

A novel ensemble deep learning model for cutting tool wear monitoring using audio sensors

Zhixiong Li^{a,d}, Xihao Liu^b, Atilla Incecik^c, Munish Kumar Gupta^{a,*}, Grzegorz M. Królczyk^a, Paolo Gardoni^e

^a Faculty of Mechanical Engineering, Opole University of Technology, Opole 45758, Poland

^b School of Engineering, Ocean University of China, Tsingtao 266001, China

^c Department of Naval Architecture, Ocean and Marine Engineering, University of Strathclyde, Glasgow, Scotland G1 1XQ, United Kingdom

^d Yonsei Frontier Lab, Yonsei University, Seoul 03722, Republic of Korea

^e Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Champaign, IL, United States of America

ARTICLE INFO

Keywords:

Machining
Tool wear
Audio signal processing
Intelligent detection
Sensors

ABSTRACT

Tool wear is an important parameter in the machining because the production, cost and performance is highly depend upon its performance. Therefore, the monitoring of cutting tool wear plays an important role in mechanical machining processes. With this aim, the present work deals with the application of novel ensemble deep learning model for cutting tool wear monitoring using audio sensors. The tool wear data during machining was extracted with an audio denoising technique combined with Fast Fourier Transform (FFT) and bandpass filters and dependent component analysis (DCA). Then, the ensemble convolutional neural networks (CNN) detection model was trained and audio signals were converted into audio images with different algorithms. Finally, the results confirm that this novel method is very accurate to predict the tool wear values under different cutting conditions.

1. Introduction

Precision high-end machine tool is the most upstream and most basic link in the modern manufacturing industry due to its wide range of applications such as grinding, milling, turning etc. [1]. However, the cutting tool used in the high-end machine tool always suffers from tool wear, which leads to low production quality, high labor and high maintenance costs etc. Therefore, there is a need to develop an effective tool monitoring system (TCM) that ensures the optimum operating conditions of the cutting tools to improve the machining quality and economics [2].

Cutting tool wear monitoring is a challenging task because many machining processes present non-linear time-varying characteristics [3]; as a result, exact theoretical models are difficult/impossible to establish for accurate monitoring, and direct measuring the tool wear during cutting is always difficult. Since the starting of industrial revolution i.e., 1951, many scholars have carried out extensive research on tool wear monitoring by employing different sensors [4]. In general, the main functions of a tool wear monitoring system include the signal

acquisition, signal pretreatment, feature extraction and decision making [5]. According to the contact degree of the measurement sensors, the tool monitoring can be divided into the direct and indirect methods [6]. The direct method includes the visual examination, laser beam and electrical resistance techniques, which involve measuring the corresponding process variables by interrupting the machining process. Consequently, most direct measuring techniques are limited to laboratory techniques and very a few can be applied to industrial applications.

The indirect method generally consists of auxiliary systems that are flexible and dependent on the machining process parameters. Many techniques [6–8] include the cutting force measuring, vibration (acceleration) measuring, acoustic emission measuring, current and power measuring, cutting temperature measuring, surface roughness measuring, etc. For instance, Chao et al. [9] predicted the average cutting temperature to judge the tool wear state by combining the shear heat source of the workpiece and the friction heat source of the tool and chip interaction with the thermal radiation sensor. Dimla et al. [10] collected the vibration signals in various directions through vibration sensors and used them to monitor tool wear. According to Silva et al.

* Corresponding author.

E-mail addresses: zhixiong.li@yonsei.ac.kr (Z. Li), 21190931169@stu.ouc.edu.cn (X. Liu), atilla.incecik@strath.ac.uk (A. Incecik), munishguptanit@gmail.com (M.K. Gupta), g.krolczyk@po.opole.pl (G.M. Królczyk), gardoni@illinois.edu (P. Gardoni).

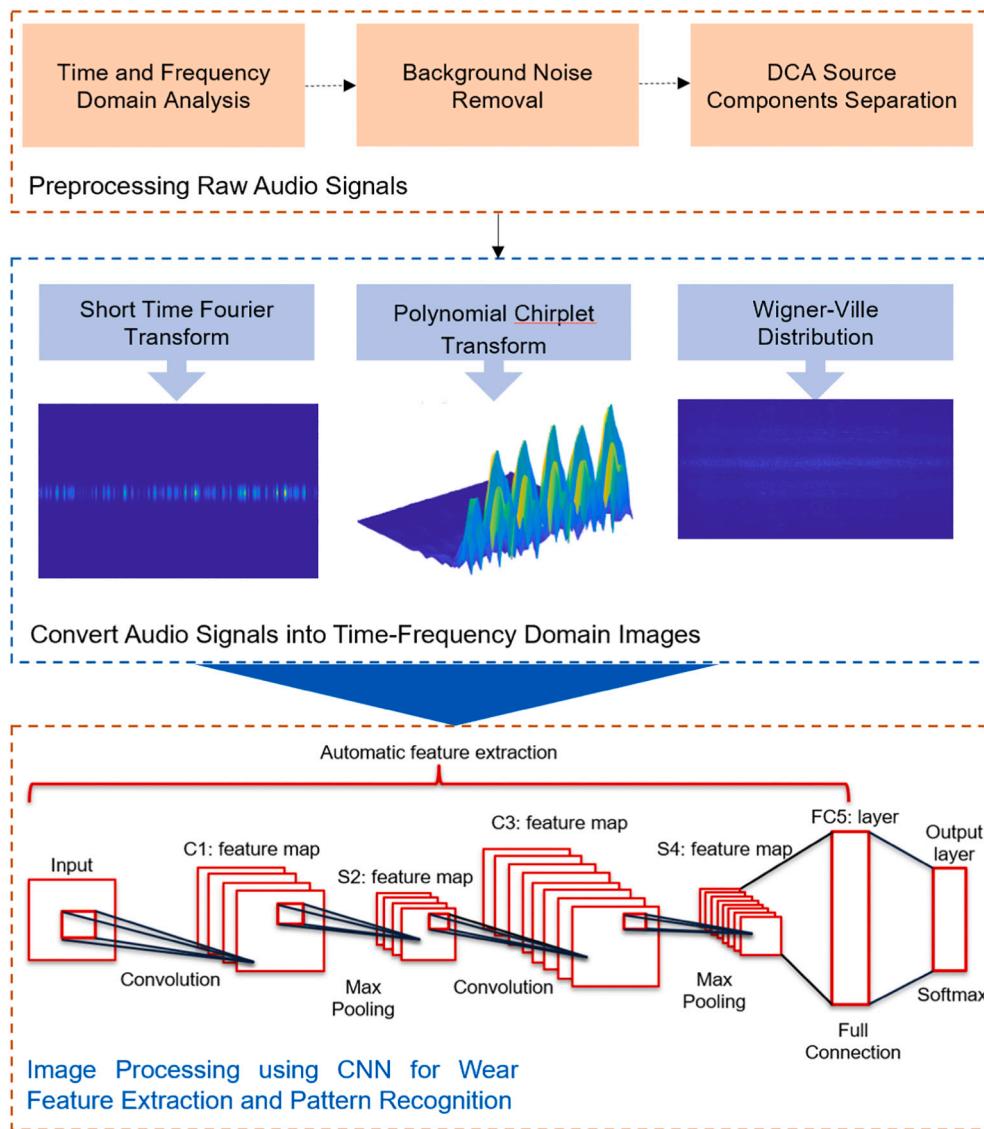


Fig. 1. An overview of the proposed tool wear monitoring method.

[11], the dynamometer is based on the measurement of the cutting force during the machining and is the most reliable method for tool condition monitoring. The dynamometer method focuses on the cutting force, which is thought to be sensitive to the friction that changes with the cutting tool wear. Twardowski et al. [12] found that when the tool is damaged, the acoustic emission became discontinuous burst signal; thus, the static characteristic parameters of the acoustic emission signal can be used to monitor the tool wear. Loenzo et al. [13] investigated the tool from the beginning of service to the failure, and found that the cumulative increase of the cutting power can effectively identify the state of tool breakage. Durmuş et al. [14] determined the tool wear range based on the surface roughness datum of the machining parts under different wear states. Although these indirect measuring techniques are able to monitor the tool wear condition, the high cost and difficult installation of these mentioned sensors significantly limit their practical applications.

An alternative way is the audible sound monitoring, which analyzes the audio waves generated during machining process to determine the tool wear conditions [15–18]. The Challenge of the Acoustic Scene and Event Detection and Classification (DCASE) provided publicly available datasets for researchers to evaluate their methods [19]. These methods include the singular value decomposition (SVD) [20], recursive feature

elimination (RFE) [21], extreme learning machine (ELM) [22], Wavelet packet decomposition (WPD) [23], Log-MEL graph analysis [24], support vector machine (SVM) [16], extended convolution bounded component analysis (ECBCA) [15], wavelet transform modulus maximum (WTMM) [25], Holder Exponents (HE) [26] and so forth. For example, Prakash and Samraj [27] used the SVD to analyze the sound signal to monitor the tool wear in small interval turning. Gomes et al. [28] adopted the SVM to process the vibration and sound signals and used the RFE to select the most important features to monitor the tool wear in a micro-milling process. Zhou et al. [29] monitored the tool wear audio based on two-layer angular kernel ELM, where only a few parameters of the sound sensor signal were used to identify the tool wear condition. Rafezi et al. [30] applied the WPD to the tool state monitoring using the sound and vibration analysis in a grinding process. Kim et al. [24] used the log-MEL spectrogram and convolutional neural networks (CNN) to analyze the sound in a remote real-time multi-device operation monitoring system. Kothuru et al. [17] employed the SVM to monitor the milling tool wear. The SVM provided the knowledge embedded model to analyze the sound signals. Cooper et al. [31] proposed a statistical model of milling acoustic signals with CNN. Li et al. [15] used the ECBCA and ensemble learning for tool wear monitoring. Zhou et al. [32] proposed a denoising algorithm based on the WTMM to eliminate noise

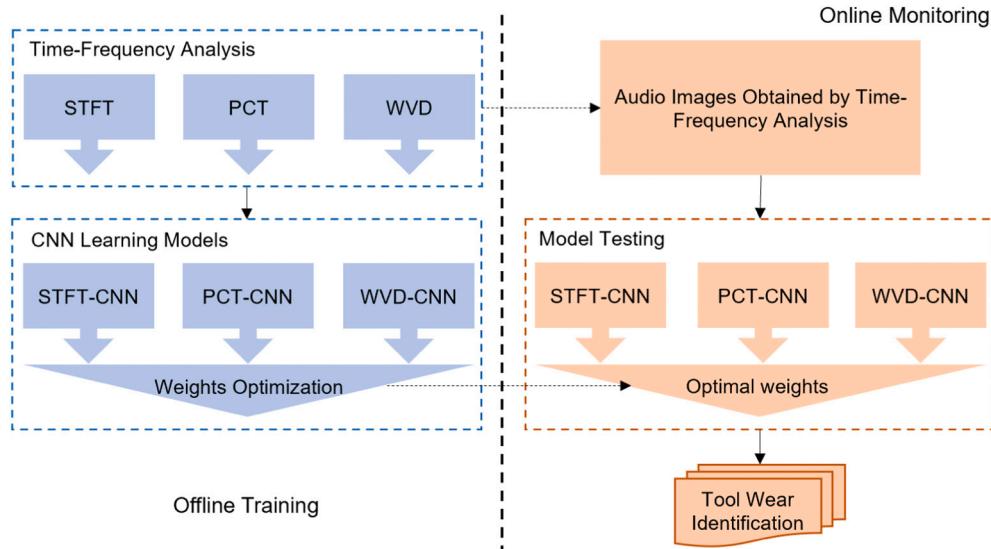


Fig. 2. Ensemble CNN model for tool wear identification.

for the milling tool wear monitoring.

Based on these existing literature survey, it concludes that the audio signal is corrupted by noise (including background noise and stochastic noise) and the sound data filtering is a must. The background noise can be processed by the band-pass filters while the random noise can be depressed by the blind source separation (BSS) [15]. Popular BSS methods include the independent component analysis (ICA) [33], dependent component analysis (DCA) [34], bounded component analysis (BCA) [35], and sparse component analysis (SCA) [36], etc. A challenging task when applying BBS to tool cutting audio denoising is the separation of dependent components because in reality various sound components may be collected during the machining process and many of them are dependent with each other. The DCA algorithm is suitable for solving this issue; however, the DCA-based tool audio denoising has not been found in literature.

Furthermore, it is far from enough to establish an effective and reliable tool wear monitoring system through regular audio signal processing without computer image vision [37–42]. Existing research results demonstrate that it is necessary to incorporate the image vision technique into the tool wear audio analysis framework to produce reliable detection results on the tool wear. However, to our best knowledge, the combination of DCA and CNN for tool wear audio processing has not been reported.

In order to address the aforementioned issues in tool wear monitoring, this study present a novel intelligent method by appropriately integrating the DCA-based denoising and CNN-based tool wear degree recognition. The feasibility of the proposed method is evaluated by experimental testing.

2. The proposed method

The proposed DCA-CNN method for tool wear monitoring is shown in Fig. 1, which consists of three part, i.e., the signal preprocessing, the signal time-frequency analysis, and the CNN-based tool wear recognition.

- (1) The signal preprocessing aims to filter both the background and stochastic noise in the audio signal, background noise and random noise are in the process of tool processing, the operator's voice, footsteps or the sound of accidentally touching the machine tool will be collected into the audio signal, because the real machining requires human control, it is difficult to avoid human noise interfering with the collection of the tool processing audio

signal. The fast Fourier transform is adopted to analyze the frequency graph of the audio signal, and a bandpass filter is built based on the Fourier analysis result to eliminate the background noise. Subsequently, the DCA is applied to the filter audio to separate the most useful source component related to the tool wear degree.

- (2) The signal time-frequency analysis aims to convert the audio signal into images that can be processed by the CNN deep learning. There are many time-frequency transform algorithms. The most popular ones are the short-time Fourier transform (STFT) and continuous Wavelet transform (CWT). However, these Fourier-based transform methods adopt the linear transform with static resolution; that is, the total resolution of the transform in the time and frequency domains is a constant (if the frequency resolution is high, then the time resolution will be low; and vice versa). The Wigner-Ville Distribution (WVD) [43] can solve this problem by the quadratic transformation strategy. In addition, the recently developed polynomial chirplets transform (PCT) [44] is able to process nonlinear instantaneous frequency signal. The theory behind the PCT is briefly introduced here.

The chirplet transform of a signal $s(t) \in L^2(\mathbb{R})$ is defined as

$$CT_s(t_0, \omega, \alpha; \sigma) = \int_{-\infty}^{\infty} z(t) \Psi^*_{(t_0, \alpha, \sigma)}(t) (t) e^{-\omega t} dt \quad (1)$$

where, $z(t)$ is the analytic signal of $s(t)$ obtained by the Hilbert transform H , i.e.,

$$z(t) = s(t) + iH[s(t)]; \text{ and } \Psi_{(t_0, \alpha, \sigma)}^*(t) \text{ is a complex window given by} \\ \Psi_{(t_0, \alpha, \sigma)}^*(t) = \omega_\sigma(t - t_0) e^{-\frac{\alpha^2}{2}(t-t_0)} \quad (2)$$

where, σ is the argument parameter, t_0 and $\alpha \in \mathbb{R}$ are the time and chirplets rate, respectively; $\omega \in L^2(\mathbb{R})$ is a non-negative, symmetric, and normalized real window.

$$\omega_\sigma(t) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{t-t_0}{\sigma} \right)^2} \quad (3)$$

In order to improve the efficiency in analyzing nonlinear instantaneous frequency trajectory signal, the PCT is proposed as

with

Table 1
Tool wear classification.

Sr. no.	Tool wear conditions	Thickness range (μm)
1	Good	0–20
2	Light wear	20–40
3	Average	40–70
4	Advanced wear1	70–100
5	Advanced wear2	100–150
6	Failure	>150

Table 2
Is an overview of actual datasheet with only a few data points per each class are displayed.

Tool wear conditions	Cutting speed (sfpm)	Feed rates (ipm)	Microphone no.
Good	176.8	9.0	1
Good	176.8	9.0	2
Good	176.8	9.0	3
Light wear	167.0	8.0	1
Light wear	167.0	8.0	2
Light wear	167.0	8.0	3
Average	147.4	7.0	1
Average	147.4	7.0	2
Average	147.4	7.0	3
Advanced wear1	137.5	12.0	1
Advanced wear1	137.5	12.0	2
Advanced wear1	137.5	12.0	3
Advanced wear2	127.7	9.0	1
Advanced wear2	127.7	9.0	2
Advanced wear2	127.7	9.0	3
Failure	196.5	24.0	1
Failure	196.5	24.0	2
Failure	196.5	24.0	3

$$PCT_s(t_0, \omega, \alpha_1, \alpha_2, \dots, \alpha_n; \sigma) = \int_{-\infty}^{\infty} z(t) \Phi_{\alpha_1, \dots, \alpha_n}^R(t) \times \Phi_{\alpha_1, \dots, \alpha_n}^M(t, t_0) \omega_{(\sigma)}(t - t_0) \exp(-j\omega t) dt \quad (4)$$

$$\Phi_{\alpha_1, \dots, \alpha_n}^R(t) = \exp\left(-j \sum_{k=2}^{n+1} \frac{1}{k} \alpha_{k-1} t^k\right) \quad (5)$$

$$\Phi_{\alpha_1, \dots, \alpha_n}^M(t, t_0) = \exp\left(j \sum_{k=2}^{n+1} \alpha_{k-1} t_0^{k-1} t\right) \quad (6)$$

Clearly, $\Phi_{\alpha_1, \dots, \alpha_n}^R(t)$ is a frequency rotating operator which rotates the analytical signal $z(t)$ by an angle θ with $\text{tg}(\theta) = -\alpha$, in the time-frequency plane; $\Phi_{\alpha_1, \dots, \alpha_n}^M(t, t_0)$ is the frequency shift operator that relocates a frequency component at ω to $\omega + \alpha t_0$, where $(\alpha_1, \dots, \alpha_n)$ are the polynomial coefficients. As can be seen, PCT uses the polynomial to replace the linear chirp kernels to present better transform performance on the nonlinear instantaneous frequency signals.

As can be seen that many algorithms can be used to transform the audio signals into images; however, it is difficult to choose a best one because each algorithm has its special advantage. Thus, in order to avoid the risk of selecting improper algorithm, the ensemble learning [15] is adopted to aggregate different time-frequency analysis algorithms using a weighted-sum average strategy.

- (3) Lastly, an ensemble CNN-based deep learning model is established based on the time-frequency images of the audio signals to identify different tool wear degrees. The ensemble CNN-based deep learning is described in Fig. 2. In the offline stage, each time-frequency analysis method will generate a set of audio time-frequency images, which will be used to train an individual CNN

model; once all the three CNN models have been well trained, the outputs of the CNN models will be weighted to minimize the tool wear recognition accuracy; as a result of the minimization process, a set of optimal weight coefficients can be obtained. Subsequently, in the online stage, the optimal weight coefficients will be assigned to the CNN models to provide an ensemble identification result. The innovative approach we insist on is to monitor tool wear by using dca-time-frequency analysis-CNN's deep learning model by acquiring audio signals for signal processing and deep learning.

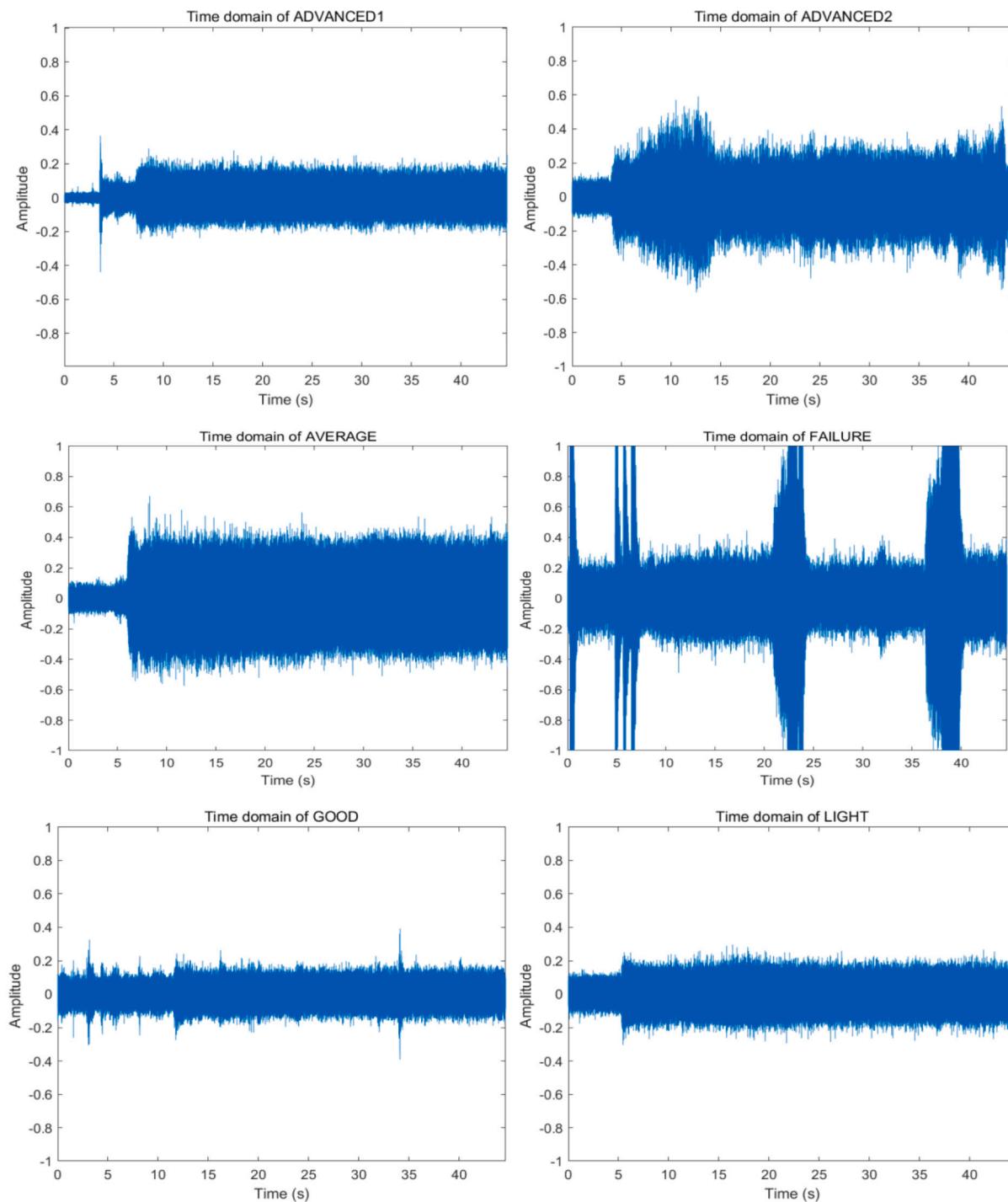
3. Experimental procedure

Tool wear is a gradual process and is subject to large shear forces and friction during cutting. The experiments were conducted on a numerically controlled TRACK K3 EMX conventional milling machine under dry, end milling conditions. All the experimental runs were performed as side milling operations with up-milling (conventional milling) configuration. In the experimental test was carried out in [16], where the uncoated high-speed steel end mill tools were used and the workpiece material adopted 6061 aluminum bar with dimensions in inches for all the experiments. Three condenser microphones (ECOOPRO EO-200) were used to collect the milling audio signals with a sampling frequency of 44.1 kHz positioned at 8, 12, and 15 in. away from the cutting zone in different angles. Table 1 summarizes the specifications for each tool state based on the average wear thickness. More specifically, Table 2 summarizes each tool condition class, five spindle speeds were chosen, and for each spindle, three different feed rates were selected to perform the experiments. Overall, 15 cutting cycles were performed for each tool condition class that totaled up to 90 cutting cycles for the whole experiment [16].

3.1. Audio signal preprocessing

The collected audio signals are in the digital format, which is unstable and often overlaps with other sound sources with unknown arrival time. In order to reduce the noise/useless signal components, a band-pass filter and a DCA denoising model have been established to extract critical tool wear information from the milling audio signals. In this study, each piece of the audio time waveform was cut into a set of sample datasets using a 1s-length window, and then, the amplitude of the signal was normalized between [-1, 1] and the specification unifies the amplitude interval for better observation of the signal characteristics and does not affect the data results. The time domain and frequency domain reproduction of the processed audio signals of the six tool wear degrees were carried out. The obtained time and frequency waveforms are depicted in Fig. 3.

As can be seen from Fig. 3(a) that, the amplitude and period of the noise in different tool wear degrees are different. In the process of tool processing, the vibration fluctuations of the audio will not be too intense, if there is noise interference, the time domain amplitude of the collected sound signal will fluctuate violently, so the sound segment with large sound amplitude fluctuations may contain background noise, then the Fourier transform of the sound segment in this segment displays its frequency domain map. In Fig. 3(b), the amplitude of the noise component varies significantly. In order to build the bandpass filter to remove the background noise, the time and frequency waveforms of the audio signal containing useful background noise components under



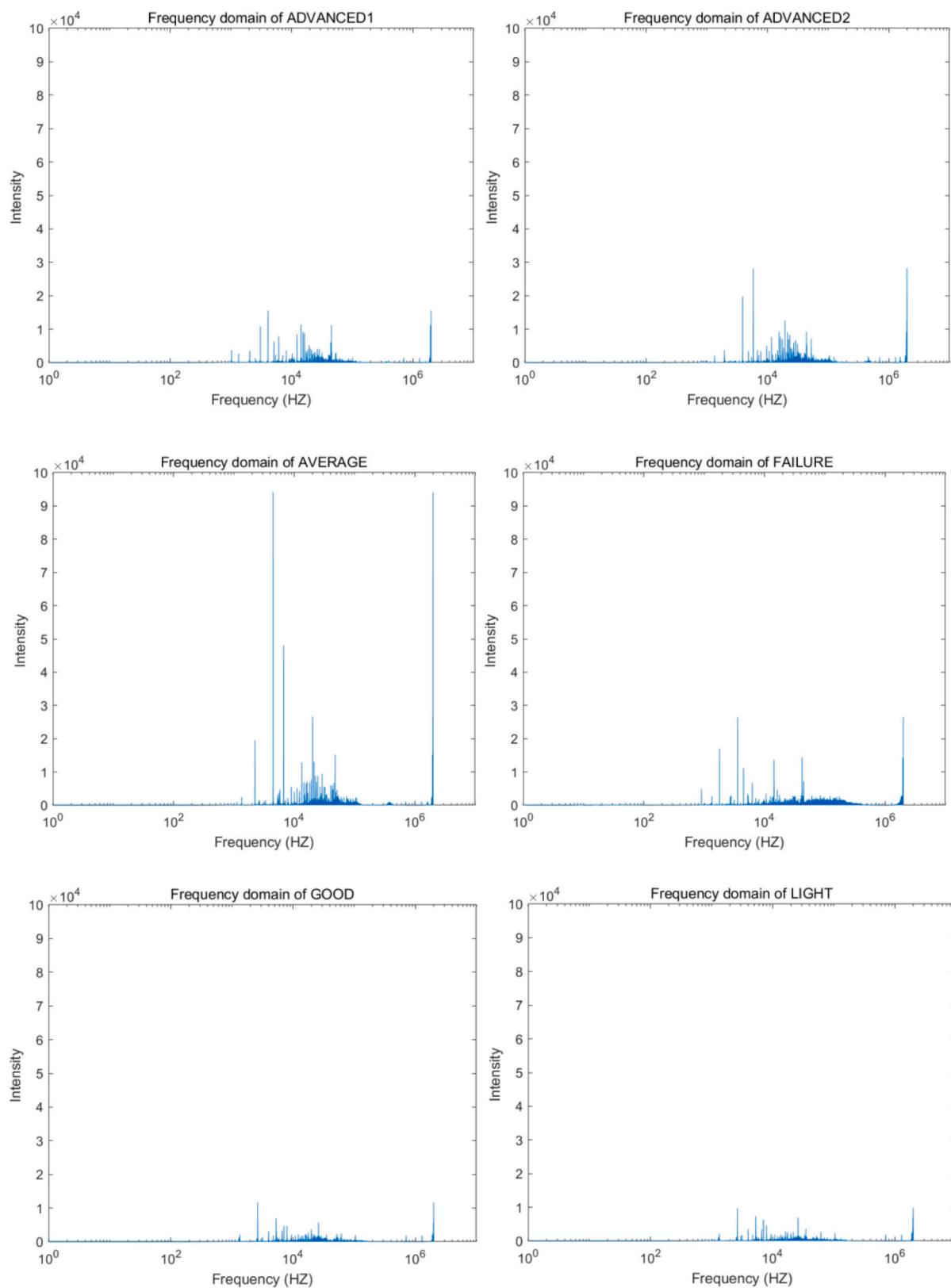
(a) Time waveforms

Fig. 3. Time and frequency curves of the six tool wear degrees.

different tool wear degrees were investigated, as shown in Fig. 4. Please note that in Fig. 4 the range in the y-axis is not the same for every tool operation condition; the reason is that a suitable scale is selected for the y-axis range in each tool wear condition to make it clearly check the details of each curves.

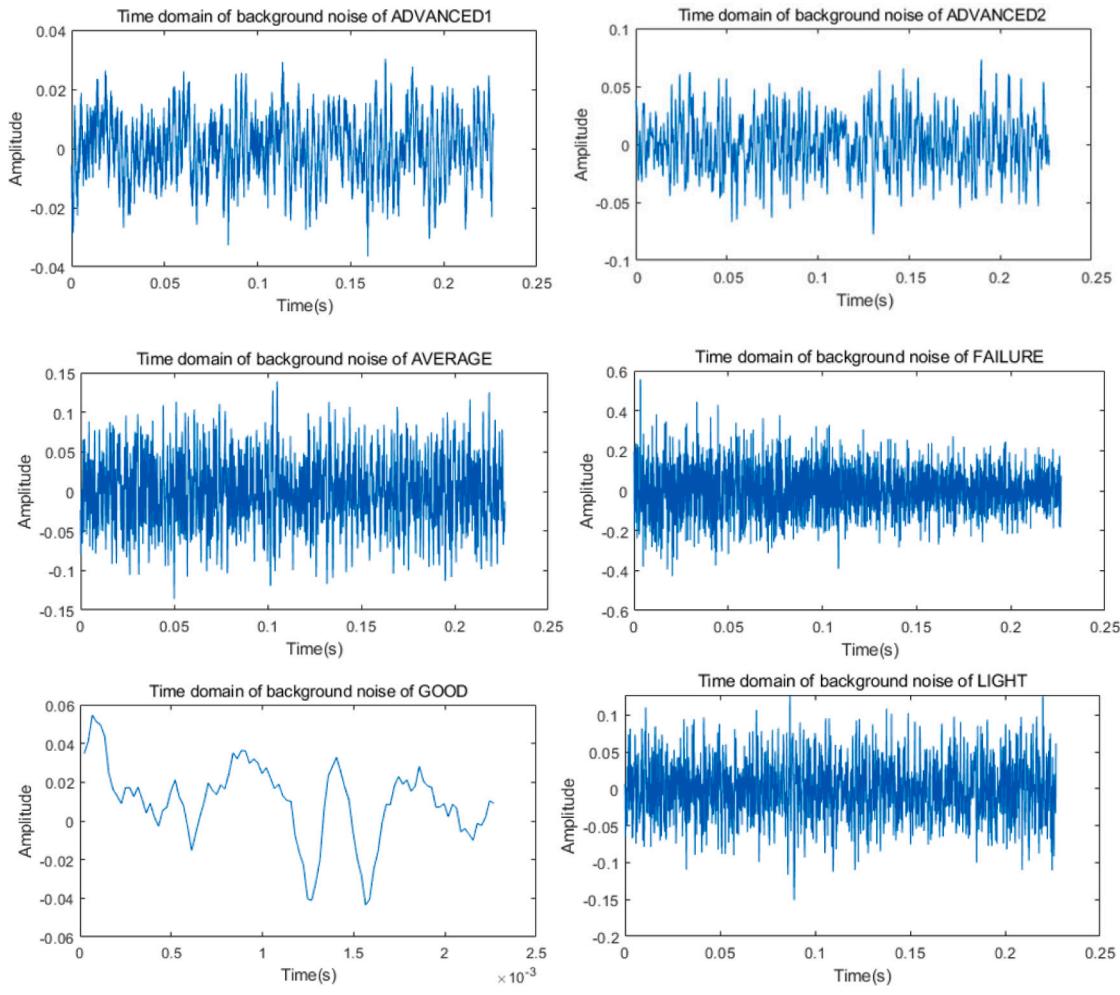
It can be seen from Fig. 4(b) that, the background noise is random and independent, each tool condition has a different background noise for the audio signal acquired during machining, except for the frequency

curve in the Good tool wear condition, the peak values of the background noise components of the other five tool wear conditions are concentrated between 16,000 Hz and 16,500 Hz. When revisit Fig. 3(b) one can note that the peak values of the frequency curves within [16,000 Hz, 16,500 Hz] in the original audio frequency domain shows little significant difference, which indicates that the frequency components in this frequency band present the characteristics of background noise. Therefore, a finite impulse response (FIR)-based bandpass filter



(b) FFT magnitude

Fig. 3. (continued).



(a) Time waveforms

Fig. 4. Time and frequency curves of background noise components.

was built to filter the frequency components from 16,000 Hz to 16,500 Hz. The order of the FIR bandpass filter was set to 70, and the Hamming window was adopted. The sampling rate of the filter was set as 44.1 kHz in order to keep consistent with the audio signal. Fig. 5 provides the time and frequency waveforms of the bandpass filtered audio signals. As can be seen in the Fig. 5, the useless background noise has been depressed effectively by the designed bandpass filter. As a result, the amplitude change of the sound signal time domain after filtering noise is small, and the concentration of the frequency domain is higher. The normalization of signal data is a common method of signal processing. Only synchronous scaling of data through mathematical methods will not change the relationship between data. Although the denoising of signal in this paper will not contaminating the signal by filtering the specific frequency band, it still cannot completely remove all types of the noise. The next step will remove the stochastic noise using the DCA technique.

The ICA is the most popular BSS technique to separate independent source components from a linear mixture of multi-channel signals. Because of its linear mixture nature, the ICA is not applicable to nonlinear mixture or dependent component separation [34]. This is the reason the DCA is employed to deal with the dependent sources in the audio signals. Fig. 6 shows the DCA denoised audio signals under the six different tool wear degrees. Compared with Fig. 3 of the original audio signals, it can be seen that in the denoised audio signals the time and frequency components become much cleaner, which indicates that the

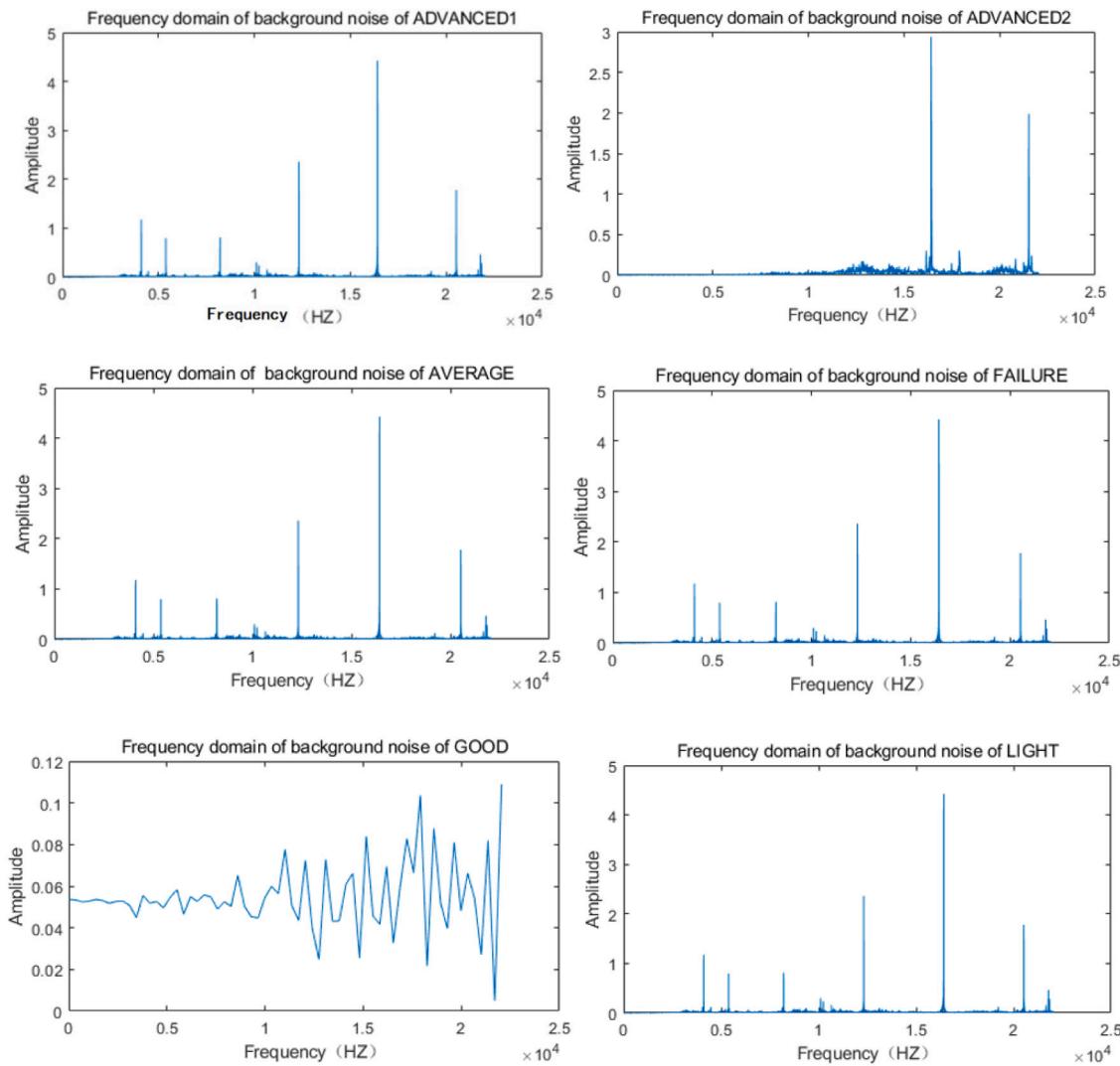
noise interference has been reduced by the proposed denoising method. More importantly, the time waveforms of different tool wear degrees present a clear trend of periodicity, which is well consistent with the theoretical milling audio characteristics. As a result, the extracted audio signal components by the DCA, to a certain degree, can directly describe the tool wear conditions, which will improve the tool wear recognition performance.

3.2. Time-frequency analysis

Firstly, the STFT was used to process the audio signals. Fig. 7 shows the time-frequency distribution of the six tool wear degrees. The energy peaks in the images varies with the tool wear degree, which may be used to distinguish different tool wear degrees. An individual CNN deep learning model will be employing to perform this task.

The polynomial chirplet transform (PCT) is able to extract the dynamic characteristics of the tool wear. Fig. 8 shows the time-frequency distribution of the six tool wear degrees. The amplitude of the energy peaks in the images varies with the tool wear degree. It is difficult to distinguish different tool wear degrees. Another individual CNN deep learning model will be employing to find the inner connection between the image feature maps and the tool wear degree.

In addition to the PCT processing, the WVD algorithm was also adopted to extract the time-frequency images of the tool wear audio



(b) FFT magnitude

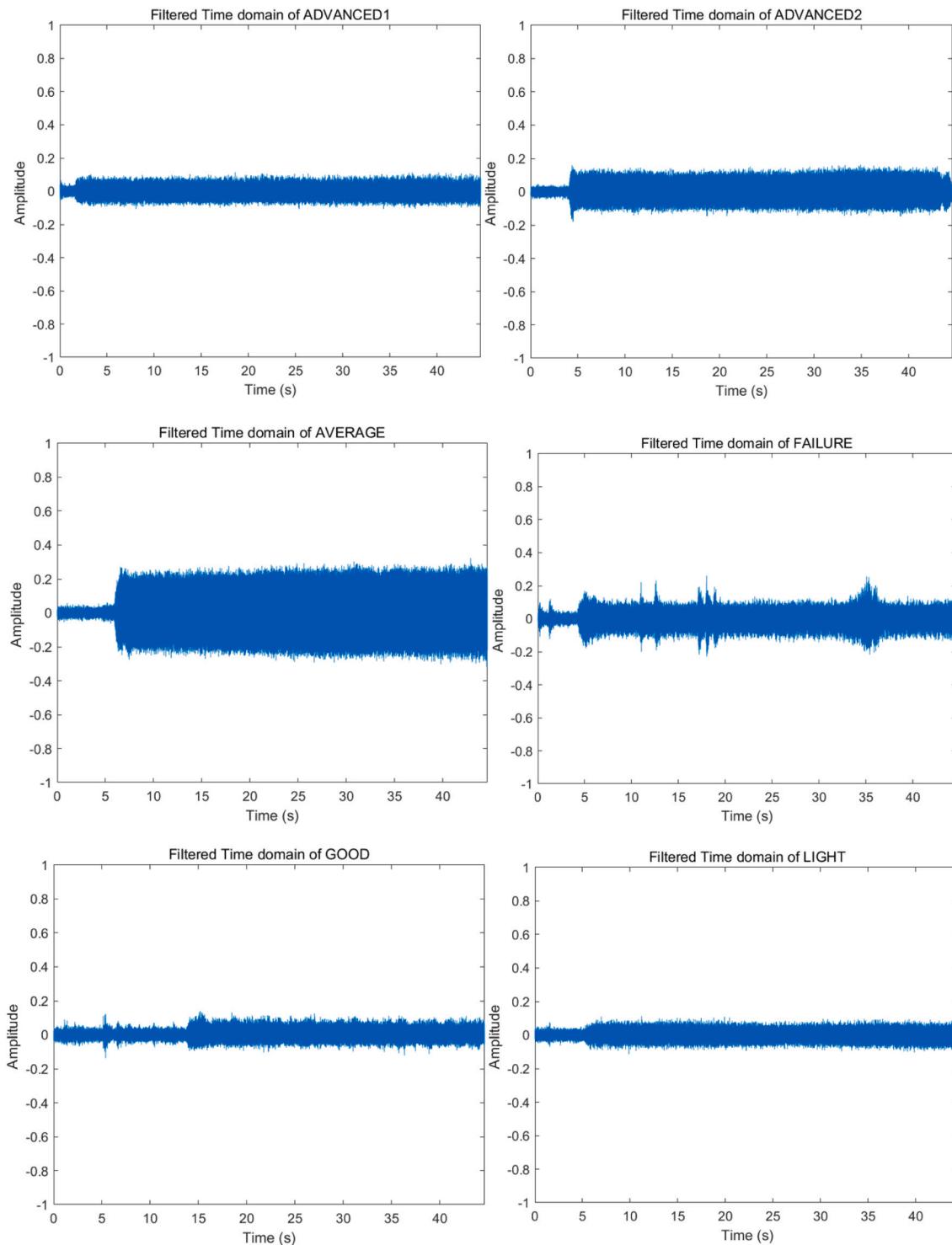
Fig. 4. (continued).

signals. Fig. 9 shows the time-frequency distribution of the six tool wear degrees using the WVD processing. Although the time-frequency pattern varies with the tool wear degree, it is not clear about the connection between the tool wear degree and the pattern. A third individual CNN deep learning model will recognize the tool wear degrees.

3.3. Tool wear degree recognition

Generally, the convolutional neural network (CNN) deep learning frameworks include LeNet5, AlexNet, GoogleNet, VGGNet, and ResNet. Since the LeNet5 framework is able to recognize different patterns directly from the original pixels of the images, this work adopts the LeNet5-based CNN framework, as shown in Fig. 10. The developed CNN has a total of 8 layers. The input layer is a 32×32 -pixel image with a total of 1024 neurons, which is significantly larger than the largest character in the database (i.e., pixels centered on a 28×28 field). The reason is that a potentially unique feature endpoint or corner point may appear at the center of the receiving region of the highest-level feature detector. The first layer (C1) is a convolutional layer, the input 32×32 pixel image is extracted by convolutional image features through 6 convolutional kernels of different sizes (5×5) and a step distance of 1,

and 6 texture maps are obtained, each texture map is 28×28 pixels, the C1 layer has a total of 784 neurons, and the input layer and the C1 layer connection have a total of $6 \times (5 \times 5 + 1) \times (28 \times 28)$ connections. The second layer (S2) is the pooling layer, which is composed of 6 28×28 -pixel texture maps for MaxPool, which has a pool size of (2×2) and a step of 2, and pools 6 feature maps (size 14×14). C3 is a convolutional layer that uses 16 different convolutional kernel sizes of (5×5) with a step of 1 to convolute S2 feature maps to generate 16 texture maps (sizes 10×10), while the S2 and C3 layers connect a total of 151,600 connections. S4 is a pooling layer, input 16 10×10 pixel texture maps for MaxPool, its pool size is (2×2) , the step is 2, pooled to get 16 feature maps (size 5×5). C5 is a convolutional layer that maps the features of S4 into 120 convolutional kernels of different sizes (1×1) in steps of 1 to obtain a convolutional operation containing 120 characteristic maps (size 5×5). F6 is a fully connected layer with 84 units, so its training parameters are $10,164 = 84 \times (120 \times (1 \times 1) + 1)$, and the F6 layer mainly uses the Sigmoid function (commonly used for multi-classification problems) for image feature classification. The neural network's learning rate is also set to $\text{lr} = 0.004$, and the optimizer of the network structure uses SGDM (SGD with momentum) and sets to momentum = 0.9, decay = 0.005, because SGDM is often used for image



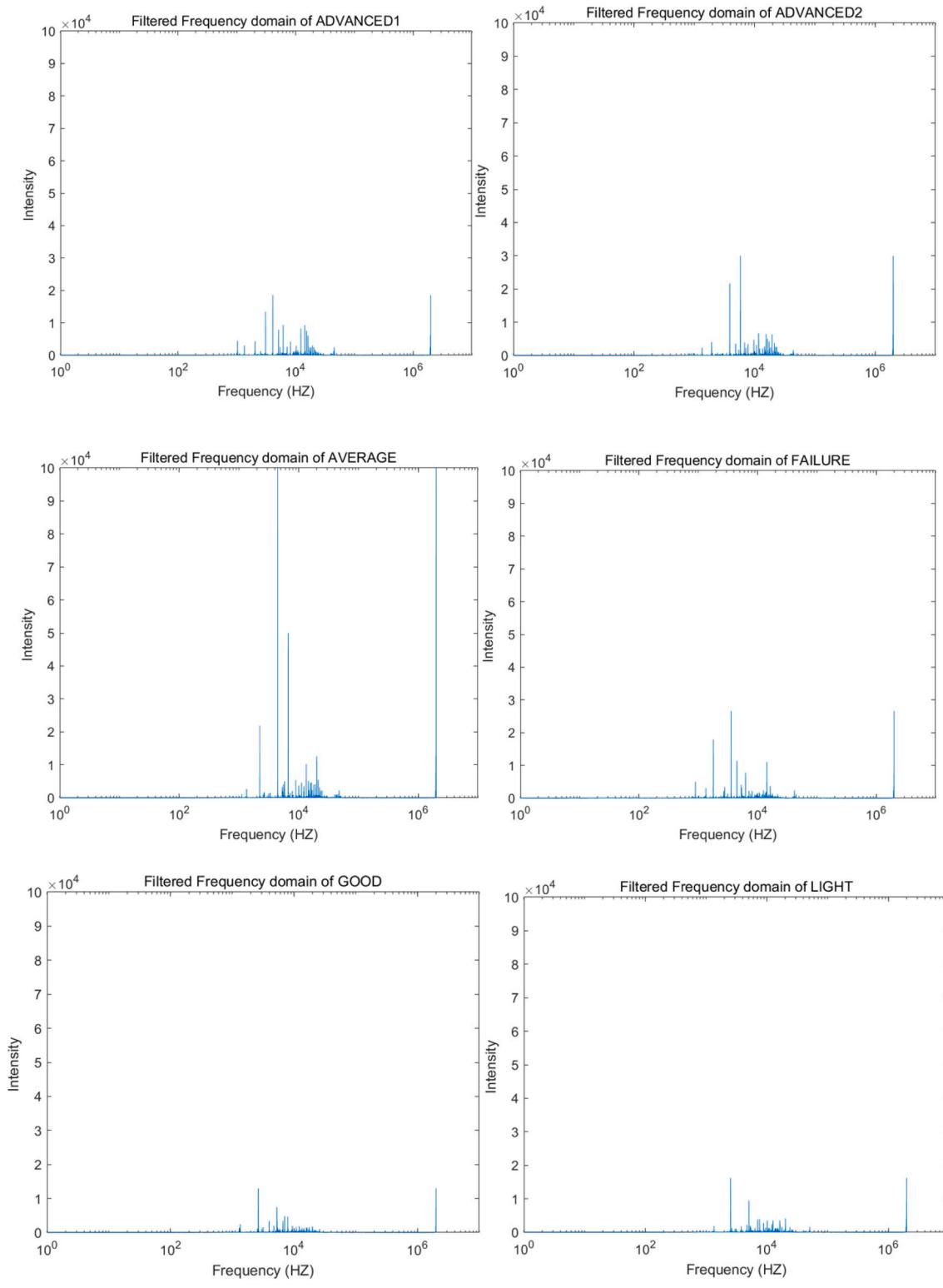
(a) Time waveforms

Fig. 5. Time and frequency curves of bandpass filtered audio signals.

classification. The output layer uses the Euclidean radial basis function. The corresponding labels for the six tool wear degrees are [1–6], see Table 1. The parameters of the above LeNet5 network structure are first used by LetNet5 to carry out transfer learning pre-training using the parameters of the neural network recognized by MNIST handwriting digital images, and then gradually optimize and adjust each parameter

from the actual neural network training, so as to obtain more satisfactory training results.

After the time-frequency analysis, 1000 time-frequency images of each type of tool wear conditions are obtained to train the convolutional neural network model, splitting the image dataset into a training dataset with a scale of 9 to 1 and a test dataset. Firstly, the image size was tailed

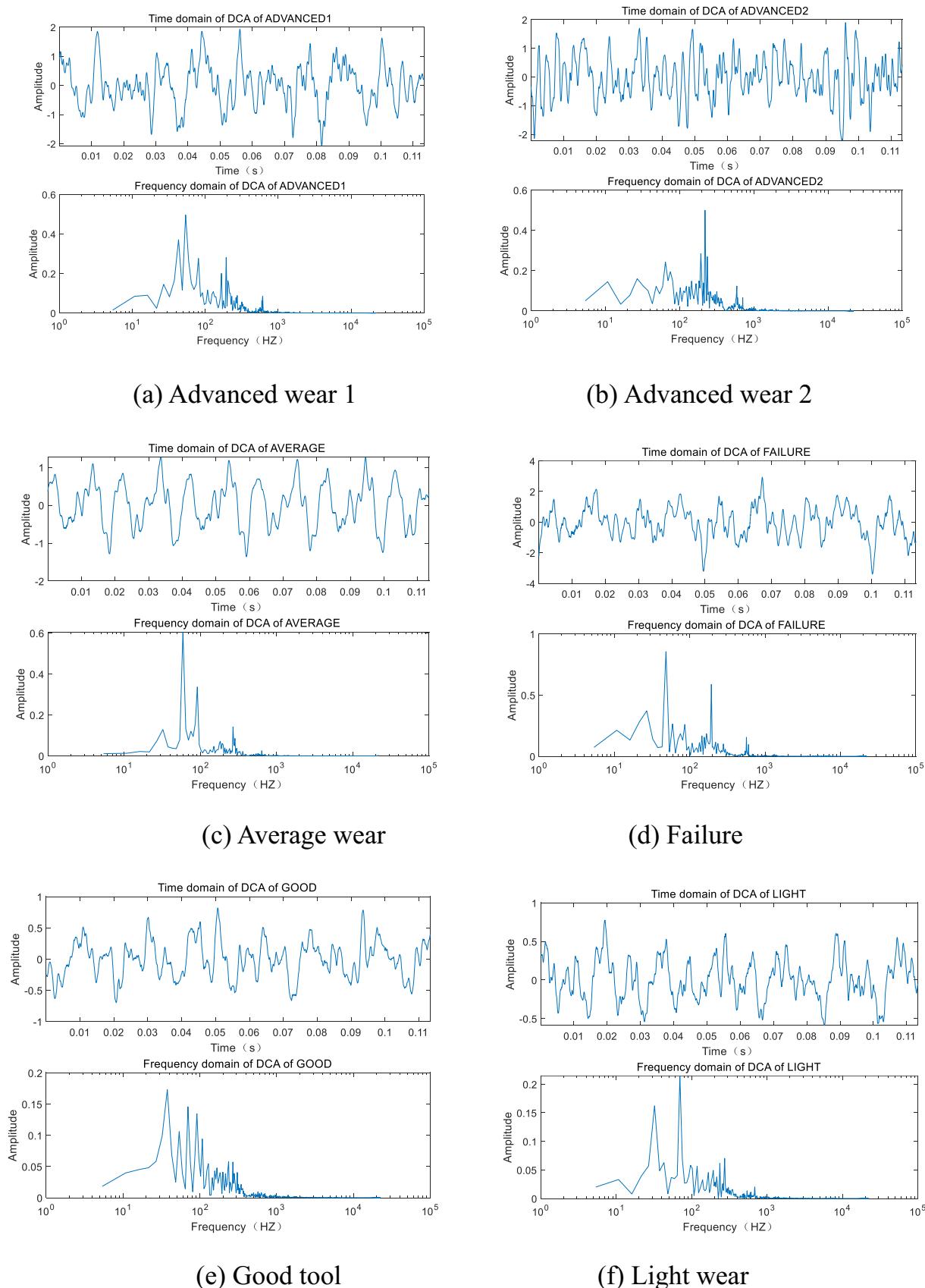


(b) FFT magnitude

Fig. 5. (continued).

according to the CNN input size; then, the gray-processing was applied to the input images to reduce the computational complexity in the convolution layers. The ReLu function was used as the activation layer to deal with the nonlinear relationship between the feature maps and the tool wear degree. In order to prevent overfitting, the Dropout function

was adopted in the CNN training. In the ensemble CNN model, the structure of each CNN model was the same. Figs. 11 and 12 show the training results of the STFT-CNN. Because the training results of the PCT-CNN and WVD-CNN are similar to the STFT-CNN, we did not give all the figures here. As can be seen in Fig. 10, the cross loss of the training

**Fig. 6.** Time and frequency curves of DCA denoised audio signals.

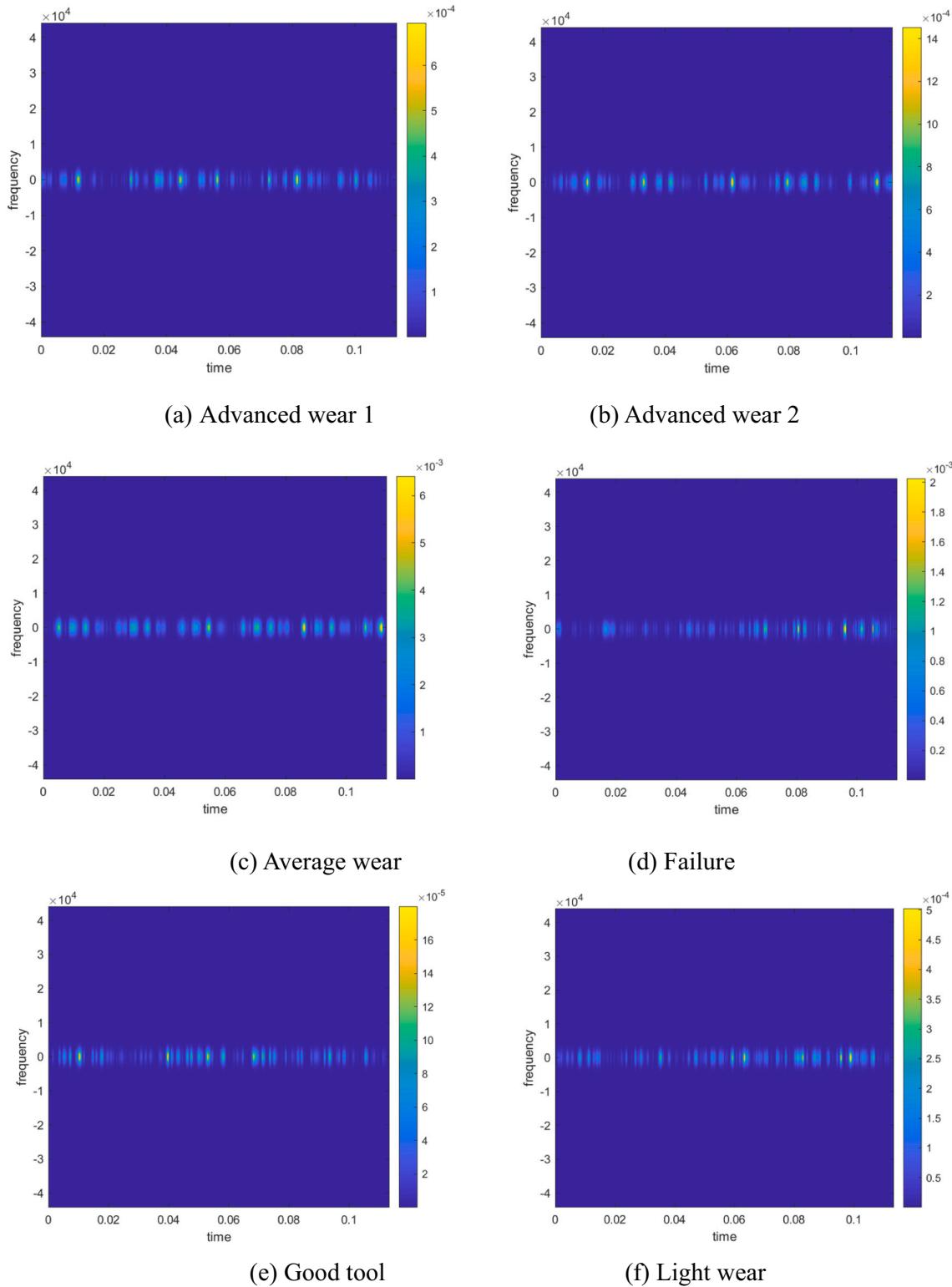


Fig. 7. Time-frequency distribution of different tool wear degrees using STFT.

gradually decreases to zero, while the training accuracy gradually increases to 1. Moreover, in Fig. 12 the identification rates of the six tool wear degrees are at least 97%. All these observations indicate the training of the STFT-CNN model is satisfactory. Then, after training the other PCT-CNN and WVD-CNN models, a weight coefficients optimization process [15] was applied to the recognition outputs of the three CNN models to yield the optimal weights for the ensemble CNN model.

Table 3 shows the tool wear identification results using different monitoring methods. As can be seen in the table, in generally, each individual CNN model performs better than the SVM model in [17] in term of identifying accuracy but worse than the ensemble learning model in [15]. When compared with the SVM model in [17], the detection rates of the individual CNN models are lower for Good tool, Advance 2 and Failure conditions but much higher in the Average and Advance 1

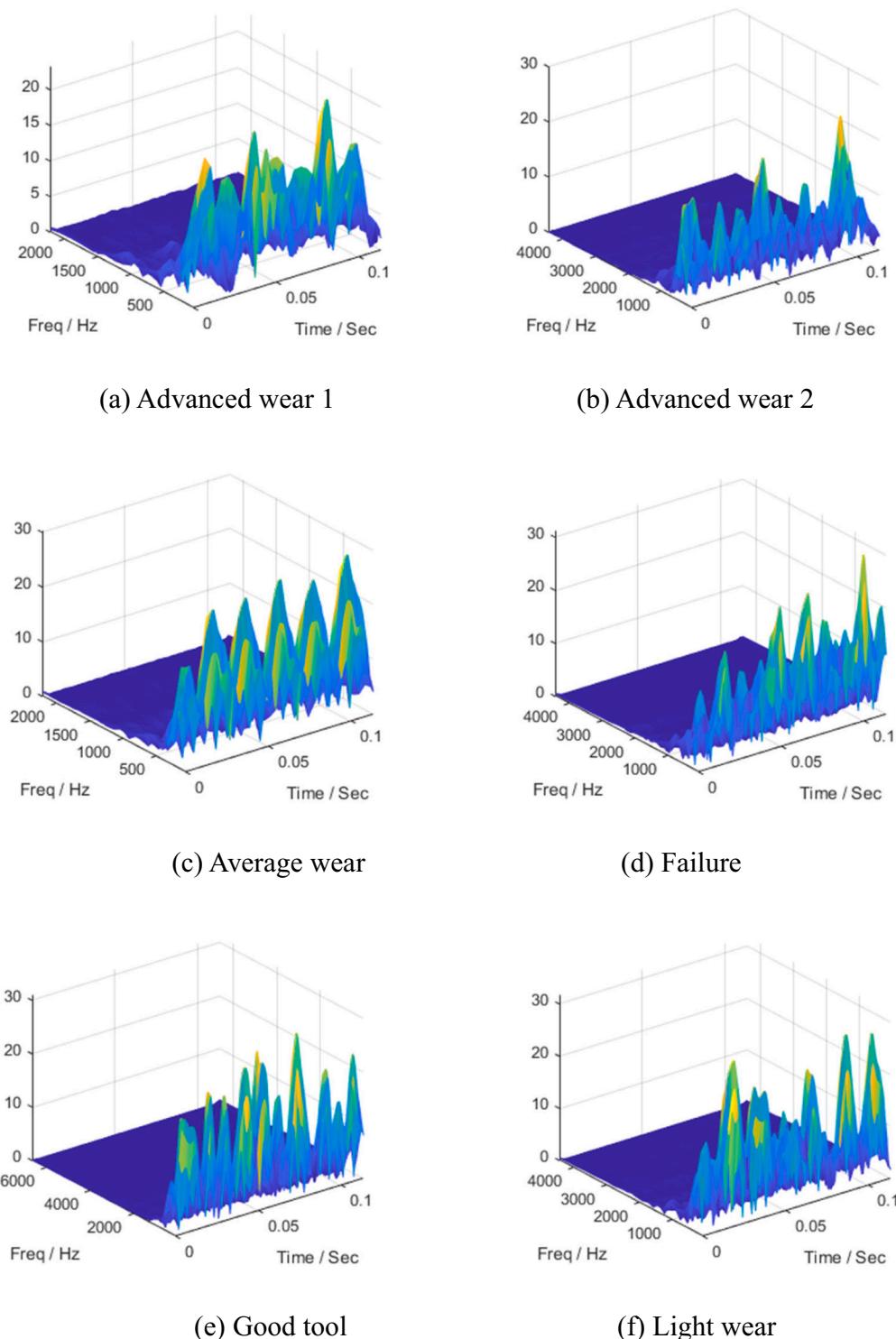


Fig. 8. Time-frequency distribution of different tool wear degrees using PCT.

conditions; as a result, the overall performance of the individual CNN models is better than that of the SVM model in [17]. By closely observing the detection results of the SVM model, one can note that the detection accuracy fluctuates significantly; on the contrary, the individual CNN models operate smoothly and maintain stable detection rate for all the six tool wear conditions. The reason is probably that the SVM does not call a deep learning strategy to comprehensively explore the hidden connection between the audio signals and the tool wear conditions. However, when comparing the ensemble learning model in [15] with

the individual CNN models, their performance is comparable; and the ensemble learning model even performs a little better than each individual CNN model. This is because the ensemble learning is able to enhance the overall capacity of several weak learners to form a stronger learner. As a result, even the ensemble learning model in [15] did not employ any deep learning strategy, the ensemble detection results can outperform a single deep learning model. Inspired by this ensemble conception, in this work an ensemble CNN model was proposed by weighting the three individual CNN models. As can be seen in the table,

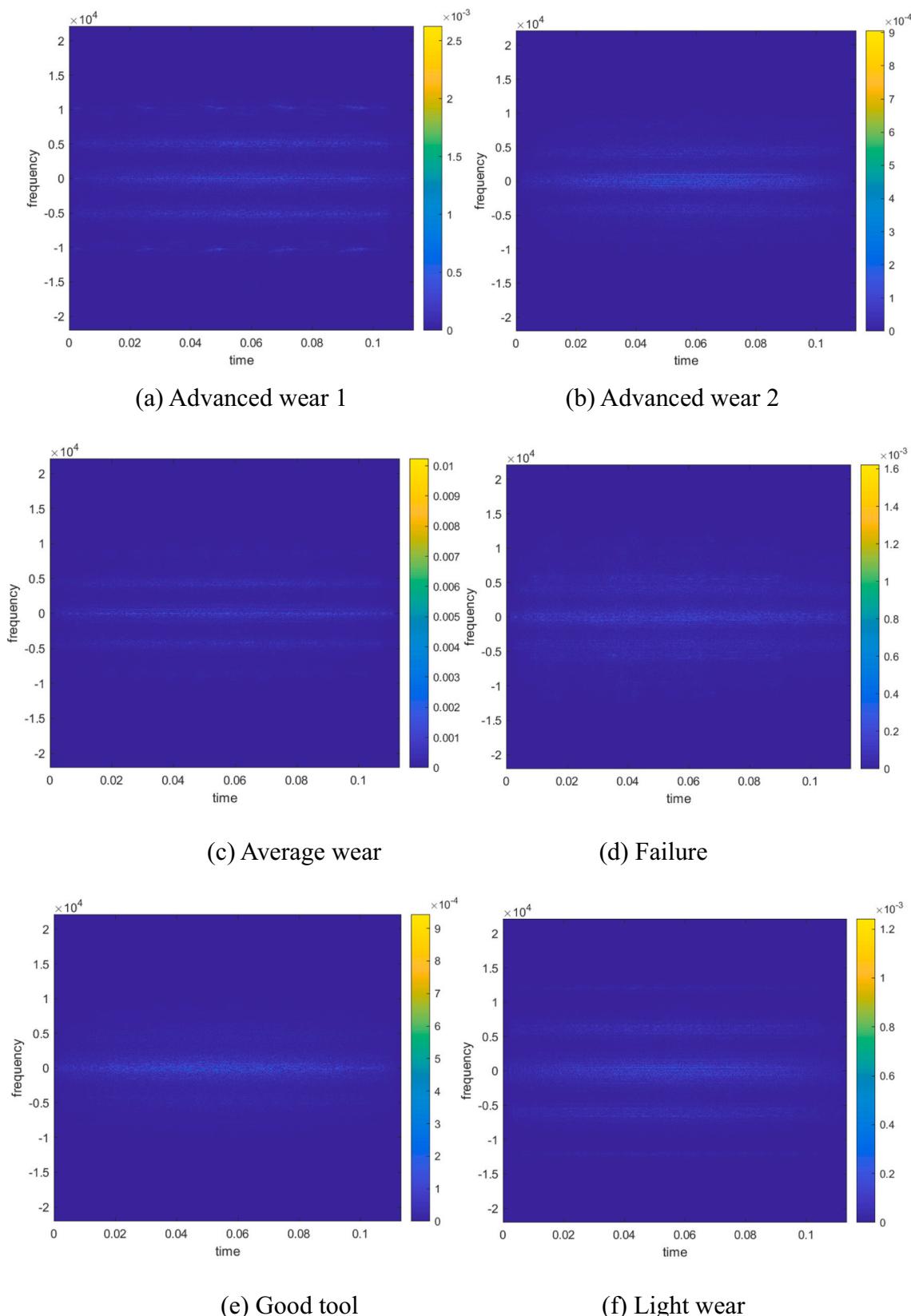


Fig. 9. Time-frequency distribution of different tool wear degrees using WVD.

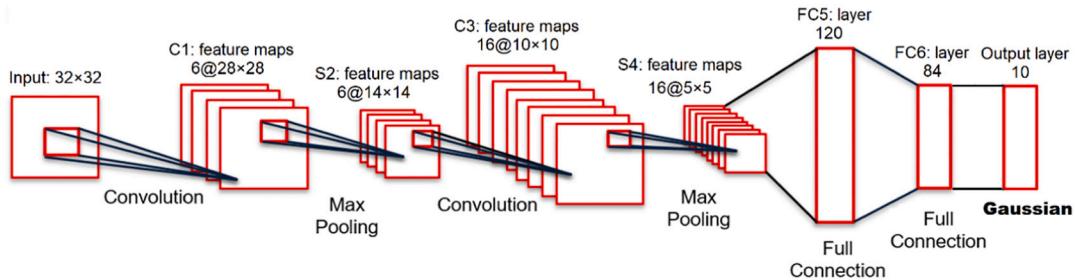


Fig. 10. Structure of the developed CNN.

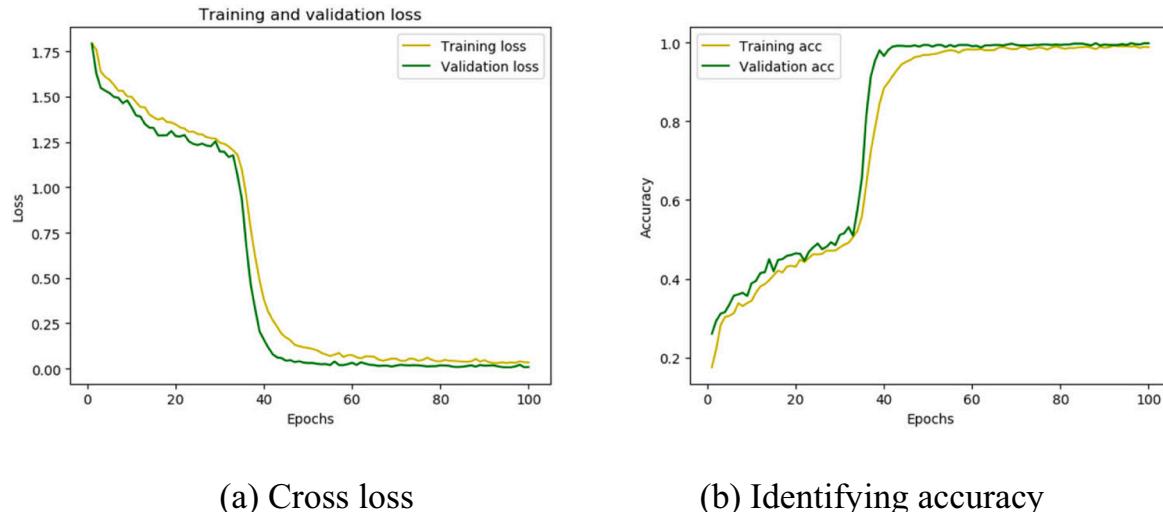


Fig. 11. Training results of STFT-CNN model.

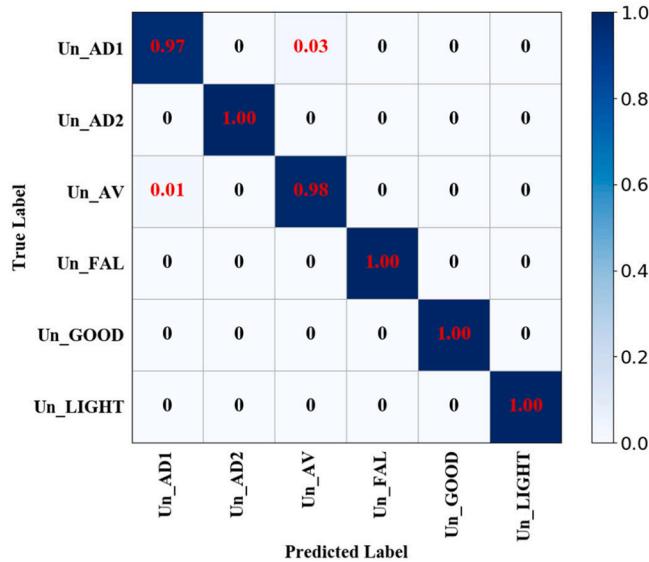


Fig. 12. Training performance of the STFT-CNN model.

thanks of the deep learning strategy, the ensemble CNN outperforms all the other five models. Comparing with the individual STFT-CNN, PCT-CNN and WVD-CNN models, the detection rate of the ensemble CNN has been improved by 2.5%, 1.3% and 1.2%, respectively. The improvement mechanism of the ensemble CNN can be explained by the following example. For one piece of audio signal in the Advance 1 condition, the

idea output of the ensemble CNN is 4; the actual outputs of the STFT-CNN, PCT-CNN and WVD-CNN were respectively [4.9, 3.8, 3.2] while the optimal weights after the offline training were [0.22, 0.63, 0.15]; then, the ensemble output was 3.952 ($=4.9 \times 0.22 + 3.8 \times 0.63 + 3.2 \times 0.15$). Thus, the output of the ensemble CNN was very close to the idea value 4, while the STFT-CNN and WVD-CNN failed to identify the tool wear condition. One can note from this example that even the individual CNN models cannot correctly identify the tool wear condition, the ensemble CNN is able to produce the correct result by weighted-sum the individual CNN models. More importantly, as well known, a 1.0% improvement in the detection rate in practical applications will eventually save millions dollar considering the huge amount of tool maintenance cost in industry. As a result, the proposed ensemble CNN method for tool wear monitoring has promising potential in practical applications.

4. Conclusions

This study proposed a new method for tool wear monitoring based on the audio signal analysis. A novel DCA-ensemble CNN model is developed for intelligent and accurate tool wear degree identification. Experimental analysis results demonstrate that (1) with the present audio denoising approach, the wear detection rate has been significantly increased; (2) the ensemble CNN performs the ‘black-boxing’ learning, which does not require any analytic solutions and professional knowledge on the tool wear degradation mechanisms; (3) as a result, with such kind of intelligent learning, the ensemble CNN model is able to establish a reliable connection between the audio signal features and the tool wear conditions; and (4) the tool wear detection performance of the proposed method has been improved by at least 1.2% when compared

Table 3

Comparisons of tool monitoring results with different monitoring methods.

Condition	Results in [17]	Results in [15]	STFT-CNN	PCT-CNN	WVD-CNN	Ensemble CNN
Good	99.4%	100%	98.2%	98.4%	97.9%	99.3%
Slight	97.5%	98.29%	98.7%	97.8%	98.5%	99.1%
Average	90.7%	98.03%	91.8%	98.3%	98.5%	98.7%
Advance 1	92.4%	98.63%	95.8%	96.4%	96.1%	98.2%
Advance 2	98.5%	98.66%	97.5%	97.7%	97.9%	98.8%
Failure	97.0%	97.20%	94.9%	95.3%	95.7%	97.7%
Overall	95.92%	98.47%	96.15%	97.32%	97.43%	98.63%

with individual CNN models and popular existing methods.

The future plan will continue modifying the proposed audio-based tool wear monitoring method. The tool wear prediction will be investigated to enable just-in-time maintenance. On-line monitoring for a real-world machine tool will be carried out in the future.

Declaration of competing interest

The work presented in this paper is, to the best of my knowledge and belief, original, except as acknowledged in the text, and the material has not been submitted, either in whole or in part, for a degree at this or any other university. There is no conflict of interest to declare.

Acknowledgement

The authors would like to express sincere thanks to Dr. R. Liu at Rochester Institute of Technology and Dr. D. Wu at University of Central Florida for sharing the experimental datasets. This work is supported by Fundamental Research Funds for the Central Universities (No. 201912036), Taishan Scholar of Shandong, China (No. tsqn201812025), Narodowego Centrum Nauki, Poland (No. 2020/37/K/ST8/02748).

References

- [1] Erden MA, Yaşar N, Korkmaz ME, Ayvaci B, Nimel Sworna, Ross K, Mia M. Investigation of microstructure, mechanical and machinability properties of milled steel produced by powder metallurgy method. *Int J Adv Manuf Technol* 2021;114:2811–27.
- [2] Kuntoğlu M, Sağlam H. Investigation of signal behaviors for sensor fusion with tool condition monitoring system in turning. *Measurement* 2021;173:108582.
- [3] Fong KM, Wang X, Kamaruddin S, Ismadi MZ. Investigation on universal tool wear measurement technique using image-based cross-correlation analysis. *Measurement* 2021;169:108489.
- [4] Nazir Q, Shao C. Online tool condition monitoring for ultrasonic metal welding via sensor fusion and machine learning. *J Manuf Process* 2021;62:806–16.
- [5] Ghosh N, Ravi YB, Patra A, Mukhopadhyay S, Paul S, Mohanty AR, Chattopadhyay AB. Estimation of tool wear during CNC milling using neural network-based sensor fusion. *Mech Syst Sign Process* 2007;21(1):466–79.
- [6] Sick B. On-line and indirect tool wear monitoring in turning with artificial neural networks: a review of more than a decade of research. *Mech Syst Sign Process* 2002;16(4):487–546.
- [7] Yaşar N, Korkmaz ME, Günay M. Investigation on hole quality of cutting conditions in drilling of CFRP composite. In: MATEC Web of Conferences. 112; 2017. p. 01013.
- [8] Boy M, Yaşar N, Ciftci I. Experimental investigation and modelling of surface roughness and resultant cutting force in hard turning of AISI H13 steel. *IOP Conf Ser Mater Sci Eng* 2016;161:012039.
- [9] Chao BT. The significance of the thermal number in metal machining. *Trans. ASME* 1953;75:109–20.
- [10] Dimla DE. The correlation of vibration signal features to cutting tool wear in a metal turning operation. *Int J Adv Manuf Technol* 2002;19(10):705–13.
- [11] Silva RG, Baker KJ, Wilcox SJ, Reuben RL. The adaptability of a tool wear monitoring system under changing cutting conditions. *Mech Syst Sign Process* 2000;14(2):287–98.
- [12] Twardowski P, Tabaszewski M, Wiciak-Pikula M, Felusiak-Czryca A. Identification of tool wear using acoustic emission signal and machine learning methods. *Precis Eng* 2021;72:738–44.
- [13] Lozano RAG, Llumbreras PDA, Troncoso RDJR, Ruiz GH. An object-oriented architecture for sensorless cutting force feedback for CNC milling process monitoring and control. *Adv Eng Softw* 2010;41(5):754–61.
- [14] Durmuş HK, Özkaya E, Meri C. The use of neural networks for the prediction of wear loss and surface roughness of AA 6351 aluminium alloy. *Mater Des* 2006;27(2):156–9.
- [15] Li Z, Liu R, Wu D. Data-driven smart manufacturing: tool wear monitoring with audio signals and machine learning. *J Manuf Process* 2019;48:66–76.
- [16] Kothuru A, Nooka SP, Liu R. Audio-based tool condition monitoring in milling of the workpiece material with the hardness variation using support vector machines and convolutional neural networks. *J Manuf Sci Eng* 2018;140(11):111006.
- [17] Kothuru A, Nooka SP, Liu R. Application of audible sound signals for tool wear monitoring using machine learning techniques in end milling. *Int J Adv Manuf Technol* 2018;95(9):3797–808.
- [18] Stuhrl B, Liu R. A flexible similarity-based algorithm for tool condition monitoring. *J Manuf Sci Eng* 2021;144(3):031010.
- [19] Ferroni G, Turpault N, Azcarreta J, Tuveri F, Serizel R, Bilen Ç, Krstulović S. Improving sound event detection metrics: insights from dcase 2020. In: ICASSP 2021–2021 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE; 2021, June. p. 631–5.
- [20] Wu X, Li J, Jin Y, Zheng S. Modeling and analysis of tool wear prediction based on SVD and BiLSTM. *Int J Adv Manuf Technol* 2020;106(9):4391–9.
- [21] Chen GS, Zheng QZ. Online chatter detection of the end milling based on wavelet packet transform and support vector machine recursive feature elimination. *Int J Adv Manuf Technol* 2018;95(1):775–84.
- [22] Ou J, Li H, Huang G, Yang G. Intelligent analysis of tool wear state using stacked denoising autoencoder with online sequential-extreme learning machine. *Measurement* 2021;167:108153.
- [23] Li X, Patri KV. Wavelet packet transforms of acoustic emission signals for tool wear monitoring. *J Manuf Sci Technol* 1999;1(2):89–93.
- [24] Kim J, Lee H, Jeong S, Ahn SH. Sound-based remote real-time multi-device operational monitoring system using a convolutional neural network (CNN). *J Manuf Syst* 2021;58:431–41.
- [25] Jie Z, Zhen-yang W, Hao M. A smoothing method of head-related transfer functions based on reconstruction from wavelet transform modulus maxima. *Dianzi Yu Xinxi Xuebao/J Electr Inform Technol* 2007;29(2):473–7.
- [26] Jiang Z, Sun C, Zhu Z. The monitoring of milling tool tipping by estimating holder exponents of vibration. *IEEE Access* 2020;8:96661–8.
- [27] Prakash K, Samraj A. Tool flank wear estimation by simplified SVD on emitted sound signals. In: 2017 Conference on Emerging Devices and Smart Systems (ICEDSS). IEEE; 2017, March. p. 1–5.
- [28] Gomes MC, Brito LC, da Silva MB, Duarte MAV. Tool wear monitoring in micromilling using support vector machine with vibration and sound sensors. *Precis Eng* 2021;67:137–51.
- [29] Zhou Y, Sun B, Sun W, Lei Z. Tool wear condition monitoring based on a two-layer angle kernel extreme learning machine using sound sensor for milling process. *J Intel Manuf* 2020;1:1–12.
- [30] Rafezi H, Akbari J, Behzad M. Tool condition monitoring based on sound and vibration analysis and wavelet packet decomposition. In: 2012 8th International Symposium on Mechatronics and its Applications. IEEE; 2012, April. p. 1–4.
- [31] Cooper C, Wang P, Zhang J, Gao RX, Roney T, Ragai I, Shaffer D. Convolutional neural network-based tool condition monitoring in vertical milling operations using acoustic signals. *Procedia Manuf* 2020;49:105–11.
- [32] Guo K, Sun J. Sound singularity analysis for milling tool condition monitoring towards sustainable manufacturing. *Mech. Syst. Signal Process.* 2021;157:107738.
- [33] Comon P. Independent component analysis, a new concept? *Signal Process* 1994;36(3):287–314.
- [34] Kuruoğlu EE, Theis FJ. Dependent component analysis. *EURASIP J Adv Sign Process* 2013;2013:185.
- [35] Erdogan AT. A class of bounded component analysis algorithms for the separation of both independent and dependent sources. *IEEE Trans Sign Process* 2013;61(22):5730–43.
- [36] Georgiev P, Theis F, Cichocki A. Sparse component analysis and blind source separation of underdetermined mixtures. *IEEE Trans Neural Netw* 2005;16(4):992–6.
- [37] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Proces. Syst.* 2012;25:1097–105.
- [38] Peng R, Liu J, Fu X, Liu C, Zhao L. Application of machine vision method in tool wear monitoring. *Int J Adv Manuf Technol* 2021;116(3):1357–72.

- [39] Bradley C, Wong YS. Surface texture indicators of tool wear-a machine vision approach. *Int J Adv Manuf Technol* 2001;17(6):435–43.
- [40] Dai Y, Zhu K. A machine vision system for micro-milling tool condition monitoring. *Precis Eng* 2018;52:183–91.
- [41] Chethan YD, Ravindra HV, Kumar SB. Machine vision for tool status monitoring in turning inconel 718 using blob analysis. *Mater Today: Proc* 2015;2(4–5):1841–8.
- [42] Ong P, Lee WK, Lau RJH. Tool condition monitoring in CNC end milling using wavelet neural network based on machine vision. *Int J Adv Manuf Technol* 2019; 104(1):1369–79.
- [43] Boashash B, Black P. An efficient real-time implementation of the wigner-ville distribution. *IEEE Trans Acoust Speech Signal Process* 1987;35(11):1611–8.
- [44] Peng ZK, Meng G, Chu FL, Lang ZQ, Zhang WM, Yang Y. Polynomial chirplet transform with application to instantaneous frequency estimation. *IEEE Trans Instrum Meas* 2011;60(9):3222–9.