Time-frequency Representation -enhanced Transfer Learning for Tool Condition Monitoring during milling of Inconel 718

Indexed by:
Web of Science Group

Yuqing Zhou[a,b,*], Wei Sun[c], Canyang Ye[d], Bihui Peng[d], Xu Fang[d], Canyu Lin[d], Gonghai Wang[b], Anil Kumar[c], Weifang Sun[c]

[a] School of information science and technology, Northwest University, China
[b] College of Mechanical and Electrical Engineering, Jiaxing Nanhu University, China
[c] College of mechanical and electrical engineering, Wenzhou University, China
[d] Zhejiang Meishuo Electric Technology Co., Ltd, China

Highlights

- A time-series dimension expansion and TL method is proposed to improve the performance of TCM for small samples.

- A time-frequency Markov transition field is proposed to encode the cutting force signal to two-dimensional color images, enriching the information of time-series dimension expansion and imaging.

- The proposed method outperforms four state-of-the-art methods for small samples by the PHM 2010 TCM dataset.

Abstract

Accurate tool condition monitoring (TCM) is important for the development and upgrading of the manufacturing industry. Recently, machine-learning (ML) models have been widely used in the field of TCM with many favorable results. Nevertheless, in the actual industrial scenario, only a few samples are available for model training due to the cost of experiments, which significantly affects the performance of ML models. A time-series dimension expansion and transfer learning (TL) method is developed to boost the performance of TCM for small samples. First, a time-frequency Markov transition field (TFMTF) is proposed to encode the cutting force signal in the cutting process to two-dimensional images. Then, a modified TL network is established to learn and classify tool conditions under small samples. The performance of the proposed TFMTF-TL method is demonstrated by the benchmark PHM 2010 TCM dataset. The results show the proposed method effectively obtains superior classification accuracies for small samples and outperforms other four benchmark methods.

Keywords

tool condition monitoring, time-frequency analysis, Markov Transition Field, transfer learning

1. Introduction

CNC machine is an important equipment for intelligent manufacturing. As a key component of CNC machines, cutting tools are most easily to be damaged and wasted 12. Unexpected damage to tools will result in additional downtime or even severe accidents. Friction between the tool and the workpiece is inevitable during the cutting process. With the increase of tool wear, cutting force, cutting heat and cutting vibration will increase or rise, thus reducing the surface quality of the workpiece. SAU *et al* 3 investigated the effect of machining parameters of Al/TiN coated carbide tools in milling Cu-B-CrC composites and the amount of powder with different weight ratios used for manufacturing machining reinforced samples on surface roughness, tool wear, chip morphology and cutting temperature, which are important for improving the machining

(*) Corresponding author.
E-mail addresses: Y. Zhou (ORCID: 0000-0002-8580-5427) zhouyq@wzu.edu.cn, W. Sun: sunwei0977@163.com, C. Ye: 57847891@qq.com, B. Peng: pengbhui2008@126.com, X. Fang: 1196011993@qq.com, C. Lin: lcy@msrelay.com, G. Wang (ORCID: 0000-0003-1581-6267) wang_gonghai@jxnhu.edu.cn, A. Kumar (ORCID: 0000-0001-6675-1657) 20210129@wzu.edu.cn, W. Sun (ORCID: 0000-0002-4181-1606) vincent_suen@126.com

quality of the parts. Frequent tool replacement to ensure machining accuracy not only seriously weakens productivity, but also easily causes waste of resources. Jae-Woong et al 4 investigated the relationship between dynamic/static cutting force components of machine tools and tool rear face wear and crescent pits wear, between cutting forces and cutting conditions for tool condition monitoring. With the increasing advancement of artificial intelligence (AI) technology, more and more researchers are applying AI to the field of tool condition monitoring. Danil *et al* 5 provides a valuable discussion of the research in the field of tool condition monitoring and a comprehensive analysis of the main advantages, disadvantages and prospects of various artificial intelligence methods. An accurate and efficient tool condition monitoring (TCM) method is necessary for ensuring the machining quality and reducing the machining costs 67.

Non-stationary signals are often produced since mechanical equipment is becoming increasingly complicated. With the continuous promotion of computing power and advanced algorithms, data-driven machine-learning (ML) methods have shown extraordinary advantage to deal with large non-stationary signals and have always been a research hot-pot in manufacturing processes 89. For instance, Yu developed a weighted HMM based method for TCM and tool remaining useful life prognosis 10. Benkedjouh *et al*. developed a method based on feature reduction and SVR for tool health assessment 11. Ma *et al*. proposed a deep convolution-based LSTM model to predict the remaining useful life of rotating machinery 12.

However, for complex non-stationary sensing data, traditional signal analysis methods are only capable of providing statistical average analysis in the time domain or frequency domain, which is difficult to reveal the comprehensive characteristics in both time and frequency domains 1314. In contrast, Time-frequency analysis can identify the frequency components of a signal and reveal its time-varying characteristics, and is an effective method for extracting mechanical health information contained in complex signals. Zhang *et al*. employed continuous wavelet transform to convert 1-D signals into 2-D images to extract wavelet power spectrum of bearing faults 15. Cheng *et al*. explored a data-driven neural network model based on continuous wavelet transform and local binary convolution for intelligent fault diagnosis of rotating machinery 16.

Unfortunately, these data-driven ML methods rely on a large number of labeled training samples, which is difficult for TCM to do in practical due to the high experimental and manual labeling costs 1718. The classification accuracy of these ML methods is not satisfactory under small labeled training samples 1920. Zhou et al21 proposed an improved edge-labeled graph neural network for tool condition monitoring with small sample data by reconstructing and expanding the processed cutting force data into multidimensional data by phase space reconstruction. Currently, transfer learning (TL) enabled methods have been considered to be an effective way to solve this problem 22. Marei *et al*. proposed a transfer learning based CNN framework for tool wear condition monitoring 23. Yang *et al*. proposed a feature-based TL network to transfer the knowledge learned from experiments to real cases for health status identification 24.

To improve the classification performance of DL-based methods for TCM, a time-frequency Markov transition field (TFMTF) algorithm is proposed to encode raw signals into two-dimensional (2D) color images containing both time and frequency information, which can make full use of the excellent ability of DL in image processing. The TFMTF-TL TCM method proposed in this paper requires only a small amount of labeled data to achieve intelligent diagnosis under the same working conditions, and the training process does not require much computation and is highly efficient. In addition, the method based on model parameter transfer does not require pre-constructed neural network, which can be directly reused, and loading pre-trained model parameters can initialize the network and accelerate the model fitting speed.

The rest of this paper is organized as follows. The proposed methodology is described in Section 2. In Sections 3, the PHM 2010 public dataset are applied to test the validity of the proposed method. The conclusion is presented in Section 4.

2. Method

2.1. Markov Transition Field

Wang and Oates first proposed the concept of the Markov Transition Field (MTF) 25, as described below. The MTF is an effective method of encoding 1D time-series data into 2D

images in timeline order. The MTF is closely related to the Markov chain 26, Figure 1 shows the state transfer diagrams corresponding to Markov chains in the three states. It uses the state-transition diagram or state-transition matrix (also called the Markov transition matrix) to describe the probabilities of a state transitioning to itself or other states.

Generally, suppose $S = \{s_1, s_2, \cdots, s_q\}$ denotes a Markov chain with q states, and its transition matrix $P_{q \times q}$ can be set as Eq.(1), in which p_{ij} denotes the one-step transition probability from state s_i to state s_j, $\sum_i p_{ij} = 1$, and $1 \leq i,j \leq q$.
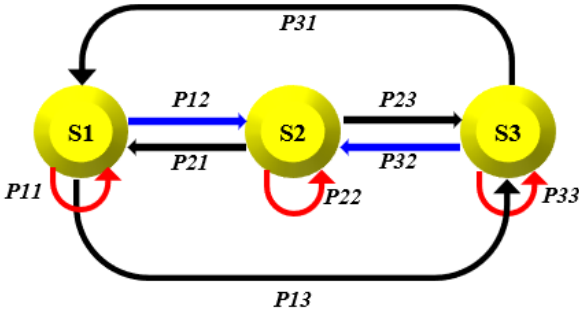

Fig. 1 A 3-state Markov chain with 9 state transition probabilities.

$$p_{q \times q} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1q} \\ p_{21} & p_{22} & \cdots & p_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ p_{q1} & p_{q2} & \cdots & p_{qq} \end{bmatrix} \qquad (1)$$

Given a time series $X = \{x_1, x_2, \cdots, x_n\}$, x_i denotes the data points at time step i. x_i can be assigned initially to a matching state s_i (or a quantile bin q_j) using the min-max standardization method, where $1 \leq j \leq q$, and q is the state number. In this case, an $q \times q$ state transition matrix $P_{q \times q}$ of the time series X can be obtained through counting c_{ij} firstly, which is the number of data points in the transition from state s_i to state s_j. Afterward, p_{ij} can be deduced as $p_{ij} = \frac{c_{ij}}{\sum_i c_{ij}}$. The MTF captures the multi-step transition probabilities between any two data points of X during the construction of the two-dimensional data. It is an $n \times n$ matrix, as shown in Eq. (2). $M_{n \times n} = p_{ij}$ is the probability of transition from the state s_i of x_k to the state s_j of x_l.

$M_{n \times n} =$

$$\begin{bmatrix} p_{ij}|x_1 \in s_i, x_1 \in s_j & p_{ij}|x_1 \in s_i, x_2 \in s_j & \cdots & p_{ij}|x_1 \in s_i, x_n \in s_j \\ p_{ij}|x_2 \in s_i, x_1 \in s_j & p_{ij}|x_2 \in s_i, x_2 \in s_j & \cdots & p_{ij}|x_2 \in s_i, x_n \in s_j \\ \vdots & \vdots & \ddots & \vdots \\ p_{ij}|x_n \in s_i, x_1 \in s_j & p_{ij}|x_n \in s_i, x_2 \in s_j & \cdots & p_{ij}|x_n \in s_i, x_n \in s_j \end{bmatrix} \quad (2)$$

2.2. Proposed Time-frequency MTF

The process of the proposed TFMTF algorithm is as follows.

Step 1: Execute time-domain signal process and obtain Xt.

Step 1.1: Time series X is first normalized by the min-max standardization algorithm (shown in Eq.(3)), and defined as X_t.

$$X_t = \frac{X - \min(X)}{\max(X) - \min(X)} \qquad (3)$$

Step 1.2: Set the q-state transition interval of the time domain signal from minimum to maximum along the amplitude.

Step 1.3 Construct an $q \times q$ state transition matrix $Pt_{q \times q}$ for X_t, as shown in Eq. (4).

$$Pt_{q \times q} = \begin{bmatrix} p_{t_{11}} & p_{t_{12}} & \cdots & p_{t_{1q}} \\ p_{t_{21}} & p_{t_{22}} & \cdots & p_{t_{2q}} \\ \vdots & \vdots & \ddots & \vdots \\ p_{t_{q1}} & p_{t_{q2}} & \cdots & p_{t_{qq}} \end{bmatrix} \qquad (4)$$

Step 1.4: Count c_{ij} ($1 \leq i,j \leq q$) (the number of X_t transiting from state s_i to state s_j) and then calculate $p_{t\,ij} = \frac{c_{ij}}{\sum_i c_{ij}}$ to obtain the time domain MTF (T-MTF) matrix $M_t = (p_{ij})_{n \times n}$ that has extra temporal information in addition to state-transition possibilities compared with the state transition matrix, as shown in Eq.(5).

$M_{t_{n \times n}} =$

$$\begin{bmatrix} p_{t_{ij}}|x_1 \in s_i, x_1 \in s_j & p_{t_{ij}}|x_1 \in s_i, x_2 \in s_j & \cdots & p_{t_{ij}}|x_1 \in s_i, x_n \in s_j \\ p_{t_{ij}}|x_2 \in s_i, x_1 \in s_j & p_{t_{ij}}|x_2 \in s_i, x_2 \in s_j & \cdots & p_{t_{ij}}|x_2 \in s_i, x_n \in s_j \\ \vdots & \vdots & \ddots & \vdots \\ p_{t_{ij}}|x_n \in s_i, x_1 \in s_j & p_{t_{ij}}|x_n \in s_i, x_2 \in s_j & \cdots & p_{t_{ij}}|x_n \in s_i, x_n \in s_j \end{bmatrix}$$

(5)

Step 2: Execute frequency-domain signal processing and obtain X_f.

Step 2.1 Calculate the frequency spectrum signal $X_f(k)$ of $X = \{x_i\}_1^N$, as shown in Eq. (6):

$$X_f(k) = \frac{abs \sum_{i=0}^{N-1} x(i) e^{-j\frac{2\pi}{N}ki}}{\frac{N}{2}} \qquad (6)$$

where k refers to the frequency axis.

Step 2.2 Compute $X_f(k)$, which is scaled to between 0 and 1 using min-max standardization, as shown in Eq. (7):

$$X_f(k) = \frac{X_f(k) - \min(X_f(k))}{\max(X_f(k)) - \min(X_f(k))} \qquad (7)$$

Step 2.3: Set the q-state transition interval of the frequency

domain signal from minimum to maximum along the frequency spectrum amplitude

Step 2.4: Construct an $q \times q$ state transition matrix $P_{f q \times q}$ for $X_f(k)$, as shown in Eq. (8).

$$P_{f_{q \times q}} = \begin{bmatrix} p_{f_{11}} & p_{f_{12}} & \cdots & p_{f_{1q}} \\ p_{f_{21}} & p_{f_{22}} & \cdots & p_{f_{2q}} \\ \vdots & \vdots & \ddots & \vdots \\ p_{f_{q1}} & p_{f_{q2}} & \cdots & p_{f_{qq}} \end{bmatrix} \quad (8)$$

Step 2.5: Obtain the frequency domain MTF (F-MTF) matrix $M_f = \left(p_{f\ ij}\right)_{n \times n}$, as shown in Eq. (9).

$$M_{f_{n \times n}} =$$

$$\begin{bmatrix} p_{f_{ij}}\big|x_1 \in s_i, x_1 \in s_j & p_{f_{ij}}\big|x_1 \in s_i, x_2 \in s_j & \cdots & p_{f_{ij}}\big|x_1 \in s_i, x_n \in s_j \\ p_{f_{ij}}\big|x_2 \in s_i, x_1 \in s_j & p_{f_{ij}}\big|x_2 \in s_i, x_2 \in s_j & \cdots & p_{f_{ij}}\big|x_2 \in s_i, x_n \in s_j \\ \vdots & \vdots & \ddots & \vdots \\ p_{f_{ij}}\big|x_n \in s_i, x_1 \in s_j & p_{f_{ij}}\big|x_n \in s_i, x_2 \in s_j & \cdots & p_{f_{ij}}\big|x_n \in s_i, x_n \in s_j \end{bmatrix} \quad (9)$$

Step 3: Construction of TFMTF

Owing to the symmetry of the MTF matrix, we intercept $M_{t\ n \times n}$ and $M_{f\ n \times n}$ into lower and upper triangular matrices, respectively, and then add the corresponding elements of the two matrices to construct the TFMTF matrix, as shown in Eq. (10). Therefore, the original vibration signals can be encoded into a time-frequency matrix, and then the elements in the TFMTF are mapped to pixel color according to a single-value function, as shown in Figure 2.

$$TFMTF_{n \times n} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ Mt_{21} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ Mt_{n1} & Mt_{n2} & \cdots & 0 \end{bmatrix} + \begin{bmatrix} Mf_{11} & Mf_{12} & \cdots & Mf_{1n} \\ 0 & Mf_{22} & \cdots & Mf_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Mf_{nn} \end{bmatrix} (10)$$

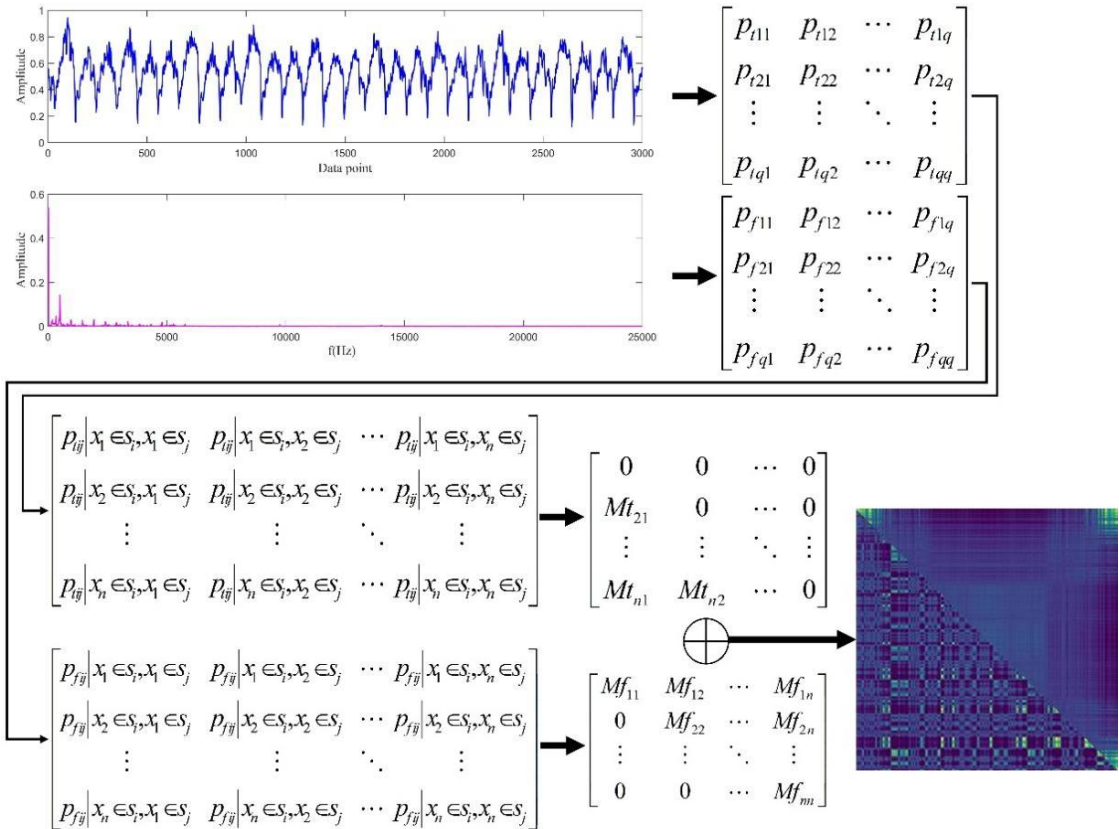Fig. 2 Image representation process of TFMTF.

2.3. Residual Neural Network

The hidden features contained in the original signal represented by a 2D image can be learned automatically based on a deep model. In this paper, we use a deep model transfer learning method (TL), which means that a general classification model with good generalization ability and compatibility is pre-trained on a large data set, and then the model can be applied to the corresponding downstream tasks according to the needs of real scenarios, and only fine-tuning of the model parameters is needed to implement a new classification model. Figure 3

illustrates a simple basic process of model transfer learning method. Using the TL method, it is possible to freeze the general feature convolution layers in the network and retrain only a few layers at the end of the network for specific feature learning or to train all layers directly without freezing, which has demonstrated to drop greatly training time and computational costs 27.
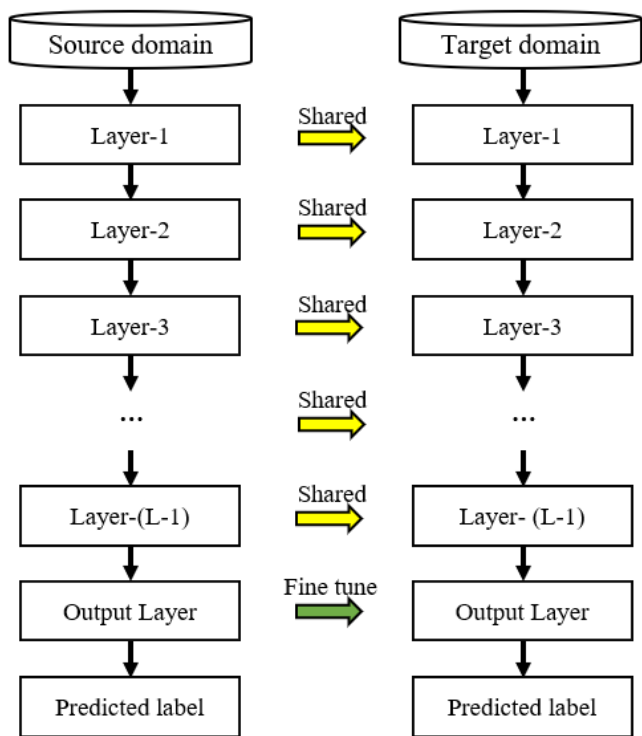

Fig. 3 Model transfer learning framework.

The residual neural network as a special kind of convolutional neural network was proposed by Kaiming He's team in 2015, and the method has achieved excellent results in major image recognition and image segmentation competitions. In recent years, the method has created a wide boom in academia and industry, and is widely used in many fields such as text classification, speech recognition, and computer vision28. Generally speaking, the deeper the neural network, the better the network performance will be accordingly, but the reality is that deeper networks have a higher probability of overfitting, also known as degradation of the network. To address this problem, the network proposes a "residual learning" mechanism that can increase the depth of the network without causing degradation in performance. The residual learning module is shown in Figure 4, from which it can be seen that the module consists of two pathways, one is a constant mapping of

the original data $x \rightarrow x$, and the other is a nonlinear mapping $F(x)$ after convolutional pooling.
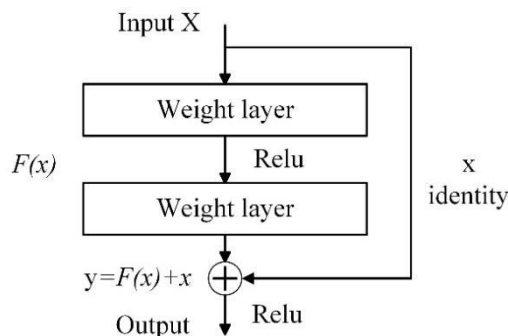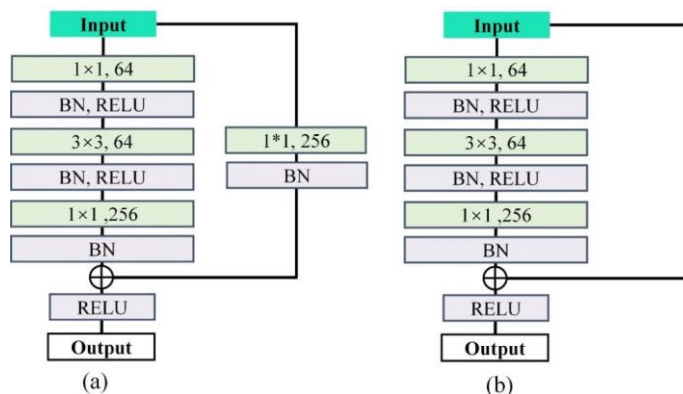

Fig. 4 Residual structure.

Fig. 5 Bottleneck structure.

The shallow residual network is mainly composed of multiple residual modules stacked as described above, but for some deeper residual networks (e.g., Resnet50, Resnet101) a Bottleneck (also known as bottleneck layer) residual module is used, as shown in Figure 5. From the figure, we can see that the nonlinear part of Bottleneck structure has a total of three convolutional layers, including two convolutional layers with 1×1 convolutional kernel size, and the middle layer with 3×3 convolutional kernel size. The convolution of the first layer will downscale the input dimension, and after that, the convolution of 3×3 and the convolution of 1×1 will be upscaled. The advantage of this is that the number of computational parameters can be reduced, making it possible to build a deeper network. Resnet network largely solves the problem of network degradation and gradient disappearance, so this paper uses Resnet to study tool condition monitoring under the same working conditions.

2.3 Proposed TCM Method

The TFMTF-based feature representation can obtain rich features of tool wear in the time-frequency domain, while the

TL method enables the transfer of generic knowledge and provides precision tool condition monitoring under small labeled data. The proposed TFMTF-TL method framework is shown in Figure 6, including data acquisition, data encoding, feature extraction, and tool condition classification.

The proposed method in this paper is to achieve the tool condition monitoring under the same working condition assuming that a small number of target domain label samples can be obtained, the method consists of four main stages.

(1) Data acquisition stage: In order to verify the accuracy of tool condition monitoring under the same working conditions we use PHM 2010 TCM public data set for analysis and validation, using a dynamometer to collect real-time sensing data during the cutting process and reserve it to the local computer database for the next data pre-processing preparation.

(2) Data pre-processing stage: Cutting force data can directly reflect the state of tool wear, and due to its high accuracy and anti-interference capability, it was widely used in the field of fault diagnosis and tool condition monitoring. Based on Markov transition field (MTF), an improved method is proposed: time-frequency Markov transition field (TFMTF), which encodes the sensor signal in the milling process into a state matrix containing both time-domain and frequency-domain information, and determines the color of the image by the size of the matrix elements, realizing the pictorial representation of time series and greatly enriching the information of imaging.

(3) Model training stage: This phase uses the model parameter transfer method by using the Resnet50 network as the feature extractor, importing the weight parameters of the network pre-trained on ImageNet into the residual network as initialization, and modifying the final number of classification layers of the model to be consistent with the number of tool wear states. The advantage of using this method is that only a small amount of labeled source domain data is needed for supervised training to achieve accurate tool condition monitoring of a new tool under the same working conditions, without the need to acquire unlabeled data from the target domain online to assist in learning, greatly improving the efficiency.
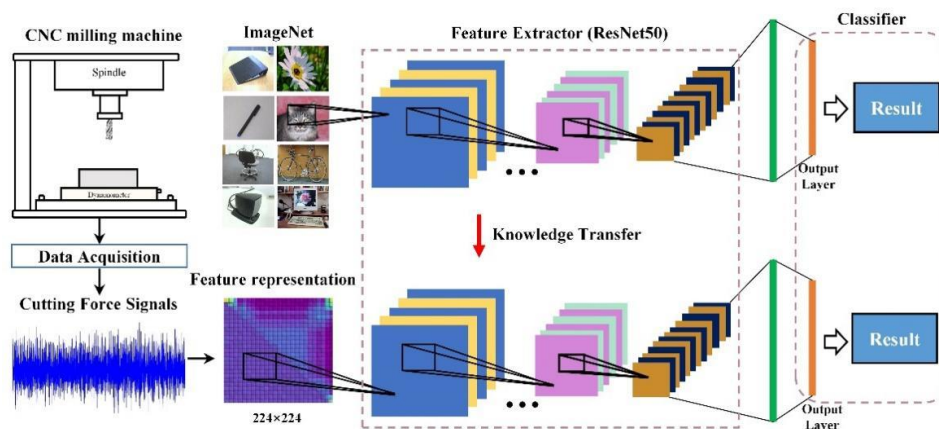

Fig. 6 Framework of the proposed TFMTF-TL method.

(4) The final test phase uses the trained classifier to classify the new data. The experimental results show that the method can quickly and effectively learn the tool wear features contained in TFMTF and achieve accurate classification results. The parameters of the deep model are listed in detail in Table 1.

Table 1 Parameters of proposed TCM method

Parameter	Learning rate	Size of image	Dropout	Batch size	Optimizer	Loss function
Value	1e⁻⁴	224× 224	0.5	16	Adam	Cross-Entropy Loss

3. Investigation with PHM 2010 TCM Experiment

3.1. Data Description

The PHM 2010 TCM experimental dataset29 was used to demonstrate the effectiveness of the proposed TCM method. Figure 7 shows the experimental platform. The dataset contains a total of six tools, of which C1, C4, and C6 are labeled and the other three tools are without label, because the main content of this paper is a tool condition monitoring study with a small number of labeled samples available in advance, so we selected the data of C1, C4, and C6 for the study. Each tool contains 315

CSV files, the data sampling frequency is 50,000 HZ, and the rotational speed is 10400 RPM. The experiment uses the workpiece size of 108 mm in length and width, and the complete distance of 108 mm after the tool is cut into the workpiece is recorded as a tool walk, after which the degree of wear will be measured using a LEICA MZI2 tool microscope and the current wear value will be recorded. The specific parameters are shown in Table 2.

Table 2 PHM 2010 TCM Cutting parameters

| Hardware	Model or parameter	Cutting condition	Parameters
CNC machine tools	Roders Tech RFM760	Sampling frequency	50k Hz
Dynamometer	Kistler 9265B	Spidle speed	10400 rpm
Charge amplifier	Kistler5019A	Feed rate	1555 mm/min
Workpiece material	Inconel 718	Y depth of cut (radial)	0.125 mm
Tool	Three flute ball end carbide milling cutter	Z depth of cut (axial)	0.2 mm
Data Acquisition Cards	NI DAQ	Feed per tooth	0.001mm
Wear measurement instrument	LEICA MZI2	Cooling condition	Dry cutting
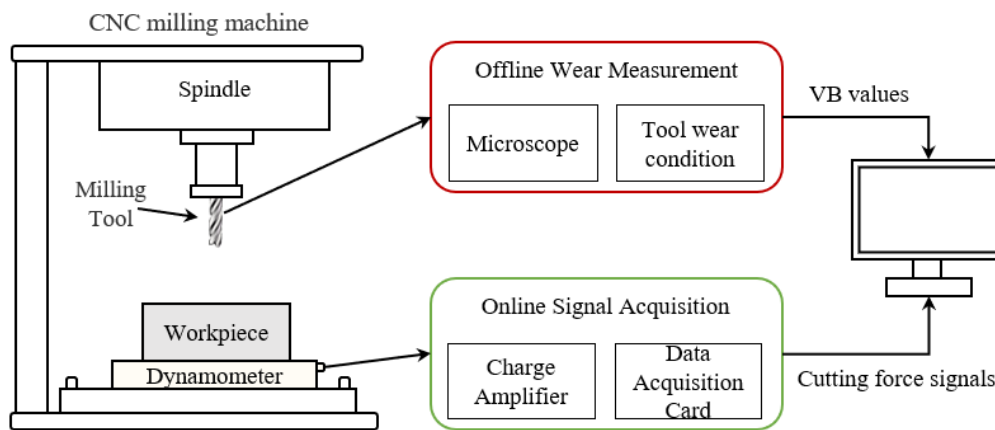
Fig. 7 Experimental platform.

Tool wear can be classified into three classes depending on the degree of tool wear: slight wear, stable wear and sharp wear. Figure 8 shows the wear degradation process of C1.
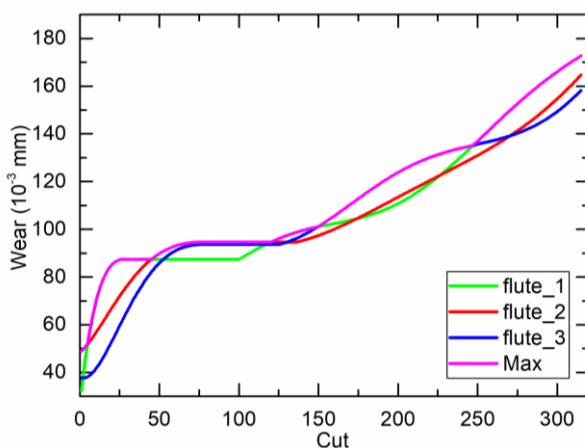

Fig. 8 Tool wear degradation process of C1.

From the figure, it can be seen that at the beginning of tool wear, the wear values change more drastically and with a larger slope, after which the change increases steadily for a long period of time and the slope of the curve becomes significantly smaller until the final tool wear reaches the critical value for tool change, when the slope starts to become larger again.

3.2. Results and Analysis
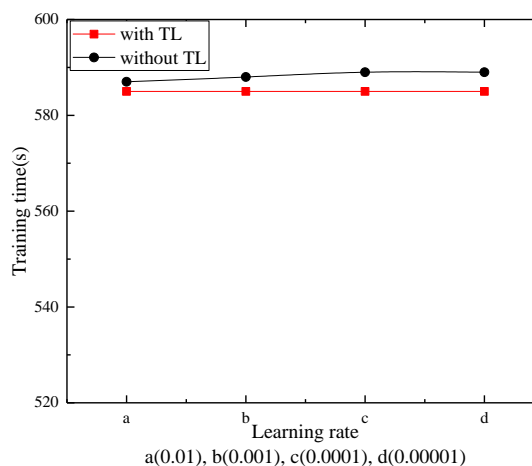
(1) Compared with not using TL strategy

To verify the effect of using the TL strategy, we use C1 of Y direction as the training set and C4&C6 of Y direction as the test set, in which the optimizer is Adam. The results are shown in Table 3. It can be found that, with the TL strategy, the model does not require a substantial iterative update of all weights of the network and only a small amount of fitting update is required on the basis of the original model parameters to obtain the optimal solution of the model, the average classification accuracy is improved by 7.8% and the average training time is shortened by 3.3 seconds compared with the method without TL strategy. Figure 9 shows the comparison of training accuracy and training time consumption before and after using the TL

strategy. It can be seen that the TL method can greatly improve the speed of model training as well as ensure the accuracy of

model training under small samples.
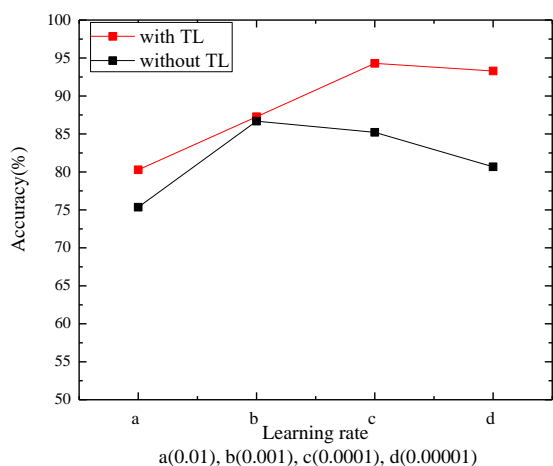
(a) accuracy comparison

(b) time consuming comparison

Fig. 9 Results of using TL strategy compared to without TL strategy.

Table 3 Results with different learning rates and number of iterations.

Method	Learning rate	Iteration number	Classification accuracy (%)	Training time (s)
With TL	0.01	10	79.13	290
	0.01	20	80.28	585
	0.01	30	87.28	871
	0.001	10	88.01	292
	0.001	20	87.26	585
	0.001	30	89.00	873
	0.0001	10	88.89	292
	0.0001	20	94.30	585
	0.0001	30	93.88	873
	0.00001	10	89.39	292
	0.00001	20	93.30	585
	0.00001	30	92.66	872
Without TL	0.01	20	75.35	587
	0.001	20	82.68	588
	0.0001	20	85.22	589
	0.00001	20	80.67	589

(2) The effect of different learning rates and iterations on the results

In order to determine the appropriate learning rate size and the effect of the number of iteration n_s on the final results, we use C1 in the Y-direction as the training set and C4&C6 in the Y-direction as the test set, with each tool containing 1200 samples, and the results were analyzed as shown in Table 3. From Table 3, it can be seen that when the learning rate is 0.01, the learning rate is large leading to the test results cannot reach a satisfactory value and no essential information is learned.

When the learning rate is 0.0001, the accuracy of the test set reaches the peak, and when the learning rate continues to become smaller, the classification accuracy transformation decreases slightly, so we set the learning rate is 0.0001. When the number of iterations is 10, it can be seen that the results are obviously low, and when the training iterations continues to 30, the results do not change significantly compared to 20 iterations, but slightly decrease. Therefore, we use training 20 iterations as the final solution.

(3) Influence of different optimizers on the results

To confirm the impact of different optimizers on the final results, we still use C1 as the training set C4 as the test set with a total sample size of 1200, including 400 samples per category, a learning rate set to 0.0001, and a number of iterations of 20. Table 4 analyzes the effect of using different optimizers on the final classification results, and it can be seen from the table that the classification accuracy using Adam optimizer can reach 94.3%, which has a significant advantage compared with RMSProp optimizer and SGD optimizer. The Adam optimizer integrates the first-order moment estimation and second-order moment estimation of the gradient to calculate the update step, and this optimization algorithm combines the advantages of both AdaGrad and RMSProp optimization algorithms.

Table 4 Classification results under different optimizers

optimizer	Adam	RMSprop	SGD
Classification accuracy/%	94.30	92.25	84.28
Training time/s	585	585	573

(4) The effect of different feature representation methods on the results

The parameters of the residual network used for the tool condition monitoring method in this paper are set as follows. The residual network after parameter transfer is optimized using the Adam optimization algorithm, the learning rate of the feature extraction part is set to 0.0001, and the learning rate of the final FC layer is set to 0.0001×10. This setting is to stabilize the initial weights, reduce the fluctuation of the loading weights, and fine-tune the classification layer. The hardware environment for the experiments is under Windows, using the Pytorch deep learning framework based on the Python language, which can improve the training speed of the model with the help of GPU for accelerated computation, and also supports dynamic neural networks, which is a simple and efficient deep learning framework. The computer processor is an AMD R5 4600H, accelerated by an NVIDIA GeForce GTX1650 GPU, and the training process requires less than 4G of GPU usage.

Due to the increasing complexity of mechanical devices, non-stationary signals are often generated. For complex nonstationary sensor data, traditional signal analysis methods can only provide statistical averaging analysis in the time or frequency domain, which is difficult to reveal the comprehensive features in the time and frequency domains30. Therefore, this method is not suitable for analyzing non-stationary signals in engineering applications. Time-frequency analysis can identify the frequency components of signals and reveal their time-varying characteristics, which is an effective method for extracting mechanical health information contained in non-stationary signals. To verify the effectiveness of the time-frequency Markov transform field (TFMTT) algorithm proposed in this paper, we compare it with the conventional Continuous Wavelet transform (CWT)31, Short-Time Fourier Transform (STFT)32 compared with MTF and frequency domain Markov transfer field (SMTF)33 as well. Considering that the data may not be abundant in the actual processing situation, we only use the data of one tool as the training set, and the data of the other two tools are used in the testing phase to test as many other data as possible with the least amount of data, and Table 5 show the TCM classification results using different feature representations.

Figure 10 shows the imaging results with five imaging algorithms: MTF, SMTF, STFT, CWT, and TFMTF. It can be found that the TFMTF-based algorithm has the richest details among all the considered algorithms that abundantly representing the information in the time and frequency domains. In addition, the color and internal shape features of the image are also richer, which can realize intelligent feature extraction and provide more effective features for deep network training.
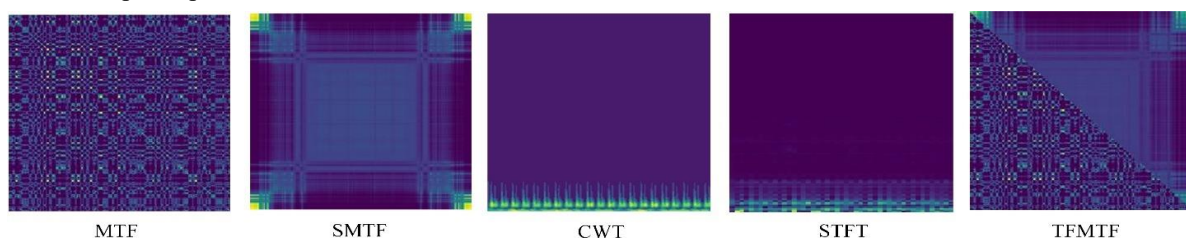

Fig. 10 Image representations.

All the results are listed in Table 5, and the average classification accuracies of different methods on the three directional datasets are given in Figure 11. It can be found that the proposed method shows better classification accuracy than

the existing methods on the Y-direction dataset with 94.30%, 95.23% and 91.98%, respectively, and the average classification accuracy reaches 93.83%, which is optimal compared with other methods. From the traditional time-frequency representation methods specifically, CWT and STFT have the highest average classification accuracy of only 87.51% in the X directions, and the results vary greatly when using different tools as the training set and are not stable. From the MTF series feature representation method, the highest performance of the

traditional MTF method can reach 85.57% classification accuracy only in the X direction, and SMTF, as a frequency domain coding method, achieves more than 90% test results in all three directions with better stability, but is still 2.52% lower than the proposed TFMTF feature representation method. Therefore, these comparative results fully reflect the superiority of the proposed method in terms of classification accuracy and generalization ability under tool state monitoring.

Table 5 Classification accuracy under different feature representation methods.

| Dataset	Training set	Testing set	MTF	SMTF	CWT	STFT	TFMTF
X	C1	C4&C6	87.75	91.35	83.02	93.33	57.77
	C4	C1&C6	87.93	91.50	77.14	89.12	69.71
	C6	C1&C4	81.03	90.59	79.36	80.08	40.60
	Average accuracy		85.57	91.95	79.84	87.51	56.03
Y	C1	C4&C6	87.15	92.39	79.53	77.61	94.30
	C4	C1&C6	87.96	94.18	87.09	86.28	95.23
	C6	C1&C4	60.93	87.38	84.53	85.81	91.98
	Average accuracy		78.68	91.31	83.72	83.23	93.83
Z	C1	C4&C6	60.36	90.13	67.42	69.92	77.45
	C4	C1&C6	66.47	91.67	72.06	66.15	63.92
	C6	C1&C4	44.92	94.48	62.34	82.73	42.32
	Average accuracy		57.25	92.09	67.27	72.93	61.23
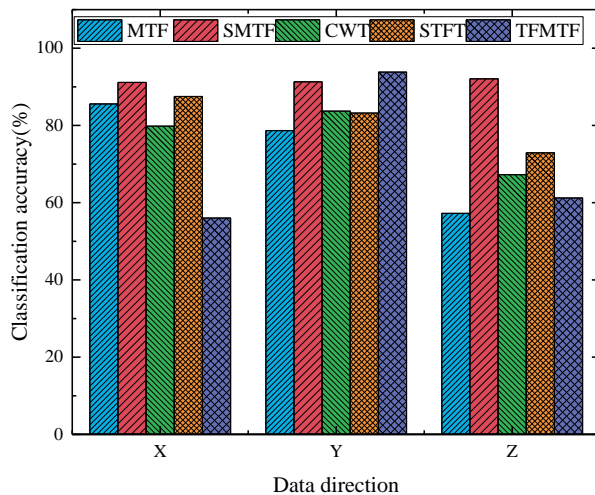
Fig. 11 Average classification results using different feature representation methods.

(5) Analysis of the results of the MTF series three-channel data fusion representation method

In addition, this paper investigates the effect of the three

methods MTF, SMTF and TFMTF after merging the data of X, Y and Z channels, as shown in Table 6. The data fusion in the table is done by MTF series methods after generating the corresponding feature representation matrix for each channel data, and then combining the three matrices into a three-dimensional 3×224×224 matrix, which is finally saved as RGB three-channel picture by computer. The data fusion process is shown in Figure 12, in which 3C stands for the abbreviation of three channels (Three Channel). From the data in Table 6, it can be seen that after the fusion of three channels, the classification accuracy of the three methods is significantly improved compared with the previous single channel. The experimental results show that the proposed TFMTF method still has obvious advantages after data fusion, which indicates that this way of combining time-frequency domain information can be applied to the actual tool condition monitoring field.
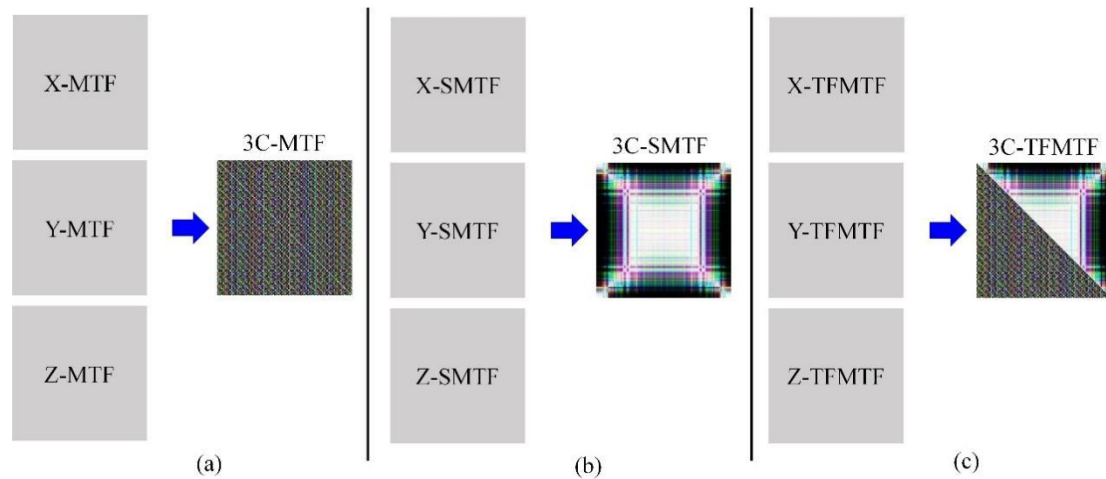
Fig. 12 MTF series three-channel data fusion representation process.

Table 6 Classification accuracy of three-channel data fusion representation.

Training set	Testing set	Classification accuracy (%)		
		3C-MTF	3C-SMTF	3C-TFMTF
C1	C4 & C6	82.86	93.46	93.97
C4	C1 & C6	88.85	95.04	95.63
C6	C1 & C4	85.32	92.41	92.55
Average accuracy		85.67	93.63	94.05

4. CONCLUSION

In this paper, a new time-frequency representation method called TFMTF is proposed, which provides an image representation of the cutting force signal and enhances the identifiability of the sample to improve the accuracy of tool condition monitoring. A learning strategy based on the transfer of model parameters is introduced to reduce the computational effort of the model, improve the training speed, and achieve tool condition monitoring under small sample conditions. We evaluated the performance of the proposed method in this paper using the PHM 2010 TCM dataset. The comparison of the results of different methods for X, Y, and Z single channel data shows that TFMTF obtains the highest classification accuracy on the Y-direction dataset, where the classification result can reach 95.23% when C4 is used as the training set C1& C6 as the test set, and the average classification accuracy of the three transfer tasks in Y-direction is 93.83%, which still achieves the highest classification result compared with several other methods. It can be proposed that the SMTF method achieves comparable results with the proposed method in this paper with test results above 90% on all three directional datasets. In order to further verify the effectiveness of the proposed method in this paper, the dataset under the three-channel data fusion was constructed for comparing the classification accuracy of the three methods MTF, SMTF and TFMTF. The results show that after the three-channel data fusion representation, the average classification accuracy of the proposed TFMTF method improved by 0.22% compared with the previous single-channel highest classification accuracy, which was higher than that of MTF, SMTF by 8.38% and 0.42%, respectively.

References

1. Yan S, Shao HD, Xiao YM, Liu B, Wan JF. Hybrid robust convolutional autoencoder for unsupervised anomaly detection of machine tools under noises. Robotics and Computer-Integrated Manufacturing 2023; 79: 102441, https://doi.org/10.1016/j.rcim.2022.102441.

2. Wang HC, Sun W, Sun WF, Ren Y. A novel tool condition monitoring based on Gramian angular field and comparative learning. International Journal of Hydromechatronics 2023; 6(2): 93-107, https://doi.org/10.1504/IJHM.2022.10048957.

3. Sau A, Mu B, Sa C, et al. Tool wear, surface roughness, cutting temperature and chips morphology evaluation of Al/TiN coated carbide cutting tools in milling of Cu-B-CrC based ceramic matrix composites. Journal of Materials Research and Technology, 2022; 16:1243-

1259, https://doi.org/10.1016/j.jmrt.2021.12.063.

4. Jae-Woong, Youn, Min-Yang, et al. A Study on the Relationships Between Static/Dynamic Cutting Force Components and Tool Wear. Journal of Manufacturing Science & Engineering, 2001; 123(2): 196-205, https://doi.org/10.1115/1.1362321.

5. Pimenov D Y, Bustillo A, Wojciechowski S, et al. Artificial intelligence systems for tool condition monitoring in machining: analysis and critical review. Journal of Intelligent Manufacturing 2023; 34: 2079-2121, https:// doi.org/10.1007/s10845-022-01923-2.

6. Selvaraj V, Xu Z, Min S. Intelligent Operation Monitoring of an Ultra-Precision CNC Machine Tool Using Energy Data. International Journal of Precision Engineering and Manufacturing- Green Technology 2022; 10: 59-69, https://doi.org/10.1007/s40684-022-00449-5.

7. Zhen D, Li DK, Feng GJ, et al. Rolling bearing fault diagnosis based on VMD reconstruction and DCS demodulation. International Journal of Hydromechatronics 2022; 5(3): 205-225, https://doi.org/10.1504/IJHM.2022.10048012.

8. Xiao YM, Shao HD, Han SY, et al. Novel joint transfer network for unsupervised bearing fault diagnosis from simulation domain to experimental domain. IEEE-ASME Transactions on Mechatronics 2022; 27(6): 5254-5263, https://doi.org/10.1109/TMECH.2022.3177174.

9. Zhou YQ, Kumar A, Parkash C, et al. A novel entropy-based sparsity measure for prognosis of bearing defects and development of a sparsogram to select sensitive filtering band of an axial piston pump. Measurement 2022; 203: 111997, https://doi.org/10.1016/j.measurement.2022.111997.

10. Yu J, Shuang L, Tang D, et al. A weighted hidden Markov model approach for continuous-state tool wear monitoring and tool life prediction. International Journal of Advanced Manufacturing Technology 2016; 91(1-4): 1-11, https://doi.org/10.1007/s00170-016-9711-0.

11. Benkedjouh T, Medjaher K, Zerhouni N, et al. Health Assessment and Life Prediction of cutting tools based on support vector regression. Journal of Intelligent Manufacturing 2015; 26(2): 213-223, https://doi.org/10.1007/s10845-013-0774-6.

12. Ma M, Mao Z. Deep-Convolution-Based LSTM Network for Remaining Useful Life Prediction. IEEE Transactions on Industrial Informatics 2021; 17(3): 1658-1667, https://doi.org/10.1109/TII.2020.2991796.

13. Zhou YQ, Kumar A, Parkash C, et al. Development of entropy measure for selecting highly sensitive WPT band to identify defective components of an axial piston pump. Applied Acoustics; 2023; 203: 109225, https://doi. org/10.1016/j.apacoust.2023.109225.

14. Vashishtha G, Kumar R. Autocorrelation energy and aquila optimizer for MED filtering of sound signal to detect bearing defect in Francis turbine. Measurement Science and Technology, 2022, 33(1), 015006, https://doi.org/10.1088/1361-6501/ac2cf2.

15. Zhang J, Kong X, Cheng L, et al, Intelligent fault diagnosis of rolling bearings based on continuous wavelet transform multiscale feature fusion and improved channel attention mechanism. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2023: 25(1), 16, https://doi.org/10.17531/ein.2023.1.16.

16. Cheng YW, Lin MX, Wu J, et al. Intelligent fault diagnosis of rotating machinery based on continuous wavelet transform-local binary convolutional neural network. Knowledge-Based Systems 2021; 216: 106796, https://doi.org/10.1016/j.knosys.2021.106796.

17. Navarro-Devia JH, Chen Y, Dao DV, et al. Chatter detection in milling processes- a review on signal processing and condition classification. International Journal of Advanced Manufacturing Technology 2023, https://doi.org/10.1007/s00170-023-10969-2.

18. Yi L, Jiang YJ, Zhang QC, et al.Remaining useful life prediction with insufficient degradation datbased on deep learning approach. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2021; 23(4): 745-756, https://doi.org/10.17531/ein.2021.4.17.

19. Liu CH, Jiao J, Li WL, et al. Tr-Predictior: An Ensemble Transfer Learning Model for Small-Sample Cloud Workload Prediction. Entropy 2022; 24(12), 1770, https://doi.org/10.3390/e24121770.

20. Nasir V, Sassani F. A review on deep learning in machining and tool monitoring: methods, opportunities, and challenges. International Journal of Advanced Manufacturing Technology 2021; 115: 2683-2709, https://doi.org/10.1007/s00170-021-07325-7.

21. Zhou Y, Zhi G, Chen W, et al. A new tool wear condition monitoring method based on deep learning under small samples. Measurement, 2022; 189, 110622, https://doi.org/10.1016/j.measurement.2021.110622.

22. Zhuang FZ, Qi ZY, Duan KY, et al. A Comprehensive Survey on Transfer Learning. Proceedings of the IEEE 2021; 109(1): 43-76, https://doi.org/10.1109/JPROC.2020.3004555.

23. Marei M, Zaatari SE, Li WD. Transfer learning enabled convolutional neural networks for estimating health state of cutting tools. Robotics and Computer-Integrated Manufacturing 2021; 71: 102145, https://doi.org/10.1016/j.rcim.2021.102145.

24. Yang B, Lei Y, Jia F, et al. An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive

bearings. Mechanical Systems and Signal Processing 2018; 122: 692-706, https://doi.org/10.1016/j.ymssp.2018.12.051.

25. Wang Z, Oates T. Imaging Time-Series to Improve Classification and Imputation. Proceedings of the 24th International Conference on Artificial Intelligence 2015: 3939-3945, https://doi.org/10.1196/annals.1333.032.

26. Lyu MZ, Wang JM, Chen JB. Closed-Form Solutions for the Probability Distribution of Time-Variant Maximal Value Processes for Some Classes of Markov Processes. Communications in Nonlinear Science and Numerical Simulation 2021; 99(8): 105803, https://doi.org/10.1016/j.cnsns.2021.105803.

27. Udmale SS, Singh SK, Singh R, et al. Multi-fault bearing classification using sensors and ConvNet-based transfer learning approach. IEEE Sensors Journal 2020; 20(3): 1433-1444, https://doi.org/10.1109/JSEN.2019.2947026.

28. He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016, https://doi.org/10.1109/CVPR.2016.90

29. The Prognostics and Health Management Society, 2010 Conference Data Challenge. https://www.phmsociety.org/competition/phm/10.

30. Feng ZP, Liang M, Chu FL. Recent advances in time-frequency analysis methods for machinery fault diagnosis: A review with application examples. Mechanical Systems and Signal Processing 2013; 38(1): 165-205, https://doi.org/10.1016/j.ymssp.2013.01.017.

31. Hess-Nielsen, N, Wickerhauser, et al. Wavelets and time-frequency analysis. Proceedings of the IEEE, 1996, https://doi.org/10.1109/5.488698.

32. Nawab S H, Quatieri T F . Short-time Fourier transform[C]// Advanced Topics in Signal Processing. Prentice-Hall, Inc. 1988.

33. Bai YL, Yang JW, Wang JH, Zhao Y, Li Q. Image representation of vibration signals and its application in intelligent compound fault diagnosis in railway vehicle wheelset-axlebox assemblies. Mechanical Systems and Signal Processing 2021; 152, 107421, https://10.1016/j.ymssp.2020.107421