Tool Wear Prediction via Multidimensional Stacked Sparse Autoencoders With Feature Fusion

Chengming Shi, Bo Luo, Songping He, Kai Li, Hongqi Liu, and Bin Li

***Abstract*—Tool wear prediction is of critical importance to maintain the desired part quality and improve productivity. Inspired by the successful application of deep learning in many condition monitoring tasks. In this article, a novel modeling framework is presented, which includes multiple stacked sparse autoencoders and a nonlinear regression function for tool wear prediction. Multiple stacked sparse autoencoders consists of two main structures. One model is designed with multidimensional stacked sparse autoencoders, which can learn more features from different feature domains in the raw vibration signal, and another single-dimensional stacked sparse autoencoders is used for feature fusion and deeper features learning. And a modified loss function is applied that improves the learning ability. In addition, due to the good properties of tool wear process in nonstationarity and complex nonlinear, a nonlinear regression function is utilized to enhance the progressive tool wear prediction tasks. A dataset from a real manufacturing process is used to evaluate the performance of the proposed modeling framework. Experimental results show that tool wear can be predicted accurately and stably by the proposed tool wear predictive model, which outperforms the already developed methods.**

***Index Terms*—Deep learning, feature extraction, feature fusion, stacked sparse autoencoders, tool wear prediction.**

NOMENCLATURE

BPNN	Back propagation neural network.
DL	Deep learning.
FD	Frequency domain.
FFT	Fast Fourier transform.
GPR	Gaussian process regression.
HF-WPC	High-frequency wavelet packet coefficients.
KPCA	Kernel principal component analysis.
LF-WPC	Low-frequency wavelet packet coefficients.
LSTM	Long-short term memory.
MAE	Mean absolute error.
MAPE	Mean absolute percentage error.
MD-SSAEs	Multidimensional stacked sparse autoencoders.
ML	Machine learning.
PCC	Pearson correlation coefficient.
RMSE	Root mean squared error.
RNN	Recurrent neural network.
RUL	Remaining useful life.
SSAEs	Stacked sparse autoencoders.
SVR	Support vector regression.
TD	Time domain.
WPD	Wavelet packet decomposition.

Manuscript received July 18, 2019; revised September 18, 2019; accepted October 20, 2019. Date of publication October 24, 2019; date of current version April 13, 2020. This work was supported in part by the National Natural Science Foundation of China under Grants 51705174 and 51625502 and in part by Major special projects in Jiangsu Province of China under Grant SBE2017020146. Paper no. TII-19-3200. *(Corresponding authors: Songping He; Bin Li.)*

C. Shi and K. Li are with the Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: 513864035@qq.com; 15527965836@163.com).

B. Luo is with the Department of Automatic Control and Systems Engineering, The University of Sheffield, Sheffield S1 3JD, U.K. (e-mail: b.luo@sheffield.ac.uk).

S. He, 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: 8562576@qq.com; liuhongqi328@163.com; li_bin_hust@163.com).

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

Digital Object Identifier 10.1109/TII.2019.2949355

I. Introduction

IN THE MANUFACTURING industry, tool wear is one of the critical factors that not only directly affects the machining accuracy and the workpiece surface quality, but also reduce production efficiency. It is estimated that the usage of worn and broken tools may lead to nearly 7%–20% of the unscheduled downtime, which can result in a significant loss of production efficiency. More seriously, tool failures can lead to the chatter of machine tool due to the sharp increase in cutting force [1]. Besides, tool wear is a highly complex phenomenon that consists of different mechanisms, such as adhesion, abrasion, chipping, diffusion, and plastic deformation [2]. Therefore, automatic and accurate tool wear predictive models are critically needed to provide early warnings for tool wear and offer autonomous decisions on cutter compensation.

In the past few decades, there are many artificial intelligent methodologies developed for tool wear prediction, in which BPNN [3]–[6], SVR [8]–[12], LSTM [13], and GPR [7] are the most widely applied. Generally, the following two essentials are necessary for tool wear prediction: 1) man-assisted feature extraction and selection and 2) ML model development. Martins *et al.* [3] researched the relationship between acoustic emission signal and the dresser wear based on neural network (NN) and verified that the combination of the frequency bands of

28–33 kHz and 42–50 kHz can best characterize the dresser wear condition. Several statistics features in the TD and FD extracted from multisensor signals via wavelet analysis were studied by Niaki *et al.* [4] and an RNN model was utilized in tool wear estimation; the authors studied the application of sensor information fusion for the purpose of increasing the generalization performance of RNN, and the results showed that there have a maximum of 13% relative error in estimating tool wear. Drouillet *et al*. [5] researched the relationship between RUL and the machine spindle power values based on the NN technique; the root mean square power value in the TD was found to be sensitive to tool wear, and the results showed a good agreement between the predicted and true RUL of tools. Corne *et al.* [6] fed spindle power and force data into NN for functional processing and, thus, predicting tool flank wear; it is found that the error predicted by NN was around 0.4–18.4% with power and force data. Dutta *et al*. [10] proposed an approach for tool wear prediction based on linear SVR and indicated that the features extracted from turned surface images can be utilized to evaluate the cutting tool flank wear with a maximum 4.9% prediction error. Kong *et al*. [11] presented a new v-SVR-based tool wear predictive model with KPCA technique to fuse effective features, which are extracted from force signals; the authors demonstrated that the model is proved effective beyond expectation. However, it has good ability only in cases with small sample size. A technique based on multisensory data (e.g., force, vibration, etc.) fusion with dimension reduction was investigated by Wang *et al*. [12], and an SVR model was utilized to estimate the tool wear width; the results showed that the performance is comparable with the actual tool wear width measured offline by a microscope instrument in terms of accuracy. Force and vibration data in three directions (x, y, and z) were discussed by Zhao *et al.* [13], and LSTM was used for predicting the actual tool wear; it is shown that LSTM model outperforms several state-of-arts baseline methods. Kong *et al*. [7] developed an approach for tool wear assessment based on KPCA_IRBF and GPR and verified that KPCA helps to compress and smooth the confidence interval of GPR; the results showed that GPR performs better accuracy than NN and SVR in tool wear prediction. It is evident in the abovementioned literatures that the workflow of feature extraction followed by the ML model has been successfully applied in tool wear prediction. However, the performance of these models highly depends on the sensitivity of the extracted features, which is generally performed via expert knowledge [14]–[16]. Moreover, such methodologies are basically demonstrated in the laboratory, rather than a real manufacturing environment.

If the already developed tool wear predictive models are to be used effectively in real manufacturing process, there will still have several significant challenges that need to be addressed. 1) Such methodologies are often used in the laboratory environment, which implies controlled conditions and low susceptibility to noise; thus, the features extracted from the signal data will have high correlation with tool wear. However, it is found that the correlation of the extracted features and tool wear is not so obvious due to too much noise in the collected signal data in a real manufacturing process. 2) BPNN, SVR, LSTM, and GPR belong to shallow-layer learning models, which are difficult to learn the complex nonlinear relationships. Thus, feature selection has become a key technology that determines the performance of most traditional ML models. However, it is a challenging and time-consuming task to select the most sensitive features. Autonomous feature extraction and selection would be preferred in this case.

DL is a new ML model that has the powerful capacity to overcome the inherent disadvantages of traditional ML models [17]–[19]. The most conspicuous difference between DL and shallow-layer models is that the former can automatically learn the valuable features from the raw data.

In the last five years, using single-feature-domain-based standard DL models is popular in the application of condition monitoring and fault diagnosis [20]–[24]. However, it is evident that different feature domains in the signal data have different properties, as well as valuable and significant information [30]; current DL models cannot meet the requirement for extracting features from multiple feature domains simultaneously. In addition, it is known that DL models are often used to address the classification problems. However, in this study, there is more in need of a DL-based regression model for predicting tool wear due to the good properties of tool wear process in nonstationarity and complex nonlinear. Thus, it is critical to research and develop a new DL-based modeling framework to address multiple feature domains, autonomous feature extraction capability, and tool wear prediction suitable for real manufacturing environments.

In this article, a novel DL-based modeling framework is proposed, which includes multiple SSAEs and a nonlinear regression function for tool wear prediction. In this study, all the vibration data are collected from a real manufacturing plant. The datasets include the raw vibration signal data and are preprocessed by FFT and WPD to create a preliminary signal set. First, a new "MD-SSAE model" is designed for four kinds of feature domains (TD data, FD data, LF-WPC, and HF-WPC) to learn more valuable features. To achieve this DL structure, a modified loss function is used to improve the learning ability. Then, a single-dimensional SSAE is used for feature fusion and deeper features learning. Finally, a nonlinear regression function is utilized to enhance the progressive tool wear prediction tasks due to the good properties of tool wear process in nonstationarity and complex nonlinear. The main contributions of this article can be summarized as follows.

1) In the phase of feature learning, via the proposed tool wear predictive model we take full advantage of the feature learning ability of DL model as well as the characteristics of different feature domains. And the modified loss function can extremely reduce the reconstruction error (for the purpose of extracting more valuable features).
2) By combining the multiple-SSAE structure with the nonlinear regression function, the performance of the proposed model will be improved because of its great complex nonlinear learning ability (for the purpose of extracting deeper, better, and more sensitive features, as well as real-time tool wear prediction).

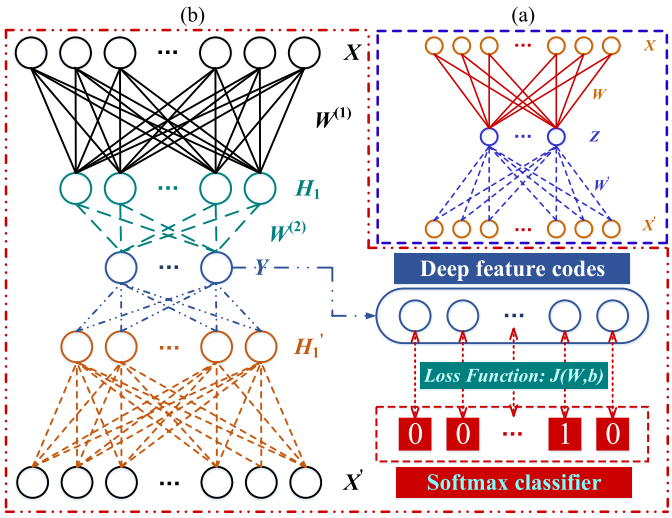

Fig. 1. Stacked autoencoders for dimension reduction and classification.

The rest of this article is organized as follows. In Section II, background theory on SSAEs is reviewed. Then, the framework and details of the proposed method for tool wear prediction are briefly introduced in Section III. In Section IV, details of experiment setup and datasets acquisition are illustrated. In Section V, analysis of the proposed method and the experimental results are presented, which includes comparative analysis and discussion. Finally, Section VI concludes this article.

II. Principle of the SSAE Model

An autoencoder is a typical type of three-layer neural network that reduce the dimensions of the input data in an unsupervised manner, which can transform the high-dimension input into a compressed code and then transform this code into a reconstruction vector [25].

Theoretically, an autoencoder consists of two parts, the encoder ϕ and the decoder φ, which can be defined as follows:

$$\phi: \quad Z = \sigma(WX + b) \tag{1}$$

$$\varphi: \quad X' = \sigma'(W'Z + b') \tag{2}$$

where W, b, σ and W', b', σ' are weight matrix, bias vector, and activation function of the encoding and decoding procedure, respectively. $X \in \mathbf{R}^N$ is the input. $Z \in \mathbf{R}^M$ is a compressed code. $X' \in \mathbf{R}^N$ is the reconstruction vector.

Fig. 1(a) represents the principal of an autoencoder, and its goal is to train W, W' to minimize reconstruction errors

$$L(X, X') = \|X - X'\|^2 = \|X - \sigma'\left[W'\sigma\left(WX + b\right) + b'\right]\|^2. \tag{3}$$

Hinton *et al.* [17] described that high-dimensional data can be converted to low-dimensional code by training a multilayer NN with a small central layer and, thus, proposed DNNs in which a layer-by-layer initialization is performed. Fig. 1(b) represents the procedure of stacked autoencoders, and its loss function [26] is defined as follows:

$$J(W, b) = \frac{1}{m}\sum_{i=1}^{m}\left[\frac{1}{2}\left\|\sigma\left(WX^{(i)} + b\right) - Y^{(i)}\right\|^2\right] + \frac{\lambda}{2}\sum_{l=1}^{n-1}\sum_{i=1}^{s_l^c}\sum_{j=1}^{s_l^r}\left[W_{ji}^{(l)}\right]^2 \tag{4}$$

where $W, b,$ and σ are weight matrix, bias vector and activation function, respectively. m is the number of samples, X is the original input, and Y is the corresponding label. λ is the regularization constant. n is the number of the layers. s_l^c and s_l^r are the columns and rows of the matrix $W^{(l)}$.

Poultney *et al.* [27] described a sparse feature learning method to achieve sparsity by additional terms in the loss function and then obtained the active hidden units. Thus, the loss function of the standard SSAE model is defined as follows:

$$L_{\text{sparse}}(W, b) = J(W, b) + \beta\sum_{j=1}^{s_2} KL\left(\rho\|\widehat{\rho}_j\right) \tag{5}$$

$$KL\left(\rho\|\widehat{\rho}_j\right) = \rho\log\frac{\rho}{\widehat{\rho}_j} + (1-\rho)\log\frac{1-\rho}{1-\widehat{\rho}_j} \tag{6}$$

where β and ρ are the divergence constant and sparseness constant, respectively. $\widehat{\rho}$ is the mean activation value. s_2 is the nodes number of the hidden layers.

III. Proposed Model for Tool Wear Prediction

A. Novel MD-SSAE Structure With a Modified Loss Function

Feature fusion technologies are always employed in the application of tool wear prediction, which is attributable to a fact that different feature domains in the raw signal data have different valuable information and characteristics. However, existing DL models have no ability to extract features from multiple feature domains simultaneously [28]. Thus, this article proposes an MD-SSAE model formed by four different SSAEs. And raw TD data, FD data, LF-WPC, and HF-WPC of the raw vibration signal data can be used as the input of the proposed model to capture more valuable and implicit features in different feature domains.

Fig. 2 represents the training process of the MD-SSAE model. The structures of four SSAEs consist of an input layer x^1, several hidden layer x^j $(j = 2, 3, \ldots, n-1)$, and output layer $y = x^n$. The input layer consists of raw TD data $x_1^1 \in \mathbf{R}^{n_1}$, FD data $x_2^1 \in \mathbf{R}^{n_2}$ following the application of FFT, LF-WPC $x_3^1 \in \mathbf{R}^{n_3}$, and HF-WPC $x_4^1 \in \mathbf{R}^{n_4}$ following WPD. Each hidden layer have four vectors $x_1^j, x_2^j, x_3^j, x_4^j \in \mathbf{R}^{m_j}$. Through calculation of all the autoencoders, four feature vectors, y_1, y_2, y_3, y_4, will be obtained via the new modeling framework.

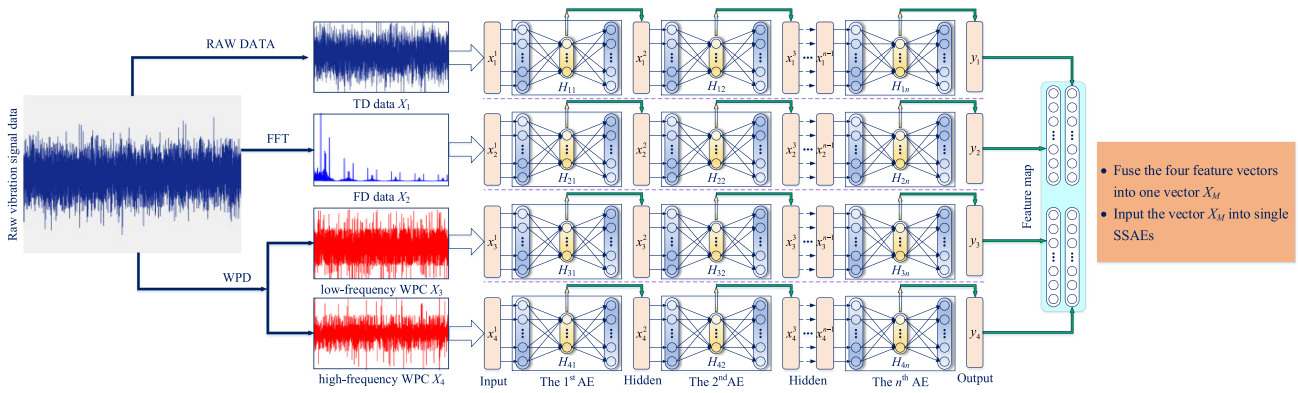

Fig. 2. Training process of the MD-SSAE model.

The loss function of the proposed MD-SSAE model in this section is modified as follows:

$$L^j_{k\ \text{sparsel}}\left(W^j_k, b^j_k\right) = \frac{1}{2m}\sum_{i=1}^{m}\left\|\sigma'_k\left[W'^j_k \cdot \sigma_k\left(W^j_k \cdot x^{j,i}_k + b^j_k\right) + b'^j_k\right] - x^{j,i}_k\right\|^2 + \frac{\lambda_k}{2}\sum_{s=1}^{j}\sum_{c=1}^{c\left(W^s_k\right)}\sum_{r=1}^{r\left(W^s_k\right)}\left\{[W^s_k]_{rc}\right\}^2 + \beta_k\sum_{t=1}^{r\left(W^s_k\right)} KL\left(\rho_k||\widehat{\rho}^j_{k,t}\right) \tag{7}$$

where $k = 1, 2, 3, 4$ and $j = 1, 2, \ldots, n_k$ are the kth SSAEs and the jth autoencoder accordingly. X is the input of autoencoder. $c(W)$ and $r(W)$ are the number of columns and rows of the weight matrix W, respectively.

The optimal solution of the weight matricesW^j_k can be calculated by using (8). And multidimensional deep features y_1, y_2, y_3, y_4 are obtained through the training of this new model

$$W^j_k = \text{argmin}\left[L^j_{k\ \text{sparse}}\left(W^j_k, b^j_k\right)\right]. \tag{8}$$

B. Feature Fusion by Combining a Single-Dimensional SSAE With a Nonlinear Regression Function

After the training of the MD-SSAE model, the four deep features vectors, y_1, y_2, y_3, and y_4, will be fused into a new vector $X_M = [y_1\ y_2\ y_3\ y_4]$, and the vector X_M will be used as the input of a single-dimensional SSAE for deeper features extraction. Due to the good properties of tool wear process in nonstationarity and complex nonlinear, a nonlinear regression function is utilized to connect to the output layer of the single-dimensional SSAEs, which will enhance the progressive tool wear prediction tasks.

Fig. 3 represents the learning process of the features fusion structure. First, the feature $X_M \in \mathbf{R}^{X_m}$ is input into the single-dimensional SSAEs, and the hidden representation $H_1 \in \mathbf{R}^{H_1}$ can be calculated by the weight matrix $W_1 \in \mathbf{R}^{H_1 \times X_m}$. Then, the deeper representation $X_F \in \mathbf{R}^{X_f}$ can be calculated by the

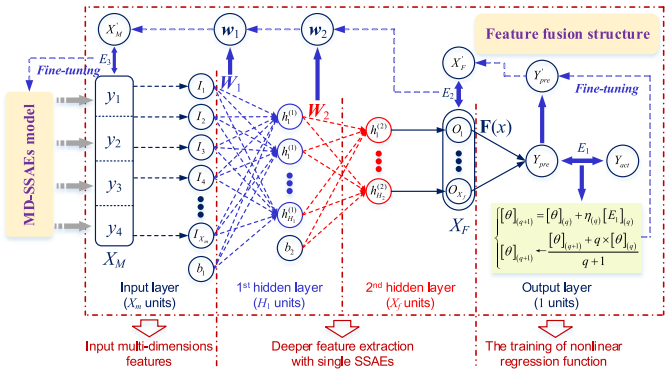

Fig. 3. Learning process of the features fusion structure.

weight matrix $W_2 \in \mathbf{R}^{X_f \times H_1}$ and H_1. Subsequently, the feature X_F is used as the input of a nonlinear regression function $\mathbf{F}(x)$ to obtain the output value Y_{pre}. Finally, the absolute error E_1 between Y_{pre} and actual tool wear value Y_{act} will be calculated

$$\mathbf{F}(x) = \frac{1}{A + e^{-Bx+C}} \tag{9}$$

$$E_1 \leftarrow |Y_{\text{pre}} - Y_{\text{act}}| = |\mathbf{F}(X_F) - Y_{\text{act}}| \tag{10}$$

where A, B, and C are the constant, the matrix, and the vector of the nonlinear regression function, respectively.

To minimize E_1, the updated equations for the parameter set $\theta = \{A, B, C\}$ are

$$\begin{cases} [\theta]_{(q+1)} = [\theta]_{(q)} + \eta_{(q)}[E_1]_{(q)} \\ [\theta]_{(q+1)} \leftarrow \frac{[\theta]_{(q+1)} + q \times [\theta]_{(q)}}{q+1} \end{cases} \tag{11}$$

with

$$\begin{cases} \eta_{(q+1)} \leftarrow R_l \times \eta_{(q)}, & \text{if } [E_1]_{(q+1)} \geq [E_1]_{(q)} \\ \eta_{(q+1)} \leftarrow R_h \times \eta_{(q)}, & \text{if } [E_1]_{(q+1)} < [E_1]_{(q)} \end{cases} \tag{12}$$

where $[\theta]_{(q)}$ and $[E_1]_{(q)}$ denote the parameter set and the error at the q-th iteration, respectively. q represents the current iteration

number. $\eta_{(q)}$ is the variation rate. R_l and R_h are decrease factor and increase factor, respectively. The empirical values of the decrease and increase factors are set as $R_l \in [0.2, 1)$ and $R_h \in (1, 5]$, respectively, in this article.

After minimizing the error E_1, the parameters A, B, and C will be fine-tuned into a, b, and c. And then the output value Y_{pre} are transformed into Y'_{pre} based on X_F and a, b, c, and the feature X'_F can be calculated by Y'_{pre} and a, b, c. Finally, the error E_2 between X'_F and X_F is input into the single-dimensional SSAEs for the purpose of the error back propagation, as detailed in the following section.

C. Error Back Propagation Rules

After inputting the error E_2 into the single-dimensional SSAEs, W_2 will be fine-tuned into w_2 based on E_2, and then W_1 will be updated into w_1 based on w_2. The hidden representation x_2 are transformed into x'_2 based on X_M and w_1. And then, the feature vector X'_M can be calculated by x'_2 and W'_1. Finally, the error E_3 between X'_M and X_M is input into MD-SSAE model.

And now, the loss function of the single-dimensional SSAEs is modified as

$$L_{\text{sparse}}(W_p, b_p) = J(W_p, b_p) + \frac{\lambda}{2}\sum_{l=1}^{p}\sum_{c=1}^{c(W_l)}\sum_{r=1}^{r(W_l)}([W_l]_{rc})^2 + \beta\sum_{j=1}^{r(W_l)} KL\left(\rho||\widehat{\rho}_{p,j}\right) \tag{13}$$

$$J(W_p, b_p) = \begin{cases} \frac{1}{2m}\sum_{i=1}^{m}\left\|X_F^i - X_F'^i\right\|^2 & p = n-1 \\ \frac{1}{2m}\sum_{i=1}^{m}\left\|\sigma\left(W_{p+1}x_{p+1}^i + b_{p+1}\right)\right. & \\ \quad \left. -\sigma\left(w_{p+1}x_{p+1}^i + b_{p+1}\right)\right\|^2 & p = n-2, \ldots, 1 \end{cases} \tag{14}$$

with

$$X_F = x_n;\ x_{p+1} = \sigma(W_p x_p + b_p);\ X_M = x_1 \tag{15}$$

where n is the layer number of the single-dimensional SSAEs. w is the updated weight matrix.

The optimal parameter set $\{w_p\}$ of the single-dimensional SSAEs is trained with stochastic gradient descent method, which can be updated as follows:

$$[w_p]_{(q+1)} = [W_p]_{(q)} - [\alpha]_{(q)} \cdot \left[\frac{\partial L_{\text{sparse}}(W_p, b_p)}{\partial W_p}\right]_{(q)} \tag{16}$$

where $[W_p]_{(q)}$ denotes the pth weight matrix at the qth iteration, and q represents the current iteration number. α is the learning rate.

Following the error back propagation rules, the weight matrices W_k^j will be fine-tuned based on the error E_3. And the loss function of MD-SSAEs is correspondingly modified as follows:

$$L^j_{k\ \text{sparse2}}\left(W_k^j, b_k^j\right) = J^j_{k2}\left(W_k^j, b_k^j\right) + \frac{\lambda_k}{2}\sum_{s=1}^{j}\sum_{c=1}^{c\left(W_k^s\right)}\sum_{r=1}^{r\left(W_k^s\right)} \times \left([W_k^s]_{rc}\right)^2 + \beta_k\sum_{t=1}^{r\left(W_k^s\right)} KL\left(\rho_k||\widehat{\rho}^j_{k,t}\right) \tag{17}$$

$$J^j_{k2}\left(W_k^j, b_k^j\right) = \begin{cases} \frac{1}{2m}\sum_{i=1}^{m}\left\|y_k^i - y_k'^i\right\|^2 & j = n_k \\ \frac{1}{2m}\sum_{i=1}^{m}\left\|\sigma_k\left(W_k^{j+1}\cdot x_k^{j+1,i} + b_k^{j+1}\right)\right. & \\ \quad \left. -\sigma_k\left(w_k^{j+1}\cdot x_k^{j+1,i} + b_k^{j+1}\right)\right\|^2 & j = n_k - 1, \ldots, 1 \end{cases} \tag{18}$$

with

$$y_k = x_k^{n_k};\ x_k^{j+1} = \sigma_k\left(W_k^j x_k^j + b_k^j\right). \tag{19}$$

The optimal parameter set $\{w_k^j\}$ of the proposed MD-SSAE model is trained with stochastic gradient descent method, which can be updated as follows:

$$\left[w_k^j\right]_{(q+1)} = \left[W_k^j\right]_{(q)} - [\alpha_k]_{(q)} \cdot \left[\frac{\partial L^j_{k\ \text{sparse2}}\left(W_k^j, b_k^j\right)}{\partial W_k^j}\right]_{(q)} \tag{20}$$

where $k = 1, 2, 3, 4$. $[W_k^j]_{(q)}$ denotes the jth weight matrix of the kth SSAEs at the qth iteration, and q represents the current iteration number. α_k is the learning rate.

By now, the novel DL modeling framework has been designed. And the construction of the proposed model can be accomplished through iterative training. Finally, the proposed modeling framework can be used for tool wear prediction.

IV. Experimental Setup and Data Acquisition

A. Experimental Setup

In order to validate the effectiveness of the proposed method, vibration data are collected from an IQ300 machining center which is used for producing high-precision molds in a production line. Fig. 4 depicts the cutting tools experimental setup and the data acquisition system. A ball-end milling cutter (material: diamond) is tested to mill a workpiece (material: stainless steel). The accelerometer with sensitivity of 101.47 mv/g is mounted on the spindle seat of the machine center for measuring the x-direction vibration signals. The vibration signals were collected at a sampling frequency of 10 kHz, and the sampling time is about 47 s. The speed of the spindle speed is set at 18 000 r/min. Based on the parameters and the speed, the characteristic fault frequencies of the cutting tool can be calculated. More details of the cutting parameters are listed in Table I.

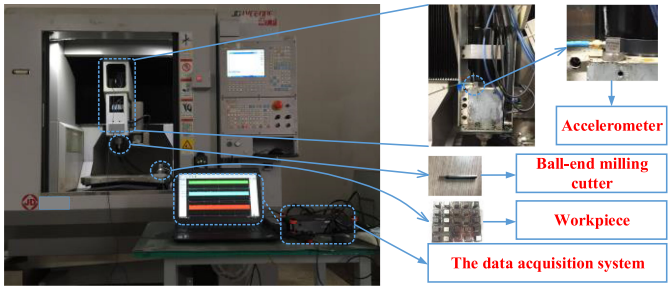

Fig. 4. Cutting tools experimental setup and the data acquisition system.

TABLE I
EXPERIMENTAL PARAMETERS OF THE CUTTING TOOLS

Test no.	Cutting speed v_s (rpm)	Feed rate f (mm/min)	Cutting depth a_p (mm)	Tool diameter d (mm)
1	18000	800	0.02	2.0
2	18000	800	0.02	1.5
3	18000	800	0.015	2.0
4	18000	800	0.015	1.5

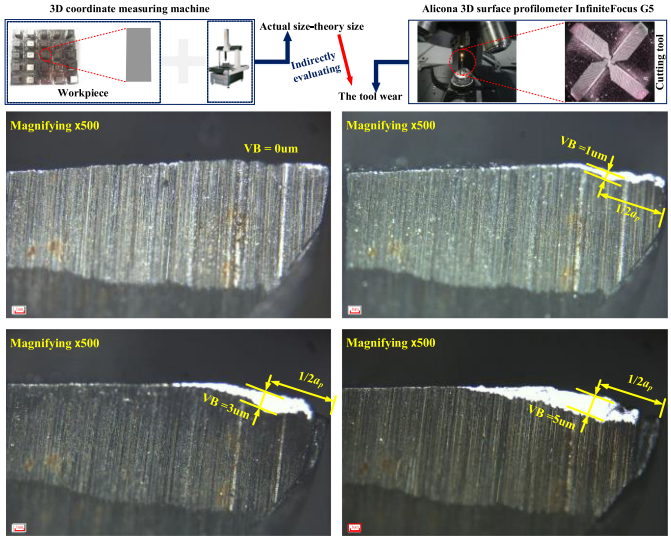

Fig. 5. Flank wear width under 3-D surface profilometer magnifying $\times$500.

B. Descriptions of Datasets

The data acquisition is described in the following points.

1) All the vibration signal dataset acquisition is performed at a real manufacturing environment, not the laboratory.
2) In this article, the tool wear value at the location of $1/2a_p$ on the flank face was taken as an object of study [32], as illustrated in Fig. 5. The flank wear width (abbreviated as VB) is measured offline using an Alicona three-dimensional (3-D) surface profilometer InfiniteFocus G5 at a magnification of $\times$500. To make sure that the tool wear is precise, a three-dimensional coordinate measuring machine is utilized to measure the size of each finished workpiece, and the difference between theory size and actual size will be the evaluation criterion of tool wear indirectly.
3) Altogether 400 data files (one data file corresponds to one cut number) are collected during the cutting tests, and a total of four sets of test data are assessed for this study, as listed in Table I. To split training/testing domains, three datasets are selected for training, and the rest one dataset for testing the proposed modeling framework.

V. PERFORMANCE EVALUATION

A. Comparison With Traditional ML Methods

BPNN, SVR, LSTM, and GPR models are widely used for tool wear prediction, as discussed in Section I. However, the following two important points need to be pointed out. 1) Different from these traditional ML models, the DL-based modeling framework proposed in this article can extract and fuse the features from the raw data without manual intervention. 2) The input of the proposed model is TD data, FD data, LF-WPC, and HF-WPC, whereas the input of traditional models is TD features, FD features, and wavelet domain features. In this section, a comparison is made between the traditional models against the proposed modeling framework.

For comparison purposes, the performance of BPNN, SVR, LSTM, and GPR will be discussed based on the same features. A total of 220 statistical features from time domain, frequency domain and time–frequency domain are extracted, more details can be seen in [29]–[37]. In general, multidomain features can be viewed as a high-dimensional multivariate matrix composed of several vectors formed by different domains. However, it is not feasible to input the abovementioned matrix to a model without dimension reduction because of the high correlation between vectors. Thus, KPCA is used for feature fusion and selection to extract discriminative features [7] and then input the sensitive features into the traditional ML models.

The test no.1 and no.4 predicted results of the five models (BPNN, SVR, LSTM, GPR, and the proposed method) are shown in Fig. 6. The error in each subfigure represents the absolute value of the difference between the predicted and measured tool wear. It can be seen that GPR and the proposed method can follow the tracks of actual tool wear effectively. BPNN, SVR, and LSTM perform not so well.

To compare the performance of the abovementioned models, four indexes are investigated to evaluate the predicted results: PCC, MAE, RMSE, and MAPE [7]. Note that a larger PCC and a smaller MAE/RMSE/MAPE imply better effectiveness of the tool wear predictive model. Performance comparison of the five models under different evaluation indexes is listed in Table II, and the detailed modeling results are shown in Fig. 7. It can be seen that among the different models, the proposed method has the maximum PCC value and the minimum MAE/RMSE/MAPE value.

These results indicate that 1) KPCA incorporated with BPNN, SVR, LSTM, and GPR can predict the tool wear; however, the abovementioned models depend heavily on human-assisted feature extraction and selection. 2) Compared with traditional ML

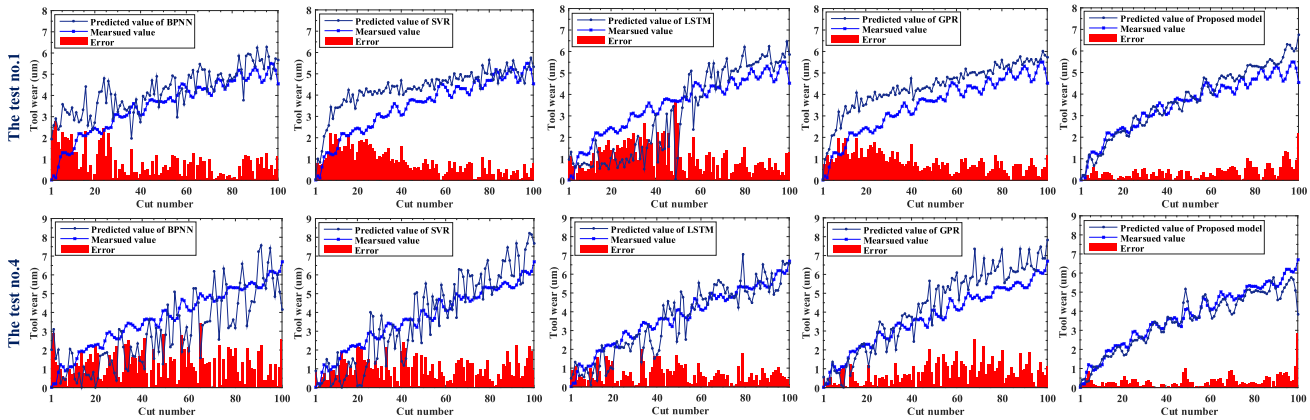

Fig. 6. Test no.1 and no.4 predicted results of the five tool wear predictive models (BPNN, SVR, LSTM, GPR, and the proposed model).

TABLE II
PERFORMANCE COMPARISON OF THE FIVE MODELS UNDER DIFFERENT EVALUATION INDEXES

Models	The test no. 1	2	3	4	1	2	3	4
	PCC				MAE			
KPCA+BPNN	0.78	0.87	0.55	0.75	0.84	1.00	1.59	1.26
KPCA+SVR	0.89	0.93	0.79	0.90	0.84	1.00	0.68	1.01
KPCA+LSTM	0.84	0.86	0.89	0.79	1.04	0.97	0.72	0.77
KPCA+GPR	0.95	0.95	0.92	0.94	0.47	0.92	0.50	0.81
Proposed model	**0.97**	**0.96**	**0.94**	**0.95**	**0.38**	**0.81**	**0.40**	**0.34**
Models	The test no. 1	2	3	4	1	2	3	4
	RMSE				MAPE			
KPCA+BPNN	6.97	8.17	13.44	7.30	0.47	0.23	0.67	0.28
KPCA+SVR	7.92	9.61	3.32	7.32	0.35	0.25	0.23	0.32
KPCA+LSTM	3.75	9.78	4.63	2.76	0.32	0.24	0.29	0.23
KPCA+GPR	3.73	7.75	2.47	5.38	0.25	0.22	0.22	0.22
Proposed model	**1.99**	**3.25**	**1.60**	**2.34**	**0.12**	**0.21**	**0.16**	**0.08**

The bold values indicate the best results, respectively.

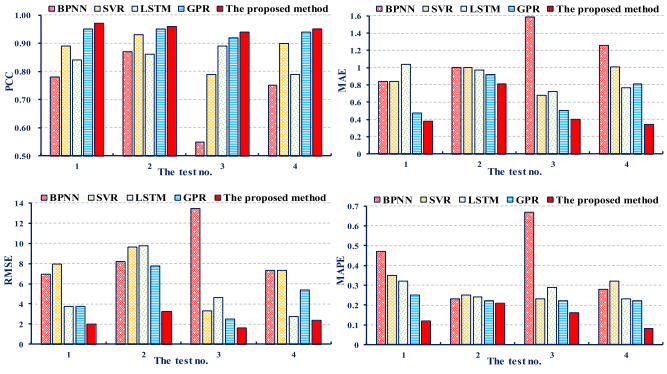

Fig. 7. Performance comparison of the five tool wear predictive models under different evaluation indexes.

methods, the proposed method can maintain higher prediction accuracy.

B. Comparison With SSAEs in Different Feature Domains

In recent years, with the increased computation capacity of computer, DL methods have been widely applied to condition monitoring and fault diagnosis [20]–[24]. However, existing DL models have no ability to extract features from multiple feature

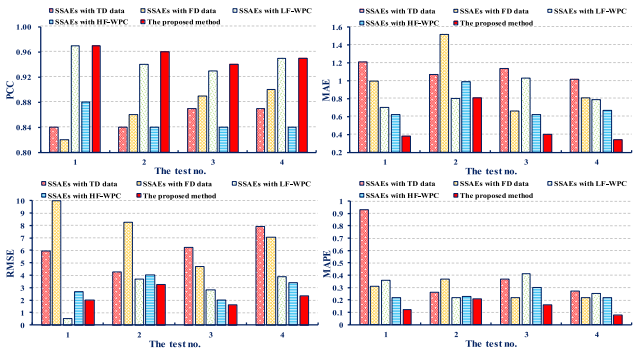

Fig. 8. Performance comparison of different feature domains on DL models under different evaluation indexes.

TABLE III
PERFORMANCE COMPARISON OF DIFFERENT FEATURE DOMAINS ON DL MODELS UNDER DIFFERENT EVALUATION INDEXES

Models	The test no. 1	2	3	4	1	2	3	4
	PCC				MAE			
SSAEs with TD data	0.84	0.84	0.87	0.87	1.21	1.07	1.14	1.02
SSAEs with FD data	0.82	0.86	0.89	0.9	1.00	1.52	0.66	0.81
SSAEs with FL-WPC	**0.97**	0.94	0.93	**0.95**	0.70	0.80	1.03	0.79
SSAEs with HL-WPC	0.88	0.84	0.84	0.84	0.62	0.99	0.62	0.67
Proposed model	**0.97**	**0.96**	**0.94**	**0.95**	**0.38**	**0.81**	**0.40**	**0.34**
Models	The test no. 1	2	3	4	1	2	3	4
	RMSE				MAPE			
SSAEs with TD data	5.93	4.29	6.24	7.94	0.93	0.26	0.37	0.27
SSAEs with FD data	9.97	8.28	4.70	7.04	0.31	0.37	0.22	0.22
SSAEs with FL-WPC	**0.52**	3.67	2.80	3.90	0.36	0.22	0.41	0.25
SSAEs with HL-WPC	2.68	4.04	2.00	3.40	0.22	0.23	0.30	0.22
Proposed model	1.99	**3.25**	**1.60**	**2.34**	**0.12**	**0.21**	**0.16**	**0.08**

The bold values indicate the best results, respectively.

domains simultaneously. In order to further scrutinize the performance of the proposed modeling framework, the effectiveness of the model and the different feature domains on comparable DL models will be studied and surveyed via several experimental trials. In this section, results will be discussed for the following four different feature domains: TD data, FD data, LF-WPC, and HF-WPC.

Performance comparison of different feature domains on DL models under different evaluation indexes are illustrated in Fig. 8 and Table III, and the detailed predicted results in the

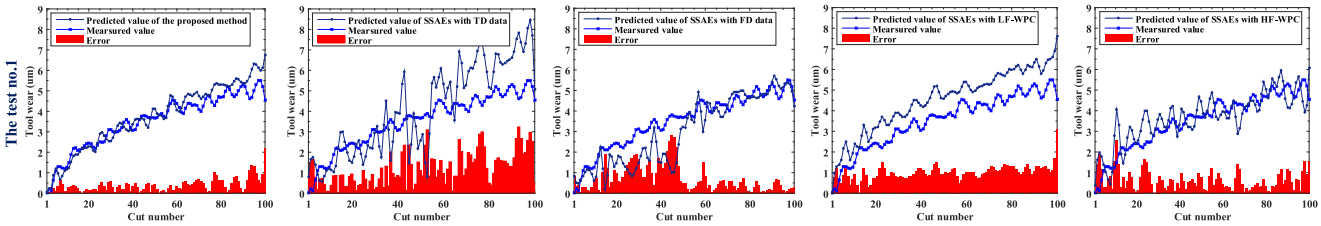

Fig. 9. Detailed predicted results of different feature domains on DL models in the test no.1.

TABLE IV
NETWORK PARAMETERS OF DIFFERENT TOOL PREDICTIVE MODELS

Models	Structure	Input	Layers	Nodes number	Learning rate	Iteration	Kernel function	Train_time (s)	Test_time (s)
The proposed method	MD-SSAEs	TD data	6	468880-46888-5000-500-80-10	0.6	2000	-	1180.59	0.042
		FD data	6	419435-46888-5000-500-80-10	0.6	2000	-		
		LF-WPC	6	234440-46888-5000-500-80-10	0.6	2000	-		
		HF-WPC	6	234440-46888-5000-500-80-10	0.6	2000	-		
	Single SSAEs	Multi-dimensions features	3	40-20-10	0.6	2000	-		
	NLR function	The final features	2	10-1	1	100	-		
Standard SSAEs	-	TD data	7	468880-46888-5000-500-80-10-1	0.6	2000	-	923.84	0.032
	-	FD data	7	419435-46888-5000-500-80-10-1	0.6	2000	-	917.65	0.031
	-	LF-WPC	7	234440-46888-5000-500-80-10-1	0.6	2000	-	864.11	0.028
	-	HF-WPC	7	234440-46888-5000-500-80-10-1	0.6	2000	-	869.35	0.028
KPCA+BPNN	-	220 statistics features	3	220-10-1	0.6	2000	-	24.98	0.019
KPCA+SVR	-	220 statistics features	-	-	-	2000	RBF	0.035	0.019
KPCA+LSTM	-	220 statistics features	3	220-10-1	0.6	2000	-	28.47	0.026
KPCA+GPR	-	220 statistics features	-	-	-	2000	Exponential	2.435	0.018

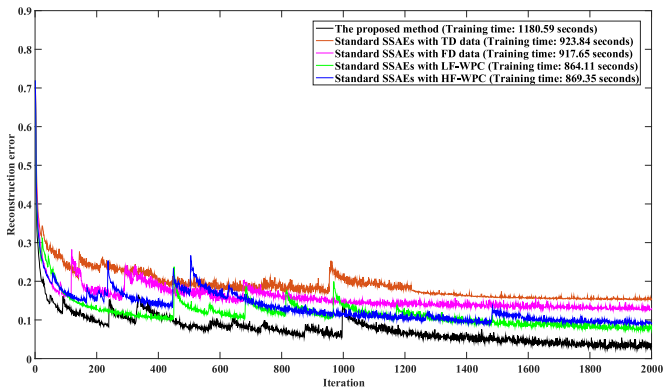

Fig. 10. Reconstruction error curves and time consumption of the proposed method and standard SSAE models based on different feature domains.

test no.1 are shown in Fig. 9. It can be concluded that the predicted results of different feature domains on DL models vary considerably. Among them, the proposed method with multiple-feature-domain fusion has the best performance (the minimum MAE/RMSE/MAPE value and the maximum PCC value), which is superior to the standard SSAE model with single feature domain.

The reconstruction error curves and time consumption of the proposed method and standard SSAE model based on different feature domains are shown in Fig. 10. Although the proposed method needs more training time (above 20 min) due to the increase in the number of network layers, hidden nodes, and parameters, the reconstruction error is smaller following convergence.

Table IV describes the structures of the proposed method. The input layer of this model is divided into four parts: TD data (raw vibration signal), FD data, LF-WPC, and HF-WPC. First, the dimension of TD data is 468 880, which can be calculated by multiplying the sampling frequency (10 kHz) by the sampling time (about 47 s). Then, FD data are obtained following FFT, and its dimension is 419 435. Finally, LF-WPC and HF-WPC are obtained following WPD, and theirs dimensions are 234 440 and 234 440, respectively. Due to over-high dimensions of the input layer in the proposed model, dimensionality reduction on the input data is necessary. However, if the decrease ratio of the layer nodes is too large, the deep model can cause overfitting results, so it is not easy to obtain the optimal model; if the decrease ratio is too small, then it will increase the complication of the model, and hence the unnecessary computing time. Thus, the decrease ratio of the layer nodes in the proposed method is controlled within [5], [10].

Table IV describes the parameters and the detailed time consumption of all the comparable tool wear predictive models. The results indicate that the SVR and GPR models perform the best in the terms of train_time and test_time, respectively. As for the proposed method, the optimization of network parameters needs to iterate repeatedly which is quite time-consuming. However, it is remarkable that the traditional models depend heavily on artificial feature extraction via expert knowledge, whereas the proposed method has the advantages of autonomous selection of implicit features.

The abovementioned results imply the following.

1) *Feature selection and extraction:* The traditional ML models rely on human-assisted feature extraction via expert expertise, whereas the proposed method has the advantages of autonomously capturing implicit representation. The performance of the traditional ML models may be further improved after selecting the sensitive features, but it is tedious and laborious.

2) *Model performance:* The proposed method shows better predicted accuracy and stability (Evaluation indexes: PCC, MAE, RMSE, and MAPE) than BPNN, SVR, LSTM, and GPR models (often used in tool wear prediction), as well as standard DL models with single feature domain. The reason is that the proposed method takes advantages of the modified loss function and enhanced network structure and can adaptively learn the implicit features from multiple feature domains of the raw data simultaneously.

VI. Conclusion

In this article, a new MD-SSAE model with feature fusion was proposed for tool wear prediction. First, a new multiple SSAE model was designed to learn more valuable features. To achieve this structure, a modified loss function was used to improve the feature learning ability. Then, a single-dimensional SSAE was used for feature fusion and deeper features learning. Finally, a nonlinear regression function was utilized to enhance the progressive tool wear prediction tasks due to the good properties of tool wear process in nonstationarity and complex nonlinear.

A dataset from a real manufacturing process was used to evaluate the performance of the proposed modeling framework. Experimental results show that the proposed model was more effective compared with the already established and developed ML methods, as well as standard SSAEs used for tool wear prediction. However, there still are certain limitations, such as the insufficient variability in setup cutting parameters of the different cutting tools, too long training time, and numerous network parameters.

With the rapid development of computational power and hardware technology, DL can be gradually applied in more industrial field. In the future work, further research and improvement in the generalization performance of our framework is needed, which includes more variability in cutting parameters and the estimation of RUL. We will continue to study this topic.

References

[1] B. Luo, H. T. Wang, H. Q. Liu, B. Li, and F. Y. Peng, "Early fault detection of machine tools based on deep learning and dynamic identification," *IEEE Trans. Ind. Electron.*, vol. 66, no. 1, pp. 509–518, Jan. 2019.

[2] C. Shi, G. Panoutsos, B. Luo, H. Q. Liu, B. Li, and X. Lin, "Using multiple feature spaces-based deep learning for tool condition monitoring in ultra-precision manufacturing," *IEEE Trans. Ind. Electron.*, vol. 66, no. 5, pp. 3794–3803, May 2019.

[3] 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.

[4] 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 Integration Intell. Syst.*, San Diego, CA, USA, 2015, pp. 271–276.

[5] C. Drouillet, J. Karandikar, C. Nath, A. C. Journeaux, M. E. Mansori, and T. Kurfess, "Tool life predictions in milling using spindle power with the neural network technique," *J. Manuf. Process.*, vol. 22, pp. 161–168, Apr. 2016.

[6] R. Corne, C. Nath, M. E. Mansori, and T. Kurfess, "Study of spindle power data with neural network for predicting real-time tool wear/breakage during inconel drilling," *J. Manuf. Syst.*, vol. 43, no. 2, pp. 287–295, Apr. 2017.

[7] D. D. Kong, Y. J. Chen, and N. Li, "Gaussian process regression for tool wear prediction," *Mech. Syst. Signal Proc.*, vol. 104, pp. 556–574, May 2018.

[8] D. F. Shi and N. N. Gindy, "Tool wear predictive model based on least squares support vector machines," *Mech. Syst. Signal Proc.*, vol. 21, no. 4, pp. 1799–1814, May 2007.

[9] W. Li, P. Fu, and W. Cao, "Study on the technology of tool wear monitoring by modifying least square support vector machine via Kalman filter," *Mech. Sci. Technol. Aerosp. Eng.*, vol. 34, no. 1, pp. 81–85, Jan. 2015.

[10] S. Dutta, S. K. Pal, and R. Sen, "On-machine tool prediction of flank wear from machined surface images using texture analyses and support vector regression," *Precis. Eng.*, vol. 43, pp. 34–42, Jan. 2016.

[11] D. D. Kong, Y. J. Chen, N. Li, and S. L. Tan, "Tool wear monitoring based on kernel principal component analysis and v-support vector regression," *Int. J. Adv. Manuf. Technol.*, vol. 89, no. 1, pp. 175–190, Mar. 2017.

[12] J. J. Wang, J. Y. Xie, R. Zhao, L. B. Zhang, and L. X. Duan, "Multisensory fusion based virtual tool wear sensing for ubiquitous manufacturing," *Robot. Comput.-Integr. Manuf.*, vol. 45, pp. 47–58, Jun. 2017.

[13] R. Zhao, J. J. Wang, R. Q. Yan, and K. Z. Mao, "Machine health monitoring with LSTM networks," in *Proc. 10th Int. Conf. Sens. Technol.*, Nanjing, China, 2016, pp. 1–6.

[14] 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," *Mod. Manuf. Eng.*, no. 12, pp. 68–72, 2014.

[15] J. A. Duro, J. A. Padget, C. R. Bowen, H. A. Kim, and A. Nassehi, "Multi-sensor data fusion framework for CNC machining monitoring," *Mech. Syst. Signal Proc.*, vol. 66/67, pp. 505–520, Jan. 2016.

[16] 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.

[17] G. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.

[18] 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, Dec. 2010.

[19] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[20] 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.

[21] Z. Q. Chen, C. Li, and R.-V. Sanchez, "Multi-layer neural network with deep belief network for gearbox fault diagnosis," *J. Vibroeng.*, vol. 17, no. 5, pp. 2379–2392, Aug. 2015.

[22] F. Jia, Y. G. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mech. Syst. Signal Proc.*, vol. 72/73, pp. 303–315, May 2016.

[23] 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.

[24] O. Janssens, V. Slavkovikj, B. Vervisch, K. Stockman, M. Loccufier, and S. Verstockt, "Convolutional neural network based fault detection for rotating machinery," *J. Sound Vib.*, vol. 377, pp. 331–345, Sep. 2016.

[25] K. J. Chen, J. Hu, and J. He, "A framework for automatically extracting overvoltage features based on sparse autoencoder," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 594–604, Mar. 2018.

[26] L. H. Meng, S. F. Ding, N. Zhang, and J. Zhang, "Research of stacked denoising sparse autoencoder," *Neural Comput. Appl.*, vol. 30, no. 7, pp. 2083–2100, Oct. 2018.

[27] C. S. Poultney, S. Chopra, and Y. Lecun, "Efficient learning of sparse representations with an energy-based model," in *Proc. 19th Int. Conf. Neural Inf. Process. Syst.*, 2007, pp. 1137–1144.

[28] M. Längkvist and A. Loutfi, "Learning feature representations with a cost-relevant sparse autoencoder," *Int. J. Neural Syst.*, vol. 25, no. 1, 2015, Art. no. 1450034.

[29] H.-C. Shin *et al.*, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.

[30] 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. 2018.

[31] 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.
[32] 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, no. 1-4, pp. 351–361, Jul. 2017.
[33] K. Zhu and T. Liu, "Online tool wear monitoring via hidden semi-Markov model with dependent durations," *IEEE Trans. Ind. Informat.*, vol. 14, no. 1, pp. 69–78, Jan. 2018.
[34] L. Hao, L. K. Bian, N. Gebraeel, and J. J. Shi, "Residual life prediction of multistage manufacturing processes with interaction between tool wear and product quality degradation," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 2, pp. 1211–1224, Apr. 2017.
[35] H. B. Sun, D. L. Cao, Z. D. Zhao, and X. Kang, "A hybrid approach to cutting tool remaining useful life prediction based on the Wiener process," *IEEE Trans. Rel.*, vol. 67, no. 3, pp. 1294–1303, Sep. 2018.
[36] O. Geramifard, J. X. Xu, J. H. Zhou and X. Li, "Multimodal hidden Markov model-based approach for tool wear monitoring," *IEEE Trans. Ind. Electron.*, vol. 61, no. 6, pp. 2900–2911, Jun. 2014.
[37] 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.

Chengming Shi received the bachelor's degree in mechanical engineering from the School of Mechanical Engineering, Xiangtan University, Xiangtan, China, in 2010. He is currently working toward the Ph.D. degree in mechatronic engineering with the Huazhong University of Science and Technology, Wuhan, China.

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

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.

Songping He received the Ph.D. degree in mechatronic engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2015.

He is currently an Associate Research Professor in Mechanical Engineering with the School of Mechanical Science and Engineering, Huazhong University of Science and Technology. His current research interests include intelligent manufacturing, high-end electronic equipment manufacturing, and digital mock-up.

Kai Li received the bachelor's degree from the School of Mechanical and Electrical Engineering, Wuhan University of Technology, Wuhan, China. He is currently working toward the Ph.D. degree in the School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan.

His current research interests include intelligent manufacturing, transfer learning, signal processing, and milling stability.

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 Computerized Numerical Control (CNC) machine tools.