The International Journal of Engineering and Science (IJES)
|| Volume || 7 || Issue || 12 Ver.II || Pages || PP 01-08 || 2018 ||
ISSN (e): 2319 – 1813 ISSN (p): 23-19 – 1805

Fused feature representation based on bottleneck sparse autoencoder for multi-source data in gear fault diagnosis

VanTrong Thai[1,2], JunSheng Cheng[1], VietHung Nguyen[2], DuyTrinh Nguyen[1,2], PhuongAnh DaoThi[3]

[1]College of Mechanical and Automotive Engineering, Hunan University, Changsha, China
[2]Faculty of Mechanical Engineering, Hanoi University of Industry, Hanoi, Vietnam
[3]Faculty of Information Technology, University Economic and Technical Industries, Hanoi, Vietnam
Corresponding author: VanTrong Thai

ABSTRACT

In this paper, we construct a deep learning network based on sparse autoencoder and optimal SVM classifier to diagnosis the gear fault conditions. The proposed spare autoencoder (SAE) is used to exploit the fused features of multi-source vibration data which is represented at the bottleneck of SAE (BtSAE). The fused features are then used to train and identify the gear fault condition. a new diagnosis technique for multi-level fault gear. This is a new approach method applied in diagnosing gear faults. The experimental results prove that the proposed method operates highly effective and mostly feasible for identifying gear faults in practice. By applying this method, the results will be more accurate and will shorten the time cost.

KEYWORDS; *Vibration Feature Extraction; Deep Learning Network; Gear; Autoencoder, Fault Diagnosis*

Date of Submission: 20-12-2018 Date of acceptance: 04-01-2019

I. INTRODUCTION

Gear and bearing components plays an important role in many of the industrial rotating and transport machinery applications. Early fault diagnosis of gear and bearings may prevent unnecessary failures of most of the rotating machinery system and there by increase operational reliability and availability of machine. Fault diagnosis techniques are important for monitor the conditions in bearing and gear [1, 2]. Currently available fault diagnosis techniques have a variety of limitations. An effective and method has to be researched and automated system has to be developed for industrial machinery component health diagnostic activities.

Deep learning has been gradually applied to image processing and feature extraction in the field of machine learning in recent years. It is an unsupervised learning, through learning a deep nonlinear network structure, achieving essential features from a small number of data sets [3]. Hinton proposed a greedy unsupervised learning method to optimize the weight of the deep network, making it a good solution to solve the problem of training the multilayer neural networks [4]. Le et al. used the sparse autoencoder algorithm to establish high-level facial feature detectors from unlabeled data sets [5]. Coates et al. showed that the number of neuronsin the hidden layer of a deep network may be more important than the feature learning algorithms and the depth of the model [6]. In addition to its profound theoretical research, deeplearning has also been successfully applied in practice [7, 8].Most current classification algorithms are the shallow structure, and one popular method of them is the support vector machines (SVM) [9]. SVM was firstly proposed by Vapnik in 1995. It is based on Vapnik-Chervonenkis (VC) dimension of statistical learning theory and structural risk minimization (SRM) principle. It has the advantage in solving the small size samples and nonlinear problems in pattern recognition [10]. But for large-scale training sets, the cost ofSVM becomes too high. Therefore, Platt made the sequential minimal optimization (SMO) algorithm to solve the problem of large training samples [11]. The suitable design of thetraining algorithm for large-scale data has become an important part of the SVM study.

Due to the advantage of the deep network in processing large-scale data, we propose a new bottleneck of SAE support vector machine (BtSAE-SVM) to solve this problem of SVM. It uses multiple layers of sparse autoencoder to build a deep network for feature learning and combine this network with SVM to classify the input data. It is able to extract multifeatures. In general, higher levels of features can better reflect the nature of the data, which is more conducive for classification. Several experiments show that this method can achieve better performance in recognition, especially improving the capacity of SVM to process large amounts of data. With the proposed method applied to feature extraction is an important stage for classification of vibration data, and the extracted features may increase the separation between similar classes, resulting in improved classification performance.

II. METHODOLOGIES

Pre-processing of vibration signal is required in this research by using fast Fourier transform (FFT) method. Transformed data reveals the transient events/ shocks defined as the mechanical disturbances of fault. Based on this data the significant features are extracted to make the best of diagnosis accuracy result.

1. Bottleneck Sparseautoencoder (BtSAE)

Sparse Autoencoder (SAE) is a particular type of neural network architecture which works as an unsupervised learning algorithm. An autoencoder consists of three layers: input layer, hidden layer, and output layer organized to two stages of the encoder and decoder as shown in **Fig. 1**. In which, the input layer $f = \{f_1, f_2, \dots, f_n\}$, and the 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 through connected which is trained to replicate its input at its output.

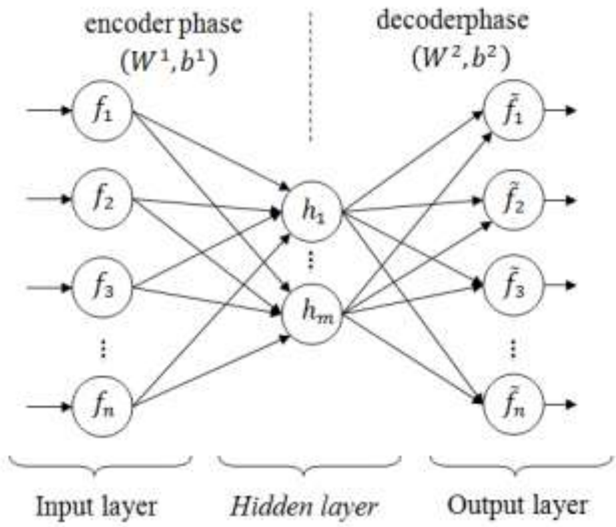

Fig. 1. Illustration of deep autoencoding

The encoder stage accomplished the feature representation from the high-dimensional input $f = \{f_1, f_2, \dots, f_n\}$ to low-dimensionality data in the hidden layer, $h = \{h_1, h_2, \dots, h_m\}(m \ll n)$. As in the mapping, the input and hidden layers are connected by the feed-forward activation function $h = \boldsymbol{Sigmoid}(W^{(1)}.f + b^{(1)})$ in which $W^{(1)}$ is weight matrix and $b^{(1)}$ is bias vector. In other words, each input of vector f is transformed into hidden representation h which compactly expressed the features of the input. On the contrary, the decoder stage was implemented to reconstruct the input f. The input data$h$is back mapped the output data $\tilde{f}$ with high-level feature representation. The activation function $\tilde{f} = \boldsymbol{Sigmoid}(W^{(2)}.l + b^{(2)})$ is used to connect the h data with $\tilde{f}$ data in this stage. In this stage, the weight matrix $W^{(2)} = \left(W^{(1)}\right)^T$ is referred to as tied weights and the $b^{(2)}$ is the bias vectorin the decoder stage.

The autoencoder is optimized in constructing it with the $\left(W^{(1)}, W^{(2)}, b^{(1)}, b^{(2)}\right)$ parameter set, which aim at the minimizing of the reconstruction error at the output. The cost function is used as follow:

$$C = \frac{1}{N}\sum_{n}^{N}\sum_{i}\left(f_i^{(n)} - \tilde{f}_i^{(n)}\right)^2 + \lambda * \Omega_W + \beta * \Omega_S \quad (1)$$

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

$$\Omega_W = \sum_{k}\sum_{i}\sum_{j}\left(W_{ij}^{(k)}\right)^2 \quad (2)$$

$$\Omega_S = \sum_{k}\sum_{j} KL\left(\rho \parallel p_j^{(k)}\right) \quad (3)$$

DOI: 10.9790/1813-0712020108

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 by **Eq. (4).**

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

It can be seen that each AE is trained independently, the feature data is extracted from the autoencoder in the hidden layer's nodes which contain most of the important information of the input which accomplishes as the mining information data. The obtained feature data can sever the next autoencoder as input in which higher-level feature representation is generated.

2. Optimal SVM Classifier

The SVM is a type of machine learning technique. The SVM relies on the theory of statistical learning. The SVM is handling the training samples as the input to a higher-dimensional characteristic space through the use of a mapping function ϕ^7 . Assuming that there is a given set of the training samples $G = \{(x_i, y_i), i = 1,2,...,l\}$ in which each sample $x_i \in R^d$ belongs to a class by $y \in \{+1,-1\}$, and the training data is not linearly separable in the space of feature, then the target function can be expressed as follows [12]:

Minimize $\phi(\omega) = \frac{1}{2}\langle \omega.\omega \rangle + C\sum_{i=1}^{l} \xi_i$

subject to $y_i(\langle \omega.\phi(x_i)\rangle + b) \geq 1 - \xi_i$, $\xi_i \geq 0$

$$i = (1,2, \dots, l) \tag{5}$$

in which ω is the normal vector of the hyperplane, C is the penalty parameter, b is the bias, $\xi_i$is non-negative slack variables, and $\phi(x)$ is the mapping function.

By introducing a set of Lagrange multipliers $\alpha_i \geq 0$, the optimization problem could be rewritten as:

Maximize

$$L(\omega, b, \alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j K(x_i x_j) \tag{6}$$

subject to $0 \leq \alpha_i \leq C$, $\sum_{i=1}^{l} \alpha_i y_i = 0$

The function of making decision can be achieved as:

$$f(x) = \operatorname{sgn}\left[\sum_{i=1}^{l} \alpha_i y_i K(x_i x) + b\right] \tag{7}$$

The SVM method used the radial basis kernel function which is the most common kernel function as it indicates in the bellow equation:

$$K(xx_i) = \exp(-\|x - x_i\|^2 / 2\sigma^2 \tag{8}$$

whereσ is the kernel parameter.

3. Proposed Technique Based On Btsae And Optimal Svm Classifier

As a method of feature learning, SAE can effectively learn the original expressions of data, and sparse expression is more effective than others. After learning feature, a good classifier is also needed. SVM is one of the best classifiers. Combining these two methods, a stacked sparse autoencoder SVM (BtSAE-SVM) model is proposed, obtaining better performance. Its structure is designed as **Fig. 2**.

A stacked sparse autoencoder is a neural network consisting of multiple layers of sparse autoencoders in which the outputs of each layer is wired to the inputs of the successive layer. A good way to obtain good parameters for a stacked sparse autoencoder is to use greedy layer-wise training. Briefly, the main idea is to train the layers of the network one at a time, so that first train a network with one hidden layer, and only after that is done, train a network withtwo hidden layers, and so on. At each step, take the old network with *k-1* hidden layers, and add an additional N-th hidden layer. Training can be unsupervised in an autoencoder.

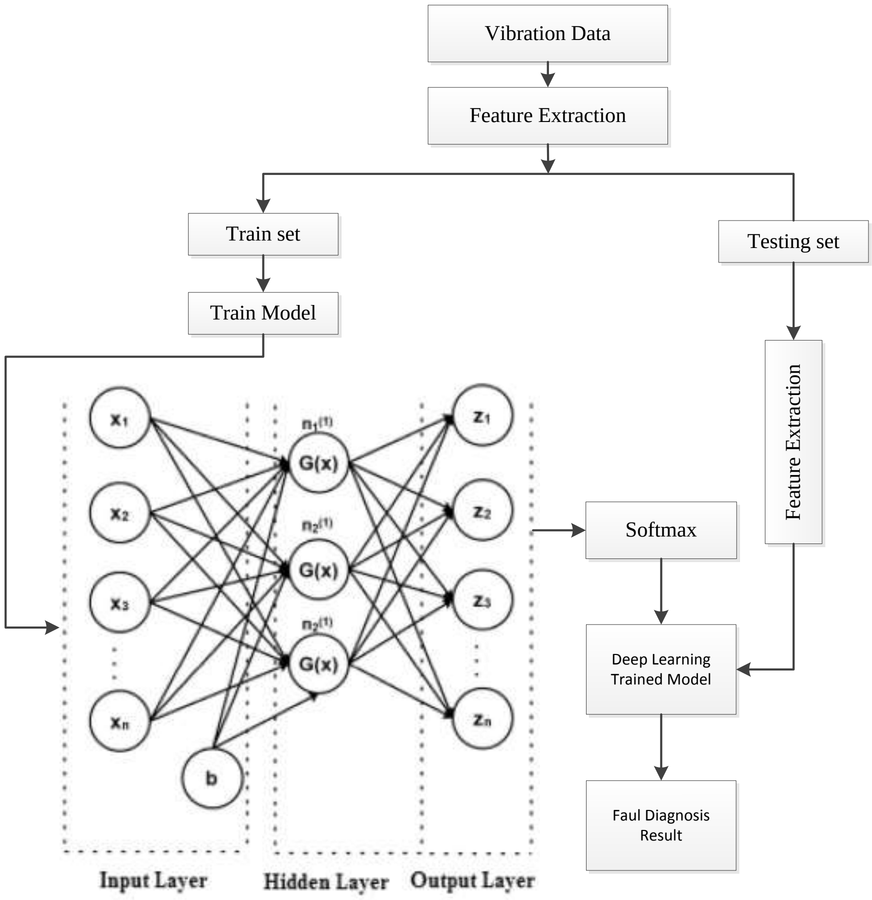

Fig.2. Stacked autoencoder model used in the study

To give a concrete example, suppose we wished to train a stacked autoencoder with two hidden layers for large-scaled data sets which is shown in Fig. 2. First, train a sparse autoencoder on the raw inputs $x^{(k)}$ to learn primary features $h_1^{(k)}$on the raw input. Next use these primary features as the raw input to another sparse autoencoder to learn secondary features $h_2^{(k)}$ on these primary features. At last, treat these secondary features as raw input to a SVM classifier, training it to map secondary features to data labels. Finally, combine all three layers together to form a stacked sparse autoencoder with two hidden layers and a final SVM classifier layer capable of classifying the large-scaled data sets as desired.

If the number of features is more, there is a choice to increase the number of hidden layers [12]. The accuracy of the classification is relatively high, so this design is relatively reasonable.

III. EXPERIMENTAL RESULTS AND ANALYSIS

3.1. Data acquisition

The experimental dataset was collected in the Gear Data Center at Case Western Reserve University (Loparo, 2013) with the experimental setup model shown in **Figure 3.** The conditions are tested in the six classes including healthy gear status, chip gear tooth, broken gear tooth. The collection of vibration signals in the time-domain of gear statuses are described in **Table 1.** As a result, 120 samples totally are acquired in the group with each gear status of 20 samples in length of 4096 sample points. In which, testing the IR fault and RE fault collected more defect conditions than the OR fault due to them being hardly identified in the small defect for research purposes.

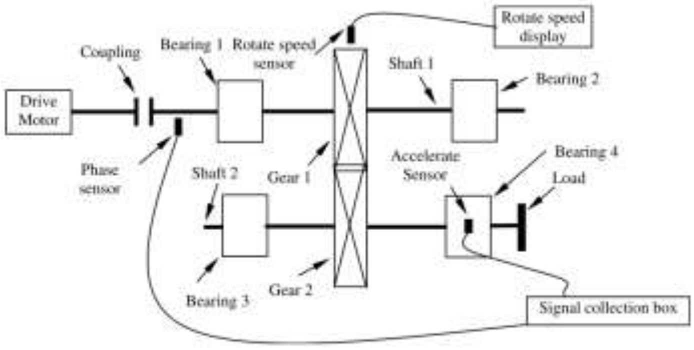

Fig. 5. Test rig.

Fig.3. Schematic of the experimental setup

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

Table 1 Collection of vibration signal samples

Gear status	Number of samples	Defect size (inch)	ID class
Normal $(x_1 - x_{20})$	20	-	1
Chip gear-tooth $(x_{21} - x_{40})$	20	0.028	2
Broken gear-tooth$(x_{41} - x_{60})$	20	0.028	3

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

3.2. Proposed Fault Diagnosis Technique based on Deep Learning Architecture

In this section, a complete fault diagnosis technique for the gear based on constructing the deep network is presented. The special combination of the unsupervised feature self-learning based autoencoder and the supervised feature learning based softmax classifier can acquire effectiveness in the diagnosis results. Its implementation is based on the high-dimensionality feature data extracted from vibration signal by EEMD method. This diagnosis technique can be described in seven steps as below and the desired process is shown bellow:

Step 1: Vibration signal acquisition.
Step 2: EEMD method decomposes the vibration signal into IMFs set and the residual.
Step 3: Feature extraction: the first several IMFs of every original signal is extracted from the features including five parameters in time-domain and three parameters in frequency domain, which formed the high-dimensionality feature data. This feature data is divided into the training set and testing set, and the training set is then used for constructing the deep feature learning which will be evaluated by the testing set.
Step 4: The unsupervised feature self-learning of each autoencoder layer is repeated to initialize the desired number of autoencoders:

Train the first autoencoder to minimize some form of the input reconstruction error. The hidden neuron outputs of the autoencoder are now used as input for the next autoencoder, also trained to be an autoencoder. Reiteration is continued until the last autoencoder.

Step 5: Train softmax classifier using the hidden layer output of last autoencoder as lower-dimensionality input. The DLA is fixed in stacking autoencoders and softmax classifier.
Step 6: Fine-tune the parameters of this DLA with respect to the class' label of supervised criterion.
Step 7: The trained deep network architecture, which identify its own actual fault statuses, belongs.

IV. RESULT ANALYSIS AND DISCUSSION

As the results presented, the collecting six classes of vibration signal samples in the gear fault statuses shown in **Table 1** section 2, totaling 120 samples, are divided into a training-testing partition, namely 90 samples in the training set and 30 samples in testing set. For every original vibration signal, we firstly use EEMD to decompose into the IMFs and then calculate the eight statistical features related to **Eqs.** (3) to (10) of the first five IMFs. The 40-elements feature vector presents the most fault information detailed in **Table 2**.

The high-dimension feature set harassesthe effect of the diagnosis method as the accuracy can be declined to a certain degree and the time is consumed for the training process. As the great result of high-dimensionality reduction, the low-dimensionality feature set is gained at the hidden layer of the final autoencoder, which is then used to train and evaluate the softmax classifier. In this study, we use two autoencoders and their parameters are shown in **Table 3.**

1	Rms_1	11	Kr_2	21	Cr_3	31	Cf_4
2	Sk_1	12	Im_2	22	Stf_3	32	Rvf_4
3	Kr_1	13	Cr_2	23	Cf_3	33	Rms_5
4	Im_1	14	Stf_2	24	Rvf_3	34	Sk_5
5	Cr_1	15	Cf_2	25	Rms_4	35	Kr_5
6	Stf_1	16	Rvf_2	26	Sk_4	36	Im_5
7	Cf_1	17	Rms_3	27	Kr_4	37	Cr_5
8	Rvf_1	18	Sk_3	28	Im_4	38	Stf_5
9	Rms_2	19	Kr_3	29	Cr_4	39	Cf_5
10	Sk_2	20	Im_3	30	Stf_4	40	Rvf_5
$i = 1, 2, \dots, 5$ is number of IMFs							

Table 2 The features of vibration signal

NO.	NUMBER OF NEURONS AT LAYERS	λ	β
Autoencoder 1	40, 20, 40	0.05	6
Autoencoder 2	20, 10, 20	0.05	4

Table 3 Parameters of Autoencoder

To demonstrate the superiority of the autoencoder in the feature dimensional reduction, we made a comparison with the PCA method to evaluate its dimensionality reduction performance on the original feature data. We used the first four features, in the reduced feature set, to present the reduction quality. It can be seen that there are overlaps of classes and features that are still scattered on the large space in the PCA method, which has a poor classification result. Meanwhile, the reduced feature set obtained from the last autoencoder represented the superior ability in the clear and concentrated distribution. There are significant features for classifying it accurately.

The gained 10-dimensionality feature set in the last autoencoder—based on the original 40-dimensionality feature set—is used to train the Softmax classifier.The probability of the class label taking on each of the k different values of the possible label is estimated related. To express the classifier effect, based on the probability of using the class label, we made the comparison between the two architecting of deep learning. The complete DLA, which is fine-tuned the parameters, and the DLA is not used in the fine-tuning of the parameters. The experimental results are shown in **Table 4**.

The results in **Table 5** showed that the classification accuracy obtained a perfect 100% on the complete DLA fine-tuned parameters, and the accuracy on the DLA excluded the stage of fine-tuning parameters just obtained at 86.7%. In which case, the results in classes 2 and 4 highlighted that our proposed technique can help to accurately identify the small defects that occurred on the chip and broken element of the test gear. These defects are known to be difficult to recognize by the traditional classification model. The reason is due to the fine-tuning phase of the complete DLA having to respect the supervised criterion aiming at minimizing the classification error.

CLASS	TESTING DATA	DLA IS NOT FINE-TUNED	DLA IS FINE-TUNED
1	$x_{16} - x_{20}$	1(5)	1(5)
2	$x_{36} - x_{40}$	2(3)1(1)4(1)	2(5)
3	$x_{56} - x_{60}$	3(5)	3(5)
4	$x_{76} - x_{80}$	4(4)6(1)	4(5)
5	$x_{96} - x_{100}$	5(5)	5(5)
6	$x_{116} - x_{120}$	6(4)2(1)	6(5)
Time (s)		10.1712	10.9972
Accuracy (%)		86.7	**100**

Table 4 Diagnosis accuracy results based on DLAs

CLASS	TESTING DATA	K-NN	FNN	Softmax
1	$x_{16} - x_{20}$	1(5)	1(4)4(1)	1(5)
2	$x_{36} - x_{40}$	2(4)6(1)	2(5)	2(4)4(1)
3	$x_{56} - x_{60}$	3(5)	3(5)	3(5)

4	$x_{76} - x_{80}$	4(0)1(3)2(1)6(1)	4(5)	4(3)6(2)
5	$x_{96} - x_{100}$	5(5)	5(5)	5(5)
6	$x_{116} - x_{120}$	6(3)1(1)2(1)	6(3)4(2)	6(5)
Time (s)		0.1507	2.5844	1.8215
Accuracy (%)		73.3	90	90

Table 5 Diagnosis accuracy results based on autoencoder and sofmax classifier

To compare the classification results with the other classifiers, we used the shallow classifiers as the K-nearest neighbor (K-NN) classifier, feed-forward neural network (FNN) classifier, and the Softmax classifier to classify the gear fault pattern, respectively. The original 40-dimensional feature data is used to accomplish the training and testing processes of classifiers. A K-NN classifier with the number of nearest neighbor of 4; a feedforward neural network classifier with hidden layer of 10, output layer of 6, and training error goal of 0.01; and a softmax classifier with maxima epochs of 1000 are established. The evaluation result of K-NN classifier, FNN classifier and softmax classifier are shown in **Table 5**. The evaluation showed that the classification accuracy of these classifiers is all lower than the classification accuracy of the proposed DLA. It is known that, the gear fault diagnosis technique, based on the proposed DLA, effected the high-level feature representation in the deep learning to gain the high classification accuracy. Nevertheless, the perfection of the proposed DLA in the classification accuracy results paid the higher price for time than the other shallow architectures as the results showed in **Table 3, 4.** The cause of this problem is that each phase in constructing deep learning is always optimized.

NUMBER	HIDDEN LAYER UNITS	λ PARAMETER	Β PARAMETER
1AE	20	0.01	6
2AE	10	0.001	4
3AE	10	0.001	4

Table3. The parameters of deep learning SAE network

Based on VFE-SAE method, the extracted feature set of training data is used to build the SVM-PSO which aimed at identification of target data with the accurate results. And to demonstrate capability of the proposed technique that are the feature representation of VFE-SAE method and the SVM-PSO classifier, we also built three SVM-GA, SVM and feedforward neural network (FFNN) classifier models for comparing the diagnosis results. The target data diagnosis results shown in the **Tables 5**, respectively.

NO.	SAMPLES OF TARGET DATA	VFE-SAE	TRANSFER LEARNING
State I	128	8.424	0.171
State II	82	7.410	0.140
State III	52	6.536	0.156

Table4. Feature representation time of target vibration data

GEAR CONDITION	TRAINING SAMPLES	TEST SAMPLES	AVERAGE COST TIME (S)	AVERAGE TEST ERRORS (%)
NORMAL	45	20	7.65	0.910
CHIP GEAR TOOTH	45	20	28.69	2.693
BROKEN GEAR TOOTH	45	20	15.38	3.872

Table5. Gear fault diagnosis cost time and average error

IV. CONCLUSIONS

A model using a deep learning algorithm to improve SVM was proposed in this paper. The BtSAE-DLN method built a deep network and extracted features by multiple layers of sparse autoencoder from the input data, then combined with SVM to classify the data, effectively improving the classification rate of data sets. The results of eight experiments have showed that this model has a better performance in classification compared with the classical DLN, even when the data set has larger size or much more features, because this model can express the characteristic of the data using multiple layers. Future work will include the improvement of this model by studying the number of hidden layers in the network. We will also apply this model to some tasks, including image processing and classification.

REFERENCE

[1] L. S. Dhamande, and M. B. Chaudhari, "Compound gear-bearing fault feature extraction using statistical features based on time-frequency method," *Measurement,* vol. 125, pp. 63-77, 2018.
[2] L. S. Dhamande, and M. B. Chaudhari, "Detection of Combined Gear-Bearing Fault in Single Stage Spur Gear Box Using Artificial Neural Network," *Procedia Engineering,* vol. 144, pp. 759-766, 2016.
[3] J. Guan *et al.*, "Polynomial dictionary learning algorithms in sparse representations," *Signal Processing,* vol. 142, pp. 492-503, 2018.
[4] M. Längkvist, L. Karlsson, and A. Loutfi, "A review of unsupervised feature learning and deep learning for time-series modeling," *Pattern Recognition Letters,* vol. 42, pp. 11-24, 2014.
[5] L. D. G. E. Hinton, D. Yu, "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups," *Signal Processing Magazine IEEE,* vol. 29, 2012.
[6] L. S. G. W. Taylor, D. J. Fleet, and G. E. Hinton, "Dynamical binary latent variable models for 3d human pose tracking," *Conference on Computer Vision and Pattern Recognition IEEE*, 2013.
[7] L. Badino *et al.*, "Integrating articulatory data in deep neural network-based acoustic modeling," *Computer Speech & Language,* vol. 36, pp. 173-195, 2016.
[8] M. Cai, and J. Liu, "Maxout neurons for deep convolutional and LSTM neural networks in speech recognition," *Speech Communication,* vol. 77, pp. 53-64, 2016.
[9] B. Richhariya, and M. Tanveer, "EEG signal classification using universum support vector machine," *Expert Systems with Applications,* vol. 106, pp. 169-182, 2018.
[10] T. D. Pham, "Possibilistic nonlinear dynamical analysis for pattern recognition," *Pattern Recognition,* vol. 46, no. 3, pp. 808-816, 2013.
[11] X. Huang, L. Shi, and J. A. K. Suykens, "Sequential minimal optimization for SVM with pinball loss," *Neurocomputing,* vol. 149, pp. 1596-1603, 2015.
[12] V. N. Vapnik, "The nature of statistical learning theory," *Springer, New York, (1995).*

VanTrong Thai. "Fused feature representation based on bottleneck sparse autoencoder for multi-source data in gear fault diagnosis" The International Journal of Engineering and Science (IJES), 7.12 (2018): 01-08