The International Journal of Advanced Manufacturing Technology (2018) 98:3217–3227
https://doi.org/10.1007/s00170-018-2420-0

ORIGINAL ARTICLE

Tool condition monitoring using spectral subtraction and convolutional neural networks in milling process

Fatemeh Aghazadeh[1] · Antoine Tahan[1] · Marc Thomas[1]

Received: 25 March 2018 / Accepted: 16 July 2018 / Published online: 31 July 2018

Abstract
Process monitoring is necessary in machining operation to increase productivity, improve surface quality, and reduce unscheduled downtime. Tool wear and breakage are important and common source of machining problems due to high temperatures and forces of the machining process. Therefore, it is highly beneficial to develop an online tool condition monitoring (TCM) system. This paper investigates a robust tool wear monitoring system for milling operation. Recent developments in machine learning, in particular deep learning methods, result in significant improvement in automation of different industries. Therefore, in this research, we employed convolutional neural network (CNN) as a well-established and powerful deep learning algorithm for tool wear estimation. Wavelet packet-based features are extracted for tool wear monitoring as a powerful time-frequency fault indicator. Moreover, a hybrid feature extraction method is proposed using wavelet time-frequency transformation and spectral subtraction algorithms to intensify the effect of tool wear in the signal and reduce the effect of other cutting parameters. CNN-based monitoring systems are compared with three other machine learning methods (support vector machine, Bayesian rigid network, and K nearest neighbor method) as the baseline. The research is validated using different datasets. The algorithms are implemented and compared using experimental force and vibration signals from LIPPS lab of ETS university as well as using current signals as the fault indicator from Nasa_Ames dataset.

Keywords Tool condition monitoring · Deep learning · Spectral subtraction · Tool wear · Wavelet transform

1 Introduction

Machining processes are fundamental part of today's competitive manufacturing industries. Due to the need for higher productivity, higher quality parts, and lower manufacturing cost, there is growing demand to make the machining operation totally automatic. Along with other directions in automation, it is necessary to automatically monitor machining online to assure the production safety and quality. Tool defects can be considered one of the most common and costly faults of the machining process. Due to contact forces and friction between cutting tool and workpiece, high temperatures in the cutting area and pressure of the chips on the tool, various defects may happen to the tool which deteriorates the surface finish or causes damage or breakage to the tool, workpiece, or machining center [1]. Therefore, there is high demand to design a reliable and robust online automatic TCM system to actively monitor the cutting process and provide live reports of tool condition status.

TCM methods can be categorized into two main groups: direct and indirect methods. Direct methods measure actual value of faults with sensors such as laser, optical and ultrasonic. Another approach is used in indirect methods by employing physical parameters of the system such as force and vibration to represent tool condition indirectly [2]. Although direct measurement methods estimate tool fault with high accuracy, they are still expensive to implement and not suitable for online applications in industrial environments. However, indirect methods can be used to

✉ Fatemeh Aghazadeh
fatemeh.aghazadehkouzekonani.1@ens.etsmtl.ca

Antoine Tahan
antoine.tahan@etsmtl.ca

Marc Thomas
marc.thomas@etsmtl.ca

1 Département de Génie Mécanique, École de Technologie Supérieure, 1100 Notre-Dame St W, Montréal, Québec H3C 1K3, Canada

fulfill TCM purposes as an alternative with accurate results and acceptable cost by using a proper descriptor signal and an appropriate modeling method [3]. Moreover, the same sensor can be used for multiple monitoring tasks.

Applicable and informative signals that are widely used for TCM includes force, vibration, acoustic emission, current, and power signals. Li et al. studied TCM for turning process by employing force signals as the fault indicator [4]. They extracted 14 time domain features from the force signals and, using v-support vector regression, developed a model for flank wear estimation. While the force signal shows promising behavior to represent tool wear variations during the machining process, it is also highly dependent on other operating conditions and relatively expensive for industry use [3]. Vibration sensors are also practical in industrial environments. For example, Harun et al. studied TCM during deep twist drilling process and compared vibration and force signals for this purpose using time and frequency domain fault descriptors.Their study suggests that both sensors are capable of performing this task, but they recommended vibration signal as the superior fault indicator [5]. Acoustic emission is also practical and informative signal which is highly used in the literature for TCM [6, 7]. Soltani Rad et al. also employed spindle current as an economic and practical indicator signal for tool breakage detection in milling process. In this paper, the authors applied least squares support vector machine (LS-SVM) as the classifier and achieved acceptable results for monitoring tool breakage [8]. To make monitoring systems more robust, sensor fusion is a powerful approach. Sensor fusion refers to combining the information of more than one sensor in a complementary way to enhance the accuracy and reliability of the system. For example, Segreto et al. employed cutting force, acoustic emission, and vibration signals for tool condition assessment in turning process. Signals are fused in feature level after processing and the results are fed to a neural network [9].

Signal processing is the next step after choosing appropriate sensors and signal acquisition to magnify the effect of monitoring parameters by removing noises. Time domain analysis, frequency domain analysis, and time-frequency domain analysis are three common approaches for this step [10]. While many researches are devoted in time domain and frequency domain analyses due to low complexity, time-frequency analysis is well suited for this application as it examines both time variant and frequency characteristics of the signal simultaneously. Based on the non-stationary nature of faulty signals, time-frequency analysis can provide discriminative information about machinery health conditions. Therefore, discriminative fault features can be extracted from a faulty signal by choosing a proper time-frequency method [11]. Rehorn et al. utilized s-transform as a time-frequency transformation method and proposed a time-frequency domain feature, selective regional correlation, for machining condition monitoring [12]. In another research, a comparative study is performed between five time-frequency transformation methods for the purposes of TCM in milling operation [13].

The signal representation in time-frequency domain has high dimensions. Therefore, after time-frequency step, dimensionality reduction methods are useful. Dimensionality reduction methods such as principal component analysis (PCA) and linear discriminant analysis (LDA) are popular among the literature to perform this task [14–16]. Spectral substraction is another method which can be implemented to enhance signal quality. It it originally used in speech enhancement to remove the effects of steady sounds in the environment [17, 18]. Similar to the sound and speech analysis applications, this method is employed as a noise reduction tool. Fault diagnosis applications can use it to reduce the steady-state part of the signal and present fault characteristics. For example, El Bouchikhi et al. proposed an algorithm for fault diagnosis of induction machine bearings using spectral subtraction method. In this study, stator current frequency response of the healthy machine is subtracted from spectrum of machine's current acquired signal to present better fault indicators [19].

During the machining process, various parameters such as operational conditions, depth of cut, feed rate, and workpiece material are changing which may degrade the monitoring performance and can reduce system robustness. Moreover, the relation between signals and monitoring parameters are often nonlinear and complex. Therefore, powerful methods are required to perform the decision-making task. Many methods such as artificial neural network (ANN), support vector machine (SVM), and Bayesian networks are employed to perform this task in the literature. Patra et al. investigated tool wear during micro drill using thrust force signals and ANN method [20]. In another study, discrete wavelet transform (DWT) and SVM are used along with sound signals for tool condition monitoring in face milling [21]. Tobon-Mejia used Bayesian network method for TCM and the estimation of its remaining useful life (RUL) in machining process [22].

Recently, powerful characteristics of deep learning methods draw attention of researchers in different fields and helped them to solve many challenges in machine learning domain [23]. Deep learning refers to machine learning techniques with deep architectures and multiple layers which enable them to learn highly complex relationships from even low-processed to raw signals [24]. In an era in which sensors are actively producing high amounts of data, such techniques are able to make the most information out of the big data. They are therefore less dependent on specific applications and frameworks and have powerful characteristics to outperform other methods when the relationship between the input data and desired output is complex [25]. Despite their high potential, they are relatively recent in the field of machinery condition monitoring. Zhao et al. employed long short-term memory networks (CBLSTM) for tool condition monitoring in milling process [26]. Jing et al. developed a CNN-based method for gearbox condition monitoring using frequency data of vibration signals and their method outperformed some of the common machine learning algorithms [27]. In another study, vibration signals of a gerbox system are pre-processed using statistical measures from the time domain and frequency band energy from frequency domain. Then, the feature vector is fed to CNN to train it to detect gearbox faults [28]. Based on high potentials of deep learning methods, further research is crucial to examine them with different signals and levels of signal processing in TCM applications.

In this study, a TCM system is proposed using convolutional neural network as a powerful and established deep learning method. In the first step, force signals and vibration signals from ETS experimental dataset are selected independently to develop the monitoring system. Wavelet packet transform is employed for signal processing step to transform the signal to time-frequency domain. The final step is the machine learning algorithm which the proposed method is compared with three common methods in the literature, support vector regression, Bayesian rigid regression, and nearest neighbor regression methods. In the next step of the study, spindle current signal from Nasa_Ames dataset [29] is used for further validation of the method. Spectral subtraction is an ideal candidate to process current signals as it is power based and not a vectorial format in contrast to force and vibration. . Spectral subtraction method is applied around tooth path frequency. Afterwards, further noise processing is performed on the output of spectral subtraction and a number of features are generated to represent the fault in the signals. Finally, the comparative study between the machine learning algorithms is performed to see the results with a different dataset, signal, and higher levels of signal processing.

This paper is organized as follows: Section 2 represents the backgrounds and formulation of the algorithms which are used in this paper. The proposed algorithm is explained in Section 3. Section 4 introduces the two datasets which are used in this study for validation of the work. Results and discussion are presented in Section 5 and Section 6 is dedicated to conclusion.

2 Background of methods

This section presents the formulation and background of the algorithms and techniques which are used in this paper.

2.1 Wavelet transform

Wavelet transform is one of the methods that is widely used for fault diagnosis and health condition monitoring. The main difference between wavelet transform (WT) and fast Fourier transform (FFT) is that in wavelet transform, wavelets are used as the basis instead of sinusoidal functions that are used in fast Fourier transforms. It is an effective tool for transient signal analysis as well as time-frequency localization since it adds a scale variable in addition to the time variable in the inner product transform. It has a better time localization but a lower frequency resolution for higher frequency components. In contrast, for lower-frequency components, the frequency resolution is higher while the time localization is worse. Following equation describes the formulation of the continuous wavelet transform [11].

$$\mathrm{WT}_x(t,a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(u)\, \psi \frac{(u-t)}{a}\, du \tag{1}$$

where wavelet $\psi(u-t)/a$ is derived by dilating and translating the wavelet basis $\psi(t)$, and $1/\sqrt{a}$ is a normalization factor to maintain energy conservation and $a > 0$.

2.2 Spectral subtraction

Spectral subtraction is a method which was originally used for speech signal enhancement. A signal is considered a combination of noise and clean speech; therefore, the noise spectrum is estimated during speech pauses, and an estimation of the noise spectrum is subtracted from the noisy speech spectrum to obtain clean speech. It can be used in fault diagnosis applications by removing the steady-state part of the spectrum from signals to obtain their anomalies and fault signatures. Consider a measured signal which consists of the steady-state normal component and additive fault [18, 19]:

$$y[n] = s[n] + d[n] \tag{2}$$

where $y[n]$, $s[n]$, and $d[n]$ are the sampled measured signals, fault, and steady-state component, respectively. The frequency domain representation of the signal is given by the following:

$$Y(jw) = S(jw) + D(jw) \tag{3}$$

Therefore, the fault component of the signal can be obtained based on the following equation:

$$\widehat{S}(jw) = Y(jw) - \widehat{D}(jw) \tag{4}$$

where $\widehat{S}(jw)$ is the fault-related spectrum estimate and $\widehat{D}(jw)$ is an estimate of the steady-state component of spectrum. $\widehat{D}(jw)$ is often obtained using the time-averaged signal spectrum using the normal healthy state of the system:

$$\widehat{D}(jw) \cong |\overline{D}(jw)| = \frac{1}{M}\sum_{i=0}^{M-1} |D_i(jw)| \tag{5}$$

where $|\overline{D}(jw)|$ is time-averaged signal spectrum and M is the number of samples in the average window.

2.3 Convolutional neural network

Deep convolutional neural networks (CNNs) have recently demonstrated a great success in many machine learning tasks, such as regression and prediction. Such CNNs have been exploited to appropriately characterize internal variations (intra-class) within a large amount of data. The special characteristic of this network is that the network learns data-driven filters to convert the data to features that describe the inputs and represent variables of interest inside the network which are usually performed separately in traditional methods [30]. Therefore, it can achieve high performance even with minimal pre-processing. The general architecture of CNN consists of one input layer, one or multiple convolutional layers, pooling layers, fully connected layers, and one output layer.

In a convolution layer, the feature maps from previous layers are convolved with learnable kernels and fed to the activation function to construct the output feature map. Each output map may combine convolutions with multiple input maps. The general formulation for a convolutional layer is as follows [30]:

$$X_j^l = f\left(\sum_{i\in M_j} X_i^{l-1} * k_{ij}^l + b_j^l\right) \tag{6}$$

where X_i^{l-1} is the feature map in the previous layer, f is a nonlinear activation function, M_j represents a selection of input maps, l is the index for each convolution layer, k is a square matrix with the size of kernels, and b is an additive bias given to each output map.

Pooling layers are generally used after the convolutional layers to produce down sampled versions of the input maps. Therefore, the number of output maps will be the same as the number of input maps, but their dimensions are decreased. In terms of formulation [30]:

$$X_j^l = f\left(\beta_j^l \mathrm{down}(X_j^{l-1}) + b_j^l\right) \tag{7}$$

where down() represents a sub-sampling function. Max pooling is an example of such functions which uses the maximum value from each cluster of neurons at the prior layer [31]. Each output map has its own multiplicative bias β and an additive bias b.

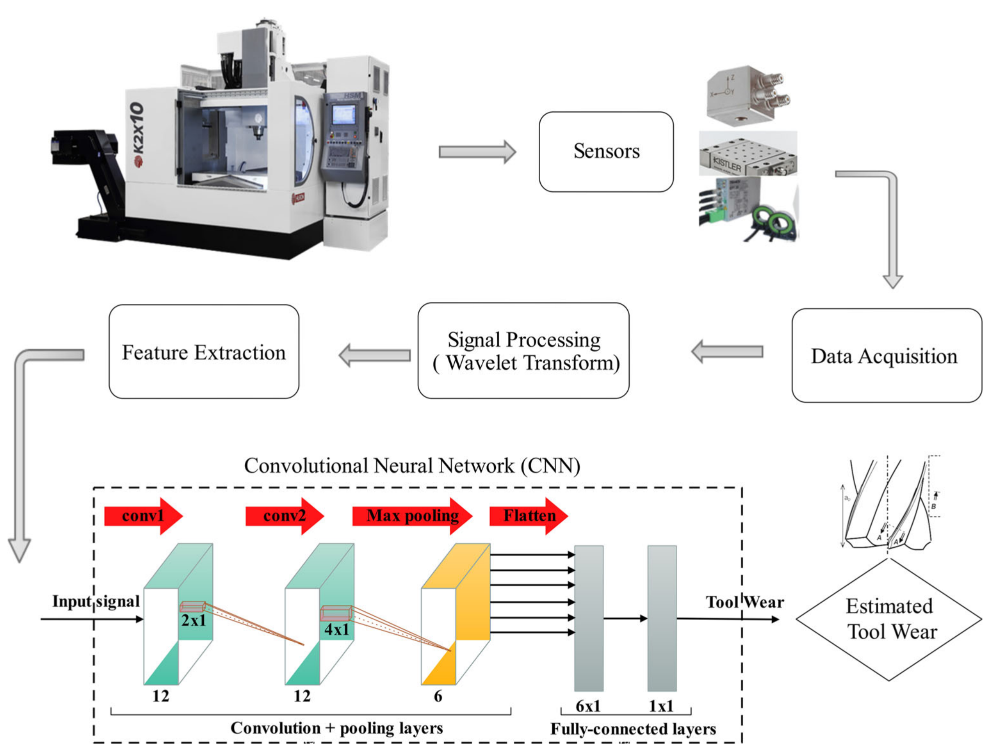

Fig. 1 The monitoring system framework

Finally, fully connected layers which are traditional multilayer perceptions (MLPs) [32] compute the desired outputs from the neurons of the previous layers.

3 Proposed methodology

In this section, proposed methodology of this paper is elaborated. The first step of the system is signal acquisition. Three different sensors (dynamometer, accelerometer, and current sensor) are examined using two different datasets. The framework of the system with its different steps is depicted in Fig. 1. After the data acquisition step, signals are processed to extract fault indicators and remove noise.

Time-frequency transformation is used for this task due to its promising capability in revealing the time-variant characteristics of the signals in frequency domain using Morlet wavelet transform method. We favored the Morlet wavelet in this research as in mechanical dynamical signals, impulses are usually the symptoms of faults and the Morlet wavelet is very similar to impulse component [33]. Furthermore, Jáuregui et al. in their research on frequency and time-frequency analysis of cutting force and vibration signals for tool condition monitoring reported the Morlet wavelet function as a good candidate for the feature extraction applications, as it provides a good balance between time and frequency resolutions [34].

The next step is to extract features from the wavelet transform that describe the fault properly. Wavelet packet transform is employed as it permits decomposing signals into uniform frequency bands [35]. WaveletPacket function from PyWavelets [36], the scientific Python module for wavelet transform calculations is used to perform the transformation. The algorithm proposed in this paper uses the Morlet as the wavelet function and with 4 as the levels of decomposition. The output of the transformation is divided into 16 uniform bands. The rms value of each grouped output frequency band is obtained and the first 12 bands of rms values are used as the input to the machine learning without further processing. Therefore, minimum pre-processing is implemented to explore the capability of CNN.

Contrary to other hand-crafted feature learning models, these data-driven models are capable of learning discriminative nonlinear feature representations. Thus, they can provide an effective prediction tool for fault detection by learning robust feature representations directly from the input signals.

A deep CNN model is proposed in this paper to accurately predict the faults in machining process. To that end, a simple yet effective architecture as shown in Fig. 1 is considered due to the constraints of tool condition monitoring system. The proposed architecture is comprised of two convolutional layers (conv1 and conv2), where each layer has 12 kernels of 2×1 and 4×1, respectively. To keep the original signal size and avoid decreasing the number of features, the stride is considered as 1 and these convolutional layers are designed consecutively. A pooling layer is used after those successive convolutional layers to handle the spatial size of the feature representation through max pooling, as well as to control the number of parameters (computational complexity of the network) and overfitting. The output of pooling layer is first flatten and then fed into two fully connected layers. The fully connected layers are responsible to compute the softmax activation with a matrix multiplication followed by a bias in order to produce the prediction value.

For the system which uses current signals, spectral subtraction method is applied to the signals. For this purpose, a local average of the spectral magnitude in different frequency bands is extracted using the dataset. For each new signal, the estimated healthy spectrum under the same cutting conditions is subtracted from the signal. Figure 2 illustrates the diagram of spectral subtraction method. Further noise canceling and signal refinement is performed after the spectral subtraction step.

In the next step, various features are extracted from the output of spectral subtraction step. Maximum energy and frequency of occurrence, variance, standard deviation, and width of frequency response are among the extracted features. Afterwards, CNN model which is used for this step is similar to the one which was explained in Fig. 1.

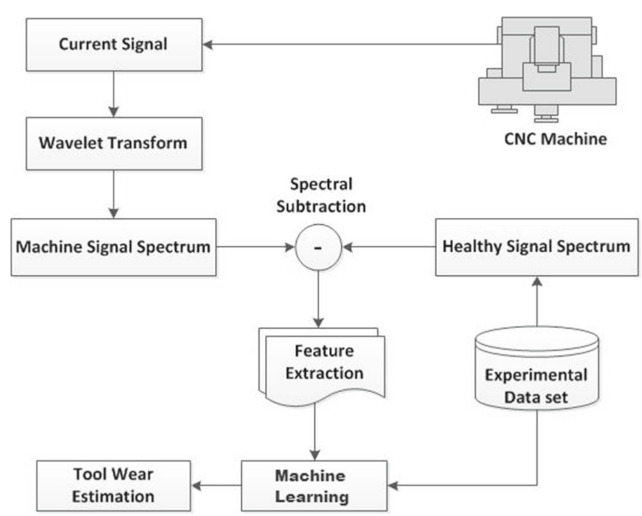

Fig. 2 Spectral subtraction method for current signal

4 Experimental datasets

4.1 ETS dataset

Experimental setup

A set of experiments are performed to measure tool flank wear during machining of hard to cut materials. K2X10 Huron high-speed CNC machine of the LIPPS laboratory at ETS is used to perform the experimental tests. A multi-component Kistler dynamometer 9255-B, coupled with charge amplifiers, was used to measure the cutting forces in three orthogonal directions (Fx/Fy/Fz). A tri-axial accelerometer was mounted on the spindle of the machine with a sensitivity of 100 mV/g for measuring acceleration.

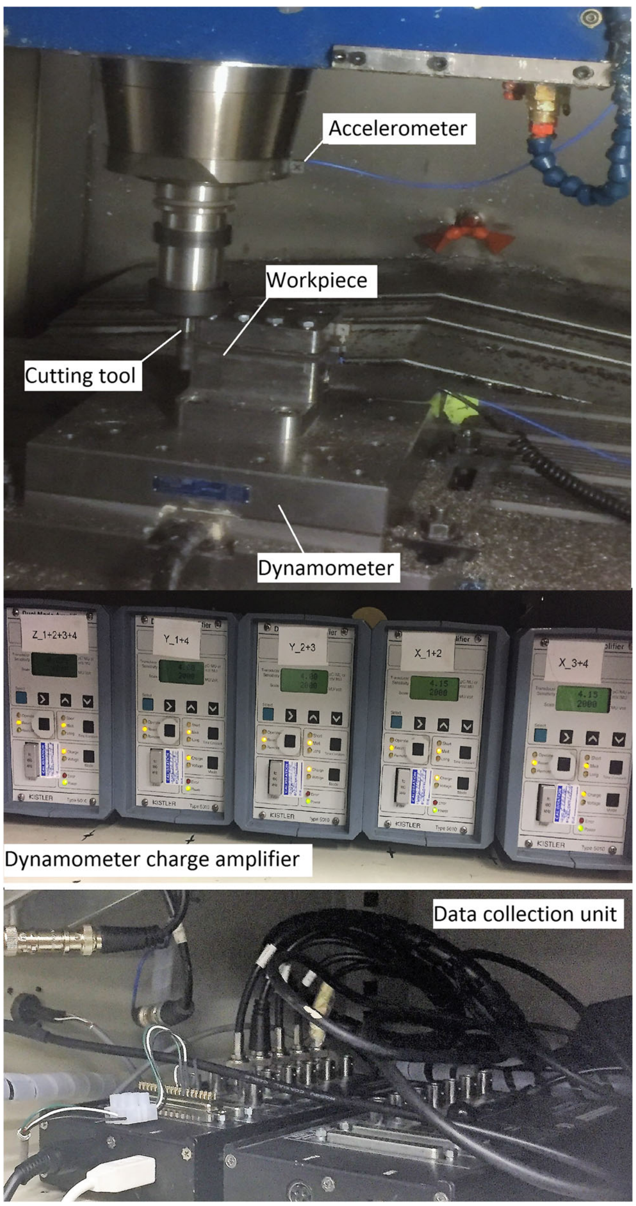

Fig. 3 Experimental setup

D2 high-speed tool steel is selected as the workpiece material with hardness of 60-62 HRC due to its high wear resistance in order to investigate tool wear in machining hard material with dimension of 200 mm × 54 mm × 4 mm. Carbide Walter End Mill Protostar H50 Ultra tool with 6 teeth is selected as the cutting tool with 50° of helix angle. Different cutting speeds of 2500 rpm and 6000 rpm and feed rates of 0.12 mm/tooth and 0.05 mm/tooth with 4-mm depth of cut and tool wear were measured at different intervals which results in 63 cases with different tool wears and cutting conditions. Figure 3 demonstrates this experimental setup.

4.2 Nasa_Ames dataset

Tool fault detection models and validation of the research method is implemented by the benchmark NASA Ames and UC Berkeley milling dataset [29]. The experiments are performed under various operating conditions using the Matsuura MC-510V machining center. In this research, spindle current sensor is selected based on its ease of use and practicality in industrial application. The dataset signals include cases with changes in depth of cut and feed rate, and therefore, the effect of these parameters can be investigated on the monitoring system accuracy and the system will be developed under varying cutting parameters.

The tool is a 70-mm face mill with six KC710 inserts based on its industrial applicability. Workpiece material in the research is cast iron. An OMRON K3TB-A1015 current converter feeds the signal from one spindle motor current phase into the cable connector and a model CTA 213 current sensor (Flexcore Div. of Marlan & Associates, Inc.) is used for data acquisition. Flank wear (VB in μm), which is defined as the distance from the cutting edge to the end of the abrasive wear on the flank face is considered as the fault and its value is reported in all the experiments using a microscope.

5 Results and discussion

In this section, results for the three monitoring systems are presented. The first one uses force signals from the ETS dataset as the monitoring signal. The second one employs the vibration signals from ETS dataset and finally the third subsection investigates the system with current signals from the Nasa_Ames dataset and spectral subtraction method.

5.1 Tool wear estimation using force signals from ETS dataset

The methodology of this system is depicted in Fig. 1. This system uses force signal as the monitoring indicator from

Table 1 Comparison between different machine learning algorithms with force signals

Regression algorithms	Average accuracy, %	RMSE
Bayesian ridge regression	73.1	0.1815
Nearest neighbors regression (KNN)	71.5	0.2021
Support vector regression (SVR)	79.0	0.1103
Convolutional neural network (CNN)	88.2	0.0709

the ETS dataset. The data is divided into two categories, training and testing. The system is trained using the training subset which consists of 70% of the data. Afterwards, the system is tested with 30% of the data. The features and tool wear are normalized before machine learning step and denormalized after tool wear prediction. MinMaxScaler function from Scikit-learn python library [37] is employed as an standard machine learning feature normalization method to normalize both inputs and tool wear. It trains an estimator using a linear scaling function to transform features .This estimator scales and translates each feature individually such that it is in the given range (between zero and one in this paper) on the training set using a linear interpolation.

The keras deep learning library is employed [38] with tensorflow as the back-end [39] to implement the proposed CNN model. Three of the common machine learning methods (SVR, KNN, and Bayesian network) in this field are also implemented using Scikit-learn machine learning library [37] as a baseline to compare the performance of the CNN-based system with these methods. For the monitoring system, average accuracy in percentage (the differences between predicted and actual tool wear values divided by average of tool wears) and RMSE are calculated as representative of the performance from the Scikit-learn machine learning performance analysis toolboxes.

Table 1 presents the results of tool wear estimation using test dataset for different machine learning algorithms. Based on the results, CNN has the highest accuracy (88.2%) and lowest root mean square error (RMSE) (0.0709) which are acceptable for most industrial applications. It can be observed from the results is a relatively high difference between the accuracy of CNN and other methods due to the fact that signals were not processed after the time-frequency transformation method. Convolution layers of the CNN was able to filter the data and convert it to more discriminative features. Figure 4 presents the predicted versus actual tool wears using the CNN-based algorithm for two tools from the no wear state up to the high tool wear values. For each tool, the experiments start with no wear (VB = 0) and the curves show gradual increase in the tool wear as cutting is continued until high tool wear values. Based on the figure, the estimated tool wear greatly correlates with actual tool wear.

5.2 Tool wear estimation using vibration signals from ETS dataset

The system in this section is similar to the force-based monitoring system of the previous section, except that the force signal is replaced with the spindle vibrations. The ETS dataset is used for this system as well. Data is divided into two categories, Training and testing with 70% and 30% of the data respectively. The keras deep learning library is employed to implement the proposed CNN model and

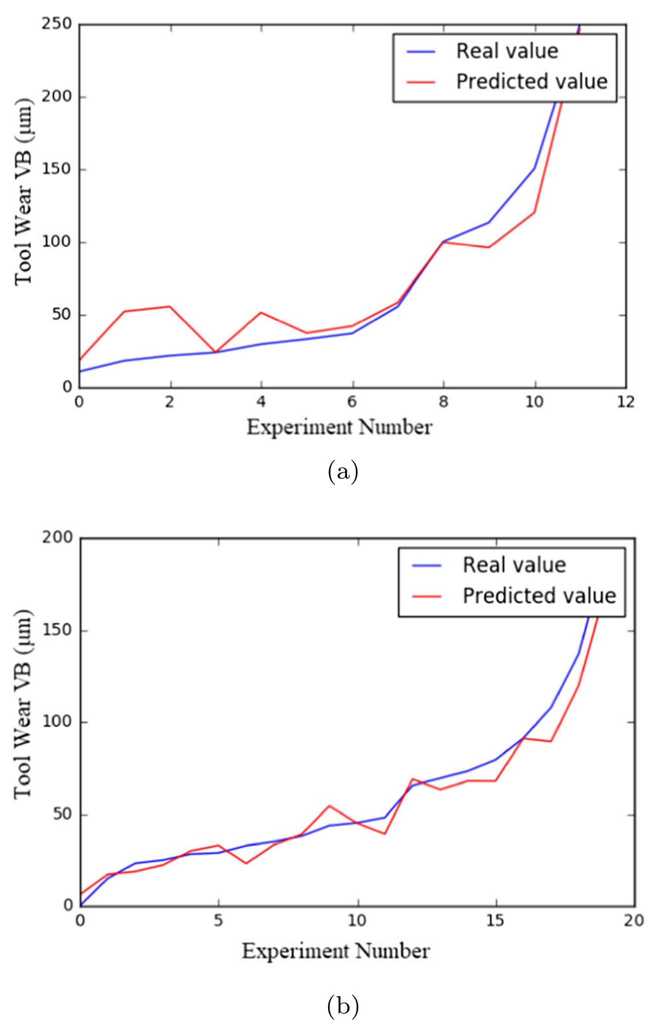

Fig. 4 Estimated and real tool wear values using force signals. **a** First cutting and **b** second cutting tool

Table 2 Comparison between different machine learning algorithms with vibration signals

Regression algorithms	Average accuracy, %	RMSE
Bayesian ridge regression	66.9	0.2251
Nearest neighbors regression (KNN)	57.0	0.281
Support vector regression (SVR)	59.6	0.2701
Convolutional neural network (CNN)	84.6	0.086

Scikit-learn for other machine learning methods similar to previous sections.

The result of tool wear estimation using vibration test dataset is provided in Table 2. The accuracy of CNN is 84% which is slightly lower than the force-based system of previous sections, but still promising for real-world applications. For the vibration-based systems, there is a high difference between the accuracy of CNN and other machine learning methods. It can be interpreted that it is due to the fact that vibration signals do not provide features directly related to tool wear without specific and hand-crafted pre-processing. However, convolution filters in CNN architecture were able to convert the data to discriminative features. RMSE value of CNN (0.086) is also significantly lower than other methods. Figure 5 illustrates the predicted versus actual tool wear using the CNN-based algorithm for two different tools. Experiments start with a new tool with no wear (Experiment 0) state up to the high tool wears.

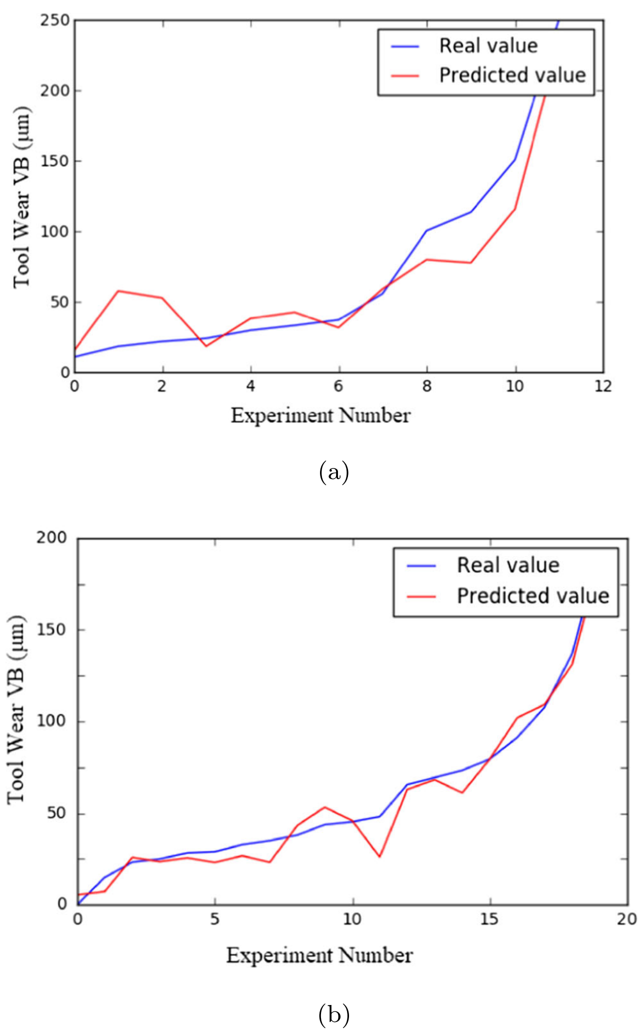

Fig. 5 Estimated and real tool wear values using vibration signals. **a** First cutting tool and **b** second cutting tool

5.3 Tool wear estimation using spindle current signals from Nasa_Ames dataset

The last section of the study investigates a system design based on the spindle current signal as the fault indicator. Nasa_Ames dataset is utilized for the validation. Two systems are trained for comparison one with spectral subtraction method and another without this method. The architecture of the first system is illustrated in Fig. 2. The architecture of the second method is similar to the force- and vibration-based systems, which directly fed machine learning with wavelet transform-based features. Therefore, the effect of higher pre-processing is investigated in the performance of the algorithms.

Figure 6 depicts the result of wavelet transform using some sample signals from the dataset. The diagram represents the WT output for the healthy signal (VB = 0) as well as four states of the fault. It can be observed from the diagrams that the signal has higher energy around tooth pass frequency. As the fault value increases, the magnitude and density of wavelet are increased. However, the data is still noisy and needs to be further processed to extract discriminative features. The processed data after the spectral subtraction is shown in Fig. 7. For this analysis, an average estimation of spectrum of the current signal around tooth path frequency for healthy case is extracted from the dataset. For each new signal of the machine, the estimated healthy spectrum under the same cutting conditions is subtracted from the signal. It normalizes the signals based on their cutting conditions and magnifies the effect of tool wear.

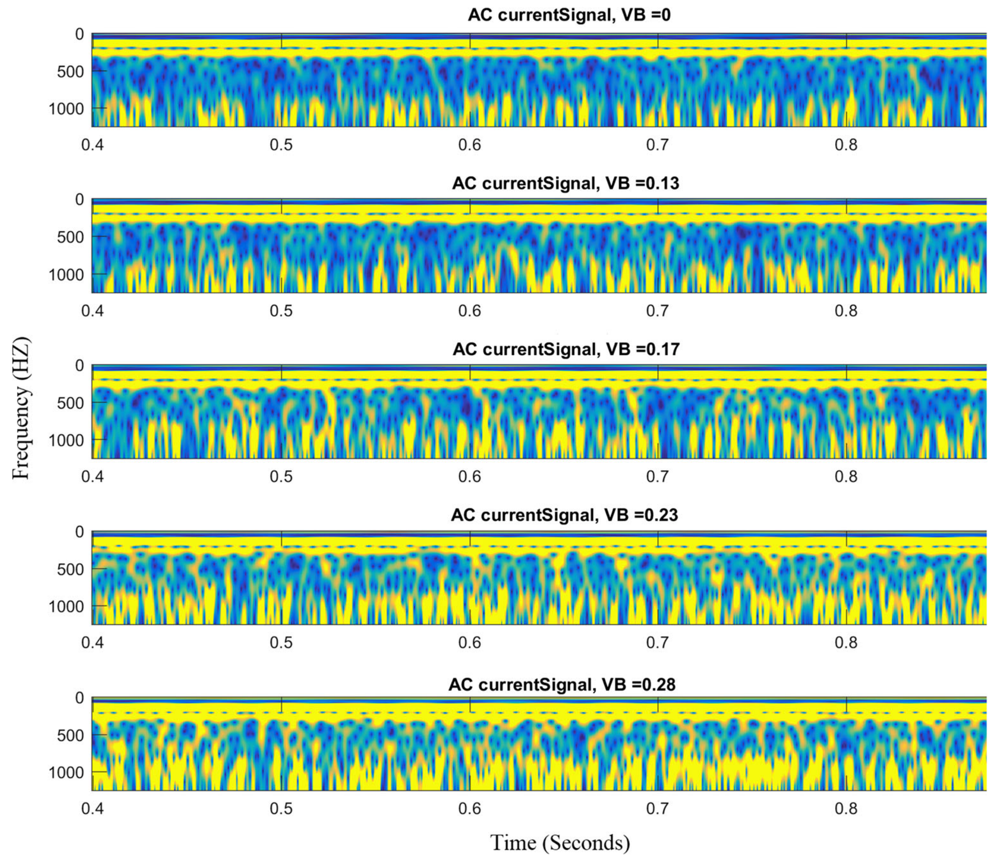

Fig. 6 Spindle current signals with different tool wears in wavelet domain

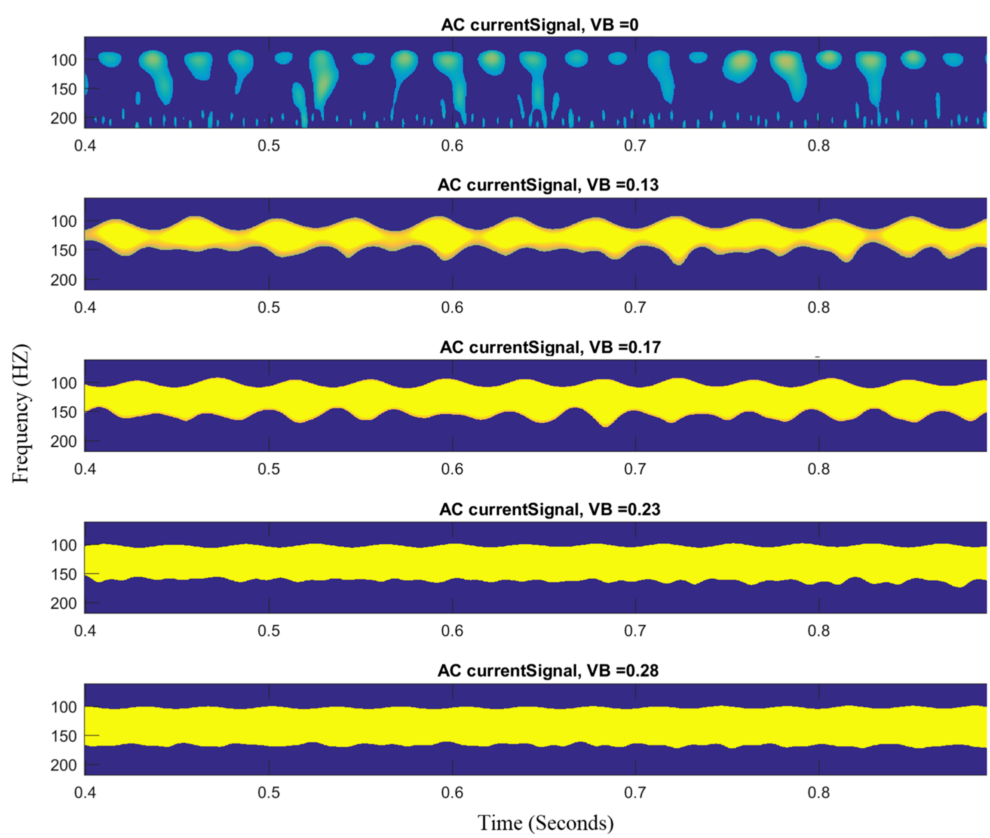

Fig. 7 Spindle current signals with different tool wears after spectral subtraction

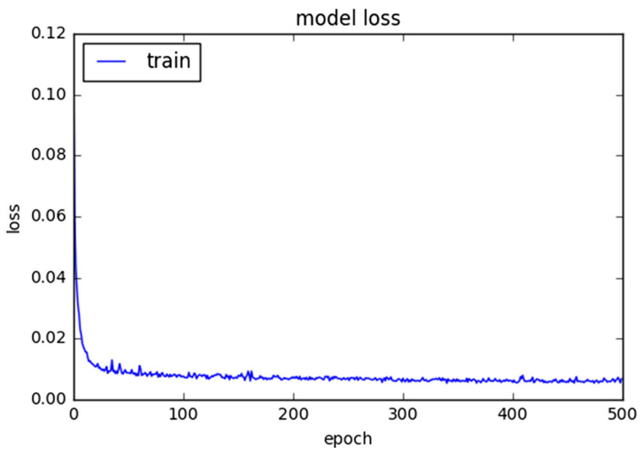

Fig. 8 Loss function during training process

Comparing the Figs. 6 and 7, the signal quality is enhanced after the spectral subtraction and the signals are more discriminative with respect to tool wear.

The CNN architecture and machine learning implementation for this step is similar to two previous systems. Figure 8 reports the loss values converging close to zero during the epochs of the training step. Afterwards, the unseen test dataset signals are fed to the system.

Based on the results of Table 3, CNN is superior to other methods with 87.2% accuracy in tool wear estimation for the system with spectral subtraction Bayesian rigid regression and support vector regression methods also have satisfactory results of 85.8% and 85.5% respectively. Based on the comparison between the systems with and without spectral subtraction, spectral subtraction increased the accuracy of each algorithm by approximately 5%. Therefore, it can be concluded that although CNN has powerful capabilities to interpret data with minimum preprocessing, it would still benefit from advanced and efficient signal processing methods, specially under limited number of training samples.

Based on the results of three studied systems, CNN consistently provided higher results than other algorithms with different datasets and signals which proves its efficiency. The major advantage of CNN is in lower processed signals which has ability to extract relevant information compared to other methods.

6 Conclusion

In this research, a tool condition monitoring methodology is proposed and tested under changing cutting parameters. Force and vibration signals from the ETS dataset and spindle motor current signals from Nasa_Ames dataset are used as monitoring signals. Wavelet transform as an advanced time-frequency transformation method is used in the signal processing step due to its great applicability to process signals and reveal rich information in both time and frequency domains simultaneously. A deep CNN method is also implemented as the last step to model the complex relationships between extracted features and tool wear values.

Wavelet analysis revealed the time-variant characteristics of frequency response of the signal and the study confirms its performance and applicability for tool wear monitoring. Spectral subtraction method is employed for the current signal which significantly improved its condition and magnified the signature of tool wear by removing the steady-state part of the signal due to normal cutting and magnified the remaining fault characteristics.The comparison between the systems with and without spectral subtractions shows approximately 5% increase in the accuracy of systems which benefit from this algorithm.

Tables 1, 2, and 3 report the comparative results of the CNN-proposed methodology of the paper with some of the common machine learning techniques. Based on the results, CNN consistently outperforms other machine learning algorithms between all three signals from two datasets which proves its robustness and high performance in this application. CNN improved the accuracy of force- and vibration-based methods significantly (around 15%). It is interpreted as the results of CNN convolution layers which filtered the signal and extracted discriminative features from the raw wavelet response. Therefore, it is beneficial in reducing the cost of specific engineering data manipulations and improving accuracy in the case of low-quality signals.

Acknowledgements The authors gratefully acknowledge the experimental mill data provided by UC Berkely BEST lab and NASA Ames Prognostic Data Repository.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Table 3 Comparison between different machine learning algorithms with current signals

Regression algorithms	With spectral subtraction		Without spectral subtraction	
	Average accuracy, %	RMSE	Average accuracy, %	RMSE
Bayesian ridge regression	85.8	0.091	79.6	0.188
Nearest neighbors regression (KNN)	79.1	0.220	78.5	0.232
Support vector regression (SVR)	85.5	0.102	80.1	0.181
Convolutional neural network (CNN)	87.2	0.088	81.5	0.156

References

1. Zhu K, Vogel-Heuser B (2014) Sparse representation and its applications in micro-milling condition monitoring: noise separation and tool condition monitoring. In J Adv Manuf Technol 70(1–4):185–199
2. Siddhpura A, Paurobally R (2013) A review of flank wear prediction methods for tool condition monitoring in a turning process. Int J Adv Manuf Technol 65(1–4):371–393
3. Abellan-Nebot JV, Subirón FR (2010) A review of machining monitoring systems based on artificial intelligence process models. Int J Adv Manuf Technol 47(1-4):237–257
4. Li N, Chen Y, Kong D, Tan S (2017) Force-based tool condition monitoring for turning process using v-support vector regression. Int J Adv Manuf Technol 91(1–4):351–361
5. Harun MHS, Ghazali MF, Yusoff AR (2017) Analysis of tri-axial force and vibration sensors for detection of failure criterion in deep twist drilling process. Int J Adv Manuf Technol 89(9–12):3535–3545
6. Bhuiyan MSH, Choudhury IA, Dahari M, Nukman Y, Dawal S (2016) Application of acoustic emission sensor to investigate the frequency of tool wear and plastic deformation in tool condition monitoring. Measurement 92:208–217
7. Rad JS, Zhang Y, Chen C (2014) A novel local time-frequency domain feature extraction method for tool condition monitoring using s-transform and genetic algorithm. IFAC Proc Vol 47(3):3516–3521
8. Lin X, Bo Z, Zhu L (2017) Sequential spindle current-based tool condition monitoring with support vector classifier for milling process. Int J Adv Manuf Technol, 1–10
9. Segreto T, Simeone A, Teti R (2013) Multiple sensor monitoring in nickel alloy turning for tool wear assessment via sensor fusion. Procedia CIRP 12:85–90
10. Rehorn AG, Jiang Jin, Orban PE (2005) State-of-the-art methods and results in tool condition monitoring: a review. Int J Adv Manuf Technol 26(7):693–710
11. Feng Z, Liang M, Chu F (2013) Recent advances in time–frequency analysis methods for machinery fault diagnosis: a review with application examples. Mech Syst Signal Process 38(1):165–205
12. Rehorn AG, Sejdić E, Jiang J (2006) Fault diagnosis in machine tools using selective regional correlation. Mech Syst Signal Process 20(5):1221–1238
13. Rad JS, Zhang Y, Aghazadeh F, Chen ZC (2014) A study on tool wear monitoring using time-frequency transformation techniques. In: Proceedings of the 2014 international conference on innovative design and manufacturing (ICIDM). IEEE, pp 342–347
14. Elgargni M, Al-Habaibeh A, Lotfi A (2015) Cutting tool tracking and recognition based on infrared and visual imaging systems using principal component analysis (PCA) and discrete wavelet transform (DWT) combined with neural networks. Int J Adv Manuf Technol 77(9–12):1965–1978
15. Shi D, Gindy NN (2007) Tool wear predictive model based on least squares support vector machines. Mech Syst Signal Process 21(4):1799–1814
16. Jin X, Zhao M, Chow TWS, Pecht M (2014) Motor bearing fault diagnosis using trace ratio linear discriminant analysis. IEEE Trans Ind Electron 61(5):2441–2451
17. Bodin P, Villemoes LF (1997) Spectral subtraction in the time-frequency domain using wavelet packets. In: 1997 IEEE workshop on speech coding for telecommunications proceeding, 1997. IEEE, pp 47–48
18. Boll S (1979) Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans Acous Speech Signal Process 27(2):113–120
19. Choqueuse V, El Hachemi Benbouzid M et al (2013) Current frequency spectral subtraction and its contribution to induction machines' bearings condition monitoring. IEEE Trans Energy Convers 28(1):135–144
20. Patra K, Jha AK, Szalay T, Ranjan J, Monostori L (2017) Artificial neural network based tool condition monitoring in micro mechanical peck drilling using thrust force signals. Precis Eng 48:279–291
21. Madhusudana CK, Kumar H, Narendranath S (2017) Face milling tool condition monitoring using sound signal. Int J Syst Assur Eng Manag 8(2):1643–1653
22. Tobon-Mejia DA, Medjaher K, Zerhouni N (2012) CNC machine tool's wear diagnostic and prognostic by using dynamic Bayesian networks. Mech Syst Signal Process 28:167–182
23. Schmidhuber J (2015) Deep learning in neural networks: an overview. Neur Netw 61:85–117
24. Deng L (2014) A tutorial survey of architectures, algorithms, and applications for deep learning. APSIPA Trans Signal Inf Process, 3
25. Jia F, Lei Y, Lin J, Zhou X, Lu N (2016) Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. Mech Syst Signal Process 72:303–315
26. Zhao R, Yan R, Wang J, Mao K (2017) Learning to monitor machine health with convolutional bi-directional LSTM networks. Sensors 17(2):273
27. Jing L, Zhao M, Li P, Xu X (2017) A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. Measurement 111:1–10
28. Chen ZQ, Li C, Sanchez R-V (2015) Gearbox fault identification and classification with convolutional neural networks. Shock Vib, 2015
29. Agogino A, Goebel K (2007) Mill data set. BEST lab, UC Berkeley. NASA Ames Prognostics Data Repository, NASA Ames, Moffett Field, CA, http://ti.arc.nasa.gov/project/prognostic-data-repository
30. Bouvrie J (2006) Notes on convolutional neural networks
31. Ciregan D, Meier U, Schmidhuber J (2012) Multi-column deep neural networks for image classification. In: 2012 IEEE conference on computer vision and pattern recognition (CVPR). IEEE, pp 3642–3649
32. Ruck DW, Rogers SK, Kabrisky M, Oxley ME, Suter BW (1990) The multilayer perceptron as an approximation to a Bayes optimal discriminant function. IEEE Trans Neural Netw 1(4):296–298
33. Lin J, Qu L (2000) Feature extraction based on Morlet wavelet and its application for mechanical fault diagnosis. J Sound Vibr 234(1):135–148
34. Jáuregui JC, Reséndiz JR, Thenozhi S, Szalay T, Jacsó Á, Takács M (2018) Frequency and time-frequency analysis of cutting force and vibration signals for tool condition monitoring. IEEE Access 6:6400–6410
35. Barros J, Diego RI (2006) Application of the wavelet-packet transform to the estimation of harmonic groups in current and voltage waveforms. IEEE Trans Power Deliv 21(1):533–535
36. Wasilewski F (2010) Pywavelets: discrete wavelet transform in python
37. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V et al (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830
38. Chollet F et al (2015) Keras
39. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M et al (2016) Tensorflow: large-scale machine learning on heterogeneous distributed systems. arXiv:1603.04467