Fault Diagnosis of Asynchronous Motors Based on LSTM Neural Network

Dengyu Xiao[1], Yixiang Huang[1], Xudong Zhang[1], Haotian Shi[1], Chengliang Liu[1], Yanming Li[1]
1: State Key Laboratory of Mechanical System and Vibration
Shanghai Jiao Tong University
Shanghai, PR China
xiaodengyu@sjtu.edu.cn

***Abstract*—Motors are one of the most critical components in industrial processes due to their reliability, low cost and robust performance. Motor failure will lead to the shutdown of a whole production line and cause great loss. Therefore, accurate, reliable and effective motor fault diagnosis must be performed. Currently, motors fault diagnosis has gained much attention to guarantee safe motor operations. In this paper, a novel fault diagnosis method is proposed for the three-phase asynchronous motor using Long Short-Term Memory (LSTM) neural network, which possesses the capacity to learn meaningful representations from raw signal without any feature engineering. Firstly, the acceleration signals of different fault motors were collected. Then, the raw data were directly fed into LSTM neural network to establish the relationship between the raw vibration signals and fault states. The whole proposition was experimentally demonstrated and discussed by carrying out the tests of six three-phase asynchronous motors under different fault conditions in the drivetrain diagnostics simulator system. Performances of other classification methods such as LR, SVM, MLP, and basic RNN, are tested and contrasted. Results show that the proposed approach achieves the highest fault diagnosis accuracy.**

Keywords—fault diagnosis; three-phase asynchronous motor; LSTM; health condition;

I. INTRODUCTION

Motors have been widely used as key machine components for the production of torque. Motors are exposed to a wide variety of environments and conditions. These factors, coupled with the natural aging process of any machine, make the motor subject to various faults. Any motor failure will cause unwanted downtime, expensive repair procedures, and even human casualties [1]. To ensure consistent and reliable operation of modern industrial systems, fault detection and diagnosis of motors has received considerable impetus from industry, leading to a spurt in research activity in industry and academia [2].

The most typical faults and their occurrence possibilities of motors from the statistical studies were conducted by the Institute of Electrical and Electronics Engineers (IEEE) and by Electric Power Research Institute (EPRI) through General Electric Corporation [3]. The study identified that faulty bearings are the most common induction motor fault (representing more than 40% of motor faults). This is followed by the stator fault (over 28%), rotor fault (over 8%), and other unspecified faults (more than 12%). Several motor fault diagnosis techniques such as vibration analysis, acoustic analysis [4], motor current analysis [5], electromagnetic field monitoring [6], and thermal analysis [7] have been developed in the last two decades for the detection and diagnosis of the motor faults. Vibration technique is one of the most frequently employed techniques for motor faults detection due to its ease of use, high accuracy, and reliability.

During recent years, many researches of motors fault diagnosis based on the data-driven approach have emerged. For instance, Talhaoui et al. [8] used a combined FFT and discrete wavelet transform (DWT) technique to evaluate faults due to broken rotor bars of an induction motor. Liang et al. [9] used a combination of power spectrum, cepstrum, bispectrum, and neural network for the data analysis and fault diagnosis of induction motors. Seera et al. [10] employed a hybrid computer model comprising a Fuzzy Min-Max (FMM) neutral network and a Classification and Regression Tree (CART). Zhao, et al. [11] used the wavelet analysis method obtain the energy ratio and train the model with optimized support vector machine (SVM). Pan, et al. [12] improved the reliability of motors using a combination of entropy feature extraction, mutual information, and support vector machine. José, et al. [13] conducted a multiple-fault diagnosis in induction motors through SVM classification at variable operating conditions using signatures from the frequency domain characteristics. Shrinathan, et al. [14] detected the fault in the bearing of a three-phase induction motor by the frequency selection in the stator-current spectrum and SVM classifier. The more recent use of artificial neural network (ANN) along with some other techniques for identification of various faults have been reported in various articles [2, 15, 16]. Vilas, et al. [2] used an optimal MLP neural network classifier for fault detection of three-phase induction motor. Soualhi, et al. [15] used an improved artificial ant clustering technique with multiple features arising from transformations made on current and voltage signals to detect the faults in induction motor. Wang, et al. [16] used short time Fourier transform (STFT) to obtain the corresponding time-frequency map and classified the fault motors by Convolutional Neural Network (CNN). These techniques are moderately successful in accurately identifying types of fault. However, all these methods do not belong to sequence models, which can obtain the intrinsic sequential characteristic directly behind sensory data. Because it can be obviously observed that these papers share the same procedure in data processing: feature extraction and selection. The methods

This project is supported by the National Key Technology R&D Program of China (No. 2017YFB1302004), and NSFC project (No. 51305258).

2166-5656/18/$31.00 ©2018 IEEE
DOI 10.1109/PHM-Chongqing.2018.00098

in feature engineering can be fast Fourier transform (FFT), short time Fourier transform, wavelet transform, Hilbert-Huang transform and so on. The feature engineering is generally time consuming and difficult, which can easily lead to false diagnosis if the selected features do not meet the requirement. To cover these shortages, deep learning, which is thought capable of discovering useful high-order feature representations, as well as the relevance of raw signals [17], is taken into consideration to skip the feature engineering. Therefore, in this paper, Long Short-Term Memory (LSTM) neural network, with the capacity to capture long-term dependencies and model sequential data, is directly used for the raw vibration data to learn meaningful representations without any feature engineering.

With the development of deep learning methods in the last few years, recurrent neural network (RNN) is one of the most state-of-art models for a variety of sequence classification and prediction problems in recent years. Long Short-Term Memory (LSTM) is a significant branch of RNN, capable of learning long-term dependencies [18][19]. LSTM is proposed to relief the problem of gradient exploding or vanishing in RNN, has emerged as one popular architectures to handle sequential data with various applications including image captioning, speech recognition, genomic analysis and natural language processing [20][21][22]. LSTM is able to address sequences of varying-length data and capture long-term dependencies. In the research field of machine health assessment, LSTM also proves to be effective. Yuan et al. investigated LSTM model for fault diagnosis and prognostics of aero engine [23]. Zhao et al. presented an empirical evaluation of LSTMs-based machine health monitoring system in the tool wear test [24]. Malhotra proposed a very interesting structure for RUL prediction. They designed a LSTM-based encoder-decoder structure, which LSTM-based encoder firstly transforms a multivariate input sequence to a fixed-length vector and then, LSTM decoder uses the vector to produce the target sequence [25]. However, so far LSTM has not been applied in fault diagnosis for asynchronous motors. In this paper, LSTM is firstly used for fault diagnosis of three-phase asynchronous motors.

This paper is organized as follows. In Section II, the basic theory of RNN and LSTM is reviewed. Then, the experiment of six three-phase asynchronous motors under different fault conditions in the drivetrain diagnostics simulator system is introduced in Section III. Then in the following Section IV, the specific process of the proposed approach is presented and the performances of other classification methods are tested and contrasted. Finally, concluding remarks are provided in Section V.

II. BASIC THEORY OF LSTM

A. Basic theory of RNN

Recurrent neural network (RNN) is the deepest of all neural networks, which can generate and address memories of arbitrary-length sequences of input patterns [26]. Compared with a feed-forward NN, a RNN is strengthened by a time-step edge that introduces a notion of time to the NN model. And edges connecting adjacent steps, called recurrent edges, form cycles that are self-connections of a neuron to itself across time. At time t , recurrent neurons are input with information not just from the previous layer x_t but also from themselves of the previous position h_{t-1}. Consequently, the output y_t is influenced not only by the current input information but also by the information at time $t-1$. Mathematically, these processes can be described using the following transition function:

$$h_t = f(w_{hx}x_t + w_{hh}h_{t-1} + b_h) \tag{1}$$

$$\hat{y}_t = f\left(w_{yh}h_t + b_y\right) \tag{2}$$

where $f(\bullet)$ is an activation function. w_{hx} is the matrix of conventional weights between an input layer and a hidden layer and w_{hh} is the matrix between a hidden layer and itself at adjacent time steps. The vectors b_h and b_y are bias parameters which allow each node to learn an offset.

The structure of the RNN across a time can be described as a deep network with one layer per time step. It can be seen that the network can be trained across time steps using backpropagation that is called backpropagation through time [27] . But the learning process is especially challenging due to the problem of gradients vanishing or exploding. This is the main reason why the training optimal process approaches to a local extreme value.

B. Basic theory of LSTM

To overcome the problem of gradient vanishing or exploding, a long short-term memory (LSTM) architecture that involves a memory cell was constructed [28]. Forget gates are introduced in LSTM to avoid the long-term dependency problem. These adopted forget gates are able to control the utilization of information in the cell states. To capture nonlinear dynamics in time-series sensory data and learn effective representations of machine conditions, LSTM should be superior compared to traditional RNN due to its capability to capture long-term dependencies. The architecture of LSTM cell can be described as Fig. 1. At each time step t, hidden state h_t is updated by current data at the same time step x_t, hidden state at previous time step h_{t-1}, input gate i_t, forget gate f_t, output gate o_t and a memory cell c_t. The following updating equations are given as follows:

$$i_t = \sigma(w_i x_t + v_i h_{t-i} + b_i) \tag{3}$$

$$f_t = \sigma\left(w_f x_t + v_f h_{t-i} + b_f\right) \tag{4}$$

$$o_t = \sigma(w_o x_t + v_o h_{t-i} + b_o) \tag{5}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot tanh(w_c x_t + v_o h_{t-i} + b_c) \tag{6}$$

$$h_t = o_t \odot \tanh(c_t) \tag{7}$$

where model parameters including all $\boldsymbol{W} \in R^{d\times k}$, $\boldsymbol{V} \in R^{d\times d}$ and $\boldsymbol{b} \in R^{d}$ are shared by all time steps and learned during model training, σ is the sigmoid activation function, $\odot$ denotes the element-wise product, k is a hyper-parameter that representing the dimensionality of hidden vectors.

In this paper, LSTM neural network, together with a flatten

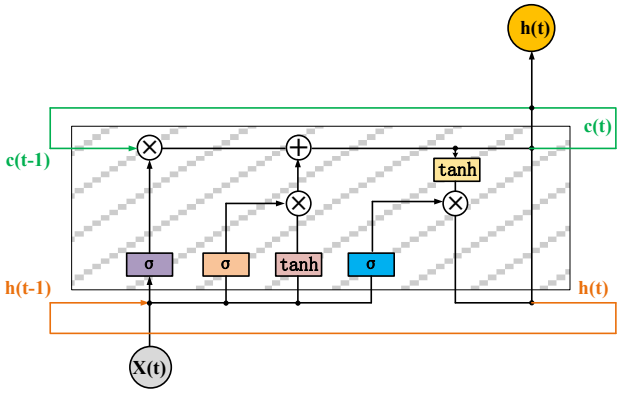


Fig. 1. The architecture of LSTM

layer and three fully connected layers, is directly used for the raw acceleration data to learn meaningful representations without any feature engineering.

III. Experiment

To experimentally verify the performance of LSTM in motors diagnosis, tests of six three-phase asynchronous motors under different fault conditions were carried out in the drivetrain diagnostics simulator system. As shown in Fig. 2, this system consists of a two-stage planetary gearbox, a two-stage fixed-axis gear box, a magnetic brake for loading, and a three-phase asynchronous motor. This test rig meets the all configuration requirement of drivetrain system and provides a practical and reliable test environment for motors diagnosis.

Six motors under different health conditions (1 healthy and 5 faulty) were used in the experiment to generate the required dataset under the load of 33.90 N.m (25 lbf.Ft). Data from the healthy motor are used as a benchmark for comparison with the experimental data from other faulty motors. The faulty motors each has a specific fault: (1) voltage imbalance, (2) rotor imbalance, (3) faulty bearing, (4) broken rotor bars, and (5) bowed rotor . The faulty conditions of these motors are also listed in Table I. Voltage imbalance is achieved by switching phases on and off introducing resistance using a control box. Rotor imbalance is achieved by taking a balanced rotor from the manufacturer and intentionally removing the balance weights. The faulty bearing motor consists of a motor with intentionally faulted bearings. The bowed rotor motor consists of a motor with an intentionally bent rotor in the center. The broken rotor bar motor consists of a motor already fitted with an intentionally broken rotor bar.

One acceleration sensor was used in the experiment and was installed at the shell of motors to measure the vibration signals in two coordinate directions (horizontal and vertical). The signal was collected by the NI-9234 data acquisition card. The sampling frequency was 10.24 kHz, the motor frequency was set at 15Hz, the test lasted approximately 120 s, and thus the number of each time sequence acquired from one condition is approximately 1228800. Fig. 3 displays a set of time waveforms of the measured acceleration signals of these faulty motors.

TABLE I. MOTORS UNDER 6 HEALTH CONDITIONS IN TESTS

Motor condition	Test time (s)	Sample	One hot label
Normal	118	1000	100000
Bowed rotor	121	1000	010000
Broken bar	122	1000	001000
Rotor imbalance	121	1000	000100
Faulty bearing	122	1000	000010
Voltage imbalance	121	1000	000001

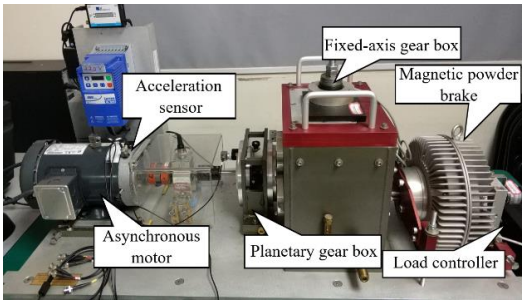


Fig. 2. The test rig

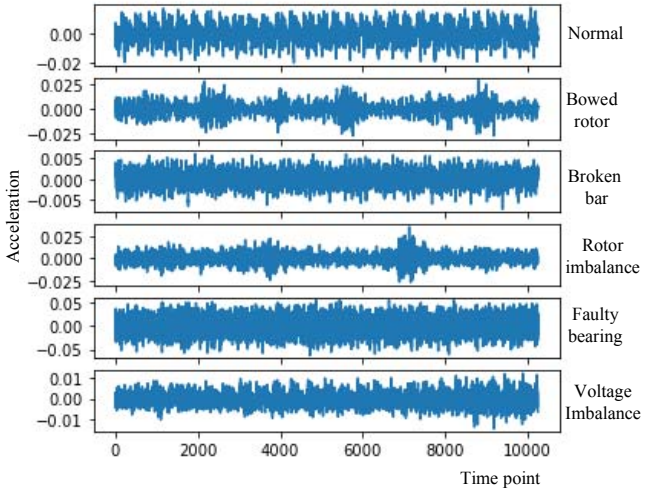


Fig. 3. The waveforms of the measured acceleration signals of 6 motors

IV. Data Processing

The flowchart of the fault diagnosis process for the six asynchronous motors is shown in Fig. 5. One LSTM neural network layer and three fully connected layers are used for the raw signals. Performances of other classification methods, such as LR, SVM, MLP, and basic RNN, are tested and contrasted.

A. Sample Preparation

From the 120 s vertical experiment acceleration signals, the middle 100 s stable signals are selected to be the train and test data sequence and every 0.1 second signals which contain 1024 data points (10.24 KHz) represent a sample. Thus there are 1000 samples for each health condition. Then the 1024 data points are divided into 64 parts and each part contains 16 data points. Thus the samples of one class can be converted into a three-dimensional tensor $S \in R[1000, 16, 64]$, where 16 is the number of time steps, 64 is the number of data points, and 1000 represents the number of samples. As the number of motors is 6,

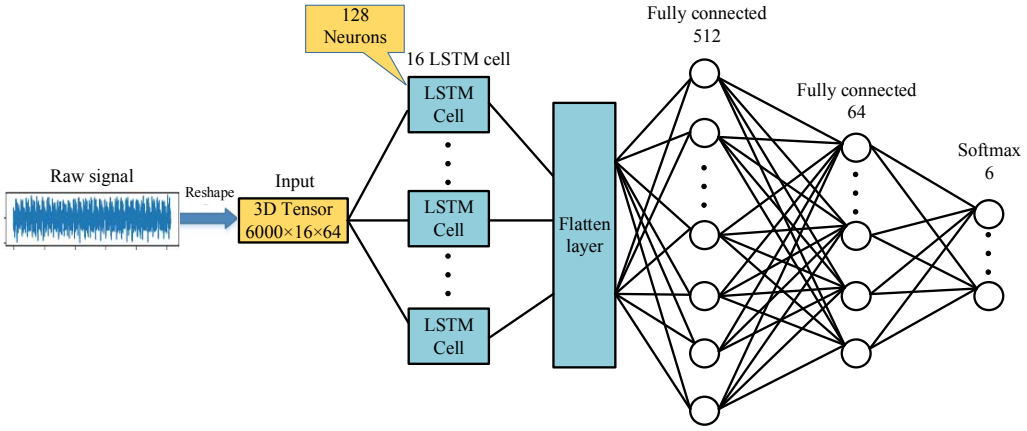


Fig. 4. The structure of the neural networks

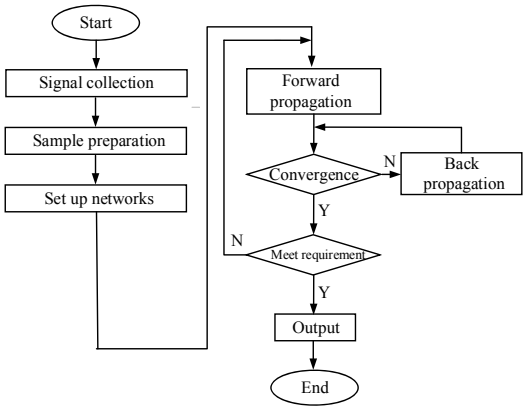


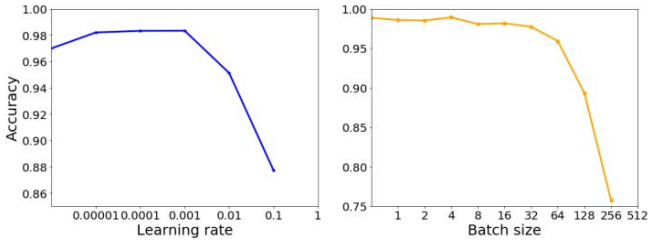
Fig. 5. The flowchart of the proposed approach

the structure of the final sample tensor is [6000, 16, 64], which means that each time step of 16 time steps contains 64 data points and the total number of samples is 6000.

To achieve higher classification accuracy, a three-folder cross-validation is adopted for training/testing splitting. The data set is divided into 3 subsets, and the holdout method is repeated 3 times. Each time, one of the 3 subsets is used as the test set and the other 2 subsets are put together to form a training set. Then the average error across all 3 trials is computed.

B. The Structure of LSTM Neural Networks

The structure of the complete network is shown as Fig. 4. The LSTM neural network includes 16 LSTM cells and each cell contains 128 layer sizes in the hiding layer. After that a flatten layer is added to convert the multidimensional output to one-dimensional output. After that three fully connected layers with a structure of [512, 64, 6] is added, where activation functions of tanh are used in the first two layers and activation functions of softmax are used in the last layers. To prevent model form overfitting, dropout and early stopping mechanism are adopted in the training process. Probability of dropping out neurons at output layer of LSTM are set to 0.5. These training samples are directly fed into the network to build the model. The architecture



(a) Accuracy under various learning rate (b) Accuracy under various batchsize

Fig. 6. The relationship between accuracy and selected parameters

of all the neural network is constructed through minimizing the following cross-entropy cost function:

$$C = -\frac{1}{n}\sum_x \|\hat{y}\, lny + (1-\hat{y})\ln(1-y)\| \qquad (8)$$

where y is an output of the model, and $\hat{y}$ is a true label value. In the testing step, testing data are input to the trained model to verify the classification accuracy.

C. Selection of Model Parameters

In the process of training the model, the gradient descent method is used for optimization. The learning rate is an important parameter, which influences the adjustment of weights and error convergence. In this experiment, different learning rates were used to train the model, and we obtained different accuracy as shown in Fig. 6 (a). By choosing an appropriate learning rate, we can speed up the convergence of the LSTM network and improve its accuracy. In this experiment, the optimized learning rate is determined as 0.001.

Due to the large size of sample data and restrictions on computer configuration, it is common to divide the samples into blocks of moderate size. The size of this block is called batch-size. In this experiment, we used different batch-sizes to train the LSTM model. The relationship between accuracy and the number of batch-size is shown in Fig. 6 (b). It can be seen that when the batch-size is small, the network testing accuracy is relatively high. However, it takes a longer time to carry out one iteration. As the batch-size becomes larger (especially when the batch-size is greater than 64), accuracy is progressively reduced. However, the time required for one iteration is also

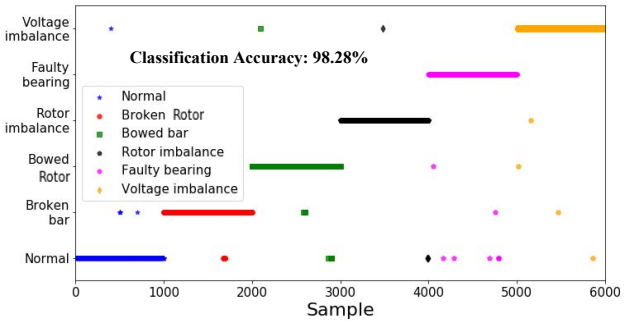


Fig. 7. Classification result of LSTM neural network

progressively reduced. By carrying out a comprehensive comparison with regard to the above situation, we find that when the batch-size is 8, we can not only ensure accuracy, but also reduce training time. So the optimized batch-size is determined as 8.

D. Result of LSTM Classification

The classification rate is calculated as the ratio of the number of correctly classified test samples to the total number of test samples. As shown in Fig. 7, LSTM neural network performs with an average classification accuracy of 98.28%, misclassifying 103 samples.

E. Comprasion with Other Methods

The performance of the proposed method is compared by the following methods:

- LR: logistic regression classifier on extracted features;
- SVM: Support Vector Machine classifier on extracted features;
- MLP: Multi-layer Neural Network on extracted features;
- RNN: Original RNN on raw data;

Since LR, SVM and MLP can not address sequential data, feature extraction is conducted firstly. Different features from time and frequency domains are extracted as shown in Table II. Six typical statistical features are extracted in time domain, including root mean square (RMS), variance, maximum value, skewness, kurtosis and peak-to-peak. RMS is a measure for the magnitude of a varying quantity. It is also related to the energy of a signal. Kurtosis indicates the spikiness of a signal. Features from the frequency domain provide another perspective of health conditions, and reveal information that are not included in the time domain. In frequency domain, spectral skewness and spectral kurtosis are extracted, where $S(f_i)$ is the power spectrum density obtained using the Welch method.

The results of other methods are shown in Fig. 8. Contrasted with other methods, the proposed method achieves the highest precision. From the result, we observe that LR performs worst. It can be explained by the limitation of linear models. SVM and MLP perform worse than the RNN group. The reasons may be the fact that the structure of SVM with *rbf* kernel and MLP with layers sizes [64-128-64-6] are too simple compared with RNN and LSTM. For a fair comparison, RNN shares the same hidden

TABLE II. EXTRACTED FEATURES

Domain	Features	Expression	Dimension
Time	RMS	$Z_{RMS}=\sqrt{\frac{1}{n}(z_1^2+z_2^2+\cdots+z_n^2)}$	1
	Variance	$Z_{\text{var}}=\frac{1}{N}\sum_{N}(z_i-\hat{z})^2$	1
	Maximum	$Z_{\max}=\max(Z)$	1
	Skewness	$Z_{skew}=E[(\frac{z-\mu}{\sigma})^3]$	1
	Kurtosis	$Z_{kurt}=\frac{1}{n}\sum_{n}(\frac{z-\mu}{\sigma})^4$	1
	Peak-to-Peak	$Z_{p-p}=\max(z)-\min(z)$	1
Frequency	Spectral skewness	$f_{skew}=\sum_{i=1}^{k}(\frac{f_i-f}{\sigma})^3 S(f_i)$	1
	Spectral Kurtosis	$f_{kurt}=\sum_{i=1}^{k}(\frac{f_i-f}{\sigma})^4 S(f_i)$	1

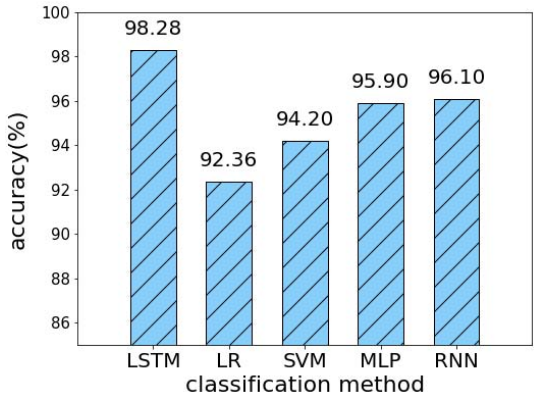


Fig. 8. The performance of all methods

layer sizes (128) with the LSTM. However, It is obvious that LSTM performs slightly better than RNN. The reasons may be the fact that gate functions employed in LSTM can enable it to capture long-term dependency better than RNN.

V. CONCLUSION

In this paper, a fault diagnosis approach with high accuracy for the three-phase asynchronous motors is presented. LSTM neural network is directly used for the raw acceleration data to learn meaningful representations without any feature engineering to classify the health conditions of faulty motors. The experiment verifies that the proposed method with the highest accuracy of health condition recognition performs better than other methods such as LR, SVM, MLP, and basic RNN. The method plays an important role for fault diagnosis of asynchronous motors and is helpful for predicting potential risks in factory manufacturing. This opens up new opportunities to advance the science base for motors health evaluation.

In future research, our effort will be devoted to make our proposed method more reliable for practical use. To achieve this goal, a knowledge database corresponding to different health conditions of asynchronous motors is considered. The thresholds can be defined by the fault diagnosis algorithm and updated with a new condition that is not included in present base.

Moreover, the database can be modified if the predicted health condition is inconsistent with the actual health condition. Our fault diagnosis algorithm can be applied to more motors conditions with better accuracy.

ACKNOWLEDGMENT

This project is supported by the National Key Technology R&D Program of China (No. 2017YFB1302004), and NSFC project (No. 51305258).

REFERENCES

[1] T. Han, B.-S. Yang,W.-H. Choi, and J.-S. Kim, "Fault diagnosis system of induction motors based on neural network and genetic algorithm using stator current signals," International Journal of Rotating Machinery,vol.2006, ArticleID61690,13 pages, 2006.

[2] Ghate V N, Dudul S V. Optimal MLP neural network classifier for fault detection of three phase induction motor[J]. Expert Systems with Applications, 2010, 37(4):3468-3481.

[3] T.Han,B.-S. Yang,and Z.-J.Yin, "Feature-based faultdiagnosis system of induction motors using vibration signal," Journal of Quality in Maintenance Engineering,vol.13, no.2,pp. 163–175, 2007

[4] Jun-Jian Hou, Wei-Kang Jiang, Wen-Bo Lu. Application of a near-field acoustic holography-based diagnosis technique in gearbox fault diagnosis. Journal of Vibration & Control, 2013, 19(1): 3–13.

[5] Huang C, Wu F, Zhao J, et al. A novel fault diagnosis method in SVPWM voltage-source inverters for vector controlled induction motor drives[J]. International Journal of Applied Electromagnetics & Mechanics, 2016, 50(1):97-111.

[6] J. R Ottewill,MOrkisz. Condition monitoring of gearboxes using synchronously averaged electric motor signals. Mechanical Sys- tems & Signal Processing, 2013, 38(2): 482–498.

[7] A M D Younus, B S Yang. Intelligent fault diagnosis of rotating machinery using infrared thermal image. Expert Systems with Applications, 2012, 39(2): 2082–2091.

[8] H. Talhaoui, A. Menacer, A. Kessal, and R. Kechida, "Fast Fourier and discrete wavelet transforms applied to sensorless vector control induction motor for rotor bar faults diagnosis," ISA Transactions,vol.53, no.5,pp. 1639–1649, 2014.

[9] B. Liang, S. D. Iwnicki, and Y. Zhao, "Application of power spectrum, cepstrum, higher order spectrum and neural net- work analyses for induction motor fault diagnosis," Mechanical Systems and Signal Processing,vol.39, no.1-2,pp. 342–360, 2013.

[10] M. Seera, C. P. Lim,D. Ishak, andH. Singh, "Offline and online faultdetectionanddiagnosis of inductionmotorsusing ahybrid soft computing model," Applied Soft Computing,vol.13, no.12, pp. 4493–4507, 2013.

[11] Hui-Min Zhao, Cai-Hua Fang, Deng Wu. Research on motor fault diagnosis model for support vector machine based on Intelligent optimization methods. Journal of Dalian Jiaotong University, 2016.

[12] Pan S, Han T, Tan A C C, et al. Fault Diagnosis System of Induction Motors Based on Multiscale Entropy and Support Vector Machine with Mutual Information Algorithm[J]. Shock and Vibration,2016,(2016-1-11), 2016, 2016(7):1-12.

[13] Martínez-Morales J D, Palacios-Hernández E R, Campos-Delgado D U. Multiple-fault diagnosis in induction motors through support vector machine classification at variable operating conditions[J]. Electrical Engineering, 2016:1-15.

[14] Pandarakone S E, Mizuno Y, Nakamura H. Distinct Fault Analysis of Induction Motor Bearing Using Frequency Spectrum Determination and Support Vector Machine[J]. IEEE Transactions on Industry Applications, 2017, 53(3):3049-3056.

[15] Soualhi, A., Clerc, G., and Razik, H., "Detection and diagnosis of faults in induction motor using an improved artificial ant clustering technique," IEEE Trans. Ind. Electron., Vol. 60, No. 9, pp. 4053–4062, 2013

[16] Wang L H, Zhao X P, Wu J X, et al. Motor Fault Diagnosis Based on Short-time Fourier Transform and Convolutional Neural Network[J]. Chinese Journal of Mechanical Engineering, 2017, 30(6):1-12.

[17] Jia Feng, Ya-Guo Lei, Jing Lin, et al. Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. Mechanical Systems & Signal Processing, 2016, 72–73: 303–315.

[18] Zhao R, Yan R, Chen Z, et al.Deep Learning and Its Applications to Machine Health Monitoring: A Survey[J]. 2016.

[19] Lecun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553):436.

[20] A. Karpathy and L. Fei-Fei, "Deep visual-semantic alignments for generating image descriptions," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3128–3137.

[21] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, "Show and tell: A neural image caption generator," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3156–3164.

[22] Graves A, Mohamed A R, Hinton G. Speech recognition with deep recurrent neural networks[C]// IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2013:6645-6649.

[23] M. Yuan, Y.Wu, and L. Lin, "Fault diagnosis and remaining useful life estimation of aero engine using lstm neural network," in 2016 IEEE International Conference on Aircraft Utility Systems (AUS), Oct 2016, pp. 135–140.

[24] R. Zhao, J. Wang, R. Yan, and K. Mao, "Machine helath monitoring with lstm networks," in 2016 10th International Conference on Sensing Technology (ICST). IEEE, 2016, pp. 1–6.

[25] Malhotra P, Vishnu T V, Ramakrishnan A, et al. Multi-Sensor Prognostics using an Unsupervised Health Index based on LSTM Encoder-Decoder[J]. 2016.

[26] Schmidhuber J. Deep Learning in neural networks: An overview.[J]. Neural Netw, 2014, 61:85-117.

[27] Lipton Z C, Berkowitz J, Elkan C. A Critical Review of Recurrent Neural Networks for Sequence Learning[J]. Computer Science, 2015.

[28] S. Hochreiter , J. Schmidhuber , Long short-term memory, Neural Comput. 9 (1997) 1735–1780 .