





Using Multiple-Feature-Spaces-Based Deep Learning for Tool Condition Monitoring in Ultraprecision Manufacturing

Chengming Shi , George Panoutsos , Bo Luo , Hongqi Liu, Bin Li , and Xu Lin

Abstract—Tool condition monitoring is critical in ultraprecision manufacturing in order to optimize the performance of the overall process, while maintaining the desired part quality. Recently, deep learning has been successfully applied to numerous classification tasks in manufacturing, often to forecast part quality. In this paper, a novel deep learning data-driven modeling framework is presented, which includes a fusion of multiple stacked sparse autoencoders for tool condition monitoring in ultraprecision machining. The proposed computational framework consists of two main structures. First, a training model that is designed with the ability to process multiple parallel feature spaces to learn the lower-level features. Second, a feature fusion structure that is used to learn the higher-level features and associations to tool wear. To achieve this learning structure, a modified loss function is utilized that enhances the feature extraction and classification tasks. A dataset from a real manufacturing process is used to demonstrate the performance of the proposed framework. Experimental results and simulations show that the proposed method successfully classifies the ultraprecision machining case study with over 96% accuracy, while also outperforming comparable methodologies.

Index Terms—Deep learning (DL), feature fusion, feature spaces, tool condition monitoring (TCM), ultraprecision manufacturing process.

NOMENCLATURE

AE	Acoustic emission.
ANFIS	Adaptive neuro-fuzzy inference system.
ANN	Artificial neural network.
BP	Back propagation.

BPNN	Backpropagation neural network.
CNC	Computer numerical control.
DL	Deep learning.
FD	Frequency domain.
FMSSAEs	Fusion of multiple stacked sparse autoencoders.
HMM	Hidden Markov model.
ML	Machine learning.
NN	Neural network.
PCA	Principal component analysis.
rms	Root mean square.
RNN	Recurrent neural network.
SAE	Sparse autoencoder.
SSAE	Stacked sparse autoencoder.
SVM	Support vector machine.
TCM	Tool condition monitoring.
TD	Time domain.
v-SVR	v-Support vector regression.
WD	Wavelet domain.
WT	Wavelet transform.

I. INTRODUCTION

ULTRAPRECISION manufacturing is being widely used in numerous industrial applications, such as in microsensors, optical elements, microsatellite components, etc. Compared with traditional machining, via ultraprecision manufacturing, one can achieve a higher precision and a better surface finish for the workpiece [1], mainly due to the diamond-based tool. Moreover, microwear of the cutting tool has a significant influence on the surface quality, which will further have a measurable impact on the production efficiency and the part yield rate. It has been shown that in CNC manufacturing processes, tool wear significantly affects the quality of parts, hence yield rate too; by monitoring the condition of the machining tool, the overall manufacturing process can be improved [2] and potentially optimized to achieve a high yield rate. Thus, there is a crucial need for methods that can accurately quantify the tool condition and offer autonomous decisions on tool life in ultraprecision manufacturing processes.

In recent decades, many artificial intelligence methodologies have been widely used for TCM in traditional manufacturing and processing. Generally, two essentials are necessary, which are: 1) man-made (expert knowledge) feature extraction and design, such as the identification of statistical characteristics, FD index,

Manuscript received February 23, 2018; revised May 14, 2018; accepted June 20, 2018. Date of publication July 26, 2018; date of current version December 28, 2018. This work was supported by the National Natural Science Foundation of China under Grant 51705174 and Grant 51625502. (Corresponding authors: Bo Luo and Hongqi Liu.)

C. Shi and X. Lin are with the Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: 513864035@qq.com; m201670426@hust.edu.cn).

G. Panoutsos is with Automatic Control and Systems Engineering, The University of Sheffield, S1 3JD Sheffield, United Kingdom (e-mail: g.panoutsos@sheffield.ac.uk).

B. Luo, H. Liu, and B. Li are with the School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: hglobo@163.com; liuhongqi328@163.com; li_bin_hust@163.com).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIE.2018.2856193

and wavelet coefficients; and 2) shallow-layer model development and study, such as NN, HMM, SVM, etc. [3]–[6]. Patra [7] proposed an approach based on the rms of wavelet packet coefficients captured from AE signals and an ANN model for TCM and indicated that rms values of the wavelet coefficients show a positive correlation to increasing drill wear. The rms and ratio of power statistics selected from AE spectra were studied by Martins *et al.* [5], and an ANN model was utilized in classifying tool wear states; it was shown that the features of specific frequency bands of the signal are effective in characterizing the wear condition of the tool. Ku *et al.* [8] researched the mapping relationship between wavelet features extracted from vibration signals and three predefined tool wear conditions based on BPNN. Statistical features in the TD and FD extracted from vibration and power signals via wavelet packet decomposition were also discussed by Niaki *et al.* [9], and an RNN was used for tool wear estimation; the authors studied the application of sensor information fusion in order to increase the estimation performance of the NN and the results showed only a maximum of 13% relative error in estimating tool wear. Mas-sol *et al.* [10] studied the relationship between tool condition and several features extracted from force and the AE signal and trained an ANFIS to monitor the wear state. The authors developed an extended Takagi Sugeno to correlate sensory signals with several cutter health conditions; however, the accuracy of the model for unknown tool parameters is still low. Ren *et al.* [11] developed multiple AE signal features to reveal the tool condition, and an ANFIS is constructed as the wear states' classifier. Meanwhile, type-2 fuzzy-logic-based tool life estimation can evaluate the tool life along the cutting process and can also predict the uncertainty in the tool life estimation. Qiu and Xie [12] developed a hybrid approach based on HMM and the rms of wavelet packets that are used to estimate the wear state in TCM. Shi and Gindy [13] combined the PCA used for feature extraction from multiple sensory signals with a least square SVM (LS-SVM) model to predict the tool state in a broaching operation. Results showed that the PCA is very efficient at capturing the underlying features, and combined with LS-SVM, it is possible to avoid local optima and yield good generalization properties. Fourteen TD features sensitive to tool wear were calculated by Li *et al.* [3], and correlation analysis was utilized for feature selection and v-SVR for tool wear condition monitoring. The authors demonstrated that the model has a good accuracy of up to 96.76%; however, it was indicated that the model is only suitable for cases with a small sample size. It is evident from the literature that the workflow of feature extraction followed by machine learning data-driven modeling has been successfully applied in TCM. However, such methodologies are only used and demonstrated in the laboratory, rather than in a real manufacturing environment, which implies controlled conditions and low susceptibility to noise, uncertainty, etc. In addition, it appears that the model's accuracy in tool condition identification and prognosis highly depends on the sensitivity of the extracted features [14], which is often performed systematically but not autonomously (via expert knowledge).

A number of significant challenges will need to be addressed if the already developed methods for TCM in traditional/standard machining processes are to be used

effectively in ultraprecision manufacturing process too. First, using specific feature extraction and selection methods for systematically designing and selecting suitable features requires prior domain knowledge and expert input [15], [16]. In addition, ultraprecision manufacturing has the characteristics of small cutting allowance, less vibration, and weak signal features. In ultraprecision manufacturing, there is very little research on expert-based feature selection and extraction. Even if there were a significant body of literature, relying on expert knowledge (human expert) would limit the potential use of the system. Autonomous feature extraction and selection would be preferred in this case. Second, In ultraprecision manufacturing, it is found that even when the cutting tool exhibits a small wear rate, the impact on part quality can be significant. In addition, the tool wear process is highly complex and nonlinear, thus challenging to identify via methods not developed specifically for ultraprecision machining, i.e., methods originally developed for traditional/classical machining [17]–[19].

Compared to traditional machine learning and intelligent system methods, DL has the most notable advantage of a powerful complex nonlinear learning ability. The conspicuous difference between DL and shallow learning neural-based methods is that the former can adaptively learn valuable features from original data [20], [24], [25]; autonomous feature extraction and selection are also possible as part of the overall data-driven modeling process.

During the last few years, DL-based models have been developed and applied in intelligent fault diagnosis, especially the gear and bearing fault diagnosis. However, it is known that the performance of deep models largely depends on the original input data (quantity and quality). Thus, choosing TD data [18], [21] or FD data [22], [23] as inputs to DL models for fault diagnosis will have the following challenges: 1) signals will have different properties in different feature spaces [14]. In fact, the influence of specific input types in different spaces on the performance of DL models is still not clear; and 2) ultraprecision manufacturing processes have the characteristics of very strict cutting tolerance and, in general, a less pronounced vibration signature. Features extracted from a single feature space in ultraprecision machining may have the limitation of being scarce, which cannot meet the requirement for tool condition identification. Thus, it is important to develop a new framework for TCM and create a bespoke DL model to address multiple feature sets, autonomous feature selection capability, and robust performance suitable for real manufacturing environments.

In this paper, a multiple feature spaces based and a bespoke DL-based framework for TCM in ultraprecision manufacturing process are proposed. In this study, all the data have been collected from a real manufacturing plant, relevant to shell machining for consumer electronics. The datasets include measurements of vibration and have been preprocessed by fast Fourier transform (FFT) and wavelet transform (WT) to create a preliminary signal set. First, a new “parallel training model” is designed, which is suitable for three kinds of feature spaces (TD data, FD data, and WD data), to learn the low-layer features of the DL structure (feature selection/extraction). Then, a feature fusion model is employed to learn to correlate the high-layer features to the tool condition. To achieve the proposed DL

structure, and parallel learning, a modified loss function and a training framework are used to improve the performance of the designed DL structure. The contribution of this paper is that via the proposed new DL modeling framework, we take advantage of the implicit feature, i.e., learning ability, of deep layer models as well as the characteristics of different feature spaces, thereby avoiding the dependence on human-assisted feature identification and over-reliance on expert knowledge. Results show that the proposed modeling framework outperforms existing machine learning model based methods for TCM, as well as standard DL methods, in terms of both accuracy and robustness. In addition, the influence of the input in different feature spaces on the performance of deep models, as well as other established machine learning (ML) methods, is discussed to further exemplify the effectiveness of the proposed modeling framework.

The rest of this paper is organized as follows. After introducing the required background on ultraprecision manufacturing as well as the application of ML on machining, a background theory on SAE is presented in Section II. In Section III, the proposed modeling framework is detailed, as well as its application for TCM in ultraprecision manufacturing. The experimental validation setup, including the data acquisition, is detailed in Section IV. In Section V, the experimental results are discussed, which includes comparative analysis and discussion. Finally, the conclusion and future work are summarized in Section VI.

II. DL FRAMEWORK

A. Dimension Reduction Principle Based on SAE

Schölkopf *et al.* [26] proposed an unsupervised feature learning theory based on sparse representations. Given an input, it uses an encoder and a decoder preceded by a nonlinearity that transforms a code vector into a sparse output vector, with both vectors being as similar as possible to the input.

The SAE theory is based on the following two main components.

- 1) The *encoder*: Given a signal data $X \in \mathbf{R}^n$, it uses a random matrix W and an all-ones bias vector b to obtain a sparse-compressed representation $Y \in \mathbf{R}^m$ ($m \ll n$) via a nonlinear sigmoid function.
- 2) The *decoder*: The sparse representation Y is transformed back to a reconstruction vector $Z \in \mathbf{R}^n$ via the sigmoid function.

The target is to find the optimal parameters W and b to minimize the distance between the reconstruction vector Z and the input vector X . Full details of the SAE model can be found in [27].

B. Stacked SAE

An SSAE is constructed with the input layer and the hidden layer of several SAEs, which can extract deeper and more implicit features than a single autoencoder. The training structure is described as follows:

$$Y^{(k)} = f_s(W^{(k)}X^{(k)} + b^{(k)}) \quad (1)$$

$$X^{(k+1)} = Y^{(k)} \quad (2)$$

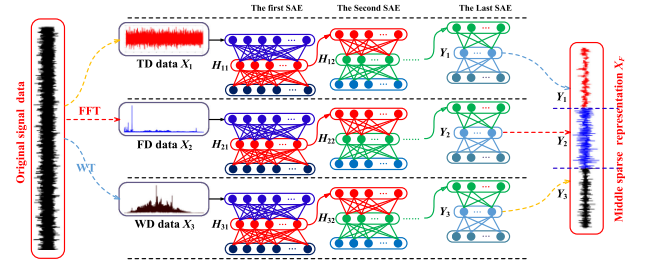


Fig. 1. Feature learning process of the parallel training model.

where $W^{(k)}$, $b^{(k)}$, $X^{(k)}$, and $Y^{(k)}$ are the weight matrix, the bias vector, the input, and the sparse representation of the k th SAE in the encoding procedure, respectively.

The loss function of the standard SSAE model is defined as follows [28]:

$$J_{\text{sparse}}(W, b) = J(W, b) + \beta \sum_{j=1}^{s_2} KL(\rho || \hat{\rho}_j) \quad (3)$$

$$J(W, b) = \left[\frac{1}{m} \sum_{i=1}^m \left(\frac{1}{2} \|h_{W,b}(x^{(i)}) - y^{(i)}\|^2 \right) \right] + \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(l)})^2 \quad (4)$$

$$KL(\rho || \hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j} \quad (5)$$

where x is the original input and y the corresponding label, m is the number of samples, and n is the number of the layers; λ , β , and ρ are the regularization constant, the divergence constant, and the sparseness constant, respectively; $\hat{\rho}$ is the mean activation value.

III. PROPOSED FMSSAEs-BASED METHOD FOR TCM

A. Novel FMSSAEs Structure

1) Parallel Training Based on Multiple SSAEs: Using a single feature space is popular in the application of fault diagnosis, via the use of DL methods in particular [18], [23]. However, different feature spaces in the vibration signal may have valuable and implicit information; existing DL methods do not simultaneously extract information from multiple spaces to characterize tool wear in machining. Moreover, in ultraprecision machining, signals and features are faint and inconspicuous, which could lead to poor overall performance of TCM. Thus, this paper proposes a parallel training model formed by using TD data, FD data, and WD data of the original vibration signal as the input of three different SSAEs to improve the capability of capturing different features in different spaces and implicit information.

Fig. 1 illustrates the feature learning process of the proposed parallel training model based on multiple SSAEs. The structures of the three SSAEs consist of the same number of network layers as well as hidden and output nodes. The input of the visible layer consists of the raw TD data $X_1 \in \mathbf{R}^{N_1}$, the FD data $X_2 \in \mathbf{R}^{N_2}$ following the application of FFT, and the WD data

$X_3 \in \mathbf{R}^{N_3}$ following a wavelet transform. Each hidden layer is divided into three groups: $H_{1k}, H_{2k}, H_{3k} \in \mathbf{R}^{M_k}$. Through calculation and sparse representation of several SSAEs, via the proposed framework, three different goal representations will be obtained: Y_1, Y_2 , and Y_3 . The learning process of any one of three SSAEs is independent of the others.

As it can be seen from the structure shown in Fig. 1, the parallel training model makes no use of any labels, which is different from most standard DL models that use labeled data to perform supervised learning. Thus, the loss function of the model in this section is modified as follows:

$$J_{j\text{sparse}}^i(W_j^i, b_j^i) = J_j^i(W_j^i, b_j^i) + \beta^i \sum_{k=1}^{t_j^i} KL(\rho^i \parallel \hat{\rho}_{k,j}^i) \quad (6)$$

$$J_j^i(W_j^i, b_j^i) = \frac{1}{2m} \sum_{h=1}^m \|f_s(W_j^i f_s(W_j^i X_j^{i,h} + b_j^i) + b_j^i) - X_j^{i,h}\|^2 + \frac{\lambda^i}{2} \sum_{r=1}^j \sum_{s=1}^{s_j^i} \sum_{t=1}^{t_j^i} (W_{ts}^{i,r})^2 \quad (7)$$

$$W_j^i, b_j^i = \operatorname{argmin} \{J_{j\text{sparse}}^i(W_j^i, b_j^i)\} \quad (8)$$

where $i = 1, 2, 3$ and j are the i th SSAE and the j th SAE, respectively; W, b and W', b' are the weight matrix and the bias vector of the encoding and decoding processes, respectively; X is the input vector; s and t are the number of nodes.

The optimal solution of the weight matrices W_j^i and the bias vectors b_j^i can be calculated by using (6)–(8), and the target representations Y_1, Y_2 , and Y_3 are obtained after parallel training.

2) Study of Feature Fusion Based on the Parallel Training Model: After transforming the representations Y_1, Y_2 , and Y_3 into a new feature vector $X_F = [Y_1 Y_2 Y_3]$, the vector will be used as the input for the next phase of our training framework, which consists of another SSAE. The difference between the fusion model and the parallel training model is that the fusion model makes use of labeled data.

Fig. 2 illustrates the learning and backpropagation processes of the fusion model. First, the feature $X_F \in \mathbf{R}^{X_F}$ is used as the input to the SSAE, and the first hidden representation $H_1 \in \mathbf{R}^{H_1}$ can be calculated by the weight matrix $W_1 \in \mathbf{R}^{H_1 \times X_F}$ and the sigmoid function f_s . Subsequently, the second hidden representation $H_2 \in \mathbf{R}^{H_2}$ can be calculated by the weight matrix $W_2 \in \mathbf{R}^{H_2 \times H_1}$ and f_s . The error E_1 between H_2 and the labels $y \in \mathbf{R}^{H_2}$ will also be calculated in this phase. The weight matrix W_2 will be fine-tuned into w_2 based on E_1 , and then W_1 will be updated to w_1 based on w_2 and E_1 . The hidden representation H_1 is transformed into H_1' based on X_F and w_1 .

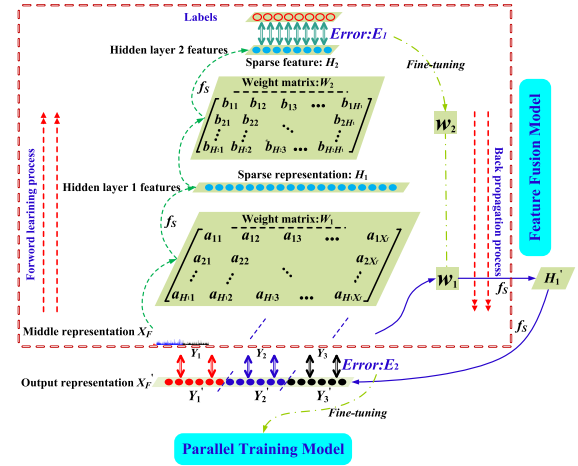


Fig. 2. Learning and backpropagation processes of the feature fusion model.

Then, the output vectors X_F' can be calculated by H_1' and w_1 . Finally, the error E_2 between X_F' and X_F is used as input to the parallel training model for the purpose of error backpropagation, as detailed in the next section.

3) Training Rules and Error Propagation: To have better convergence and reconstruction ability as well as deeper features, the proposed FMSSAEs framework makes use of an enhanced learning regime that involves twice the training and twice the fine tuning on each learning iteration. Thus, the loss function needs to be modified accordingly.

The learning process of the model is demonstrated as follows. First, the raw TD data X_1 , FD data X_2 , and WD data X_3 are used as the input vectors to the parallel training model to capture the target representations Y_1, Y_2 , and Y_3 , respectively. As a second step, the target representations are transformed into a new feature vector $X_F = [Y_1 Y_2 Y_3]$, which is subsequently used to obtain the hidden representation H_2 via another SSAE. In the third step, the first (of the two) fine-tuning processes will be carried out in the fusion model based on the error E_1 between H_2 and the labels y and the loss function $J_{\text{sparse2}}(W_p, B_p)$. Finally, the second fine-tuning will be carried out in the parallel training model based on the error E_2 between X_F' and X_F and the loss function $J_{j\text{sparse1}}^i(W_j^i, b_j^i)$.

The above mentioned training function is summarized as follows:

$$J_{\text{sparse2}}(W_p, B_p) = J_2(W_p, B_p) + RL + \beta \sum_{q=1}^{N_p+1} KL(\rho \parallel \hat{\rho}_{p,q}) \quad (9)$$

Eq. (10) shown at the bottom of this page,

$$RL = \frac{\lambda}{2} \sum_{k=1}^{n_k-1} \sum_{m=1}^{s_k} \sum_{n=1}^{s_{k+1}} (W_{nm,p}^{(k)})^2 \quad (11)$$

$$J_2(W_p, B_p) = \begin{cases} \frac{1}{2m} \sum_{q=1}^m \|f_{(W_p, B_p)s}(x_p^{(q)}) - \text{Labels}^{(q)}\|, & p = n-1 \\ \frac{1}{2m} \sum_{q=1}^m \|f_{(W_{p+1}, B_{p+1})s}(x_p^{(q)}) - f_{(w_{p+1}, b_{p+1})s}(x_p^{(q)})\|^2, & p = n-2, \dots, 1 \end{cases} \quad (10)$$

$$w_p, b_p = \operatorname{argmin}\{J_{\text{sparse2}}(W_p, B_p)\} \quad (12)$$

$$J_{j\text{sparse1}}^i(W_j^i, B_j^i) = J_j^i(W_j^i, B_j^i) + RL_j + \beta \sum_{k=1}^{N_j^i} KL(\rho || \tilde{\rho}_j^{i,k}) \quad (13)$$

Eq. (14) shown at the bottom of this page,

$$RL_j = \frac{\lambda}{2} \sum_{r=1}^{n_j-1} \sum_{s=1}^{N_j^i} \sum_{t=1}^{N_j^{i+1}} (W_{j,ts}^r)^2 \quad (15)$$

$$w_j^i, b_j^i = \operatorname{argmin}\{J_{j\text{sparse1}}^i(W_j^i, B_j^i)\} \quad (16)$$

where m is the number of samples, n and n_j are the layer number of the fusion model and the parallel training model, respectively, and N is the number of nodes.

B. Algorithmic Procedure for Proposed Methodology

To apply the proposed FMSSAEs modeling framework to the case study of ultraprecision machining, the following data processing and algorithmic process is followed.

- 1) First, the vibration signal of the tool in ultraprecision machining is collected from the manufacturing facility. Next, the TD data X_1 , FD data X_2 , and WD data X_3 are captured by preprocessing the signal via FFT and WT. Subsequently, all the data are divided into the training and testing samples based on the labels, which have been recorded by measuring the surface quality of the finished workpieces. This is for the purpose of model validation.
- 2) The target representations Y_1 , Y_2 , and Y_3 of the training samples are extracted via the unsupervised parallel training model and converted to a feature vector X_F .
- 3) Construct the fusion model based on the inputs/features of the parallel training model. The fusion model is used for feature learning of the feature vector. The learned deep features H_2 are fed into a back propagation (BP) algorithm for fine-tuning the parameters W and b of the fusion model based on the labels y in a supervised learning fashion.
- 4) The parameters W^i and b^i of the parallel training model are updated based on Step 3) to ensure that the parameters of the fusion model have been updated.
- 5) In the FMSSAEs modeling framework, Steps 3) and 4) will be repeated iteratively to train the overall model until the framework reaches the intended iterative steps. Finally, the performance of the proposed model is validated using the testing samples. Additionally, the generalization of the proposed method is verified.

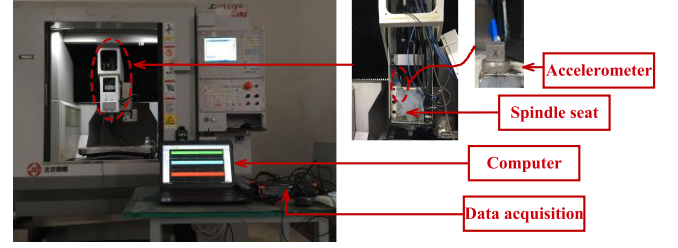


Fig. 3. Cutting tools experimental setup for ultraprecision machining.

IV. EXPERIMENT SETUP AND DATASET ACQUISITION

A. Experiment Setup

In this study, vibration data are collected continuously from a JDLVM550T_A13S CNC machine tool that is used to machine the external shell of a portable electronic device in a production line, i.e., a real manufacturing environment. Fig. 3 depicts the cutting tools experimental setup for ultraprecision machining and the data acquisition system. The machine tool is used to process and shape the external shell. The accelerometer with a sensitivity of 100.9 mV/g is mounted on the spindle seat of the machine tool for measuring the x -direction vibration signals. The vibration signals were collected at a sampling frequency of 10 kHz, and the overall sampling time is 20 s. The feed rate is set at 800 mm/min and the spindle speed at 5500 r/min. Based on the parameters and the spindle speed (v_s), the characteristic frequency ($f = v_s/60$) of the cutting tool can be calculated.

B. Data Acquisition

The data acquisition process is described by the following points.

- 1) All the data acquisition is performed in a manufacturing environment, not in the laboratory.
- 2) For consistency, all the workpieces are machined with the same cutting parameters. In order to indirectly evaluate the tool condition (for the purpose of sample labeling and supervised learning), the surface quality of each finished workpiece is measured under an Olympus microscope STM6 at a magnification of 1000 \times .
- 3) Three cutting tools were assessed for this study (tools A, B, and C) for the purpose of experiment validation and generalization verification. The parameters for the cutting tools are listed in Table I. There are slight differences in the tool setup parameters, which represent how the tools are used in a real manufacturing environment. These differences are to account for various uncertainties in the manufacturing process and yield a different number of parts processed by each tool. While the manufacturing cell continuously records data,

$$J_j^i(W_j^i, B_j^i) = \begin{cases} \frac{1}{2m} \sum_{h=1}^m \left\| f_{(W_j^i, B_j^i)s} \left(x_j^{i,(h)} \right) - Y_j^{i,(h)} \right\|^2, & i = n_j - 1 \\ \frac{1}{2m} \sum_{h=1}^m \left\| f_{(W_j^{i+1}, B_j^{i+1})s} \left(x_j^{i+1,(h)} \right) - f_{(w_j^{i+1}, b_j^{i+1})s} \left(x_j^{i+1,(h)} \right) \right\|^2, & i = n_j - 2, \dots, 1 \end{cases} \quad (14)$$

TABLE I
PARAMETERS OF THE CUTTING TOOLS

Work material	Tool material	Tool number	Relief angle	Flank width	Maximum workpiece number	Cutting allowance
Aluminium alloy	Diamond	A	3.1°	69.8mm	596	0.01mm
		B	3.4°	71.8mm	593	
		C	4.7°	91.8mm	436	

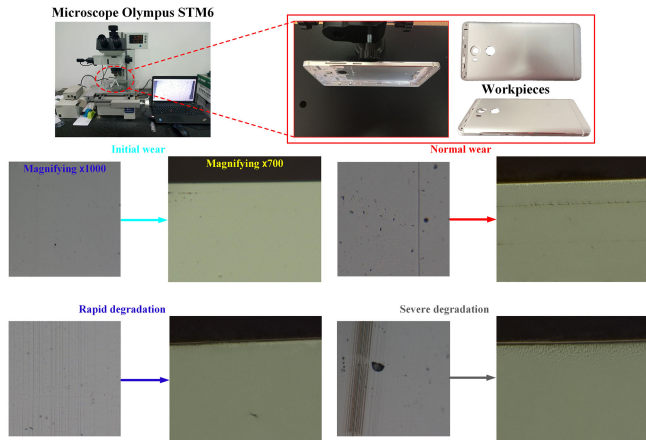


Fig. 4. Four surface quality states and the tool wear conditions under the microscope at a magnification of 1000 \times and 700 \times .

TABLE II
DESCRIPTION OF THE CUTTING TOOLS' WEAR CONDITIONS

Wear conditions of the cutting tools	Size of samples		Tool C	Total	Label of conditions
	Tool A	Tool B			
Initial wear	64	54	43	161	1
Normal wear	307	306	204	817	2
Rapid degradation	138	132	105	375	3
Severe degradation	87	101	84	272	4

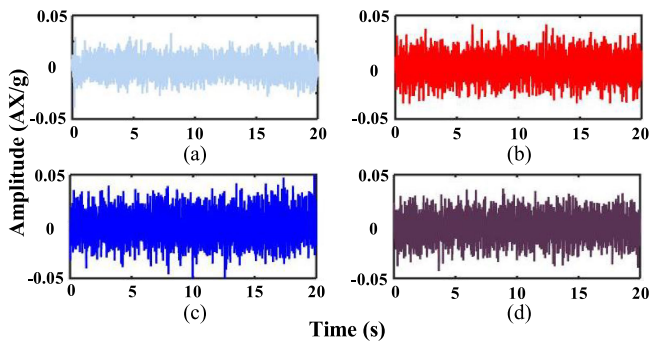


Fig. 5. Amplitudes of the raw vibration signal for the four wear conditions. (a) Initial wear. (b) Normal wear. (c) Rapid degradation. (d) Severe degradation.

some samples are randomly selected for training and the remaining for testing the proposed modeling framework.

In the experimental process, four different surface quality states have been defined for the finished workpieces, corresponding to four kinds of gradual tool wear conditions, namely, initial wear, normal wear, rapid degradation, and severe degradation. This categorization also matches expert knowledge on

TABLE III
DESCRIPTION OF THE PARAMETERS OF THE TRADITIONAL METHODS

Description	Parameters	
The traditional methods	BPNN	SVM
Transfer function	logsig	-
Kernel function	-	linear
Penalty parameters	-	0.25
Learning rate	0.8	-
Momentum constant	0.9	-
Epochs	700	700

TABLE IV
IDENTIFICATION RESULTS OF DIFFERENT METHODS

Methods	Average testing accuracy	Standard deviation
The proposed method	96.63% (3503/3625)	0.007242
BPNN with raw TD data	58.51% (2118/3625)	0.011796
BPNN with TD features	79.34% (2878/3625)	0.019776
BPNN with FD features	83.77% (3038/3625)	0.008259
BPNN with the fusion features	84.70% (3069/3625)	0.009812
SVM with raw TD data	57.76% (2088/3625)	0.014111
SVM with TD features	78.39% (2835/3625)	0.014092
SVM with FD features	84.45% (3061/3625)	0.009059
SVM with the fusion features	91.97% (3333/3625)	0.008334

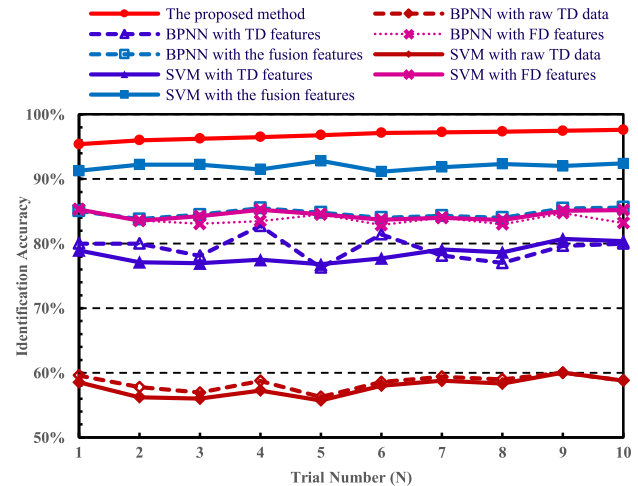


Fig. 6. Identification results of the ten trials using different methods.

the particular mechanisms of tool wear for the process under investigation. Fig. 4 depicts the four surface quality states and the tool wear conditions under the microscope at a magnification of 1000 \times and 700 \times , respectively.

Table II presents the size of sample sets and the label of conditions. Each condition contains 161, 817, 375, and 272 samples, respectively. Each sample is a raw vibration signal containing 20 000 data points, following some filtering to remove noise. Meanwhile, FFT and WT are applied to each sample signal in order to obtain FD and WD data. Therefore, the raw data, FD data, and WD data contain 20 000, 16 285, and 20 000 data points, respectively. Because of the strict cutting tolerances and less overall vibration in ultraprecision machining (compared to normal machining), the amplitude of the raw vibration signal of the four health conditions is quite similar, on first inspection, as shown in Fig. 5.

TABLE V
NETWORK PARAMETERS OF DIFFERENT DL METHODS

Deep learning methods	Structure	Input	Layers	Nodes number	Learning rate	Iteration
The proposed method	Parallel learning model	TD data	4	20000-1870-187-41	0.8	700
		FD data	4	16385-1870-187-41	0.8	700
		WD data	4	20000-1870-187-41	0.8	700
	Fusion model	The fusion features	3	123-20-4	0.8	700
Standard SAE	-	TD data	4	20000-1870-187-41	0.8	700
	-	FD data	4	16385-1870-187-41	0.8	700
	-	WD data	4	20000-1870-187-41	0.8	700

TABLE VI
INFLUENCE OF DIFFERENT FEATURE SPACES ON DL METHODS

Deep learning methods	Input	Average testing accuracy	Standard deviation
The proposed method	Data fusion	96.63% (3503/3625)	0.007242
	TD data	72.86% (2641/3625)	0.011826
Standard SAE	FD data	85.32% (3093/3625)	0.014646
	WD data	90.04% (3264/3625)	0.010881

V. EXPERIMENT RESULTS

A. Comparison With Established ML Methods

BPNN and SVM are widely used for TCM and fault diagnosis, as discussed in Section I. However, these machine learning, or intelligent systems, methods often rely on engineering experience and expert knowledge to artificially design and extract features, which are not trivial tasks in the whole modeling workflow. In this section, a comparison is made between a BPNN having a shallow-layer structure and an SVM structure against the proposed method.

Each of the different feature spaces in the original signal may have different properties; hence, these will pose a different challenge to any data-driven modeling framework. For comparison purposes, the performance of BPNN and SVM will be discussed based on the TD feature space and also the FD feature space, as well as the fusion/combination of the two. A number of statistical features from TD (mean, standard deviation, skewness factor, etc.) and FD (variance, kurtosis factor, etc.) are extracted; more details can be found in [3] and [30]. Table III presents the model parameters for the BPNN and SVM model structures. The average identification accuracy and the standard deviation comparison of the different methods and features using ten experiments are presented in Table IV, and the detailed modeling results are shown in Fig. 6.

From Table IV, it can be noted that, among the different methods, the proposed method has the best performance (average accuracy: 96.63%, standard deviation: 0.007242), which is superior to the traditional ML methods (best result: 91.97%, 0.008259). In terms of the different methods based on different spaces, the SVM with the fusion features exhibits a better performance (average accuracy: 91.91%) compared with BPNN (average accuracy: 84.70%). This is still inferior to the proposed method of parallel processing the feature spaces.

From Fig. 6, it can be seen that, from the first trial to the last trial, with the increase of the training samples, the accuracy of the proposed method keeps growing, which is typical of continuous learning in DL networks.

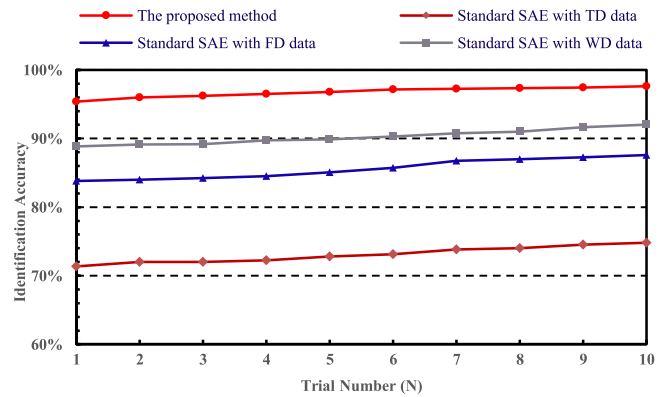


Fig. 7. Identification results of the ten trials using different feature spaces on DL methods.

Another interesting observation is that following the same use of features, BPNN and SVM have a similar performance, such as when using raw data and TD features, especially in terms of the average accuracy.

These results indicate that: 1) compared with established ML methods (BPNN and SVM), the proposed method has several advantages, which include the ability to separately extract features autonomously despite the complexity and dimensionality of the feature space set, due to its deep-layer structure and nonlinear mapping ability; 2) the traditional methods depend on human-assisted feature extraction. From Table IV, it can be seen that if the selected feature set is uncorrelated to the labels, the model will have poor accuracy in classifying tool wear; and 3) from the first trial to the last trial, the accuracy of BPNN and SVM keeps fluctuating, as expected, due to the batch training process. The DL framework takes advantage of inherent sequential training to improve its overall prediction accuracy.

B. Results in Different Features Spaces

In recent years, due to advances and access to significant computational power, DL methods have been applied to gear and bearing fault diagnosis. However, the influence of different feature spaces on DL methods is still unclear in fault detection. In order to further scrutinize the performance of the proposed FMSSAs model, the effectiveness of the model and the different feature spaces on comparable DL methods will be studied and surveyed via several experimental trials. Table V presents the structure of the studied DL methods for features extraction. In this section, results will be discussed for the following three different feature sets: the TD data, FD data, and WD data.

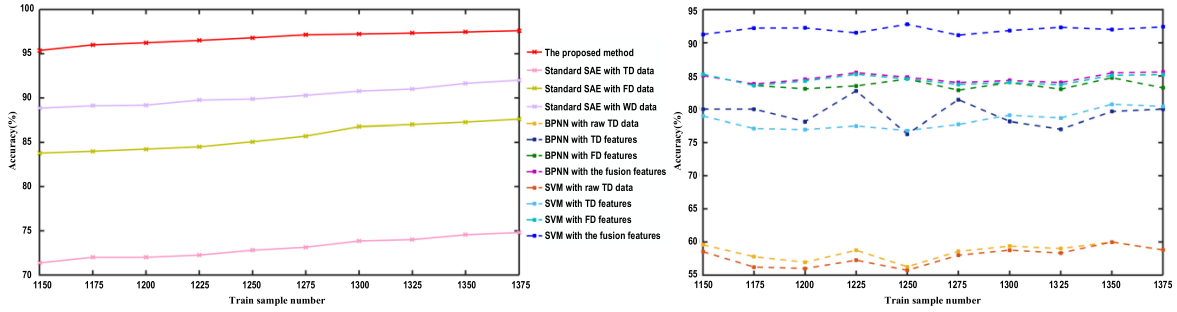


Fig. 8. Reconstruction error curves of DL algorithms based on different feature spaces.

The average identification accuracy and the standard deviation of the different feature spaces on DL methods are given in Table VI, and the detailed results with ten experiments in each trial are shown in Fig. 7. It can be concluded that the identification accuracy of separate feature spaces on DL methods vary considerably. Among them, the proposed method with data fusion has the best performance (average accuracy: 96.63%, standard deviation: 0.007242), which is superior to standard DL methods with a single feature space (best result: 90.04%, 0.010881).

The reconstruction error curves of the SAE model based on the different inputs and the proposed method are shown in Fig. 8. Although the proposed DL model needs more training time (above 4 min.) due to the increase of units and network layers, the reconstruction error is smaller following convergence. Meanwhile, the variability of the weight update can be reflected on the reconstruction curve, which is an indicator of efficient training.

Fig. 9 shows the details of the identification accuracy of all the methods based on different training samples. The results indicate that the more the training sample sizes, the better the performance of all DL methods (which is expected), while the traditional methods have no such characteristic.

The abovementioned results imply the following.

- 1) *Feature selection and extraction*: The traditional ML methods depend on artificial feature extraction via expert knowledge, while DL methods have the advantage of autonomous selection of implicit representation through nonlinear mapping. There may be equally good combinations of selecting the feature set (compared with the autonomous method) for the traditional ML methods, but this would require tedious and laborious processes.
- 2) *Generalization ability*: The performance of DL methods can be improved with an increase of the training samples, while the traditional methods do not have this property. This shows that the DL methods can have better results if there are sufficient samples, which is somewhat expected due to the complexity of the DL structure.
- 3) *Model performance*: The proposed method is superior to the established popular ML methods, often used in tool wear prediction, as well as when compared with DL methods due to the modified loss function and the enhanced training algorithm that takes advantage of separate feature spaces to extract features but combines them to link them to the labeled outcomes (tool wear).

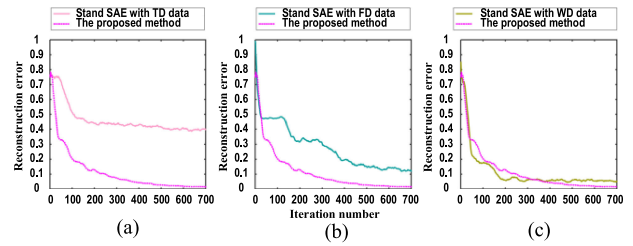


Fig. 9. Details of results of all the methods based on different training samples.

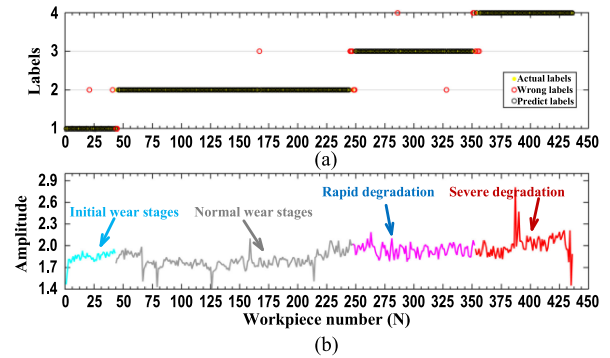


Fig. 10. Identification results of the proposed method on tool C dataset.

C. Generalization Verification

In this section, the datasets of tools A, B, and C are used to further scrutinize (beyond using testing data from the same tool) the generalization ability of the proposed modeling framework. The acquired data samples are separated into two parts: the training sets (tools A and B) used for model training and the testing sets (tool C) for assessing generalization.

Fig. 10(b) shows the rms of the vibration signal of tool C (from the 1st workpiece to the 436th workpiece) and the four kinds of tool operation conditions, i.e., initial wear stages, normal wear stages, rapid degradation, and severe degradation. Table VII presents the training sets, testing sets, and the input features of the three methods.

Fig. 10(a) shows the prediction performance of the testing sets. It can be observed that the proposed method shows very good generalization properties for TCM in ultraprecision machining. In addition, Table VII reveals that the proposed methodology outperforms the standard SAE DL as well

TABLE VII
IDENTIFICATION RESULTS OF DIFFERENT METHODS

Methods	Training sets		Testing sets	Input	Testing accuracy
	Tool A	Tool B	Tool C		
The proposed method				Multiple features spaces-based	93.18% (406/436)
Standard SAE	596	593	436	WD data	86.82% (378/436)
SVM				14 TD features and 11 FD features	88.24% (385/436)

as the traditional SVM method in terms of identification accuracy. These positive results demonstrate that the proposed methodology can potentially be used in a real manufacturing environment to accurately identify tool wear states for TCM in ultraprecision manufacturing.

VI. CONCLUSION AND FUTURE WORK

In this paper, a novel multiple feature spaces based DL model (FMSSAEs) was proposed for tool condition identification in ultraprecision manufacturing. First, a new parallel training model structure was designed to learn the low-layer features in terms of feature extraction and selection. Then, a fusion model was employed to learn the deep features and to fine-tune the parallel training model to further adjust the network's parameters. To achieve this structure, the proposed method makes use of a modified loss function and an improved overall training framework.

The proposed method was applied to real manufacturing data, consisting of cutting tool vibration signals captured during an ultraprecision machining process. Results show that the proposed method is more effective compared with established and popular ML methods for tool wear prediction, as well as standard DL methods used for TCM. Generalization properties appear to be good for the new proposed methodology; however, there are certain limitations when a different tool is used because of the variability in setup tool parameters of the different cutting tools, such as the relief angle, flank width, etc. Further research work is needed in this area to perhaps focus the model training regime on creating uncertainty-tolerant features rather than aiming for an ultimate overall prediction accuracy.

With the rapid development of hardware technology and computational power, DL structures can find applications in more industrial fields; however, modification and enhancement of the DL framework to suit a particular application, including efficient error propagation and heuristic training regime, are not trivial tasks.

REFERENCES

- [1] L. B. Kong, C. F. Cheung, W. B. Lee, and S. To, "An integrated manufacturing system for the design, fabrication, and measurement of ultra-precision freeform optics," *IEEE Trans. Electron. Packag. Manuf.*, vol. 33, no. 4, pp. 244–254, Oct. 2010.
- [2] A. J. Torabi, M. J. Er, X. Li, B. S. Lim, and G. O. Peen, "Application of clustering methods for online tool condition monitoring and fault diagnosis in high-speed milling processes," *IEEE Syst. J.*, vol. 10, no. 2, pp. 721–732, Jun. 2016.
- [3] N. Li, Y. J. Chen, D. D. Kong, and S. L. Tan, "Force-based tool condition monitoring for turning process using v-support vector regression," *Int. J. Adv. Manuf. Technol.*, vol. 91, pp. 351–361, Jul. 2017.
- [4] M. S. H. Bhuiyan and I. A. Choudhury, "Review of sensor applications in tool condition monitoring in machining," *Comprehensive Mater. Process.*, vol. 13, pp. 539–569, Apr. 2014.
- [5] C. H. R. Martins, P. R. Aguiar, A. Frech, and E. C. Bianchi, "Tool condition monitoring of single-point dresser using acoustic emission and neural networks models," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 3, pp. 667–679, Mar. 2014.
- [6] O. Geramifard, J. X. Xu, J. H. Zhou, and X. Li, "A physically segmented hidden Markov model approach for continuous tool condition monitoring: Diagnostics and prognostics," *IEEE Trans. Ind. Inform.*, vol. 8, no. 4, pp. 964–973, Jun. 2012.
- [7] K. Patra, "Acoustic emission based tool condition monitoring system in drilling," *Lecture Notes Eng. Comput. Sci.*, vol. 3, no. 1, pp. 2126–2130, 2011.
- [8] X. C. Ku, Y. M. Zhou, P. L. Gao, and M. D. Duan, "Recognition of tool wear state based on wavelet packet and BP neural network," *Modern Manuf. Eng.*, vol. 12, pp. 68–72, 2014.
- [9] F. A. Niaki, D. Ulutan, and L. Mears, "Wavelet based sensor fusion for tool condition monitoring of hard to machine materials," in *Proc. IEEE Int. Conf. Multisensor Fusion Integr. Intell. Syst.*, Oct. 2015, pp. 271–276.
- [10] O. Massol, X. Li, R. Gouriveau, J. H. Zhou, and O. P. Gan, "An exTS based neuro-fuzzy algorithm for prognostics and tool condition monitoring," in *Proc. 11th Int. Conf. Control Automat. Robot. Vis.*, Dec. 2010, pp. 1329–1334.
- [11] Q. Ren, M. Balazinski, L. Baron, K. Jemielniak, R. Botez, and S. Achiche, "Type-2 fuzzy tool condition monitoring system based on acoustic emission in micromilling," *Inf. Sci.*, vol. 255, no. 1, pp. 121–134, Jan. 2014.
- [12] Y. Qiu and F. Y. Xie, "Tool wear monitoring based on wavelet packet coefficient and hidden Markov model," *Mach. Tool Hydraul.*, vol. 12, pp. 40–44, 2014.
- [13] D. F. Shi and N. N. Gindy, "Tool wear predictive model based on least squares support vector machines," *Mech. Syst. Signal Process.*, vol. 21, no. 4, pp. 1799–1814, May 2007.
- [14] J. J. Wang, J. Y. Xie, R. Zhao, K. Z. Mao, and L. B. Zhang, "A new probabilistic kernel factor analysis for multisensory data fusion: Application to tool condition monitoring," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 11, pp. 2527–2537, Nov. 2016.
- [15] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen, and J. Wang, "Machine health monitoring using local feature-based gated recurrent unit networks," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1539–1548, Feb. 2018.
- [16] E. Kannatey-Asibu, J. Yum, and T. H. Kim, "Monitoring tool wear using classifier fusion," *Mech. Syst. Signal Process.*, vol. 85, pp. 651–661, Feb. 2017.
- [17] C. Shang, F. Yang, D. X. Huang, and W. X. Lyu, "Data-driven soft sensor development based on deep learning technique," *J. Process Control*, vol. 24, no. 3, pp. 223–233, Mar. 2014.
- [18] Y. G. Lei, F. Jia, J. Lin, S. B. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Trans. Ind. Electron.*, vol. 63, no. 5, pp. 3137–3147, May 2016.
- [19] 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.
- [20] H. D. Shao, H. K. Jiang, H. Z. Zhang, and T. C. Liang, "Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network," *IEEE Trans. Ind. Electron.*, vol. 65, no. 3, pp. 2727–2736, Mar. 2017.
- [21] H. D. Shao, H. K. Jiang, F. Wang, and H. W. Zhao, "An enhancement deep feature fusion method for rotating machinery fault diagnosis," *Knowl.-Based Syst.*, vol. 119, pp. 200–220, Mar. 2017.
- [22] V. T. Tran, F. AlThobiani, and A. Ball, "An approach to fault diagnosis of reciprocating compressor valves using Teager–Kaiser energy operator and deep belief networks," *Expert Syst. Appl.*, vol. 41, no. 9, pp. 4113–4122, Jul. 2014.
- [23] O. Janssens, V. Slavkovikj, B. Vervisch, K. Stockman, M. Loccufer, and S. Verstockt, "Convolutional neural network based fault detection for rotating machinery," *J. Sound Vib.*, vol. 377, pp. 331–345, Sep. 2016.
- [24] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, Jun. 2009, pp. 609–616.
- [25] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.
- [26] B. Schölkopf, J. Platt, and T. Hofmann, "Efficient learning of sparse representations with an energy-based model," in *Advances in Neural Information Processing Systems 19*. Cambridge, MA, USA: MIT Press, 2007, pp. 1137–1144.

- [27] J. D. Sun, C. H. Yan, and J. T. 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.
- [28] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P. A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. 12, pp. 3371–3408, Mar. 2010.
- [29] Y. G. Lei, Z. J. He, and Y. Y. Zi, "EEMD method and WNN for fault diagnosis of locomotive roller bearings," *Expert Syst. Appl.*, vol. 38, no. 6, pp. 7334–7341, Jun. 2011.
- [30] K. Jemielniak, T. Urbański, J. Kossakowska, and S. Bombiński, "Tool condition monitoring based on numerous signal features," *Int. J. Adv. Manuf. Technol.*, vol. 59, no. 1–4, pp. 73–81, Mar. 2012.



Chengming Shi is currently working toward the Ph.D. degree in mechatronic engineering at the Huazhong University of Science and Technology, Wuhan, China.

His current research interests include intelligent manufacturing, sensor data mining, signal processing, and machine learning.



George Panoutsos received the Ph.D. degree in automatic control and systems engineering from the University of Sheffield, Sheffield, U.K., in 2007.

He was a Lecturer with the Department of Automatic Control and Systems Engineering, University of Sheffield, in 2010. He is currently the Deputy Head of Department of Automatic Control and Systems Engineering (University of Sheffield, UK). He has a research grant portfolio of over £2M from the UK EPSRC, Innovate UK,

EU Horizon 2020, and direct industry funding, as well as over 60 research publications in theoretical as well as applied contributions in the areas of computational intelligence, data-driven modeling, optimization, and decision support systems.



Bo Luo received the bachelor's degree in machinery manufacturing and automation and the Ph.D. degree in mechatronic engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2014, respectively.

He is currently a Postdoctoral Researcher with the School of Mechanical Science and Engineering, Huazhong University of Science and Technology. His current research interests include intelligent manufacturing, big data mining,

signal processing, and machine learning.



Hongqi Liu received the Ph.D. degree in mechatronic engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2008.

He is currently an Associate Research Professor with the School of Mechanical Science and Engineering, Huazhong University of Science and Technology. His current research interests include machinery health monitoring, diagnosis and prognosis, complex systems failure analysis, quality and reliability engineering, and manufacturing systems design, modeling, scheduling, and planning.



Bin Li received the B.S., M.S., and Ph.D. degrees in mechanical engineering from the Huazhong University of Science and Technology, Wuhan, China, in 1982, 1989, and 2006, respectively.

He is currently a Professor with the School of Mechanical Science and Engineering, Huazhong University of Science and Technology. His current research interests include intelligent manufacturing and CNC machine tools.



Xu Lin received the M.S. degree in mechanical engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2018.

His current research interests include intelligent manufacturing, sensor data mining, signal processing, and machine learning.