MM Science Journal | www.mmscience.eu
ISSN 1803-1269 (Print) | ISSN 1805-0476 (On-line)
Special Issue | HSM 2025
18th International Conference on High Speed Machining
October 15-16, 2025, Metz, France
DOI: 10.17973/MMSJ.2025_12_2025162

HSM2025-44901

RESEARCH ON CHATTER DETECTION FOR THIN-WALLED WORKPIECE MACHINING BASED ON META-REINFORCEMENT AND HYBRID DEEP CONVOLUTIONAL NEURAL NETWORK

Yongliang LU[1,2,3], Jun Zhao[1,2,3*], Xujie Tang[1,2,3], Shihua Zhang[1,2,3]

[1]Key Laboratory of High Efficiency and Clean Mechanical Manufacture of MOE, School of Mechanical Engineering, Shandong University, Jinan 250061, China

[2]National Key Laboratory of High-end Equipment and Advanced Technology for Metal Forming, School of Mechanical Engineering, Shandong University, Jinan 250061, China

[3]National Demonstration Center for Experimental Mechanical Engineering Education, Shandong University, Jinan 250061, China

*Corresponding author; e-mail: zhaojun@sdu.edu.cn

Abstract

Cutting chatter is a common phenomenon during metal cutting operations, which can severely impair the surface integrity of machined components and reduce overall production efficiency. Therefore, the precise detection of chatter during the cutting process is of critical importance. In recent years, the widespread integration of artificial intelligence (AI) techniques into chatter detection has led to satisfactory results. Although deep learning methods exhibit excellent capabilities in feature learning and classification, their generalization and accuracy are highly dependent on data labeling quality and training procedures. To address these issues, this study introduces a chatter detection approach for thin-walled parts milling based on the hybrid deep convolutional neural network (HDCNN) and names it as Chatter-CNN. The Chatter-CNN model combines two Inception-Chatter modules and two Squeeze-and-Excitation ResNet blocks (SR-blocks). The Inception-Chatter module is capable of automatically extracting multi-scale features from cutting force signals. The SR-block mitigates the risk of gradient vanishing, accelerates network convergence, and adaptively assigns weights to different feature channels. The proposed method is benchmarked against two convolutional neural networks (CNNs) that have been extensively applied in the field of chatter detection over the past decade. A series of milling experiments on wedge-shaped thin-walled workpieces are conducted under various cutting conditions. The experimental results indicate that the proposed Chatter-CNN model achieved classification accuracies of 100% on the validation set and 97.5% on the test set, outperforming existing methods.

Keywords: Chatter detection, Artificial intelligence (AI), Deep learning, HDCNN

1 INTRODUCTION

Milling is among the most extensively employed machining processes in the manufacturing of thin-walled parts. However, the inherent characteristics of thin-walled parts, including low damping, inadequate stiffness, and time-varying dynamic parameters, make them susceptible to chatter during machining [Zhu 2020]. Chatter not only generates harmful noise but also accelerates tool wear, degrades cutting surface quality, and may result in workpiece scrapping or machine tool damage, ultimately diminishing the economic efficiency of production [Navarro-Devia 2023]. With the rapid advancement of computing power over the past few years, artificial intelligence (AI) technologies, particularly deep learning, have been widely applied in vibration detection and have yielded promising results [Jeon 2020]. The implementation of effective chatter detection strategies can significantly improve both production efficiency and machining quality.

Current research on chatter detection is generally categorized into two main approaches: offline chatter prediction and online chatter detection [Yuan 2018]. Offline chatter prediction is a traditional method that optimizes cutting parameters to avoid chatter by utilizing the stability lobe diagram (SLD) [Cordes 2019]. Offline chatter prediction methods have inherent limitations because SLD relies on the frequency response function (FRF) of the interaction between the cutting tool holder, tool, workpiece material and machine tool. Since the FRF often varies during the cutting process and difficult to measure accurately, it becomes extremely difficult to accurately obtain SLD [Postel 2018]. Online chatter detection utilizes various sensors to gather data generated during the cutting process, enabling real-time identification of the machining status. Compared to offline chatter prediction methods, online chatter detection methods directly determine the machining status from sensor signals without requiring manual intervention, thereby enhancing interference resistance [Chu 2022]. If chatter can be accurately identified and effectively suppressed during the cutting process, its occurrence can be minimized while maintaining production efficiency [Bahtiyar 2022].

Chatter detection can be formulated as a classification problem, and online detection can be effectively achieved through the integration of machine learning and signal processing techniques. In such methods, features are typically extracted and processed from the time domain, frequency domain, and time–frequency domain, followed by classification using machine learning algorithms. Commonly used traditional machine learning approaches include artificial neural networks (ANN) [Friedrich 2017], random forests (RF) [Oleaga 2018], gradient tree boosting (GTB) [Zhang 2024], and support vector machines (SVM) [Yao 2010]. However, shallow machine learning methods primarily depend on manual feature extraction and selection, which is both labor-intensive and time-consuming. Moreover, models developed using traditional chatter detection approaches often struggle to extract highly generalizable features, limiting their ability to adapt to varying operating conditions. Deep learning, as an important branch of machine learning, exhibits powerful capabilities in automatic feature extraction and classification. Although often regarded as a "black-box" approach, it is highly effective in modeling complex and nonlinear physical systems. These advantages have contributed to its widespread application in fault diagnosis and related fields [Lei 2020] [Zhang 2022]. In recent years, deep learning techniques have been increasingly applied to milling vibration detection. These methods typically rely on combining feature extraction approaches—such as eigenvalue analysis—with deep learning models. By integrating multi-sensor data and multiple feature types, deep learning significantly improves the accuracy of vibration state identification [Stavropoulos 2023] [Zhang 2023]. Unlike traditional approaches, deep learning-based vibration detection does not require manual threshold setting to determine machining states. However, the accuracy of data labeling has a substantial impact on model training and performance. Moreover, the issues of model interpretability and generalization capability remain key challenges that require further investigation.

In summary, a variety of advanced machine learning models and signal processing methods have been extensively applied to chatter detection. Traditional chatter detection methods that rely on chatter-sensitive features typically require manual threshold setting, thereby limiting their applicability across diverse machining conditions. Deep learning demonstrates exceptional performance in feature learning and classification, and its application in vibration detection is becoming increasingly prevalent. Currently, deep learning-based vibration detection methods primarily depend on annotating full-process data during modeling, yet accurately labeling data in transition states remains challenging. Moreover, the test dataset and training dataset are typically divided after shuffling the total dataset, ensuring that their corresponding processing parameters remain consistent. Although this approach achieves high accuracy, extending the model to different machining conditions remains challenging, highlighting the need to enhance and validate its generalization performance.

The main contribution of this paper is to propose a novel hybrid deep convolutional neural network (HDCNN) for online chatter detection in thin-walled part milling to overcome the challenges identified in current research. The key features of the proposed novel approach are outlined as follows:

1) To ensure the accuracy of data labeling for dataset construction, all transition-state data in the milling force signal are discarded;

2) To eliminate the need for manually extracting chatter indicators and designing thresholds, a HDCNN combining the Inception module and the SR-block module is proposed and named Chatter-CNN;

3) The training set and test set are constructed using the three-dimensional milling force data under different cutting conditions (i.e., different spindle speed n, radial cutting depth a_e and number of tool edges). To evaluate the generalization performance of the model, its chatter detection capability is benchmarked against two widely used deep learning networks (i.e., DenseNet [Huang 2017] and ResNet [He 2016]).

2 CHATTER DETECTION METHOD FOR THIN-WALLED PARTS BASED ON HDCNN

2.1 Architecture and workflow of chatter-CNN model

The model architecture of the Chatter-CNN chatter detection method proposed in this study mainly consists of three parts: data preprocessing, model preprocessing, and online milling chatter detection. The details of these three parts will be explained in detail in ***Sections 2.1.1*** to ***2.1.3***.

2.1.1 Data pre-processing

First, the three-dimensional cutting force signals collected from each milling experiment are preprocessed and labeled. Since accurately labeling transition state data is highly challenging, the proposed method discards it during dataset construction. After data selection, the cutting force signal is segmented into blocks to construct the dataset. Since frequency characteristics and energy variations are effective indicators of chatter, the original signal is transformed using the Fast Fourier Transform (FFT). The Z-score normalization method is applied to standardize input data with varying amplitudes onto a unified scale, thereby enhancing the model's convergence speed. Finally, a dataset comprising training, validation, and test subsets is constructed.

2.1.2 Model construction

The Chatter-CNN model proposed in this study for vibration detection in thin-walled parts milling primarily comprises two Inception-Chatter modules (i.e., Inception-Chatter_A and Inception-Chatter_B), a max pooling layer, two SR-blocks (i.e., SR-block_Ⅰ and SR-block_Ⅱ) an average pooling layer, and three FC layers. The Softmax classifier is employed to generate the probability distribution over all possible machining states.

2.1.3 Online detection of chatter in milling operations

The established Chatter-CNN model is trained, validated, and tested to assess its predictive accuracy and generalization capability. It should be emphasized that the validation set and test set differ in two aspects: the radial cutting depth and the number of tool edges. Next, the complete cutting force data obtained from each milling experiment is processed using the data preprocessing method illustrated in ***Section 3.1.1***, with transition state data discarded. Finally, the chatter detection model takes the constructed dataset as input to generate probability outputs and perform state classification.

2.2 Dataset preparation for the Chatter-CNN Model

Prior to constructing the Chatter-CNN model, the process of building the individual dataset is described, as illustrated in Figs. 1 and 2. Fig. 1 illustrates the design of the milling experiment and the acquisition process of three-dimensional (3D) cutting forces, while Fig. 2 presents the dataset construction procedure. In the milling experiments, the three-dimensional cutting force signal is collected and then one frame of data is extracted. In this study, each data frame consists of 250 data points, corresponding to a sampling time of 0.05s. The three-dimensional milling force signal is transformed from the time domain to the frequency domain by applying FFT and Z-score normalization, resulting in a dataset with dimensions of $1\times125\times3$.

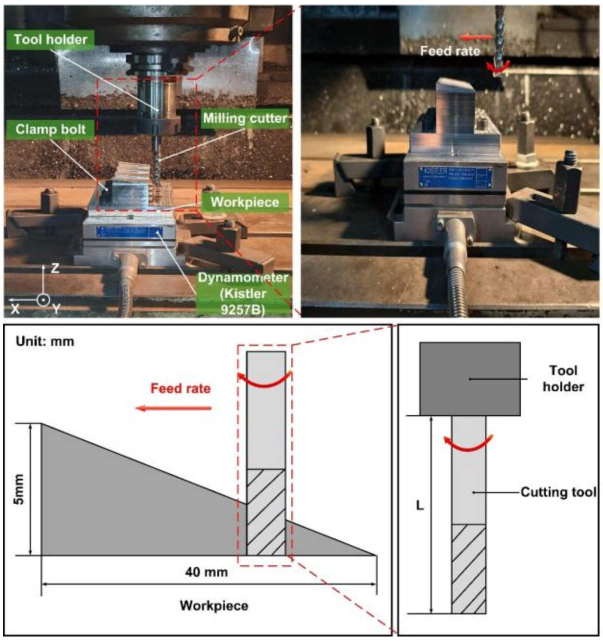

Fig. 1: Milling process design and 3D force acquisition.

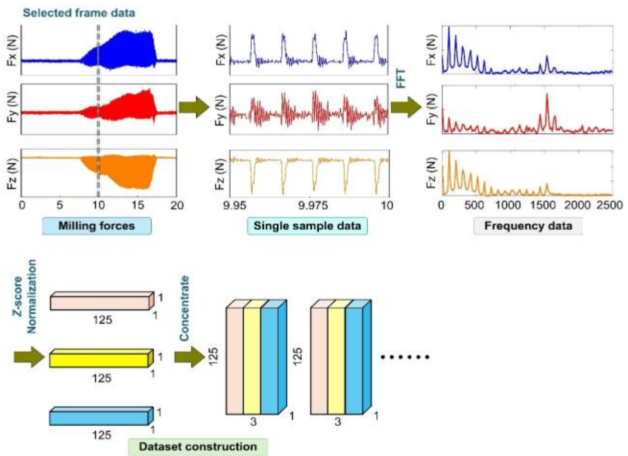

Fig. 2: Dataset construction for the proposed Chatter-CNN model.

2.3 Architecture and Parameter Configuration of the Chatter-CNN Model

Fig. 3 illustrates the architecture and parameter configuration of the Chatter-CNN model. According to Fig. 3, the parameters specify the kernel size, padding, and stride for each convolutional layer. Input data of dimensions $1\times125\times3$ is provided to the Chatter-CNN model, which consists of a series of operations including Inception-Chatter modules, max pooling, SR-blocks, average pooling, and fully connected layers. Following these operations, the model is able to classify the input into three machining states. In the established architecture, the Inception-Chatter modules (i.e., Inception-Chatter_A and Inception-Chatter_B) capture multi-scale spatial characteristics through multiple convolutions with diverse kernel sizes, thereby improving the model's accuracy and generalization capability. Within the Inception-Chatter_A module, convolution kernels (i.e., 3×1, 7×1 and 9×1) are utilized, significantly enriching the extracted features. Different features are concatenated along the channel dimension, and convolutions are performed to reduce the model's parameters, thereby lowering its complexity. Upon receiving input data from the Inception-Chatter_A module, a feature map of size $64\times125\times3$ is generated and subsequently used as input for the next module. Moreover, following the Inception-Chatter_B module, a feature tensor with shape $256\times25\times3$ is generated. The dimensionality of the input feature map is reduced by a max pooling layer, resulting in a $256\times62\times3$ feature map. The SR-block consists of a ResNet block and an SE block. Compared to conventional CNNs, ResNet modules are capable of extracting more detailed features while mitigating gradient vanishing during network training [Yan 2021]. The SE block performs adaptive feature calibration by assigning different weights to feature channels, thereby enhancing the relevant feature channels and improving chatter detection accuracy. Through the sequential application of SR-block_ Ⅰ and SR-block_Ⅱ, the extracted features are progressively enriched along the channel dimension, yielding a feature map with dimensions of $512\times31\times2$. Compared to the structure of SR-block_ Ⅰ, SR-block_Ⅱ incorporates a 1×1 convolutional layer within the skip connection to transform the input features. After feature extraction by the two SR-blocks, the resulting feature channels are concatenated to form the final feature map. Subsequently, global average pooling is conducted to capture channel-level feature information. Finally, three FC layers are employed for dimensionality reduction, transforming the feature size from $512\times1\times1$ to $3\times1\times1$, corresponding to the output of the three machining states.

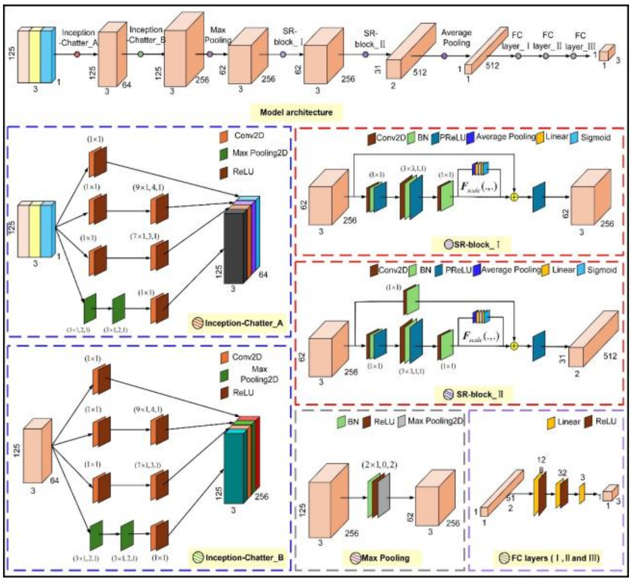

Fig. 3: Architecture and parameters setting of the proposed Chatter-CNN model.

3 DESIGN OF EXPERIMENTS AND PERFORMANCE EVALUATION OF THE MODEL

3.1 Experimental procedure and signal preprocessing

The overall experimental design process is depicted in Fig. 10. To validate the effectiveness of the proposed chatter detection method in thin-walled part milling, a milling experiment was conducted on a three-axis CNC milling center, as illustrated in Fig. 10(a). The machining center used in this study is the DAEWOO ACE-V500, equipped with a FANUC 18M CNC system. It offers precise positioning and smooth operation, ensuring the accuracy of the experiment. The dynamometer employed in the experiment is the Kistler

9257B, which is capable of simultaneously measuring the instantaneous milling force and torque in the X-, Y-, and Z-directions. Its primary components consist of a charge amplifier (Kistler, type 5080, Switzerland), a data collector (Kistler, type 5697A1, Switzerland), a dynamometer sensor (Kistler, type 9257B, Switzerland) and associated circuitry. Once the dynamometer sensor is connected to the computer, the measured data can be stored and analyzed using the supporting analysis software, Dyno Ware 3.2.2.0. To ensure result accuracy, the dynamometer and machining center are preheated for over 30 minutes prior to the experiments. As shown in Fig. 10(b), all milling experiments are conducted under dry cutting and down-milling conditions. Moreover, in accordance with the Shannon-Nyquist sampling theorem [Lu 2025], the sampling frequency f_s of this experiment is set to 5000 Hz.

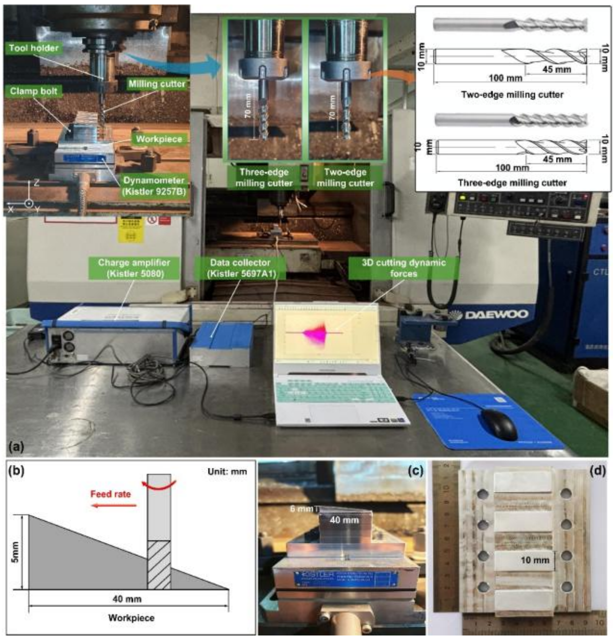

Fig. 4: The overall design process of the experiment.

To evaluate the generalization performance, training and test sets are constructed using data from various cutting conditions. Variations in cutting conditions included not only differences in cutting parameters (i.e., different spindle speed *n* and radial cutting depth a_e) but also differences in the number of flutes. The selected cutting parameters are shown in Tab. 1, where tools used in ***No. 1~No. 3*** and ***No. 4~No. 5*** differ in both the number of cutting edges and radial cutting thicknesses. Variations in the number of cutting edges alter tool stiffness, which consequently influences the dynamic characteristics of the cutting system. The experimental tools include two-edge flat-end milling cutters (JINLU, type UA100-SL2, Xiamen, China) and three-edge flat-end milling cutters (JINLU, type UA100-SL3, Xiamen, China), both with a diameter of 10 mm and a helix angle of 45°. As shown in Fig. 10(c) and (d), the wedge-shaped thin-walled parts machined in the experiment are made of AL6061-T6 with a cross-sectional area of 40×5 mm2. The workpieces are mounted on the force sensor of the dynamometer using eight bolts to measure the three-directional (3D) cutting forces. During these milling tests (***No.1~No. 5***), the 3D cutting force data are employed to train, validate and test the proposed Chatter-CNN model.

Before developing the milling chatter detection model, the collected cutting force data are marked to identify three machining states in this study. The classification accuracy of the machining states has a direct impact on the trained model's ability to identify these states. However, current data annotation depends on manual judgment of workpiece surface features and spectral characteristics within the vibration detection algorithm, making it prone to misclassification, particularly during transition states including air cut to stable cutting, stabile cutting to chatter, chatter to stabile cutting and stabile cutting to air cut. Fig. 5 illustrates the procedure for signal selection and labelling. As shown in Fig. 5, this research excludes transition state data based on the Short Time Fourier Transform (STFT) spectrum to improve labeling accuracy. In this study, data corresponding to the transition state was excluded in the trained model, but thanks to the deep learning network, more advanced features can be extracted, so the established model can still be used to determine the transition state. In order to facilitate effective training, validation, and testing of the proposed model, the data corresponding to the three machining states (air cut, stable, and chatter) are labeled with 0, 1, and 2, respectively.

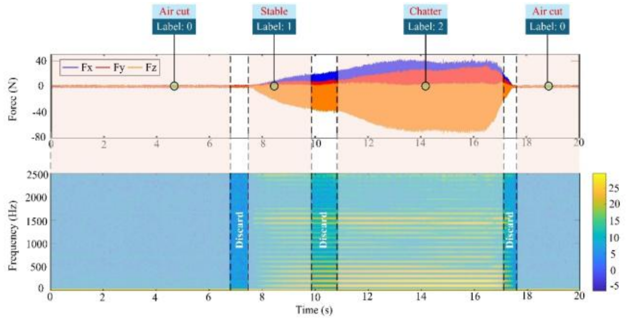

Fig. 5: Procedure for signal selection and labelling.

After data selection and labeling, the milling force signals were segmented using a sliding time-window approach to construct the dataset, as illustrated in Fig. 6. Fig. 6(a) illustrates the principal diagram of the sliding time-window method, while Fig. 6(b) presents the parameter settings adopted in this research. To satisfy the real-time constraints of online chatter identification, the sliding window length is set to 250 data points, corresponding to a time duration of 0.05 s based on the sampling rate. Each window overlaps the next by 200 data points, with a stride of 50 data points, enabling the online chatter detection to update results every 0.01 s.

Based on the aforementioned method, milling experiments (***No. 1~No. 5***) are selected for analysis. The transition state data are excluded, and the remaining data were categorized into three states: air cut, stable and chatter. Using the sliding time window approach proposed in this study, an independent dataset is constructed for each experiment. Subsequently, Fast Fourier Transform (FFT) and Z-score normalization were applied to the segmented data to generate the final datasets used for model training and evaluation. Experiments (***No. 1~No. 3***) were used as model datasets to train and validate the model's prediction accuracy, while experiments (***No. 4~No. 5***) served as test sets to evaluate the model's generalization performance. To facilitate model training and evaluation, the dataset was randomly divided into training and validation subsets at a ratio of 7:3. Importantly, the training and testing data were sourced from distinct experimental setups, which introduced differences in cutting conditions and cutter configuration

Tab. 1: Experimental machining parameters.

Trial No.	Spindle speed n (r/min)	Feed per revolution f_z (mm/r)	Axial cutting depth a_p (mm)	Radial cutting depth a_e (mm)	Overhang length L (mm)	Number of flutes S
No.1	6000	0.06	0-5	2	70	2
No.2	4800	0.06	0-5	2	70	2
No.3	4400	0.06	0-5	2	70	2
No.4	4400	0.06	0-5	2.5	70	3
No.5	6000	0.06	0-5	2.5	70	3

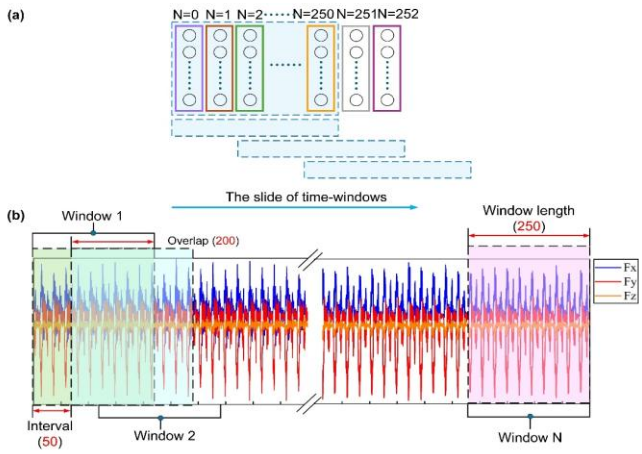

Fig. 6: A sliding time-window method for segmenting milling force signals: (a) The principle of sliding time window method; (b) Parameter settings.

3.2 Parameters setting for chatter detection model

Tab. 2 illustrates the training and validation procedure of the established Chatter-CNN model. During training, the weights and biases of each network layer are randomly initialized based on specific probability distributions to facilitate effective parameter updates. The deep learning network is constructed using the PyTorch framework. The AdamW optimizer is utilized for gradient descent backpropagation and weight parameters update.

Tab. 2: Training Strategy and Validation Results of the Chatter-CNN model.

Input: Training set $\{x_s, y_s\}_{s=1}^{n}$ and validation set $\{x_m\}_{m=1}^{t}$

Output: The results of classification for machining state $\{y_m\}_{m=1}^{t}$

Initialize: The parameters of the Inception modules, SR-blocks, pooling layers, and fully connected layers are randomly initialized based on appropriate probability distributions.

Forward computation in the Chatter-CNN model:

- Multiscale features are extracted from the FFT spectrum of the signal through the use of two Inception modules (i.e., Inception-Chatter_A and Inception-Chatter_B).
- The feature map is downsampled via a max pooling layer to compress spatial information while retaining salient features.
- Two SR-blocks (i.e., SR-block_Ⅰ and SR-block_Ⅱ) are employed to extract fine-grained features, adaptively assign channel-wise weights, and mitigate the network degradation issue.
- An average pooling layer is used to synthesize the generated feature representations and reduce spatial dimensions.
- Three fully connected layers and a final Softmax layer are utilized to classify the machining states.

Back Propagation:

The network gradient is computed using the AdamW optimizer, and the network parameters are subsequently updated.

Tab. 3 presents the hyperparameter settings used during model training. The learning rate is set to 0.0001, the weight decay to 0.001, the batch size to 32, and the learning rate scheduler is configured as StepLR with a step size of 10. Model training is terminated after 60 epochs.

Tab. 3: Training Hyperparameters.

Hyperparameters	Values
Learning rate	0.0001
Batch size	32
Weight decay	0.001
Eproch	60
StepLR.gamma	0.8
StepLR.step_size	10

In order to effectively evaluate the performance of the proposed Chatter-CNN model, it is compared with two widely adopted deep learning models, both of which have been extensively applied in chatter detection, fault diagnosis and tool wear detection [Liu 2022]. One of them is the DenseNet [Huang 2017], which primarily consists of four densely connected blocks, three transition layers, fully connected layers, a convolutional layer, and a max pooling layer. The other model is the ResNet [He 2016]. To adapt with the experimental research in this paper, the traditional DenseNet and ResNet models have been modified. Both models utilize the same dataset as the proposed Chatter-CNN model and directly process frequency domain data as input.

The proposed Chatter-CNN, DenseNet and ResNet models are trained respectively. Fig. 7(a)~(c) shows the loss and accuracy of the training and validation processes of the three models respectively. As shown in Fig. 7, the training and validation losses decrease rapidly to near 0 within the initial epochs as training progresses, and then gradually stabilize. Similarly, the training and validation accuracies increase rapidly to nearly 100% within the initial epochs and then stabilize, indicating that the established Chatter-CNN, DenseNet, and ResNet models do not suffer from gradient vanishing during the training process. Moreover, the training and validation accuracies stabilize close to 100%, demonstrating the good robustness of the three models.

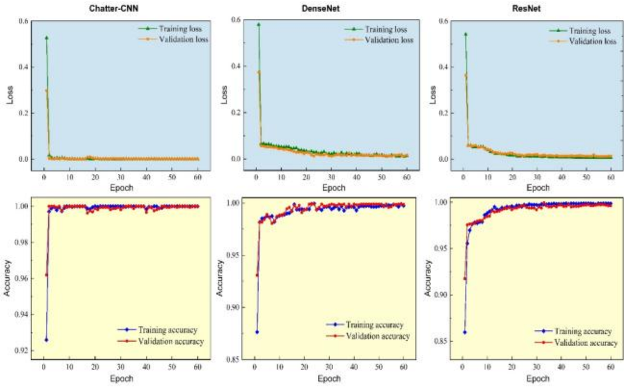

Fig. 7: Training and validation performance (Loss & Accuracy): (a) Chatter-CNN; (b) DenseNet; (c) ResNet.

3.3 Performance evaluation of chatter detection model

3.3.1 Confusion matrices of validation and test sets

In this section, the confusion matrices of the validation and test sets for each model are obtained, as illustrated in Fig. 8(a)~(c). The validation set results indicate that the proposed Chatter-CNN model achieves 100% classification accuracy across all categories; the DenseNet model achieves 100% classification accuracy in all states except the air cut condition. However, the ResNet model achieved a classification success rate of 100% solely in the chatter condition. The validation set results indicate that the proposed Chatter-CNN model achieves higher classification accuracy compared to the DenseNet and ResNet models. The test set results indicate that the proposed Chatter-CNN model achieves a classification accuracy exceeding 90% across all categories, with classification accuracies of 99.4% for the air cut state, 90.5% for the stable state and 97.6% for the chatter state. Although the DenseNet model achieved 100% classification accuracy for the air cut state, its accuracy for the stable state and chatter state is only 51.3% and 71.8%, respectively, resulting in poor discrimination between these two conditions. The ResNet model achieves classification accuracies of 99.1% and 90.1% for the air cut state and stable state, respectively. However, its classification accuracy for the chatter state is only 32.4%, which is significantly lower than that of the proposed Chatter-CNN model. The test set results indicate that the proposed Chatter-CNN model achieves higher classification accuracy compared to the DenseNet and ResNet models. In summary, it can be concluded that the proposed Chatter-CNN model demonstrates higher classification accuracy on both the validation set and the test set. Additionally, the test set varies from the training set with respect to machining parameters, while the validation set shares identical parameters with the training set. As a result, all three models exhibit lower classification accuracy on the test set in comparison to the validation set.

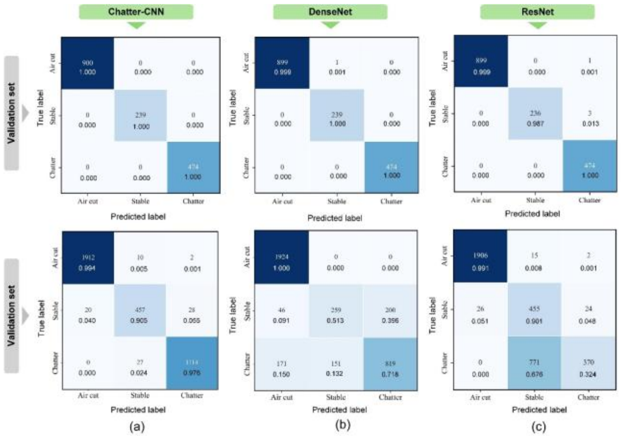

Fig. 8: Classification outcomes: confusion matrices for the validation and test sets: (a) Chatter-CNN; (b) DenseNet; (c) ResNet.

3.3.2 Performance indicators on validation and test sets

Based on the confusion matrices of the three models obtained in ***Section 3.3.1***, evaluation indicators such as precision, recall, and accuracy on both the validation and test sets are further calculated. The computed performance metrics for the validation set and test set are presented in Tabs. 4 and 5, respectively. To account for the uneven sample distribution across machining states, the F1 score is employed, offering a harmonic balance between precision and recall. As shown in Tab. 4, regardless of the machining state, all three models achieve high F1 scores. The proposed Chatter-CNN model achieves an F1 score of 1 across all machining states, outperforming the DenseNet model and the ResNet model. Moreover, the chatter detection accuracy rates achieved by the Chatter-CNN, DenseNet, and ResNet models on the validation set are 100%, 99.9%, and 99.7%, respectively. As shown in Tab. 5, the F1 score achieved by the proposed Chatter-CNN model in each machining state surpasses those obtained by the DenseNet and ResNet models. Among all evaluated models, the proposed Chatter-CNN demonstrates superior generalization performance, attaining a test accuracy of 97.5%. A comparison of the F1 scores and classification accuracy of the three models on the validation set and test set demonstrates that the proposed Chatter-CNN model outperforms the DenseNet and ResNet models.

Tab. 4: Validation set classification performance.

Models	Precision			Recall			F1 score			Accuracy
	Air cut	Stable	Chatter	Air cut	Stable	Chatter	Air cut	Stable	Chatter	
Chatter-CNN	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DenseNet	1.000	0.996	1.000	0.999	1.000	1.000	0.999	0.998	1.000	0.999
ResNet	1.000	1.000	0.992	0.999	0.987	1.000	0.999	0.994	0.996	0.997

Tab. 5: Test set classification performance.

Models	Precision			Recall			F1 score			Accuracy
	Air cut	Stable	Chatter	Air cut	Stable	Chatter	Air cut	Stable	Chatter	
Chatter-CNN	0.990	0.925	0.974	0.994	0.905	0.976	0.992	0.915	0.975	0.975
DenseNet	0.899	0.632	0.804	1.000	0.513	0.718	0.947	0.566	0.758	0.851

ResNet	0.986	0.367	0.934	0.991	0.901	0.324	0.989	0.521	0.481	0.765

4 CONCLUSION

In the past few years, the widespread adoption of AI technology in vibration detection has yielded promising results. Deep learning approaches demonstrate superior capabilities in feature learning and classification, effectively addressing the limitations associated with manual feature extraction in traditional chatter detection approaches. However, data labeling and training conditions significantly impact the accuracy and generalization performance of the chatter detection model. Currently, deep learning-based vibration detection methods primarily depend on full-process data labeling during model training; however, accurately labeling data in transitional states remains challenging, thereby impacting the accuracy of chatter detection. Moreover, the test and training sets are typically partitioned after shuffling the entire dataset, resulting in identical processing parameters for both sets. While this approach yields high accuracy, it often limits the model's applicability to diverse machining conditions, necessitating further enhancement and validation of the model's generalization performance. To address these issues, this study introduces a chatter detection approach for thin-walled parts milling based on the HDCNN and names it as Chatter-CNN. The performance of the proposed Chatter-CNN model is systematically evaluated and verified through a series of comparative milling experiments. Experimental results demonstrate that the proposed Chatter-CNN model outperforms existing CNN methods in chatter detection accuracy.

5 ACKNOWLEDGMENTS

This work is supported by the Key Technology Research and Development Program of Shandong Province (2023CXPT074) and the National Natural Science Foundation of China (51775315).

6 REFERENCES

[Zhu 2020] Zhu, L. and Liu, C. Recent progress of chatter prediction, detection and suppression in milling. Mech Syst Signal Process, 2020, Vol.143.

[Navarro-Devia 2023] Navarro-Devia, J. H., et al. Chatter detection in milling processes—a review on signal processing and condition classification. International Journal of Advanced Manufacturing Technology, 2023, Vol.125, pp 3943–80.

[Jeon 2020] Jeon, B., et al. The architecture development of Industry 4.0 compliant smart machine tool system (SMTS). J Intell Manuf, 2020, Vol.31, pp 1837–59.

[Yuan 2018] Yuan, L., et al. A review on chatter in robotic machining process regarding both regenerative and mode coupling mechanism. IEEE/ASME Transactions on Mechatronics, 2018, Vol.23, pp 2240–51.

[Cordes 2019] Cordes, M., et al. Chatter stability in robotic milling. Robot Comput Integr Manuf, 2019, Vol.55, pp 11–18.

[Postel 2018] Postel, M., et al. High speed tooltip FRF predictions of arbitrary tool-holder combinations based on operational spindle identification. Int J Mach Tools Manuf, 2018, Vol.129, pp 48–60.

[Chu 2022] Chu, et al. Research on the recognition of machining conditions based on sound and vibration signals of a CNC milling machine. IEEE Sens J, 2022, Vol.22, pp 6364–77.

[Bahtiyar 2022] Bahtiyar, K., et al. Lightweight and robust chatter detection algorithms for milling. Manufacturing Letters, 2022.

[Friedrich 2017] Friedrich, J., et al. Estimation of stability lobe diagrams in milling with continuous learning algorithms. Robot Comput Integr Manuf, 2017, Vol.43, pp 124–34.

[Oleaga 2018] Oleaga, I., et al. A machine-learning based solution for chatter prediction in heavy-duty milling machines. Measurement (Lond), 2018, Vol.128, pp 34–44.

[Zhang 2024] Zhang, H., et al. A tree regression algorithm based on incremental gradient boosting. In: Proceedings of an IEEE conference, Institute of Electrical and Electronics Engineers (IEEE), 2024, pp 1–7.

[Yao 2010] Yao, Z., et al. On-line chatter detection and identification based on wavelet and support vector machine. J Mater Process Technol, 2010, Vol.210, pp 713–9.

[Lei 2020] Lei, Y., et al. Applications of machine learning to machine fault diagnosis: a review and roadmap. Mech Syst Signal Process, 2020, Vol.138.

[Zhang 2022] Zhang, P., et al. Cutting tool wear monitoring based on a smart toolholder with embedded force and vibration sensors and an improved residual network. Measurement (Lond), 2022, Vol.199.

[Stavropoulos 2023] Stavropoulos, P., et al. An adaptive, artificial intelligence-based chatter detection method for milling operations. International Journal of Advanced Manufacturing Technology, 2023, Vol.124, pp 2037–58.

[Zhang 2023] 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 Process, 2023, Vol.193.

[Huang 2017] Huang, G., et al. Densely connected convolutional networks. In: Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Institute of Electrical and Electronics Engineers Inc., 2017, pp 2261–9.

[He 2016] He, K., et al. Deep residual learning for image recognition. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, 2016, pp 770–8.

[Yan 2021] Yan, B., et al. Tool wear monitoring of TC4 titanium alloy milling process based on multi-channel signal and time-dependent properties by using deep learning. J Manuf Syst, 2021, Vol.61, pp 495–508.

[Lu 2025] Lu, Y., et al. Milling force prediction in titanium alloy thin-walled components side milling based on Tri-Dexel model with comprehensive consideration of tool runout and workpiece deflection. J Manuf Process, 2025, Vol.141, pp 1211–34.

[Liu 2022] Liu, Z., et al. An improved burr size prediction method based on the 1D-ResNet model and transfer learning. J Manuf Process, 2022, Vol.84, pp 183–97