



Depth-wise separable convolutional neural-network-based intelligent chatter monitoring for thin-walled polish grinding

Yuan Zhao, Chunxia Zhu, and Guofa Xu

School of Mechanical Engineering, Shenyang Jianzhu University, Shenyang 110168, China

Correspondence: Chunxia Zhu (team_zhcx@163.com)

Received: 16 May 2025 – Revised: 15 August 2025 – Accepted: 19 August 2025 – Published: 22 October 2025

Abstract. To overcome the limitations of traditional convolutional neural networks in monitoring polishing and grinding chatter, this paper proposes a fusion approach combining deep separable convolutional neural networks (DCNNs) with gated recurrent units (GRUs). The grinding force signal undergoes preprocessing via an optimised variational model decomposition (VMD) algorithm and wavelet threshold denoising, with model hyperparameters optimised using a hybrid genetic–particle swarm algorithm (MGP). The designed squeeze-and-excitation (SE)–DCNN–GRU model employs deep separable convolution for multi-scale feature extraction, utilises the SE attention module to enhance key feature representation, and incorporates residual connections to address network degradation issues. Experiments demonstrate that this model achieves up to 98.8 % recognition accuracy for thin-walled titanium alloy components, maintaining over 95 % accuracy even under -5 dB noise conditions. Convergence speed increases by 39 % with significantly enhanced stability, providing an effective solution for chatter monitoring under complex operating conditions.

1 Introduction

In modern manufacturing, thin-walled components are extensively employed in aerospace, energy, and other industries due to their lightweight advantages, yet the occurrence of chatter can severely impact economic efficiency and impede the advancement of high-speed, high-precision machining (Feng et al., 2016; Cao et al., 2017). Traditional prediction methods based on stability lobe diagrams are plagued by various limitations, underscoring the urgent need for real-time intelligent monitoring technologies. Polish-grinding chatter is essentially the result of a positive feedback coupling between cutting forces and the displacement of the machining system, and it can be broadly classified into free vibration, forced vibration, and self-excited vibration (Altintas and Weck, 2004). Among them, regenerative chatter is induced by the periodic fluctuation of the cutting thickness. In 1954, based on Arnold's theory, Tlusty further deepened the dynamic model of friction-induced chatter and proposed that the relative kinematic characteristics between the micro-elements of the cutting edge of the tool and the surface to

be machined of the workpiece during the dynamic cutting process, the dynamic modulation effect of the tool rake angle, and the phase coupling relationship among the cutting force vectors are the core mechanisms inducing friction-type chatter (Arnold, 1946). Liu et al. (1998) constructed a polish-grinding chatter dynamic model that incorporated clearance nonlinearity, and through a combination of numerical simulations and asymptotic analysis, they revealed the interactive coupling phenomena of two nonlinear modes within the machine tool structure, ultimately deriving explicit expressions for the critical conditions of cutting speed and chatter instability using a multi-scale averaging method (Tlusty and Ismail, 1981).

Chatter monitoring in titanium alloy thin-walled component processing presents multiple challenges. Existing feature extraction techniques are primarily categorised into time-domain analysis (Altintas and Budak, 1995), frequency-domain analysis (Altintas, 2002), and time–frequency analysis (Kuljanic et al., 2008). Traditional machine learning methods, such as support vector machines (SVMs), decision trees, and artificial neural networks (ANN), have been ap-

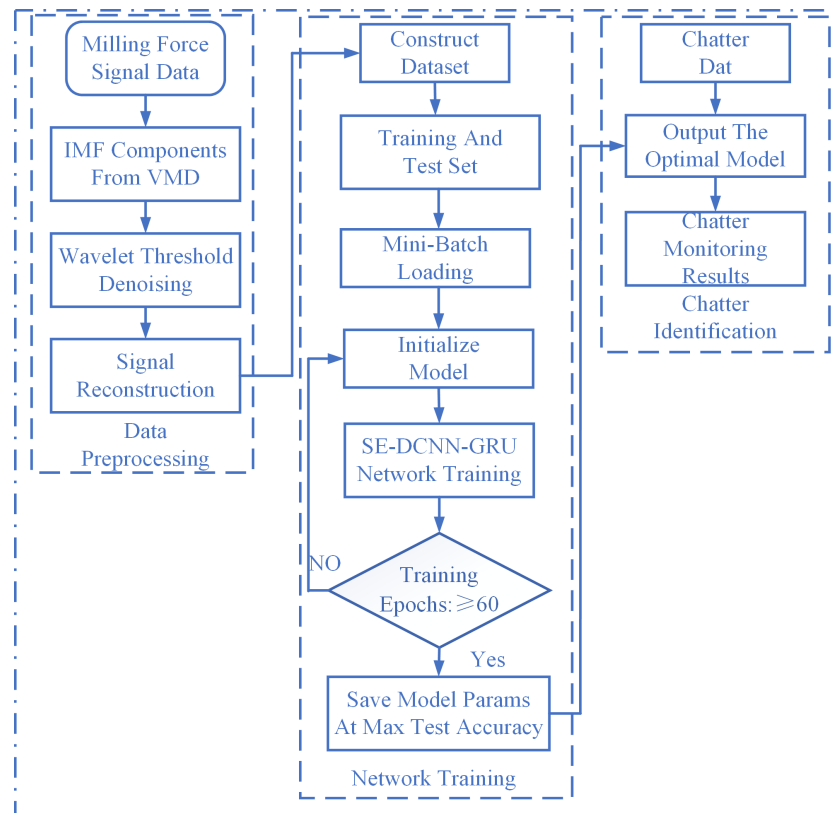


Figure 1. Flowchart of chatter monitoring based on SE-DCNN-GRU.

plied; for instance, Yang et al. (2019) employed simulated annealing (SA) to adaptively determine the optimal number of variational model decomposition (VMD) layers, and Liu et al. (2017) proposed a method for automatically selecting the VMD decomposition layers based on peaks. However, these approaches, which depend heavily on manual feature engineering, often struggle to generalise effectively under complex operating conditions.

In recent years, deep learning technology has made significant breakthroughs. Conventional neural network (CNN) models extract local features through convolution operations, although their performance can be limited by issues such as gradient vanishing (Zhang et al., 2023). Recurrent neural network (RNN) models, including long short-term memory (LSTM) and gated recurrent unit (GRU) models, are adept at capturing temporal dependencies but tend to be inefficient in processing high-dimensional data (Han et al., 2022). Hybrid models, such as CNN-LSTM, have shown promise in vibration monitoring applications (Zhou et al., 2022), yet the mechanisms for effective feature fusion remain to be optimised. The limitations of traditional convolutional neural networks (CNNs) in vibration signal detection are primarily manifested in three aspects: first, the one-sided nature of feature extraction; second, insufficient ability to capture temporal dynamic features; and third, weak noise resis-

tance. These limitations are particularly evident in the context of chatter monitoring during milling of thin-walled components. For example, in milling experiments on TC4 titanium alloy thin-walled components, when using a traditional CNN model for chatter identification of milling force signals, the highest accuracy rate achieved was only 90.2 %. This result stems from the fact that the convolution kernels in traditional CNNs are primarily designed to target local spatial features, making it difficult to effectively extract the multi-scale frequency components inherent in chatter signals. When the milling process transitions from a stable state to mild chatter, the high-frequency components of the signal exhibit non-stationary fluctuations. However, the fixed convolution kernels in traditional CNNs cannot adaptively capture such dynamic changes, leading to incomplete feature extraction.

Existing research shows that deep separable CNNs (DCNNs) reduce the parameter load by decomposing the convolution kernels (Tariq et al., 2022), while the squeeze-and-excitation (SE) attention mechanism strengthens the correlation between channels, providing new insights for overcoming the limitations inherent in traditional CNNs (Cao et al., 2013). Motivated by these challenges and advancements, this study presents an intelligent monitoring approach for polish-grinding chatter that integrates DCNN with GRU (Yang et al., 2019). A VMD parameter optimisation strategy based

on fuzzy entropy is proposed, which, in conjunction with wavelet threshold denoising (Zhu et al., 2020), significantly improves signal quality in the preprocessing stage. The resultant SE-DCNN-GRU model achieves multi-scale feature extraction via deep separable convolution, leverages an SE module to enhance key feature representation, and employs a novel MGP (Cherukuri et al., 2019) hybrid algorithm (genetic algorithm and particle swarm optimisation) designed to achieve global optimisation of network hyperparameters (Zacharia and Krishnakumar, 2020).

In summary, the main contributions of this study can be outlined as follows:

1. An optimised signal preprocessing method combining VMD with wavelet threshold (Zhu et al., 2010) denoising is developed (Li et al., 2025). By rigorously determining the optimal values for the VMD decomposition level and penalty factor, the method effectively suppresses noise interference during signal reconstruction, thereby providing a robust foundation for subsequent neural network training and testing (Gao et al., 2020).
2. The SE-DCNN-GRU network model is constructed to enhance polish-grinding chatter monitoring. The model employs an inception module for multi-scale feature extraction and leverages deep separable convolution to accelerate training. Subsequently, the integration of an SE attention mechanism deepens feature extraction, and the softmax function is utilised to generate the final classification results (Liu et al., 2025).
3. The MGP fusion optimisation algorithm proposed in this paper is used to optimise the hyperparameters of the SE-DCNN-GRU network model, and finally a network model with the optimal solution is fully constructed (Zhou et al., 2023). Through various means such as loss function evaluation and accuracy curve plotting, it is fully verified that this model has a good recognition effect. Finally, through a comprehensive comparison with different models and algorithms, the significant advantages of the algorithm proposed in this paper in terms of performance, accuracy, etc. are highlighted.

The remainder of this paper is organised as follows: Sect. 2 introduces the components of the attention-enhanced DCNN-GRU monitoring model; Sect. 3 details the hyperparameter optimisation process for the monitoring model; Sect. 4 presents experimental validation and analysis of the model's recognition accuracy; and Sect. 5 concludes the study with a summary of the findings.

2 Attention-enhanced DCNN-GRU chatter monitoring model

2.1 Overview

This paper proposes a chatter monitoring system based on the SE-DCNN-GRU neural network model. The flowchart is shown in Fig. 1. By adopting the depth-wise separable convolution technology and comprehensively considering the overall situation, the entire process of the model, from preprocessing to network training and chatter identification, is obtained. Taking full advantage of the superiority of CNN in local feature extraction and the ability of GRU to extract the temporal characteristics of time series, a CNN-GRU monitoring model is constructed. Meanwhile, innovatively, the MGP algorithm, that is, the genetic algorithm (GA)-particle swarm optimisation (PSO) fusion algorithm, is used to optimise the hyperparameters of the DCNN-GRU neural network model, avoiding the limitations of the traditional method of setting hyperparameter values based on experience. Finally, the optimal values of the model hyperparameters are obtained. As a result, the accuracy of the proposed model in monitoring chatter reaches 98.8 %.

2.2 Improved VMD-wavelet denoising preprocessing

2.2.1 Principle of the improved VMD method

In VMD, the selection of the decomposition level k and the penalty factor α significantly influences the quality of the resulting mode decomposition. In this study, a tailored parameter selection strategy is adopted to determine the optimal values of k and α .

(i) Selection of decomposition level k .

This study proposes a criterion for selecting the optimal decomposition level k based on the difference in correlation coefficients between adjacent intrinsic mode function (IMF) components. The detailed procedure for determining k is as follows:

- a. Initialise k : set the initial value of $k = 2$.
- b. Apply VMD: perform VMD on the original signal to obtain k IMFs.
- c. Compute correlation coefficients: calculate the correlation coefficient k between each pair of IMF components using Eq. (1).

$$\rho_k = \frac{\sum_{t=0}^{\infty} \sum_{k=1}^K u_k(t) x(t)}{\sqrt{\sum_{t=0}^{\infty} \sum_{k=1}^K u_k^2(t) x^2(t)}} \quad (1)$$

Here, $u_k(t)$ denotes the k th IMF component obtained from the VMD decomposition.

- d. The difference in correlation coefficients δ between two adjacent IMF components is calculated as follows:

$$\delta_k = |\rho_k - \rho_{k+1}|. \quad (2)$$

- e. Compare the difference in correlation coefficients δ_k with the decision threshold θ (where $\theta = 0.1$). If $\delta_k < \theta$, it indicates over-decomposition, and K is decremented by 1 until the criterion is satisfied. Conversely, if $\delta_k > \theta$, K is incrementally increased by 1 until over-decomposition occurs, at which point the value of $K - 1$ is output.

The polish-grinding force signal is selected for VMD decomposition preprocessing. Starting with $K = 2$, the value of K is gradually increased. Based on extensive experimental results, it is found that setting $K = 4$ yields the optimal outcome for the VMD decomposition parameter.

(ii) Selection of the penalty factor α .

This study employs a method based on fuzzy entropy to select the penalty factor (α), effectively enhancing the impact components in the chatter signal. By optimising (α), the fuzzy entropy (FE) value of the VMD reconstructed signal is minimised, thereby improving the signal quality and making the decomposition more effective in capturing the essential features of the chatter signal. The definition of FE is as follows:

$$FE(m, n, r) = [\ln \varphi^m(n, r) \ln \varphi^{m+1}(n, r)]. \quad (3)$$

When N is finite, the fuzzy entropy $FE(mnr)$ can be expressed as

$$FE(m, n, r, N) = \ln \varphi^m(n, r) \ln \varphi^{m+1}(n, r). \quad (4)$$

Based on the above formula for signal analysis, when the penalty factor $\alpha = 2300$, the minimum value of the signal's fuzzy entropy gradually stabilises. This indicates the moment when the signal contains the most significant impact components and exhibits the highest similarity. Therefore, the signal is decomposed using $K = 4$ and $\alpha = 2300$ as the optimal parameter settings for VMD decomposition.

2.2.2 Basic principle of wavelet transform

Due to its localisation properties, wavelet transform can simultaneously analyse a signal in both the time and the frequency domains, effectively capturing local features of the signal. The continuous wavelet transform (CWT) expression for a function $f(t)$ is given by

$$W_\varphi f(a, b) = \langle f(x), \varphi_{a,b}(x) \rangle = |a|^{-\frac{1}{2}} \int f(x) \varphi\left(\frac{x-b}{a}\right) dx \quad f \in L^2(R), \quad (5)$$

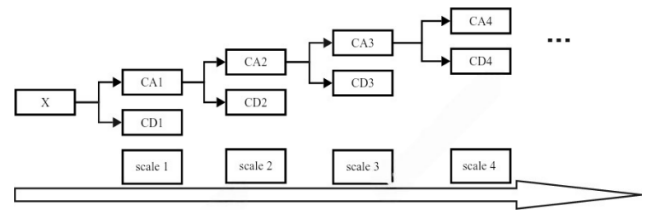


Figure 2. Wavelet transform structure.

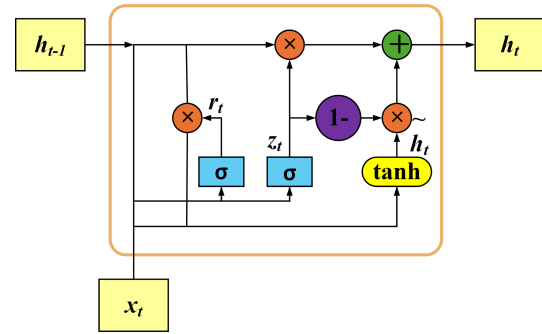


Figure 3. GRU network basic unit structure.

$$\varphi_{a,b}(x) = |a|^{-\frac{1}{2}} \varphi\left(\frac{x-b}{a}\right). \quad (6)$$

In the expression, a represents the scale factor and b represents the translation factor. The projection of the function $f(x)$ onto the function $\varphi_{a,b}(x)$ is given by the wavelet transform $W_\varphi f(a, b)$, which can be expressed as

$$f(x) = \frac{\int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} (W_\varphi f)(a, b) \varphi_{a,b}(x) db \right] \frac{da}{a^2}}{2 \int_0^\infty \frac{|\hat{\varphi}(\omega)|^2}{\omega} d\omega}. \quad (7)$$

Figure 2 illustrates the structure of the wavelet transform, where CA represents the high-frequency coefficients and CD represents the low-frequency coefficients.

3 Attention-enhanced DCNN–GRU model for vibration monitoring

3.1 GRU neural network

GRU, a variant of RNN, was proposed by Chung et al. (2014). Utilising a gating mechanism, it effectively retains long-term temporal information. Compared to LSTM, which uses three gates – input, output, and forget gates – GRU simplifies the architecture by employing only two gates: the update gate and the reset gate. This simplification reduces model complexity, improves computational efficiency, and minimises potential redundancy between gates. The basic unit structure of the GRU network is shown in Fig. 3.

As shown in Fig. 3, (z_t) denotes the update gate and (r_t) represents the reset gate. The specific calculation formulas are as follows:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z), \quad (8)$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r). \quad (9)$$

In these equations, σ denotes the sigmoid activation function. W_z , W_r , U_z , and U_r represent the corresponding weight matrices, while b_z and b_r denote the respective bias vectors. x_t is the input at the current time step, and h_{t-1} is the hidden state from the previous time step.

As can be seen from the above formulations, both the update gate and the reset gate perform parameter transformations on (x_t) and (h_{t-1}) , followed by the sigmoid activation to constrain the outputs within the range $[0, 1]$. This mechanism enables fine-grained control over the flow of information.

After obtaining the two gating signals, the reset gate is used to compute the candidate hidden state $(\tilde{h}_t)$, which is then employed to calculate the current hidden state (h_t) . The specific formulas are as follows:

$$\tilde{h}_t = \tanh(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h), \quad (10)$$

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \tilde{h}_t. \quad (11)$$

In these equations, $\tanh$ denotes the hyperbolic tangent function. W_h and U_h represent the corresponding weight matrices, and b_h is the associated bias vector. The symbol $\odot$ denotes the element-wise (Hadamard) product.

When $z_t = 0$, the current hidden state h_t exhibits a non-linear relationship with the previous hidden state h_{t-1} ; when $z_t = 1$, the relationship becomes linear. Similarly, when $r_t = 0$, the candidate state $\tilde{h}_t$ depends solely on the current input x_t , excluding any influence from past hidden states. In contrast, when $r_t = 1$, the computation of $\tilde{h}_t$ mirrors that of a conventional RNN.

3.2 DCNN

This study innovatively integrates the inception module into the CNN architecture. The inception module performs multiple convolution and pooling operations in parallel and concatenates all outputs to enable the model to autonomously learn features at multiple scales. As illustrated in Fig. 4, the CNN architecture primarily consists of an input layer, convolutional layers, pooling layers, fully connected layers, and an output layer.

As a core component of CNNs, the convolutional layer performs feature extraction through convolution operations. The specific computation is given by the following formula:

$$x_j^l = f\left(\sum_{i \in M_j}^{l-1} x_i^{l-1} * k_{ij}^l + b_j^l\right). \quad (12)$$

In this equation, x_j^l denotes the j th feature map in the l th layer and $f(\cdot)$ represents the nonlinear activation function. M_j is the set of input feature maps contributing to the

j th output, and x_i^{l-1} denotes the i th feature map from the $(l-1)$ th layer. k_{ij}^l is the convolution kernel applied between the i th input and the j th output feature map, $*$ denotes the convolution operation, and b_j^l is the bias term.

Furthermore, this study introduces a deeper optimisation of the inception module by incorporating depth-wise separable convolution. Traditional 3×3 and 5×5 kernels are decomposed into multiple smaller kernels such as 1×3 , 3×1 , 1×5 , and 5×1 , significantly accelerating the training process. The detailed architecture of the optimised inception module is illustrated in Fig. 5.

3.3 SE attention mechanism module

In the field of deep learning, the attention mechanism is an extremely effective tool. The SE attention mechanism, proposed by squeeze-and-excitation networks (SE networks), is illustrated in Fig. 6.

Final output of the SE module $\tilde{x}_c$ is calculated as follows:

$$\tilde{x}_c = F_{\text{scale}}(u_c, s_c) = s_c u_c. \quad (13)$$

3.4 Hyperparameter optimisation for vibration monitoring model

In traditional neural network models, hyperparameter values are often selected based on the experience of researchers, which can lead to model uncertainty and negatively impact vibration monitoring performance. Optimisation algorithms can effectively select the optimal hyperparameter values, reducing model redundancy and improving operational efficiency. Among the commonly used optimisation algorithms, genetic algorithms (GAs) and particle swarm optimisation (PSO) have shown superiority in handling time series, though each has its own strengths and weaknesses. This study innovatively applies the MGP algorithm, a GA-PSO hybrid approach, to optimise the hyperparameters of the DCNN-GRU neural network model.

3.4.1 MGP algorithm for neural network optimisation

To address the issues of premature convergence and local minima in the PSO algorithm, this study integrates the GA algorithm with PSO to construct the MGP optimisation method. By leveraging the strengths of both algorithms, the hyperparameters of the neural network are optimised. During PSO execution, GA's selection and mutation operations are incorporated. The global search capability of the PSO algorithm is enhanced by integrating the selection and mutation mechanisms of GA. Specifically, introducing the selection and mutation operations of GA into the PSO iteration process can optimise the movement trajectory of particles while maintaining population diversity, thereby effectively improving the accuracy and stability of the machine tool vibration monitoring model.

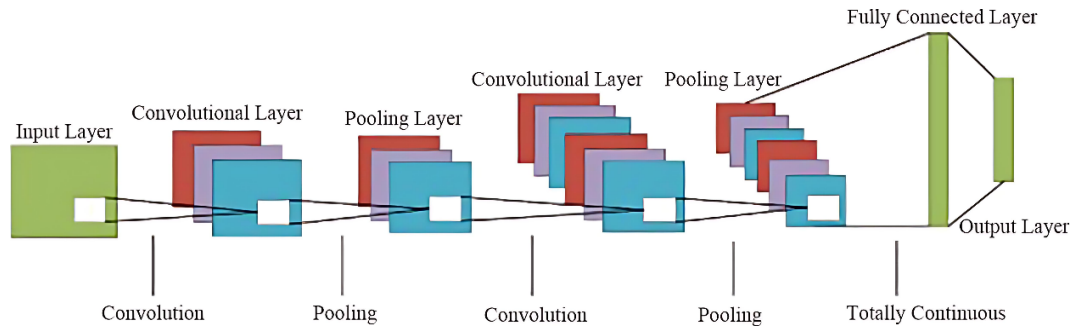


Figure 4. Basic structure of convolutional neural network.

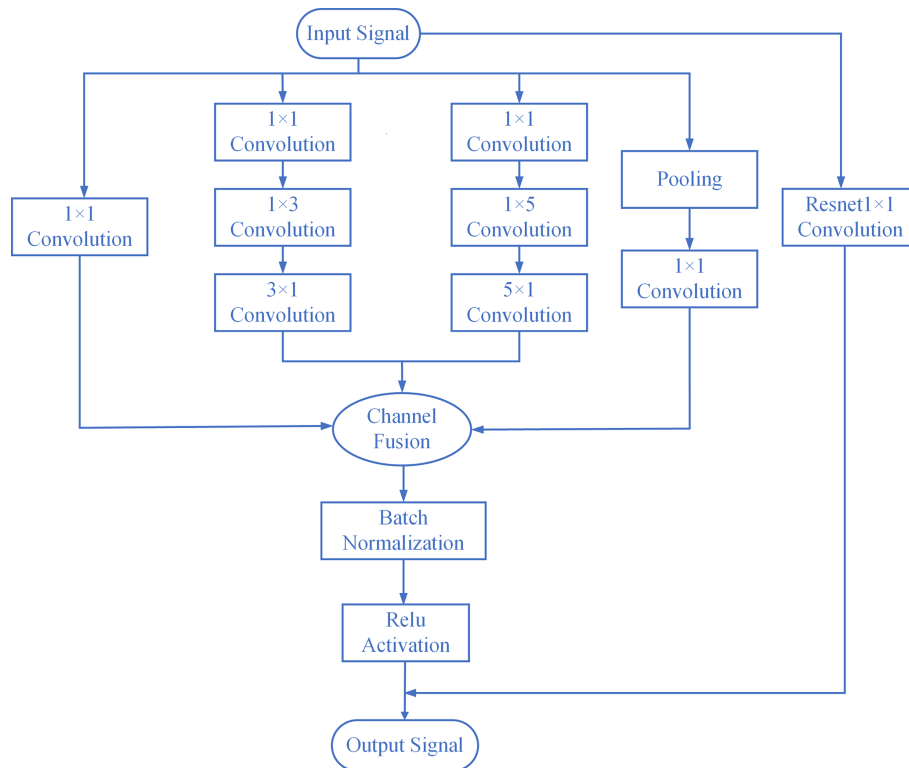


Figure 5. Illustration of depth-wise separable convolution.

The optimisation process begins by defining the hyperparameters to be tuned, specifically the learning rate, batch size, and number of iterations, with other parameters pre-sets. The MGP algorithm is then applied to optimise these hyperparameters, leading to the identification of the optimal network model. First, initialise the particle swarm containing hyperparameters such as learning rate, batch size, and iteration count. Each particle corresponds to a set of hyperparameter combinations. During iteration, first execute the PSO velocity and position update mechanism; i.e. particles adjust their search direction based on their own historical best solution (pbest) and global best solution (gbest). Next, introduce the selection and mutation operators from GA to optimise the population: sort the particle swarm by fitness, retain the top

50 % of high-quality particles to maintain their superior characteristics, and use the roulette wheel selection operator to screen out individuals for optimisation from the bottom 50 % of particles. Then perform velocity and position mutations on these individuals, introducing random perturbations to avoid getting stuck in local optima. Finally, the fitness of particles before and after mutation is compared, and the best ones are selected to form a new population, achieving the synergistic effect of PSO's rapid convergence and GA's maintenance of population diversity.

The overall workflow of the MGP-based network optimisation is shown in Fig. 7.

This paper employs the MGP algorithm to optimise the hyperparameters of the DCNN-GRU neural network, improv-

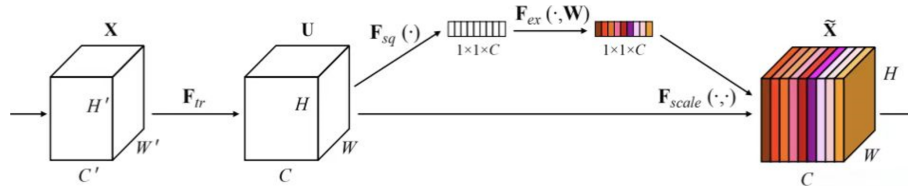


Figure 6. SE attention mechanism module. X: input feature map; U: output feature map; specifications: $C' \times W' \times H'$.

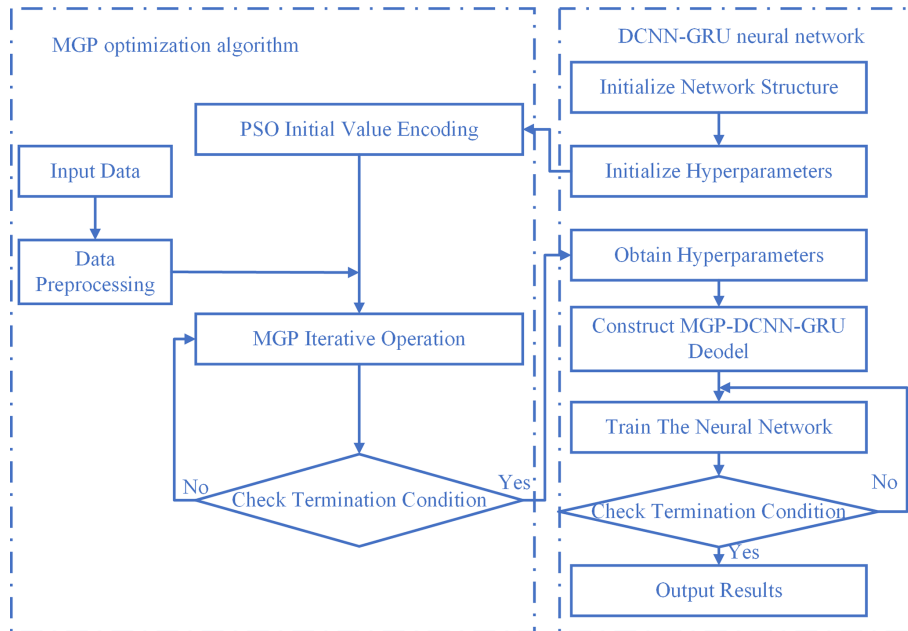


Figure 7. MGP-DCNN-GRU algorithm flowchart.

ing convergence speed, overcoming premature convergence, and enhancing network performance. The optimal solution from the MGP algorithm defines the hyperparameters to be optimised within the DCNN-GRU network. These hyperparameters are then applied to the DCNN-GRU model for training and validation, resulting in chatter monitoring accuracy and the loss function curve.

The fitness function, designed based on the optimisation problem, evaluates particle quality during the iteration process. The MGP-DCNN-GRU model is used to monitor chatter in machine tool machining, with the mean squared error (MSE) chosen as the fitness function, as defined by the following formula:

$$E(y) = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2. \quad (14)$$

Here, n denotes the number of training samples, $\hat{y}_i$ represents the monitoring output of the i th instance, and y_i is the corresponding expected value (i.e. the ground truth).

The proposed MGP algorithm employs two genetic operators, the selection operator and the mutation operator, as detailed below.

1. Selection operator: this study adopts the roulette wheel method to perform the selection operation. The selection probability p_s is defined by the following formula:

$$p_s = \frac{1/E(y)}{\sum_1^N (1/E(y))}. \quad (15)$$

Here, $E(y)$ denotes the fitness value of an individual in the population.

2. Mutation operator: in the MGP algorithm, mutation operations are applied to both the velocity and the position of particles. The specific mutation formulas are as follows.

Velocity mutation operation.

$$v_i(k+1) = \begin{cases} v_i(k) + (v_i(k) - v_{\max}) \times f(g) & r_1 \geq 0.5 \\ v_i(k) + (v_{\min} - v_i(k)) \times f(g) & r_1 < 0.5 \end{cases} \quad (16)$$

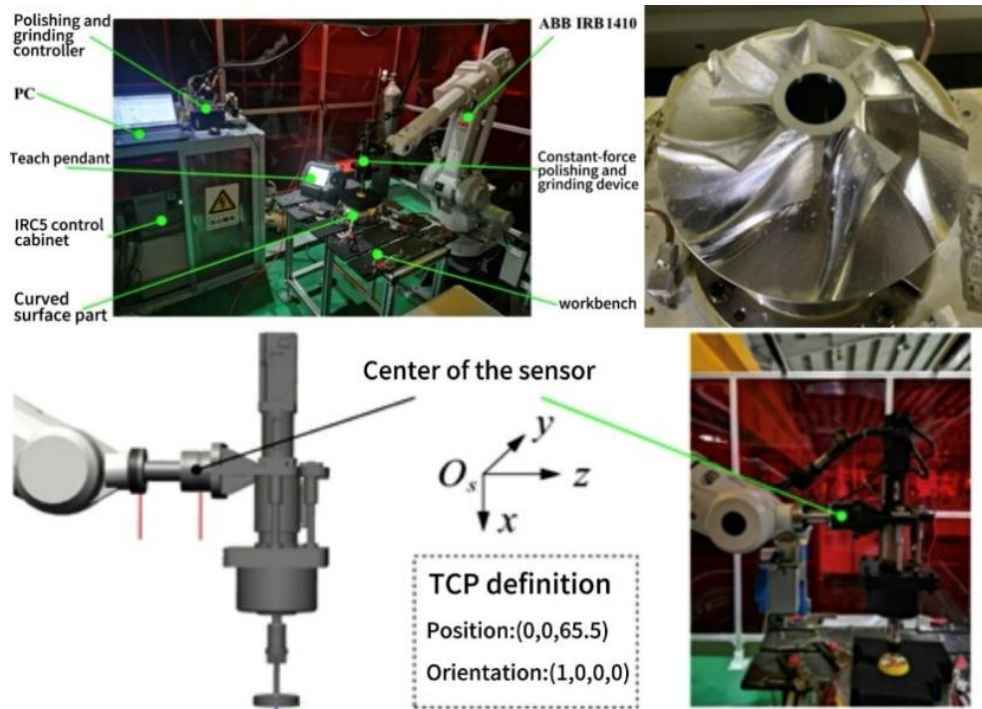


Figure 8. Single-degree-of-freedom constant-force polishing and grinding experimental platform.

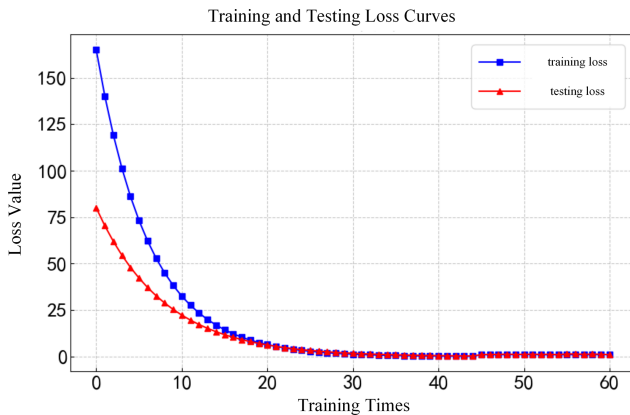


Figure 9. Loss function change plot.

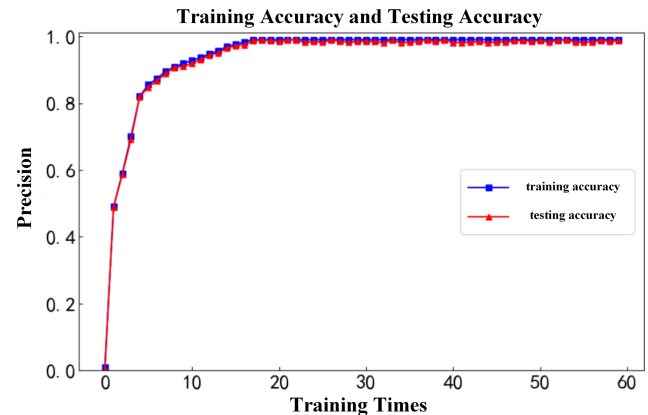


Figure 10. Accuracy change plot.

Position mutation operation.

$$x_i(k+1) = \begin{cases} x_i(k) + (x_i(k) - x_{\max}) \times f(g) & r_2 \geq 0.5 \\ x_i(k) + (x_{\min} - x_i(k)) \times f(g) & r_2 < 0.5 \end{cases} \quad (17)$$

$$f(g) = r_3(1 - g/P_{\max}) \quad (18)$$

Here, $v_{\max}$ and $v_{\min}$ represent the upper and lower bounds of particle velocity, respectively, and $x_{\max}$ and $x_{\min}$ denote the upper and lower bounds of particle position. The variables, r_1 , r_2 , and r_3 are random values uniformly distributed in the range $[0, 1]$, and g denotes the current iteration number.

The parameter design in particle swarm optimisation (PSO) plays a critical role, particularly concerning the inertia weight and maximum velocity, as the effectiveness of these parameters directly influences the algorithm's search capability. In this study, the particle velocity is constrained as follows: if a particle's velocity $v_i > v_{\max}$, then v_i is set to $v_{\max}$; if $v_i < -v_{\max}$, then v_i is set to $-v_{\max}$. This velocity regulation allows particles to maintain a high-speed movement state when exceeding the threshold, which facilitates further exploration based on already promising solutions. As a result, the likelihood of the PSO algorithm identifying the global optimum is significantly enhanced.

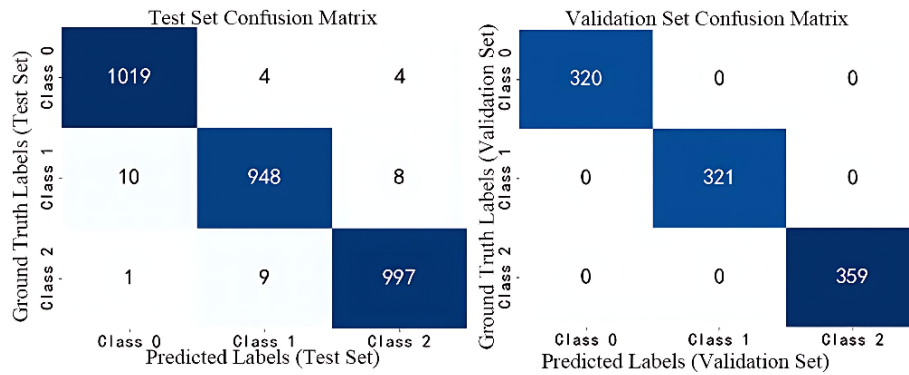


Figure 11. Confusion matrices for the validation and test sets.

4 Experimental results and analysis

4.1 Experimental design for data collection

To facilitate subsequent data preprocessing and network training, this paper conducts milling experiments under different dry cutting conditions using an industrial robot with ABB IRB1410 and a single-degree-of-freedom constant-force grinding and polishing device as the end effector. The industrial robot control system is the IRC5 body controller, and the end effector is controlled by a Siemens PLC. The controller includes the IRC5 body controller of the mechanical arm, and the grinding tool used is a diamond-coated grinding head. The experimental workpiece is a titanium alloy TC4 impeller with a size of $\varnothing 200 \text{ mm} \times 50 \text{ mm}$.

To thoroughly validate the performance of the proposed MGP-SE-DCNN-GRU model, the accuracy and loss changes for both the training and the test sets were recorded during the training process. After optimisation of the hyper-parameters of the SE-DCNN-GRU model using the MGP algorithm, the final selected parameters were a learning rate of 0.001, 60 iterations, and a batch size of 64, as shown in Table 1. The specific results for vibration monitoring accuracy and loss are depicted in Figs. 9. and 10.

The experimental results demonstrate that the MGP-SE-DCNN-GRU model initially exhibits relatively low accuracy and high loss. As training progresses, by approximately the 20th epoch the model's performance stabilises, achieving a peak accuracy of 98.8 % with negligible subsequent variation and reducing the loss to around 1 %, thereby validating the stability and convergence of the proposed network architecture. To visually illustrate classification performance, confusion matrices for both the validation and the test sets are presented in Fig. 11.

The experimental results reveal that the proposed model achieved 100 % correct classification of chatter types on the validation set. Furthermore, it was able to accurately distinguish chatter modes on the test set to a very high degree. These findings unequivocally demonstrate the model's excellent accuracy on this dataset.

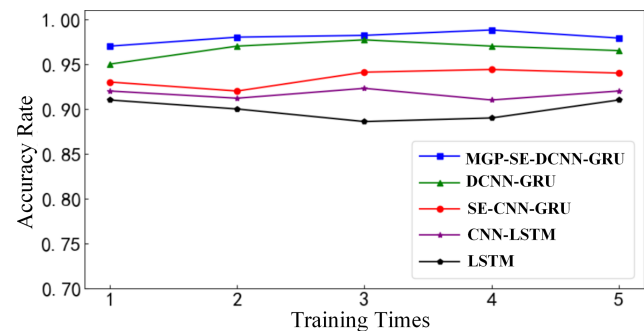


Figure 12. Comparison of different models.

4.2 Comparative experiments

In this study, a series of comparative experiments were conducted on the dataset to benchmark the proposed model against several representative methods for polish-grinding force chatter monitoring. The models selected for comparison were the GA-CNN-GRU network, the PSO-CNN-LSTM network, the SE-CNN-GRU network, and a conventional CNN-GRU network. Chatter vibration data were input into each model, and performance was evaluated over five independent trials. The comparative results are shown in Fig. 12.

As illustrated in Fig. 13, the MGP-SE-DCNN-GRU model achieves markedly higher accuracy and stability, demonstrating exceptional classification performance on the target dataset and significantly outperforming the other compared networks.

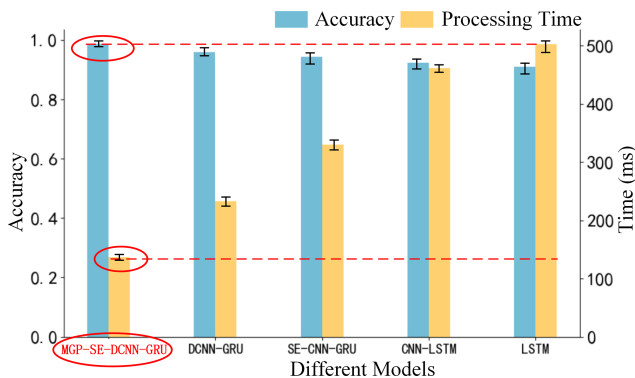
The results in Fig. 13 further demonstrate that the proposed network architecture not only achieves the highest classification accuracy (98.8 %) but also converges significantly faster during training than the compared models, highlighting its overall advantages in feature learning efficiency and parameter optimisation.

Table 1. Structural mass statistics of the polishing and grinding device.

Serial number	Name	Mass (g)	Quantity (pcs)
1	SGM7J – C2AFC6S Servo Motor	500	1
2	TMEC0050 – 012 – 00Z Voice Coil Motor	645	1
3	YuLi M3813D Six – axis Force Sensor	90	1
4	Hiemosen IKS9.1 Magnetic Scale Read Head	75	1

Table 2. Network hyperparameters.

Parameter	Value
Learning rate	0.001
Batch size	64
Optimiser	Adam
Epochs	60
Loss function	MES
Activation function	ReLU

**Figure 13.** Comparison of classification accuracy and inference time for different models.

5 Conclusions

To enable effective chatter monitoring, this study proposes a novel method based on a deep separable convolutional neural network (DCNN) integrated with a gated recurrent unit (GRU) network. The approach begins with data pre-processing of the acquired signals, followed by precise determination of the optimal number of VMD decomposition layers and the corresponding penalty factor. Subsequently, a wavelet thresholding technique is applied for denoising and signal reconstruction. A SE-DCNN-GRU chatter monitoring model is then constructed, achieving a classification accuracy of 98.8 %.

1. The inception module is enhanced by introducing depth-wise separable convolutions, decomposing traditional 3×3 and 5×5 kernels into smaller kernels such as 1×3 , 3×1 , 1×5 , and 5×1 . This reduces the total number of model parameters while enabling multi-

scale feature extraction, improving training speed and mitigating gradient vanishing and explosion. Batch normalisation is also incorporated to normalise the input of each layer, enhancing robustness and generalisation.

2. The model further integrates the squeeze-and-excitation (SE) attention mechanism to improve feature extraction, while ResNet modules are introduced to enable feature reuse of chatter characteristics. The final classification is obtained through a softmax function.
3. A multi-strategy global optimisation algorithm (MGP) is proposed to optimise the hyperparameters of the SE-DCNN-GRU network, leading to a well-tuned model. The effectiveness of the model is verified through loss function evaluation and accuracy curve analysis. Comprehensive comparisons with alternative models and algorithms highlight the superior accuracy and fast convergence of the proposed DCNN-GRU model in chatter detection based on polish-grinding force signals, confirming its overall advantages.

Data availability. This paper conducts milling experiments under different dry cutting conditions using an industrial robot with ABB IRB1410 and a single-degree-of-freedom constant force grinding and polishing device as the end effector. The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Author contributions. YZ and CZ conceived the idea. YZ and GX performed all the experiments. YZ drafted the manuscript, and CZ interpreted, discussed, and edited the manuscript. GX finalised the manuscript, including preparing the detailed response letter. YZ supervised the work.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibil-

ity lies with the authors. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Acknowledgements. This work was supported by the National Natural Science Foundation of China (52475115), Project of the Basic Scientific Research Program for Higher Education Institutions of the Liaoning Provincial Department of Education (LJ212410153037), and Project of the Liaoning Province Research Foundation for Applied Basic Research, China (2023JH2/101300222).

Financial support. This work was supported by the National Natural Science Foundation of China (grant no. 52475115), Project of the Basic Scientific Research Program for Higher Education Institutions of the Liaoning Provincial Department of Education (grant no. LJ212410153037) and Project of the Liaoning Province Research Foundation for Applied Basic Research, China (grant no. 2023JH2/101300222).

Review statement. This paper was edited by Pengyuan Zhao and reviewed by two anonymous referees.

References

- Arnold, N. R.: The mechanism of tool vibration in the cutting of steel, *ARCHIVE Proceedings of the Institution of Mechanical Engineers* 1847–1982, 154, 261–284, 1946.
- Altintas, Y.: Analytical Prediction of Three Dimensional Chatter Stability in Milling, *JSME International Journal Series C Mechanical Systems, Machine Elements and Manufacturing*, 44, 717–723, <https://doi.org/10.1299/jsmec.44.717>, 2002.
- Altıntaş, Y. and Budak, E.: Analytical Prediction of Stability Lobes in Milling, *CIRP Annals*, 44, 357–362, [https://doi.org/10.1016/S0007-8506\(07\)62342-7](https://doi.org/10.1016/S0007-8506(07)62342-7), 1995.
- Altintas, Y. and Weck, M.: Chatter Stability of Metal Cutting and Grinding, *CIRP Annals*, 53, Arnold and N. R.: The mechanism of tool vibration in the cutting of steel, *ARCHIVE Proceeding*, 619–642, [https://doi.org/10.1016/S0007-8506\(07\)60032-8](https://doi.org/10.1016/S0007-8506(07)60032-8), 2004.
- Cao, H. R., Lei, Y. G., and He, Z. J.: Chatter identification in end milling process using wavelet packets and Hilbert-Huang transform, *International Journal of Machine Tools & Manufacture*, 69, 11–19, <https://doi.org/10.1016/j.ijmachtools.2013.02.007>, 2013.
- Cao, H. R., Yue, Y. T., Chen, X. F., and Zhang, X. W.: Chatter detection in milling process based on synchrosqueezing transform of sound signals, *International Journal of Advanced Manufacturing Technology*, 89, 2747–2755, <https://doi.org/10.1007/s00170-016-9660-7>, 2017.
- Cherukuri, H., Perez-Bernabeu, E., Selles, M. A., and Schmitz, T. L.: A neural network approach for chatter prediction in turning, *Procedia Manufacturing*, 34, 885–892, 2019.
- Chung, J., Gulcehre, C., Cho, K. H., and Bengio, Y.: Empirical evaluation of gated recurrent neural networks on sequence modeling, *arXiv [preprint]*, arXiv:1412.3555, <https://doi.org/10.48550/arXiv.1412.3555>, 2014.
- Feng, J. L., Sun, Z. L., Jiang, Z. H., and Yang, L.: Identification of chatter in milling of Ti-6Al-4V titanium alloy thin-walled workpieces based on cutting force signals and surface topography, *International Journal of Advanced Manufacturing Technology*, 82, 1909–1920, <https://doi.org/10.1007/s00170-015-7509-0>, 2016.
- Gao, H. N., Shen, D. H., Yu, L., and Zhang, W. C.: Identification of cutting chatter through deep learning and classification, *International Journal of Simulation Modelling*, 19, 667–677, <https://doi.org/10.2507/ijssimm19-4-co16>, 2020.
- Han, Z. Y., Zhuo, Y., Yan, Y. Z., Jin, H. Y., and Fu, H. Y.: Chatter detection in milling of thin-walled parts using multi-channel feature fusion and temporal attention-based network, *Mechanical Systems and Signal Processing*, 179, <https://doi.org/10.1016/j.ymssp.2022.109367>, 2022.
- Kuljanic, E., Sortino, M., and Totis, G.: Multisensor approaches for chatter detection in milling, *Journal of Sound and Vibration*, 312, 672–693, <https://doi.org/10.1016/j.jsv.2007.11.006>, 2008.
- Li, X. Y., Liu, R. L., and Zhu, Z. Y.: Data Acquisition and Chatter Recognition Based on Multi-Sensor Signals for Blade Whirling Milling, *Machines*, 13, <https://doi.org/10.3390/machines13030206>, 2025.
- Liu, C. F., Zhu, L. D., and Ni, C. B.: The chatter identification in end milling based on combining EMD and WPD, *International Journal of Advanced Manufacturing Technology*, 91, 3339–3348, <https://doi.org/10.1007/s00170-017-0024-8>, 2017.
- Liu, R. Y., Liu, L. Y., Wang, X. Z., Huang, L., and Wang, Z. H.: Online milling chatter detection based on signal correlation and optimized variational mode decomposition, *Measurement*, 244, <https://doi.org/10.1016/j.measurement.2024.116530>, 2025.
- Liu, X. J., Wang, D. J., and Chen, Y. S.: Approximate analytical solution of the self-excited vibration of piecewise-smooth systems induced by dry friction, *Acta Mech. Sin.*, 14, 78–84, 1998.
- Tariq, S. A., Zia, T., and Ghafoor, M.: Towards counterfactual and contrastive explainability and transparency of DCNN image classifiers, *Knowledge-Based Systems*, 257, <https://doi.org/10.1016/j.knosys.2022.109901>, 2022.
- Trusty, J. and Ismail, F.: Basic Non-Linearity in Machining Chatter, *CIRP Annals*, 30, 299–304, [https://doi.org/10.1016/S0007-8506\(07\)60946-9](https://doi.org/10.1016/S0007-8506(07)60946-9), 1981.
- Yang, K., Wang, G. F., Dong, Y., Zhang, Q. B., and Sang, L. L.: Early chatter identification based on an optimized variational mode decomposition, *Mechanical Systems and Signal Processing*, 115, 238–254, <https://doi.org/10.1016/j.ymssp.2018.05.052>, 2019.
- Zacharia, K. and Krishnakumar, P.: Chatter Prediction in High Speed Machining of Titanium Alloy (Ti-6Al-4V) using Machine Learning Techniques, *Materials Today: Proceedings*, 24, 9, 2020.
- Zhang, P. F., Gao, D., Hong, D. B., Lu, Y., Wu, Q., Zan, S. S., and Liao, Z. R.: Improving generalisation and accuracy of on-line milling chatter detection via a novel hybrid deep convolutional neural network, *Mechanical Systems and Signal Processing*, 193, <https://doi.org/10.1016/j.ymssp.2023.110241>, 2023.
- Zhou, L., Zhang, T. Y., Zhang, Z. D., Lei, Z. L., and Zhu, S. L.: Monitoring of resistance spot welding expulsion based on machine learning, *Science and Technology of Welding and Joining*, 27, 292–300, <https://doi.org/10.1080/13621718.2022.2051408>, 2022.
- Zhu, H., Chen, R., and Li, R.: Wavelet Neural Network - Based Research on Online Wearing Prediction of Ti6Al4V Cutter in

- High Speed Milling, *Key Engineering Materials*, 431–432, 205–208, 2010.
- Zhu, L. D., Liu, C. F., Ju, C. Y., and Guo, M. X.: Vibration recognition for peripheral milling thin-walled workpieces using sample entropy and energy entropy, *International Journal of Advanced Manufacturing Technology*, 108, 3251–3266, <https://doi.org/10.1007/s00170-020-05476-7>, 2020.
- Zhou, Z. X., Wang, H., Li, Z. X., and Chen, W.: Fault diagnosis of rolling bearing based on deep convolutional neural network and gated recurrent unit, *Journal of Advanced Mechanical Design Systems and Manufacturing*, 17, <https://doi.org/10.1299/jamdsm.2023jamdsm0017>, 2023.