

Received February 19, 2019, accepted February 25, 2019, date of publication March 6, 2019, date of current version March 25, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2903295

A New Snapshot Ensemble Convolutional Neural Network for Fault Diagnosis

LONG WEN¹, LIANG GAO¹, (Member, IEEE), AND XINYU LI¹

State Key Laboratory of Digital Manufacturing Equipment and Technology, School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China

Corresponding author: Xinyu Li (lixinyu@mail.hust.edu.cn)

This work was supported in part by the National Natural Science Foundation for the Distinguished Young Scholars of China under Grant 51825502, in part by the Natural Science Foundation of China (NSFC) under Grant 51805192 and Grant 51775216, in part by the Natural Science Foundation of Hubei Province under Grant 2018CFA078, in part by the China Postdoctoral Science Foundation under Grant 2017M622414, and in part by the Program for the HUST Academic Frontier Youth Team.

ABSTRACT Deep learning (DL) has been widely applied in fault diagnosis field. The ensemble of DL can significantly improve the generalization ability of DL. Snapshot ensemble learning, which uses the cyclic learning rate scheduler (CLR) and then combines the local minima in once training to form an ensemble method, has shown powerful performance. However, the learning rate range of CLR should be pre-defined by experiences, which may limit its prediction accuracy. In this paper, a new snapshot ensemble convolutional neural network (SECNN) is proposed, which can find the proper range of learning rate for SECNN automatically when facing a new dataset. First, a max-min cosine cyclic learning rate scheduler (MMCCLR) is designed to avoid learning rate range being affected by other parameters. Then, a new learning rate testing (logLR Test) is applied to estimate the proper learning rate range for MMCCLR. Finally, the SECNN with MMCCLR is developed. The SECNN is tested on three famous datasets, including bearing dataset of Case Western Reserve University, self-priming centrifugal pump dataset, and bearing dataset provided by the Machinery Failure Prevention Technology. The results have been improved a lot by the proposed SECNN methods. The SECNN has also achieved the start-of-the-art prediction accuracies by comparing with other DL and traditional methods.

INDEX TERMS Ensemble learning, learning rate estimation, deep learning, fault diagnosis, convolutional neural network.

I. INTRODUCTION

Fault diagnosis plays a vital role in modern complicated system. Accurate fault detection of engineering systems can save invaluable time and cost to avoid considerable loss and catastrophic consequences [1]. Because of the growing convenience in data acquiring and storing, the data-driven fault diagnosis methods have been increasingly studied by researchers from both academic and engineering fields [2], [3]. Data-based advanced methods become the important issues in data-driven fault diagnosis [4] to effectively mine data features and accurately predict fault models. And these data-based advanced methods can establish fault models using historical data without prior physical model, making them very suitable for modern complex industrial process and equipment [5].

The associate editor coordinating the review of this manuscript and approving it for publication was Xi Peng.

Artificial neural network (ANN) and support vector machine (SVM) are two of the most popular methods in data-driven fault diagnosis [6]. These two methods are generally combined with other feature extraction processes [7]. But, it is difficult to know what handcrafted features should be extracted for ANN and SVM methods [8], and the upper-bound performances of ANN and SVM methods are defined by these handcrafted features [9]. Deep learning (DL) uses the data representation learning rather than explicit feature extraction process to extract data features [8], and it enables the features to be learnt automatically. With this advantage, DL based fault diagnosis methods have shown great potential [10], [11]. However, there are still some problems when using the individual DL for fault diagnosis, including low generalization ability [12].

Ensemble learning is an effective way to improve the generalization ability of individual learner (denoted by

base learners) [13]. Ensemble learning can use multiple base learners and a certain combination strategy to get better results than individual base learner. Lots of ensemble learning techniques which apply DL methods as base learners have been applied in fault diagnosis. Shao *et al.* [12] constructed the ensemble DL method containing various auto-encoders with different activation functions and achieved good results.

Almost all of the ensemble learning methods should train multiple base learners, which can consume much time. Huang *et al.* [14] proposed a new paradigm of ensemble learning, named snapshot ensemble learning (SEL). It uses the cosine cyclic learning rate (CCLR) scheduler, and divides the whole training process to M cycles. At each cycle end, the learning rate of CCLR is small to encourage the model convergence toward to the local minimum. Then SEL combines these local minima and has shown a powerful performance. SEL can obtain M different base learners at once training, which can overcome the above shortcomings. It can promote its efficiency without increasing the computing budget. However, the learning rate range of CCLR is pre-defined by experiences, which may limit its prediction accuracy. To select a proper learning rate range of CCLR is vital for the further improvement.

In this research, a new max-min cosine cyclic learning rate scheduler (MMCCLR) is proposed. In MMCCLR, the upper and lower boundaries of learning rate ranges have no interaction with other parameters, and can be assigned separately. Then, a new learning rate range estimation method, named as LogLR Test, is proposed to estimate the upper and lower boundaries of MMCCLR when facing a new dataset. Thirdly, as Convolutional Neural Network (CNN) has been widely applied in fault diagnosis field, a new snapshot ensemble convolutional neural network (SECNN) based on MMCCLR is developed for fault diagnosis. It is tested on three famous datasets, and the results validate its good performance.

The remainder of this paper is organized as following. Section II discusses the related works. Section III presents the snapshot ensemble CNN for fault diagnosis. Section IV presents the proposed max-min cosine cyclic learning rate scheduler and the learning rate range estimation method. Section V shows the results and discussions. The conclusion and future researches are presented in Section VI.

II. RELATED WORKS

The related works are described in this section, including ensemble deep learning based fault diagnosis methods and learning rate selection for deep learning.

A. ENSEMBLE DEEP LEARNING BASED FAULT DIAGNOSIS METHODS

Fault diagnosis could be roughly classified into three categories: model-driven, signal-based and data-driven [1]. Data-driven fault diagnosis methods can learn from historic data without prior physical models and explicit signal patterns about the system, which are suitable for the modern complex industrial process and equipment, and they have attracted

more and more attentions from both scientific and engineering fields [15].

ANN and SVM are two of the most popular methods in data-driven fault diagnosis [6]. Ren *et al.* [16] proposed a fault diagnosis method by jointing sparse representation with SVM. Prieto *et al.* [17] developed a hierarchical ANN based method for bearing fault detection. Ahmed *et al.* [18] studied an ANN based fault detection and classification method of an automotive internal-combustion-engine. However, ANN and SVM should be generally combined with other feature extraction processes [7], but it is difficult to know what handcrafted features should be extracted for them.

Different from traditional machine learning methods, DL uses the data representation learning rather than explicit features extraction to extract data features [8]. With these advantages, DL has been widely applied in the clustering of unlabeled data set [19], deep hashing for cross-modal retrieval [20], [21], dimension reduction [22], similarity search [23]. DL has also been applied in the fault diagnosis field. Zhao *et al.* [24] reviewed the various DL methods based fault diagnosis, including deep belief network (DBN), stacked auto-encoder (SAE), convolutional neural network (CNN) and recurrent neural network. Liu *et al.* [7] surveyed the artificial intelligent methods for fault diagnosis of rotating machinery. Qin *et al.* [25] studied an optimized DBN for the fault diagnosis on planetary gearboxes of wind turbines. Shao *et al.* [26] studied an optimized DBN using particle swarm optimization for fault diagnosis of rolling bearing. Gan and Wang [27] applied a DBN based hierarchical diagnosis network on the fault diagnosis of rolling element bearings. Chen and Li [28] studied a SAE-DBN based multi-sensor feature fusion for bearing fault diagnosis. Qi *et al.* [29] proposed a SAE based deep network for fault diagnosis. Wen *et al.* [30] proposed SAE based deep transfer learning for fault diagnosis. Sun *et al.* [31] investigated a stacked SAE for intelligent fault diagnosis. Zhang *et al.* [32] investigated a subset based SAE for intelligent fault diagnosis. Zhang *et al.* [33] proposed a CNN with Wide First-layer Kernels (WDCNN) for the fault diagnosis of bearing. Xia *et al.* [34] applied multi-sensor signals and CNN to the fault diagnosis of bearing and gearbox. Zhang *et al.* [35] studied the intelligent fault diagnosis under varying working conditions using domain adaptive CNN method. Wen *et al.* [11] applied LeNet-5 CNN to the fault diagnosis of rolling bearing, self-priming centrifugal pump and axial piston hydraulic pump. Even though DL has been studied widely, there are still some problems when using DL for fault diagnosis, such as the low generalization ability [12].

Ensemble learning can improve the generalization ability of individual learner (denoted by base learners) [13]. Ensemble DL method for fault diagnosis always shows better performance than individual method. Zhang *et al.* [36] developed CNN with Training Interference (TICNN) and the ensemble version of TICNN can obtain high accuracy under noisy environment. Wu *et al.* [37] proposed integrated ensemble learning models for imbalanced fault diagnosis.

Zhang *et al.* [38] studied a multi-objective DBN based ensemble model for remaining useful life estimation. Wang *et al.* [39] investigated a feature ensemble learning using stacked denoising Auto-encoder for induction motor fault diagnosis. However, almost all of the ensemble learning methods for fault diagnosis have trained multiple base learners, which are very time-consumption.

Snapshot ensemble learning (SEL) proposed by Huang *et al.* can overcome the above shortcomings [14]. It uses a cosine cyclic learning rate (CCLR) scheduler and divides the whole training process into M cycles. At each cycle end, the learning rate is decreased to a small value to let the training process converge toward to the local minimum. Then SEL combines these local minima to form an efficiency ensemble method. It has shown the good performance without increasing computing budget. With this advantage, a new snapshot ensemble convolutional neural network (SECNN) is proposed for fault diagnosis in this research.

However, as most deep learning methods, the learning rate range for SECNN is selected by experiences. Since the selection of a proper learning rate range is vital, a new learning rate test method is proposed to estimate the proper learning rate of SECNN automatically in this research.

B. LEARNING RATE SELECTION FOR DEEP LEARNING

Learning rate is one of the most important hyper parameter of DL [40], as large learning rate will make the training process diverge and small learning rate will affect the convergence of the training process. The most common learning rate schedulers are setting the learning rate as a global constant value or decreasing the learning rate monotonically during the training process. Some researchers also investigated the improved learning rate schedulers for DL in fault diagnosis field. Jing *et al.* [41] applied an adaptive learning rate based CNN method for gearbox fault diagnosis, which set a large learning rate at beginning and decreased the learning rate adaptively. Guo *et al.* [42] proposed an adaptive learning rate scheduler based CNN method for bearing fault diagnosis, and it used the gradient of the weight to increase or decrease the learning rate.

Smith [40] studied a cyclical learning rate scheduler (CLR). It used the triangular CLR which let the learning rate cyclically vary between a reasonable boundary values. Loshchilov and Hutter [43] proposed the cosine CLR (CCLR) and it has achieved good results on image classification. However, the learning rate is decreased to a small value at each cycle end, leading the training process converge toward to the local minimum. Huang *et al.* [14] proposed a novel SEL by combining these local minima to form a powerful ensemble method without increasing the computing budget.

Like other DL methods, the range of learning rate in CCLR should be pre-defined, and how to select a proper learning rate range is vital for further improvement. Xia *et al.* [34] investigated the fault diagnosis on roller bearing by testing different learning rate from 0.005 to 0.05 with an interval

of 0.005. Xie and Zhang [44] set the learning rate varied from 0.01 to 0.001 for selecting a proper learning rate. Gan and Wang [27] compared four different DBN models with different learning rate for fault diagnosis. These learning rate test methods are labor consumption, and may also limit the prediction accuracy of DL methods.

Linear learning rate test (LLR Test) [40] can provide enormously valuable information about the learning rate when facing a new dataset. The learning rate linearly increases from a low value to high value and LLR Test notices the values when the accuracy starts to increase and becomes diverse. However, LLR Test has a low resolution ratio when the learning rate range is large. In this research, a new log linear LR Test (LogLR Test) is proposed to remedy this drawback.

Since LogLR Test can both detect the lower and upper boundaries, it is better to set the learning rate range of CCLR using the results of LogLR Test. However, the lower boundary of learning rate range in CCLR is calculated result of upper boundary and cyclical numbers, so a new max-min cosine cyclic learning rate scheduler (MMCCLR) is proposed. In MMCCLR, both of lower and upper boundaries are not affected by other parameters, and they will be estimated using LogLR Test.

III. PROPOSED SNAPSHOT ENSEMBLE CNN FOR FAULT DIAGNOSIS

In this section, the proposed SECNN for fault diagnosis is presented. Firstly, the data preprocessing method based on S-Transform is introduced to convert the time-domain signal to 2D matrix format. Secondly, a new structure of CNN is given. Thirdly, the cosine cyclic learning rate scheduler is introduced. Finally, the SECNN is presented.

A. DATA PREPROCESSING METHOD BASED ON S-TRANSFORM

The S-Transform method proposed by Stockwell *et al.* [45] in 1996 is a famous time-frequency technique. It can be viewed as an extension of short-time Fourier transform (STFT). Let τ denote the time of spectral localization, f denote the Fourier frequency, $w(t)$ denote the window function, which is presented in equation (1). Then the S-Transform can be calculated by equation (2).

$$w(t) = \frac{|f|}{\sqrt{2\pi}} e^{-t^2 f^2 / 2} \quad (1)$$

$$S(\tau, f) = \int_{-\infty}^{+\infty} x(t) \frac{|f|}{\sqrt{2\pi}} e^{-(t-\tau)^2 f^2 / 2} e^{-j2\pi f t} dt \quad (2)$$

The advantage of S-Transform is that the window function is a function of both frequency and time. The width of window is inversely proportional to the frequency. In other words, the window provides good localization in the frequency domain for low frequencies while providing good localization in time domain for higher frequencies.

B. CNN STRUCTURE BASED ON LENET-5

LeNet-5 is an effective classical CNN, and it contains several alternative convolutional and pooling layers and a two-layer fully connected (FC) layer. In this research, a new CNN structure based on LeNet-5 is proposed for fault diagnosis, and the detail of network structure is presented in TABLE 1.

TABLE 1. The layer configurations of CNN models.

Layer name	CNN Operator and Parameters		
	Operator	Filter parameters	Stride parameters
L1	Conv	$7 \times 7 \times 64$	1×1
L2	MaxPool	2×2	2×2
L3	Conv	$5 \times 5 \times 96$	1×1
L4	MaxPool	2×2	2×2
L5	Conv	$3 \times 3 \times 128$	1×1
L6	MaxPool	2×2	2×2
L7	Conv	$3 \times 3 \times 256$	1×1
L8	MaxPool	2×2	2×2
L9	Conv	$3 \times 3 \times 256$	1×1
L10	MaxPool	2×2	2×2
L11	Conv	$3 \times 3 \times 256$	1×1
L12	MaxPool	2×2	2×2
L13	2560, L2-regulation, dropout rate 0.3		
L14	512, L2-regulation, dropout rate 0.3		

The denotation of ‘Conv’ means that it is the convolution layer. The parameter ‘ $7 \times 7 \times 64$ ’ means the filter size is 7×7 and the depth of filter is 64. Maxpool layer is adopted and its pool size is 2×2 . The stride size of Maxpool layer is 2×2 , and the padding type is zero-padding.

After six alternative convolutional and pooling layers, a two-layer FC layer is followed. For FC layer, the L2-regulation and dropout techniques are applied to improve the generalization of proposed CNN method.

C. COSINE CYCLIC LEARNING RATE SCHEDULER

Cyclical learning rate scheduler (CLR) which lets learning rate vary between reasonable boundary values has shown good performance. Cosine CLR (CCLR) applies a shift cosine function along with increase of training epoch, as shown in Fig. 1.

The equation of CCLR can be denoted as equation (3).

$$\eta = \frac{\eta_{\max}}{2} \left(1 + \cos \left(\frac{\pi \bmod ([t, b], [T, Mb])}{[T, Mb]} \right) \right) \quad (3)$$

η stands for the learning rate of current epoch. $\eta_{\max}$ is the pre-defined upper boundary of learning rate. t stands for the current iteration number and T is the total iterations. M denotes the number of cycles. b is the batch size and $\lceil \cdot \rceil$ is the exact division function.

D. SNAPSHOT ENSEMBLE LEARNING METHODS

In CCLR, the learning rate is decreased to a small value at each cycle end, leading the training process converge toward to the local minimum. Let M denote the number of the cycle. Snapshot ensemble learning (SEL) which combines these M

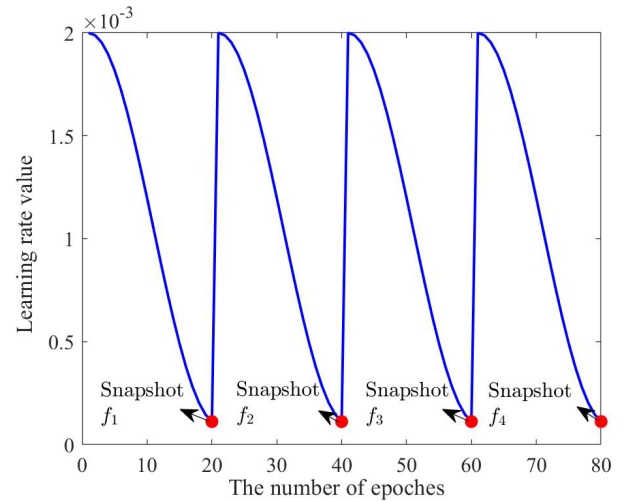


FIGURE 1. The learning rate curve of CCLR.

local minima to form an ensemble method can improve a lot without increasing of the computational budget. As shown in Fig. 1, let $f_i, i = 1 \cdots M$ denote each local minimum. The SEL model f_{ens} can be formulated by equation (4).

$$f_{\text{ens}} = \frac{1}{M} \sum_1^M f_i \quad (4)$$

The extension SEL can also combine N repeated running of SEL methods, and it can be represented by equation (5). But it should be noted that the N repeated running of SEL will consume almost N times of training times.

$$f_{\text{ens}} = \frac{1}{M \times N} \sum_1^M \sum_1^N f_{i,j} \quad (5)$$

In this research, SECNN is proposed for fault diagnosis which combines the new CNN structure, MMCCLR and SEL.

IV. PROPOSED LEARNING RATE SCHEDULER AND LEARNING RATE RANGE ESTIMATION METHOD

A. MAX-MIN COSINE CYCLIC LEARNING RATE SCHEDULER

As shown in equation (3), the upper boundary of CCLR is pre-defined by $\eta_{\max}$, but the lower boundary is calculated using $\eta_{\max}$ and M . In this research, a max-min cosine cyclic learning scheduler (MMCCLR) is proposed. The upper and lower boundaries will have no interaction with other parameters. $\eta_{\min}$ denotes the lower boundary of learning rate. The formulation of MMCCLR can be formulated by equation (6).

$$\eta = \eta_{\min} + \frac{(\eta_{\max} - \eta_{\min})}{2} \times \left(1 + \cos \left(\frac{\pi \bmod ([t, b], [T, Mb])}{[T, Mb]} \right) \right) \quad (6)$$

The advantage of MMCCLR is that it has more accurate control on learning rate. Both of $\eta_{\max}$ and $\eta_{\min}$ can be assigned separately. Then, they will be estimated by the LogLR Test.

B. PROPOSED LOG LINEAR LEARNING RATE TEST METHOD

To determine the reasonable learning rate range for MMCCLR is vital, especially when facing new datasets. Linear learning rate test (LLR Test) which trains the model using several epochs with linearly increasing the learning rate from a low value to a high value, can provide enormously valuable information about the learning rate. It notices about the minimum learning rate when the accuracy starts to increase and the max learning rate when the accuracy becomes diverse. LLR Test can be formulated by equation (7).

$$\eta = \eta_{\min}^{LR} + \left(\eta_{\max}^{LR} - \eta_{\min}^{LR} \right) \times \frac{[t, b]}{[T, b]} \quad (7)$$

LLR Test has low resolution ratio when the learning rate range is large. In this research, a new log linear LR Test (LogLR Test) is proposed and presented in equation (8).

$$\log_{10} \eta = \log_{10} \eta_{\min}^{LR} + \left(\log_{10} \eta_{\max}^{LR} - \log_{10} \eta_{\min}^{LR} \right) \times \frac{[t, b]}{[T, b]} \quad (8)$$

LogLR Test has a better resolution ratio when the detected learning rate range is large. It is a two-step learning rate estimation. In the first estimation, the learning rate is varied in a large scale to find the minimum learning rate and maximum learning rate area. In the second estimation, the learning rate is scaled to search the max learning rate area and finds the value of maximum learning rate. As shown in Fig. 2, in the first estimation, the learning rate range is set to $1e-5$ to $1e-2$, and it finds that the min learning rate $\eta_{\min}$ is near $1e-4$. In second estimation, the learning rate range is set to $1e-3$ to $7e-3$, and the peak of learning rate (Safe max learning rate $\eta_{\max}^S$) and the valley of learning rate (Escape max learning rate $\eta_{\max}^E$) would be recognized.

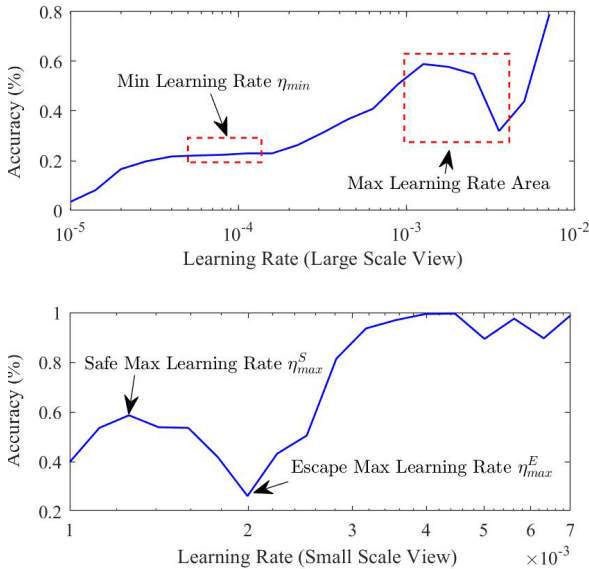


FIGURE 2. The learning rate range estimation method.

For most tasks, $\eta_{\max}^S$ can provide the safe upper boundary of the learning rate. But in this research, $\eta_{\max}^E$ is preferred since it can help MMCCLR to escape from the local minimum at each cyclic beginning to improve the final global accuracy. The effective of $\eta_{\max}^E$ would be shown in the case study section.

After determining the upper and lower boundary of learning rate, the proposed SECNN can be implemented.

V. CASE STUDIES AND RESULTS

In this section, the performance of SECNN is tested on three famous datasets for fault diagnosis, including CWRU bearing dataset, SPCP pump dataset and MFPT bearing dataset.

A. LEARNING RATE RANGE ESTIMATION USING LOGLR TEST

In the case studies section, three datasets are conducted for validating the performance of MMCCLR and SECNN. Here, the learning rate range estimation results of LogLR Test on three datasets are presented in TABLE 2 and Fig. 3, including the values of $\eta_{\min}$, $\eta_{\max}^S$ and $\eta_{\max}^E$.

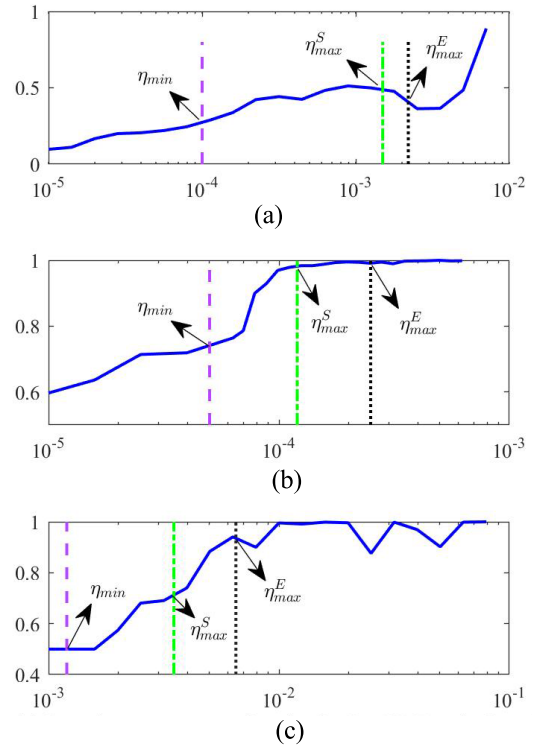


FIGURE 3. Learning rate range estimation by LogLR test. (a) Learning rate estimation by LogLR test in case 1. (b) Learning rate estimation by LogLR test in case 2. (c) Learning rate estimation by LogLR test in case 3.

Fig. 3(a) shows the learning rate range estimation results by LogLR Test on Load 3 in Case 1. Since the Load 1 and Load 2 have the similar curve with Load 3, they are not plotted in Fig. 3(a). Fig. 3(b) and Fig. 3(c) present the learning rate range estimation results by LogLR Test in Case 2 and Case 3. From Fig. 3, it can be seen that the most sensitive range can be

TABLE 2. Learning rate range estimation results on three case studies.

	Case 1: CWRU Dataset			Case 2:	Case 3:
	Load 1	Load 2	Load 3	SPCP	MFPT
$\eta_{\max}^E$	0.0022	0.002	0.002	0.00025	0.0065
$\eta_{\max}^S$	0.0015	0.00125	0.00125	0.00012	0.0035
$\eta_{\min}$	0.0001	0.0001	0.0001	0.00005	0.0012

recognized by LogLR Test, and this can promote the effective of MMCCLR. The values of $\eta_{\min}$, $\eta_{\max}^S$ and $\eta_{\max}^E$ for three cases are shown in TABLE 2.

B. ALGORITHM SETUP AND BASELINE METHODS

Ten different learning rate schedulers are implemented to show the effective of the MMCCLR. The baseline methods are based on stochastic gradient descent (SGD), and they are SGD with decay ratio (SGD-Decay) and stepwise decay (Step-Decay). Triangular CLR (TCLR) [40] is also used for comparison with MMCCLR, and its learning rate scheduler is denoted by equation (9). $s = [T, 2M]$ denotes the number of epochs of a half cycle.

$$\eta = \eta_{\min} + (\eta_{\max}^E - \eta_{\min}) \times \left(1 - \text{abs} \left(\frac{\text{mod}(t, 2bs)}{bs} - 1 \right) \right) \quad (9)$$

The details of these ten learning rate schedulers are:

- 1) SGD-Decay: SGD with Decay Ratio
- 2) Step-Decay: Step-Decay Stepwise Decay
- 3) T-Cyclic: TCLR method
- 4) C-Cyclic: MMCCLR method
- 5) SnapT-5: TCLR with Ensembles ($M = 5$)
- 6) SnapC-5: MMCCLR with Ensembles ($M = 5$)
- 7) SnapT-5-2: TCLR with Ensembles ($M = 5, N = 2$)
- 8) SnapC-5-2: MMCCLR with Ensembles ($M = 5, N = 2$)
- 9) SnapT-5-2S: TCLR with Ensemble using $\eta_{\max}^S$ ($M = 5, N = 2$)
- 10) SnapC-5-2S: MMCCLR with Ensembles using $\eta_{\max}^S$ ($M = 5, N = 2$)

T-Cyclic and C-Cyclic are TCLR and MMCCLR without SEL. SnapT-5 and SnapC-5 are TCLR and MMCCLR using SEL but no repeat running. SnapT-5-2 and SnapC-5-2 are TCLR and MMCCLR using SEL with repeat running. All these six implements use $\eta_{\max}^E$ as the maximum learning rate. SnapT-5-2S and SnapC-5-2S are TCLR and MMCCLR using SEL with repeat running using $\eta_{\max}^S$, and they are used for showing the effectiveness of $\eta_{\max}^E$.

C. CASE STUDY 1: CWRU BEARING DATASET

The bearing dataset of Case Western Reserve University (CWRU) is a popular dataset [33]. The health conditions of bearings contain the normal condition (NO) and three fault conditions. The three fault conditions are roller fault (RF), outer race fault (OF) and inner race fault (IF). Each fault

condition has three different damage sizes, which are 0.18mm, 0.36mm and 0.54mm. So, there are ten healthy conditions in this dataset. The experimental dataset is collected under four working conditions with different motor loads (0, 1, 2, 3 hp).

The task of this case study comes from Zhang *et al.* [33]. Six configurations of $1 \rightarrow 2$, $1 \rightarrow 3$, $2 \rightarrow 1$, $2 \rightarrow 3$, $3 \rightarrow 1$ and $3 \rightarrow 2$ are conducted in this case study. ' $a \rightarrow b$ ' denotes that the SECNN model is trained using dataset collected on load a , and tested the prediction accuracy on load b . The prediction accuracy (Acc) and the standard deviation (Std) are applied to measure the performance of SECNN models. During the training process, 5-fold cross validation is applied to obtain the reliable results.

1) RESULTS WITH DIFFERENT LEARNING RATE SCHEDULERS

The prediction results of SECNN with different learning rate schedulers are presented in TABLE 3, Fig. 4 and TABLE 4. From TABLE 3, the average prediction accuracies (AVG) of SGD-Decay and Step-Decay are 96.613% and 96.907%. The results show that all eight SECNN with CLR methods outperform these two baseline methods. The AVG of T-Cyclic, C-Cyclic, SnapT-5, SnapC-5, SnapT-5-2, SnapC-5-2, SnapT-5-2S and SnapC-5-2S are 96.923%, 97.308%, 97.087%, 97.360%, 97.198%, 97.543%, 97.110% and 97.518%. SnapC-5-2 obtains the best results among these methods. It also can be seen that the Std of SnapC-5-2 is significant, and this can validate the generalizability of SnapC-5-2 is better than other seven SECNN models.

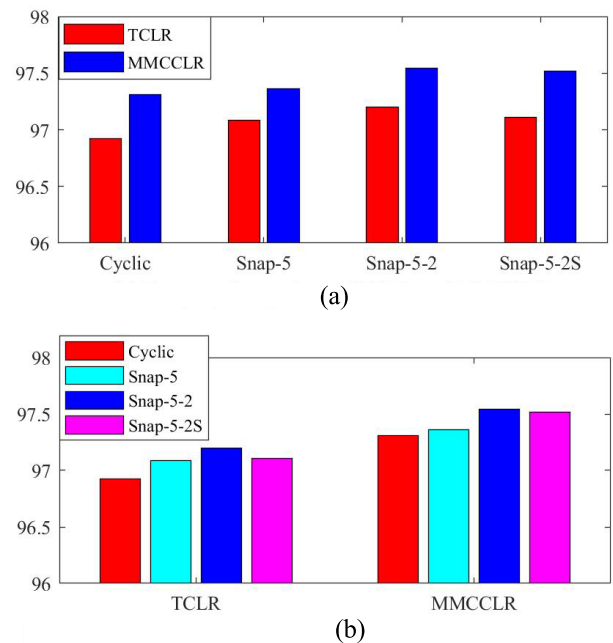


FIGURE 4. Category comparison results of SECNN with different learning rate schedulers in case 1. (a) Category comparison of TCLR and MMCCLR. (b) Category comparison of cyclic, snap-5, snap-5-2 and snap-5-2S.

TABLE 3. The results of SECNN with different learning rate schedulers in case 1 (%).

Methods	1→2		1→3		2→1		2→3		3→1		3→2		AVG
	Acc	Std	Acc	Std	Acc	Std	Acc	Std	Acc	Std	Acc	Std	
SGD-Decay	99.93	0.0447	96.47	1.3065	97.75	0.2475	98.1	0.8116	90.43	2.3344	97.00	2.4930	96.613
Step-Decay	99.96	0.0418	95.93	1.8461	97.74	0.1782	98.6	0.2622	90.86	1.2867	98.35	1.0624	96.907
T-Cyclic	99.97	0.0084	95.26	0.4661	97.75	0.2615	98.38	0.3966	92.11	1.2618	98.46	0.5293	96.988
C-Cyclic	99.96	0.0173	96.278	0.5492	97.84	0.0311	98.34	0.4657	92.85	0.9560	98.66	0.5661	97.321
SnapT-5	100	0	96.93	1.6404	98.00	0.0612	99.04	0.2080	90.71	1.4271	98.10	0.7211	97.130
SnapC-5	100	0	97.53	1.6855	98.03	0.0908	98.91	0.5920	91.93	1.0462	98.71	0.3110	97.518
SnapT-5-2	100	0	97.04	1.0991	98.00	0.1118	98.93	0.2564	90.8	0.6072	98.42	0.2253	97.198
SnapC-5-2	100	0	98.47	0.4382	98.06	0.0224	98.97	0.2490	91.27	1.1206	98.49	0.5042	97.543
SnapT-5-2S	100	0	96.05	1.6283	97.98	0.0570	99.04	0.1140	91.23	1.5291	98.22	0.9011	97.087
SnapC-5-2S	99.99	0.0224	96.63	1.7087	98.02	0.0570	99.00	0.1768	91.77	0.9622	98.75	0.4287	97.360

Fig. 4 presents the category comparison of eight learning rate schedulers. Fig. 4(a) presents the TCLR and MMCCLR, and the categories are CLR (denoted as Cycle), snapshot ensemble (denoted as Snap-5), snapshot ensemble with repeated running (denoted as Snap-5-2) and snapshot ensemble and repeated running using $\eta_{\max}^S$ (denoted as Snap-5-2). It can be seen that the proposed MMCCLR is superior to TCLR on all categories. Fig. 4(b) presents the comparison results of different snapshot ensemble models. It can be seen that Snap-5-2 is better than snap-5 and cyclic methods in both TCLR and MMCCLR categories. It should be noted that snap-5-2 is better than snap-5-2S in both TCLR and MMCCLR categories, showing that $\eta_{\max}^E$ performs better than $\eta_{\max}^S$ in this case study.

TABLE 4. Training times of SECNN with different learning rate schedulers in case 1 (seconds).

Methods	Training Time of different loads		
	1	2	3
SGD-Decay	2048.2	2108.03	2121.68
Step-Decay	2004.55	2052.22	2067.47
T-Cyclic	2130.06	2097.2	2083.76
C-Cyclic	2011.67	1971.42	2066.31
SnapT-5	2211.61	2197.35	2128.16
SnapC-5	2185.06	2191.92	2152.6
SnapT-5-2	4424.05	4375.14	4456.99
SnapC-5-2	4271.81	4324.99	4325.77
SnapT-5-2S	4437.48	4420.32	4381.47
SnapC-5-2S	4155.99	4312.4	4236.26

TABLE 4 presents the training times of SECNN with different learning rate schedulers in this case study. From the results, it can be seen that the training time of SGD-Decay, Step-Decay, T-Cyclic, C-Cyclic, SnapT-5 and SnapC-5 are

very close to each other. The training times of SnapT-5 and SnapC-5 are higher than SGD-Decay, Step-Decay, T-Cyclic, C-Cyclic, but the increase of training times are less than 11% (2211.61/2004.55-100%). However, the training times of SnapT-5-2, SnapC-5-2, SnapT-5-2S and SnapC-5-2S are almost the twice of T-Cyclic, C-Cyclic, because they run the SECNN for two time repeatedly. It should be noted that the training times of MMCCLR variants (C-Cyclic, SnapC-5, SnapC-5-2 and SnapC-5-2S) is less than the TCLR variants (T-Cyclic, SnapT-5, SnapT-5-2 and SnapT-5-2S).

2) COMPARED WITH OTHER DL METHODS

In order to validate the performance of proposed SECNN, the results of SECNN are compared with other DL methods in literature. They are WDCNN, WDCNN(AdaBN) from Zhang *et al.* [33] and TICNN, Ensemble TICNN from W. Zhang *et al.* [36]. SVM and ANN are selected for the comparison [36]. The comparison results are presented in TABLE 5.

TABLE 5. Comparison results of SECNN with other DL methods in case 1 (%).

Methods	1→2	1→3	2→1	2→3	3→1	3→2	AVG
SECNN	100	98.47	98.06	98.97	91.27	98.49	97.543
WDCNN	99.2	91.0	95.1	91.5	78.1	85.1	90.0
WDCNN (AdaBN)	99.4	93.4	97.5	97.2	88.3	99.9	95.9
TICNN	99.1	90.7	97.4	98.8	89.2	97.6	95.5
Ensemble TICNN	99.5	91.1	97.6	99.4	90.2	98.7	96.1
SVM	68.6	60.0	73.2	67.6	68.4	62.0	66.6
ANN	82.1	85.6	71.5	82.4	81.8	79.0	80.4

In this comparison, SECNN uses the results of SnapC-5-2. As shown in TABLE 5, SECNN has achieved a good prediction accuracy. The average prediction accuracy of SECNN

is 97.543%, which outperforms Ensemble TICNN, TICNN, ADCNN(AdaBN) and WDCNN, whose AVG are 96.1%, 95.5%, 95.9% and 90.0% respectively. SECNN also outperforms SVM, and ANN. These results indicate the performance of SECNN in this case study.

D. CASE STUDY 2: SPCP PUMP DATASET

This case study presents the results of SECNN on self-priming centrifugal pump dataset [46]. In the data acquisition system, the acceleration sensor is installed on a specific pedestal above the motor housing and the rotation speed is 2,900 per minute. The sampling frequency of vibration signals is 10,240Hz. There are four fault conditions and one normal condition (NO). The fault conditions are bearing roller wearing (BR), inner race wearing (IR), outer race wearing (OR), and impeller wearing (IW) fault condition.

1) RESULTS WITH DIFFERENT LEARNING RATE SCHEDULERS

The prediction results of SECNN with different learning rate schedulers are presented in TABLE 4 and Fig. 5. 5-fold cross validation is also applied. From TABLE 4, the results of SGD-Decay and Step-Decay are 99.872% and 99.880%. All eight SECNN with CLR methods outperform two baseline methods in both *Acc* and *Std* terms except T-Cyclic. The *Acc* of T-Cyclic, C-Cyclic, SnapT-5, SnapC-5, SnapT-5-2, SnapC-5-2, SnapT-5-2S and SnapC-5-2S are 99.864%, 99.888%, 99.952%, 99.968%, 99.968%, 99.976%, 99.904% and 99.912%. SnapC-5-2 obtains the best result and has the best generalizability among these methods.

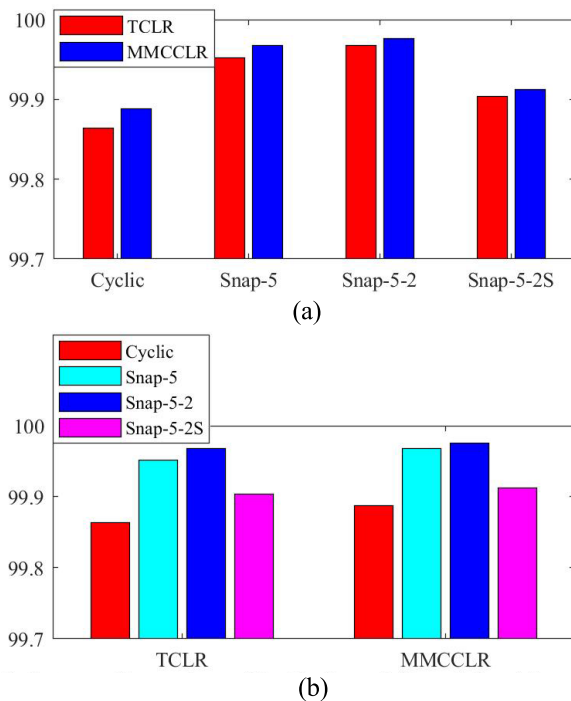


FIGURE 5. Category comparison results of SECNN with different learning rate schedulers in case 2. (a) Category comparison of TCLR and MMCCLR. (b) Category comparison of cyclic, snap-5, snap-5-2 and snap-5-2S.

TABLE 6. The results of SECNN with different learning rate schedulers in case 2 (%).

	<i>Acc</i>	<i>Std</i>	<i>Training Times</i>
SGD-Decay	99.872	0.0593	1267.5
Step-Decay	99.880	0.0566	1217.16
T-Cyclic	99.864	0.0518	1279.09
C-Cyclic	99.888	0.0901	1250.09
SnapT-5	99.952	0.0438	1343.49
SnapC-5	99.968	0.0335	1251.99
SnapT-5-2	99.968	0.0438	2681.9
SnapC-5-2	99.976	0.0219	2587.8
SnapT-5-2S	99.904	0.0358	2564.79
SnapC-5-2S	99.912	0.0438	2603.75

The training times of SECNN with these ten learning rate schedulers are also presented in TABLE 6. It can be seen that the training time of SGD-Decay, Step-Decay, T-Cyclic, C-Cyclic, SnapT-5 and SnapC-5 are close. The maximum increase of training times is nearly 10% (1343.49/1217.16-100%). The SECNN with two repeated running variants (SnapT-5-2, SnapC-5-2, SnapT-5-2S and SnapC-5-2S) consumes almost the twice training times compared with SGD-Decay.

Fig. 5 presents the category comparison of these eight learning rate schedulers. Fig. 5(a) presents the TCLR and MMCCLR, it can be seen that MMCCLR is also superior to TCLR on all categories in this case. Fig. 5(b) presents the comparison results of different snapshot ensemble models. It can be seen that Snap-5-2 is better than snap-5, snap-5-2S and cyclic methods in both TCLR and MMCCLR categories. These comparison results agree with the results of case study 1.

2) COMPARED WITH OTHER METHODS

The proposed SECNN is compared with the Speeded Up Robust Features based Probabilistic Neural Networks (SURF based PNN) proposed in Lu *et al.* [46], CNN models proposed by Wen *et al.* [11] and subset based deep auto-encoder (SBTDA) from Zhang *et al.* [32]. SECNN uses the results of SnapC-5-2, and the comparison results are presented on TABLE 7.

TABLE 7. Comparison results with other methods in case 2 (%).

Methods	Mean Accuracy
SECNN	99.976
SBTDA	99.60
CNN	99.48
SURF based PNN	98.33

The prediction accuracy of proposed SECNN is 99.976%, while these of SBTDA, CNN and SURF based PNN are 99.60%, 99.48% and 98.33%. These results show that SECNN has obtained a good result and it outperforms these two methods in this case study.

E. CASE STUDY 3: MFPT BEARING DATASET

This case study presents the results of SECNN on the dataset provided by the Machinery Failure Prevention Technology (MFPT) Society [47]. A test rig with a NICE bearing gathered acceleration data for baseline conditions at 270 lbs of load and a sampling rate of 97,656Hz for six seconds. In total, ten outer-raceway and seven inner-raceway fault conditions were tracked. Three outer race faults included 270 lbs of load and a sampling rate of 97,656Hz for six seconds. Seven additional outer race faults were assessed at varying loads: 25, 50, 100, 150, 200, 250, and 300lbs. The sample rate for the faults was 48,828Hz for three seconds. Seven inner race faults were analyzed with varying loads of 0, 50, 100, 150, 200, 250, and 300 lbs. The sample rate for the inner race faults was 48,848Hz for three seconds. There are three health conditions, including normal condition (NO), outer-raceway fault (OU) and inner-raceway fault (IN).

1) RESULTS WITH DIFFERENT LEARNING RATE SCHEDULERS

The prediction results of SECNN with different learning rate schedulers are presented in TABLE 8 and Fig. 6. 5-fold cross validation is also applied. The *Acc* of SGD-Decay and Step-Decay are both 99.85%. All eight SECNN with CLR methods outperform these two baseline methods. The *Acc* and *Std* of SnapC-5, SnapT-5-2, SnapC-5-2, SnapT-5-2S and SnapC-5-2S are all 99.94% and 0.0224, which are better than T-Cyclic, C-Cyclic, and SnapT-5.

TABLE 8. The results of SECNN with different learning rate schedulers in case 3 (%).

	<i>Acc</i>	<i>Std</i>	<i>Training Times</i>
SGD-Decay	99.85	0.0791	1009.87
Step-Decay	99.85	0.0707	991.394
T-Cyclic	99.88	0.0652	1001.52
C-Cyclic	99.90	0.1000	977.473
SnapT-5	99.93	0.0274	1036.05
SnapC-5	99.94	0.0224	1027.46
SnapT-5-2	99.94	0.0224	2138.14
SnapC-5-2	99.94	0.0224	2081.48
SnapT-5-2S	99.94	0.0224	2131.85
SnapC-5-2S	99.94	0.0224	2077.87

The training times of SECNN with these ten learning rate schedulers are also presented in TABLE 8. The training time of SGD-Decay, Step-Decay, T-Cyclic, C-Cyclic, SnapT-5 and SnapC-5 are also close in this case study. The maximum increase of training times is nearly 6% (1036.05/977.473-100%). The SECNN with two repeated running

variants (SnapT-5-2, SnapC-5-2, SnapT-5-2S and SnapC-5-2S) also consumes almost the twice training times compared with SGD-Decay. The training times of SECNN variants with MMCCLR (C-Cyclic, SnapC-5, SnapC-5-2 and SnapC-5-2S) is less than SECNN variants with TCLR (T-Cyclic, SnapT-5, SnapT-5-2 and SnapT-5-2S) in this case study.

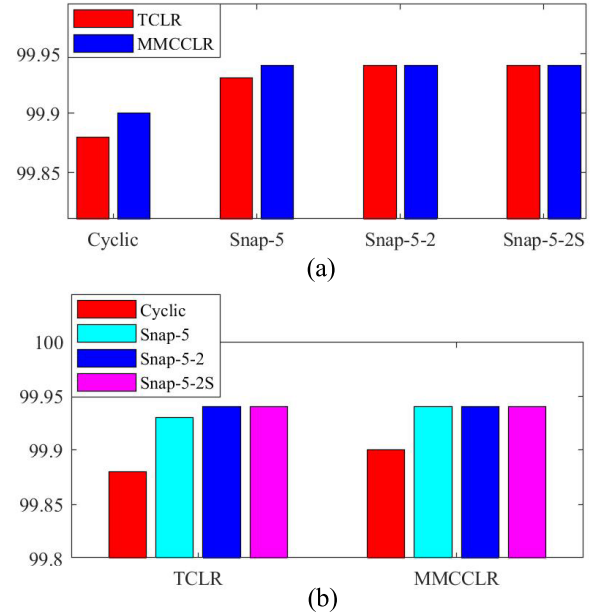


FIGURE 6. Category comparison results of SECNN with different learning rate schedulers in case 3. (a) Category comparison of TCLR and MMCCLR. (b) Category comparison of cyclic, snap-5, snap-5-2 and snap-5-2S.

Fig. 6 presents the category comparison of these eight learning rate schedulers. Fig. 6(a) presents the TCLR and MMCCLR, it can be seen that MMCCLR is better than TCLR on Snap-5 category and as good as TCLR on snap-5-2 and snap-5-2S. Fig. 6(b) presents the comparison results of different snapshot ensemble models. It can be seen that Snap-5-2 is better than cyclic methods in both TCLR and MMCCLR categories, and as good as TCLR to Snap-5-2S. These results also indicate that the $\eta_{\max}^E$ has equal effect as $\eta_{\max}^S$ in this case study.

2) COMPARED WITH OTHER DL METHODS

The proposed SECNN is compared with CNN with Scalogram 32*32, CNN with Scalogram 96*96, CNN with Spectrogram 32*32, CNN with Spectrogram 96*96, CNN with HHT 32*32 and CNN with HHT 96*96 proposed by Verstraete *et al.* [47], and CNN-VAE from San Martin *et al.* [48]. SECNN uses the prediction accuracies of SnapC-5-2, and the results are presented on TABLE 9.

The prediction accuracy of proposed SECNN is 99.94%, which is better than CNN with Scalogram 32*32, CNN with Spectrogram 32*32, CNN with Spectrogram 96*96, CNN with HHT 32*32, CNN with HHT 96*96 and CNN-VAE, and as good as CNN with Scalogram 96*96.

TABLE 9. Comparison results with other DL methods in case 3 (%).

Methods	Mean Accuracy
SECNN	99.94
CNN with Scalogram 32*32	99.7
CNN with Scalogram 96*96	99.9
CNN with Spectrogram 32*32	82.0
CNN with Spectrogram 96*96	91.3
CNN with HHT 32*32	75.9
CNN with HHT 96*96	92.9
CNN-VAE	92.58

F. DISCUSSION

In this section, the proposed SECNN with ten different learning rate schedulers are implemented, and the results show that the proposed MMCCLR outperforms TCLR in case 1 and case 2 and is better than or as good as TCLR in case 3. These results show the effectiveness of MMCCLR.

It also can be seen that SECNN with MMCCLR obtains better prediction accuracies compared with baseline methods on all case, and its results also outperform many other DL and traditional methods. The advantage of SECNN is that it can improve the performance and generalizability without increasing the computation time.

What's more, the results also indicate that $\eta_{\max}^E$ is superior to $\eta_{\max}^S$. In Snap-5-2 category, $\eta_{\max}^E$ performs better than $\eta_{\max}^S$ in case 1 and case 2, and is as good as $\eta_{\max}^S$ in case 3. These results provide experimental support for the effective of $\eta_{\max}^E$.

VI. CONCLUSION AND FUTURE RESEARCHES

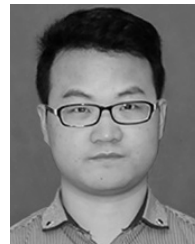
This research presents a new snapshot ensemble convolutional neural network (SECNN) for fault diagnosis. The main contributions are developing MMCCLR, proposing a new LogLR Test and developing SECNN for fault diagnosis. The proposed SECNN is conducted on three famous datasets. The results show the effectiveness of SECNN. SECNN with MMCCLR obtains better prediction accuracies compared with baseline methods on all cases. The results show that SECNN could improve the performance and generalizability without increasing the computation time. SECNN obtains start-of-art results, and outperforms several other DL and traditional methods.

The limitations of the proposed SECNN include the following aspects. Firstly, the fault data in real industrial applications is limited, but SECNN ignores the imbalance of the volume of the normal data and fault data. Secondly, most fault types should be pre-defined and the fault which has not been learned would be misclassified into one of the known ones. Based on the limitations, the future researches can be conducted in the following ways. Firstly, the SECNN could be extended to handle with the imbalanced data in fault diagnosis field. Secondly, the method can be modified to detect and recognize the unknown fault conditions.

REFERENCES

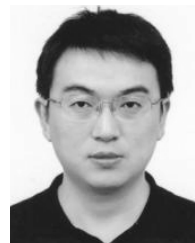
- [1] X. Dai and Z. Gao, "From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis," *IEEE Trans. Ind. Informat.*, vol. 9, no. 4, pp. 2226–2238, Nov. 2013.
- [2] W. Qian, S. Li, and J. Wang, "A new transfer learning method and its application on rotating machine fault diagnosis under variant working conditions," *IEEE Access*, vol. 6, pp. 69907–69917, 2018.
- [3] S. Yin, S. X. Ding, X. Xie, and H. Luo, "A review on basic data-driven approaches for industrial process monitoring," *IEEE Trans. Ind. Electron.*, vol. 61, no. 11, pp. 6418–6428, Nov. 2014.
- [4] M. He and D. He, "Deep learning based approach for bearing fault diagnosis," *IEEE Trans. Ind. Appl.*, vol. 53, no. 3, pp. 3057–3065, May/Jun. 2017.
- [5] S. Yin, H. Gao, J. Qiu, and O. Kaynak, "Fault detection for non-linear process with deterministic disturbances: A just-in-time learning based data driven method," *IEEE Trans. Cybern.*, vol. 47, no. 11, pp. 3649–3657, Nov. 2017.
- [6] H. Shao, H. Jiang, H. Zhang, W. Duan, T. Liang, and S. Wu, "Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing," *Mech. Syst. Signal Process.*, vol. 100, pp. 743–765, Feb. 2018.
- [7] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mech. Syst. Signal Process.*, vol. 108, pp. 33–47, Aug. 2018.
- [8] J. Wang, Y. Ma, L. Zhang, R. X. Gao, and D. Wu, "Deep learning for smart manufacturing: Methods and applications," *J. Manuf. Syst.*, vol. 48, pp. 144–156, Jul. 2018.
- [9] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.
- [10] C. Sun, M. Ma, Z. B. Zhao, and X. Chen, "Sparse deep stacking network for fault diagnosis of motor," *IEEE Trans. Ind. Informat.*, vol. 14, no. 7, pp. 3261–3270, Jul. 2018. doi: [10.1109/TII.2018.2819674](https://doi.org/10.1109/TII.2018.2819674).
- [11] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5990–5998, Jul. 2018.
- [12] H. Shao, H. Jiang, Y. Lin, and X. Li, "A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders," *Mech. Syst. Signal Process.*, vol. 102, pp. 278–297, Mar. 2018.
- [13] H. Chen and X. Yao, "Regularized negative correlation learning for neural network ensembles," *IEEE Trans. Neural Netw.*, vol. 20, no. 12, pp. 1962–1979, Dec. 2009.
- [14] G. Huang, Y. Li, G. Pleiss, Z. Liu, J. E. Hopcroft, and K. Q. Weinberger, "Snapshot ensembles: Train 1, get M for free," in *Proc. ICLR*, Toulon, France, Apr. 2017.
- [15] L. Zhao and X. Wang, "A deep feature optimization fusion method for extracting bearing degradation features," *IEEE Access*, vol. 6, pp. 19640–19653, 2018.
- [16] L. Ren, W. Lv, S. Jiang, and Y. Xiao, "Fault diagnosis using a joint model based on sparse representation and SVM," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 10, pp. 2313–2320, Oct. 2016.
- [17] M. D. Prieto, G. Cirrincione, A. G. Espinosa, J. A. Ortega, and H. Henao, "Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks," *IEEE Trans. Ind. Electron.*, vol. 60, no. 8, pp. 3398–3407, Aug. 2013.
- [18] R. Ahmed, M. El Sayed, S. A. Gadsden, J. Tjong, and S. Habibi, "Automotive internal-combustion-engine fault detection and classification using artificial neural network techniques," *IEEE Trans. Veh. Technol.*, vol. 64, no. 1, pp. 21–33, Jan. 2015.
- [19] X. Peng, J. Feng, S. Xiao, W.-Y. Yau, J. T. Zhou, and S. Yang, "Structured autoencoders for subspace clustering," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 5076–5086, Oct. 2018.
- [20] E. Yang, C. Deng, W. Liu, X. Liu, D. Tao, and X. Gao, "Pairwise relationship guided deep hashing for cross-modal retrieval," in *Proc. 31st AAAI Conf. Artif. Intell. (AAAI)*, 2017, pp. 1618–1625.
- [21] C. Deng, Z. Chen, X. Liu, X. Gao, and D. Tao, "Triplet-based deep hashing network for cross-modal retrieval," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 3893–3903, Aug. 2018.
- [22] Z. Huang, H. Zhu, J. T. Zhou, and X. Peng, "Multiple marginal Fisher analysis," *IEEE Trans. Ind. Electron.*, to be published. doi: [10.1109/TIE.2018.2870413](https://doi.org/10.1109/TIE.2018.2870413).
- [23] E. Yang, C. Deng, C. Li, W. Liu, J. Li, and D. Tao, "Shared predictive cross-modal deep quantization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5292–5303, Nov. 2018.

- [24] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mech. Syst. Signal Process.*, vol. 115, pp. 213–237, Jan. 2019.
- [25] Y. Qin, X. Wang, and J. Zou, "The optimized deep belief networks with improved logistic Sigmoid units and their application in fault diagnosis for planetary gearboxes of wind turbines," *IEEE Trans. Ind. Electron.*, vol. 66, no. 5, pp. 3814–3824, May 2019. doi: [10.1109/TIE.2018.2856205](https://doi.org/10.1109/TIE.2018.2856205).
- [26] H. D. Shao, H. K. Jiang, X. Zhang, and M. G. Niu, "Rolling bearing fault diagnosis using an optimization deep belief network," *Meas. Sci. Technol.*, vol. 26, no. 11, Nov. 2015, Art. no. 115002.
- [27] M. Gan, C. Wang, and C. Zhu, "Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings," *Mech. Syst. Signal Process.*, vols. 72–73, pp. 92–104, May 2016.
- [28] Z. Chen and W. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1693–1702, Jul. 2017.
- [29] Y. Qi, C. Shen, D. Wang, J. Shi, X. Jiang, and Z. Zhu, "Stacked sparse autoencoder-based deep network for fault diagnosis of rotating machinery," *IEEE Access*, vol. 5, pp. 15066–15079, Jul. 2017.
- [30] L. Wen, L. Gao, and X. Li, "A new deep transfer learning based on sparse auto-encoder for fault diagnosis," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 49, no. 1, pp. 136–144, Jan. 2017. doi: [10.1109/TSMC.2017.2754287](https://doi.org/10.1109/TSMC.2017.2754287).
- [31] J. Sun, C. Yan, and J. Wen, "Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 1, pp. 185–195, Jan. 2018.
- [32] Y. Zhang, X. Li, L. Gao, and P. Li, "A new subset based deep feature learning method for intelligent fault diagnosis of bearing," *Expert Syst. Appl.*, vol. 110, pp. 125–142, Nov. 2018.
- [33] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, p. 425, 2017.
- [34] M. Xia, T. Li, L. Xu, L. Liu, and C. W. de Silva, "Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 1, pp. 101–110, Feb. 2018.
- [35] B. Zhang, W. Li, X.-L. Li, and S.-K. Ng, "Intelligent fault diagnosis under varying working conditions based on domain adaptive convolutional neural networks," *IEEE Access*, vol. 6, pp. 66367–66384, 2018.
- [36] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," *Mech. Syst. Signal Process.*, vol. 100, pp. 439–453, Feb. 2018.
- [37] Z. Wu, W. Lin, and Y. Ji, "An integrated ensemble learning model for imbalanced fault diagnostics and prognostics," *IEEE Access*, vol. 6, pp. 8394–8402, 2018.
- [38] C. Zhang, P. Lim, A. K. Qin, and K. C. Tan, "Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2306–2318, Oct. 2017.
- [39] J. W. Wang, C. Sun, Z. Zhao, and X. Chen, "Feature ensemble learning using stacked denoising autoencoders for induction motor fault diagnosis," in *Proc. IEEE Prognostics Syst. Health Manage. Conf. (PHM-Harbin)*, Jul. 2017, pp. 1–6.
- [40] L. N. Smith, "Cyclical learning rates for training neural networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Santa Rosa, CA, USA, Mar. 2017, pp. 464–472.
- [41] L. Jing, T. Wang, M. Zhao, and P. Wang, "An adaptive multi-sensor data fusion method based on deep convolutional neural networks for fault diagnosis of planetary gearbox," *Sensors*, vol. 17, no. 2, p. 414, 2017.
- [42] X. Guo, L. Chen, and C. Shen, "Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis," *Measurement*, vol. 93, pp. 490–502, Nov. 2016.
- [43] I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," in *Proc. ICLR*, Toulon, France, Apr. 2017.
- [44] Y. Xie and T. Zhang, "Fault diagnosis for rotating machinery based on convolutional neural network and empirical mode decomposition," *Shock Vib.*, vol. 2017, Aug. 2017, Art. no. 3084197.
- [45] R. G. Stockwell, L. Mansinha, and R. P. Lowe, "Localization of the complex spectrum: The S transform," *IEEE Trans. Signal Process.*, vol. 44, no. 4, pp. 998–1001, Apr. 1996.
- [46] C. Lu, Y. Wang, M. Ragulskis, and Y. Cheng, "Fault diagnosis for rotating machinery: A method based on image processing," *PLoS ONE*, vol. 11, no. 10, Oct. 2016, Art. no. e0164111.
- [47] D. Verstraete, A. Ferrada, E. L. Droguett, V. Meruane, and M. Modarres, "Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings," *Shock Vib.*, vol. 2017, Oct. 2017, Art. no. 5067651.
- [48] G. S. Martin, E. L. Droguett, V. Meruane, and M. das Chagas Moura, "Deep variational auto-encoders: A promising tool for dimensionality reduction and ball bearing elements fault diagnosis," *Structural Health Monitor.*, 2018. doi: [10.1177/1475921718788299](https://doi.org/10.1177/1475921718788299).



LONG WEN received the Ph.D. degree in industrial engineering from the Huazhong University of Science and Technology, China, in 2014, where he holds a postdoctoral position with the State Key Laboratory of Digital Manufacturing Equipment and Technology, Department of Industrial and Manufacturing Systems Engineering, School of Mechanical Science and Engineering.

His research interests include machine learning and intelligent algorithm.



LIANG GAO (M'08) received the Ph.D. degree in mechatronic engineering from the Huazhong University of Science and Technology, China, in 2002, where he is currently a Professor with the State Key Laboratory of Digital Manufacturing Equipment and Technology, Department of Industrial and Manufacturing Systems Engineering, School of Mechanical Science and Engineering.

He had published more than 170 refereed papers. His research interests include operations research and optimization, big data, and machine learning.



XINYU LI received the Ph.D. degree in industrial engineering from the Huazhong University of Science and Technology, China, in 2009, where he is currently an Associate Professor with the State Key Laboratory of Digital Manufacturing Equipment and Technology, Department of Industrial and Manufacturing Systems Engineering, School of Mechanical Science and Engineering.

He had published more than 60 refereed papers. His research interests include intelligent algorithm, big data, and machine learning.

...