2019 IEEE Symposium Series on Computational Intelligence (SSCI)
December 6-9 2019, Xiamen, China

Remaining Useful Life Prediction of Machining Tools by 1D-CNN LSTM Network

Jiahe Niu, Chongdang Liu, Linxuan Zhang, Yuan Liao
Department of Automation
Tsinghua University
Beijing 100084, P. R. China
Email:{njh17; liucd16; lxzhang}@mails.tsinghua.edu.cn; thu_ly@126.com

***Abstract*—In the field of machining, machining tool life (degree of wear) is a key factor affecting the quality of the machined workpiece. Over-protection strategies may increase production costs and cause unnecessary machining tool downtime. Therefore, if the remaining useful life (RUL) of the machining tool can be accurately predicted, the work schedule will be effectively optimized and the machining tool procurement cost will be reduced. In this paper, we propose a system schema that integrates programmable logic controller (PLC) signals with sensor signals for online RUL prediction of machining tools. The preprocessed sensor signals are segmented and we propose ensemble discrete wavelets transform (EDWT) to eliminate the noise of three-dimensional vibration signals and get time-frequency information. Then statistics features are extracted based on time domain and frequency domain analysis. Further, we use spearman's coefficient, autocorrelation and monotonicity indicators for feature selection to reduce feature dimensions. Finally, we use a 1D-CNN LSTM network architecture for machining tools RUL prediction. The evaluation results show that our system schema is feasible for the industrial field, and has a better performance than other common methods.**

Keywords—machining tools; remaining useful life; 1D-CNN; LSTM.

I. INTRODUCTION

The failure of machining tools may result in an increase in the surface roughness and a decrease in dimensional accuracy of the workpiece, more seriously, the workpieces may be scrapped or the computer numerical control (CNC) machine may be damaged. Therefore, the remaining useful life (RUL) of the machining tool is a practical problem to be solved in the factory.

In recent years, studies on machining tools' RUL prediction can be divided into two types: model-based methods and data-driven methods. The model-based method mainly uses the domain knowledge and physical principle model of the system or component to perform calculations, and the failure behavior of the machining tools can be quantitatively characterized. However, model-based methods are often difficult to achieve due to the uncertainty of model parameters and the complexity of failure mechanisms of the machining tools in the cutting process [1]. In this case, data-driven methods are receiving more and more attention.

Data-driven methods can be divided into statistics-based methods, traditional machine learning methods, and deep learning methods. Common statistics-based methods include wiener processes [2], gamma processes [3], markov models [4], etc. Common methods based on traditional machine learning mainly include support vector regression (SVR) [5], artificial neural networks (ANN) [6], extreme learning machines (ELM) and neuro-fuzzy system [7]. In fact, for a large amount of data in the actual machining process, traditional machine learning algorithms are sometimes difficult to extract hidden information that characterizes the degradation process of the tool. In this respect, deep learning methods tend to have better effects, as it has powerful adaptive learning and anti-noise ability, and it can automatically extract deep features, which is more versatile than traditional machine learning methods. Common RUL prediction methods based on deep learning include recurrent neural network (RNN) [8], Long Short-Term Memory network (LSTM) [9], convolutional neural networks (CNN) [10], and deep belief networks (DBN) [11], etc.

Among the commonly used deep learning models, CNN has a very important position in the field of image recognition. Due to its capacity to automatically extract features, it is also being used in the field of fault diagnosis and process monitoring today [13]. Liang [14] used the one-dimensional convolutional neural network (1D-CNN) to extract the deep features of high-speed train fault signal which can achieve the classification accuracy of 96.4%. Turker [10] also used 1D-CNN on real motor data for real-time motor condition monitoring. And LSTM can effectively mine the hidden degradation trend in time series. Zheng [9] proposed an LSTM approach for RUL estimation. In fact, we can combine CNN's high-dimensional feature extraction capacity and LSTM's advantage on time series problems. After CNN extracts features, we input them into the LSTM for training, then some improvements in accuracy and speed can be achieved.

Based on the working condition information and sensor data collected by the programmable logic controller (PLC) and external sensors, this paper constructs the machining tool wear state evaluation and life prediction model to diagnose the wear state of the machining tool. First, data preprocessing is performed on the training dataset and the testing dataset, including denoising, outlier culling, and data structure defragmentation. The cleaned data is then decomposed by ensemble discrete wavelets transform (EDWT) to obtain a

978-1-7281-2485-8/19/$31.00 ©2019 IEEE

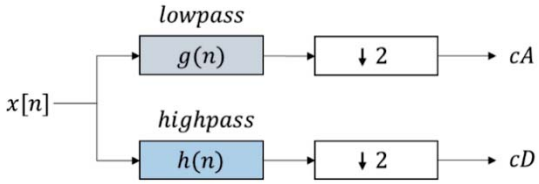

Fig. 1. Discrete wavelet transform.

time-frequency signal. Then, a series of statistical features such as mean, median, variance, energy, etc. are extracted according to the time-frequency signal. After the statistical features are acquired, the optimal feature selection is performed by using the spearman's coefficient, autocorrelation and monotonicity indicators to obtain the feature set with better performance. Finally, we constructs a 1D-CNN LSTM network to predict the remaining life of the machining tools.

II. FEATURE ENGINEERING

A. Sensor signal preprocessing based on EDWT

Discrete Wavelet Transform (DWT) [19] is a signal analysis algorithm with time domain and frequency domain analysis capabilities, which is achieved by discretization of continuous wavelet transform. And it is often used to feature mining of mechanical vibration signals. Each stage of discrete wavelet transform is mainly related to two filters as shown in "Fig. 1": low-pass filters $g(n)$and high-pass filter $h(n)$. When decomposing, the signal pass $g(n)$ and $h(n)$ respectively, and then performs 1/2 downsampling to get coefficient cA and coefficient cD [15].

The meanings of the symbols are as follows:

$x[n]$: Input signal.

$g[n]$: Low-pass filter that filters out the high-frequency portion of the input signal and outputs the low-frequency portion.

$h[n]$: High-pass filter, in contrast to the low-pass filter, filters out the low-frequency part and outputs the high-frequency part.

$\downarrow Q$: Downsampling filter, if $x[n]$ is used as input, output $y[n] = x[Qn]$. Here Q is 2.

Due to the influence of the actual processing environment and the defects of the sensor devices, the obtained tool data may contain a variety of noises. With reference to the design of ensemble empirical mode decomposition (EEMD) [16], this paper proposes an improved DWT method EDWT. During each iteration, it adds different white noise to the original vibration signal, then DWT decomposition is performed on the generated data. Finally, we calculate the average of all iteration results. The following figure shows the process of signal preprocessing based on EDWT. The specific steps are shown in "Fig. 2".

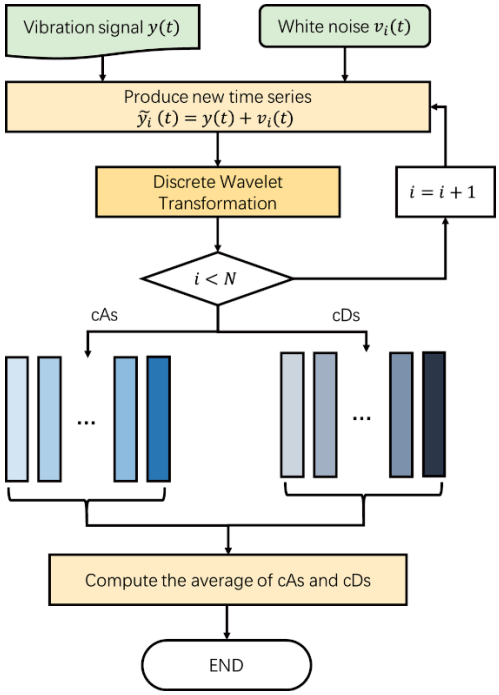

Fig. 2. Ensemble eiscrete wavelets transform.

B. Feature extraction

In this paper, classical features and Trigonometric features are extracted from the time-frequency domain signal [1][15]. The extracted features can be divided into time domain features and frequency domain features. Mean value (MV), variance value (VV), median value (MDV), mean square error (MSE), square mean root (SMR), root mean square (RMS), maximum absolute value (MA), energy value and margin factor (MF) are features in the time domain. Root mean square frequency (RMSF) and variance frequency (VF) are features in the frequency domain. The trigonometric features include the standard deviation of inverse hyperbolic sine (asinh) and standard deviation of inverse tangent (atanh). The calculating equations of different features are shown in "Table I".

C. Feature selection

For RUL prediction in machining tools, the number of extracted features are sizeable, and there are some irrelevant and redundant features. So the selection of features is very important for improving the prediction results. This paper selects the features by the following three indicators, which are widely used in the field of prognostic and health management[1][15]:

The first indicator is Spearman's coefficient [18].

$$R_s(x_i) = \frac{\left|\sum_{j=1}^{N}(t_j-\bar{t_j})(x_i(t_j)-\bar{x_i})\right|}{\sqrt{\sum_{j=1}^{N}(t_j-\bar{t_j})^2 \sum_{j=1}^{N}(x_i(t_j)-\bar{x_i})^2}} \quad (1)$$

TABLE I. STATISTICS FRA TURES TABLE

Feature Name	Equation
MV	$x_1 = \sum_{i=1}^{N} y(i)/N$
VV	$x_2 = \sum_{i=1}^{N} (y(i) - x_1)^2/N$
MDV	$x_3 = median(y(i))$
MSE	$x_4 = \sqrt{\sum_{i=1}^{N} (y(i) - x_1)/N}$
SMR	$x_5 = \left(\sum_{i=1}^{N} \frac{\lvert y(i)\rvert}{N}\right)^2$
RMS	$x_6 = \sqrt{\sum_{i=1}^{N} (y(i))^2/N}$
MA	$x_7 = max\lvert y(i)\rvert$
Energy	$x_8 = \sum_{i=1}^{N} y(i)^2$
Kurtosis	$x_8 = \frac{\sum_{i=1}^{N}(y(i) - x_1)^4}{(N-1)\sigma^4}$
Entropy	$x_{10} = -\sum_{i=1}^{K} p(y_k)\, log_b\, p(y_k)$
CF	$x_{11} = x_7/x_6$
MF	$x_{12} = x_7/x_5$
RMSF	$x_{13} = \sqrt{\frac{\sum_{i=2}^{N} \dot{y}^2(i)}{4\pi^2 \sum_{i=1}^{N} y^2(i)}}$
VF	$x_{14} = \frac{\sum_{i=2}^{N} \dot{y}^2(i)}{4\pi^2 \sum_{i=1}^{N} y^2(i)} - \frac{\sum_{i=2}^{N} \sum_{i=2}^{N} \dot{y}(i) y(i)}{2\pi \sum_{i=1}^{N} y^2(i)}$
SD of asinh	$x_{15} = \sigma(log[y_i + (y_i^2 + 1)^{1/2}])$
SD of atan	$x_{16} = \sigma((i/2) log((i + y_j)/(i - y_j)))$

The second indicator is Autocorrelation [1].

$$R_A(x_i) = \frac{\sum_{i=1}^{N} \left(x_i(t_j) - x_i(t_{j-1})\right)^2}{N} \tag{2}$$

The third indicator is Monotonicity [1].

$$\mathrm{R_M}(x_i) = \frac{1}{N-1} \left| \sum_{i=1}^{N} \delta\left(x_i(t_{j+1}) - x_i(t_j)\right) - \sum_{i=1}^{N} \delta(x_i(t_j) - x_i(t_{j+1})) \right| \tag{3}$$

In these equations, x_i is the ith extracted statistics feature. Finally, this paper selects the top 30 attributes of each indicator, and combines the results of the three indicators as a collection of high-quality features.

III. 1D-CNN LSTM NETWORK FOR RUL ESTIMATION

A. 1D-Convolutional and pooling

The data collected during the actual machining of the tool can be represented as a two-dimensional matrix with a time axis and a sensor variable axis. For time-series problems, one-dimensional convolutional neural network (1D-CNN) is more suitable than common convolution neural network. One of the characteristics of the 1D-CNN is that for time-series data, the receptive field moves only in the direction of time, so the local inter-variable correlation can be extracted. [13]. "Figure. 3" shows the architecture of the 1D-CNN. Each convolutional layer consists of several convolutional units whose parameters are optimized by backpropagation algorithms.

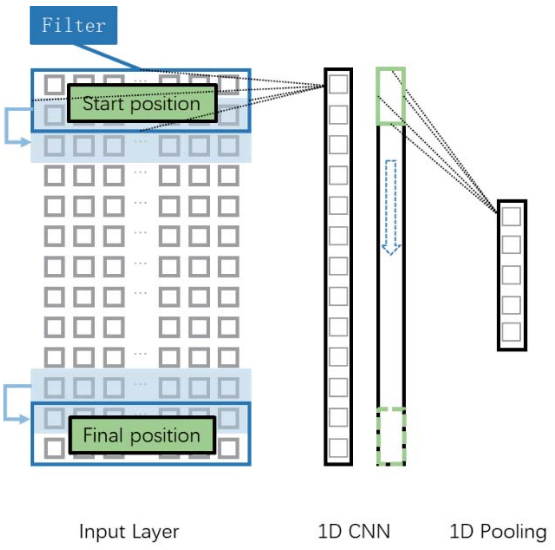

Fig. 3. One-dimensional convolutional neural network.

The process of the convolution layer can be represented as [14]:

$$x_j^l = f(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l) \tag{4}$$

Where x_i^{l-1} is the input, k_{ij}^l is the kernel weights, b_j^l is the biases, $f(\bullet)$ is the activation function, x_j^l is the output of the j-th kernel in the l-th convolutional layer.

The feature map obtained by the convolution layer continues to carry pooling operation with maximum pooling function, which can effectively reduce the amount of data and increase the calculation speed. The max pooling function's formula can be represented as:

$$\widehat{x_j^l} = f(\beta_j^l down(x_j^l) + c_j^l) \tag{5}$$

Where $\widehat{x_j^l}$ is the input from convolution layer, β_j^l is the weight matrix, $down(\bullet)$ is the down sampling function, c_j^l is the baise. $\widehat{x_j^l}$ is the output of the j-th kernel in the l-th pooling layer.

B. LSTM

Long short term memory network (LSTM) [20] is a special type of recurrent neural network (RNN) structure. In traditional RNN, each unit of RNN is a simple chain structure, it processes the input sequence $\{x_1, x_2, \ldots, x_T\}$ sequentially to

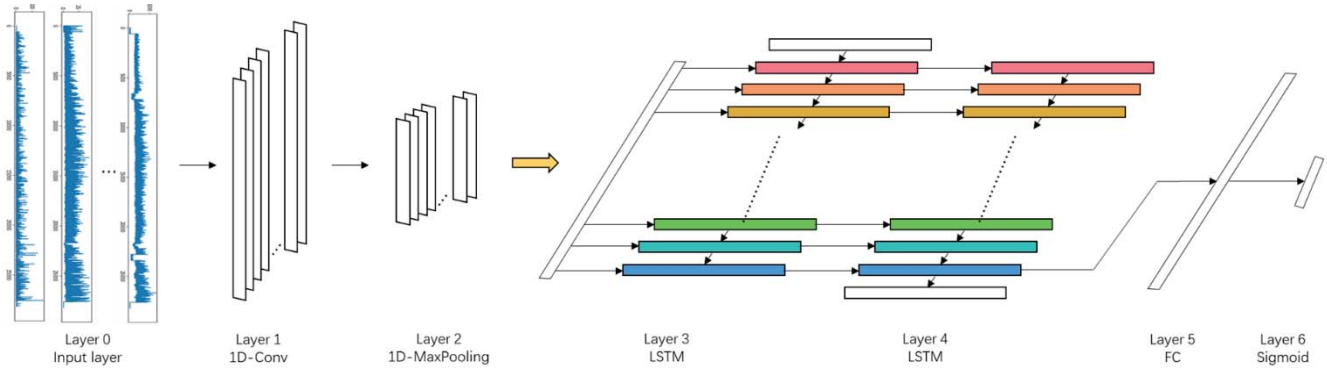

Fig. 4. 1D-convolutional LSTM Network architecture.

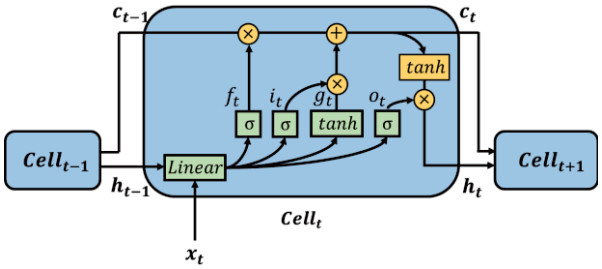

Fig. 5. Long short term memory network cell.

construct a corresponding sequence of hidden states $\{h_1, h_2, \ldots, h_T\}$. In LSTM, a mermory cell c_t is introduced in addition to the hidden state h_t at the timestep t [17].

As shown in "Figure. 5", x_t, h_t, and c_t represent the input, output, and status information of the $cell_t$. They are computed via three gate functions, including forget gate, input gate, and output gate. The forget gate function f_t realizes the information discarding of the state from the previous memory cell $cell_{t-1}$, the input gate controls the information that will participate in the calculation of the memory cell $cell_t$, and the output gate determines the output information of the memory cell $cell_t$. Due to the gate structure, the LSTM has the ability to "memorize" and thus exhibits better performance in processing some time series data. We use f, i and o to represent the forget function, the input gate function and the output gate function respectively. The subscript of the parameter W, h indicates which gate it corresponds to. According to the above notations, an LSTM is formally defined as:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \quad (6)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \quad (7)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \quad (8)$$

$$g_t = tanh(W_{xg}x_t + W_{hg}h_{t-1} + b_g) \quad (9)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \quad (10)$$

$$h_t = o_t \odot \tanh(c_t) \quad (11)$$

Where $\sigma(\bullet)$ is the sigmoid function and $\odot$ is the element-wise multiplication.

C. 1D-convolutional LSTM Network architecture

The architecture of 1D-CNN LSTM network consists of 6 layers as shown in "Fig. 4": 1 convolutional layer, 1 max-pooling layer, 2 LSTM layers, 1 fully connected layer and 1 activation layer. First, the original signals are passed into the first one-dimensional convolution layer after feature extraction, feature selection, and time window division. The filter and kernel size of the one-dimensional convolution layer are both 32. The size of the time window taken in the data preprocessing process is 200, so after passing through the layer of convolution the output shape is changed to (None, 169, 32). The second layer is the one-dimensional max pooling layer. The pooling size of this layer is 40, and the output shape of this layer is (None, 4, 32). Layer 3 and 4 are the LSTM layers with 64 neurons and 16 neurons, and output shape after these two layers is (None, 16). Layer 5 is a fully connected and dropout layer which consists of 1 neuron. The main purpose of the dropout layer is to reduce over-fitting, and in the fully connected layers, a sigmoid activation function is been used at the end for RUL prediction.

IV. Experiment

The proposed RUL prediction system is tested using CNC machining tools data collected from actual machining process, including the data from PLC and vibration sensor. We compare the results of 1D-CNN, LSTM and 1D-CNN LSTM networks after DWT transformation and EDWT transformation respectively.

A. Data set description

The dataset comes from the 2nd Industrial Big Data Innovation Competition organized by China Academy of Information and Communications Technology. According to the PLC and external sensor in cyber-physical systems, the working condition information and sensor data during the machining process are collected to achieve the online monitoring and remaining useful life prediction of tool wear. PLC data is the complete processing history data, including

Fig. 6. The installation location of the vibration sensor.

TABLE II. FEATURE FIELD DESCRIPTION

Data Sources	*Feature*	*Description*
PLC data	time	Recording time
	spindle_load	Spindle load
	x	X-axis mechanical coordinates
	y	Y-axis mechanical coordinates
	z	Z-axis mechanical coordinates
	csv_no	Corresponding sensor_file
Sensor data	vibration_1	X-axis vibration signal
	vibration_2	Y-axis vibration signal
	vibration_3	Z-axis vibration signal
	current	First phase current signal

recording time, spindle load, X-axis vibration, Y-axis vibration, Z-axis vibration and other information. Since the amount of raw signal data of the sensor is extremely large, one minute of data is taken every 5 minutes to form one csv file. In terms of the data sampling frequency, the sampling frequency of the PLC signal is 33 Hz, and the sampling frequency of the vibration sensor is 25600 Hz.

The data set contains a total of 7 sets of CNC machining data from the actual CNC machining process from the start of a new tool until the end of the tool life. Three of them are training sets, and the remaining four are testing sets.

The installation location of the vibration sensor is shown in "Fig. 6", where the angle of view is in the direction of the front of the machine. The fields included in the PLC and Sensor data are shown in "Table II".

B. Data preprocessing

The sampling frequency of the PLC signal is 33 Hz, and the sampling frequency of the vibration sensor is 25.6 KHz. Therefore, it is necessary to first align them in the time dimension. This article chooses to align them evenly. In addition, some abnormal values exist in the original data. For example, the current contains abnormal maximum value and abnormal minimum value (absolute value is even greater than 10^{10}), these values may affect the extraction of statistics. So this article discards the row containing the kind of outliers (this article divides the outliers with the absolute value 10^2 as the limit).

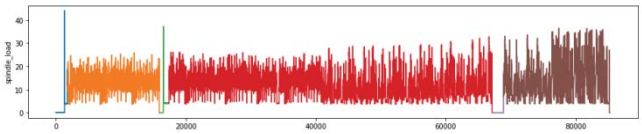

Fig. 7. Spindle load variable.

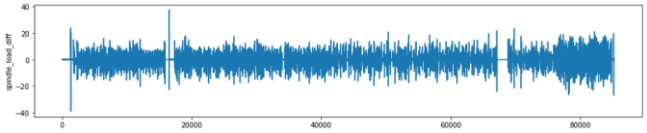

Fig. 8. Difference of the spindle load variable.

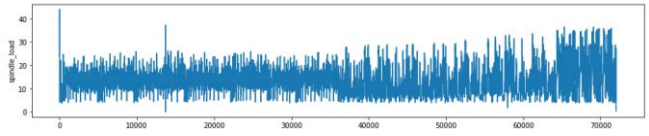

Fig. 9. Spindle load in normal processing state.

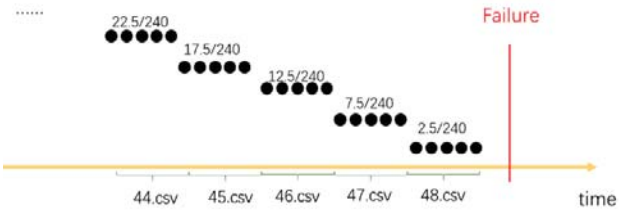

Fig. 10. The label of remaining life ratio.

According to the data provided in the PLC, the spindle load change of the tool during the actual machining process can be obtained. As shown in "Fig. 7", it can be found that there is intermittent load value at a lower value, the data of this part may not work in the normal processing state, so it needs to be eliminated. The solution adopted in this paper is to calculate the absolute value of the difference of the spindle load variable in the time dimension as shown in "Fig. 8", and then add the value to the spindle load value. If the sum is less than the threshold value 5, it is considered that the current time tool is not working in the normal processing state. "Fig. 9" represents the result of the data culling by the threshold, and it can be found that the data of the abnormal working state is substantially eliminated.

The processed data is then evenly divided into 600 parts (each corresponding time is about 0.1 seconds), and the time-frequency feature is extracted after performing EDWT transformation on each data segment. Finally, the size of the

feature matrix extracted by each tool is 600*k*n, where k is the number of .csv files, n is the number of extracted features.

Since the data in each csv is fragment data of 1 minute every 5 minutes, when the training set is subjected to RUL labeling, as shown in "Fig. 10", the 5 minutes expected value is used as the label of the entire csv. For example, the RUL value corresponding to the last csv file of each tool in the training set should be 2.5 minutes. For the three tools in the training set, the complete life cycle is 240mins, 240mins, 180mins, even if

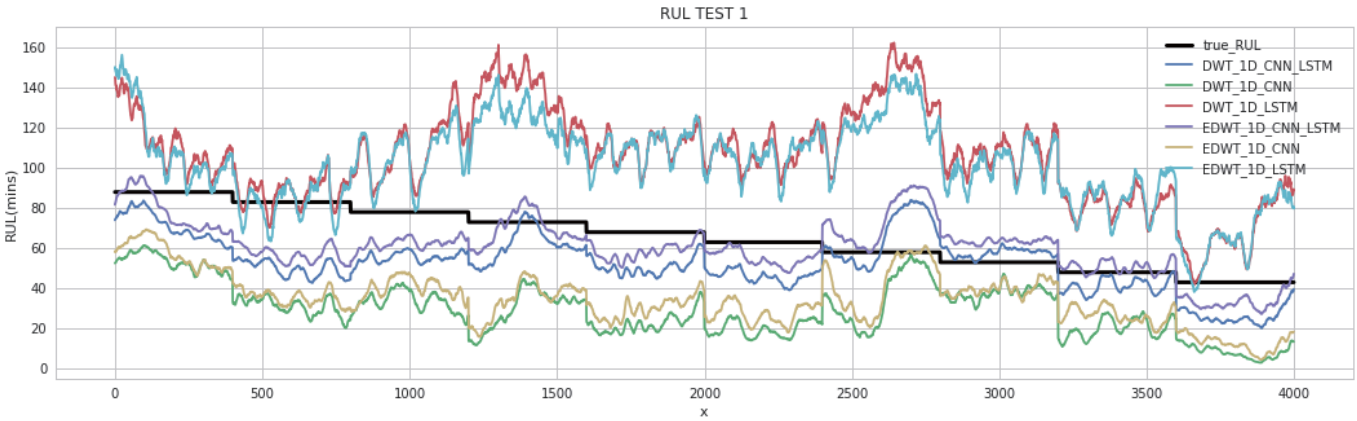

Fig. 11. Results of the first machining tool in the test data set using different methods.

they have same remaining life of 100mins, their tool wear states are not consistent . Therefore, the concept of remaining life ratio is introduced here. We normalize the remaining useful life of each tool and renormalize it to obtain the true remaining useful life of the tool.

C. Results

The remaining life ratio of the machining tool's begin and end of the prediction can be obtained from the 1DCNN-LSTM network. Then the remaining life of the test set can be calculated by the following formula:

$$RUL = \frac{5*R_e*(n_{csv}-1)}{R_b-R_e} \tag{12}$$

Where R_e is the remaining life ratio of the predicted end time, R_b is the remaining life ratio of the predicted begin time, n_{csv} is the number of csv files.

The scoring function given by the organizer is:

$$A_i = \begin{cases} exp^{-ln(0.5)*(\frac{Er_i}{5})} & ,if\ Er_i \le 0 \\ exp^{+ln(0.5)*(\frac{Er_i}{20})} & ,if\ Er_i > 0 \end{cases} \tag{13}$$

$$Er_i = true_i - pred_i \tag{14}$$

Where i is the i-th tool of the test, $true_i$ is the true value of the i-th tool of the test, $pred_i$ is the predicted value of the i-th tool of the test. The curve of the score function is shown in "Fig. 12".

In the aspect of comparative experiments, we compares the results of 1D-CNN, LSTM and 1D-CNN LSTM networks after DWT transformation and 1D-CNN, LSTM and 1D-CNN LSTM networks after EDWT transformation. The model parameters of CNN and LSTM are manually searched to get better results. Since the scoring function is sensitive to prediction results and the network output has some volatility,

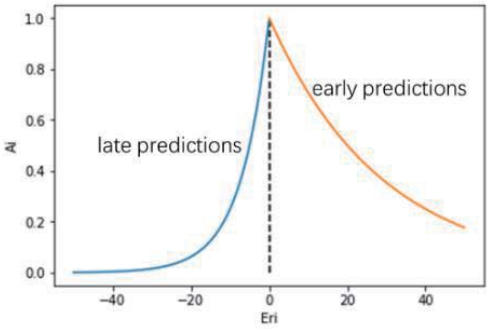

Fig. 12. Score function.

we need to use multiple experimental results to evaluate the predictive ability of different methods. This paper conducts training and testing for 20 times each scheme. And "Fig. 11" shows the mean of 20 times predictions of each methods, which based on the first machining tool in the test dataset. The x-axis represents the line number of the test data, the y-axis represents the predicted RUL in minutes and the black line is the true value.

We use two evaluation indicators in our experiment, which are the score from organizer and the mean square error(MSE). For each indicator, we calculate the results' mean and standard deviation after 20 experiments. The results are shown in detail

in “Table III” as “MSE mean” and “Score mean”, etc. In the case of using score as the evaluation indicator, we can see that the 1D-CNN LSTM model using EDWT has the highest average score, and the standard deviation is very close to the lowest value, indicating that EDWT can effectively improve the accuracy while ensuring the stability of the prediction. When using mean square error as the evaluation indicator, the 1D-CNN LSTM model using EDWT completely defeated other methods. And regardless of which feature extraction method is used, the network structure using 1D-CNN LSTM has better performance than the 1D-CNN model or the LSTM

TABLE III. 20 TIMES TRAINING-TESTING RESULTS

Preprocess	*Model*	*Score mean*	*Score std*	*MSE mean*	*MSE std*
EDWT	CNN LSTM	**0.6406**	**0.1889**	**12.335**	**2.503**
	CNN	0.4313	0.2158	62.036	73.401
	LSTM	0.3671	0.3056	32.456	18.044
DWT	CNN LSTM	0.5454	0.2541	15.106	4.315
	CNN	0.3854	**0.1860**	39.152	26.460
	LSTM	0.4385	0.3037	22.653	8.697

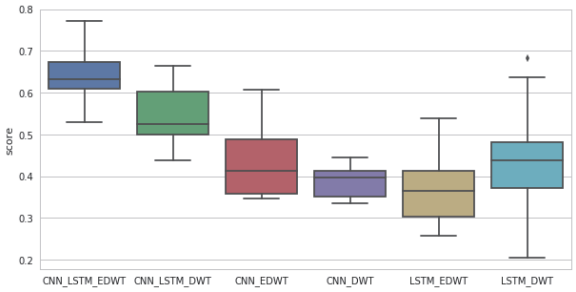

Fig. 13. Boxplot of different methods.

model alone. These results are shown in “Fig. 13” in the form of a box diagram.

V. CONCLUSIONS

In this paper, we propose a system schema that integrates PLC signals with sensor signals for online RUL prediction of machining tools. The preprocessed sensor signals are segmented and we use EDWT to eliminate the noise of three-dimensional vibration signals. Then statistics features are extracted based on time domain and frequency domain analysis. Further, we use spearman’s coefficient, autocorrelation and monotonicity indicators for feature selection to reduce feature dimensions. Finally, we propose a 1D-CNN LSTM network architecture for machining tools RUL prediction. The evaluation results show that our system schema is feasible for the industrial field, and has a better performance than other methods.

In the future, the proposed method can also add more analysis of PLC signals. Thus, better results may be obtained by analyzing the working conditions. At the same time, the practicality of the system can be improved by automatic parameter optimization.

ACKNOWLEDGMENT

The authors would like to thank China Academy of information and Communications Technology, Foxconn, CyberInsight for providing the raw data of CNC cutting tools. Moreover, assistance provided by Dr. Zhao was greatly appreciated.

REFERENCES

[1] J. Wu, Y. Su, Y. Cheng, X. Shao, C. Deng, and C. Liu, “Multi-sensor information fusion for remaining useful life prediction of machining tools by adaptive network based fuzzy inference system,” *Applied Soft Computing,* vol. 68, pp. 13-23, 2018.

[2] Q. Zhai and Z. Ye, “RUL Prediction of Deteriorating Products Using an Adaptive Wiener Process Model,” *IEEE Transactions on Industrial Informatics*, vol. 13, no. 6, pp. 2911-2921, 2017.

[3] Q. Wei and D. Xu, “Remaining useful life estimation based on gamma process considered with measurement error,” in *2014 10th International Conference on Reliability, Maintainability and Safety (ICRMS)*. 2014, pp. 645-649.

[4] T. T. Le, C. Berenguer, and F. Chatelain, “Multi-branch Hidden semi-Markov modeling for RUL prognosis,” in *2015 Annual Reliability and Maintainability Symposium (RAMS)*. 2015, pp. 1-6.

[5] Y. Guo, “MKLS-SVR based remaining useful life prediction for avionics,” in *2015 12th IEEE International Conference on Electronic Measurement & Instruments (ICEMI).* 2015, pp. 223-227.

[6] P. Lall, S. Deshpande, and L. Nguyen, “ANN based RUL assessment for copper-aluminum wirebonds subjected to harsh environments,” in *2016 IEEE International Conference on Prognostics and Health Management (ICPHM),* 2016, pp. 1-10.

[7] X. Li, “Remaining Useful Life Prediction of Bearings Using Fuzzy Multimodal Extreme Learning Regression,” in *2017 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC).* 2017, pp. 499-503.

[8] Ü. Şentürk, I. Yücedağ, and K. Polat, “Repetitive neural network (RNN) based blood pressure estimation using PPG and ECG signals,” in *2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT).* 2018, pp. 1-4.

[9] S. Zheng, K. Ristovski, A. Farahat, and C. Gupta, “Long Short-Term Memory Network for Remaining Useful Life estimation,” in *2017 IEEE International Conference on Prognostics and Health Management (ICPHM).* 2017, pp. 88-95.

[10] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, “Real-Time Motor Fault Detection by 1-D Convolutional Neural Networks,” *IEEE Transactions on Industrial Electronics*, vol. 63, no. 11, pp. 7067-7075, 2016.

[11] D. A. Tobon-Mejia, K. Medjaher, and N. Zerhouni, “CNC machine tool's wear diagnostic and prognostic by using dynamic Bayesian networks,” *Mechanical Systems and Signal Processing*, vol. 28, pp. 167-182, 2012.

[12] F. Pacheco, M. Cerrada, R.-V. Sanchez, D. Cabrera, C. Li, and J. V. de Oliveira, “Attribute clustering using rough set theory for feature selection in fault severity classification of rotating machinery,” *Expert Systems with Applications*, vol. 71, pp. 69-86, 2017.

[13] K. B. Lee, S. Cheon, and C. O. Kim, “A Convolutional Neural Network for Fault Classification and Diagnosis in Semiconductor Manufacturing Processes,” *IEEE Transactions on Semiconductor Manufacturing*, vol. 30, no. 2, pp. 135-142, 2017.

[14] K. Liang, "1D Convolutional Neural Networks For Fault Diagnosis of High-speed Train Bogie," in *2018 IEEE 23rd International Conference on Digital Signal Processing*. 2018, pp. 1-5.

[15] K. Javed, R. Gouriveau, N. Zerhouni, and P. Nectoux, "Enabling Health Monitoring Approach Based on Vibration Data for Accurate Prognostics," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 1, pp. 647-656, 2015.

[16] X. Chen, G. Cai, H. Cao, and W. Xin, "Condition assessment for automatic tool changer based on sparsity-enabled signal decomposition method," *Mechatronics,* vol. 31, pp. 50-59, 2015.

[17] Z. Li, H. Di, F. Tian, W. Chen Q. Tao, L. Wang, and T. Liu, "Towards Binary-Valued Gates for Robust LSTM Training," *arXiv preprint arXiv*, 2018.

[18] X. Liu, P. Song, C. Yang, C. Hao, and W. Peng, "Prognostics and Health Management of Bearings Based on Logarithmic Linear Recursive Least-Squares and Recursive Maximum Likelihood Estimation," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 2, pp. 1549-1558, 2018.

[19] M. J. Shensa, "The discrete wavelet transform: wedding the a trous and Mallat algorithms," *IEEE Transactions on Signal Processing*, vol. 40, no. 10, pp. 2464-2482, 1992.

[20] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997.