x_{46} - x_{60}$)	15	0.021	4
Inner race fault 4 ($x_{61} - x_{75}$)	15	0.028	5
Outer race fault 1 ($x_{76} - x_{90}$)	15	0.007	6
Outer race fault 2 ($x_{91} - x_{105}$)	15	0.014	7
Outer race fault 3 ($x_{106} - x_{120}$)	15	0.021	8
Roller element fault 1 ($x_{121} - x_{135}$)	15	0.007	9
Roller element fault 2 ($x_{136} - x_{150}$)	15	0.014	10
Roller element fault 3 ($x_{151} - x_{165}$)	15	0.021	11

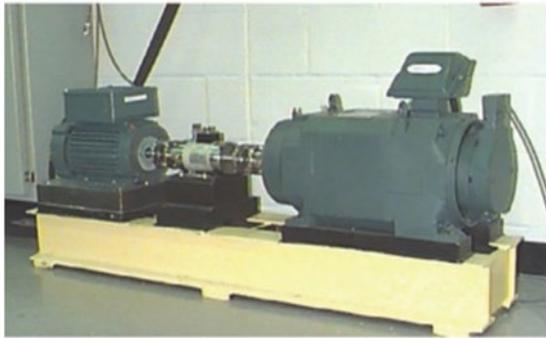


Fig.6. Schematic of the experimental setup.

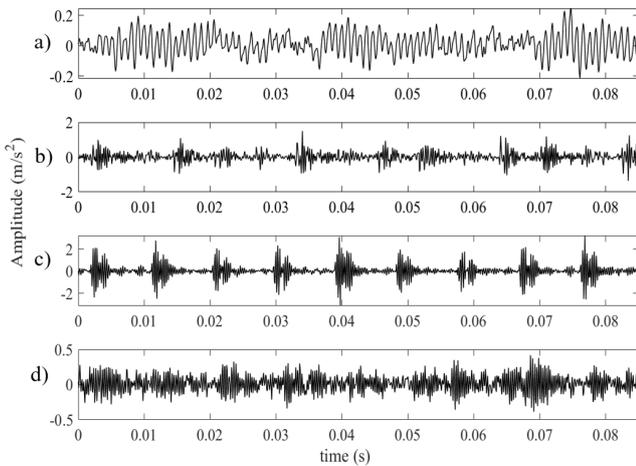


Fig.7. A sample of roller bearing vibration data at different statuses a) normal bearing, b) inner race fault, c) outer race fault, and d) roller element fault.

B. Result analysis and discussion

According to the purpose of bearing conditions diagnosis with multi-level fault based on target data, this study used 80 %, 60 %, and 40 % of the above collected vibration data to respectively demonstrate three states (I, II, and III) of the target datasets in the testing process.

To extract the vibration data features, three deep learning SAE networks were constructed with layer-wise training, including one hidden layer (1AE), two hidden layers (2AE), and three hidden layers (3AE). The deep learning SAE network parameters are detailed in Table 3. The vibration data feature sets were used to train and evaluate the proposed technique which were extracted by three different SAE networks. The final classification results demonstrate the capability of feature representation learning for the proposed VFE-SAE. Especially, for extensive vibration data, the vibration training and target datasets were pre-processed by FFT. An illustration for pre-processing vibration data is depicted in Fig.8. which presents the important characteristics of original vibration data. Additionally, we used transfer learning to represent the features of three target dataset states; the achieved diagnosis accuracy is high, as shown in Table 5. According to the results in Table 4., the time cost only pays with low price which is often computationally intensive in undertaking transfer learning. The obtained features can be then used to evaluate the classifier model.

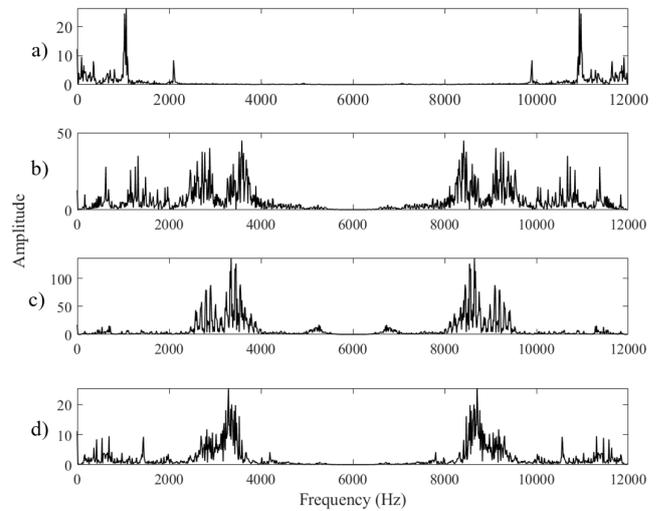


Fig.8. A sample of pre-processing data at different statuses a) normal bearing, b) inner race fault, c) outer race fault, and d) roller element fault.

Based on VFE-SAE, the extracted feature set of the training data can be used to build the LSSVM-PSO, which, in turn, can be used to identify the target data with the accurate results. Furthermore, to demonstrate capability of the proposed technique, which is the feature representation based on the VFE-SAE method and the LSSVM-PSO classifier, we also built three LSSVM-GA, LSSVM and feed forward neural network (FFNN) classifier models for comparing the diagnosis results. The target data-based diagnosis results are shown in Table 5. - Table 8., respectively.

Table 3. The parameters of deep learning SAE network.

	Hidden layer units	λ	β
1AE	20	0.01	6
2AE	10	0.001	4
3AE	10	0.001	4

Table 4. Target data feature representation time.

	Samples of target data	VFE-SAE (s)	Transfer learning (s)
State I	144	8.424	0.171
State II	108	7.410	0.140
State III	72	6.536	0.156

In these tables, the results show that pre-processing data is superior. The obtained classification results based on the original vibration data are lower than the pre-processed data, which is compared at the columns on the left side with the right side, respectively. This means that the original vibration data is inefficient for feature representation learning and classification. In contrast, the pre-processed data revealed clear local structures in the original data and the purity, which generated significant classification results. Furthermore, the implementation of transfer learning plays an important role in extracting significant features, which are then used to obtain

highly accurate diagnosis results, higher than 97 % accuracy for 80 % target data of Stage I. Thus, in comparison with other techniques, the proposed VFE-SAE - LSSVM-PSO diagnosis technique obtained more accurate classification results, as shown in Table 5.

From Table 5., it is evident that the combination of VFE-SAE (high-level extracted features) and LSSVM-PSO (optimization) generates the superior classification results at all three stages of target data. At Stage III, the maximum efficiency for VFE-SAE–LSSVM-PSO is 97.94 % for 2AE architecture, for 1AE and 3AE architecture, it is 97.09 % and 96.20 %, respectively. In light of this information, it is evident that the deep learning SAE network with two layers of wise training is constructed for extracting the high-level features of bearing vibration data and classification accuracy of optimal model. There is a noticeable problem for the 3AE architecture-based VFE: The obtained classification accuracy results are lower than other 1AE and 2AE architectures for all three stages of target data. This is probably due to the fact that the features were generated in a large number of hidden layers and were generally invariant to local changes in the data. Each extracted feature contains more abstract of the original vibration signal. More information can be found in [17]. Finally, the proposed VFE-SAE - LSSVM-PSO technique is very efficient in terms of VFE-SAE feature representation and LSSVM-PSO classification model, which is illustrated in Fig.9.

Table 5. Classification accuracy results of LSSVM-PSO (%).

	Using original vibration data			Using pre-processed data		
	1AE	2AE	3AE	1AE	2AE	3AE
Stage I	85.57	85.26	85.55	97.10	97.76	96.46
Stage II	86.41	87.09	87.24	95.08	95.74	94.03
Stage III	85.64	84.32	86.72	97.09	97.94	96.20

Table 6. Classification accuracy results of LSSVM-GA (%).

	Using original vibration data			Using pre-processed data		
	1AE	2AE	3AE	1AE	2AE	3AE
Stage I	85.02	85.24	84.79	96.58	96.24	96.04
Stage II	86.17	85.13	86.13	94.48	95.07	94.48
Stage III	86.34	86.06	88.24	96.58	96.22	96.58

Table 7. Classification accuracy results of LSSVM (%).

	Using original vibration data			Using pre-processed data		
	1AE	2AE	3AE	1AE	2AE	3AE
Stage I	17.93	18.98	17.36	86.81	87.96	84.72
Stage II	25.31	23.46	19.14	89.50	91.66	89.50
Stage III	28.70	25.46	15.74	86.57	85.18	85.65

Table 8. Classification accuracy results of FFNN (%).

	Using original vibration data			Using pre-processed data		
	1AE	2AE	3AE	1AE	2AE	3AE
Stage I	12.97	11.57	13.20	75.47	75.03	59.73
Stage II	9.90	11.10	12.03	63.90	60.17	52.80
Stage III	10.63	7.43	9.23	78.23	74.10	56.03

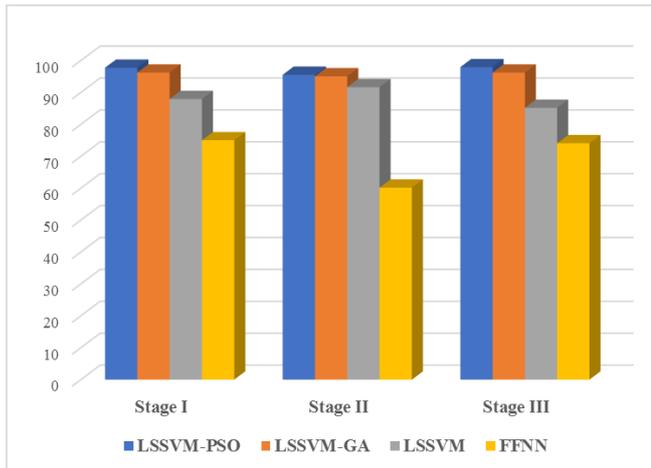


Fig.9. Comparisons of classification accuracy results.

6. CONCLUSIONS

In this paper, the high-level features were extracted from vibration data by deep learning SAE network constructed with two layers of wise training. The optimized SAE architecture with an activation function was used to exploit the vibration features, which improved the separation capability in classifying the vibration data. Transfer learning was applied to the feature extraction of vibration target data. Additionally, the optimal LSSVM-PSO classifier model was then constructed to fine-tune the extracted training feature set and classify the vibration target data. Experiments were conducted to obtain roller bearing vibration data with multi-level faults under four conditions. The proposed VFE-SAE - LSSVM-PSO diagnosis technique outperformed the other classifier models. In particular, the extracted features based on VFE-SAE were highly accurate of classification. The result was compared to other feature extraction methods. The discriminative ability of extracted features also demonstrated the effective performance of LSSVM-PSO. In future work, we would like to improve the feature representation learning (based on deep learning networks) to identify the random target fault status of roller bearings without training data or under different conditions.

ACKNOWLEDGMENT

This research is supported by National Key Research and Development Program of China (2016YFF0203400) and the National Natural Science Foundation of China (51575168 and 51875183). The authors also would like to thank the support from the Department of Mechanical Engineering, Hanoi University of Industry, Vietnam.

REFERENCES

- [1] Chen, Z., Deng, S., Chen, X., Li, C., Sanchez, R.-V. Qin, H. (2017). Deep neural networks-based rolling bearing fault diagnosis. *Microelectronics Reliability*, 75, 327-333. <https://doi.org/10.1016/j.microrel.2017.03.006>
- [2] Hong, L., Dhupia, J.S. (2014). A time domain approach to diagnose gearbox fault based on measured vibration signals. *Journal of Sound and Vibration*, 333 (7), 2164-2180. <https://doi.org/10.1016/j.jsv.2013.11.033>
- [3] Wang, X., Zheng, Y., Zhao, Z., Wang, J. (2015). Bearing fault diagnosis based on statistical locally linear embedding. *Sensors*, 15 (7), 16225-16247. <https://doi.org/10.3390/s150716225>
- [4] Li, B., Chow, M.-Y., Tipsuwan, Y., Hung, J.C. (2000). Neural-network-based motor rolling bearing fault diagnosis. *IEEE Transactions on Industrial Electronics*, 47 (5), 1060-1069. <https://doi.org/10.1109/41.873214>
- [5] Cheng, J., Yu, D., Yang, Y. (2006). A fault diagnosis approach for roller bearing based on EMD method and AR model. *Mechanical Systems and Signal Processing*, 20 (2), 350-362. <https://doi.org/10.1016/j.ymsp.2004.11.002>
- [6] Cheng, J., Zheng, J., Yang, Y. (2012). A nonstationary signal analysis approach - The local characteristic-scale decomposition method. *Zhendong Gongcheng Xuebao/Journal of Vibration Engineering*, 25 (2), 215, DOI: 10.16385/j.cnki.issn.1004-4523.2012.02.002.
- [7] Cheng, J., Yang, Y., Yang, Y. (2012). A rotating machinery fault diagnosis method based on local mean decomposition. *Digital Signal Processing*, 22 (2), 356-366. <https://doi.org/10.1016/j.dsp.2011.09.008>
- [8] Lei, Y., Lin, J., He, Z., Zuo, M.J. (2013). A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mechanical Systems and Signal Processing*, 35 (1-2), 108-126. <https://doi.org/10.1016/j.ymsp.2012.09.015>
- [9] Nguyen, V.H., Cheng, J.S., Yu, Y., Thai, V.T. (2019). An architecture of deep learning network based on ensemble empirical mode decomposition in precise identification of bearing vibration signal. *Journal of Mechanical Science and Technology*, 33, 41-50. <https://doi.org/10.1007/s12206-018-1205-6>
- [10] Martinez, A.M., Kak, A.C. (2001). PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23 (2), 228-233. <https://doi.org/10.1109/34.908974>
- [11] Prieto-Moreno, A., Llanes-Santiago, O., Garcia-Moreno, E. (2015). Principal components selection for dimensionality reduction using discriminant information applied to fault diagnosis. *Journal of Process Control*, 33, 14-24. <https://doi.org/10.1016/j.jprocont.2015.06.003>
- [12] Nguyen, V.H., Cheng, J.S., Thai, V.T. (2017). An integrated generalized discriminant analysis method and chemical reaction support vector machine model (GDA-CRSVM) for bearing fault diagnosis. *Advances in Production Engineering & Management*, 12 (4), 321-336. <https://doi.org/10.14743/apem2017.4.261>
- [13] Zhang, L., Zhang, Q., Zhang, L., Tao, D., Huang, X., Du, B. (2015). Ensemble manifold regularized sparse low-rank approximation for multiview feature embedding. *Pattern Recognition*, 48 (10), 3102-3112. <https://doi.org/10.1016/j.patcog.2014.12.016>

- [14] Yao, B., Zhen, P., Wu, L., Guan, Y. (2017). Rolling element bearing fault diagnosis using improved manifold learning. *IEEE Access*, 5, 6027-6035. <https://doi.org/10.1109/ACCESS.2017.2693379>
- [15] Bengio, Y. (2009). Learning deep architectures for AI. *Foundations and Trends® in Machine Learning*, 2 (1), 1-127. <http://dx.doi.org/10.1561/22000000006>
- [16] Hou, P., Wen, C., Dong, D. (2017). Rolling bearing fault diagnose based on stacked sparse auto encoder. In *2017 36th Chinese Control Conference (CCC)*. IEEE, 7027-7032. <https://doi.org/10.23919/ChiCC.2017.8028463>.
- [17] Bengio, Y., Courville, A., Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35 (8), 1798-1828. <https://doi.org/10.1109/TPAMI.2013.50>
- [18] Bengio, Y. (2012). Deep learning of representations for unsupervised and transfer learning. *JMLR: Workshop and Conference Proceedings*, 27, 17-37.
- [19] Dou, D., Zhou, S. (2016). Comparison of four direct classification methods for intelligent fault diagnosis of rotating machinery. *Applied Soft Computing*, 46, 459-468. <https://doi.org/10.1016/j.asoc.2016.05.015>
- [20] Kankar, P.K., Sharma, S.C., Harsha, S.P. (2011). Fault diagnosis of ball bearings using machine learning methods. *Expert Systems with Applications*, 38 (3), 1876-1886. <https://doi.org/10.1016/j.eswa.2010.07.119>
- [21] Patel, J.P., Upadhyay, S.H. (2016). Comparison between artificial neural network and support vector method for a fault diagnostics in rolling element bearings. *Procedia Engineering*, 144, 390-397. <https://doi.org/10.1016/j.proeng.2016.05.148>
- [22] Yu, Y., YuDejie, Junsheng, C. (2006). A roller bearing fault diagnosis method based on EMD energy entropy and ANN. *Journal of Sound and Vibration*, 294 (1-2), 269-277. <https://doi.org/10.1016/j.jsv.2005.11.002>
- [23] Ben Ali, J., Fnaiech, N., Saidi, L., Chebel-Morello, B., Fnaiech, F. (2015). Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals. *Applied Acoustics* 89, 16-27. <https://doi.org/10.1016/j.apacoust.2014.08.016>
- [24] Suykens, J.A.K., Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural Processing Letters*, 9, 293-300. <https://doi.org/10.1023/A:1018628609742>
- [25] Liu, X., Bo, L., Luo, H. (2015). Bearing faults diagnostics based on hybrid LS-SVM and EMD method. *Measurement*, 59, 145-166. <https://doi.org/10.1016/j.measurement.2014.09.037>
- [26] Zhang, Y., Qin, Y., Xing, Z., Jia, L., Cheng, X. (2013). Roller bearing safety region estimation and state identification based on LMD-PCA-LSSVM. *Measurement*, 46 (3), 1315-1324. <https://doi.org/10.1016/j.measurement.2012.11.048>
- [27] Nguyen, V., Hoang, T.D., Thai, V., Nguyen, X. (2019). Big vibration data diagnosis of bearing fault base on feature representation of autoencoder and optimal LSSVM-CRO classifier model. In *2019 International Conference on System Science and Engineering (ICSSE)*. IEEE, 557-563. <https://doi.org/10.1109/ICSSE.2019.8823332>
- [28] Yunlong, Z., Peng, Z. (2012). Vibration fault diagnosis method of centrifugal pump based on EMD complexity feature and least square support vector machine. *Energy Procedia*, 17 (Part A), 939-945. <https://doi.org/10.1016/j.egypro.2012.02.191>
- [29] Su, Z., Tang, B., Liu, Z., Qin, Y. (2015). Multi-fault diagnosis for rotating machinery based on orthogonal supervised linear local tangent space alignment and least square support vector machine. *Neurocomputing*, 157, 208-222. <https://doi.org/10.1016/j.neucom.2015.01.016>
- [30] Budiman, A., Fanany, M.I., Basaruddin, C. (2014). Stacked Denoising Autoencoder for feature representation learning in pose-based action recognition. In *2014 IEEE 3rd Global Conference on Consumer Electronics (GCCE)*. IEEE, 684-688. <https://doi.org/10.1109/GCCE.2014.7031302>
- [31] Eberhart, R., Kennedy, J. (1995). A new optimizer using particle swarm theory. In *Proceedings of the Sixth International Symposium on Micro Machine and Human Science (MHS '95)*. IEEE, 39-43. <https://doi.org/10.1109/MHS.1995.494215>

Received October 10, 2021

Accepted March 28, 2022