



OPEN Multi-modal denoised data-driven milling chatter detection using an optimized hybrid neural network architecture

Haining Gao^{1,3}✉, Haoyu Wang², Hongdan Shen³, Shule Xing³, Yong Yang³, Yinlin Wang³, Wenfu Liu³, Lei Yu^{3,4}, Mazhar Ali⁴ & Imran Ali Khan⁴

Chatter, a type of self-excited vibration, deteriorates surface quality and reduces tool life and machining efficiency. Chatter detection serves as an effective approach to achieve stable cutting. To address the low accuracy in chatter detection caused by the limitations of both one-dimensional temporal and two-dimensional image modal information, this study proposes a multi-modal denoised data-driven milling chatter detection method using an optimized hybrid neural network architecture. A data denoising model combining Complementary Ensemble Empirical Mode Decomposition (CEEMD) and Singular Value Decomposition (SVD) is established. The Ivy algorithm is employed to optimize the hyperparameters of CEEMD-SVD. Multi-modal data features of different machining states are then obtained using time–frequency domain methods and Markov transition field methods. Sensitivity analysis of time–frequency domain features is conducted using Pearson correlation coefficient analysis. A hybrid neural network model (DBMA) for chatter detection is constructed by integrating dual-scale parallel convolutional neural networks, bidirectional gated recurrent units, and multi-head attention mechanisms. The Ivy algorithm is utilized to optimize the hyperparameters of DBMA. The t-SNE algorithm is employed to visualize features extracted from different network layers of the chatter detection model. Results demonstrate that effective denoising of machining signals and the use of multi-modal data can significantly improve the accuracy of state detection. Compared with other methods, the proposed model exhibits superior stability and robustness.

Keywords Chatter detection, Multi-modal data, Optimized hybrid neural network, Data denoising

Machining processes play a vital role in modern manufacturing, particularly in the aerospace industry where high precision and reliability are essential¹. However, critical components in aerospace applications, such as blades, are prone to chatter during milling due to their thin-walled structure, complex geometry, and the use of difficult-to-machine materials like titanium alloys. Chatter not only degrades machining surface quality and shortens tool life but also poses serious safety risks to operators. To prevent chatter occurrence during milling, researchers have conducted studies on chatter prediction² and chatter detection^{3,4}. Chatter prediction first requires establishing a multi-degree-of-freedom dynamic model of the research object. Then, milling experiments or cutting force models are conducted to obtain parameters such as modal parameters and cutting force coefficients⁵. Next, stability lobe diagrams are generated using stability analysis methods. Based on the stability lobe diagrams, stable machining parameters are selected⁶. However, due to the complexity of machining conditions and variability of machining parameters, chatter may still occur even when stable cutting parameters are selected⁷. Chatter detection has garnered widespread attention due to its advantages in real-time performance and accuracy.

Signal acquisition is the first step in chatter detection. Common detection signals include vibration signals^{8,9}, cutting force signals^{10,11}, current signals^{12,13}, and fused signals^{3,14}. However, fused signals face challenges such as temporal alignment and synchronization difficulties of multi-source data, information redundancy leading to

¹School of Mechanical and Power Engineering, Hennan Polytechnic University, Jiaozuo 454000, China. ²Mode Institute of Technology, Henan University of Science and Technology, Luoyang 471000, China. ³Henan Province International Joint Laboratory of New Energy Digitalization Technology, Huanghuai University, Zhumadian 463000, China. ⁴Department of Computer Science, COMSATS University Islamabad, Abbottabad 22060, Pakistan. ✉email: 20191908@huanghuai.edu.cn

increased feature selection and dimensionality reduction complexity, and difficulties in optimization and adaptive adjustment of fusion weights¹⁵. Kuljanic et al.¹⁶ found that rotary dynamometers performed well in laboratory tests but had limited industrial applications. Spindle power signals showed lower reliability for chatter detection. Wang et al.¹⁵ discovered that force signals and acceleration signals were more reliable for chatter detection. Matthew et al. conducted comparative experiments between cutting force signals and acceleration signals. Their research revealed that acceleration signals, with lower computational complexity, were more suitable for chatter detection¹⁷. However, experimental data obtained through sensors has limitations in processing condition range, high costs, and time-consuming nature. To address this issue, Yin et al.¹⁸ generated simulated machining signals based on dynamic milling process models. Kuo et al.¹⁹ employed generative adversarial networks to produce abundant training data.

Feature extraction is the second step in chatter detection. Common feature extraction methods include time-domain methods^{3,20}, frequency-domain methods²¹, and time–frequency domain methods^{22–24}. Time-domain feature extraction is widely applied due to its relative simplicity and convenience. With the rapid development of deep learning methods, time–frequency domain methods have become a research hotspot. Song et al.²² proposed a second-order synchrosqueezing transform method for feature extraction. Compared with traditional methods such as EMD, EEMD, VMD, and SET1, their proposed method exhibited the lowest noise in envelope spectra. Jauhari et al.²³ improved the sharpness of continuous wavelet transform images using wavelet-based synchrosqueezing transform, enhancing the accuracy of feature extraction. The accuracy of recurrence plots in reflecting cutting system dynamics depends on hyperparameter settings. To address this issue, Chen et al.²⁴ proposed an adaptive particle swarm algorithm to calculate hyperparameters, thereby automatically generating recurrence plots. Matthew et al.¹⁷ applied time–frequency masking techniques to process short-time Fourier transforms, effectively reducing noise interference and highlighting chatter frequency components.

Pattern recognition is the third step in chatter detection. Pattern recognition methods consist of threshold methods^{8,14,25} and intelligent classification methods. The essence of threshold methods lies in selecting suitable chatter indicators and thresholds. Chatter indicators include various measures such as entropy values^{14,25} and statistical indicators⁸. However, most existing indicators are directly or indirectly correlated with machining parameters, causing indicators to fluctuate with changes in machining parameters and making early chatter detection challenging. To address these issues, researchers have conducted several studies. Lu et al.²⁶ introduced a chatter correlation factor indicator that eliminated the effects of modal mixing and machining parameter influences. Chen et al.²⁷ developed a chatter indicator based on cyclic steady-state energy ratio, which is independent of cutting conditions. Zhao et al.²⁸ proposed thresholds based on dual risk levels, effectively overcoming the limitations of single thresholds. The sparse dictionary-based chatter detection method proposed by Wang et al.²⁹ can automatically adjust chatter detection thresholds. Similarly, sequential probability ratio tests can also achieve automatic adjustment of chatter detection thresholds³⁰.

Although traditional intelligent classification algorithms such as support vector machines and random forests have achieved certain success in chatter detection, their practical applications are limited by high dependence on manual feature engineering, restricted generalization capability, and unstable performance under complex working conditions. Wan et al.²¹ improved support vector machines using adaptive boosting algorithms, enhancing chatter monitoring accuracy. Deep learning has been widely applied in chatter detection due to its advantages in automatic feature extraction, end-to-end learning capabilities, and efficient utilization of large-scale data. Convolutional neural networks (CNNs) are the most extensively applied deep learning methods. Current research primarily focuses on incorporating different attention mechanism modules into CNNs. Zhang et al.³¹ proposed a hybrid deep convolutional neural network combining SE ResNet modules and Inception modules. Lu et al.³² developed a noise-resistant convolutional neural network based on adaptive frequency band attention modules and discriminative feature attention modules. Sun et al.³³ constructed a novel deep attention information fusion network by integrating SE modules, MAConv modules, and CapsNet modules.

Multimodal data fusion has been widely applied in fault diagnosis³⁴, geological disaster prediction³⁵, and vehicle detection³⁶. Existing multimodal data fusion methods can be classified into three levels: data-level fusion, feature-level fusion, and decision-level fusion. Data-level fusion often encounters synchronization issues, while decision-level fusion may lose detailed feature information. Feature-level fusion is adopted to balance information preservation and fusion efficiency. Currently, single-modal features (one-dimensional temporal features or two-dimensional image features) are predominantly used in chatter detection processes. Meanwhile, the hyperparameters in signal denoising algorithms and convolutional neural networks are often set based on human experience. Research on chatter detection based on multi-modal data and optimized hybrid neural networks remains relatively limited, with their detection capabilities and application value requiring further exploration.

An optimized CEEMD-SVD model for data denoising is first constructed in this study. Subsequently, time–frequency domain analysis techniques and Markov transition field methods are applied to extract multi-modal data related to machining states. Then, A DBMA for chatter detection is constructed by integrating dual-scale parallel convolutional neural networks (DSPCNN), bidirectional gated recurrent units (BiGRU), and multi-head attention mechanisms (MSA). To enhance model performance, the Ivy algorithm is employed to optimize the hyperparameters of the DBMA model. Finally, the effectiveness of the proposed method is validated through multiple evaluation metrics including accuracy, precision, recall, and F1-score.

Ivy algorithm

The Ivy algorithm is proposed by Ghasemi et al.³⁷. The algorithm primarily simulates different life stages of Ivy plants, including growth, climbing, and propagation within ivy populations.

At the beginning of the algorithm, the initial positions of the Ivy population in the search space are randomly determined using Eq. (1).

$$I_i = I_{\min} + rand(1, D) \odot (I_{\max} - I_{\min}), i = 1, \dots, N_{pop} \quad (1)$$

where, I_{\max} and I_{\min} represent the upper and lower bounds of the search space, respectively. \odot denotes the Hadamard product of two vectors. $rand(1, D)$ is a D -dimensional vector of uniformly distributed random numbers in the interval $[0,1]$. N_{pop} represents the total population size of Ivy plants.

Based on data-intensive experiments and simulation processes, the difference equation for growth rate $Gv_i(t)$ of member I_i is shown in Eq. 2.

$$\Delta Gv_i(t+1) = rand^2 \odot (N(1, D) \odot \Delta Gv_i(t)) \quad (2)$$

Vectors $\Delta Gv_i(t+1)$ and $\Delta Gv_i(t)$ represent the growth rates of the discrete-time system (at time t and $t+1$). $rand^2$ is a random number of the random variable with probability density function equal to $1/(2\sqrt{x})$. $N(1, D)$ denotes a D -dimensional random vector.

The equation describing member I_i 's climbing and logical movement along the light source direction utilizing member I_i is shown in Eq. 3.

$$\begin{cases} I_i^{new} = I_i + |N(1, D)| \odot (I_i - I_i) + N(1, D) \odot \Delta Gv_i, & i = 1, 2, \dots, N_{pop} \\ \Delta Gv_i = \begin{cases} I_i \oslash (I_{\max} - I_{\min}), & Iter = 1 \\ rand^2 \odot (N(1, D) \odot \Delta Gv_i), & Iter > 1 \end{cases} \end{cases} \quad (3)$$

where, $|N(1, D)|$ is the absolute value of $N(1, D)$. \oslash represents the Hadamard division of two vectors.

Member I_i searches for better optimal solutions in the vicinity of member I_{Best} . The computational process is shown in Eq. 4.

$$\begin{cases} I_i^{new} = I_{Best} \odot (rand(1, D) + N(1, D) \Delta Gv_i) \\ \Delta Gv_i^{new} = I_i^{new} \oslash (I_{\max} - I_{\min}) \end{cases} \quad (4)$$

When the objective function value $f(I_i)$ of member I_i is less than $f(I_{Best})$ multiplied by a factor, $\beta = (2 + rand)/2$. Subsequently, the ivy plant begins to expand the width of its branches and leaves (given by Eq. 3). Otherwise, the ivy grows and climbs upward (given by Eq. 4).

Data denoising and feature extraction

Ivy-CEEMD-SVD data denoising

CEEMD overcomes the mode mixing problem of EMD by adding white noise to the original signal, performing multiple EMD decompositions, and then taking their average.

The reconstructed signal is obtained by adding and subtracting n Gaussian white noise signals $n(t)$ with zero mean, constant amplitude, and standard deviation from the original vibration signal $x(t)$.

$$\begin{cases} x_{ja}(t) = x_j(t) + n_j(t) \\ x_{jb}(t) = x_j(t) - n_j(t) \end{cases} \quad (5)$$

where, $x_{ja}(t)$ represents the composite signal after adding j Gaussian white noise signals, $x_{jb}(t)$ denotes the composite signal after subtracting j Gaussian white noise signals.

EMD decomposition is performed on $x_{ja}(t)$ and $x_{jb}(t)$ to obtain IMF_{ji}^+ and IMF_{ji}^- , which are two sets of integrated IMF components, and their average value is calculated.

SVD can effectively separate the principal components of signals from noise components, as larger singular values typically correspond to the main structure of the signal, while smaller singular values are generally associated with noise.

The IMFs obtained through CEEMD are reconstructed to obtain $P = [IMF_1 MF_2, \dots, MF_n]^T$. Its Hankel matrix is presented in Eq. 6.

$$P = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{11} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{m(m+n-1)} \end{bmatrix} \quad (6)$$

where IMF_1 can be expressed as $\{x_{11}, x_{12}, \dots, x_{1n}\}$.

P is decomposed using singular value decomposition.

$$P = \mathbf{U}\mathbf{S}\mathbf{V}^T = \sum_{i=1}^k \sigma_i \mu_i v_i \tag{7}$$

where, \mathbf{U} and \mathbf{V}^T represent the left and right singular matrices respectively. \mathbf{S} is the diagonal matrix, $\mathbf{S} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_k)$, $\{\sigma_1, \sigma_2, \dots, \sigma_k\}$ denotes the singular values of matrix P . μ_i and v_i are the i -th eigenvectors of square matrices $\mathbf{P}\mathbf{P}^T$ and $\mathbf{P}^T\mathbf{P}$ respectively.

The denoising performance of CEEMD and SVD depends on several parameters, including ensemble size, noise standard deviation, and the number of retained singular values. To reduce manual intervention and enhance both the efficiency and effectiveness of the denoising process, the IVY-CEEMD-SVD method is proposed by combining intelligent optimization algorithm with CEEMD-SVD. In the signal denoising process, the design of the fitness function is crucial for evaluating the denoising performance. The ratio of Shannon entropy between the denoised signal and the original signal is adopted as the fitness function. Signal-to-Noise Ratio (SNR), Kurtosis Ratio (KR), and Correlation Coefficient (CL) are employed as evaluation metrics.

$$fit = \frac{-\sum (p_y \cdot \log 2(p_y))}{-\sum (p_x \cdot \log 2(p_x))} \tag{8}$$

$$SNR = 10 \lg \frac{\sum_{j=1}^n (x_j)^2}{\sum_{j=1}^n (x_j - y_j)^2} \tag{9}$$

$$KR = \frac{E[(y - \bar{y})^4] / (E[y - \bar{y}]^2)^2 - 3}{E[(x - \bar{x})^4] / (E[x - \bar{x}]^2)^2 - 3} \tag{10}$$

$$CL = \frac{\text{cov}(x, y)}{\text{std}(x) \cdot \text{std}(y)} \tag{11}$$

where, x represents the original signal, y denotes the denoised signal, $E[\]$ represents the expected value, p denotes the probability distribution of signal amplitude, cov represents the covariance, std denotes the standard deviation.

Feature extraction

Comprehensive and accurate acquisition of state characteristics during the machining process is a prerequisite for precise identification of milling chatter. Currently, time-domain, frequency-domain, and time-frequency domain methods are commonly employed to obtain state characteristics during the machining process³⁸. Time-domain features, frequency-domain features, and time-frequency domain features are selected as state features, as shown in Table 1.

The Markov Transition Field is a method that transforms time-series data into spatial image data. This method extends the Markov state transition matrix by sequentially expressing the state transition matrix, fully preserving discretized temporal dynamic information, and ultimately generates two-dimensional images through fuzzy kernel aggregation.

Let $X = \{x_1, x_2, \dots, x_n\}$ denote the time series, where n represents the total number of time points. The range X is divided into Q equal quantile intervals. The Markov state transition matrix W is then constructed.

$$W = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1Q} \\ w_{21} & w_{22} & \dots & w_{2Q} \\ \vdots & \vdots & & \vdots \\ w_{Q1} & w_{Q2} & \dots & w_{QQ} \end{bmatrix} \tag{12}$$

where, w_{ij} represents the transition probability from quantile interval i to quantile interval j , $\sum_{i=1}^Q w_{ij} = 1$.

A dynamic probability transition matrix M across multiple time scales was generated based on the Markov state transition matrix.

$$M = \begin{bmatrix} w_{ij}|_{x_1 \in q_i, x_1 \in q_j} & w_{ij}|_{x_1 \in q_i, x_1 \in q_j} & \dots & w_{ij}|_{x_1 \in q_i, x_1 \in q_j} \\ w_{ij}|_{x_1 \in q_i, x_1 \in q_j} & w_{ij}|_{x_1 \in q_i, x_1 \in q_j} & \dots & w_{ij}|_{x_1 \in q_i, x_1 \in q_j} \\ \vdots & \vdots & & \vdots \\ w_{ij}|_{x_1 \in q_i, x_1 \in q_j} & w_{ij}|_{x_1 \in q_i, x_1 \in q_j} & \dots & w_{ij}|_{x_1 \in q_i, x_1 \in q_j} \end{bmatrix} \tag{13}$$

Category	Feature	Expression	Feature	Expression
Time domain	Peak to peak value (PTP)	$X_1 = \max x_i - \min x_i $	Root mean square (RMS)	$X_2 = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$
	Absolute mean (ABS)	$X_3 = \frac{1}{N} \sum_{i=1}^N x_i $	standard deviation (STD)	$X_4 = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - X_3)^2}$
	Kurtosis (KUR)	$X_5 = \frac{1}{N} \sum_{i=1}^N \left(\frac{ x_i - X_3}{X_4} \right)^4$	Skewness (SKE)	$X_6 = \frac{1}{N} \sum_{i=1}^N \left(\frac{ x_i - X_3}{X_4} \right)^3$
	Peak factor (PAF)	$X_7 = \frac{\max x_i }{X_2}$	Pulse factor (PUF)	$X_8 = \frac{\max x_i }{X_3}$
	Waveform factor (WAF)	$X_9 = \frac{X_2}{X_3}$	Clearance factor (CF)	$X_{10} = \frac{\max(x_i)}{X_2^2}$
Frequency domain	Mean of frequency spectrum (MFA)	$X_{11} = \frac{1}{N} \sum_{i=1}^N P(f_i)$	Gravity frequency of frequency spectrum (GFFS)	$X_{12} = \frac{\sum_{i=1}^N (f_i \cdot P(f_i))}{\sum_{i=1}^N P(f_i)}$
	Mean square frequency of frequency spectrum (MSFS)	$X_{13} = \frac{\sum_{i=1}^N (f_i \cdot f_i \cdot P(f_i))}{\sum_{i=1}^N P(f_i)}$	Root mean square frequency of frequency spectrum (RMSFS)	$X_{14} = \sqrt{X_{13}}$
	Frequency spectrum variance (FSV)	$X_{15} = \frac{\sum_{i=1}^N ((f_i - FC^2) \cdot P(f_i))}{\sum_{i=1}^N P(f_i)}$		
Time -frequency	Wavelet energy entropy	$X_{16} = \sum_{i=1}^N x_{(j,i)}^p ^2$		

Table 1. Time frequency domain characteristics.

Chatter detection model

DSPCNN

CNN is a biologically-inspired feedforward neural network primarily composed of convolutional layers, pooling layers, and fully connected layers. The convolutional layers primarily extract implicit feature information from data through convolution operations, then incorporate nonlinear factors using activation functions to enhance feature expressiveness. The computational process of the convolutional layer is as follows.

$$X_{i,j}^{l+1} = \sum_{j=1}^L \sum_{i=1}^m (X_{i,j}^l \times w_{i,j}^l) + b \tag{14}$$

where, $X_{i,j}^{l+1}$ represents the features of the $l + 1$ -th layer after convolution operation, $w_{i,j}^l$ denotes the convolution kernel, $X_{i,j}^l$ represents the j -th feature value of the i -th feature map, b is the bias term, and L indicates the kernel size.

A DSPCNN is proposed to effectively extract milling chatter components and enhance model identification performance. The architecture is illustrated in Fig. 1. The DSPCNN model parallelly connects convolutional layers at two different scales to enhance feature extraction capabilities. The DSPCNN model also incorporates max pooling layers and global average pooling layers. The max pooling layers scan the pooling regions of feature maps and extract maximum values as feature representations, thereby reducing the number of parameters in the model. The global average pooling layers combine and flatten the features from each feature map into feature vectors to achieve feature concatenation of the two branches.

BiGRU

BiGRU addresses the propagation direction issue by employing two GRU layers. The forward GRU processes the input sequence from start to finish, while the backward GRU processes it in the reverse direction. Each GRU layer maintains its own hidden state and updates it based on the current input and the hidden state from the previous time step.

The computational formula for the internal hidden layer units of BiGRU is shown in Eq. 15.

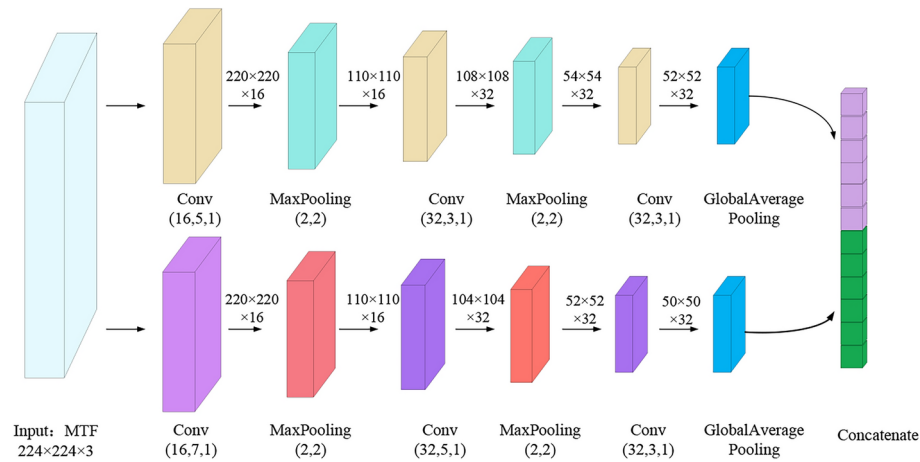


Fig. 1. Dual scale parallel convolutional neural network structure.

$$\begin{cases} \vec{h}_t = GRU(x_t, \vec{h}_{t-1}) \\ \overleftarrow{h}_t = GRU(x_t, \overleftarrow{h}_{t-1}) \\ h_t = f(W_{\vec{h}_t} \cdot \vec{h}_t + W_{\overleftarrow{h}_t} \cdot \overleftarrow{h}_t + b_t) \end{cases} \quad (15)$$

where, \vec{h}_t and \overleftarrow{h}_t represent the forward and backward states of the hidden layer at time t , $W_{\vec{h}_t}$ and $W_{\overleftarrow{h}_t}$ denote the forward and backward weights of the hidden layer at time t , and b_t is the bias of the hidden layer at time t .

BiGRU

MSA consists of multiple self-attention mechanism structures that process the same feature information simultaneously, with its output being the concatenation of multiple self-attention outputs. The computation is as follows.

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h) \cdot W^o \quad (16)$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (17)$$

where, W_i^Q , W_i^K , and W_i^V are the weights of the mapping matrices, W^o is the weight matrix of the output.

Ivy-DBMA model

An Ivy-DBMA milling chatter monitoring model is proposed in this study. The architecture is illustrated in Fig. 2. Subsequently, the processing state feature values are input into the BiGRU layer to extract temporal features. The processed two-dimensional feature map is input into the DSPCNN layer to extract Markov transition field image features. The high-dimensional image features and one-dimensional temporal features are fused to obtain temporal-spatial features. These temporal-spatial features are then input into the multi-head attention mechanism layer, which effectively captures the contribution degree of extracted features and emphasizes key features. Finally, using the validation set recognition accuracy as the objective function and combining the Ivy algorithm with DBMA for parameter optimization, the optimal chatter classification model is obtained.

To objectively evaluate the prediction results of different models, this paper employs the following evaluation metrics: Accuracy, Precision, Recall, F1 Score, and Specificity. Accuracy is the ratio of correctly classified samples to the total number of samples. Precision is the proportion of true positive samples among all samples classified as positive. Recall is the proportion of correctly identified positive samples among all actual positive samples. Specificity measures the proportion of correctly classified negative samples among all actual negative samples. The F1 Score is the harmonic mean of precision and recall, comprehensively considering the performance of both metrics. The closer all five metrics are to 1, the better the performance.

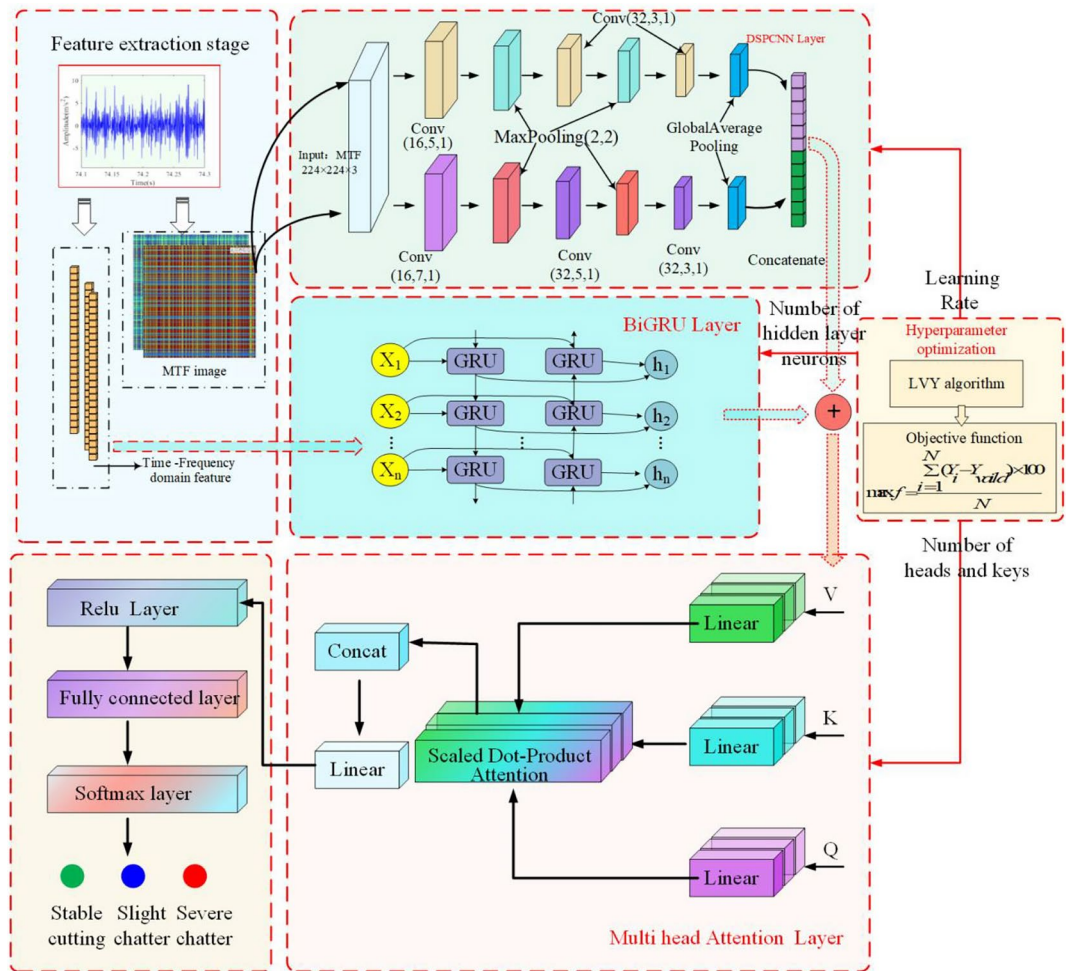


Fig. 2. Milling Chatter Detection Model.

$$\begin{cases}
 Accuracy = (TP + TN)/(TP + TN + FP + FN) \\
 Precision = TP/(TP + FP) \\
 Recall = TP/(TP + FN) \\
 F1\ Score = 2 * (Precision * Recall)/(Precision + Recall) \\
 Specificity = TN/(TN + FP)
 \end{cases}
 \tag{18}$$

Milling experiment

Experimental setup

The machine tool used is a VDL-1000E three-axis CNC machine tool manufactured by Dalian Machine Tool Group. The cutting tool is a flat-end mill with 4 teeth and a diameter of 10 mm. The workpiece material is TC4 titanium alloy with dimensions of 200×200×5 mm. A PCB accelerometer with a sensitivity of 10.42 mV/g and a Donghua DH5922 acquisition system.

were used to collect acceleration signals during the machining process. Sampling frequency is 5000 Hz. The workpiece stiffness was modified through the installation of spacer blocks on both sides. The milling mode is down-milling with dry cutting. The experimental setup is shown in Fig. 3. The simulation experiments are conducted on a Windows 10 (64-bit) operating system, with hardware specifications including an Intel(R) Core (TM) i9-12900 k CPU, NVIDIA GeForce RTX3080, 3.2 GHz clock speed, and 32 GB RAM. All simulations were executed using Matlab R2023b.

Experimental parameter settings

The modal test was conducted using a force hammer with a sensitivity of 3.41 pC/N to obtain the natural frequency of the titanium alloy thin-walled component, which was 864 Hz. The cutting parameters and machining states selected based on the stability lobe diagram drawn from reference² are shown in Table 2. Among them, the radial cut width is 0.5 mm, and the feed per tooth is 0.1 mm/tooth.

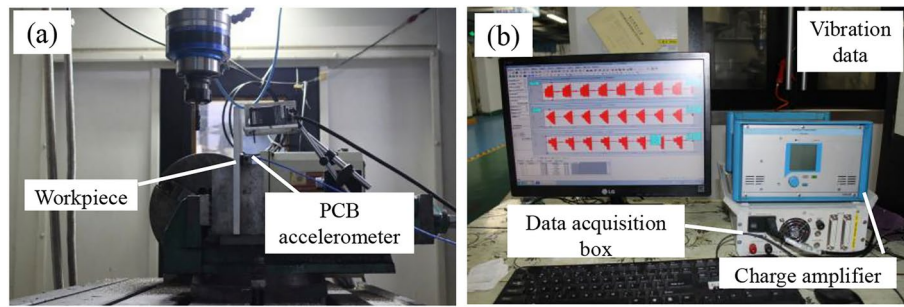


Fig. 3. Experimental setup (a) Experimental site (b) Vibration data acquisition.

No	Cutting parameters		Machining state	No	Cutting parameters		Machining state	
	n (r/min)	ap (mm)			n (r/min)	ap (mm)		
1	750	6.8	Stable	13	1050	7.2	Slight chatter	
2	750	7.2		14	1050	7.5		
3	900	5.9		15	1200	5.1		
4	900	6.3		16	1200	5.4		
5	1050	6.0		Severe chatter	17	750	8.2	Slight chatter
6	1050	6.4			18	750	8.5	
7	1200	4.3			19	900	7.6	
8	1200	4.7			20	900	7.8	
9	750	7.9	Slight chatter		21	1050	8.0	Severe chatter
10	750	8.0			22	1050	8.2	
11	900	7.0			23	1200	5.8	
12	900	7.1			24	1200	6.5	

Table 2. Processing status under different conditions.

Parameters	Ensemble size	Noise standard deviation	Singular value
Range	[500,1000]	[0.5,1.5]	[50,100]
Default value	600	0.95	50
Optimal value	818	1.2	80

Table 3. Parameters of CEEMD-SVD.

Results analysis

Signal denoising analysis

Based on Experiment 4, the denoising effect of the IVY-CEEMD-SVD method is analyzed. The detailed information of the IVY algorithm optimizing the CEEMD-SVD parameters, including the range of values and the optimized values, is reflected in Table 3. The comparison results of different denoising methods after optimizing the model parameters are shown in Fig. 4. The evaluation results of different denoising methods are presented in Table 4.

As shown in Table 4, compared to other methods, the proposed method achieves up to a 23.9% improvement in SNR, a 14.3% reduction in KR, and a 16.67% increase in CL. The significant improvement in SNR indicates excellent denoising performance. The slight decrease in KR indicates that while removing noise, the denoising process largely preserved the peak characteristics of the signal. The significant increase in CL demonstrates that the denoised signal maintains high similarity with the original signal. After optimization of the CEEMD-SVD model, the SNR and CL increased by 0.23 and 0.05, respectively, while the KR decreased by 0.06. This demonstrates the effectiveness of the optimization method.

Markov transition field

The original data under different states is segmented using a sliding window. The window size is set to 5000. The Markov transition field images of signals under different states are shown in Fig. 5.

As shown in Fig. 5, the color variation characteristics of the Markov transition field provide rich information about the machining states. Under stable conditions, the transition field exhibits light colors with good predictability. As chatter emerges and intensifies, the colors gradually darken and form concentrated regions,

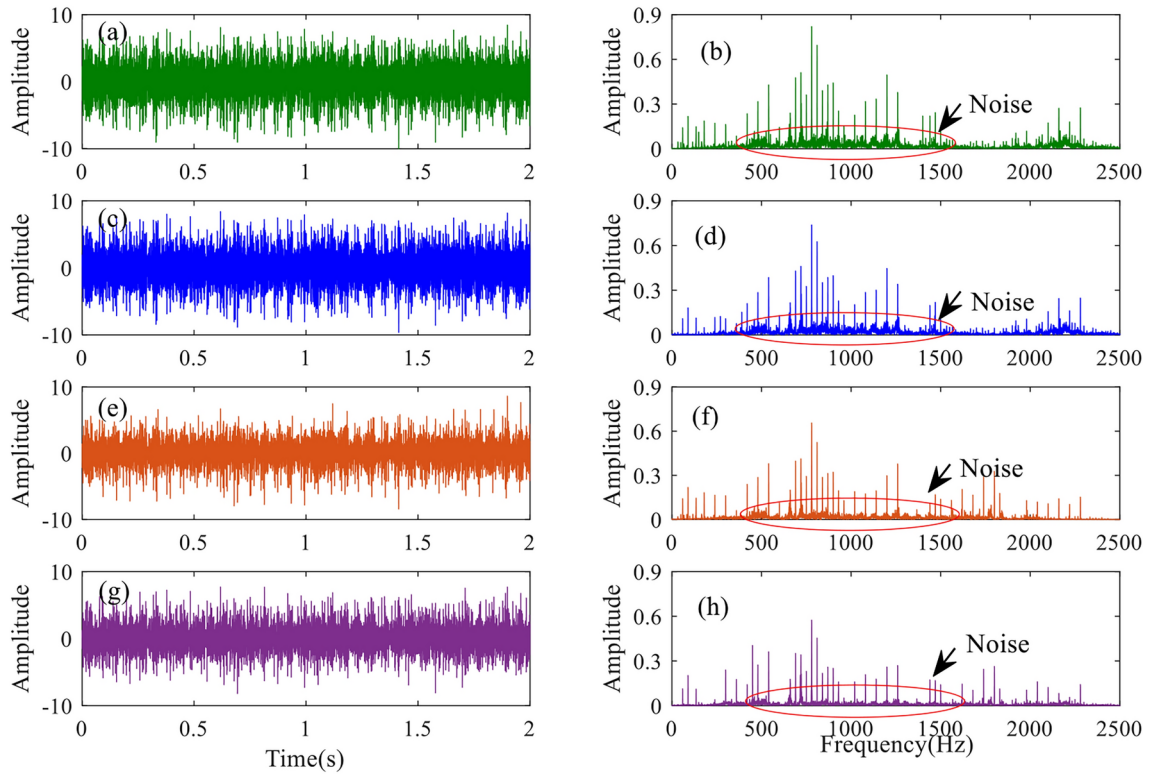


Fig. 4. Comparison results of different noise reduction methods (a) Original signal, (b) Fourier transform of the original signal, (c) CEEMD denoised signal, (d) Fourier transform of CEEMD denoised signal, (e) SVD denoised signal, (f) Fourier transform of SVD denoised signal, (g) Optimize the CEEMD-SVD denoised signal, (h) Fourier transform of CEEMD-SVD denoised signal.

Method	SNR	KR	CL
SVD	3.22	1.19	0.84
CEEMD	3.67	1.11	0.91
CEEMD-SVD	3.76	1.08	0.93
IVY-CEEMD-SVD	3.99	1.02	0.98

Table 4. Comparison of evaluation results for different denoising methods.

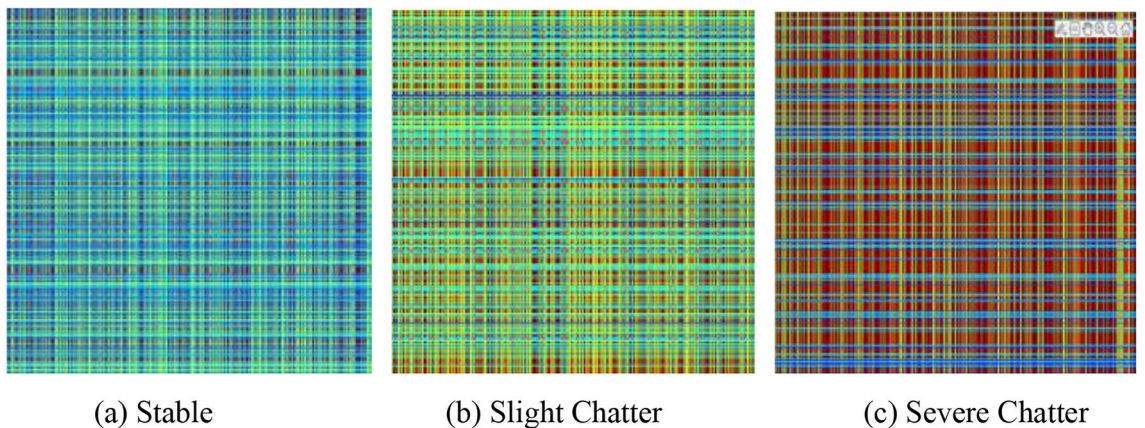


Fig. 5. Markov transition fields of different states.

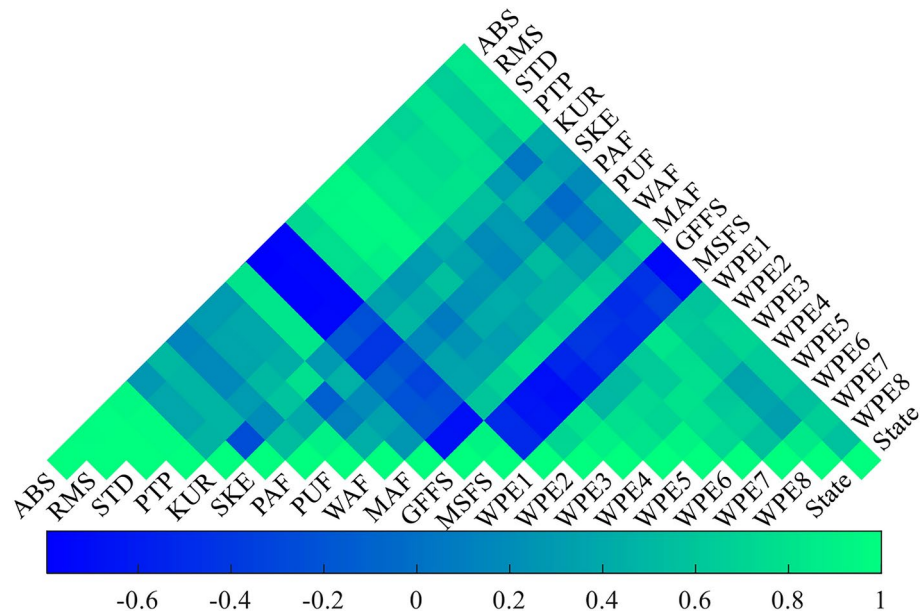


Fig. 6. Correlation analysis between feature extraction and processing status.

Rank	1	2	3	4	5	6	7	8	9	10
Name	STD	RMS	ABS	PTP	GFFS	WPE3	MSFS	MAF	WPE2	WPE7
Value	0.83524	0.83522	0.83492	0.82036	-0.74217	0.70495	-0.7003	0.6924	0.68684	0.65797
Rank	11	12	13	14	15	16	17	18	19	20
Name	WPE4	WPE5	WPE1	WPE8	WPE6	KUR	PUF	WAF	SKE	PAF
Value	0.64128	0.63091	0.58704	0.56585	0.56067	0.39129	0.3659	0.35911	0.33939	0.18326

Table 5. Ranking of feature importance.

reflecting increased frequency and volatility in state transitions. The light-colored regions during stable machining correspond to small amplitude variations and stable machining behavior. The medium-intensity colors during slight chatter indicate the onset of irregularities in tool-workpiece interactions. Under severe chatter conditions, the darkest regions represent strong correlations between adjacent states, indicating sustained and intense vibration patterns.

Correlation analysis

Correlation analysis is used to select features that are sensitive to changes in the machining state. Figure 6 displays the Pearson correlation coefficient matrix of time–frequency domain features. Table 5 summarizes the results of the correlation analysis between time–frequency domain features and the machining state. It can be observed that the GFFS and MSFS features are negatively correlated with the machining state. GFFS negative correlation indicates that chatter causes significant shifts in the spectral centroid. MSFS negative correlation reflects the discretization trend of frequency domain energy distribution. In the chatter state, energy is dispersedly distributed across a wide frequency band. The remaining time–frequency domain features are positively correlated with the machining state. The correlation coefficients of STD, RMS, and ABS are 0.83524, 0.83522, and 0.83492, respectively. The high positive correlation of RMS reflects the cumulative effect of system energy during the chatter process. STD reflects the degree of vibration amplitude fluctuation in the cutting process. ABS characterizes the average amplitude deviation from zero point. Their positive correlations reflect the decrease in cutting system stability during chatter state. Overall, time domain features show the strongest correlation, followed by frequency domain features. The impact of chatter on the system is first manifested in amplitude changes, followed by frequency characteristic alterations.

A threshold Pearson correlation coefficient of 0.6 is established for feature selection, with features exceeding this threshold being retained for further analysis. The results indicate that STD, RMS, ABS, PTP, GFFS, WPE3, MSFS, MAF, WPE2, WPE7, WPE4, and WPE5 show a strong correlation with the machining state. The selection of 12 features designed for the time domain, frequency domain, and time–frequency domain provides rich information for the machining state indicators.

Analysis of classification results

The detailed information of the IVY algorithm optimizing the DBMA model parameters, including the range of values and the optimized values, is reflected in Table 6. All data are divided into training, validation, and test sets, with proportions of 80%, 10%, and 10%, respectively. Under the optimized hyperparameters of the DBMA model, the identification results of the training, validation, and test sets are shown in Fig. 7. In the classification scheme, the numerical labels 1, 2, and 3 correspond to stable machining, slight chatter, and severe chatter conditions, respectively.

Analysis of Fig. 7a and b reveal that the model achieved a recognition accuracy of 98.94% on the training set. Two instances of stable cutting are misidentified as slight chatter. One instance of slight chatter is misidentified as stable cutting, and another as severe chatter. One

instance of severe chatter is misidentified as slight chatter. Examination of Fig. 7c and d indicate that the model attained a recognition accuracy of 98.31% on the validation set. In this set, a single instance of slight chatter is misclassified as stable cutting. From Fig. 7e and f, it is observed that the recognition accuracy of the test set is 96.61%, with one instance of slight chatter misidentified as stable cutting and another as severe chatter.

The ten fold cross validation results can eliminate the bias caused by random segmentation of the dataset³⁹. The detection results before and after hyperparameter optimization are shown in Table 7.

As can be seen from Table 7, the optimization effect is significant, and all indicators have been significantly improved. In terms of accuracy, the optimized method has improved by about 5–6 percentage points compared to the default value. In terms of ten fold cross validation, the optimized method also improved by about 6–9 percentage points. Overall, the optimized model has good generalization ability. The obtained results are consistent with the conclusion of reference^{39,40}.

The ten fold cross validation mean results of machining state detection accuracy before and after denoising are shown in Table 8.

As shown in Table 8, the machining state detection accuracy is highest using the IVY-CEEMD-SVD denoised signal, followed by the CEEMD-SVD denoised signal, with the raw signal showing the lowest accuracy. Compared to other methods, the proposed approach achieves up to a 2.18% improvement in detection accuracy on the training set. It demonstrates up to a 1.89% improvement in detection accuracy on the validation set. On the test set, it shows up to a 1.34% improvement in detection accuracy. These results demonstrate that removing noise from machining signals can improve detection accuracy. These findings are consistent with the conclusions in reference⁴¹.

The CEEMD-SVD method with IVY optimization demonstrates significant advantages compared to the traditional CEEMD-SVD method. Specifically, the IVY-CEEMD-SVD method improved detection accuracy by 1.06, 1.03, and 0.20 percentage points on the training set, validation set, and test set, respectively.

The ten fold cross validation mean results of machining state detection accuracy under different modal data are shown in Table 9.

As shown in Table 9, the highest machining state detection accuracy is achieved using multi-modal data, followed by temporal data. Compared to other data types, multi-modal data demonstrates up to a 1.17% improvement in detection accuracy on the training set. It shows up to a 1.35% improvement in detection accuracy on the validation set. On the test set, it achieves up to a 1.0% improvement in detection accuracy.

The features extracted from different network layers of DBMA are visualized using the t-SNE algorithm, and the results are shown in Fig. 8.

Figure 8a shows the t-SNE visualization of test set data samples at the model input layer. It can be observed that various data samples are intermixed, showing poor clustering performance. Figure 8b presents the visualization results after feature extraction by the DSPCNN layer. As evident, samples of the same class have largely converged together, though some minor mixing still exists. Figure 8c shows the visualization results after feature extraction by the BiGRU layer. The data is fundamentally differentiated, but significant intermixing still exists with poor convergence. Comparing Fig. 8a and b, it is clear that the recognition accuracy of the DSPCNN layer is higher than that of the BiGRU layer. Figure 8d presents the visualization results after feature extraction by the MSA layer. It is evident that all data samples are completely segregated into three classes, corresponding to three machining states. The DBMA model effectively classifies and recognizes milling states.

The Davies Bouldin index (DBI), silhouette coefficient (SC), and Calinski Harabasz index (CHI) are used to evaluate the tightness and separation of feature distributions. The results are shown in Table 10.

As shown in Table 10, the model demonstrates continuously improving clustering performance from the initial to final stages. The DBI metric decreased from 0.82 to 0.58, a reduction of 29.3%, indicating significant improvement in intra-cluster compactness and inter-cluster separation. The SC metric increased from 0.65 to 0.73, an improvement of 12.3%, reflecting gradually enhanced accuracy in sample classification. The CHI metric increased from 2500 to 3200, a 28% improvement, suggesting continuous optimization of the ratio between inter-cluster and intra-cluster variance.

Parameters	DSPCNN learning rate	Number of hidden layer neurons in BiGRU	Number of MSA heads
Range	[0.001,0.1]	[0,100]	[2,6]
Default value	0.001	16	2
Optimal value	0.029581	25	4

Table 6. Parameters of the DBMA model.

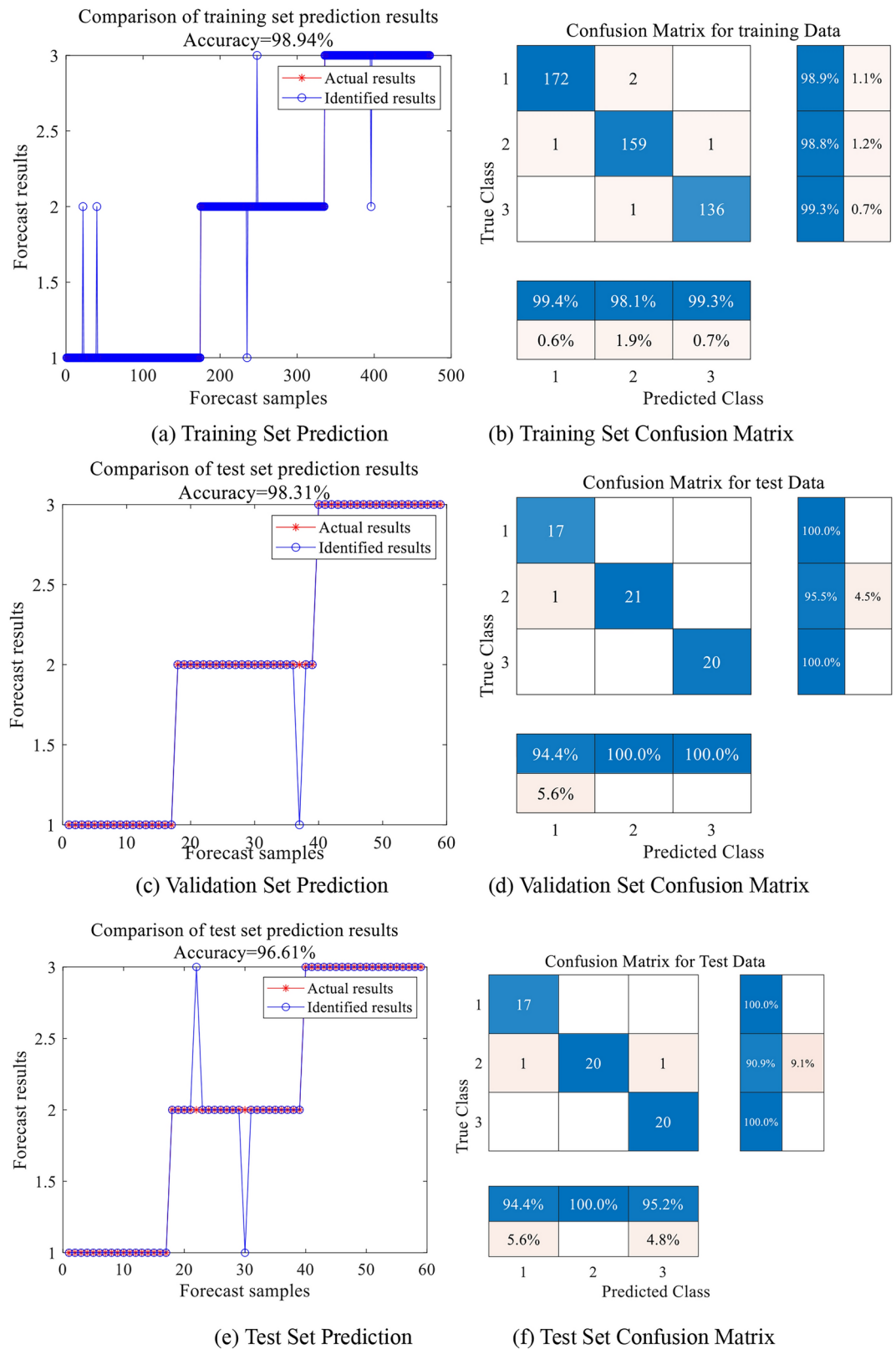


Fig. 7. Recognition results of different machining states.

To verify the reliability and superiority of the model, comparisons are made with CNN-BiGRU, CNN-BiGRU-MSA, ISR-CNN³¹, and ILR-DNN¹¹. To minimize error, the ten fold cross validation mean results of five repeated experiments are used as the final evaluation results. The comparison results for the training set are shown in Fig. 9. The comparison results for the test set are shown in Fig. 10. The ten fold cross validation mean

Categories		Training set (%)	Validation set (%)	Test set (%)
Accuracy	Default	93.05	91.48	90.95
	Optimal	98.94	98.31	96.61
Ten fold cross validation mean	Default	90.76	88.63	87.06
	Optimal	97.57	97.14	95.73

Table 7. Detection results before and after hyperparameter optimization.

Signal type	Training set (%)	Validation set (%)	Test set (%)
Original signal	96.75	96.27	95.19
SVD denoised signal	97.11	96.91	95.80
CEEMD denoised signal	97.72	97.07	96.32
CEEMD-SVD denoised signal	97.87	97.13	96.33
IVY-CEEMD-SVD denoised signal	98.93	98.16	96.53

Table 8. Machining state detection results under different signals.

Data type	Training set (%)	Validation set (%)	Test set (%)
Temporal data	97.79	97.16	96.17
Image data	97.65	96.97	95.49
Multi-modal data	98.82	98.32	96.49

Table 9. Machining state detection results under different modal data.

results of different methods for calculating time are shown in Table 11. Among them, the prediction is divided into four scenarios with sample sizes of 10, 20, 30, and 40.

As shown in Fig. 9, the proposed method outperforms all comparison methods across various evaluation metrics. The accuracy reaches 98.9%, representing an improvement of up to 6.7% compared to the other four methods. The precision achieves 99.0%, showing an increase of up to 6.4%, while the recall attains 98.6%, demonstrating an improvement of up to 5.7%. Furthermore, the average F1-score and specificity reach 98.8% and 99.2%, respectively, exhibiting improvements of up to 6.4% and 2.7% compared to the baseline methods.

As shown in Fig. 10, the proposed method achieves the highest accuracy and recall of 96.2% and 96.8% among all methods. The maximum improvement compared to the comparative method is 7.6% and 5.0%, respectively. The specificity value is 98.2%, showing improvements of 0.3–7.9% over other methods. The method demonstrates differentiated performance in terms of precision and F1-score. The precision and F1-score of the proposed method are 95.1% and 96.1%, outperforming three comparative methods but 0.9% and 0.2% lower than the ILR-DNN method.

As can be seen from Table 11, the proposed method shows significant differences in computation time compared to the other four methods. During the training phase, the proposed method took 183.40 s, an increase of 44.5% compared to the fastest CNN BiGRU method (126.88 s). In the prediction stage, when processing sample sizes of 10, 20, 30, and 40, the proposed method requires 25.61 ms, 48.7 ms, 67.26 ms, and 83.23 ms, respectively. Although the proposed method has a longer computation time, considering its significantly improved detection performance (accuracy increased by 7.6%), this increase in computational overhead is acceptable. Moreover, with a prediction time of 83.23 ms for 40 samples, it still meets the requirements for real-time monitoring.

Vibration signals under three states in variable stiffness conditions are selected to verify the robustness of the proposed method. The initial stiffness of the workpiece is 1.26×10^5 N/m. The terminal stiffness of the workpiece is 2.23×10^5 N/m. The radial cut width is 0.5 mm, and the feed per tooth is 0.1 mm/tooth. The spindle speed is 1000 r/min, and the axial cutting depth is 7.8 mm. The signals and detection results under different machining states are shown in Fig. 11.

The experimental results show an overall machining state detection accuracy of 94.94%. The detection accuracy for severe chatter is 93.84%. The detection accuracy for slight chatter is 87.5%. The detection accuracy for stable cutting is 97.14%. Four cases of severe chatter are misidentified as slight chatter. Two cases of slight chatter are misidentified as severe chatter. One case of slight chatter is misidentified as stable cutting. Two cases of stable cutting are misidentified as slight chatter. These results demonstrate that the proposed method exhibits good robustness and stability.

Conclusion

An Ivy-CEEMD-SVD data denoising model is constructed in this paper. Time–frequency domain methods and Markov transition fields are used to extract multi-modal data features related to machining states. A DBMA for chatter detection is constructed by integrating DSPCNN, BiGRU, and MSA. Finally, the advantages of the

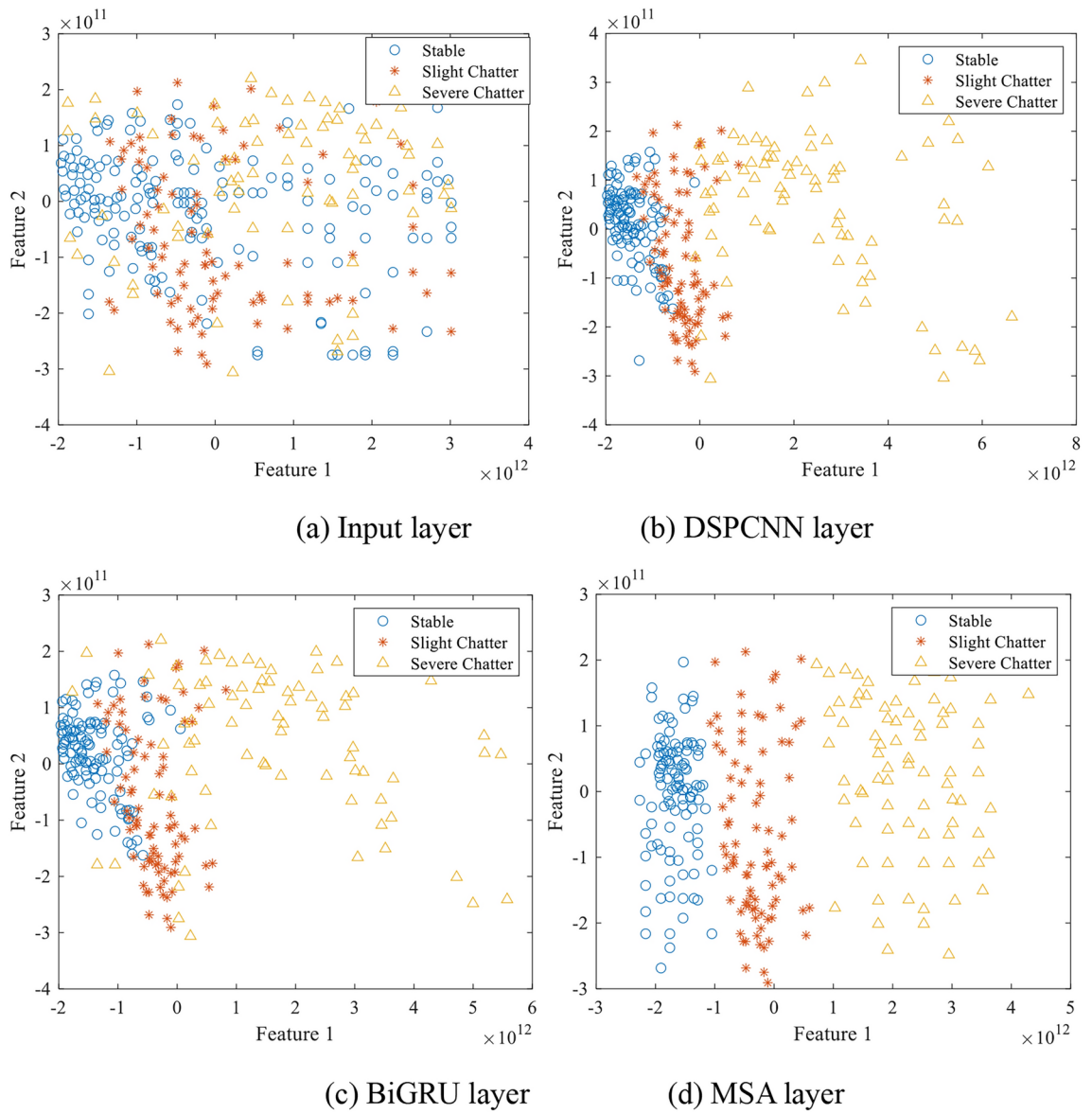


Fig. 8. Visualization results of classification model.

Stage	DBI	SC	CHI
Input layer	0.82	0.65	2500
DSPCNN layer	0.65	0.72	2800
BiGRU layer	0.71	0.69	2650
MSA layer	0.58	0.74	3200

Table 10. Cluster quality evaluation indicators at different training stages.

proposed method are validated through evaluation metrics such as accuracy. The specific conclusions are as follows.

- (1) Compared to SVD and CEEMD, the proposed denoising method achieved up to 23.9% improvement in SNR, 14.3% reduction in KR, and 16.67% improvement in CL. The detection accuracy of machining conditions can be improved by using the denoised signal.
- (2) Through Pearson correlation analysis of time-frequency domain features, it is found that STD, RMS, ABS, PTP, GFFS, WPE3, MSFS, MAE, WPE2, WPE7, WPE4, and WPE5 have a strong correlation with the machining state.

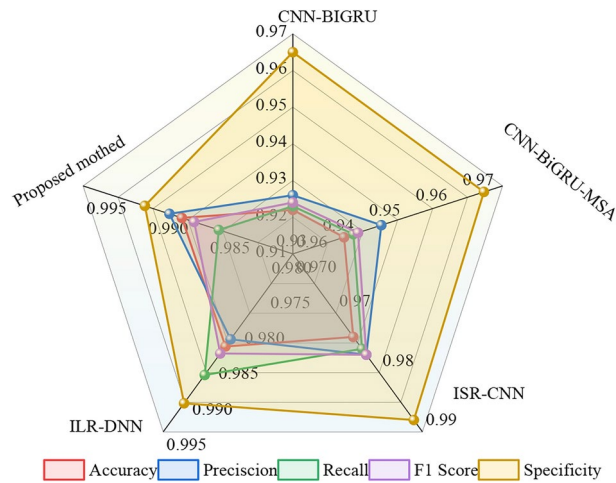


Fig. 9. Comparison results on the training set.

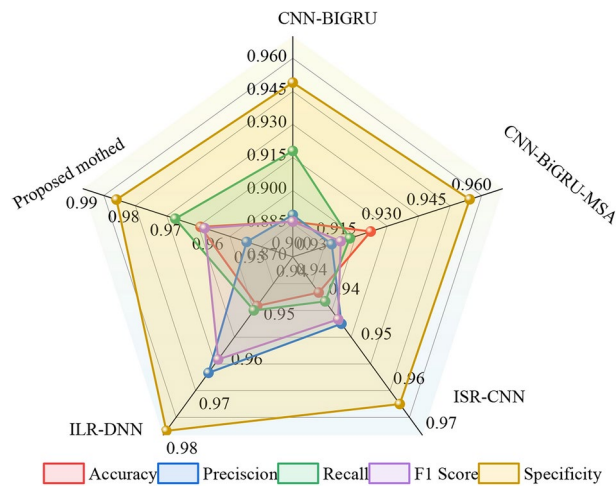


Fig. 10. Comparison results on the test set.

Time	CNN-BiGRU	CNN-BiGRU-MSA	ISR-CNN	ILR-DNN	Proposed method
Training (S)	126.88	161.35	155.6	149.31	183.40
Prediction ₁₀ (ms)	16.82	20.84	14.47	12.3	25.61
Prediction ₂₀ (ms)	31.53	39.09	27.6	22.56	48.7
Prediction ₃₀ (ms)	44.13	54.62	37.87	31.53	67.26
Prediction ₄₀ (ms)	54.61	67.65	46.82	39.08	83.23

Table 11. Calculation time for different methods.

- (3) The t-SNE algorithm is used to visualize features extracted from different network layers. The results indicate that the recognition accuracy of the DSPCNN layer is higher than that of the BiGRU layer.
- (4) Compared to unimodal data, multimodal data achieved up to 1.17% improvement in detection accuracy on the training set, 1.35% on the validation set, and 1.0% on the test set.
- (5) Compared to CNN-BiGRU, CNN-BiGRU-MSA, ACNN-BiLSTM, and ILR-DNN, the proposed method achieved improvements of up to 7.6% in accuracy, 6.4% in precision, 5.7% in recall, 6.4% in F1 score, and 2.7% in specificity.
- (6) The proposed method still meets the requirements of real-time monitoring with a prediction time of 83.23 ms in the case of 40 samples.
- (7) The accuracy of the proposed method in detecting machining status under variable working conditions is 94.94%, which proves that the proposed method has good robustness.

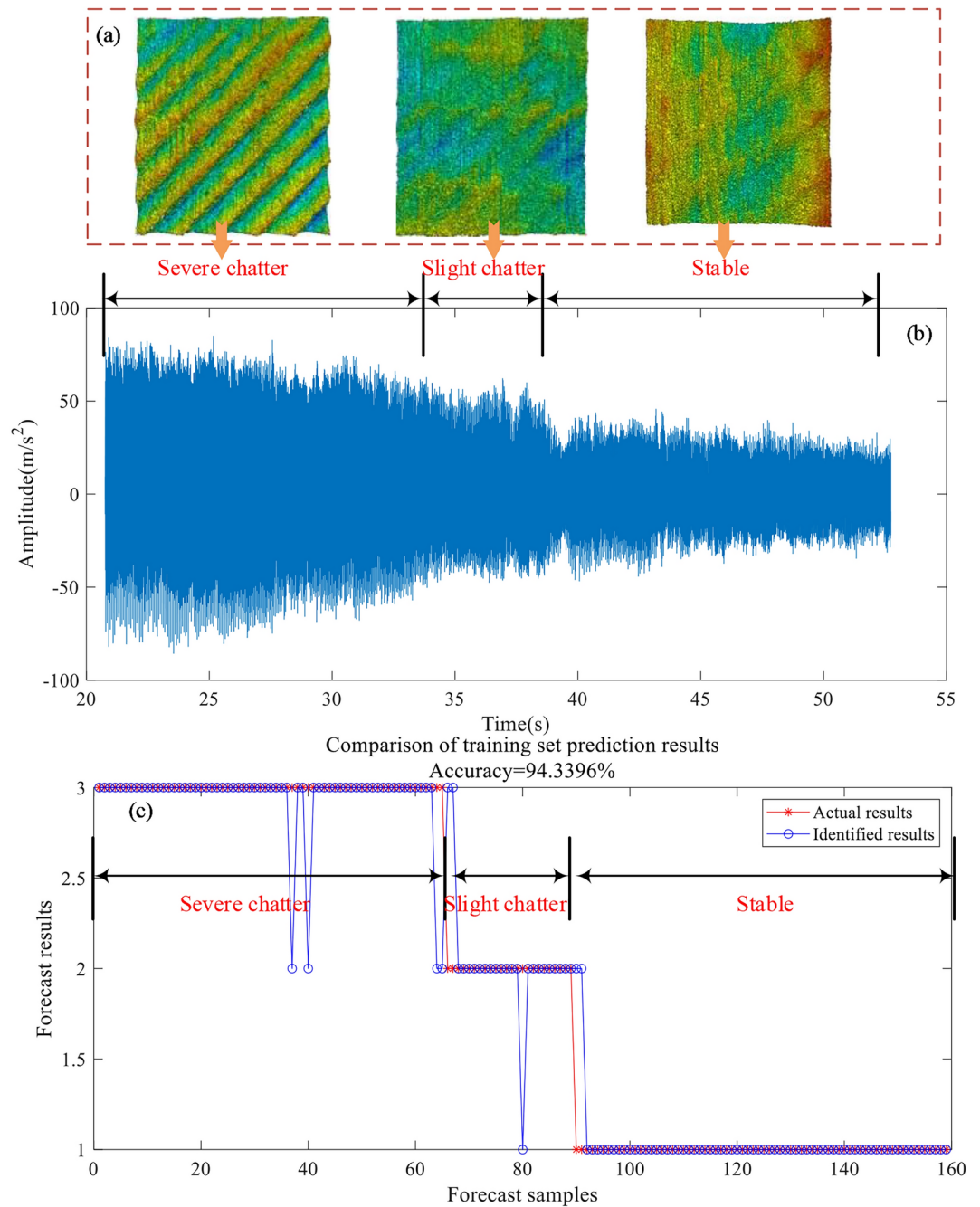


Fig. 11. Signals and detection results under different machining states (a) Surface topography (b) Vibration signal (c) Maching state detection.

Multi-signal fusion can overcome the limitations of individual signals and improve chatter detection accuracy. Future work will explore multi-source signal and multi-modal data fusion for machining state detection.

Data availability

Data is provided within the manuscript or supplementary information files.

Received: 16 December 2024; Accepted: 28 January 2025

Published online: 31 January 2025

References

1. Wang, Z. et al. Feasibility and mechanism of atmospheric pressure cold plasma jet (APCPJ) assisted micro-milling of bulk metallic glasses (BMGs). *Ceram Int.* **50**(7), 11094–11105. <https://doi.org/10.1016/j.ceramint.2024.01.011> (2024).

2. Yang, Y. et al. Generalized model for dynamics and stability of milling of titanium alloys by integrating process damping, multiple modes and multiple delays. *J. Manuf. Processes*. **102**, 718–735. <https://doi.org/10.1016/j.jmapro.2023.07.071> (2023).
3. Gao, H. et al. Milling chatter detection system based on multi-sensor signal fusion. *IEEE Sens. J.* **21**(22), 25243–25251. <https://doi.org/10.1109/JSEN.2021.3058258> (2021).
4. Cui, X. et al. Performance and reliability of Al₂O₃/TiC ceramic tool with multi-bionic surface microstructure in intermittent dry turning of AISI 52100 hardened steel. *J. Mater. Res. Technol.* **25**, 7219–7240. <https://doi.org/10.1016/j.jmrt.2023.07.132> (2023).
5. Li, W., Wang, L. & Yu, G. Chatter prediction in flank milling of thin-walled parts considering force-induced deformation. *Mech. Syst. Signal Pr.* **165**, 108314. <https://doi.org/10.1016/j.ymsp.2021.108314> (2022).
6. Li, Z. et al. Grinding force model for ultrasonic assisted grinding of γ -TiAl intermetallic compounds and experimental validation. *Rev. Adv. Mater. Sci.* **63**(1), 20230167. <https://doi.org/10.1515/rams-2023-0167> (2024).
7. Yesilli, M. C., Khasawneh, F. A. & Otto, A. Chatter detection in turning using machine learning and similarity measures of time series via dynamic time warping. *J. Manuf. Processes*. **77**, 190–206. <https://doi.org/10.1016/j.jmapro.2022.03.009> (2022).
8. Albertelli, P. et al. Development of a generalized chatter detection methodology for variable speed machining. *Mech. Syst. Signal Pr.* **123**, 26–42. <https://doi.org/10.1016/j.ymsp.2019.01.002> (2019).
9. Song, Y., Cao, J. & Hu, Y. In-process feature extraction of milling chatter based on second-order synchroextracting transform and fast kurtogram. *Mech. Syst. Signal Pr.* **208**, 111018. <https://doi.org/10.1016/j.ymsp.2023.111018> (2024).
10. Li, K. et al. A novel online chatter detection method in milling process based on multiscale entropy and gradient tree boosting. *Mech. Syst. Signal Pr.* **135**, 106385. <https://doi.org/10.1016/j.ymsp.2019.106385> (2020).
11. Sun, Y. et al. Online chatter detection considering beat effect based on Inception and LSTM neural networks. *Mech. Syst. Signal Pr.* **184**, 109723. <https://doi.org/10.1016/j.ymsp.2022.109723> (2023).
12. Aslan, D. & Altintas, Y. On-line chatter detection in milling using drive motor current commands extracted from CNC. *Int. J. Mach. Tool Manu.* **132**, 64–80. <https://doi.org/10.1016/j.ijmactools.2018.04.007> (2018).
13. Ji, J. et al. Grafting-inspired bionic hierarchical structure for efficiently enhancing cemented carbide tool performance in green turning. *J. Mater. Process. Tech.* **324**, 118274. <https://doi.org/10.1016/j.jmatprot.2023.118274> (2024).
14. Tran, M. Q., Liu, M. K. & Elsis, M. Effective multi-sensor data fusion for chatter detection in milling process. *ISA t* **125**, 514–527. <https://doi.org/10.1016/j.isatra.2021.07.005> (2022).
15. Wang, W. K. et al. Chatter detection methods in the machining processes: A review. *J. Manuf. Processes*. **77**, 240–259. <https://doi.org/10.1016/j.jmapro.2022.03.018> (2022).
16. Kuljanic, E., Sortino, M. & Totis, G. Multisensor approaches for chatter detection in milling. *J. Sound Vib.* **312**(4–5), 672–693. <https://doi.org/10.1016/j.jsv.2007.11.006> (2008).
17. Matthew, D. E. et al. Improved STFT analysis using time-frequency masking for chatter detection in the milling process. *Measurement* **225**, 113899. <https://doi.org/10.1016/j.measurement.2023.113899> (2024).
18. Yin, C. et al. Attention-driven transfer learning framework for dynamic model guided time domain chatter detection. *J. Intell. Manuf.* **35**(4), 1867–1885. <https://doi.org/10.1007/s10845023021330> (2024).
19. Kuo, P. H. et al. Machine tool chattering monitoring by Chen-Lee chaotic system-based deep convolutional generative adversarial nets. *Struct. Health. Monit.* **22**(6), 3891–3907. <https://doi.org/10.1177/14759217231159865> (2023).
20. Yan, S. & Sun, Y. Early chatter detection in thin-walled workpiece milling process based on multi-synchrosqueezing transform and feature selection. *Mech. Syst. Signal Pr.* **169**, 108622. <https://doi.org/10.1016/j.ymsp.2021.108622> (2022).
21. Wan, S. et al. Milling chatter detection by multi-feature fusion and Adaboost-SVM. *Mech. Syst. Signal Pr.* **156**, 107671. <https://doi.org/10.1016/j.ymsp.2021.107671> (2021).
22. Jauhari, K. et al. Building digital-twin virtual machining for milling chatter detection based on VMD, synchro-squeeze wavelet, and pre-trained network CNNs with vibration signals. *J. Intell. Manuf.* **35**(7), 3083–3114. <https://doi.org/10.1007/s10845023021950> (2024).
23. Chen, K., Zhang, X. & Zhao, W. Automatic feature extraction for online chatter monitoring under variable milling conditions. *Measurement* **210**, 112558. <https://doi.org/10.1016/j.measurement.2023.112558> (2023).
24. Wan, M. et al. Chatter detection for micro milling considering environment noises without the requirement of dominant frequency. *Mech. Syst. Signal Pr.* **199**, 110451. <https://doi.org/10.1016/j.ymsp.2023.110451> (2023).
25. Lu, Y. et al. An early chatter detection method based on multivariate variational mode decomposition and chatter correlation factor. *IEEE/ASME T Mech.* **27**(6), 5724–5735. <https://doi.org/10.1109/TMECH.2022.3188680> (2022).
26. Chen, D. et al. Development of a novel online chatter monitoring system for flexible milling process. *Mech. Syst. Signal Pr.* **159**, 107799. <https://doi.org/10.1016/j.ymsp.2021.107799> (2021).
27. Zhao, Y. et al. Incipient chatter fast and reliable detection method in high-speed milling process based on cumulative strategy. *ISA t*. **131**, 397–414. <https://doi.org/10.1016/j.isatra.2022.05.039> (2022).
28. Wang, C., Zhang, Y. & Hu, J. Adaptive milling chatter identification based on sparse dictionary considering noise estimation and critical bandwidth analysis. *J. Manuf. Processes*. **106**, 328–337. <https://doi.org/10.1016/j.jmapro.2023.10.012> (2023).
29. Yang, B. et al. Early chatter detection in robotic milling under variable robot postures and cutting parameters. *Mech. Syst. Signal Pr.* **186**, 109860. <https://doi.org/10.1016/j.ymsp.2022.109860> (2023).
30. Zhang, P. et al. Improving generalisation and accuracy of on-line milling chatter detection via a novel hybrid deep convolutional neural network. *Mech. Syst. Signal Pr.* **193**, 110241. <https://doi.org/10.1016/j.ymsp.2023.110241> (2023).
31. Lu, Y. et al. An interpretable anti-noise convolutional neural network for online chatter detection in thin-walled parts milling. *Mech. Syst. Signal Pr.* **206**, 110885. <https://doi.org/10.1016/j.ymsp.2023.110885> (2024).
32. Sun, L. et al. An intelligent chatter detection method for high-speed milling under variable tool-workpiece systems and cutting parameters. *Mech. Syst. Signal Pr.* **224**, 111960. <https://doi.org/10.1016/j.ymsp.2024.111960> (2025).
33. Zhang, Y. et al. Multi-modal data cross-domain fusion network for gearbox fault diagnosis under variable operating conditions. *Eng. Appl. Artif. Intel.* **133**, 108236. <https://doi.org/10.1016/j.engappai.2024.108236> (2024).
34. Liang, R. et al. Multimodal data fusion for geo-hazard prediction in underground mining operation. *Comput. Ind. Eng.* <https://doi.org/10.1016/j.cie.2024.110268> (2024).
35. Roy, D. et al. Multi-modality sensing and data fusion for multi-vehicle detection. *IEEE T Multim.* **25**, 2280–2295. <https://doi.org/10.1109/TMM.2022.3145663> (2022).
36. Ghasemi, M. et al. Optimization based on the smart behavior of plants with its engineering applications: Ivy algorithm. *Knowl-Based Syst.* **295**, 111850. <https://doi.org/10.1016/j.knosys.2024.111850> (2024).
37. Gao, H. et al. A monitoring method of milling chatter based on optimized hybrid neural network with attention mechanism. *Facta Univ, Ser: Mech.* <https://doi.org/10.22190/FUME240804047G> (2024).
38. Vakharia, V. & Gujar, R. Prediction of compressive strength and portland cement composition using cross-validation and feature ranking techniques. *Constr. Build Mater.* **225**, 292–301. <https://doi.org/10.1016/j.conbuildmat.2019.07.224> (2019).
39. Vakharia, V. et al. Estimation of lithium-ion battery discharge capacity by integrating optimized explainable-AI and stacked LSTM model. *Batteries* **9**(2), 125. <https://doi.org/10.3390/batteries9020125> (2023).
40. Shawki, N., Nunez, R. R., Obeid, I., et al. On automating hyperparameter optimization for deep learning applications. in *2021 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*. IEEE, 1–7. <https://doi.org/10.1109/SPMB52430.2021.9672266> (2021).
41. Kounta, C. A. K. A. et al. Deep learning for the detection of machining vibration chatter. *Adv. Eng. Softw.* **180**, 103445. <https://doi.org/10.1016/j.advengsoft.2023.103445> (2023).

Author contributions

Haining Gao: calculation, simulation, experiment validation, and writing—original draft preparation. Haoyu Wang: software programming, and experiment validation. Hongdan Shen: calculation, and writing—review and editing. Shule Xing: simulation, and writing—review and editing. Yong Yang: data collection, and writing—review and editing. Yinlin Wang: experiment testing, and writing—review and editing. Wenfu Liu: Lei Yu: experiment testing, and writing—review and editing. Experiment testing, and writing—review and editing. Mazhar Ali: writing—review and editing. Imran Ali Khan: writing—review and editing.

Funding

This work is supported in part by the Henan Province Young Backbone Teachers Support Program in Higher Education (No. 2023GGJS157), the Natural Science Foundation of Henan province (No. 222300420239), the Science and Technology Planning Project in Henan Province (Nos. 212102210326, 242102240132), Program for Innovative Research Team (in Science and Technology) in University of Henan Province (No. 24IRTSTHN020) and the key research and development Program of Henan Province (No. 241111111600).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-88242-7>.

Correspondence and requests for materials should be addressed to H.G.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025