Sensors and Materials, Vol. 35, No. 7 (2023) 2321–2335
MYU Tokyo

S & M 3325

Tool Wear Prediction Based on Attention Long Short-term Memory Network with Small Samples

Weiwei Yu, Hua Huang,* Runlan Guo, and Pengqiang Yang

School of Electrical and Mechanical Engineering, Lanzhou University of Technology,
36 Pengjiaping road, Qilihe, Lanzhou 730050, China

(Received May 11, 2023; accepted June 20, 2023)

Keywords: attention long short-term memory network, data augmentation, state recognition, k-nearest neighbor classifier, tool wear prediction

In tool wear monitoring, the environment for signal collection is always complex, which leads to insufficient signal state samples and unbalanced category labels. Moreover, the hidden state features extracted by neural networks in conventional methods are mixed together, resulting in the low prediction accuracy of tool wear. Therefore, a tool wear prediction method based on an attention long short-term memory (LSTM) network with data imbalance is proposed. First, a generative adversarial network (GAN) is used to improve the imbalance of state category labels and expand data samples. Then, an extended data sample is used as the input of the stacked sparse autoencoder network (SSAE) to adaptively extract features, and the k-nearest neighbor classifier is used to identify the different stages of tool wear. Finally, on the basis of the state identification results, the time series features with different tool expansion data samples are extracted and input into the attention LSTM network to map the tool wear values for different tool wear processes. The experimental results show that the proposed method can improve the imbalance of category labels, increase the selection of more informative components in sequence data, and obtain excellent prediction accuracy and generalization.

1. Introduction

As the key component of computerized numerical control machine tools, cutting tools play a crucial role in the processing quality and accuracy of machined parts. According to statistics, 20% of the total downtime of machine tools is caused by tool failure.(1) Therefore, to improve tool utilization and reduce the failure rate during the machining process, a reliable tool wear status monitoring method is necessary.

Currently, there are two popular methods of monitoring tool wear: direct and indirect.(2) The direct method evaluates tool wear by directly measuring the change in physical properties such as the tool's volume and mass. It is possible to determine the state of tool wear, but the direct monitoring method is susceptible to machining conditions (e.g., light, cutting fluid, and material debris) and requires downtime for inspection. Moreover, the optical measurement equipment is

*Corresponding author: e-mail: huanghua@lut.edu.cn
https://doi.org/10.18494/SAM4509

ISSN 0914-4935

always expensive and complex to install. On the other hand, the indirect monitoring method only needs one or more sensors to collect the cutting force, vibration, acoustic emission, current, and other signals during the cutting process of the tool, which has become the mainstream method of tool wear monitoring. In addition, the development of machine learning offers a wider scope for the intelligent monitoring of tool wear. Wu *et al.* used a data-driven random forest network to predict tool wear values and evaluated the algorithm's performance using mean square error, variance, and training time on a dataset of 315 milling experiments.[3] Meng *et al.* collected multisensor signals and established a gravitational search algorithm back propagation neural network tool wear monitoring model to improve the prediction accuracy of tool wear.[4] Wang *et al.* proposed a tool-remaining-life prediction method with multichannel signal fusion and Bayesian updating, which solved the problems of high fluctuation of single sensor prediction accuracy and low reliability.[5] Deep learning has become an indispensable "medium" for tool monitoring research in recent years because of its end-to-end learning characteristics and the ability to adaptively mine deep data features. Huang *et al*. constructed an adaptive multidomain feature fusion tool online monitoring model based on deep convolutional neural network, which solved the problems of the low efficiency and accuracy of manual feature fusion monitoring.[6] Ou *et al.* combined online sequential extreme learning machine and stacked denoising autoencoder networks to develop a tool wear state identification model and proved the effectiveness of the developed model using spindle current.[7] Wang and Zhou used a stacked autoencoder to extract features and used long short-term memory network to solve time series problems under different working conditions, realizing the prediction of tool wear.[8] These tool wear modeling approaches have improved tool wear prediction and identification accuracy considerably, but the following problems still remain.

(1) The tool wear signal acquisition environment is usually characterized by complex working conditions, which result in limited signal state samples and insufficient number of samples. In the case of high dimension and loss ratio, the tool category label is unbalanced, resulting in the poor recognition and prediction of tool wear state, so that the tool cannot be fully utilized. In view of this, the modeling methods need to be improved.
(2) Tool wear is predicted according to the original signal characteristics and its nonlinear mapping relationship, so the features are mixed together. While the features closely related to tool wear are not strengthened, the characteristics less related to tool wear are not weakened, resulting in a considerable difference between the predicted and actual tool wear values.

To address these problems, a tool wear prediction method based on the attention long short-term memory (LSTM) network with data imbalance is proposed in this study, which can expand tool wear data samples, balance the category label of wear stage, and extract the important time series information of the wear process. First, a generative adversarial network (GAN) is used to expand the data samples, which solves the problem of insufficient data samples and unbalanced category labels due to the complex processing conditions. Through the "zero sum game" of the generator and discriminator, a data signal highly similar to the original cutting signal is generated. Second, the extended data is divided into training set and test set according to the tool wear stage, and the divided data set is input into the stacked sparse autoencoder network (SSAE) network for adaptive feature extraction. Then, the extracted deep data features are used to

identify the state using the k-nearest neighbor classifier. Finally, on the basis of the results of state recognition and accuracy, we extract feature vector sequences from the enhanced samples and input them into the attention LSTM network to achieve accurate prediction of tool wear.

2. Methods

In this study, a tool wear prediction model is established with data imbalance by combining the GAN and attention LSTM network. The model framework is shown in Fig. 1. The model is divided into three stages and includes the following three steps:

(1) Data augmentation, the GAN is used to enhance the cutting force signal samples of tool wear. A limited sample of tool wear cutting force data is extracted, and the extracted data is fed into the GAN to generate new sample data, then the generated data is combined with the real data to form the required sample data.

(2) State recognition using the SSAE network to identify tool wear state.
The data generated by the GAN is fed into the SSAE network to mine the data features. Afterwards, the feature vectors with high characterization capability are selected, and dimensionality reduction is performed using the SSAE network.[9] In addition, the features of dimensionality-reduced data are fed into the k-nearest neighbor classifier to achieve the accurate identification of tool wear states with unbalanced data.

(3) Tool wear prediction using the attention LSTM network to predict tool wear.

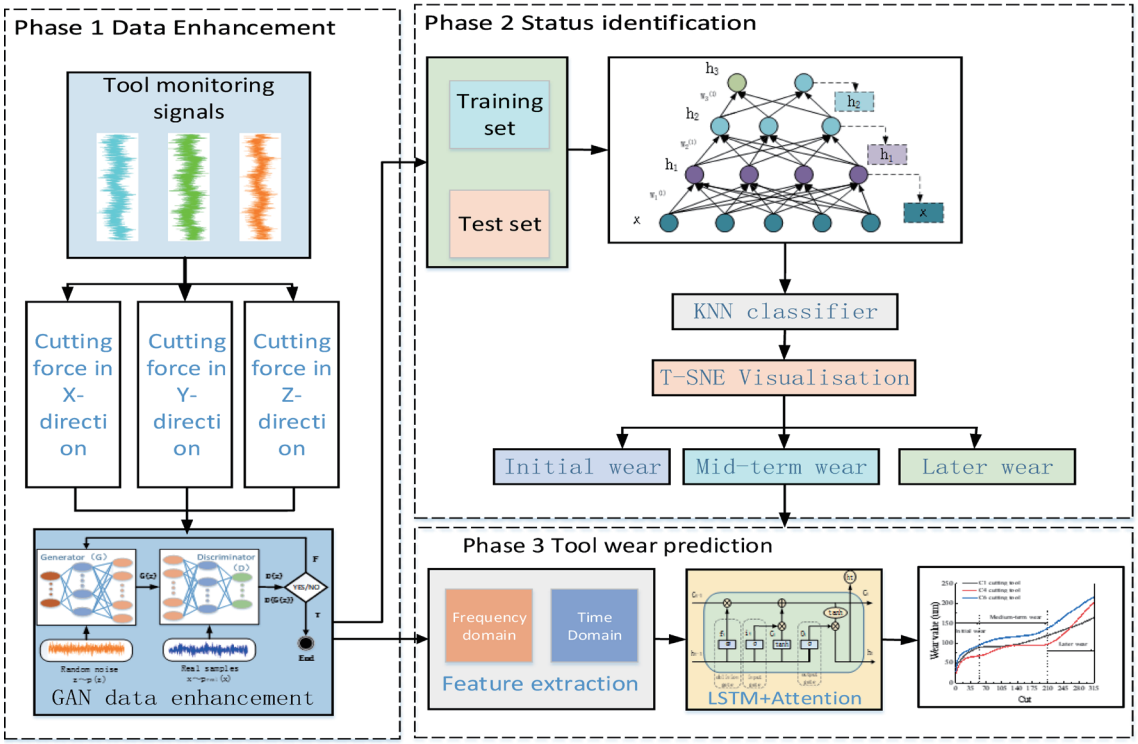

Fig. 1. (Color online) Framework of tool wear prediction model for data enhancement attention mechanism.

According to the results of state recognition, the features (time domain and frequency domain features) of data samples expanded by the GAN are extracted, and the feature vector sequence is processed by feature normalization technology. Then, the extracted features are input into the attention LSTM network to map the relationship between the signal and the tool wear value. Finally, the tool's wear values in different cutting processes are obtained.

3. Theory

3.1 Generative adversarial network

The GAN is an unsupervised learning paradigm that combines game theory ideas with machine learning to train the distribution of data feature samples through effective confrontation between network models, which in turn results in data features as close to the real ones as possible.[10] The GAN structure is shown in Fig. 2.

The GAN consists of a discriminator (D) and a generator (G), forming a dual network structure. Random noise z is fed into G, and by learning from real samples, the input noise is mapped to generative samples similar to the real samples by multilayer upsampling. The input of D is either the true sample x or the generative sample $G(z)$, and after multiple layers of feature extraction, the output probability $D(x)$ represents the magnitude of the probability that the input sample belongs to the true sample. During the training process, G and D are trained alternately. Through the adversarial learning between them, D cannot distinguish the source of the data. The objective function is defined as

$$\min_G \max_D V(G,D) = E_{z-P_{z(z)}}[\log(1-D(G(z)))] + E_{x-P_{r(x)}}[\log(D(X))], \tag{1}$$

where $P_{r(x)}$ and $P_{z(z)}$ are the true sample distribution and Gaussian distribution, respectively.

At each iteration of the training process, G is first fixed, and D is trained to distinguish the

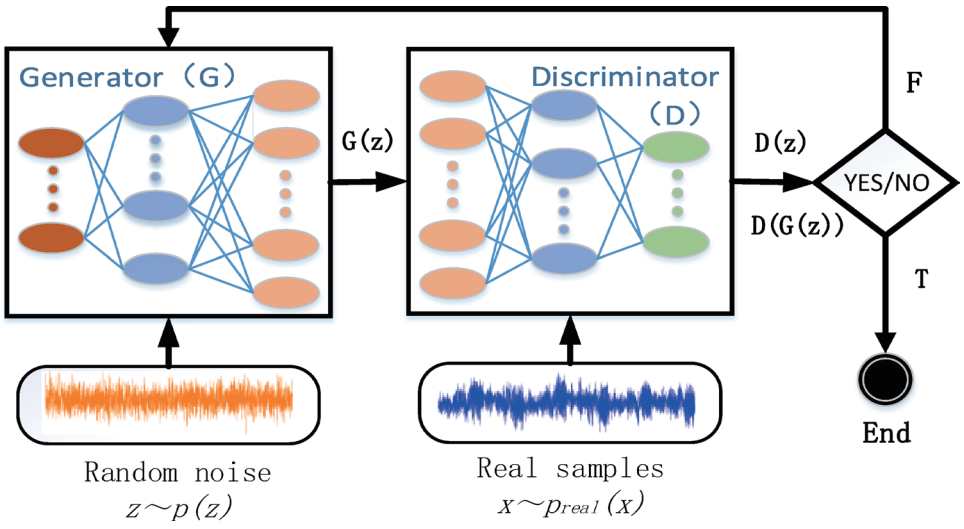

Fig. 2. (Color online) GAN structure.

source of the input data, thereby obtaining a binary classifier. Then, D is fixed, and G is trained so that the generated data is close to the real data. The alternating iteration maximizes the error of the other side, and G can estimate the distribution of real data through confrontation to make up for the absence of real data.(11)

3.2 Long short-term memory network

The acceleration signal data used in this study is a type of time series, which reflects the changing trend of tool wear with processing time. Therefore, it is suitable to use LSTM for time series prediction.(12) LSTM is derived from a recurrent neural network, which has more advantages in dealing with long time series problems than a general recurrent neural network. LSTM introduces the addition operation into the network through gate control, which solves the gradient problem and short-term memory problem of the recurrent neural network to a certain extent. The data flow and transmission process are shown in Fig. 3.

On the basis of input x_t at the current moment and the output h_{tl} at the previous moment, the self-circulating weights in the forgetting gate control unit are used to control the update of the memory unit:

$$f_t = \sigma(W_f^T \cdot [h_{t-1}, x_t] + b_f). \tag{2}$$

The weight of input gate control unit information inflow is as follows:

$$i_t = \sigma(W_i^T \cdot [h_{t-1}, x_t] + b_i). \tag{3}$$

On the basis of oblivion gate and the input gate, the internal state of the unit is updated as

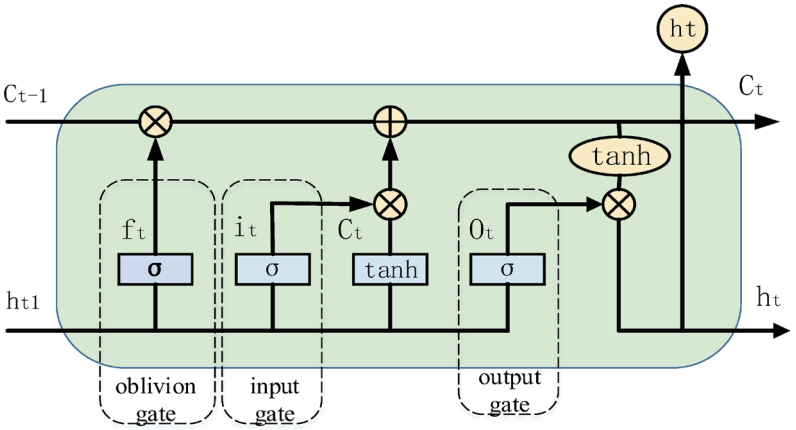

Fig. 3. (Color online) Long short-term memory network structure.

$$C_t = \tanh(W_C^T \cdot [h_{t-}, x_t] + b_C), \tag{4}$$

$$c_t = f_t \times C_{t-1} + i_t \times C_t. \tag{5}$$

The output gate control unit output weight is

$$O_t = \sigma(W_o^T \cdot [h_{t-}, x_t] + b_o) \cdot \tag{6}$$

From the output gate and internal state of the unit, the final output of the LSTM unit is as

$$h_t = O_t \times \tanh(c_t), \tag{7}$$

where the σ is the activation function that maps the value of the gating unit to between 0 and 1. b_f, b_i, b_o, and b_C are the bias terms, and $W_f{}^T$, W_i^T, $W_o{}^T$, and $W_C{}^T$ denote the weights.

3.3 Attention mechanism

The attention mechanism is an indispensable part of deep learning methods. It can not only automatically focus on the relevant information of network input features but also adaptively combine high-dimensional features. Moreover, it can select the most suitable features for the current task. This can reduce the difficulty of network hyperparameter selection, improve the adaptability of the network, and save on computing resources.

There are generally two ways to implement the attention mechanism in the LSTM network. The first is to add an attention mechanism before the LSTM network layer. The input sample passes through the attention mechanism layer before entering the LSTM network layer. This implementation essentially adds a feature attention layer, which makes the model pay more attention to the input features that have a significant impact on tool wear and give them higher weights, thereby ensuring that the wear value prediction is more accurate. The second implementation is to add an attention mechanism after the LSTM network layer. The hidden state output after the LSTM network layer does not directly enter the output layer but first enters the attention mechanism layer. After the LSTM network layer, each time step of the time series generates a hidden state. The degradation information learned by LSTM is included in these hidden states, and then different weights are assigned to each hidden state to ensure that the model learns more information related to degradation. Therefore, in this study, we add an attention mechanism to the output layer.[13]

4. Experimental Setup

To prove the effectiveness of the tool wear monitoring model based on the attention LSTM network with conditions of data imbalance, the high-speed milling tool wear data set shown in PHM 2010 is used in this study. The experiment is carried out on a high-speed computer numerical control (CNC) milling machine, the cutting tool is a three-flute ball nose tungsten carbide cutter, and the workpiece is an HRC 52 stainless steel plate.[14] The experimental platform is shown in Fig. 4.

The three-component dynamometer Kistler 81552 is installed on the experimental platform to collect the cutting force signals in three directions during the milling process, the Kistler 8636C piezoelectric accelerometers collect the vibration signals in three directions during the milling process, and the Kistler 9265B acoustic emission sensor collects the acoustic emission signals during the milling process. A total of seven-dimensional channel signals are formed. The experimental equipment and experimental processing parameters are shown in Tables 1 and 2, respectively.

The output signal of the sensor is amplified by a Kistler charge amplifier and collected by NI DAQ PCI 1200. In the experimental machining, the feed rate of each tool is 0.001 mm. The tool

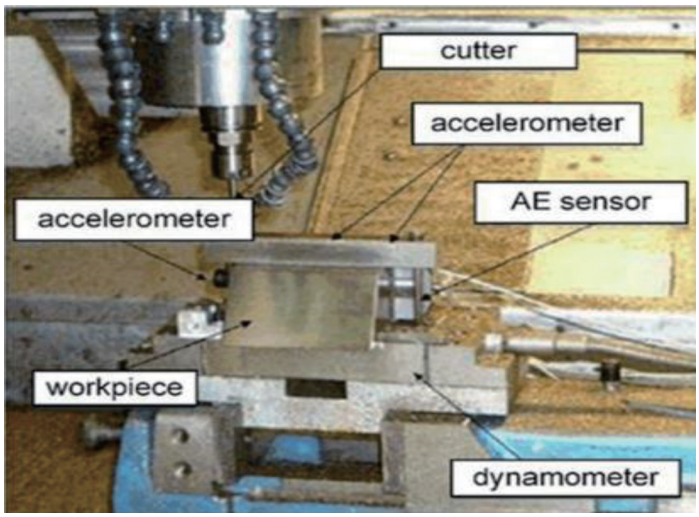

Fig. 4. (Color online) Experimental table.

Table 1
Experimental equipment.

Vertical CNC machines	Force sensors	Vibration sensors	Acoustic emission sensors	Amplifiers	Digital mining equipment	Wear measurement equipment	Milling tools
Roders Tech RFM760	Kistler three-component dynamometers	Kistler three-axis acceleration sensors	Kistler acoustic emission sensors	Kistler charge amplifier	NI DAQ PCI 1200	LEICA MZ12	Ball end carbide milling cutters

Table 2
Experimental processing parameters.

Spindle speed (r/min)	Feeding speed (mm/min)	Depth of cut in *Y* direction (mm)	*Z*-directional depth of cut (mm)	Milling method	Cooling method	Workpiece material
10400	1555	0.125	0.2	Smooth milling	Dry cutting	Stainless steel HRC52

backface wear was measured offline using a LEICA MZ12 microscope after the tool had completed 108 mm of face milling in the *X*-direction. Each tool measurement yielded 315 tool wear values, each corresponding to a raw signal of the (*n*,7) tensor. To predict the tool wear value more accurately, the data of the first 6000 sampling points of the unbalanced cutting force signal after each cutting was expanded. The expanded data samples were extracted according to the correlation coefficient method. The calculation formula is

$$\rho(A,B)=\frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{A_i-\mu_A}{\sigma_A}\right)\left(\frac{B_i-\mu_B}{\sigma_B}\right), \tag{8}$$

where ρ is the degree of correlation between A and B, with a value between −1 and 1.

With $|\rho_{xy}| \geq 0.95$ as the screening standard, 20 qualified characteristic parameters were preliminarily screened out of all the characteristic parameters.[15] There are 14 qualified characteristic parameters in the time domain and six qualified characteristic parameters in the frequency domain. A 60-dimensional feature can be obtained in three directions and used as the initial feature data set. The specific characteristic indicators are shown in Table 3.

5. Results and Discussion

5.1 Status identification

The accurate identification of tool wear state is an indispensable part of tool wear prediction. First, the quality of extended data can be judged through state recognition, and second, the accurate identification of tool wear status directly affects the prediction of tool wear values.

According to the characteristics of tool wear, the process of tool wear is divided into three stages: initial wear, medium wear, and later wear. The three wear stages are represented by labels “1”, “2”, and “3”, respectively. In this study, according to the actual wear process of the tool, the 1–50 tool passes are categorized into the initial wear phase, the 51–200 tool passes are categorized into the medium wear phase, and the 201–315 tool passes are categorized into the later wear phase.[16] Finally, the wear samples for each tool are categorized as shown in Table 4.

5.1.1 Data enhancement

Tool monitoring generally uses the raw signals collected by sensors as input to the network features to identify and predict the tool wear status. Almost all deep learning networks require a

Table 3
Characteristic indicators.

Serial number	Feature name	Expression	Serial number	Feature name	Expression
1	Peak to peak	$A=\left\|\max\left(x_i\right)\right\|-\left\|\min\left(x_i\right)\right\|$	11	Peak index	$C=\frac{A}{X_R}$
2	Variance	$s^2=\frac{1}{N}\sum_{i=1}^{N}\left(x_i-\overline{x}\right)^2$	12	Pulse index	$\mathrm{X}_{P_i}=\frac{\max\left(x_i\right)}{\left\|X_A\right\|}$
3	Mean value	$\overline{x}=\frac{1}{N}\sum_{i=1}^{N}x_i$	13	Tolerance index	$\mathrm{X}_{M_i}=\frac{\max\left(x_i\right)}{R}$
4	Skewness	$C_S=\frac{\sum_{i=1}^{N}x_i^3}{N}$	14	Kurtosis index	$X_K=\frac{K}{X_R{}^4}$
5	Kurtosis	$K=\frac{1}{N}\sum_{i=1}^{N}\frac{\left(x_i-\overline{x}\right)^4}{S^4}$	15	Mean frequency	$f_{\overline{x}}=\frac{1}{k}\sum_{k=1}^{k}x_k$
6	Mean square value	$X_{RMS}=\frac{1}{N}\sum_{i=1}^{N}x_i{}^2$	16	Frequency second-order distance	$f_{S_m}=\frac{\sum_{k=1}^{k}f\left(k\right)^2\times x\left(k\right)}{\sqrt{\sum_{k=1}^{N}f\left(k\right)}}$
7	Square root amplitude	$R=(\sum_{i=1}^{N}\frac{\sqrt{\left\|x_i\right\|}}{N})^2$	17	Standard deviation frequency	$f_{S_d}=\frac{\sum_{k=1}^{k}(x_k-f_{\overline{x}}{}^2)}{k-1}$
8	Root mean square value	$X_R=\sqrt{\frac{1}{N}\sum_{i=1}^{N}x_i{}^2}$	18	Kurtosis frequency	$f_{S_b}=\frac{1}{k}\sum_{k=1}^{k}\frac{\left(x_k-f_{\overline{x}}\right)}{f_{RMS}{}^4}$
9	Absolute mean	$X_A=\sum_{i=1}^{N}\frac{\left\|x_i\right\|}{N}$	19	Root mean square frequency	$f_{RMS}=\sqrt{\frac{\sum_{k=1}^{k}x_k{}^2}{k}}$
10	Waveform index	$W=\frac{X_R}{X_A}$	20	Center frequency	$f_c=\frac{\sum_{k=1}^{k}f_k\times x_k}{\sum_{k=1}^{k=k}x_k}$

Table 4
Classification of wear stages.

Wear phase	Initial wear	Medium wear	Later wear
Number of tool walks	1–50	51–200	201–315
Wear and tear labels	1	2	3
Feature size	50 × 60	150 × 60	115 × 60

large amount of data for training; however, if the network training sample data is insufficient, the model training will be over-fitted, reducing the training accuracy of the model. In the practical machining process, the environment for signal collection is always complex, and the whole process monitoring signal of tool wear is difficult to collect, which leads to most of the

tool monitoring models being developed with small sample data or unbalanced data. Therefore, to solve the problem of limited tool wear sample data and unbalanced state labels, the GAN is used to generate more recognizable tool wear category label samples. These high-resolution generated sample data are used for the expansion and improvement of unbalanced and small sample training data sets. Then, the SSAE network is used to reduce the dimensionality of the generated data samples and mine features. Finally, these features are input into the k-nearest neighbor classifier to realize the state recognition with the unbalanced data state of the tool.

First, 6000 original cutting force signals of each tool are extracted from the 315 original signals of the C1, C4, and C6 tools as the input of the GAN. Through the iterative cycles of G and D in the GAN, which counterbalance each other so that the simulated data obtained by the generator can be infinitely close to the real data and the discriminator cannot identify the source of the input data, the model training achieves the desired goal. At the end of the continuous adversarial training of the network model, new samples with a similar distribution to the input samples are obtained, thus enabling the expansion of the training samples. The GAN adopts a three-layer coding and three-layer discriminant structure. The functions used are all sigmoid functions, so that the node output falls within the (0,1) interval. The numbers of three-layer coding nodes are as follows: the first layer is 120, the second layer is 240, and the third layer is 480. The numbers of nodes in the three-layer discriminant structure are as follows: the first layer is 240, the second layer is 120, and the third layer is 1. Finally, 12000 original data samples of cutting force in three directions are generated, and a total of 315 × 3 × 12000 sample data points are generated in 315 cutting processes. The generated sample data is used as the input of the SSAE network.

5.1.2 State identification

The 1st, 150th, and 315th tool walking cutting force enhancement data points were selected as inputs to the SSAE network, where the ratio of the training set to the test set of the selected data was 7:3. The SSAE network uses a three-layer network structure and selects the sigmoid function as the encoding and decoding function of the automatic coding network. The number of iterations is set to 1000, and the sparsity parameter is 0.01. In addition, the first, second, and third hidden layer nodes are set to 100, 20, and 1, respectively. The tool wear features determined by the adaptive feature extraction of the SSAE network are input into the k-nearest neighbor classifier for state recognition. Finally, the t-distributed stochastic neighbor embedding (T-SNE) visualization technology is used to visualize the identified state features. The three tool classification results are visualized as shown in Fig. 5.

From the visual map, the three wear stages of a tool, namely, initial wear, medium wear, and later wear, can be clearly seen; all the stages can be well separated, and the clustering effect of the three wear states is clear. The confusion matrix used to represent the recognition results is shown in Fig. 6.

For the three tools of C1, C4, and C6, the average accuracies of state recognition in the *X*, *Y*, and *Z* directions reached 96.3, 96.0, and 95.3%, respectively. From the recognition accuracy, it can be proved that the original data of cutting force enhanced by the GAN can achieve a good

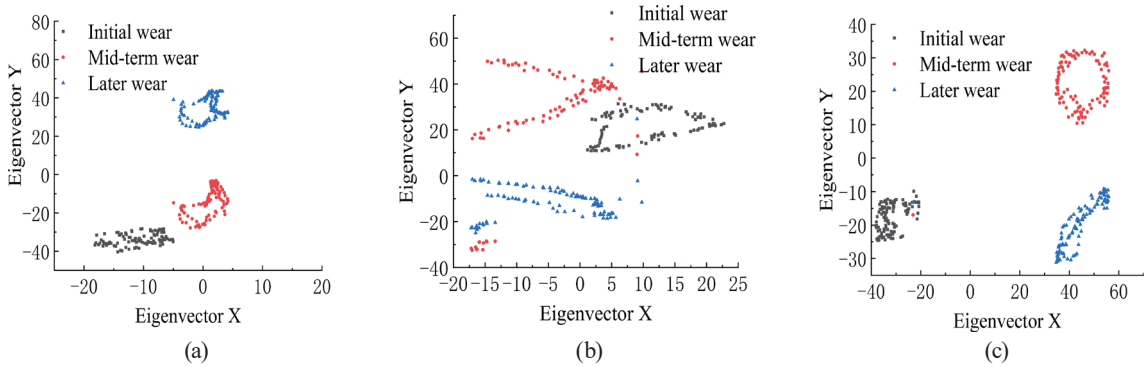

Fig. 5. (Color online) Visualization of state recognition: (a) C1, (b) C4,and (c) C6.

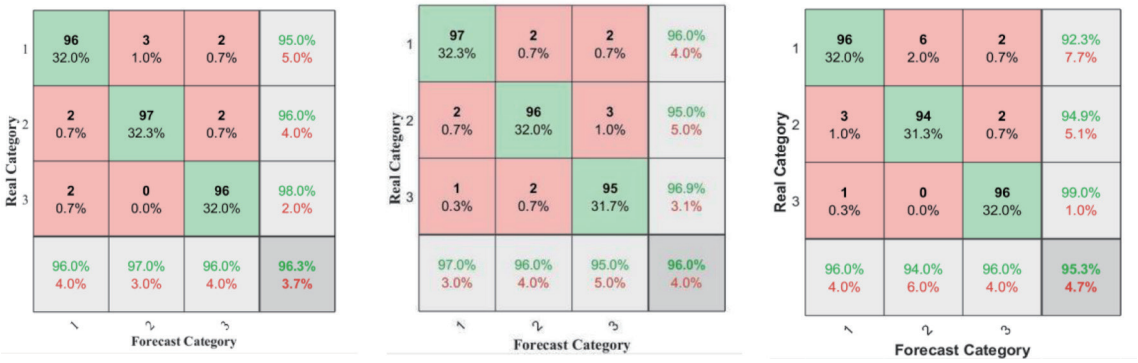

Fig. 6. (Color online) Confusion matrix: (a) C1, (b) C4, and (c) C6.

recognition effect, improve the data imbalance, and expand the data samples with insufficient and missing samples. The data enhanced by the GAN can be used as tool wear samples to predict the wear value.

5.2 Tool wear prediction

Through the accurate identification and classification of the omit tool wear stage, it can be clearly seen from the figure that the clustering effect of each state in different phases of tool wear is very good. In addition, the average accuracy of the expanded data classification is more than 95.3%. That is , the expanded data sample can be used as the wear signal to map the tool wear value. Since the regression model cannot process sequence data, feature extraction is required first. According to the correlation coefficient method, 20-dimensional feature vectors are selected from the expanded data and input into the linear regression model to predict the tool wear value.

In this study, the attention mechanism is incorporated in the output layer of the LSTM network for the prediction of tool wear values. After extracting features from the enhanced data samples, the three directional features are merged to obtain a 60-dimensional data feature set for

each tool. Before entering the linear regression model, the extracted data features are normalized. Then the normalized feature matrix is divided by time step and input into the LSTM network model to extract the hidden state of each time step of the degradation information output. After the input sample passes through the LSTM network, the hidden state of the output does not directly enter the output layer, but it is first entered into the attention network, and the hidden state weight of the LSTM learning degradation information is calculated and extracted in the attention network. Then the hidden states with different weights are output to ensure that the model learns more information related to degradation.

In the model, the training and validation numerical value ratio is set to 7:3, and other network structure parameters are set as shown in Table 5. In this study, the C1, C4, and C6 tool life cycle extraction features are used as data sets to verify the model. Each tool has 315 actual tool side wear values, each wear value corresponds to 60 data features, and each tool forms a total of 315×60 tool wear data. By taking the linear regression model as the "medium" and the historical tool wear data as the "guidance", the wear values of the three tools with different cutting processes are predicted. The accuracy of the model prediction is evaluated by combining the variance (R^2), mean absolute percentage error (*MAPE*), and the root mean square error (*RMSE*) with the following three evaluation equations:[17]

$$RMSE=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i-\hat{y}_i\right)^2}\ , \tag{9}$$

$$MAPE=\frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i-\hat{y}_i}{y_i}\right|, \tag{10}$$

$$R\ =\ -\frac{\sum_{i=}^{N}\left(y_i-\hat{y}_i\right)}{\sum_{i=}^{N}\left(\bar{y}_i-y_i\right)}, \tag{11}$$

where N is the total number of samples, and y_i, $\hat{y}_i$, and $\bar{y}_i$ are the actual wear values, the predicted values from the model, and the average of the actual wear values, respectively.

Figure 7 shows the predicted results of C1, C4, and C6 tools using the tool wear prediction regression model proposed in this study. It can be seen from the figure that the tool wear prediction curve of the model proposed in this study agrees well with the actual tool wear curve, and the three different wear stages are clearly divided. In addition, owing to the large error in the

Table 5
LSTM network structure parameter settings.

Activation function	Optimizer	Training batches	Learning decline factor	Number of cross-validations	Number of nodes	Learning rate	Number of iterations
Sigmoid	Adam	10	0.9	10	30	0.005	100

manual measurement of tool wear value, there is a certain deviation between the model predicted results and the actual wear value in the later stage of wear, but the error range is small, which has little effect on the overall predicted results of the model, thus verifying the effectiveness and accuracy of the proposed model. The accuracy indicator values of the three tools are shown in Table 6.

It can be seen from Table 6 that the *MAPE* values of C1, C4, and C6 are 6.8887, 10.4061, and 8.9716, respectively. The *RMSE* values are 8.1790, 12.4661, and 12.5025, respectively. The R^2 values are 0.9676, 0.9792, and 0.9547, respectively. The accuracy values of *MAPE* and *RMSE* are relatively small, and R^2 is greater than 0.8. This shows that the model proposed in this study has better prediction accuracy for any tool in the experiment. To prove the effectiveness and advantages of the model proposed, *RMSE* is employed as the accuracy evaluation index and compared with the prediction results of the GAN + BPNN (BP neural network), GAN + FNN (fuzzy neural network), GAN + RF (random forest), and GAN + LSTM regression models. The *RMSE* values of different prediction models are shown in Table 7.

It can be seen from the table that the average *RMSE* values of the four tools of GAN + BPNN, GAN + FNN, GAN + RF, and GAN + LSTM are 16.7734, 17.2965, 18.0509, and 14.0442, respectively, and the average *RMSE* values of the three tools of GAN + LSTM + attention model is 11.0492. Compared with those of the commonly used regression models of BPNN, FNN, and RF with enhanced data, the *RMSE* values of the tool wear prediction model proposed in this study are reduced by 34.12, 36.11, and 38.78%, respectively. This indicates that the regression model proposed in this study is superior to the commonly used machine learning networks in

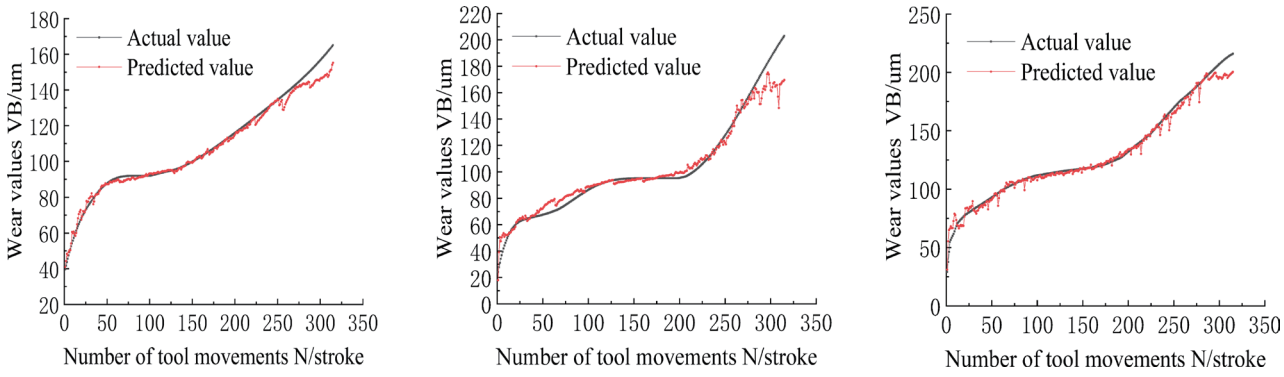

Fig. 7. (Color online) Tool wear prediction results: (a) C1, (b) C4, and (c) C6.

Table 6
Accuracy indicator values.

	MAPE	R^2	*RMSE*
C1	6.8888	0.9676	8.1790
C4	10.4061	0.9792	12.4661
C6	8.9716	0.9547	12.5025

Table 7
RMSE values of different prediction models.

	C1	C4	C6	Average
GAN + BPNN	13.8286	14.5738	21.9178	16.7734
GAN + FNN	11.2262	17.1180	23.5452	17.2965
GAN + RF	14.4621	15.5162	24.1989	18.0509
GAN + LSTM	10.6165	12.1779	19.3384	14.0442
GAN + LSTM + attention	8.1790	12.4661	12.5025	11.0492

tool wear prediction. Compared with those of the LSTM regression model with expanded data without attention mechanisms, the *RMSE* values of the model proposed in this study is reduced by 21.32%. It is proved that the combination of attention mechanisms and LSTM can better predict the tool wear value. In summary, the *RMSE* values of the proposed model are reduced by at least 21.32% compared with those of the above four commonly used regression prediction models, which again proves the effectiveness and practicality of the proposed model.

6. Conclusions

A tool wear prediction method based on the attention LSTM network with conditions of data imbalance is proposed in this study. The GAN is used to solve the problem of unbalanced tool state category labels and missing wear samples. The attention LSTM network is used to solve the problem that the hidden state of the deep learning network is miscellaneous and the features strongly related to the tool wear value cannot be fully mapped. The specific conclusions are as follows:

(1) The tool wear data samples were supplemented with the GAN, and the k-nearest neighbor classifier was used to identify the tool wear status by adaptively extracting the features of the supplemented data samples using the SSAE network. The average recognition accuracies of the three tools reached 96.3, 96.0, and 95.3%. It can be proved from the recognition accuracy that the enhanced tool wear data can achieve a good recognition effect, improve the data imbalance, and have a good expansion of the missing data samples, which can solve the problem of an unbalanced label for complex working conditions.
(2) The attention mechanism is added to the output layer of the LSTM network to predict the tool wear value. Compared with the commonly used tool wear regression prediction network model, the accuracy evaluation index (ie., the *RMSE* value) is reduced by at least 21.32%. Compared with the traditional regression model, the stability is stronger and the accuracy is higher, which proves the accuracy and feasibility of monitoring tool wear state.

The method proposed in this study is still limited to the identification and prediction of tool wear state with different parameters of a single working condition. In future research, the identification and prediction of tool wear states under different working conditions and different parameters should be considered to improve the actual utilization rate and generalization performance of the model.

Data Availability Statement

All data generated or analyzed during this study are included in this article. The data used in this article can be accessed at https://www. phmsociety.org/competition/phm/10.

Acknowledgments

Financial support from the National Natural Science Foundation of China (Nos, 51965037 and 51565030) is appreciated.

References

1 M. H. Cheng, L. Jiao, and P. Yan: J. Manuf. Syst. **62** (2022) 286. https://doi.org/10.1016/j.jmsy.2021.12.002
2 M. Shah, V. Vakharia, and R. Chaudhari: Int. J. Adv. Manuf. Tech. **121** (2022) 723. https://doi.org/10.1007/S00170-022-09356-0
3 D. Z. Wu , C. Jennings, and J. Terpenny: J. Manuf. Sci. Eng. **139** (2017) 071018. https://doi.org/10.1115/1.4036350
4 X. F. Meng, J. J. Zhang, and G. C. Xiao: Int. J. Adv. Manuf. Tech. **114** (2021) 3793. https://doi.org/10.1007/S00170-021-07152-W
5 Y. W. Wang, L. Deng, and L. Y. Zheng: Chin. J. Mech. Eng. **57** (2021) 214. https://doi.org/10.3901/JME.2021.13.214
6 Z. W. Huang , J. M. Zhu, and J. T. Lei: J. Intell. Manuf. **31** (2019) 1. https://doi.org/10.1007/s10845-019-01488-7
7 J. Y. Ou, H. K. Li, and G. J. Huang: Measurement **167** (2021). https://doi.org/10.1016/j.measurement.2020.108153
8 M. W. Wang, and J. T. Zhou: IEEE Access. **8** (2020) 140726. https://doi.org/10.1109/access.2020.3010378
9 C. Sun, M. A. Meng, and Z. B. Zhao: IEEE Trans. Ind. Inf. **15** (2019) 2416. https://doi.org/10.1109/TII.2018.2881543
10 H. D. Shao and W. Li: Fault diagnosis of rotor-bearing system based on dual-threshold attention generation adversarial network and small sample under time-varying speed, http://kns.cnki.net/kcms/detail/11.2187.TH.20221006.1458.002.html (accessed May 2022).
11 Q. S. Zhu and B. T. Sun: IEEE Trans. Instrum. Meas. **70** (2021). https://doi.org/10.1109/TIM.2021.3077995
12 F. Wu, H. Y. Nong, and C. H. Ma: J. Jilin University (Engineering and Technology Edition) **53** (2023) 989. https://doi.org/10.13229/j.cnki.jdxbgxb20220419
13 J. B. Zheng, J. Liao, and Z. B. Chen: Sensors **22** (2022) 6489. https://doi.org/10.3390/S22176489
14 Y. Yin, S. X. Wang, and J. Zhou: Appl. Intell. **53** (2022) 4448. https://doi.org/10.1007/S10489-022-03773-0
15 X. F. Meng, J. J. Zhang, and G. C. Xiao: Int. J. Adv. Manuf. Tech. **114** (2021) 3793. https://doi.org/10.1007/S00170-021-07152-W
16 R. W. Li, X. L. Ye, and F. Q. Yang: Machines **11** (2023) 297. https://doi.org/10.3390/MACHINES11020297
17 Y. R. Qin, J. F. Li, and C. X. Zhang: Entropy **24** (2022) 1733. https://doi.org/10.3390/E24121733