Gearbox Fault Diagnosis Classification with Empirical Mode Decomposition based on Improved Long Short-Term Memory

Sheng-Nan Chen
School of Computer and Information Technology
Beijing Jiaotong University
Beijing, China
e-mail: snchen@bjtu.edu.cn

Feng Liu
School of Computer and Information Technology
Beijing Jiaotong University
Beijing, China
e-mail: fliu@bjtu.edu.cn

Chang-Xia Gao
School of Computer and Information Technology
Beijing Jiaotong University
Beijing, China
e-mail: cxgao@bjtu.edu.cn

Jinyang Li
Department of Electrical Engineering
Columbia University in the New York City
Beijing, China
e-mail: jl5173@columbia.edu

***Abstract*—As the main part of the Rotary Machinery System, the health of the gearbox's internal gears and bearings are essential when the machine is running. It is necessary to analyze the vibration data of the gearbox healthy state in time, find the fault, exact fault locations and types. Some researchers are only taking advantage of Long Short-Term Memory (LSTM) to deal with fault diagnosis gearbox dataset, which has a poor precision actually. In this paper, a deep neural network which combines Empirical Mode Decomposition (EMD) method, Long Short-Term Memory (LSTM) model and Particle Swarm Optimization (PSO) algorithm to achieve a high precision rate of machine fault diagnosis. Compared with existing methods, the proposed method is faster on training and more accurate. Firstly, original sensor data is pre-trained with EMD, then EMD's output as input of the LSTM to identify the gearbox fault types. At the same time, the PSO algorithm optimizes the LSTM's hyper-parameter automatically to avoid the problem that random initialization makes the network fall into local optimum. In the meanwhile, the Root Mean Square Error (RMSE) is established for better performance the EMD-PSOLSTM model. Therefore, the proposed method can learn a robust and discriminative representation from the raw gearbox dataset. Compared with other machine learning methods, such as Back Propagation (BP) and Support Vector Machine (SVM), the proposed method shows state-of-the-art results on gearbox dataset, and it is effective and efficient for gearbox fault diagnosis.**

Keywords-gearbox; Long Short-Term Memory (LSTM), Particle Swarm Optimization (PSO), Root Mean Square Error (RMSE), Empirical Mode Decomposition (EMD), fault diagnosis.

I. INTRODUCTION

As one of the most essential part in Rotary Machinery System for industrial and machinery equipment, the gearbox is worthy for study. Once the gear breaks down, it will be leading to corrosion during machine operation. Therefore, intelligent fault diagnosis system for vibration signals is significant, it contains three main parts generally: data acquisition, feature extraction and selection, and fault diagnosis.

However, traditional intelligent fault diagnosis techniques still have some limitations obviously as follows:

1) Signal processing techniques and diagnosis expertise require much priori knowledge, while conventional fault diagnosis methods are based on handcrafted to select features, which is time consuming and labor intensive.

2) Most existing methods are almost domain-specific and cannot timely updated or adapted well on new fault diagnosis domains due to limitations of the structure, thus it is highly desirable for a general-purpose approach, which is not an easy work.

3) Compared with other classification tasks, manual features are specific, that means it is unsuitable for some circumstances of other scenarios, and it cannot achieve a high precision prediction in all situations.

Alternatively, as a powerful feature learning ability for intelligent fault diagnosis, deep learning provides efficient operational scheme and promising results. Deep learning methods are able to identify high-grade data representation and feature classification by non-linear information processing units of multiple stacked layers in a hierarchical architecture. In more recent years, as a powerful tool, deep learning successfully applied to all kinds of domains, such as health informatics [1], particle physics [2], natural language processing [3], computer vision [4], and speech recognition [5]. Different deep models such as stacked denoising autoencoders [6], deep belief network [7], convolutional neural networks (CNNs) [8], sparse filtering [9], recurrent neural network [10], organic chemistry [21], games [22] and so on. Compared to handcraft selected features of traditional methods, deep learning methods show a promising performance in diagnosis areas.

In terms of Deep Learning models, it has been used to fault diagnosis. For instance, Wen et al. [19] proposed the method of three-layer sparse auto-encoder and deep transfer learning (DTL) to extract original data features for induction motor diagnosis and achieved higher prediction accuracies. The

result is achieved as 99.82%. Xu et al. [20] designed a deep transfer convolutional neural network (TCNN) framework and achieved the desired diagnostic accuracy within limited training time to successfully address the online and offline issues. To improve diagnostic performance, Zhao et al. [13] combined a deep residual network and dynamic weighted wavelet coefficients (DRN+DWWC) method.

Long-Short-Term Memory (LSTM) as a deep learning method has been used to diagnose the research on machine fault. For instance, Fu proposed convolutional neural network (CNN) combined with the long short-term memory network (LSTM) to establish the logical relationship between observed variables and verify the high practicability and generalization [11]. Based on actual time-varying non-stationary operating conditions, Cao designed a deep bi-directional Long Short-term Memory (DB-LSTM) and achieved a higher accuracy than other four existing diagnosis methods [12]. To reduce the covariate shift problem, Zhao et al. [17] proposed a batch normalization procedure of the LSTM.

For obtain the high prediction accuracy, more and more researchers use the hybrid model to classify the fault diagnosis nowadays due to the limitation of the single model. To gain a better prediction result on the non-linear and non-stationary characteristic, Tian et al. [13] proposed a hybrid ACLCD-PSOLSTM model. Kim and Cho [14] designed the particle swarm optimization (PSO), which can automatically determine different kinds of hyperparameters in the CNN-LSTM network and achieve the perfect prediction performance.

In this paper, we design a hyper-parameter training and selection method by using PSO algorithm to optimize the model of EMD-LSTM. EMD is used to extract local characteristics by decomposing complex signals into multiple simple components. LSTM is performing well in dealing with sequence data. PSO is a population-based heuristic algorithm as a simulate social behavior algorithm that formats and expresses the movement of animal species living in large colonies like birds, ants and fish. The summarize of this paper's main contributions is as follows.

1) We proposed a framework of machine fault diagnosis based on EMD-PSOLSTM model. Empirical Mode Decomposition (EMD) has the decomposition capacity to improve the accuracy. The characteristic diagnosis datasets have nine types and two working conditions. Because of each component has the various depth effect on the performance, we select the first five relative effective components.

2) An LSTM solves the continue time problems and extracts the temporal features of time-series data. We divide millions of gearbox fault data into 2000 data each group, then use LSTM to indicate the fault types of every group classification.

3) PSO algorithm optimize hyper-parameters, which is a best method to avoid the problem of getting into local minimum. The matching performance is determined by the number of hidden units in LSTM, the layer units of the LSTM and the learning rate. Particles are according their external information to update their positions and directions. There are highly optimized performance and the training speed of the LSMT model with PSO, the result is higher 6.3% compared with the hyperparameters are not fine-tuning.

The structure of this paper is organized as follows: The theoretical background of the EMD, LSTM and PSO algorithms represents in Section II. Then Section III describes the proposed approach. The experimental verification is demonstrated in Section IV. Finally, Section V is the summary of the conclusion and the discussion of the future works.

II. THEORETICAL BACKGROUND

A. *Empirical Mode Decomposition (EMD)*

The Empirical Mode Decomposition (EMD), as a method for analyzing non-linear and non-stationary signal processing technique, which was proposed by Huang et al. [15] in 1998. EMD is a breakthrough of wavelet and Fourier analysis. According to the time scale characteristics of high signal-to-noise ratio and self-adaptation, EMD is used to analysis and process the non-stationary and non-linear signals. Based on local characteristic of different time scales signals, EMD decomposes the complicated signals into a limited number of nearly orthogonal and complete Intrinsic Mode Function (IMFs). Each IMF component has a specific signal by a discrete frequency vibration mode. The IMF components show the signal's internal features, which is the reason why EMD strategy is widely used to diagnosis the fault and feature extraction. Fig. 1 shows the different conditions of the gearbox vibration signals. Obviously, it is hard to identify the fault types just by intuition. As is shown in Fig. 2, an original signal sample is decomposed by EMD into six IMFs and a residual. It is more clearly to identify the different features. Then select the first five components as the optimal input of the LSTM.

The original time series sample data is set as $x(t)$. First, obtain the local maxima and minima values, then the fitting operation is carried out to obtain the upper envelope $x_{max}(t)$ and lower envelope $x_{min}(t)$ sequences of $x(t)$, and finally produce the upper and lower envelopes average value as the input sequence, the mean is designated as $M_1(t)$:

$$M_1(t) = \frac{x_{max}(t)+x_{min}(t)}{2} \quad (1)$$

The first component $I_1^1(t)$ is the difference between $x(t)$ and $M_1(t)$.

$$I_1^1(t) = x(t) - M_1(t) \quad (2)$$

Generally, on condition that $I_1^1(t)$ satisfies the two requisite conditions of the IMF, it is regarded as the first IMF from the original signal. $I_1^1(t)$ is obtained by decomposition with the highest frequency component of the IMFs. Otherwise, it can be treated as the original sequences, and repeated the above steps (1) ~ (2) k times until the average curve approaches zero before stopping. The criterion is designated as $C_1(t)$.

$$sd = \sum_{t=0}^{T} \frac{\left|I_1^{k-1}(t) - I_1^k(t)\right|^2}{\left|I_1^{k-1}(t)\right|^2} \quad (3)$$

Repeat the above steps (1) ~ (3), the step (3) is used as the termination criterion. The result is usually reasonable when the value is maintained between 0.2 and 0.3. By subtracting $C_1(t)$ from $x(t)$, the highest frequency component can be

obtained by the residual sequence $r_1(t) = x(t) - C_1(t)$. The IMF components can be achieved by repeating the above screening process. It can be stopping the process when the residue $R_n(t)$ becomes a monotonic function, or the pre-error becomes larger than $C_n(t)$. The original sequences of the n IMFs and residual $r_n(t)$ can be shown as:

$$x(t) = \sum_{i=1}^{n} C_i(t) + r_n(t) \tag{4}$$

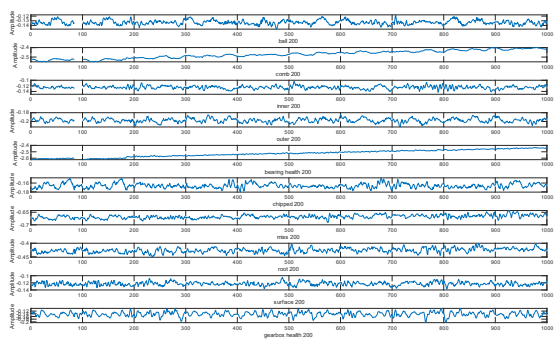


Figure 1. Vibration signal for different gearbox conditions.

(x-axis, y-axis show 1000 sampling points and the amplitude of signals respectively)

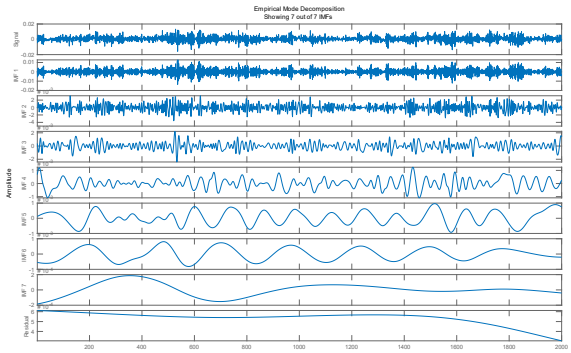


Figure 2. Decomposition of the first sample vibration by signal using EMD.

B. *Long-Short-Term Memory (LSTM)*

There are different machine learning approaches in the vibration signal prediction. Recurrent neural networks (RNNs), the information flows from each neuron to every other neuron in its layers. It's the extension of the traditional feedforward neural networks. However, the problem of the gradient vanishing and exploding limits the model's ability to learn long-term sequential data.

As a special kind of RNNs, Long-Short-Term Memory (LSTM) networks are capable of handling the problems of long-term dependence and the gradient disappearance [16]. It roots the memory unit (Memory Cell) into neural intercept point in the hidden layer of the recurrent neural network and realizes the recording of the historical information by adding (Input, Forget, Output) three kinds of gate structures. It can receive and delete cell state's information selectively.

$(x_1, x_2, \ldots x_t)$ is designed as the input sequence, and hidden layer's state is $(h_1, h_2, \ldots h_t)$, as shown in Fig. 3. At the t moment, the corresponding equations are given as follows:

$$i_t = \sigma(W_{hi} h_{t-1} + W_{xi} x_t) \tag{5}$$
$$f_t = \sigma(W_{hf} h_{t-1} + W_{xf} x_t) \tag{6}$$
$$\widetilde{C}_t = tanh(W_{hc} h_{t-1} + W_{xc} x_t) \tag{7}$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \widetilde{C}_t \tag{8}$$
$$O_t = \sigma(W_{xo} x_t + W_{ho} h_{t-1}) \tag{9}$$
$$h_t = O_t \odot \tanh(c_t) \tag{10}$$

Where i_t, f_t, O_t are gates of input, forget and output respectively, h_{t-1} means the output information of the hidden layer unit at the previous moment and h_t represents the current output information. Cell unit is designed as c_t. W_h, W_x, W_c denote the weight matrix of different connection layers respectively. σ and $tanh$ are sigmoid and tanh activation functions.

The determination of the new input gate information is added to the cell unit. The forget gate determines the cell retain or lost at the state of the previous moment. The update cell status is composed of two parts, they are the new input information at this time and the cell status at the last time, which are controlled by the above input and forget gate respectively. The hidden cell value about the final output is based on the status of the cell unit at the current time, while the output gate is determined the special output ratio. The cooperation of the three control gates and the cell unit make the LSTM network have the information memory function. When the input gate weight is zero, no information enters the cell unit; when the forget gate weight is zero, information on cell state is forgotten and cannot be transferred to the next moment. When output gate weight is zero, cell unit information cannot output. When input gate and output gate close weight are both zero, cell unit information is locked and passed to later moment via forget gate.

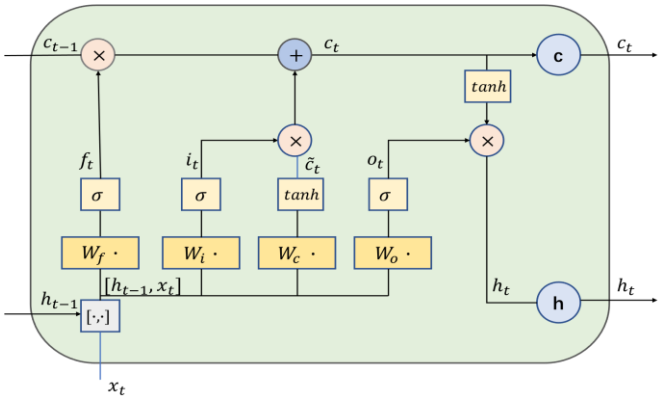


Figure 3. The structure of the LSTM memory block.

C. *Particle Swarm Optimization (PSO)*

PSO is a powerful tool for dealing with global optimization problems. It is a swarm intelligence optimization algorithm on the basis of birds foraging behaviors, which was first proposed by Eberhart and Kennedy in 1995. PSO

possesses several advantages, such as simplicity, robustness to control parameters, easily implemented, computational effect and efficiency. Each particle is influenced by two factors of its 'best' achieved position and the group 'best' position. All particles in populations are correspond to the fixed search space. Following the steps (11) ~ (12), two formulas are optimized by adjusting their speed and position constantly until they satisfy convergence termination conditions. Every iteration, according to the new velocity, each particle's position is changed as shown:

$$V_{i,j}^{t+1} = \omega V_{i,j}^{t} + c_1 r_{1,i,j}^{t}\left(\widehat{y_i^t} - x_{i,j}^{t}\right) + c_2 r_{2,i,j}^{t}\left(y_{i,j}^{t} - x_{i,j}^{t}\right) \quad (11)$$

$$x_{i,j}^{t+1} = x_{i,j}^{t} + V_{i,j}^{t+1} \quad (12)$$

Where $V_{i,j}^{t}$ denotes the velocity of particle i in the j dimension at tth iteration, $x_{i,j}^{t}$ represents the ith position, ω is the inertia weight, c_1 and c_2 are positive constants (learning factor). $\widehat{y_i^t}$ is the particle's individual best solution, $y_{i,j}^{t}$ is the extreme point in the t iteration of the particle swarm, $r_{1,i,j}^{t}$, $r_{2,i,j}^{t}$ are random parameters uniformly within [0,1], $V_{i,j}^{t} \in [-V_{max}, V_{max}]$, and V_{max} is constant, which encourage the search in all possible locations. Maximum and minimum velocity values also define the randomly distributed particles.

III. Proposed Approach

A new approach based on vibration signal analysis for rotating machinery faults diagnosis is presented, that the model is combined EMD method into PSOLSTM structure. First, decomposing the raw vibration signal data into some similar components by EMD and using the EMD's output as input of the LSTM to identify the gearbox fault types. In theory, designing LSTM process network, within a certain range, the more neurons and hyperparameters in the hidden layer, the higher prediction accuracy of the model. Nevertheless, as the amount of data collected, more challenges and slower computing can raise. Hence, an improved LSTM by PSO optimization strategy we proposed, which is performance well for the prediction.

Learning rate, hidden num and batch size determine the updating speed of the parameters. The parameters control the model training speed and converge. Likewise, the curve becomes more smoother by a small learning rate. During the training network, it will cause the speed too slow to converge, but the speed problem can be resolved by a large learning rate, however, the network may fall into local maximum and not stable easier. Therefore, it is desirable using PSO optimization method to avoid the above problems and improving the accuracy of the fault diagnosis. We proposed EMD-PSOLSTM approach flow chart in Fig. 4.

EMD-PSOLSTM Methodology

The main idea of this approach is to optimize LSTM network hyper-parameters via PSO algorithm, which effectively develop the prediction performance of LSTM network.

The following main steps are as follows:

Step 1: Data preprocessing. The original vibration signal is normalized by the map-min-max normalization, and it is used to limit the dataset in the interval value [-1,1], which is defined by the following equation:

$$y = (y_{max}\text{-}y_{min}) \times \frac{x\text{-}x_{min}}{x_{max}\text{-}x_{min}} + y_{min} \quad (13)$$

Where y denotes the standardized data after the normalization. y_{max} and y_{min} are the interval normalization values. x is the original vibration signal. The maximum value is x_{max} and the minimum value is x_{min} from the raw signal.

Step 2: EMD decomposition. First, EMD decomposes the pre-processed data into some IMF_s and $resiude$, then items IMF_s and $resiude$ are recombined into n components. The high frequency components manifest the mildly change of the curves, and the low frequency components manifest the amplitude. Between IMF_1 and IMF_5 are regarded as high frequency components, and the rest $n-5$ IMF_s and $resiude$ are as low frequency components. Therefore, we selected the first five representative samples as the input results of the LSTM, then determine structure parameters of the LSTM.

Step 3: Initialize the particle swarm parameters of LSTM network. PSO includes the number of iterations, number of layers, and the limited range of population size, learning factor, particle position and velocity. The random value is the initial value of particle velocity and position. The LSTM network structure initialization mainly refers to determine the mini-batch size, max-epochs, number of neurons and hidden layers in each layer of the network.

Step 4: Confirm the evaluation function of the particles. The fitness function of swarm particles is defined as:

$$fit_i = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i\text{-}y_i}{y_i}\right| \quad (14)$$

Where n indicates the number of populations, the sample output value is Y_i and the actual output value is y_i.

Step 5: Calculating each fitness value of particles and constructing the population rule tree structure. Each particle of the fitness value is calculated and sorted by the formula (14), then the particle swarm population of the rule tree structure is to construct.

Step 6: Update the local and global best positions of particles in real time.

Step 7: Update the speed and position of the particles themselves according to (11).

Step 8. It is stopped when the condition reached the end of the iteration (the maximum number of iterations). Otherwise, skip into step 3 and continue the iteration.

Step 9: Obtain the optimum results and assign the connection weights of the LSTM network, then train the LSTM prediction model. Finally, we use the 10-fold cross-validation to perform this model and output the optimal solution of time series prediction.

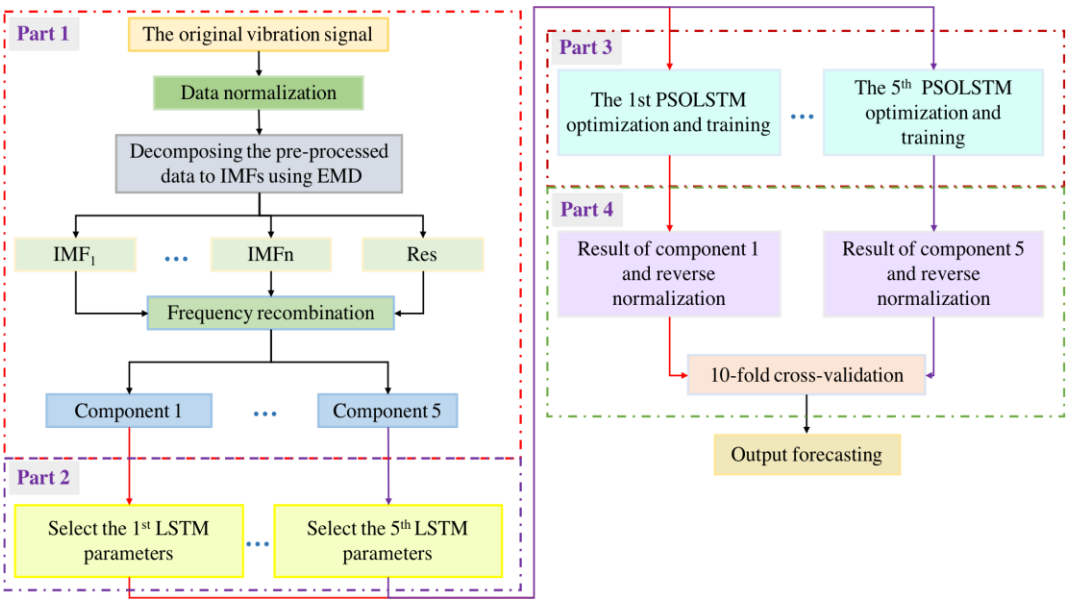
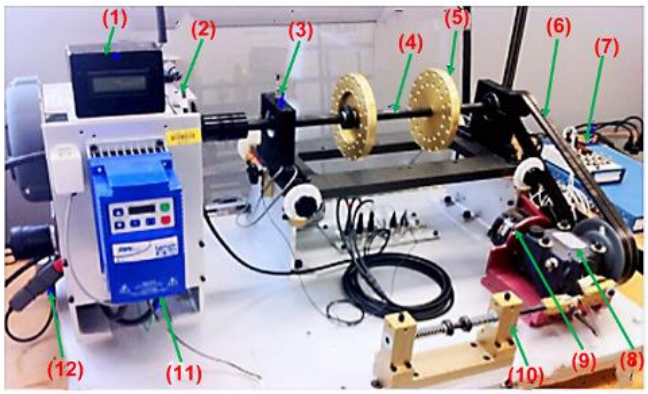


Figure 4. Flow chart of the proposed EMD-PSOLSTM approach.

IV. Experimental and Analysis

In this part, to verify the performance and effectiveness of the EMD-PSOLSTM method, we evaluate the EMD-PSOLSTM approach using the gearbox and bearing datasets. They are from the drivetrain dynamic simulator (DDS), which was obtained by SEU and collected by the research group of Pf. Ruqiang Yan, SouthEast University, China, as shown in Fig. 5. This dataset under the speed to be 20 HZ-0V and 30 HZ-2V based on two different working conditions.

TABLE I shows the different gearbox and bearing types. There are including one health state and four failure types which is a task classification with 5-class in gearbox dataset, and that include one health state and four failure types in bearings dataset. Therefore, there are the total one heath state and eight fault types with 9-class in vibration signals. Hence, we proposed the other three models for comparative experiments. The comparative models contain EMD-PSOLSTM, LSTM, Back Propagation Neural Network (BP), and support vector machines (SVM) models.



(1) Opera meter, (2) Induction Motor, (3) Bearing, (4) Shaft, (5) Loading Disc, (6) Driving Belt, (7) Data Acquisition Board, (8) Bevel gearbox, (9) Magnetic Load, (10) Reciprocating Mechanism, (11) Variable Speed Controller, (12) Current Probe[18]

Figure 5. Photo of experimental facility.

TABLE I. Description of Gearbox and Bearing Types under 20HZ-0V or 30HZ-2V [18].

Location	Label	Type	Description
Gearbox	1	Surface	Wear occurs in the surface of gear
	2	Miss	Missing one of feet in the gear
	3	Root	Crack occurs in the root of gear feet
	4	Chipped	Crack occurs in the gear feet
	9	Normal	Health
Bearing	5	Ball	Crack occurs in the ball
	6	Inner	Crack occurs in the inner ring
	7	Outer	Crack occurs in the outer ring
	8	Combination	Crack occurs in the both inner and outer ring
	9	Normal	Health

In the cases, the parameters of the values are designed in TABLE II.

TABLE II. PARAMETERS OF THE EMD-PSOLSTM METHOD.

Parameters	*Values*
LSTM input size	2000
Num-Hidden-Units	220
LSTM layers	5*1
Max-Epochs	5
Mini-Batch-Size	100,100,100
Learning rate	0.001

Each type vibration signal of planetary gearbox in three directions is selected 1048000 data respectively and there have 9 types. Therefore, there are total 62880000 data. We set 2000 data in each sample and each sample has five IMFs, so there are total 157200 cells. The training set has 141480 cells and the testing set has 15720 cells. In order to avoid the overfitting problems, we carry out the EMD-PSOLSTM approach to prove the performance by using the 10-fold cross validation. The gate activation function is the sigmoid and the state is the hyperbolic tangent function(tanh). In addition, the stability of the convergence of the PSO algorithm is significant important. Hence, we analyze the impact of the number of particles in the range on each iteration of the algorithm's convergence. As the number of particles is large enough, the results can avoid fall into local optimum. In this paper, we select 5 particles and 10 particles for 10 times simulation experiments, the results are shown in Fig. 6. From the figure we can see that when the number of the particles is 5, the curve of the result is unstable due to the small number of the particles. When the number of particles is increased to 10, all the curves of the results reached the optimal value can stable converge after 100 iterations. Therefore, the PSO algorithm strategy can help LSTM get ideal optimization results stably as long as the number of particles is large enough. Therefore, the population of the swarm is set to 10, the spatial dimension is defined as 3, the iteration of max number is set to 10, the range of the position parameters are set to [215, 225], [95, 105], [1, 10] respectively, and the velocity are set to [-1, 1]. The inertia weight is set to 0.5, the self-learning and the group learning are set to 2, 2 respectively. Through the PSO optimization, we selected the optimal results, that the input size is set as 2000, the num of hidden layer unit is set as 220, the max epoch is set as 5, the minibatch size is set as 100,100,100 respectively, and the optimal learning rate is set as 0.001.

Furthermore, to assess the manifestation of the EMD-PSOLSTM hybrid model, three common evaluation factors are employed. They are the Root Mean Square Error (RMSE), the Mean Absolute Error (MAE) and the Mean Absolute Percentage Error (MAPE), which are defined respectively as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{P_i}\text{-}\mathrm{T_i})^2} \tag{15}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|P_i - T_i| \tag{16}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{P_i - T}{T_i}\right| \times 100\% \tag{17}$$

Where N denotes the sample size, P_i means the predicted value and T_i means Test value.

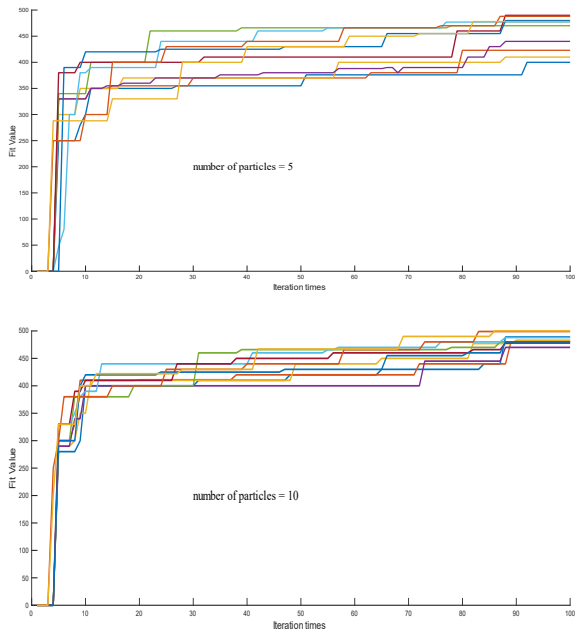


Figure 6. Convergence preformance under different particle numbers.

In TABLE III, the comparison results contain three indicators, that are RMSE, MAPE and MAE. The lower results, the test more effective. This table demonstrates the EMD-PSOLSTM model's performance outbalance the LSTM, BP and SVM models.

TABLE IV shows the experimental results. From this result, the proposed model of the fault classification accuracy outperforms better than the other three methods. The EMD-PSOLSTM approach has achieved approximate 6% accuracy than only using the LSTM, about 7% accuracy than BP and 9% than SVM. Moreover, in order to display a visual performance of the accuracy, the boxplot of the Fig. 7 shows the result. It indicates that among the EMD-PSOLSTM, the LSTM, the BP, the SVM four models, the LSTM, the BP and the SVM have a lower accuracy, the LSTM and the SVM are not stable than the EMD-PSOLSTM. The BP deviates greatly from the maximum and the minimum values. The EMD-PSOLSTM achieves the highest average accuracy.

In Fig. 8, we evaluate the performance by the root mean square error (RMSE) using the 10-fold cross validation. Obviously, the boxplot demonstrates that the EMD-PSOLSTM has a lower error and stable performance compared with other models.

TABLE III. THE ERROR RESULTS OF FOUR MODELS FOR THE VIBRATION SIGNAL.

Models	*RMSE (mm/s)*	*MAPE (mm/s) (%)*	*MAE (m)*
EMD-PSOLSTM	0.4598	1.7349	0.0464
LSTM	1.1089	8.4519	0.2704
BP	1.2915	4.6214	0.3751
SVM	1.3357	14.4095	0.408

TABLE IV. TEN-FOLD CROSS-VALIDATION AND THEIR AVERAGE ACCURACY FOR FOUR DIFFERENT MODELS.

Model	*Each time the value of the accuracy (%)*										*Average*
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	9	10	
EMD-PSOLSTM	96.97	96.55	97.75	97.09	97.32	98.55	98.47	95.67	98.5	97.53	97.44
LSTM	91.6	95.99	90.9	90.97	91.54	90.97	87.6	91.09	89.38	91.28	91.132
BP	92.17	86.276	93.51	93.67	86.342	92.33	93.09	91.24	85.41	91.08	90.5118
SVM	87.2	88.0	94.4	87.67	88.4	86.0	87.0	89.33	86.0	87.6	88.16

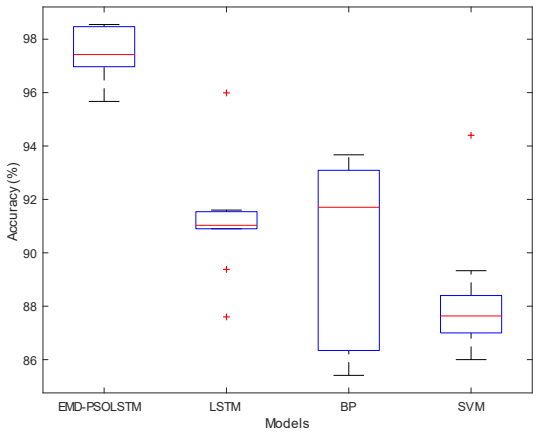


Figure 7. The accuracy in the boxplot of different models.

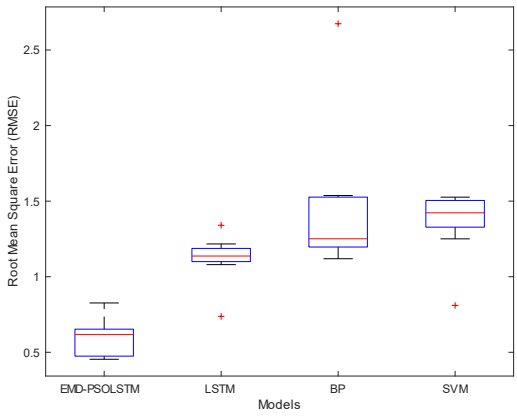


Figure 8. The Root Mean Square Error (RMSE) in the boxplot for four different models.

So as to indicate the EMD-PSOLSTM performance in fault classes, the confusion matrix is shown in Fig. 9, which consists of misclassification error percentage and classification accuracy percentage. As the result, the num of 15331 correct samples with the total 15720 testing labeled set samples. Compared with the different types, it is clearly that classification accuracies about the proposed method are all up to 97% except label 2 and label 4. The result was validated the proposed approach has an effect and efficiency performance of the fault classification.

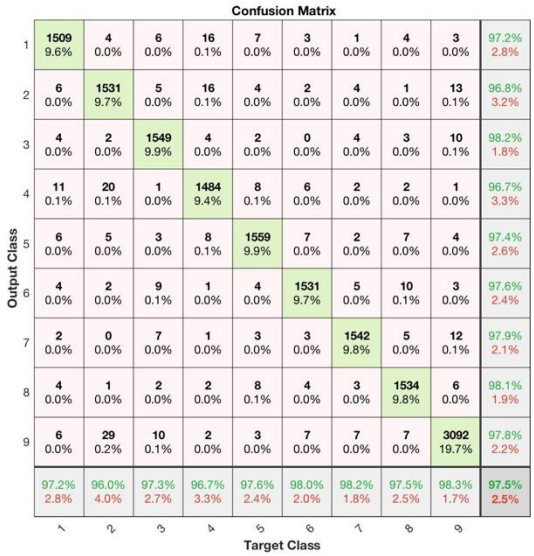


Figure 9. The confusion matrix of the EMD-PSOLSTM method

V. CONCLUSION

Fault diagnosis classification plays an essential part in rotary machinery system. This paper, we propose a hybrid deep learning approach based on the EMD, the LSTM and the

PSO models to diagnose and classify the fault types. By comparison with other three models (LSTM, BP, SVM), the proposed approach has achieved a better performance, the higher stability and the faster training speed. It has to verify the effectiveness of the new hybrid approach and get state-of-the-art results by using the gearbox dataset. Further research will combine with transfer learning to the real machinery system scenario as well as other fault diagnosis domains.

ACKNOWLEDGMENT

This work was supported in part by the Research on the Standard of Data Acquisition and Processing of EMU Remote Operation and Maintenance Service Program of China under Grant no. BZYJ2018-03.

REFERENCES

[1] A. H. H. M. Mohamed, H. Tawfik, L. Norton and D. Al-Jumeily, "Does e-Health technology design affect m-Health informatics acceptance? A case study," Proceedings of 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics, Hong Kong, 2012, pp. 968-971.

[2] L. Wu et al., "Calibration and Status of the 3-D Imaging Calorimeter of DAMPE for Cosmic Ray Physics on Orbit," in IEEE Transactions on Nuclear Science, vol. 65, no. 8, pp. 2007-2012, Aug. 2018.

[3] Q. Xie, X. Zhou, J. Wang, X. Gao, X. Chen and C. Liu, "Matching Real-World Facilities to Building Information Modeling Data Using Natural Language Processing," in IEEE Access, vol. 7, pp. 119465-119475, 2019.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Proc. Int. Conf. Neural Inf. Process. Syst., 2012, pp. 1097–1105.

[5] F. U. Khan, B. P. Milner and T. Le Cornu, "Using Visual Speech Information in Masking Methods for Audio Speaker Separation," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 26, no. 10, pp. 1742-1754, Oct. 2018.

[6] F. N. Khan and A. P. T. Lau, "Robust and efficient data transmission over noisy communication channels using stacked and denoising autoencoders," in China Communications, vol. 16, no. 8, pp. 72-82, Aug. 2019.

[7] L. Xiuli, Z. Xueying and W. Liyong, "Fault Diagnosis Method of Wind Turbine Gearbox Based on Deep Belief Network and Vibration Signal," 2018 57th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), Nara, 2018, pp. 1699-1704.

[8] Z. Zhao et al., "Exploring Deep Spectrum Representations via Attention-Based Recurrent and Convolutional Neural Networks for Speech Emotion Recognition," in IEEE Access, vol. 7, pp. 97515-97525, 2019.

[9] W. Xiong, Q. He, K. Ouyang and Z. Peng, "Combining Spatial Filtering and Sparse Filtering for Coaxial-Moving Sound Source Separation, Enhancement and Fault Diagnosis," in IEEE Access, vol. 7, pp. 25150-25162, 2019.

[10] T. de Bruin, K. Verbert and R. Babuška, "Railway Track Circuit Fault Diagnosis Using Recurrent Neural Networks," in IEEE Transactions on Neural Networks and Learning Systems, vol. 28, no. 3, pp. 523-533, March 2017.

[11] J. Fu, J. Chu, P. Guo and Z. Chen, "Condition Monitoring of Wind Turbine Gearbox Bearing Based on Deep Learning Model," in IEEE Access, vol. 7, pp. 57078-57087, 2019.

[12] L. Cao, Z. Qian, H. Zareipour, Z. Huang and F. Zhang, "Fault Diagnosis of Wind Turbine Gearbox Based on Deep Bi-Directional Long Short-Term Memory Under Time-Varying Non-Stationary Operating Conditions," in IEEE Access, vol. 7, pp. 155219-155228, 2019.

[13] M. Zhao, M. Kang, B. Tang and M. Pecht, "Deep Residual Networks With Dynamically Weighted Wavelet Coefficients for Fault Diagnosis of Planetary Gearboxes," in IEEE Transactions on Industrial Electronics, vol. 65, no. 5, pp. 4290-4300, May 2018.

[14] T. Kim and S. Cho, "Particle Swarm Optimization-based CNN-LSTM Networks for Forecasting Energy Consumption," 2019 IEEE Congress on Evolutionary Computation (CEC), Wellington, New Zealand, 2019, pp. 1510-1516.

[15] Norden E. Huang, Zheng Shen, Steven R. Long, Manli C. Wu, Hsing H. Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung and Henry H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," Proc. R. Soc. Lond. A.454903–995, 1998.

[16] S. Hochreiter, J. Schmidhuber, "Long short-term memory, Neural computation", vol. 9, pp. 1735-1780, 1997.

[17] H. Zhao, S. Sun and B. Jin, "Sequential Fault Diagnosis Based on LSTM Neural Network," in IEEE Access, vol. 6, pp. 12929-12939, 2018, doi: 10.1109/ACCESS.2018.2794765.

[18] Siyu Shao, Stephen McAleer, Ruqiang Yan, Pierre Baldi. "Highly Accurate Machine Fault Diagnosis Using Deep Transfer Learning", IEEE Transactions on Industrial Informatics, 2019.

[19] L. Wen, L. Gao and X. Li, "A New Deep Transfer Learning Based on Sparse Auto-Encoder for Fault Diagnosis," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 49, no. 1, pp. 136-144, Jan. 2019.

[20] G. Xu, M. Liu, Z. Jiang, W. Shen and C. Huang, "Online Fault Diagnosis Method Based on Transfer Convolutional Neural Networks," in IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 2, pp. 509-520, Feb. 2020.

[21] C. Lo and K. Tang, "Blended Learning with Multimedia e-Learning in Organic Chemistry Course," 2018 International Symposium on Educational Technology (ISET), Osaka, 2018, pp. 23-25.

[22] Y. Jang and S. Ryu, "Exploring Game Leadership and Online Game Community," 2009 Conference in Games and Virtual Worlds for Serious Applications, Coventry, 2009, pp. 178-181.