IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 1, 2020, accepted December 21, 2020, date of publication December 24, 2020, date of current version January 5, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3047205

Tool Wear Status Recognition and Prediction Model of Milling Cutter Based on Deep Learning

YANG XIE, CHAOYONG ZHANG, AND QIONG LIU
School of Mechanical Science and Engineer, Huazhong University of Science and Technology, Wuhan 430074, China

Corresponding author: Chaoyong Zhang (zcyhust@mail.hust.edu.cn)

This work was supported in part by the Project of International Cooperation and Exchanges NSFC under Grant 51861165202, in part by the National Natural Science Foundation of China under Grant 51705263 and Grant 51805330, and in part by the 111 Project of China under Grant B16019.

ABSTRACT In order to ensure the reliability and stability of the manufacturing process, tool wear state should be realized real-time and accurate monitoring. This paper proposes a tool wear state recognize and predictive framework model based on Stacking Sparse De-noising Auto-Encoder (SSDAE), the Particle Swarm Optimization (PSO) and the Least Squares Support Vector Machine (LSSVM). The Stacking Sparse De-noising Auto-Encoder (SSDAE) technique is utilized to realize multi-feature signal dimension reduction with the aim of improving the prediction accuracy, which reduces the dependence on the prior knowledge of feature selection and greatly improves the modeling efficiency. PSO technique is helpful for adaptive optimization of kernel parameters, which greatly improved computing power and LSSVM model prediction accuracy. A dataset from a real machining process is utilized to verify the effectiveness of proposed model in improving the prediction accuracy. The experimental results show that a high correlation coefficient greater than 0.95 is used to extract feature vector from time domain, frequency domain and time-frequency domain three directions, and the proposed SSDAE-PSO-LSSVM model performs better than Partial Least Squares Regression (PLSR), Back Propagation Neural Network (BPNN) and Extreme Learning Machine (ELM) in terms of prediction accuracy.

INDEX TERMS Tool condition monitoring, deep learning, SSDAE, feature fusion, PSO-LSSVM algorithm.

I. INTRODUCTION

Tool wear state is considered as a crucial factor to ensure the reliability and stability of the manufacturing process. However, tool wear is the bane of manufacturing users due to its ubiquity in machining processes caused by the inevitable frictions between tool edges and workpieces. According to previous studies, the cost of maintenance performed to lessen the influence of tool failure can range from 15 to 40% of the cost of goods produced [1]. In modern manufacturing systems, tool failure results in up to 20% downtime, prescribing a tremendous loss of productivity and profits [2]. Thus, real-time and accurate assessment of tool wear status can not only reduce production costs, but also effectively improve machining tool utilization rates. The tool wear state must be accurately predicted to guarantee adequate replacement and maintenance.

The associate editor coordinating the review of this manuscript and approving it for publication was Cesar Vargas-Rosales.

With the development of sensing techniques, by the fused information of indirect measurements (e.g. current, power, cutting force, vibration, and acoustic emission, etc.) generate in the machining process, many efforts have been devoted to the measurement of indirect indicators which can conceal tool wear. In recent decades, many artificial intelligence methodologies have been widely used for tool wear monitoring in manufacturing and processing base on these indirect indicators. Commonly used artificial intelligence algorithms (shallow-layer learning model) for tool wear monitoring, such as artificial neural networks (ANN), hidden Markov model (HMM) and support vector machines (SVM), etc [3]. Yen *et al.* [4] applied a self-organization feature map (SOM) neural network (NN) to acoustic emission (AE) signal-based tool wear monitoring for a micro-milling process. Karali P *et al.* [5] proposed an approach based on the mean square root(MSR) of wavelet packet coefficients captured from AE signals and an ANN model for tool wear monitoring and indicated that MSR values of the wavelet coefficients show a positive correlation to increasing drill wear. Ertunc *et al.* [6] selected HMM for the identification of tool wear state based on the measurement of cutting force and power signals. Benkedjouh *et al.* [7] applied two non-linear feature dimension-reduction techniques to SVM for tool wear state monitoring and predicting. Dutta *et al.* [8] utilized e-support vector regression (e-SVR) to realize tool flank wear prediction; it was shown that the texture features obtained from the turned surface images, with a 4.9% maximum predicted error. Zhang *et al.* [9] presented a tool wear prediction model based on a least squares support vector machine (LSSVM) for a ball-end milling tool and found that the LSSVM based model performed better than the ANN-based model in prediction accuracy. Although there is clear evidence from the above literature that the feature identification followed by artificial intelligence algorithms (shallow-layer learning model) has been successfully applied in tool wear prediction and monitoring. However, the above-mentioned researches not only need the measurement of indirect indicators, but also need complex feature extraction and selection based on prior knowledge. In addition, manual feature extraction often causes some loss of the original signal, which lowers sensitivity of the extracted features. Therefore, it is necessary to utilize adaptive features mining instead of manually features extraction for tool wear recognition.

Compared to traditional machine learning and intelligent system methods, deep learning model has huge advantages in data processing scale, non-linear ability, and convergence. It can effectively avoid the limitation of artificial feature extraction and obtain high monitoring accuracy and good generalization performance. The significant difference between deep learning and traditional machine learning methods is that the former can adaptively learn valuable features from raw data. In other words, deep learning models can get rid of the dependence on the prior knowledge about advanced signal processing techniques and domain experts [10].

Deep learning was first proposed in 2006 when an article written by Hinton published in Science [11]. With the advent of big data era, deep learning has becoming a hot spot in the field of artificial intelligence, and it has achieved much progress in the field of image recognition, speech recognition, natural language processing and so on [12]–[14]. In the last three years, artificial intelligence methodologies depend on deep learning have been developed and applied in fault diagnosis and life prediction. Various deep learning models can be divided into three main types [15]: Stacked Auto-Encoder Network (SAE) [16], [17], deep belief networks (DBN) [18], [19] and convolution neural network (CNN) [20], [21]. Sannino G *et al.* [22] present a Deep Neural Network (DNN) for the heartbeat classification by using Tensor Flow and Google deep learning library. Sherin MM. [23] used Restricted Boltzmann Machine (RBM) and DBN for the classification of single-lead electrocardiogram (ECG) signals. Vincent P *et al.* [24] explored an original strategy for building deep networks based on stacking denoising auto-encoder (SDAE) and clearly established the value of using a denoising criterion. Song K *et al.* [25] proposed network uses the deep convolutional neural network (DCNN) model to extract features from the spindle current clutter signal (SCCS) as the wear state of milling cutter classification index. Chen *et al.* [26] used deep belief network (DBN) to predict the flank wear of a cutting tool and the result was shown that the performance of the DBN have the absolutely advantage compared with the performances of ANNs and SVR in terms of the mean-squared error (MSE). The aforementioned literatures demonstrate that SDAE is easy to train and it is a purely unsupervised feature learning model, besides, it is also a promising tool for handling massive amounts of data. Although SDAE have been gradually applied in the intelligent manufacturing industry, few studies have focused on their application in tool condition monitoring, particularly using fusion data collected from multiple sensors.

This paper aims to propose a novel method for automatic and accurate identification of tool wear state by recurring to SSDAE and PSO-LSSVM. First, fast Fourier transform (FFT) and wavelet transform (WT) is used to preprocess the collected vibration data for creating an initial data set. The process noise data and redundancy data are bound to appear in the extracted signal features which will affect the prediction accuracy of LSSVR, Pearson correlation coefficient method is employed to correlate the high-layer features to the tool wear from time domain (TD), frequency domain (FD) and time-frequency domain (TFD) three directions. Then, the SSDAE technique is utilized to realize multi-feature signal dimension reduction with the aim of improving the prediction accuracy, which reduces the dependence on the prior knowledge of feature selection and greatly improves the modeling efficiency. Finally, a PSO-LSSVM model is established with the aim of improving optimization performance and convergence speed. Besides, the comparison between PSO-LSSVR and the traditional methods such as PLS, ANN and SVM will also be carried out to further show the advantages of PSO-LSSVM in tool wear prediction.

The rest of this paper is organized as follows. Section II introduces the background theory on SSDAE and the LSSVM. The proposed modeling framework about tool wear predictive model based on SSDAE and PSO-LSSVR is detailed in Section III. Analysis of the presented model and the experimental results are discussed in Section IV. Finally, the conclusions are summarized in Section V.

II. BACKGROUNDS

A. STACKING SPARSE DE-NOISING AUTO-ENCODER

1) AUTO-ENCODER

Stack de-nosing auto-encoder (SDAE) is widely accepted as the main deep learning methods. In essence, AE network is a one of unsupervised learning domains, and its goal is to reconstruct the input signal. The error of the input signal and the reconstructed signal is the error of the whole network. According to the principle of error back propagation, the network can be trained to reproduce the input signal as much as possible. What is ultimately needed is the eigenvectors of the

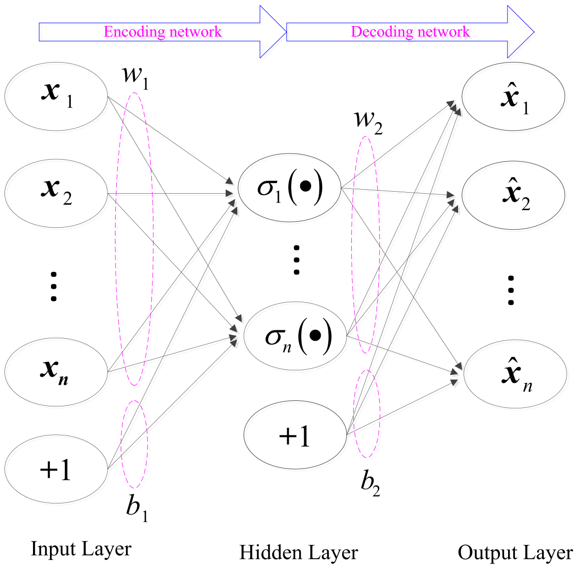

FIGURE 1. **Network structure of AE.**

compressed hidden layer of dimensions. Network structure of AE as shown in FIGURE.1.

The AE network can be divided into two parts: coding network and decoding network. For the coding network, $\{x^n | n = 1, 2, \ldots, n\}$ denotes the sample data set that training for the AE network, the encoding network process from the input layer to the hidden layer can be expressed by Eq. (1)

$$\mathrm{h}^n = f_{\theta 1}(\mathrm{x}^n) = \sigma(w_1 \mathrm{x}^n + b_1) \tag{1}$$

where x and σ denote the original input and activation function, $\theta_1 = \{\omega_1, b_1\}$ represents the parameter set of the encoding network, h denote the hidden feature, ω_1 is the weight matrix and b_1 is the bias vector.

The decoding network process from the hidden layer to the output layer can be expressed by Eq. (2).

$$\hat{\mathrm{x}}^n = g_{\theta_2}(\mathrm{h}^n) = \sigma(w_2 \mathrm{h}^n + b_2) \tag{2}$$

where σ denote sparse representation of the sample; $\theta_2 = \{\omega_2, b_2\}$ represents the parameter set of the decoding network, ω_2 is the weight matrix and b_2 is the bias vector.

The goal of the AE network is to find the final reconstructed data set as close as possible to the original input, in other words, optimal parameters $\theta = \{\omega_1, b_1, \omega_2, b_2\}$ are obtained to enable the output to be equal to the input data. The training process of AE network is to minimize the reconstruction error loss function $L = \{\omega_1, b_1, \omega_2, b_2\}$. The reconstruction error loss function of AE can be expressed as:

$$L(\omega_1, b_1, \omega_2, b_2) = \left[\frac{1}{n}\sum\nolimits_{i=1}^{n} J\left(\mathrm{x}^{(i)}, \hat{\mathrm{x}}^{(i)}\right)\right] + \frac{\lambda}{2}\sum\nolimits_{l=1}^{nl-1} \times \sum\nolimits_{i=1}^{sl}\sum\nolimits_{j=1}^{sl+1}\left(W_{ij}^{(l)}\right)^2 \tag{3}$$

$$J\left(\mathrm{x}^{(i)}, \hat{\mathrm{x}}^{(i)}\right) = \frac{1}{2}\left\|\mathrm{x}^{(i)} - \hat{\mathrm{x}}^{(i)}\right\|^2 = \frac{1}{2}\left\|\mathrm{x}^{(i)} - \sigma\left(w_2 \cdot \sigma(w_1 \mathrm{x}^{(i)} + b_1) + b_2\right)\right\|^2 \tag{4}$$

As shown in Eq. (4), $J\left(\mathrm{x}^{(i)}, \hat{\mathrm{x}}^{(i)}\right)$ stands for the mean square error (MSE) between $\mathrm{x}^{(i)}$ and $\hat{\mathrm{x}}^{(i)}$. In Eq. (3), $\mathrm{x}^{(i)}$ and $\hat{\mathrm{x}}^{(i)}$ represent the input vector and sparse representation of the ith samples, second item is the regularization constraint item, which is used to prevent over fitting and local minimum. When the second term is too large, it is hard to keep sparse; when the second item is very close to 0, the penalty is too strong; it is difficult to get the eigenvector of complete information. In generally, this item set value 0.05, a more appropriate result can be obtained. In order to punish those neurons with high activity, suppress their expression and finally realize the sparse structure of the whole hidden layer. The actual activation is adjusted by using KL divergence, as shown in the Eq. (5).

$$KL(\rho||\rho_j) = \rho \times \log\frac{\rho}{\rho_j} + (1-\rho) \times \log\frac{\rho}{\rho_j} \tag{5}$$

where ρ and ρ denote the sparsity parameter of the hidden layer and mean activation, they are similar to the second item in Eq. (3).

Finally, the reconstruction error loss function of SAE can be expressed as

$$\mathrm{J}_{SAE}(\theta) = \min_{\theta,\theta',\beta,\rho}\left[\mathrm{L}(x^m, g) + \beta\sum_{i=1}^{m}\mathrm{KL}(\rho\|\hat{\rho}_j)\right] \tag{6}$$

2) DE-NOISING AUTO-ENCODER

The existence of noise is inevitable in signal acquisition, which puts forward higher requirements for feature extraction. If the test data set containing noise is used to test the network, the input data itself is not subject to the original data distribution. Due to the above factor, the features obtained by AE method may be unreliable. In order to solve the problem of data deviation caused by noise, de-noising auto-encoder network (DAE) was designed, which is to add noises to the training data on the basis of the Auto-encoder.

The random noise $\boldsymbol{x}^n$ will be added in the sample by the q_D distribution, as shown in the Eq. (7)

$$\widetilde{\mathrm{x}} \sim q_D\left(\widetilde{\mathrm{x}}|\mathrm{x}\right) \tag{7}$$

where $\widetilde{x}$ denotes the corrupted form of x by adding random noise and is achieved by stochastic mapping $q_D\left(\widetilde{\mathrm{x}}|\mathrm{x}\right)$.

Differ from the traditional auto-encoder network; DAE uses the eigenvectors containing noise in the training process. In DAE method, we add random noise into the training samples. DAE must learn to remove noise in order to obtain input characteristics which are not contaminated by noises. Therefore, the generalization ability has a significant enhancement and the robustness of feature expression can be improved.

3) STACKED AUTO-ENCODER

Traditional auto-encoder are generally divided into three simple layers; however, the learning ability is limited when dealing with dimension reduction problems of high-dimensional eigenvectors. Stacked auto-encoder (SAE) is stacked by many DAEs, which depend on multiple hidden layer stacks, can

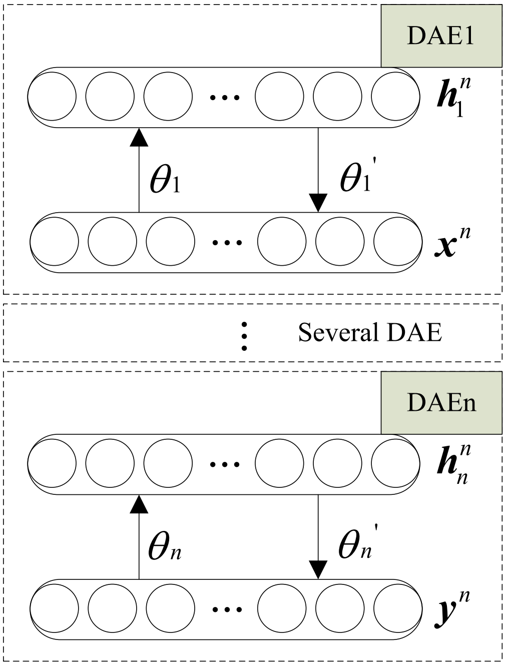

FIGURE 2. Network structure of SDAE.

handle more abstract and complex tasks and effectively achieves the objective of deep learning.

When the number of hidden layers is greater than 1, the auto-coder is treated as a deep structure, which is called SAE. The training of SAE network is shown in FIGURE. 2: (1) Input the initial data, train DAE1 network and gets feature encoding according to the unsupervised training method; (2) The hidden layer output of the first layer of the automatic encoder is taken as the input data of the second layer of the AE, and the second layer of the AE is trained in the same way; (3) Repeat the second step until you have completed all of the AE training; (4) Training the DAEn network and get the final feature encoding, the output of the hidden layer is taken as the final dimension reduction feature.

B. THE PARTICLE SWARM OPTIMIZATION ALGORITHM

The particle swarm optimization (PSO) is an evolutionary optimization algorithm, where a population of particles or proposed solutions evolves by each iteration, moving towards the optimal solution of the problem. Instead of having crossover and mutation as in genetic algorithms, PSO follows the optimal particle in the solution space for searching. Compared with genetic algorithms, PSO has the advantage of being simple and easy to implement without many adjusting parameters.

The algorithm updates the positions and the velocities of the particles following the equations

$$v_i^{k+1} = \beta v_i^k + c_1 r_1 \left(p_{best} - s_i^k\right) + c_2 r_2 \left(g_{best} - s_i^k\right) \quad (8)$$

$$s_i^{k+1} = s_i^k + s_i^{k+1} \quad (9)$$

where $v_i = (v_{i1}, v_{i2}, \ldots, v_{id})$ and $s_i = (s_{i1}, s_{i2}, \ldots, s_{id})$ are the velocity and position of particle; k is the number of iterations, and β is the inertia Weight; d is the total number of particles; r_1 and r_2 denote the random numbers distributed uniformly in the interval [0, 1]; c_1, c_2 denote the learning factor; $p_{best} = (p_{i1}, p_{i2}, \ldots, p_{id})$ represents the optimal position found by the ith particle search (optimal solution); $g_{best} = (g_1, g_2, \ldots, g_d)$ denotes the optimal location of group search (optimal solution).

C. LEAST SQUARES SUPPORT VECTOR REGRESSION

SVM is a machine-learning tool and especially useful for the classification and prediction with small sample cases. SVM is typically used to estimate the relationship between input and output variables. Instead of solving the complex quadratic problems as in SVM, LSSVR training only obtains the solutions of a set of linear equations.

$P = \{(x_i, y_i), i = 1, 2 \ldots n\}$ is the given training set where $x_i \in R^n, y_i \in R$. The following regression model can be constructed by using non-linear mapping function $\varphi(\cdot)$

$$f(x) = \omega^{\mathrm{T}} \varphi(x) + b \quad (10)$$

where ω, b are the weight vector and the bias term. The objective optimization function of LSSVR algorithm is given as follows:

$$\begin{cases} \min \frac{1}{2}\omega^{\mathrm{T}} \cdot \omega + \frac{1}{2}\lambda \sum_{i=1}^{n} \xi_i^2 \\ s.t.\ y_i = \omega^{\mathrm{T}} \phi(x) + b + \xi_i, \quad i = 1, 2 \ldots n \end{cases} \quad (11)$$

where λ denotes penalty coefficient, ξ is the error variance. To solve this optimization problem, by introducing Lagrange multipliers α. Lagrange function is constructed as

$$L(\omega, b, \alpha, \xi) = \frac{1}{2}\omega^{\mathrm{T}} \cdot \omega + \frac{1}{2}\lambda \sum_{i=1}^{n} \xi_i^2 + \sum_{i=1}^{n} \alpha_i [\omega^{\mathrm{T}} \varphi(x) + b + \xi_i - y_i] \quad (12)$$

The solution of Eq. (7) can be obtained by partially differentiating with respect to ω, b, ξ_i and α_i

$$\begin{cases} \frac{\partial L}{\partial \omega} = 0 \rightarrow \omega = \sum_{i=1}^{n} \alpha_i \varphi(x_i) \\ \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{n} \alpha_i = 0 \\ \frac{\partial L}{\partial \xi_i} = 0 \rightarrow \alpha_i = \gamma \xi_i \\ \frac{\partial L}{\partial \alpha_i} = 0 \rightarrow \ \omega^{\mathrm{T}} \varphi(x) + b + \xi_i - y_i = 0 \end{cases} \quad (13)$$

The Eq. (10) - (13) can be rewritten as

$$\begin{bmatrix} 0 & S^{\mathrm{T}} \\ S & K + \gamma^{-1} E \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ Y \end{bmatrix} \quad (14)$$

where $y = (y_1, y_2 \ldots y_n)^{\mathrm{T}}$; $\alpha = (\alpha_1, \alpha_2 \ldots \alpha_n)^T$; $i, j = 1, 2 \ldots n$, $k\left(x_i, x_j\right)$ is the non-linear kernel function, by

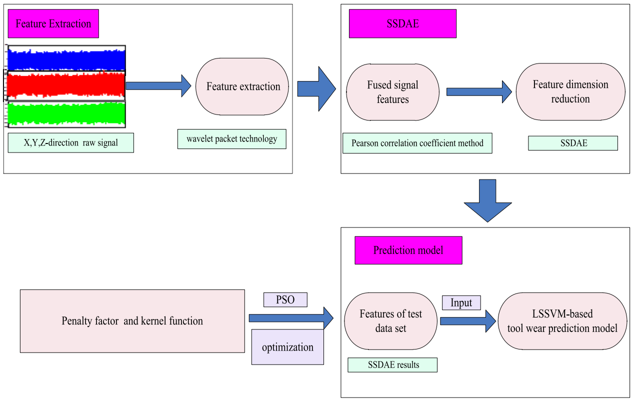

FIGURE 3. Tool wear predictive model based on SSDAE and PSO-LSSVM.

introducing kernel parameter σ, the expression is given as follow:

$$k(x_i, x_j) = \exp\left(-\frac{1}{2\sigma^2}||x_i - x_j||^2\right) \tag{15}$$

Finally, the resulting LSSVM model can be expressed as:

$$f(x) = \sum_{i=1}^{n} \alpha_i k(x_i, x_j) + b \tag{16}$$

III. TOOL WEAR PREDICTIVE MODEL BASED ON SSDAE AND PSO-LSSM

This work aims at real-timely and accurately monitoring the tool wear in machining process by utilizing the proposed SSDRE technique and the PSO-LSSVM model. It is well known that the raw monitoring signals cannot be adopted directly as the inputs of LSSVM due to the existing of noise and other components. In this paper, the original vibration signal is extracted from three directions: TD, FD, TFD; Pearson correlation coefficient method is used to determine the correlation between the characteristic quantities and the tool wear. Then, SSDAE as dimension-reduction technique have been used to remove noises and enhance the computational efficiency. Finally, the feature vectors after dimension reduction can be input into LSSVM the training data set, in the meantime,the radial basis kernel function σ and the penalty factor λ in LSSVM are optimized by PSO algorithm. The complete process for the tool wear predictive model based on SSDAN and PSO-LSSVM is illustrated in FIGURE. 3.

A. FEATURE EXTRACTION AND CORRELATION ANALYSIS

Signal features need to be extracted from the raw vibration signals, which can powerfully reflect the change of tool wear in machining process. In this work, the raw vibration signals obtain through the three-dimensional acceleration sensor, and the signal features are extracted from TD, FD, TFD three directions. Besides, the signal features that need to be extracted are selected and determined according to the wavelet packet technology.

TABLE 1. Mathematical equations of the extracted signal features.

Domain	Signal Features	Expression
TD	Mean	$\bar{x} = \sum_{i=1}^{N} x_i / N$
	RMS	$X_{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$
	STD	$X_{STD} = \sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$
FD	FC	$FC = \left[\sum_{F=1}^{N/2} fw(f)\right] \Big/ \sum_{f=1}^{N/2} w(f)$
	FV	$FV = \left[\sum_{F=1}^{N/2} (f - FC)^2 w(f)\right] \Big/ \left[\sum_{f=1}^{N/2} w(f)\right]$
	MSF	$MSF = \left[\sum_{F=1}^{N/2} f^2 w(f)\right] \Big/ \left[\sum_{f=1}^{N/2} w(f)\right]$

Comprehensively compare the three aspects of signal analysis in time domain, frequency domain and time-frequency domain, three characteristic parameters in time domain and frequency domain were selected, totaling 18 characteristic vectors. The TD, FD features as listed in Table 1 are extracted from the original vibration signals, 6 wavelet-domain features can be extracted from each vibration signal by recurring to wavelet packet decomposition. The previous 32 frequency band energy is used as the detection feature and altogether 96 signal features are obtained.

Pearson correlation coefficient method was used to calculate the correlation between feature vectors and tool wear, the formula is given as follow:

$$\rho_{xy} = \sum_n (x_n - \bar{x})(y_n - \bar{y}) \Big/ \sqrt{\sum_n (x_n - \bar{x})^2} \sqrt{\sum_n (y_n - \bar{y})^2} \tag{17}$$

where x_n, y_n are the nth of the column vectors X, Y; $\bar{x}, \bar{y}$ are the mean value of the column vectors X, Y; ρ_{xy} is the correlation coefficient, and the value range is between the interval $[-1, 1]$.

B. SSDAE

In general, dimension-reduction is a practical manner to remove noises and enhance the computational efficiency. In this work, the SSDAE technique is utilized to obtain the dimension-reduction features to weaken the negative effects of in-process noises. When the hidden layer number is greater than 1, the DAE can be called a SSDAE, along with the increase of the number of hidden layers generally can obtain better dimension reduction effect, but at the same time, the performance of ascension along with the number of iterations and the training time of extended rapidly, and thus the hidden layer with the best combination of time and precision shall be selected as the final hidden layer of SSDAE.

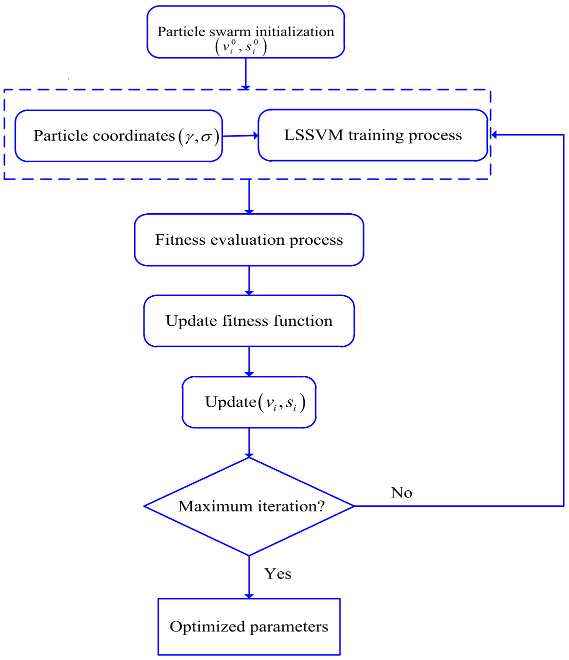

FIGURE 4. The flowchart of the PSO-LSSVM-based model.

C. PREDICTION MODEL

In this study, an LSSVM technique in combination with the PSO approach has been implemented in order to predict the milling tool flank wear values. The flowchart of the PSO-LSSVM-based model is shown in FIGURE.4. The working procedures of the proposed model are as follows:

a. Prepare training samples and test samples by SSDAE, then, initialize the parameters (the velocity and position of particle) of the PSO algorithm.
b. Take the penalty coefficient λ and kernel parameter σ of LSSVM as the two-dimensional coordinates of each particle. Set the mean relative wear error $\alpha = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y(t)-\hat{y}(t)}{y(t)}\right|$ as the fitness function of PSO algorithm, where $y(t)$ is the actual measured value, $\hat{y}(t)$ denotes the predicted wear value, N represents the total number of predicted points.
c. By comparing fitness with self-optimal solution and global optimal solution, update the best fitness.
d. According to the Eq.8, update the velocity and position of particle.
e. Set the maximum number of iterations as the end condition, the location of the optimal particle and its fitness value are preserved. If not, return to step (b) to continue the iteration
f. The particle position obtained by PSO algorithm as an optimization parameter is substituted into LSSVR algorithm, to recognition tool wear status by the model.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

This paper presents a tool wear predictive model based on SSDAE and PSO-LSSVM. The experimental platform used to verify the prediction model proposed in the previous section consists of a high-speed milling machine, the experimental setup for tool wear is shown in FIGURE.5. In this paper, data acquisition includes two parts: one is online measurement of vibration signals; the other is offline measurement of tool flank wear width.

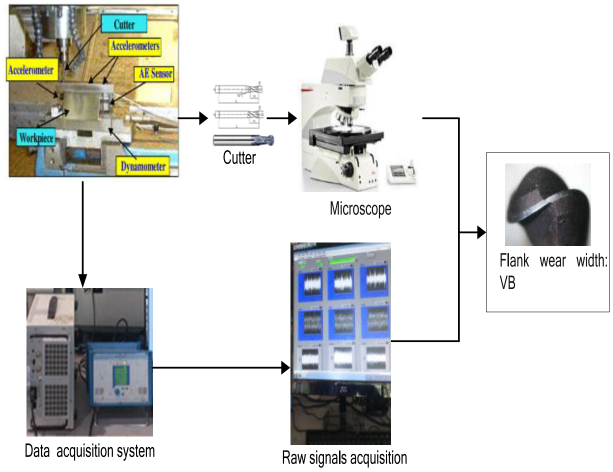

FIGURE 5. Experimental setup for tool wear.

TABLE 2. Experimental instruments and their types.

Instruments	Type
Dynamometer	Kistler 9265B
Acceleration sensor	Kistler 8636C
Charge amplifier	Kistler5019A
Microscope	LEICA MZ12
Data acquisition card(DAQ)	NI DAQ PCI 1200

A. EXPERIMENTAL INTRODUCTION

The vibration signals acquisition system that composed of acceleration sensor and DAQ boards used to collect the vibration value in three directions during the milling process. The related details of experimental instruments and their types are given in Table2. The three-dimensional acceleration sensor is installed between the work-piece and machining table to record the vibration signals; the charge amplifier and data acquisition card is used for data real-time collection and transmission; tool flank wear width that adopted to quantify the tool wear degree is measured by microscope offline.

The cutting tool material is high-speed steel, while the work-piece material is stainless steel. The vibration signals are collected at a sampling rate of 50 kHz by the DAQ board. For a new tool insert, 315 cuts are carried out with the same cutting condition and technological parameters. The machining parameters of these experiments are given with spindle speed of 10,400 RPM, Feeding speed of 1555 mm/min, cutting width of 0.125 mm, and cutting depth of 0.2 mm.

Each cutting process of ball end milling cutter is a 108mm facing milling due to the same tool path, after each cutting, the wear value of ball end milling cutter in three flutes was measured by microscope, and the average wear value was taken as the actual wear amount of ball end milling cutter. The validation data used in this article is from the PHM society

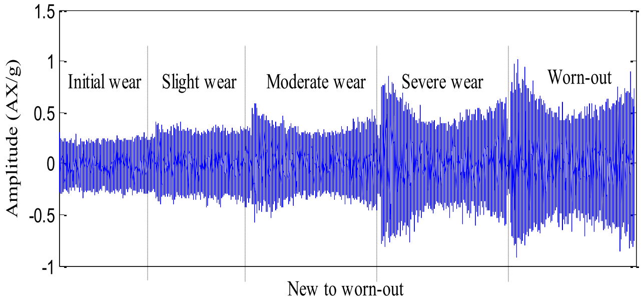

FIGURE 6. Amplitudes of the raw vibration signal for the five states.

tool 2010 life prediction challenge, and amplitudes of the raw vibration signal for the five states are shown in FIGURE.6. From the FIGURE.6, the five wear stages commonly called initial wear, slight wear, moderate wear, severe wear, and worn-out. The initial wear stage and after severe wear stage wear faster, the reason is that the initial cutter surface tissue wear and temperature rise lead to the tool wear aggravation. The normal wear stage that contains slight wear stage and moderate wear stage gently grows due to cutting force on the cutting surface is uniform and decreases.

B. EFFECTIVENESS OF MODEL

1) FEATURE EXTRACTION AND FUSION

Every 15 s, there are approximately 200,000 sampling signals that can be recorded from the vibration sensors in X, Y, Z direction. The sampling points are too hugeness to directly recognize the relationship between the raw vibration signals and the tool wear by utilizing LSSVM. Therefore, it is necessary to extract signal features from the raw vibration signals, which can powerfully reflect the change of tool wear in machining process.

According to the above mentioned, altogether 18 + 96 signal features can be extracted from time domain, frequency domain and time-frequency domain three directions. After introducing the correlation coefficient, the signal features in three domains are reduced as follows:

(1). From the TD and FD: the correlation coefficient is greater than 0.95. Finally, 6 feature vectors in the time domain are extracted.

(2). From the TFD: the correlation coefficient is greater than 0.97. Finally, 10 feature vectors in the time domain are extracted.

At last, altogether 16 signal features can be extracted and are adopted as the pre-fusion features of the SSDAE-PSO-LSSVM. The tool wear curves of six feature vectors (X, Y, Z Standard deviation (STD) and X, Y, Z Root mean square (RMS)) in the time domain are drawn in FIGURE.7. It can be seen from the figure that there is a strong positive correlation between the feature vector and the wear value due to the preliminary screening by Pearson correlation coefficient method.

2) DIMENSION REDUCTION TECHNIQUES

In this paper, 16 feature vectors were screened out from 96 initial feature vectors, each of which can be used as the input training sample. The 16 feature vectors are normalized and integrated into a matrix of 16*315 dimensions. Set the parameters of sparse to 0.05 and the iteration times as 1000, etc. SSDAE with different number of hidden layers are set up and trained, the best time and precision of which are selected as the final number of hidden layers of SSDAE.

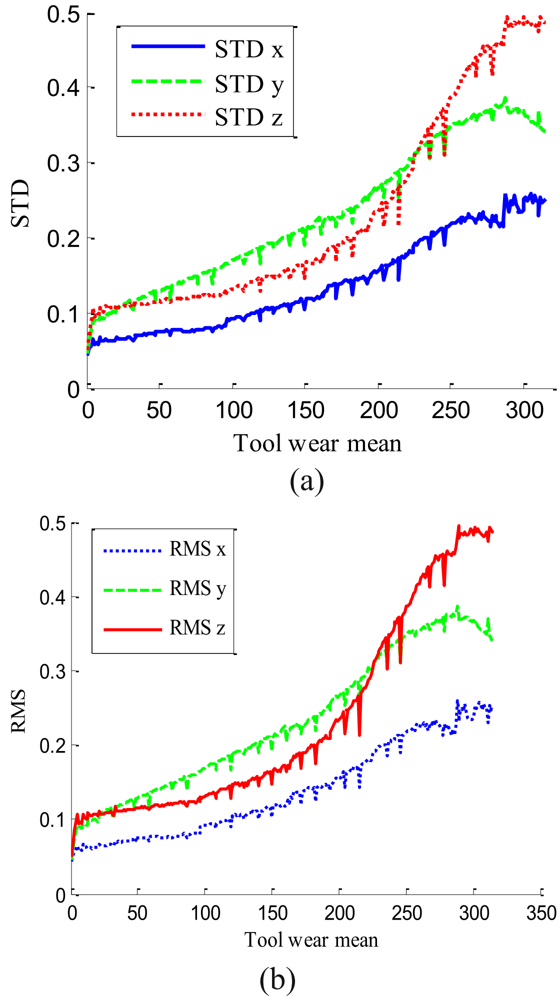

FIGURE 7. (a) X, Y, Z Standard deviation and (b) X, Y, Z Root mean square.

SSDAE is a dimension-decrement technique and can realize the reduction of feature dimension. The 16 feature vectors screened are merged into a matrix of 16*315, and the dimensionality is finally reduced to 1*315. On the one hand, the dimension-decrement vector is guaranteed to be consistent with the initial feature vector dimension, which is more comparable as the input of the subsequent prediction model. On the other hand, dimension-decrement can effectively avoid the requirement of prior knowledge in the feature selection process, and greatly improve the efficiency of feature modeling. The dimension-decrement interval is within a moderate range; thus, choose 2-4 layers for the hidden layer.

In order to remove noises and enhance the computational efficiency, this section test mean relative error (MRE) and mean absolute error (MAE) of SSDAE and the training time when the number of hidden layers of SSDAE is 2-4. The

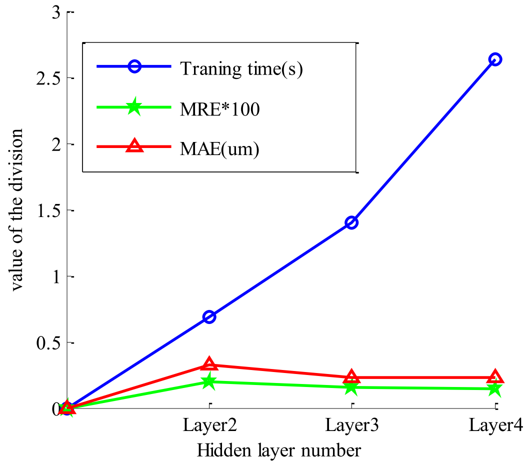

FIGURE 8. Different hidden layer training process and effect.

training process and effect as shown in FIGURE.8, when the third-layer network structure is selected, the training time rank second. At the same time, it can be found that with the increase of training layers, the training time of SSDAE will increase significantly. The MRE ratio of the predictive indicators is optimized by 25% in layer 2 and is very close to the minimum error of layer 4, the MAE of the third-layer network structure is the smallest among the three hidden layers. Considering the model accuracy and training time, SSDAE with 3 hidden layers network structure is selected in this paper.

3) PSO AND LS-SVM

As it has been previously pointed out, in order to optimize the penalty coefficient λ and kernel parameter σ of LSSVM, the PSO algorithm was proposed. After normalizing the extracted feature vectors, the method in section II.C is used for parameter optimization. Initialize the parameters of the PSO algorithm as follow: number of particles $i = 30$; learning factor $c_1 = 2, c_2 = 2$; maximum number of iterations $k_{max} = 200$.

Finally, predicted results of the constructed LSSVM-based tool wear recognition model for milling process by using the fused features of SSDAE as shown in FIGURE.9, the hyper parameters are optimized previously by using the PSO technique. The FIGURE.9a shows the results of PSO optimize LSSVM classification model, and FIGURE.9b shows the recognition effect of the proposed model, where sampling classifications are as follow: 1, early tool wear (initial wear); 2, mid-term tool wear (slight wear and moderate wear); 3, late-term tool wear (severe wear and worn-out). The best penalty coefficient $\lambda = 28.4031$ and the best kernel parameter $\sigma = 0.1$ .The predicted tool wear of LSSVM well coincide with the actual results, and the total tool wear recognition rate is up to 97.2366%. The training dataset and test dataset without SSDAE dimension reduction is adopted to test and verify the predictive performance of LSSVM and PSO-LSSVM. The total tool wear recognition rate is 89.2234% and 91.3461%, which is less effective than the

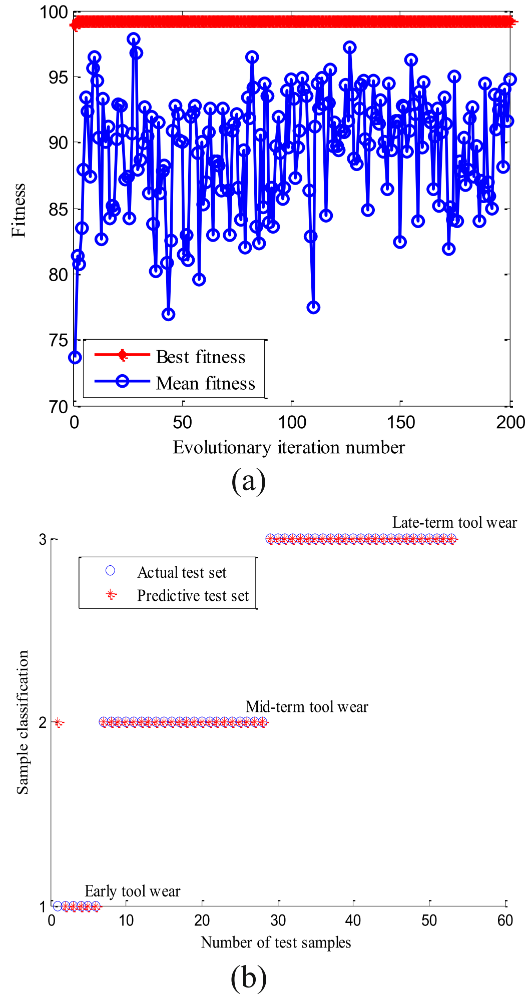

FIGURE 9. (a) The results of PSO optimize LSSVM and (b) PSO-LSVM predicting results.

proposed model. Therefore, the effectiveness of SSDAE and PSO technique in improved the performance of LSSVM is intuitively reflected.

C. COMPARISON WITH OTHER METHODS

Predicted results of the constructed LSSVM-based tool wear predictive model for milling process by using the fused features of SSDAE as shown in FIGURE.9, the hyper parameters were optimized previously by using the PSO technique. The FIGURE.10a shows the comparison between the actual tool wear in micrometers and the tool wear predicted using LSSVM-based model, and FIGURE.10b shows the relative error of the predicted model. The predicted tool wear of LSSVR well coincide with the actual results, and the average relative error is only reached 0.0021, and the average absolute error is 0.3745 um. Therefore, the effectiveness of SSDAE and PSO technique in improving the performance of LSSVM is intuitively reflected.

To further show the superiority of LSSVM, the traditional methods such as partial least squares regression (PLSR), Back Propagation Neural Network (BPNN) and Extreme

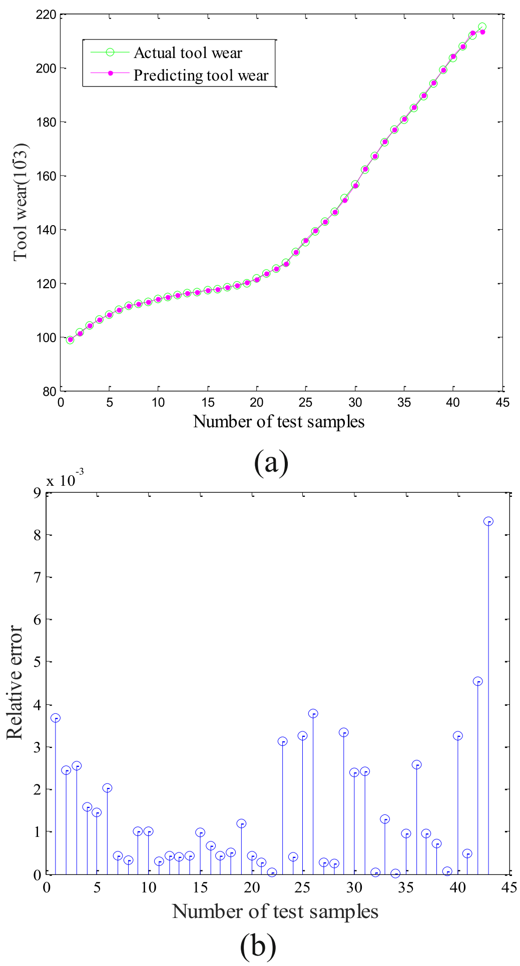

FIGURE 10. (a) The results of the predicted model and (b) The relative error of the predicted model.

TABLE 3. Experimental results of different models for tool prediction.

Methods	MRE	MAE (um)
PLSR	0.0312	4.1298
BPNN	0.0108	1.7655
ELM	0.0293	3.8917
LSSVM	0.0021	0.3745

Learning Machine (ELM) are also utilized to realize tool wear prediction. The same training dataset and test dataset in Section IV.B.1 is adopted to test and verify the predictive performance of PLSR, BPNN and ELM. The number of PLSR-constituent is set to 30. In the back propagation neural network (BPNN), the number of hidden layer neurons was set as 7, the number of iterations was set as 100, and the learning rate was set as 0.1. The hidden layer neuron of ELM was set as 20, and the activation function was set as sigmoid function.

The experimental results of different models are given out in Table 3. BPNN performs better than PLSR and ELM in terms of MRE/MAE, however, far below the proposed PSO-LSSVM model. Therefore, the prediction accuracy of the three shallow learning models are far lower than that of the proposed model in this paper, which once again indicates the effectiveness of the SSDAE-PSO-LSSVM model.

V. CONCLUSION

Based on the experimental and predicted results, the main contributions of this research work can be summarized as follows:

1) A high correlation coefficient greater than 0.95 is used to extract feature vector from time domain, frequency domain and time-frequency domain three directions.
2) The SSDAE technique is utilized to realize multi-feature signal dimension reduction with the aim of improving the prediction accuracy, which reduces the dependence on the prior knowledge of feature selection and greatly improves the modeling efficiency.
3) PSO technique is helpful for adaptive optimization of kernel parameters, which greatly improve computing power and LSSVM model prediction accuracy.
4) The proposed SSDAE-PSO-LSSVM model performs better than PLSR, BPNN and ELM in terms of prediction accuracy. The total tool wear recognition rate of LSSVM and PSO-LSSVM is less effective than the proposed model.

In summary, a new tool wear predictive model based on the SSDAE technique and the PSO-LSSVM model is presented in this paper. Considering the urgent demand for tool online condition monitoring. In subsequent studies, we will conduct experiments to compare the performance of different types of deep learning networks and develop practical applications of deep learning. However, optimization and improvement of the deep learning framework to suit a particular application, including efficient error feedback and heuristic search mechanism, are not trivial tasks.

REFERENCES

[1] R. K. Mobley, *An Introduction to Predictive Maintenance*. New York, NY, USA: VanNostrand Reinhold, 1990.
[2] S. Kurada and C. Bradley, "A review of machine vision sensors for tool condition monitoring," *Comput. Ind.*, vol. 34, no. 1, pp. 55–72, Oct. 1997.
[3] C. Shi, G. Panoutsos, B. Luo, H. 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.
[4] C.-L. Yen, M.-C. Lu, and J.-L. Chen, "Applying the self-organization feature map (SOM) algorithm to AE-based tool wear monitoring in micro-cutting," *Mech. Syst. Signal Process.*, vol. 34, nos. 1–2, pp. 353–366, Jan. 2013.
[5] P. Karali, "Acoustic emission-based tool condition monitoring system in drilling," in *Proc. World Congr. Eng.*, London, U.K., 2011, vol. 3, no. 1, pp. 2126–2130.
[6] H. M. Ertunc, K. A. Loparo, and H. Ocak, "Tool wear condition monitoring in drilling operations using hidden Markov models (HMMs)," *Int. J. Mach. Tools Manuf.*, vol. 41, no. 9, pp. 1363–1384, Jul. 2001.
[7] T. Benkedjouh, K. Medjaher, N. Zerhouni, and S. Rechak, "Health assessment and life prediction of cutting tools based on support vector regression," *J. Intell. Manuf.*, vol. 26, no. 2, pp. 213–223, Apr. 2015.
[8] 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.
[9] C. Zhang and H. Zhang, "Modelling and prediction of tool wear using LS-SVM in milling operation," *Int. J. Comput. Integr. Manuf.*, vol. 29, no. 1, pp. 76–91, 2016.
[10] S. MM, K. Chandra, and K. EB, "Maximum correntropy based dictionary learning framework for physical activity recognition using wearable sensors," in *Proc. Int. Symp. Visual Comput.*, vol. 10073, 2016, pp. 123–132.

[11] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006.

[12] D. Cabrera, F. Sancho, C. Li, M. Cerrada, R.-V. Sánchez, F. Pacheco, and J. V. de Oliveira, "Automatic feature extraction of time-series applied to fault severity assessment of helical gearbox in stationary and non-stationary speed operation," *Appl. Soft Comput.*, vol. 58, pp. 53–64, Sep. 2017.

[13] S. M. Mathews. (2017). *Dictionary and Deep Learning Algorithms With Applications to Remote Health Monitoring Systems*. [Online]. Available: http://udspace.udel.edu/handle/19716/21241

[14] S. M. Mathews. (2015). *Leveraging a Discriminative Dictionary Learning Algorithm for single-Lead ECG Classification*. [Online]. Available: http://udspace.udel.edu/handle/19716/17337

[15] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.

[16] G. Liu, H. Bao, and B. Han, "A stacked autoencoder-based deep neural network for achieving gearbox fault diagnosis," *Math. Probl. Eng.*, vol. 201, 8pp. 85–117, Jul. 2018.

[17] G. Jiang, H. He, P. Xie, and Y. Tang, "Stacked multilevel-denoising autoencoders: A new representation learning approach for wind turbine gearbox fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 9, pp. 2391–2402, Sep. 2017.

[18] S. Lee, "Tool condition monitoring system in turning operation utilizing wavelet signal processing and multi-learning ANNs algorithm methodology," *Int. J. Eng. Res. Innov.*, vol. 138, no. 5, pp. 37–49, 2015.

[19] D. M. D'Addona, A. M. M. S. Ullah, and D. Matarazzo, "Tool-wear prediction and pattern-recognition using artificial neural network and DNA-based computing," *J. Intell. Manuf.*, vol. 28, no. 6, pp. 1285–1301, Aug. 2017.

[20] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," *Mech. Syst. Signal Process.*, vol. 100, pp. 439–453, Feb. 2018.

[21] W. Fuan, J. Hongkai, S. Haidong, D. Wenjing, and W. Shuaipeng, "An adaptive deep convolutional neural network for rolling bearing fault diagnosis," *Meas. Sci. Technol.*, vol. 28, no. 9, Sep. 2017, Art. no. 095005.

[22] G. Sannino and G. De Pietro, "A deep learning approach for ECG-based heartbeat classification for arrhythmia detection," *Future Gener. Comput. Syst.*, vol. 86, pp. 446–455, Sep. 2018.

[23] S. M. Mathews, C. Kambhamettu, and K. E. Barner, "A novel application of deep learning for single-lead ECG classification," *Comput. Biol. Med.*, vol. 99, pp. 53–62, Aug. 2018.

[24] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P. A. Manzagol, L. and Bottou, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, pp. 3371–3408, Dec. 2010.

[25] K. Song, M. Wang, L. Liu, C. Wang, T. Zan, and B. Yang, "Intelligent recognition of milling cutter wear state with cutting parameter independence based on deep learning of spindle current clutter signal," *Int. J. Adv. Manuf. Technol.*, vol. 109, nos. 3–4, pp. 929–942, Jul. 2020.

[26] Y. Chen, Y. Jin, and G. Jiri, "Predicting tool wear with multi-sensor data using deep belief networks," *Int. J. Adv. Manuf. Technol.*, vol. 99, nos. 5–8, pp. 1917–1926, Nov. 2018.

YANG XIE received the M.S. degree in mechanical engineering from the Wuhan University of Science and Technology, Wuhan, China, in 2013. He is currently pursuing the Ph.D. degree in mechanical engineering with the Huazhong University of Science and Technology, Wuhan, China. His research interests include signal analysis, signal processing, tool wear prediction, tool condition monitoring, and sustainable manufacturing.

CHAOYONG ZHANG received the B.S. degree in mechanical engineering from the Tianjin University of Science and Technology, Tianjin, China, in 1993, the M.S. degree in mechatronics engineering from the University of Science and Technology Beijing, Beijing, China, in 1999, and the Ph.D. degree in mechatronics engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2007. He is currently a Professor with the School of Mechanical Science and Engineering, Huazhong University of Science and Technology. His research interests include modeling, optimization, scheduling for production manufacturing systems, effective resource utilization of manufacturing systems, sustainable manufacturing including clean and highly efficient manufacturing processes, and the power consumption model of machine tools. He has authored two books and authored or coauthored over 70 academic papers. He has received the 2008 Ministry of Education Science and Technology Progress First Place, the 2010 China Machinery Industry Federation First Place, and the 2013 Ministry of Education Natural Science First Place.

QIONG LIU is currently a Professor and a Doctoral Supervisor with the Department of Industrial and Manufacturing Systems Engineering, School of Mechanical Science and Engineering, Huazhong University of Science and Technology. She is also a member of the State Key Laboratory of Digital Manufacturing Equipment and Technology and the National Engineering Research Center of Digital Manufacturing Equipment. From 2014 to 2015, she was a Senior Visiting Scholar with Purdue University, USA. From 2004 to 2005, she was engaged in research work at the University of Cambridge in Britain.

Professor Liu's research focuses on low carbon manufacturing, manufacturing system optimization, and reverse logistics. She has been committed to promoting low carbon emissions in manufacturing systems through integrated optimization of production scheduling, process, cutting parameters, material handling, and reverse logistics. A more accurate calculation method of carbon footprint of mixed flow manufacturing system products is studied. An extended MES system was developed to obtain data related to carbon emissions in manufacturing systems. She has published papers on production scheduling, implementation layout, order-to-transport issues, project scheduling, reverse logistics network design, and resource allocation.