IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 30, 2021, accepted August 8, 2021, date of publication August 13, 2021, date of current version August 19, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3104668

Multi-Step-Ahead Tool State Monitoring Using Clustering Feature-Based Recurrent Fuzzy Neural Networks

JIACHEN YAO, BAOCHUN LU, AND JUNLI ZHANG
School of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing 210002, China
Corresponding author: Baochun Lu (nlglbc@126.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1308300.

ABSTRACT Reliable and precise multi-step-ahead tool wear state prediction is significant to modern industries for maintaining part quality and reducing cost. This study proposes a Clustering Feature-based Recurrent Fuzzy Neural Network (CFRFNN) for tool wear state monitoring and remaining useful life (RUL) prediction based on K-means Clustering, Recurrent Fuzzy Neural Network (RFNN) and Genetic Algorithm (GA). K-means Clustering method is utilized to realize tool wear state definition and input signal division, which reduces the dependence on the prior knowledge of tool wear degree and improves the prediction accuracy. Then, an enhanced RFNN model is designed and applied on the clustered features to predict tool wear state. The optimized GA technique is helpful for adaptive optimization of model parameters, which significantly improves convergence rate and prediction accuracy. The experiments on tool state prediction are performed to validate superiority of CFRFNN, and the results demonstrate that the proposed network could reasonably configure the complex non-stationary tool wear process and have high prediction accuracy of tool wear state.

INDEX TERMS Multi-step-ahead tool state prediction, remaining useful life (RUL), recurrent fuzzy neural network (RFNN), tool wear monitoring.

I. INTRODUCTION

Tool wear is a pervasive problem in manufacturing process, which also has an extremely negative effect on the perfor-mance and productivity of computerized numerical control (CNC) machines. Effective methods that can process big data in manufacturing and monitor tool wear state are meaningful to improve processing quality and reduce cost [1]–[3]. Meanwhile, it promotes the research and of tool state management systems (TSMS), which have capacity to effectively evaluate tool wear state and make prediction of tool remaining useful life (RUL) in real time.

Data-driven method has been verified a reliable approach for tool wear state monitoring and RUL prediction [4]. There are three stages in this kind of method: the first is extracting discriminant features, and the second is training machine learning models by using historical feature data-sets, the final is applying the trained model on real-time data to predict tool state and RUL. Therefore, the premise of data-driven methods is extracting discriminant tool features from input signals to train machine learning models [5]–[8]. Zhang *et al.* [5] utilized wavelet packet to extract the features of nonstationary vibrations in three vertical directions. Zhu *et al.* [6] discovered that the Holder exponents extracted from wavelet transform modulus maxima can reflect tool wear state. After extracting features from original signals, in order to enhance feature space and reduce input dimension, principal component analysis [7] and factor analysis [8] were adopted. Finally, the extracted features are utilized as input to different models, such as auto-regression [9], manifold learning [10], hidden Markov [11], sparse decomposition [12], and deep learning [13].

Among these methods, deep learning models have been extensively applied in tool state diagnosis and prognosis due to the powerful capabilities in complex sequence modeling. Sun *et al.* [14] utilized features learned by sparse auto-encoder (SAE) to train deep neural network (DNN) for induction motor fault diagnosis. Jia *et al.* [15] proposed a DNN with deep architectures to mine key information from raw signals for intelligent fault diagnosis. Ren *et al.* [16] constructed a prediction framework based on auto-encoder and DNN for prognostics of bearing RUL. Rai *et al.* [17]

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

encouraged a method based on a hybrid of empirical mode decomposition (EMD) and k-medoids clustering to monitor the degradation in bearings. Li *et al.* [18] developed a convolutional neural network (CNN) to capture multi-source data for industrial production monitoring. Wang *et al.* [19] proposed a novel SAE-LSTM for tool wear prediction under variable working conditions.

In addition to the above deep learning models for fault monitoring and diagnosis, recurrent neural network (RNN) as crucial branches of deep learning models, have been widely applied in multi-step-ahead tool wear state and RUL prediction. RNN is designed for processing sequential data and has good capabilities in time series data modeling. Current RNN models can be generally classified into three categories according to the structures: RNN [20], [21], long short term memory (LSTM) networks [22]–[24], and gated recurrent units (GRU) networks [25], [26]. LSTM is a variant of RNN and adds an approach to carry information across multiple timesteps, which can prevent early signals from gradually vanishing during processing. Zhao *et al.* [22] proposed a convolutional bidirectional LSTM to predict tool wear by analyzing raw sensory data. Zhang *et al.* [23] presented a data-driven method based on LSTM to predict RUL of engines. Hinchi *et al.* [24] proposed a deep frame-work based on convolutional LSTM recurrent units for ball bearing RUL estimation. GRU networks can be regarded as a streamlined version of LSTM networks with cheaper computational cost. Peng *et al.* [25] proposed a novel bidirectional gated recurrent unit (BGRU) to realize fault diagnosis utilizing cost sensitive active learning. Liu *et al.* [26] utilized GRU-based non-linear predictive denoising autoencoders to detect anomalous conditions and classify fault types of rolling bearing. Generally, prediction methods based on deep learning models have the capability to directly learn features from input datasets, which may not make full use of extensive expert knowledge. The adaptive network based fuzzy inference system (ANFIS) proposed by Jang *et al.* [27] combined adaptive learning ability of artificial neural network (ANN) with knowledge expression ability of FIS and can effectively use expert knowledge to deal with complex problems, such as fault diagnosis [28]–[30]. However, tool wear is a time-dependent dynamic process, whose input-output modes are difficult to be fully recognized only by static ANFI. Recently, amounts of researches have been concentrating on time series analysis ability of RNN with integrating knowledge expression ability of FNN, which have been effectively applied in multi-step-ahead time series prediction tasks [31]–[33].

In this paper, a clustering feature-based recurrent fuzzy neural network (CFRFNN) is proposed for automatic identification and prediction of tool wear state. In the proposed framework, features of each time step are first extracted from segments of original sensor signals. Meanwhile, orthogonal wavelet packet transform (OWPT) is used to realize multi-feature signal dimension reduction and noise elimination for creating initial datasets. Then, through K-means algorithm, the extracted tool features are divided into three wear degrees to provide a judgement basis for initial state and RUL prediction. Finally, an enhanced RFNN model is proposed with the aim of improving optimization performance and convergence speed. Besides, several tradi-tional models are also trained for comparison to prove the superiority of the proposed CFRFNN in tool wear moni-toring and prediction. The main contributions of this paper can be summarized as follows.

1) The proposed framework can be regarded as a hybrid approach of unsupervised feature division and supervised feature learning. The multi-timestep feature extraction and K-means clustering scheme can reduce the model size and the enhanced RFNN model is able to process datasets of different wear degrees separately to improve prediction accuracy.

2) An enhanced CFRFNN is proposed. A bidirectional stacking recurrent structure is adopted to increase network capacity. The clustering feature-based output averaging operation can directly calculate node output, which can supplement the output of the LSTM layers to ensure the full connection to the fuzzification layer. Finally, the proposed genetic-simulated annealing algorithm can determine the optimal fuzzy rules and initial parameters of CFRFNN to accelerate convergence and improve prediction accuracy.

3) The experimental studies in multisensory scenario and comparisons between CFRFNN and some other prediction methods are conducted to verify effectiveness and genera-lization capability of the proposed framework.

The remainder of this paper is organized as follows. In Section II, the LSTM model is reviewed, and then the proposed CFRFNN is shown in Section III. Experimental results are discussed in Section IV. More experiments are conducted in Section V, and the conclusions are drawn in Section VI.

II. BACKGROUNDS

The main objective of this study is to explore an optimized hybrid method for multi-step-ahead prediction. The recurrent mechanism is responsible for time series analyzing and temporal processes. In this study, Long-Short Term Memory (LSTM), a RNN variant, is adopted.

In LSTM layers, the core component is the state unit $c_j^{(t)}$ with a linear self-loop. The weight of the self-loop is modu-lated by a forget gate $f_j^{(t)}$, which adjusts the weight to a value between 0 and 1 by a sigmoid unit. There is also an input gate $i_j^{(t)}$ to control the new memory content that needs to be added but with its own parameters. Gates are computed by

$$f_j^{(t)} = \sigma(W_f x^{(t)} + U_f h^{(t-1)} + b_j^f) \tag{1}$$

$$i_j^{(t)} = \sigma(W_i x^{(t)} + U_i h^{(t-1)} + b_j^i) \tag{2}$$

where $x^{(t)}$ is the input signals and W_f, W_i, U_f, U_i, b_j^f, b_j^i are the input weights diagonal matrices, recurrent weight diagonal matrices and biases, respectively. $h^{(t)}$ is the output of LSTM unit that can also be modulated by the output gate $o_j^{(t)}$, which can be expressed by:

$$h_j^{(t)} = o_j^{(t)} \tanh(c_j^{(t)}) \tag{3}$$

$$o_j^{(t)} = \sigma(W_o x^{(t)} + U_o h^{(t-1)} + b_j^o) \tag{4}$$

which has the parameters W_o, U_o, b_j^o for its input weights matrices, recurrent weight diagonal matrices and biases. The illustration of LSTM model is given in Figure. 1.

In the model proposed in this study, the evolutionary fuzzy neural layers are embedded into an enhanced LSTM network. First, the enhanced LSTM layers are used to internally loops over entire time series elements, and the output can be calcu-lated by two components: the output of bidirectional stacking recurrent layers and the clustering feature-based averaging output, which can complement each other. Furthermore, an optimized genetic-simulated annealing algorithm is proposed to select the optimal set of initial weights and rules with high classification ability, which can enhance accuracy and acce-lerate the convergence of CFRFNN.

III. TOOL STATE MONITORING AND PREDICTION WITH CFRFNN

The core task of multi-step-ahead tool wear state prediction is to estimate the expected time from current cutting features to severe wear features. The accuracy of RUL prediction is mainly affected by the following two ingredients: the current tool wear features and the definition of severe wear. In this paper, the K-Means clustering method is used to divide and identify the tool wear state into three degrees according to the features extracted from experimental datasets for realizing real-time judgement of tool wear state as well as improving prediction accuracy through datasets segmentation. Figure. 2 shows the diagram of prediction through CFRFNN. After extracting local features of each timestep from raw sensor signals, the CFRFNN is applied on the current features to predict the features of subsequent timesteps.

A. CLUSTERING BASED ON K-MEANS METHOD

In the practical application of multi-step-ahead tool wear state prediction, the first problem is defining tool wear state and determine the initial state of monitored tool. In this study, tool wear state is divided into three states by K-means method referring to the Taylor tool life curve. Compared with other clustering methods, K-means method has advan-tages of good clustering effect and fast convergence, which is suitable for the real-time identification of tool wear state. The purpose of K-means method is to classify the input features into several clusters, and update the cluster centers on premise of minimizing the objective function O. Therefore, it can be utilized to determine tool wear degree. The distance between input features x_i and cluster centers c_j can be defined by:

$$d(x_i, c_j) = \sum_{l=1}^{f} \left\| x_{il} - c_j \right\| \tag{5}$$

where $\|*\|$ is the general Euclidean distance and f is the number of input features. Then the objective function can be

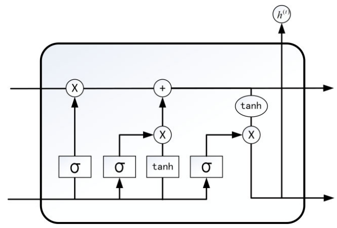


FIGURE 1. Illustrations of LSTM networks.

expressed as:

$$d(x_i, c_j) = \sum_{l=1}^{f} \left\| x_{il} - c_j \right\| \tag{6}$$

where p is the dimension of input features and q is the number of cluster centers.

B. ENHANCED LSTM WITH CLUSTERING OUTPUT AVERAGING

Although fuzzy layers have unique characteristics of strong knowledge expression and learning ability. In the real-world RUL prediction, there also exist temporal correlations and vanishing gradient problems [34], [35], which result in degradation of the performance of fuzzy layers. LSTM networks are time dependent and can process the timesteps in order, which is why they perform well on RUL prediction tasks. In CFRFNN, the enhanced LSTM layers are added to memorize relations among features of various timesteps.

1) RECURRENT DROPOUT REGULARIZATION

To prevent the problem of overfitting with limited data, the dropout regularization is applied to randomly zeroing-out the input unit, thus to break happenstance correlation in training data. The core idea of dropout is to randomly drop units (set to 0) from neural network during training. Each unit in each timestep has same dropout mask p to avoid disrupting error signals. Dropout generally utilized to the outputs of a fully connected LSTM can be written as:

$$h_t = Z * \mathrm{H}(x_t, h_{t-1}) \tag{7}$$

where $*$ denotes the element wise product. Z is a binary mask vector with each element generated from $Z_r \sim Bernoulli(p)$.

2) BIDIRECTIONAL STACKING RECURRENT LAYERS

To increase the model capacity and flexibility as well as providing sufficient information for subsequent fuzzy layers, bidirectional stacking recurrent layers are introduced to expand the network capacity. The bidirectional structure is able to memorize both the past and future signals during processing time series information by dividing the regular LSTM neurons into two directions: one for forward direction and the other for backward direction, which also enhance the role of time series information by capturing potentially richer patterns. The hidden layer concatenated vectors of the final

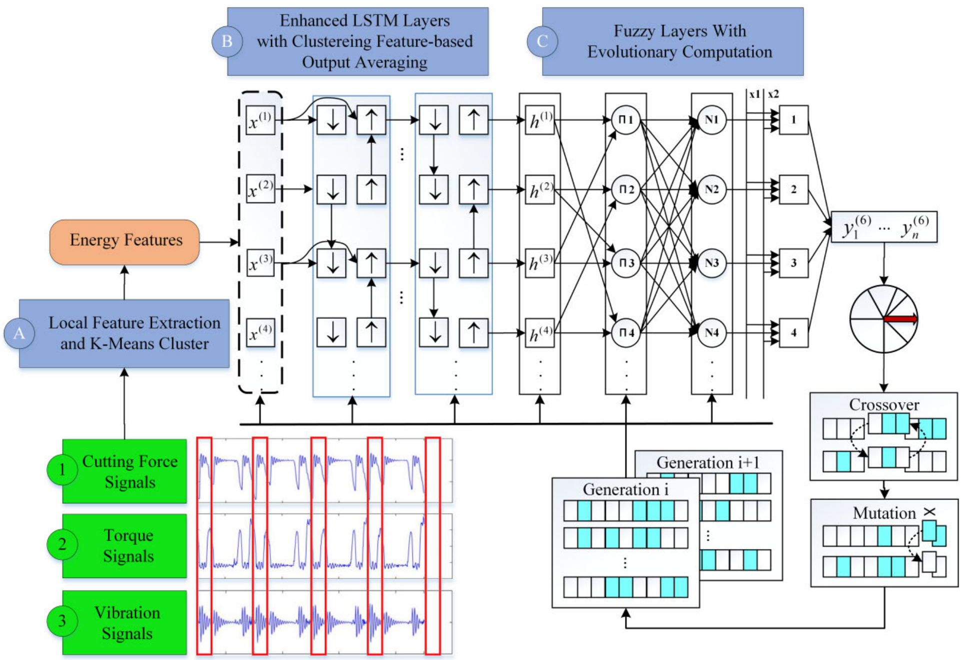


FIGURE 2. Architecture of the proposed CFRFNN.

timestep h_{t+n} can be written as follows:

$$h_{t+n} = \vec{h}_1 \otimes \overleftarrow{h}_{t+n} \tag{8}$$

where $\vec{h}$ and $\overleftarrow{h}$ represent the forward and backward layers output, which are both calculated through the standard LSTM updating equations as follows:

$$\vec{h}_t = \vec{F}\left(\vec{x}_t, \vec{h}_{t-1}; \vec{\nabla}_{LSTM}\right) \tag{9}$$

$$\overleftarrow{h}^{(t)} = \overleftarrow{F}\left(\overleftarrow{x}_t, \overleftarrow{h}_{t+1}; \overleftarrow{\nabla}_{LSTM}\right) \tag{10}$$

where F is the standard LSTM updating equations defined by (1)-(4) and ∇_{LSTM} is the parameter set.

In addition, to construct a more powerful recurrent structure, the stacking recurrent layers are adopted to extract more generalized sequence features. Based on the above idea, a method of stacking multiple LSTM hidden layers is proposed to increase the capacity of CFRFNN. Each intermediate layer returns a full sequence of outputs and the averaging method is introduced to optimize the final output at each timestep t. The final output can be expressed as:

$$\bar{h}_t = \frac{1}{n}\sum_{l=1}^{n} h_t^{(l)} \tag{11}$$

where n denotes the labels of hidden layers.

3) CLUSTRING FEATURE-BASED OUTPUT AVERAGING

The final output of the above LSTM layers at each timestep can be regarded as the input of next fuzzy layers. However, in the real process of data acquisition, feature extraction and enhanced LSTM layers, sequence information might be lost. Considering different output characteristics among several wear states, clustering feature-based output averaging is proposed to ensure integrity of output sequence. The average output is expressed as:

$$\bar{h}_t = \sum_{i}^{S_n} \omega^{(i)} h_t^{(i)} \tag{12}$$

where $S_n = [T_n^{(s)}, T_n^{(e)}]$ denotes the different wear states divided by K-means algorithm, $T_n^{(s)}$ and $T_n^{(e)}$ denote the start timestep and end timestep, respectively. n is the clustering type of timestep t, i is the index for timestep.

As the LSTM layers are bidirectional and the extracted features have been clustered by K-Means method, which means the intermediate time series information of each cluster has a larger impact on the output, thus the impact of clustering middle output is highlighted to improve reliability of the algorithm, weights are designed as follows:

$$\omega_i = \frac{\exp(l(i))}{\sum\limits_{k=T_n^{(s)}}^{T_n^{(e)}} \exp(l(k))} \tag{13}$$

where the formula $l(*)$ denotes the minimum distance from current timestep to cluster boundary, which is expressed as:

$$l(i) = \min(i - T_n^{(s)}, T_n^{(e)} - i) \tag{14}$$

C. FUZZY LAYERS WITH EVOLUTIONARY COMPUTATION

At last, the learned results $h = [h_1, h_2 \ldots, h_n]$ are transmitted into the fuzzy layers for further fitting. In fuzzy inference system, the main difficulties are choosing the optimal weights and selecting the suitable fuzzy rules. Therefore, an optimized genetic-simulated annealing algorithm is proposed to optimize the initial weights and choose the fuzzy rules with high classification ability. To prevent CFRFNN from falling into local optimum and accelerate the convergence speed, the proposed genetic algorithm increases probability of dominant individual selection and accepts the worsening solution with a certain probability.

1) FIVE-LAYER FEEDFORWARD STRUCTURE

In this study, the proposed CFRFNN can effectively integrate the learning dexterity of neural layers with the humanlike knowledge expression of fuzzy layers. In fuzzy layers, first is fuzzification layer, in which neurons convert input features into membership degrees through activation function. Then is rule layer, in which each neuron represents a fuzzy rule. The connection of rule antecedents can be calculated by operator product, which can be expressed by:

$$y_i^{(2)} = \omega_{Ai} \times \omega_{Bi} \tag{15}$$

where $y_i^{(2)}$ is the output of ith rule neuron. The value of ω_i corresponds to the true value of ith rule.

As shown in Figure. 2, the third layer is normalization layer, in which each neuron evaluates normalized excitation intensity to represent contribution rate to final result. Hence, the output can be updated as:

$$y_i^{(3)} = \frac{x_{ii}^{(3)}}{\sum_{j=1}^{m} x_{ji}^{(3)}} = \frac{\omega_i}{\sum_{j=1}^{m} \omega_j} = \overline{\omega_i} \tag{16}$$

where m is the sum of rule neurons, and $x_{ji}^{(3)}$ represents the ith input of jth neuron.

The main function of next layer is defuzzification, each neuron is connected to the corresponding normalized neuron and initial inputs. The defuzzification layer calculates the weighted consequent value, which is determined as:

$$y_i^{(4)} = \overline{\omega_i}(p_{i1}x_1 + p_{i2}x_2 + q_i) \tag{17}$$

where p_i are the consequent parameters.

The final layer is a neuron to synthesize the output of the previous layer then to provide the actual outputs.

2) OPTIMIZING WEIGHT MATRICES VIA GENETIC ALGORITHMS

In training process, each epoch can be divided into forward transmission and backward propagation. In forward transmission, each neuron output is calculated layer by layer. The output of defuzzification layer, Eq. (17), is a linear function, which can be established according to the consequent parameters in the matrix notation:

$$y^{(4)} = Mp \tag{18}$$

where M is the output. p are the consequent parameters, which are expressed by Least Square Estimation (LSE). And the LSE of p, $p*$ can minimize the squared error $\left\|Mp - y^{(5)}\right\|^2$, which can be realized by pseudo-inverse technique:

$$p* = (M^T M)^{-1} M^T y^{(5)} \tag{19}$$

where M^T is the transpose of M.

In backward propagation, when the optimal rule consequent parameters are determined, the error signals will be propagated backward. Antecedent parameters are evaluated in terms of the chain rule:

$$f = \sum_{i=1}^{l} e_i^2 = \sum_{i=1}^{l} (y_i^{(5)} - y_i) \tag{20}$$

where f is the fitness value of population, e_i is the error vector of ith variable, l is the total number of input variables, $y_i^{(5)}$ and y_i are the actual output and desired overall output of ith variable.

In back propagation learning algorithm, there may be the problem of converging to sub-optimal weights, which indicates the optimal solution may not be reached. To get better initial weight parameters, an optimized genetic method is proposed, and the procedure involves:

a. Chromosome population initialization and fitness function definition, including population size, mutation probabi-lities and learning rate.
b. Roulette method has been employed as selection strategy to choose optimal individuals with high fitness. As well adjusting fitness with simulated annealing algorithm to stick out the excellent individuals, the probability of individual selection is as follows:

$$p_i = \frac{e^{1/f_i T}}{\sum_{i=1}^{n} e^{1/f_i T}} \tag{21}$$

where n is the size of population and T denotes the temperature in simulated annealing algorithm.
c. Two genetic operators, crossover and mutation, are used to produce offspring chromosomes. In order to enable the algorithm to escape local extremum and avoid premature convergence, the worsening solution with better fitness value is accepted with probability in process of mutation. The acceptance criteria are as follows:

$$A = \begin{cases} 1, & df \leq 0 \\ e^{df}, & df > 0 \end{cases} \tag{22}$$

where df is the difference between offspring and parent chromosome fitness values.
d. Repeat the process (b) and (c) until the specified number of iterations is reached.

3) SELECTING APPROPRIATE FUZZY RULES VIA EVOLUTIONARY COMPUTATION

In fuzzy layers, evolutionary computation is employed to choose and optimize fuzzy IF-THEN rules for more precise

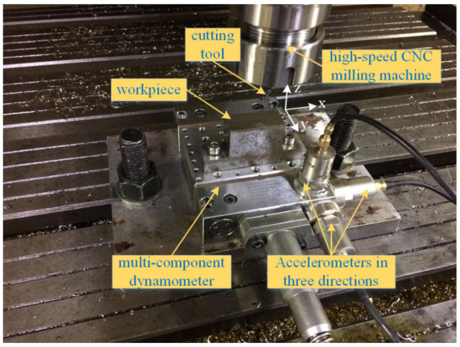


FIGURE 3. **Experimental setup for tool state prediction.**

prediction. Since the number of multiple fuzzy rule tables required may be quite large, genetic algorithm is adopted to minimize the number of rule sets. The steps selecting fuzzy IF-THEN rules by genetic algorithm are similar to the steps described above, and fitness function satisfying optimization requirements is as follows:

$$f(P) = \omega_A \frac{A_S}{A_{ALL}} - \omega_R \frac{R_S}{R_{ALL}} \tag{23}$$

where A_s is the amount of successful prediction, A_{ALL} is the total amount of prediction datasets, R_s and R_{ALL} are the amount of IF-THEN rules in the set P and P_{ALL}. The importance of prediction accuracy and size of a ruleset are reflected by specifying the weights ω_A and ω_R.

The whole framework of CFRFNN has been illustrated in Figure. 2.

IV. EXPERIMENTAL VALIDATION

To test the performances of the proposed CFRFNN, real-time series data of tool wear in same working conditions are applied. In this study, according to the Taylor tool life curve, the tool wear state can be classified into three stages: slight wear, moderate wear, and severe wear. The specific degree of tool wear is measured by K-means method.

A. BENCHMARKING DATA DESCRIPTION

1) EXPERIMENT SETUP

The experiment was conducted on a high-speed CNC milling machine without cutting fluid and the signals were collected synchronously. The experimental platform is shown in Figure. 3. The cutting work-piece material was CR12moV and cutting tools were carbide end-milling cutters with 4 teeth, as shown in Figure. 4(b). The total length and the diameter of the cutting tools were 50mm and 6mm, respec-tively. Cutting force, torque and vibration signals were selected as observational signals. Other processing parameters were as follows: The spindle rotation speed was 3000 r/min; the feed speed in x direction was 240 mm/min; the cutting depth in y direction (radial direction) and z direction (axial direction) were both 0.5 mm; the sampling frequency of signals was 20 kHz and the duration of each cutting segment was 10 seconds. To get adequate information,

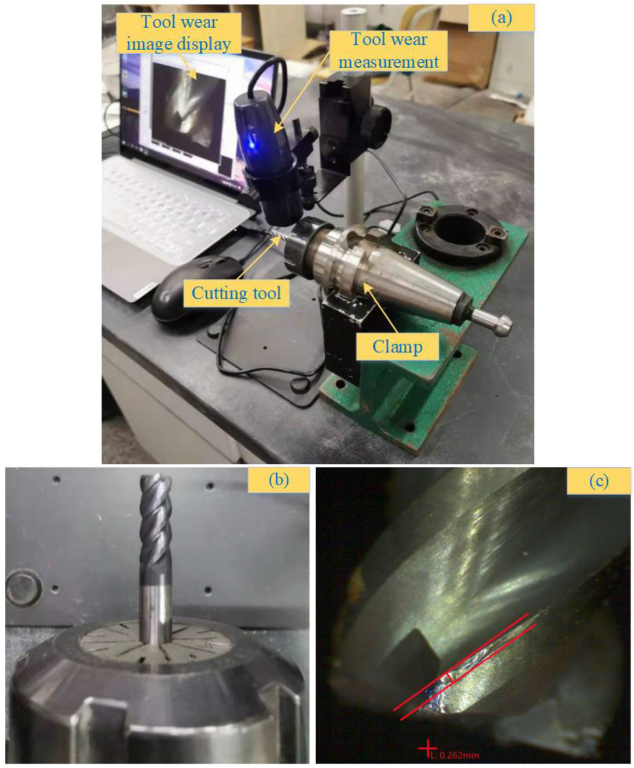


FIGURE 4. **Tool wear. (a) Tool wear monitoring system. (b) Tool used for processing. (c) Tool wear measurement.**

the program was set to generate 389 cutting segments from each cutting tool. Then, to acquire real-time cutting force and torque signals in three directions, a Kistler compact multi-component dynamometer was mounted under the cutting work-piece. At the same time, to collect real-time vibration signals in three directions, a DAQ Elsys TraNET 404S8 was adopted and piezo accelerometers were installed on Kistler dynamometer with a same sampling frequency of 20 kHz during the cutting experiment. The accelerometer sensitivity was 100 mV/g and the frequency range was 0.5 Hz to 13 kHz. The flank wear of the cutting tools measured by a digital measuring microscope INSIZE ISM-WF200 is shown in Figure. 4. To ensure the continuity of the production process, K-means method is adopted to analyze the extracted features instead of measuring the flank wear width. The main type of tool wear in this paper is mechanical wear and the tool life end criterion is based on the boundary determined by K-means method. According to the Taylor tool life curve, the tool wear range of slight wear state is under 0.15 mm and the tool wear range of severe wear state is over 0.3 mm. The features of the tool wear boundary are extracted and then the number of clustering kernels is constantly adjusted to fit the boundary features, thus K-means can be correlated to the real tool wear. The input sensor signals contained nine types of data: force, torque and vibration in three directions. Three sets of tool wear experiments were carried out: two experi-ment acquisition signals were used as training datasets, and the third were used as test datasets.

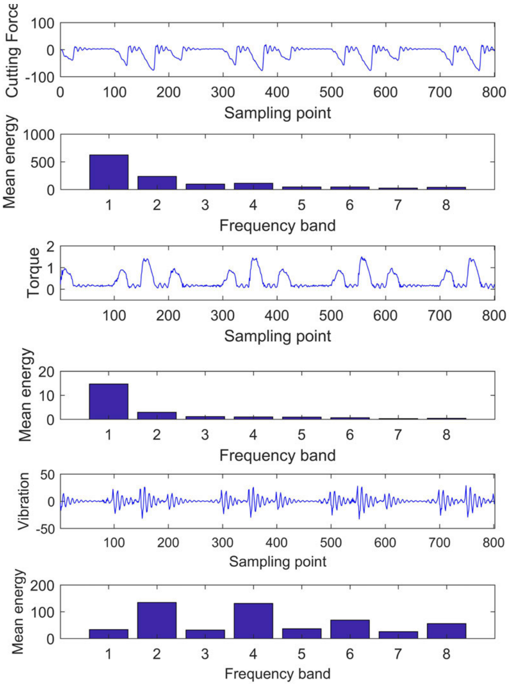


FIGURE 5. Cutting force, torque, and vibration signals and their energy wavelet packet.

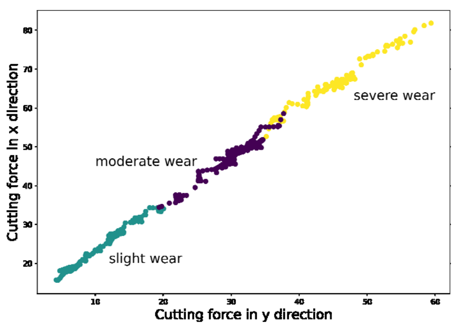


FIGURE 6. Two-dimensional (2-D) projection clustered by K-means method.

2) FEATURE EXTRACTION AND DATA CLUSTERING

To reduce data size and extract features without losing information, orthogonal wavelet packet transform (OWPT) is used to extract the wavelet energy feature, whose energy spectrum distribution is related to tool wear state. OWPT shows a good effect on noise elimination and dimension reduction, which can improve the prediction accuracy and calculating speed. The input signals can be decomposed into 2^i sub-bands by i-level OWPT, and the energy of each sub-band can be expressed by:

$$E_i(t_i) = \sum_{k=1}^{m} \left| x_{i,k} \right|^2 \tag{24}$$

where $x_{i,k}$ is the amplitude of reconstructed signals.

Cutting force, torque and vibration signals and their mean energies are shown in Figure. 5. Three sets of acquired signals were transformed into feature datasets by OWPT and the feature of each timestep contains 81-dimensional vectors. Then, in order to define the boundary of three wear degrees and determine the tool wear degree in real time, K-means method is used to cluster the training feature datasets. With an increase in amount of training datasets, the boundary becomes clearer and the tool wear diagnosis is more accurate. The 2-dimensional projection of clustering results is shown in Figure. 6. Two vectors are chosen as coordinate axes and 2-D projection can show the boundaries of different wear state, which provides foundation for RUL prediction. The mapping process to determine the tool wear boundary is shown in Figure. 7.

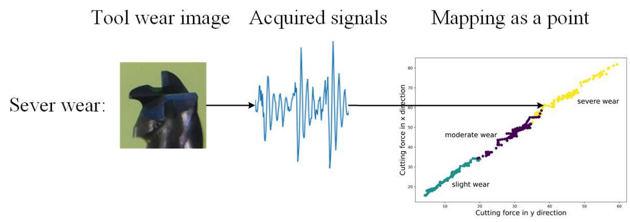


FIGURE 7. Mapping process to determine the tool wear boundary.

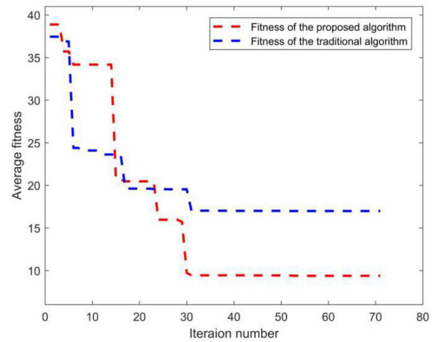


FIGURE 8. Average fitness calculated by CFRFNN.

B. CFRFNN TRAINED BY RUN-TO-FAILURE DATASETS

The practical application of multi-step-ahead tool wear state prediction goes through two phases: offline training and online application. The primary objective of training stage is fitting the network offline, which means the more training time cost due to the more complicated of methods will not affect the real-time performance. Therefore, the optimization method in training stage focus on the accuracy improvement.

After data clustering, the training datasets can be input into CFRFNN for network training. The first step is determining network structure, the amount of input and output nodes is adjusted according to the extracted features. Furthermore, initial population of two kinds of chromosomes is randomly generated: each gene in one kind of chromosome represents a true value of each initial weight and the threshold of whole network, each gene in the other kind represents a specific fuzzy IF-THEN rule. During training process, the average fitness of each generation is calculated and recorded by averaging the fitness of all individuals in current population. The relationship between the average fitness and the iteration number is shown in Figure. 8. It can be seen that an obvious decreasing of the average fitness exists along with the increase of iteration number, which means the training error is decreasing gradually. Compared with traditional genetic algorithms, the proposed genetic algorithm is a more effective optimization method for the proposed CFRFNN network, which has higher prediction accuracy and convergence speed.

C. TOOL WEAR STATE AND RUL PREDICTION RESULTS

Different from training stage, the main task of application stage is predicting wear state and RUL of a new tool. During data preprocessing phase, the amount of input parameters is reduced by the feature extraction and the K-means method. As all networks have been trained, only one forward trans-mission process is conducted. Therefore, CFRFNN is able to make multi-step-ahead prediction in real time.

In application stage, the validation dataset is input into the trained network. As run-to-failure tool wear state describes the wear percentage and increases along with cutting step forward. To quantitatively measure the change of tool wear degree, the wear degree formula is defined as follows:

$$W = \frac{\bar{f} - \bar{f}_{\min}}{\bar{f}_{\max} - \bar{f}_{\min}} \tag{25}$$

where $\bar{f}$ denotes the average feature. To highlight the impact of extracted features containing major information. Root mean square (RMS) values are adopted to calculate the average feature:

$$\bar{f} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} f_i^2} \tag{26}$$

where n is the number of extracted features.

In this study, the interval between cutting segments, and timesteps, is set to 6min. Each tool is designed to create 389 segments (horizons 0 up to t + 7, each horizon = 4.86h) and 20 steps ahead tool wear state prediction is studied by recursive method to demonstrate the effectiveness of CFRFNN. More specifically, the information of 20 timesteps is predicted and then complemented with the existing time series for the feature prediction of subsequent 20 timesteps. The prediction results of tool cutting force sub-feature, vibration sub-feature, torque sub-feature and wear degree in online application stage are shown in Figure. 9. Time steps T_1 and T_2 are the boundaries of different wear degree, which are determined by K-means method. More specifically, the number of clustering kernels is constantly adjusted to match the required boun-daries, and the number of tool wear stages reorganized according to the demand, which is set at 3 in this paper. It can be seen that the predicted values approximate real values very well, which shows CFRFNN is an effective multi-step-ahead prediction method for tool wear monitoring. At last, RUL can

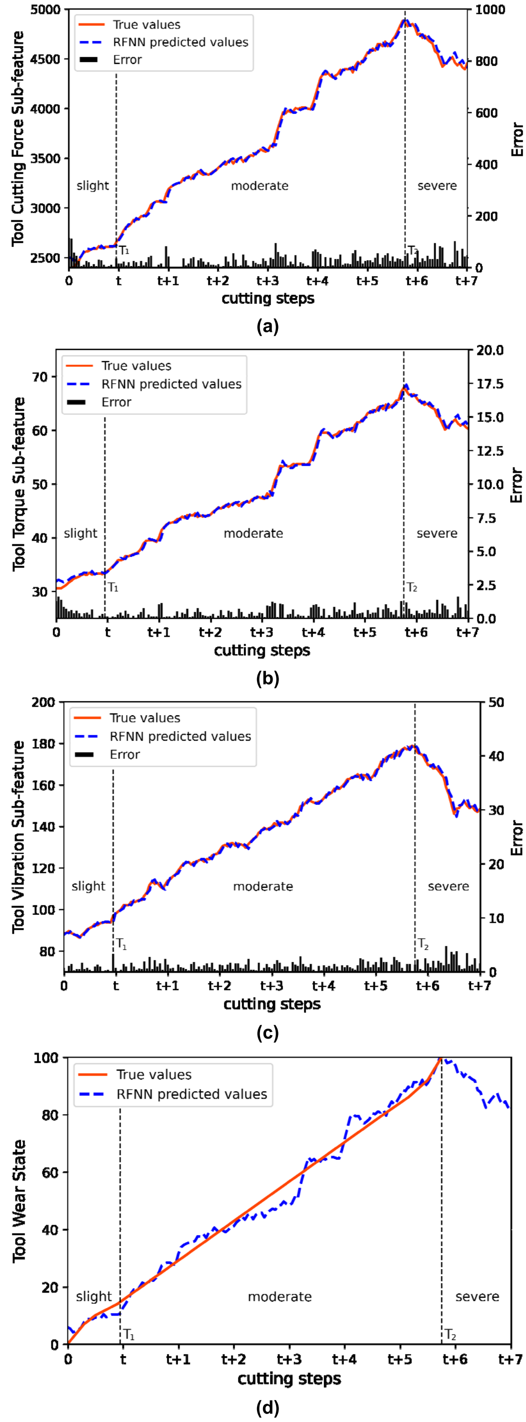


FIGURE 9. Eight cutting steps ahead prediction results learned by CFRFNN. (a) tool cutting force sub-feature and RUL prediction. (b) tool torque sub-feature and RUL prediction. (c) tool vibration sub-feature and RUL prediction. (d) tool wear degree and RUL prediction.

be calculated as follows:

$$RUL = T_2 \tag{27}$$

To investigate the relationship between the outputs of the above variables, the correlation analysis is conducted and the scatter plots is shown in Figure. 10. The results show

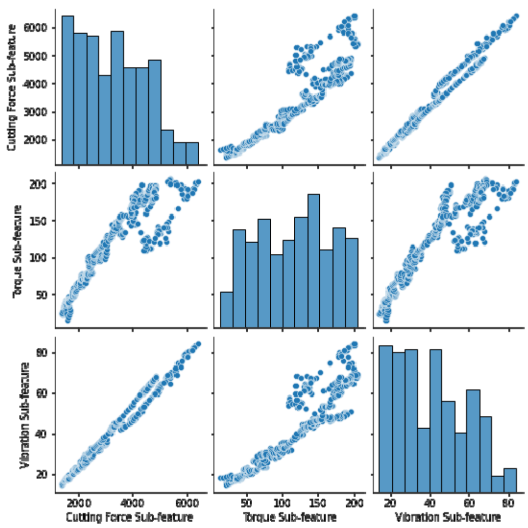


FIGURE 10. The relationship between the outputs of multiple variables.

TABLE 1. MAE achieved by compared algorithms in multi-step-ahead prediction tasks.

	Horizon							
	t	t+1	t+2	t+3	t+4	t+5	t+6	t+7
CFRFNN	0.66	0.65	0.70	0.78	0.92	1.08	1.21	2.13
RNN	0.83	0.87	0.95	1.22	1.31	1.51	1.59	2.74
LSTM	0.61	0.71	0.83	1.14	1.25	1.34	1.47	2.43
GRU	0.72	0.70	0.85	1.18	1.27	1.45	1.54	2.59
FNN	0.84	0.91	1.13	1.28	1.38	1.55	1.71	3.01
EFNN	0.71	0.81	0.87	1.16	1.24	1.38	1.52	2.86

a strong linear correlation between the predicted outputs, which means each output vector can effectively represent and predict the tool wear state.

Furthermore, to quantify the performance of all prediction models, two indicators for the evaluation of model robustness and prediction ability are utilized including mean absolute error (MAE) and root mean squared error (RMSE). The corresponding equations of the two indicators are expressed as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|T_i - P_i| \tag{28}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(T_i - P_i)^2} \tag{29}$$

where T_i and P_i are the true and predicted tool wear features, respectively. The errors of tool state and RUL prediction are shown in Table 1 and Table 2. It shows that low prediction errors are acquired by CFRFNN, which shows effectiveness for RUL prediction.

D. COMPARISON AND DISCUSSION

To investigate the performance enhancement of multi-step-ahead prediction caused by CFRFNN, six algorithms for the time series analysis are conducted to compare the prediction accuracy. For accurate display and comparison, the MAE and RMSE are shown in Table 1 and Table 2.

TABLE 2. RMSE achieved by compared algorithms in multi-step-ahead prediction tasks.

	Horizon							
	t	t+1	t+2	t+3	t+4	t+5	t+6	t+7
CFRFNN	0.67	0.69	0.82	1.09	1.26	1.39	1.95	5.51
1RNN	0.80	1.12	1.26	1.79	2.23	2.56	2.78	6.79
LSTM	0.66	0.69	1.04	1.61	2.07	2.23	2.48	6.26
GRU	0.75	0.71	1.06	1.69	2.18	2.43	2.65	6.61
FNN	0.92	1.19	1.58	1.90	2.40	2.75	2.87	7.48
EFNN	0.72	0.76	1.05	1.65	2.13	2.38	2.57	6.98

TABLE 3. The cutting parameters for extend experiments.

Case	spindle speed (r/min)	cutting depth (mm)	feed speed (mm/min)
1	3000	0.5	240
2	4000	0.5	240
3	3000	1	240

In RNN and its variants, RMSprop optimizer is adopted as the gradient descent method to optimize the learning rate. The online prediction errors of all algorithms have been calculated. From the results comparison, the performance of CFRFNN shows the improved accuracy and generalization ability in multi-step-ahead tool wear state prediction. By comparing the prediction errors, several significant findings have been acquired and discussed as follows:

1) As expected, the prediction accuracy of all algorithms deteriorates with the increasing of cutting steps. In multi-step-ahead tool wear prediction, CFRFNN gradually shows its superiority compared with other algorithms. As cutting steps forward, the difference of prediction perfor-mance among various algorithms becomes larger, but CFRFNN can maintain the stable operation and consis-tently provide the lowest MAE and RMSE values. Owing to the fact that the wear state of same tools has a common trend, the bidirectional structure can fully consider the trend and capture the missed aspects. Besides, the clustering feature-based output ensures the full connection of the network and the evolutionary computation can provide the parameters with higher accuracy. Therefore, CFRFNN has the stronger learning ability and the better performance in tool wear prediction.
2) As a hybrid model, CFRFNN can combine the advantages of sub algorithms to improve the performance. In addition, as RNN and its variants can process series information incrementally thus to give a fluid representation, the performance of RNN is better than FNN.

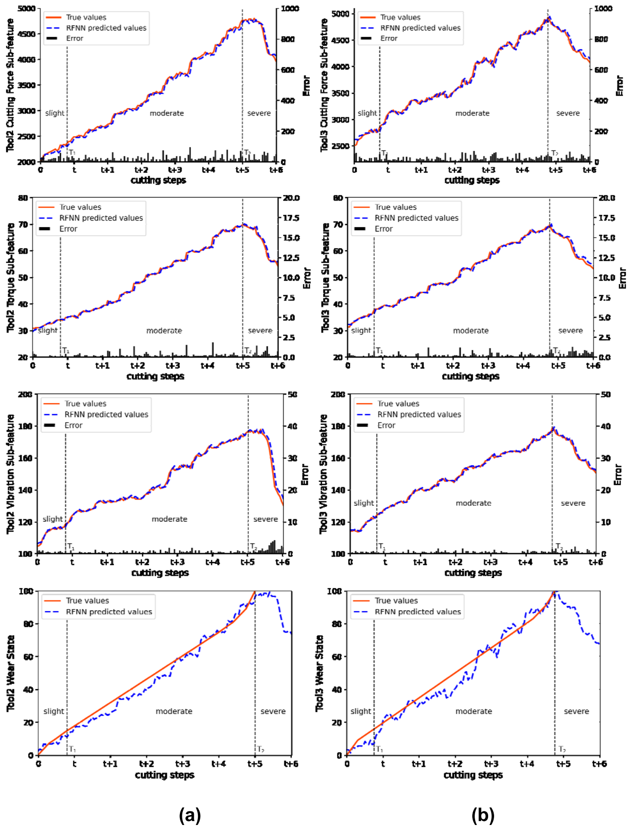


FIGURE 11. Predicted and true values of two extend tools. (a) tool sub-features and wear state of tool2. (b) tool sub-features and wear state of tool3.

V. MODEL ANALYSIS AND PERFORMANCE EVALUATION

A. MODEL ANALYSIS

In Section IV, the tool wear state and RUL prediction is performed by the proposed CFRFNN method and other compared algorithms. However, the tool cutting conditions are rather limited. To further verify the effectiveness and generalization ability of the proposed method, more experiments with different cutting parameters are conducted. The specific parameters during the cutting process are shown in Table 3, in which the Case 1 is the initial experiment and the cutting depth represent both radial and axial.

B. MODEL PERFORMANCE EVALUATION

The feature vectors of tool cutting force, torque and vibration are used as input for RUL prediction in this section. During cutting process, the cutting parameters, including spindle speed and cutting depth, may lead to different dynamic ranges, which will affect prediction accuracy. Two cutting parameters commonly used in practical manufacturing are chosen to verify the generalization ability of CFRFNN. The feature and RUL prediction results are shown in Figure. 11. The predicted value has similar trend and fluctuates around the actual value, which indicates that the proposed method has good accuracy and adaptability. It can be found that tool RUL is different along with the change of cutting parameters, which can be reflected in location moving of the point T_2.

As wavelet packet decomposition is adopted to calculate energy coefficient, the input for prediction models is stable and low dimensional, which ensure the strong generalization ability to various cutting parameters. At the same time, the growth of extracted features in real machining will provide the ample training data, which is conducive to the further improvement of prediction accuracy through continuous learning.

VI. CONCLUSION

In this paper, we have proposed a clustering feature-based recurrent fuzzy neural network (CFRFNN) based on TSMS for multi-step-ahead tool wear state and RUL prediction. After extracting features from sensor signals, K-means algorithm is first adopted to cluster for real-time tool state monitoring. Then, an enhanced CFRFNN is utilized to learn the extracted features for prediction. The experimental results verify that the proposed CFRFNN model possesses a more accurate prediction precision and better generalization capabilities.

This method could be extended to the other systems whose degradation process resembles tool wear process, and can also be generalized to determine new tools quality. In the future work, the evolutionary capability of CFRFNN weights will be expanded so that it may adapt to the sudden change of working parameters.

REFERENCES

[1] J. Wan, S. Tang, D. Li, S. Wang, C. Liu, H. Abbas, and A. V. Vasilakos, "A manufacturing big data solution for active preventive maintenance," *IEEE Trans. Ind. Informat.*, vol. 13, no. 4, pp. 2039–2047, Aug. 2017.

[2] Y. Lei, F. Jia, J. Lin, S. 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.

[3] X. Liang, Z. Liu, and B. Wang, "State-of-the-art of surface integrity induced by tool wear effects in machining process of titanium and nickel alloys: A review," *Measurement*, vol. 132, pp. 150–181, Jan. 2019.

[4] K. Zhong, M. Han, and B. Han, "Data-driven based fault prognosis for industrial systems: A concise overview," *IEEE/CAA J. Automatica Sinica*, vol. 7, no. 2, pp. 330–345, Mar. 2020.

[5] C. Zhang, X. Yao, J. Zhang, and H. Jin, "Tool condition monitoring and remaining useful life prognostic based on a wireless sensor in dry milling operations," *Sensors*, vol. 16, no. 6, p. 795, May 2016.

[6] K. Zhu, T. Mei, and D. Ye, "Online condition monitoring in micromilling: A force waveform shape analysis approach," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3806–3813, Jun. 2015.

[7] P. Cai and X. Deng, "Incipient fault detection for nonlinear processes based on dynamic multi-block probability related kernel principal component analysis," *ISA Trans.*, vol. 105, pp. 210–220, Oct. 2020.

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

[9] Y. Chen, X. Liang, and M. J. Zuo, "Sparse time series modeling of the baseline vibration from a gearbox under time-varying speed condition," *Mech. Syst. Signal Process.*, vol. 134, Dec. 2019, Art. no. 106342.

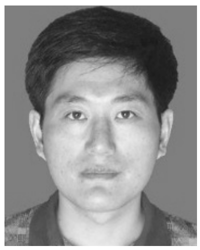
[10] W. Li, S. Zhang, and S. Rakheja, "Feature denoising and nearest–farthest distance preserving projection for machine fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 12, no. 1, pp. 393–404, Feb. 2016.

[11] O. Geramifard, J.-X. Xu, J.-H. Zhou, and X. Li, "A physically segmented hidden Markov model approach for continuous tool condition monitoring: Diagnostics and prognostics," *IEEE Trans. Ind. Informat.*, vol. 8, no. 4, pp. 964–973, Nov. 2012.

[12] C. Wang, H. Li, J. Ou, R. Hu, S. Hu, and A. Liu, ''Identification of planetary gearbox weak compound fault based on parallel dual-parameter optimized resonance sparse decomposition and improved MOMEDA,'' *Measurement*, vol. 165, Dec. 2020, Art. no. 108079.
[13] H. Wu and J. Zhao, ''Self-adaptive deep learning for multimode process monitoring,'' *Comput. Chem. Eng.*, vol. 141, Oct. 2020, Art. no. 107024.
[14] C. Sun, M. Ma, Z. B. Zhao, and X. Chen, ''Sparse deep stacking network for fault diagnosis of motor,'' *IEEE Trans. Ind. Informat.*, vol. 14, no. 7, pp. 3261–3270, Jul. 2018.
[15] 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 Process.*, vols. 72–73, pp. 303–315, May 2016.
[16] L. Ren, Y. Sun, J. Cui, and L. Zhang, "Bearing remaining useful life prediction based on deep autoencoder and deep neural networks,'' *J. Manuf. Syst.*, vol. 48, pp. 71–77, Jul. 2018.
[17] A. Rai and S. H. Upadhyay, ''Bearing performance degradation assessment based on a combination of empirical mode decomposition and K-medoids clustering,'' *Mech. Syst. Signal Process.*, vol. 93, pp. 16–29, Sep. 2017.
[18] S. Li, H. Wang, L. Song, P. Wang, L. Cui, and T. Lin, ''An adaptive data fusion strategy for fault diagnosis based on the convolutional neural network,'' *Measurement*, vol. 165, Dec. 2020, Art. no. 108122.
[19] M. Wang, J. Zhou, J. Gao, Z. Li, and E. Li, ''Milling tool wear prediction method based on deep learning under variable working conditions,'' *IEEE Access*, vol. 8, pp. 140726–140735, 2020.
[20] J. Schmidhuber, ''Deep learning in neural networks: An overview,'' *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.
[21] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, ''Deep learning and its applications to machine health monitoring,'' *Mech. Syst. Signal Process.*, vol. 115, pp. 213–237, Jan. 2019.
[22] R. Zhao, R. Yan, J. Wang, and K. Mao, ''Learning to monitor machine health with convolutional bidirectional LSTM networks,'' *Sensors*, vol. 17, no. 2, pp. 273–290, 2017.
[23] J. Zhang, P. Wang, R. Yan, and R. X. Gao, ''Deep learning for improved system remaining life prediction,'' *Procedia CIRP*, vol. 72, pp. 1033–1038, Jan. 2018.
[24] A. Z. Hinchi and M. Tkiouat, ''Rolling element bearing remaining useful life estimation based on a convolutional long-short-term memory network,'' *Procedia Comput. Sci.*, vol. 127, pp. 123–132, Jan. 2018.
[25] P. Peng, W. Zhang, Y. Zhang, Y. Xu, H. Wang, and H. Zhang, ''Cost sensitive active learning using bidirectional gated recurrent neural networks for imbalanced fault diagnosis,'' *Neurocomputing*, vol. 407, pp. 232–245, Sep. 2020.
[26] H. Liu, J. Zhou, Y. Zheng, W. Jiang, and Y. Zhang, ''Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders,'' *ISA Trans.*, vol. 77, pp. 167–178, Jun. 2018.
[27] J.-S. R. Jang, ''ANFIS: Adaptive-network-based fuzzy inference system,'' *IEEE Trans. Syst., Man, Cybern.*, vol. 23, no. 3, pp. 665–685, May/Jun. 1993.
[28] S. A. Khan, M. D. Equbal, and T. Islam, ''A comprehensive comparative study of DGA based transformer fault diagnosis using fuzzy logic and ANFIS models,'' *IEEE Trans. Dielectr. Elect. Insul.*, vol. 22, no. 1, pp. 590–596, Feb. 2015.
[29] C.-J. Lin, J.-Y. Jhang, S.-H. Chen, and K.-Y. Young, ''Using an interval type-2 fuzzy neural network and tool chips for flank wear prediction,'' *IEEE Access*, vol. 8, pp. 122626–122640, 2020.
[30] M. Motahari-Nezhad and S. M. Jafari, ''ANFIS system for prognosis of dynamometer high-speed ball bearing based on frequency domain acoustic emission signals,'' *Measurement*, vol. 166, Dec. 2020, Art. no. 108154.
[31] Z. Li, K. Goebel, and D. Wu, ''Degradation modeling and remaining useful life prediction of aircraft engines using ensemble learning,'' *J. Eng. Gas Turbines Power*, vol. 141, no. 4, Apr. 2019.
[32] S. D. Nguyen, S.-B. Choi, and T.-I. Seo, ''Recurrent mechanism and impulse noise filter for establishing ANFIS,'' *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 2, pp. 985–997, Apr. 2018.
[33] J.-J. Xiong and G. Zhang, ''Improved stability criterion for recurrent neural networks with time-varying delays,'' *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5756–5760, Nov. 2018.
[34] L. Wen, Y. Dong, and L. Gao, ''A new ensemble residual convolutional neural network for remaining useful life estimation,'' *Math. Biosci. Eng.*, vol. 16, no. 2, pp. 862–880, 2019.
[35] Y. Yu, C. Hu, X. Si, J. Zheng, and J. Zhang, ''Averaged bi-LSTM networks for RUL prognostics with non-life-cycle labeled dataset,'' *Neurocomputing*, vol. 402, pp. 134–147, Aug. 2020.

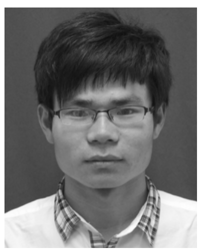
JIACHEN YAO received the B.S. degree in mechanical engineering from Nanjing University of Science and Technology, Nanjing, China, in 2015, where he is currently pursuing the Ph.D. degree in mechanical engineering.

His current research interests include machine learning, pattern recognition, mechanical fault diagnosis and prognosis, and remaining useful life prediction.



BAOCHUN LU received the Ph.D. degree in mechanical engineering from Nanjing University of Science and Technology, Nanjing, China, in 2002.

He is currently a Professor with Nanjing University of Science and Technology. His research interests include signal processing, sensor networks for the condition monitoring, and health diagnosis of dynamical systems.



JUNLI ZHANG received the B.S. degree in mechanical engineering from Northeast Agricultural University, Harbin, China, in 2015. He is currently pursuing the Ph.D. degree in mechanical engineering with Nanjing University of Science and Technology, Nanjing, China.

His current research interests include machine learning, pattern recognition, and robust adaptive control, and simulation analysis of electrohydraulic serve systems.