Sound & Vibration 2025, 59(3), 3502.
https://doi.org/10.59400/sv3502

Article

Small-sample and imbalanced milling chatter detection: Improved GAN with attention and hybrid deep learning

Haining Gao[1,2,*], Xinli Xiong[1], Hongdan Shen[1], Yong Yang[1], Yinlin Wang[1]

[1] Henan Province International Joint Laboratory of New Energy Digitalization Technology, Huanghuai University, Zhumadian 463000, China
[2] School of Mechanical and Power Engineering, Hennan Polytechnic University, Jiaozuo 454000, China
*** Corresponding author:** Haining Gao, 20191908@huanghuai.edu.cn

CITATION

Gao H, Xiong X, Shen H, et al. Small-sample and imbalanced milling chatter detection: Improved GAN with attention and hybrid deep learning. Sound & Vibration. 2025; 59(3): 3502. https://doi.org/10.59400/sv3502

ARTICLE INFO

Received: 17 April 2025
Accepted: 28 May 2025
Available online: 15 June 2025

Abstract: Chatter detection during milling processes plays a pivotal role in ensuring machining quality and efficiency. While the accuracy of chatter detection heavily relies on experimental data, systems tend to exhibit overfitting phenomena under conditions of limited training samples, resulting in diminished detection precision. To address this limitation, this study presents a data augmentation algorithm based on an Improved Generative Adversarial Network (IGAN). This algorithm integrates advanced techniques including Wasserstein distance metrics, cycle consistency constraints, and channel attention mechanisms, effectively enhancing the quality of generated data. An innovative milling chatter detection deep learning model (MNBGA) is constructed, synthesizing cutting-edge architectures such as multi-scale convolutional neural networks, bidirectional gated recurrent neural networks, and attention mechanisms. To optimize model performance, the Ivy algorithm is employed for hyperparameter optimization of the MNBGA model. When the training dataset comprises 40 or more samples, the proposed method achieves detection accuracy exceeding 90%. Notably, under extreme imbalanced data conditions (24:1:1 ratio), the detection accuracy maintains 84.32%. The processing time for 40 samples requires only 76.17 ms, meeting real-time monitoring requirements. This research presents a novel technical solution for addressing the challenge of milling chatter detection under small-sample conditions.

Keywords: milling chatter detection; small-sample learning; imbalanced data; improved GAN; attention mechanism; hybrid deep learning; data augmentation; machining monitoring

1. Introduction

Within aerospace manufacturing, critical components including turbine blades are highly prone to chatter phenomena during their machining processes, which is primarily attributed to their distinctive geometric characteristics and material properties [1]. The occurrence of chatter severely compromises workpiece surface integrity, accelerates cutting tool wear, and poses safety hazards to on-site personnel. To address this issue, researchers have primarily conducted investigations from two directions: chatter prediction [2] and chatter detection [3]. Given the complex and variable nature of industrial machining environments as well as the dynamic characteristics of operational parameters, chatter can still occur even when employing theoretically derived stable cutting conditions [4]. Consequently, chatter detection methods offering real-time capability and high accuracy have garnered substantial attention from researchers.

Research in chatter detection primarily encompasses three aspects: data preprocessing, chatter threshold determination, and machine learning detection methodologies. The variation of chatter frequencies due to noise interference presents

a significant challenge to detection accuracy. Cao et al. [3] employed wavelet packet transformation for signal denoising. Liu et al. [4] eliminated extraneous noise through cross-wavelet transformation. Wan et al. [5] and Chen et al. [6] utilized matrix notch filters to remove noise components. Wang et al. [7] proposed a filtering method based on noise estimation. Wan et al. [8] an advanced variable forgetting factor recursive least squares (VFF-RLS) algorithm specifically designed to effectively eliminate environmental noise interference from collected vibration signals. Lu et al. [9] designed an adaptive frequency band attention module capable of dynamically enhancing frequency bands containing rich chatter information while suppressing noise interference through learning time-frequency domain characteristics. Xiao et al. [10] leveraged the order difference of normalized square envelope spectra to suppress noise interference and enhance the detection capability of grinding chatter features.

Addressing the limitation of single data sources in comprehensively reflecting complex machining states and chatter characteristics, several multi-source approaches have been developed. Zhang et al. [11] proposed a multi-domain data-driven method based on directional attention mechanisms, incorporating temporal cutting vibration data, frequency domain transmission rates, and spatial domain point cloud data. Deng et al. [12] utilized both externally acquired signals and internal control signals for chatter detection and analysis. External signals included acceleration and displacement measurements, while internal controller signals encompassed joint angles, velocities, accelerations, torques, and tool center point positions. Gao et al. [13] achieved chatter detection through the fusion of one-dimensional temporal and two-dimensional image modal information.

Traditional chatter threshold determination methods rely heavily on empirical knowledge and lack adaptive capabilities. To address this limitation, researchers have conducted various investigations. Albertelli et al. [14] introduced an automated chatter threshold determination method based on control chart theory. Zhao et al. [15] implemented a dual-level risk threshold approach, overcoming the constraints inherent in single-threshold systems. Matthew et al. [16] proposed an adaptive thresholding technique for chatter detection by integrating local statistical analysis with time-frequency masking in the improved STFT framework. Yang et al. [17] developed an automatic chatter threshold determination method incorporating maximum entropy principles and sequential probability ratio testing. Shi et al. [18] introduced an adaptive threshold discrimination algorithm based on local statistical features, achieving real-time chatter detection in high-speed milling processes through the implementation of a dynamic mean-standard deviation threshold mechanism within sliding windows.

Machine learning approaches to chatter detection primarily encompass enhanced traditional machine learning methodologies and deep learning techniques. Wan et al. [19] leveraged the adaboost algorithm to construct a series of SVM weak classifiers through iterative processes, combining them into a robust classifier to enhance chatter classification accuracy. Lu et al. [9] introduced an interpretable noise-resistant convolutional neural network for online chatter detection (AD-CNN). Zhang et al. [20] proposed an innovative hybrid deep learning architecture that combines convolutional neural network structures with advanced Inception modules and specialized SR-blocks. This integrated approach leverages the feature extraction capabilities of Inception networks alongside the enhanced representational power of SR-blocks to achieve

improved performance in complex pattern recognition tasks. Sun et al. [21] proposed an innovative deep neural network (ILR-DNN) combining Inception modules, Long Short-Term Memory networks, and residual networks. Li et al. [22] and Wei et al. [23] investigated the application of multi-source information fusion and multi-dimensional attention mechanisms in machining state recognition, achieving superior recognition accuracy. Gao et al. [24] presented an optimized hybrid neural network approach for milling chatter monitoring that incorporates attention mechanisms.

Jauhari et al. [25] proposed a hybrid deep learning method that integrates deep Stem-Inception networks with residual channel-spatial attention mechanisms. Wang et al. [26] developed a milling chatter detection approach for thin-walled components by synthesizing multi-channel signal fusion, channel attention mechanisms, and Bayesian variational attention networks. Sun et al. [27] introduced a chatter detection method that combines deep learning models with meta-learning training strategies. The core principle of the meta-learning training strategy lies in enhancing the generalization capability of the chatter detection model across varying workpieces and cutting conditions through rapid model adaptation.

Existing chatter detection methods heavily rely on extensive experimental data, while experiments covering a wide range of machining parameter combinations are both costly and time-intensive, significantly limiting the industrial application of chatter detection technologies. To address the challenge of insufficient experimental data, data augmentation techniques have emerged as an important research direction. Currently, data augmentation methods primarily encompass physics-based approaches [28], data-driven approaches [29], and hybrid model approaches [30,31]. Physics-based methods generate simulated signals across extensive machining parameter ranges by establishing mathematical models of dynamic milling processes [32], but suffer from high modeling complexity and difficulties in parameter calibration. Hybrid model approaches integrate physical constraints into data-driven frameworks [33], which improve prediction accuracy to some extent but face challenges including difficulties in physical constraint design and increased model complexity. Data-driven methods exhibit minimal dependence on prior knowledge and demonstrate capability in handling complex nonlinear problems, thus gaining widespread adoption. Kuo et al. [34] introduced Generative Adversarial Networks (GANs) into chatter detection data augmentation, addressing the problem of insufficient experimental data through training data generation. However, GANs require relatively substantial original datasets for feature extraction, and when confronted with severely limited original data, GAN training exhibits significant instability, resulting in inconsistent quality of generated samples. To overcome the limitations of GANs, Auxiliary Classifier Generative Adversarial Networks [35] (ACGANs) and Wasserstein Generative Adversarial Networks [36] (WGANs) have been proposed.

In the domain of milling chatter, traditional augmentation methods struggle to achieve satisfactory results due to data scarcity and high-dimensional features, where training instability can lead to generated samples that are either overly similar or extreme. This thesis introduces an IGAN data augmentation method that enhances the traditional GAN framework by incorporating Wasserstein distance metrics, cycle consistency, and channel attention mechanisms. Additionally, we construct an MNBGA model for milling chatter detection that integrates Multi-Scale

Convolutional Neural Networks (MSCNN), Bidirectional Gated Recurrent Units (BiGRU), and Multi Head Self-Attention (MSA). The model's hyperparameters are optimized using the Ivy algorithm. Multiple assessment criteria, such as precision, are employed to confirm the efficacy of the presented approach. The overall architecture is shown in **Figure 1**.

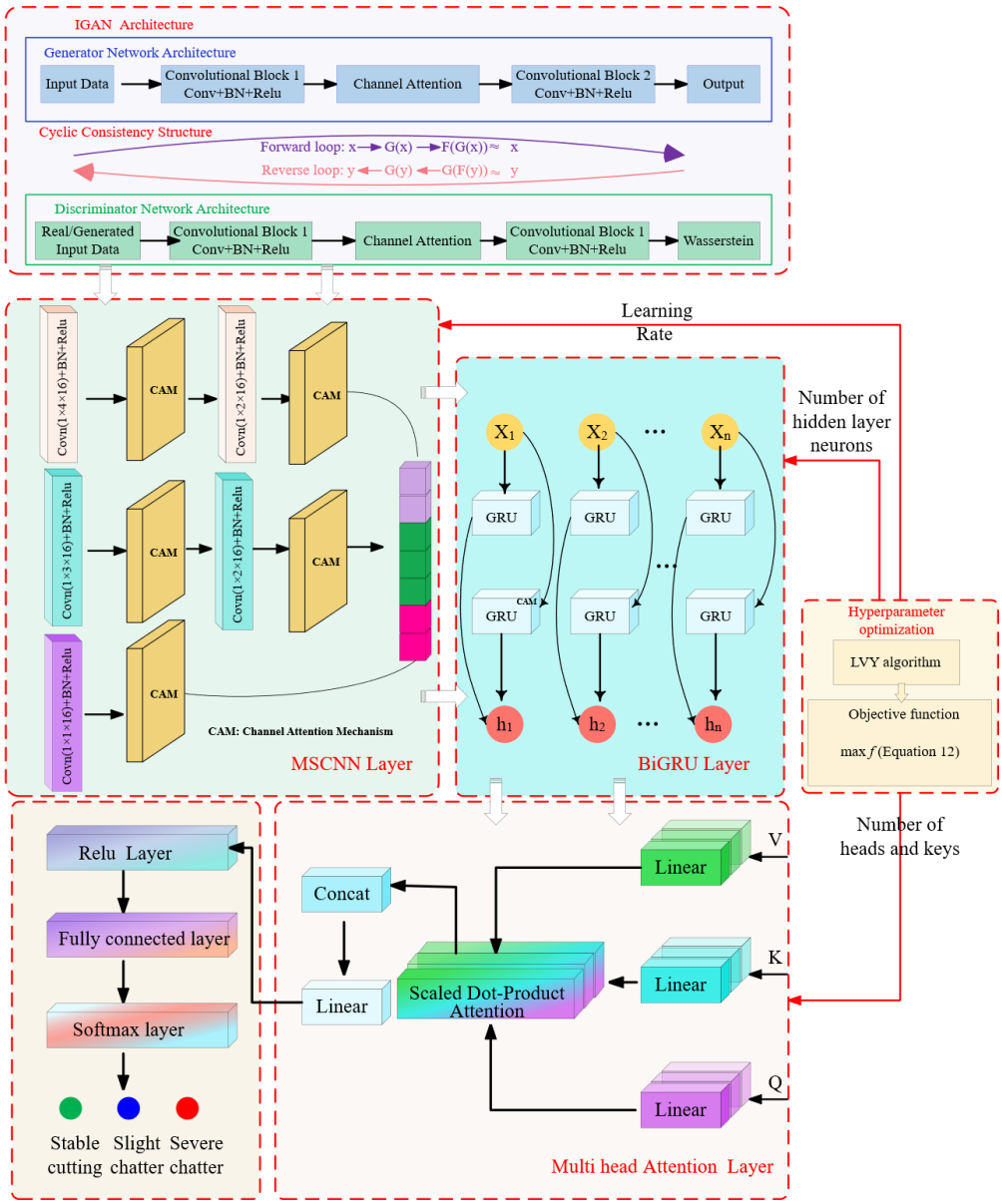

Figure 1. Overall framework of the thesis.

2. IGAN

GAN represents a deep learning architecture inspired by zero-sum game theory [37]. It comprises a generator network and a discriminator network. The generator's loss function is expressed in Equation (1).

$$G(Loss) = \frac{1}{N\sum_{i=1}^{N} log(D(G(z_i)))} \tag{1}$$

where N represents the data volume size, z_i denotes the input noise of the generator,

and $G(z_i)$ indicates the data produced by the generator.

The discriminator's loss function is presented in Equation (2).

$$D(Loss) = \frac{-1}{N\sum_{i=1}^{N} log(D(x_i)+log(1-D(G(z_i))))} \tag{2}$$

where, x_i represents the real data, and $D(x_i)$ denotes the discriminator function.

The objective adversarial loss function for the entire GAN is expressed in Equation (3).

$$L_{GAN}(G,D) = G(Loss) + D(Loss) \tag{3}$$

To address the issue of training termination caused by gradient vanishing in the generator, the Jensen-Shannon divergence is replaced with a Wasserstein function, and a gradient penalty term is incorporated to ensure training stability. The resulting objective loss function is presented in Equation (4).

$$L_{WGAN-GP}(G,D) = L_{GAN}(G,D) + \lambda E_{x\sim P_m}(\|\nabla_x D_\omega(x)-1\|)^2 \tag{4}$$

where, λ represents the gradient penalty coefficient, P_m denotes the sample space interpolated between the real data distribution and noise distribution, and $(\|\nabla_x D_\omega(x)-1\|)^2$ indicates the gradient penalty term after sampling.

In the data augmentation process, two domains, X and Y, are established. The generator network G is tasked with mapping input data from domain X to domain Y, while the discriminator network D aims to differentiate between actual data $F(y)$ from domain Y and the data $G(x)$ generated by generator G.

Vibration data from milling processes exhibits both temporal and periodic characteristics. Additionally, the generated data must preserve the physical properties of the original data. Therefore, a cycle consistency loss function is incorporated. The cycle consistency loss is expressed in Equation (5).

$$L_{cycle}(G,F) = \lambda_{cycle}(E_{x\sim P(x)}\|F(G(x)-x)\|_1 + E_{y\sim P(y)}\|F(G(y)-y)\|_1 \tag{5}$$

where $G(x)$ represents the generated data mapped from domain X to domain Y by generator G. $F(\cdot)$ denotes the generated data mapped from domain Y back to domain X by generator F. λ_{cycle} represents the weighting coefficient of the cycle consistency loss, and $\|\cdot\|_1$ indicates the L_1 norm used to measure the absolute differences between datasets.

Channel attention weights are incorporated into the loss functions of both generator and discriminator to enable more effective weight-based learning during model training. The channel attention loss term can be integrated into the loss functions of both the generator and discriminator in the GAN framework. The attention loss is defined in Equation (6).

$$L_{Attention}(A_X, A_Y) = \lambda_{Attention}(E_{x\sim P(x)}\|A_X(x)-A_Y(G(x))\|_1 + E_{y\sim P(y)}\|A_Y(y)-A_X(F(y))\|_1 \tag{6}$$

where $\lambda_{Attention}$ represents the weighting coefficient of the attention loss, $A_X(x)$ denotes the attention data predicted by the attention network for input x in domain X, and $A_Y(G(x))$ indicates the attention data obtained from attention network A_Y.

The comprehensive loss function of the IGAN is expressed in Equation (7).

$$L(G, D_Y, F) = L_{WGAN-GP}(G, D) + L_{cycle}(G, F) + L_{Attention}(A_X, A_Y) \quad (7)$$

3. Chatter detection model

3.1. MSCNN

Convolutional Neural Networks (CNN), drawing inspiration from the visual processing mechanisms found in biological systems, represent a class of feed-forward artificial neural architectures that are fundamentally constructed through the integration of three primary component types: convolutional layers for feature extraction, pooling layers for spatial dimensionality reduction, and fully connected layers for classification tasks [38]. The computational process of the convolutional layer is expressed in Equation (8).

$$X_{i,j}^{l+1} = \sum_{j=1}^{L} \sum_{i=1}^{m} (X_{i,j}^{l} \times w_{i,j}^{l}) + b \quad (8)$$

where $X_{i,j}^{l+1}$ signifies the post-convolution characteristics of layer $l+1$, $w_{i,j}^{l}$ stands for the convolutional kernel matrix, $X_{i,j}^{l}$ identifies the j-th feature element belonging to the i-th feature plane, b represents the bias parameter, and L denotes the kernel dimensions.

Based on the Inception module architecture, channel attention mechanisms are integrated into each convolutional neural network module to generate channel-specific weights that are continuously optimized during the training process, facilitating network parameter updates. MSCNN iteratively optimizes the significance of different features extracted from machining state data, reinforcing critical features while suppressing interference signals based on their relative importance, thereby enhancing both the diagnostic efficiency and effectiveness of the model.

3.2. BiGRU

BiGRU employs a parallel dual-module architecture to overcome the limitations of unidirectional information flow [39]. The forward module executes stepwise analysis of input data following the progressive order of time series, while the reverse module implements backward information extraction through reverse traversal methodology. Both modules maintain independent internal state vectors and perform updates at each temporal node based on current input features and state information from adjacent time steps. The comprehensive feature output is ultimately generated through concatenation or weighted fusion of hidden representations from both directions.

The mathematical expression for BiGRU's internal hidden units is presented in Equation (9).

$$\begin{cases} \vec{h}_t = GRU(x_t, \vec{h}_{t-1}) \\ \overleftarrow{h}_t = GRU(x_t, \overleftarrow{h}_{t-1}) \\ h_t = f(W_{\vec{h}_t} \cdot \vec{h}_t + W_{\overleftarrow{h}_t} \cdot \overleftarrow{h}_t + b_t) \end{cases} \quad (9)$$

where $\vec{h}_t$ and $\overleftarrow{h}_t$ respectively denote the forward and backward state vectors at

temporal node t, $W_{\vec{h}_t}$ and $W_{\overleftarrow{h}_t}$ correspond to the forward and backward parametric weights of the hidden layer at this timestep, and b_t indicates the bias term of the hidden layer at time t.

3.3. MSA

MSA employs multiple parallel self-attention sub-module architectures that enable simultaneous multi-dimensional processing of identical feature information [40]. The comprehensive output is ultimately generated through concatenation of computational results from individual self-attention sub-modules. This is expressed in the following equation.

$$MultiHead(Q,K,V) = Concat(head_1,\cdots,head_h)\cdot W^o \tag{10}$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \tag{11}$$

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

3.4. Ivy algorithm optimizes MNBGA

Ghasemi et al. [41] introduced the Ivy optimization algorithm in 2024. This approach mimics the developmental behavior of Ivy vegetation through synchronized population expansion and the propagation dynamics of these climbing plants. Mathematical modeling via differential equations and extensive empirical studies captures the developmental velocity of Ivy species. The methodology leverages information from adjacent plant specimens to establish directional growth patterns and accomplishes enhancement through identifying the closest and most influential neighboring elements.

The MNBGA model contains numerous hyperparameters such as learning rate, number of neural network layers, and number of attention mechanism heads, which significantly influence model training effectiveness. The Ivy algorithm is implemented to adjust model hyperparameters, thereby enhancing detection performance. This thesis presents optimization of the MNBGA network's learning rate, number of BiGRU hidden layer neurons, and number of attention mechanism heads. The accuracy rate of validation set samples is employed as the fitness function for the Ivy algorithm in optimizing MNBGA model hyperparameters.

$$max\ f = \frac{\sum_{i=1}^{N}(Y_i == Y_{vaild})\times 100}{N} \tag{12}$$

Various metrics including accuracy are employed to evaluate prediction results across different models. Details of various indicators can be found in reference [42]. The specific calculation for these metrics is shown in Equation (13).

$$\begin{cases} Accuracy = (TP+TN)/(TP+TN+FP+FN) \\ Pr\,e\,cision = TP/(TP+FP) \\ Re\,c\,all = TP/(TP+FN) \\ F1\ Score = 2*(Pr\,e\,cision * Re\,c\,all)/(Pr\,e\,cision + Re\,c\,all) \\ Specificity = TN/(TN+FP) \end{cases} \tag{13}$$

4. Milling experiment

4.1. Experimental setup

The experimental apparatus utilizes a VDL-1000E triaxial CNC milling machine produced by Dalian Machine Tool Corporation. A four-tooth flat-bottom end mill with 10 mm diameter serves as the cutting implement. The tool rake angle is 9° and the helix angle is 35°. Workpiece specimens consist of TC4 titanium alloy plates sized 200 × 200 × 5 mm. The vibration measurement apparatus encompasses a PCB accelerometer (10.42 mV/g sensitivity) integrated with a Donghua DH5922 acquisition system, configured for 5000 Hz data sampling. The experimental configuration is illustrated in **Figure 2**. Numerical simulations are conducted under a Windows 10 (64-bit) operating environment, with hardware specifications including an Intel Core i9-12900K microprocessor, NVIDIA GeForce RTX3080 graphics processing unit, operating frequency of 3.2 GHz, system memory of 32 GB, and computational software implemented on the MATLAB R2023b platform.

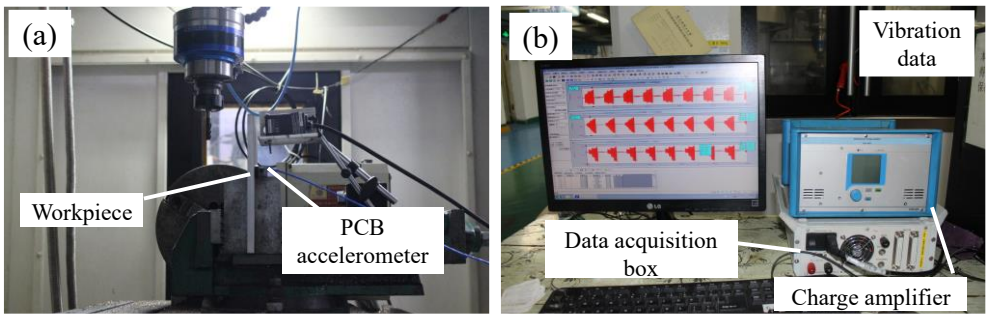

Figure 2. Experimental setup: **(a)** Testing environment; **(b)** vibration measurement system.

4.2. Experimental parameter settings

Modal characterization conducted with an instrumented hammer (sensitivity: 3.41 pC/N) determines the first-order natural frequency of the thin-walled titanium specimen to be 864 Hz. The time-frequency domain characteristics and cutting parameters are detailed in references [24,43]. The radial depth of cut is set at 0.5 mm, with a feed per tooth of 0.1 mm/tooth. The spindle speed ranges from 750 to 1200 rpm, while the axial depth of cut varies between 4.3 and 8.5 mm. For chatter detection, a total of 16 features are extracted, comprising 10 time-domain features, 5 frequency-domain features, and 1 time-frequency domain feature.

5. Results and analysis

5.1. Chatter detection analysis

The detailed parameter optimization results of the MNBGA model using the Ivy algorithm, including parameter ranges and optimized values, are presented in **Table 1**.

Based on an initial sample size of 100 for each condition, the IGAN method is employed to augment each state sample to 200. Following hyperparameter optimization of the MNBGA framework, the complete data collection is split into training, validation, and testing components with corresponding ratios of 80%, 10%,

and 10%. The model is subsequently trained and evaluated for chatter detection accuracy. The detection results for the training, validation, and test sets across the three states are illustrated in **Figure 3**.

Table 1. DBMA architecture parameter settings.

Parameters	**Learning rate 1**	**Learning rate 2**	**Learning rate 3**	**Learning rate 4**	**Learning rate 5**	**Number of hidden layer neurons in BiGRU**	**Number of MSA heads**
Range	[0.001, 0.1]	[0.001, 0.1]	[0.001, 0.1]	[0.001, 0.1]	[0.001, 0.1]	[0, 100]	[2, 6]
Default value	0.001	0.001	0.001	0.001	0.001	16	2
Optimal value	0.0296	0.0439	0.0399	0.0124	0.0765	25	4

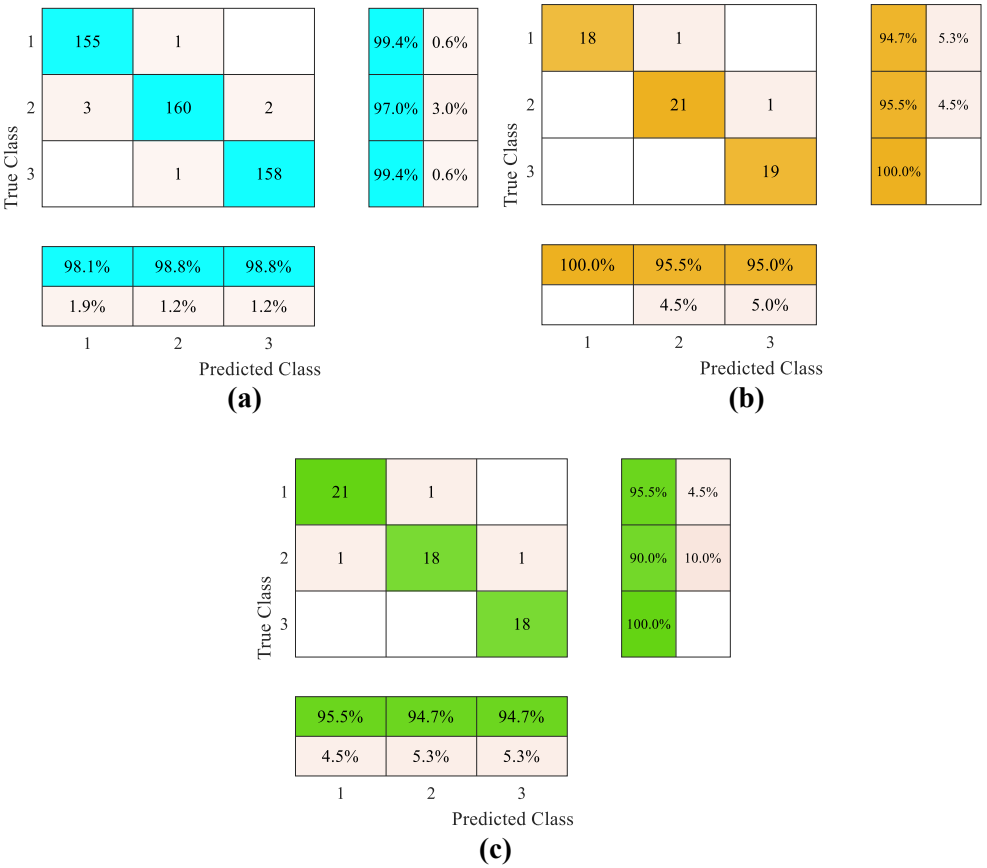

Figure 3. Identification results of various machining conditions: **(a)** Training set confusion matrix; **(b)** validation set confusion matrix; **(c)** test set confusion matrix.

As shown in **Figure 3a**, the recognition accuracy for the training set achieves 98.54%. Specifically, one stable cutting condition is misclassified as slight chatter, three instances of slight chatter are incorrectly identified as stable cutting, two cases of slight chatter are misclassified as severe chatter, and one severe chatter condition is incorrectly identified as slight chatter. As illustrated in **Figure 3b,** the validation set demonstrates a recognition accuracy of 96.67%. The misclassifications consist of one slight chatter condition being incorrectly identified as severe chatter, and one stable cutting condition being misclassified as slight chatter. From **Figure 3c**, the test set exhibit a recognition accuracy of 95%. The misclassifications include one slight chatter condition being incorrectly identified as severe chatter, one instance of slight

chatter being misclassified as stable cutting, and one stable cutting condition being incorrectly identified as slight chatter.

Ten-fold cross-validation approach reduces bias stemming from stochastic dataset segmentation [44]. **Table 2** displays the classification performance before and after parameter tuning.

Table 2. Identification performance before and after hyperparameter tuning.

Category		Training set	Validation set	Test set
Ten-fold Cross-validation Mean	Default	93.85%	91.73%	90.24%
	Optimal	98.54%	96.67%	95.0%

As evident from **Table 2**, model performance demonstrates comprehensive improvement following hyperparameter optimization. The training set accuracy increases from 93.85% to 98.54%, representing an enhancement of 4.69 percentage points. The validation set accuracy improves from 91.73% to 96.67%, showing an increase of 4.94 percentage points. The test set accuracy rises from 90.24% to 95.0%, indicating an advancement of 4.76 percentage points. Overall, hyperparameter optimization not only significantly enhances the model's performance across all datasets but also improves its generalization capabilities, resulting in superior robustness and reliability for practical applications.

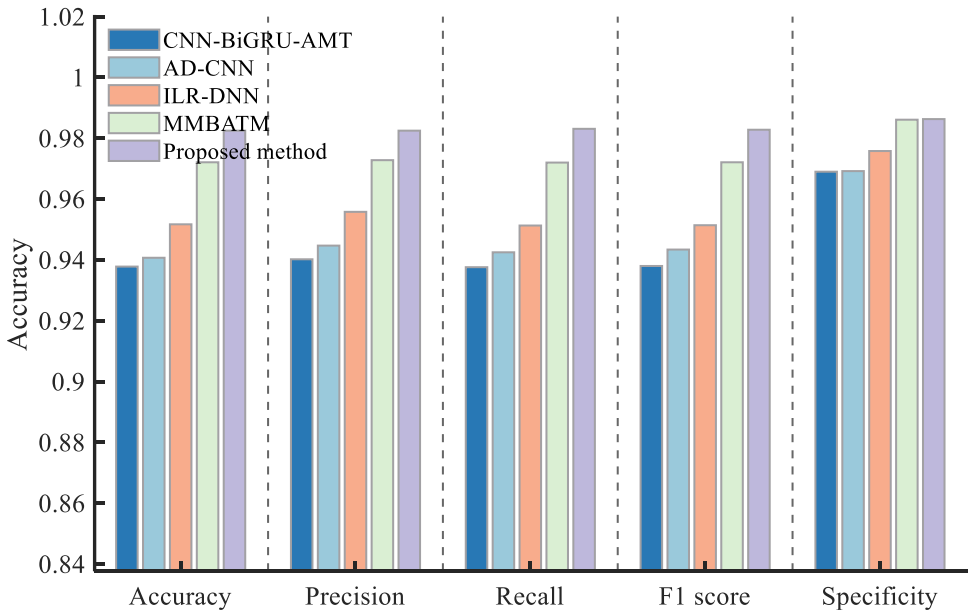

Figure 4. Training set detection results.

To validate the model's reliability and superiority, the proposed method is compared with CNN-BiGRU-AMT, AD-CNN [9] ILR-DNN [21], and MMBATM [24]. For error reduction purposes, mean values from ten-fold cross-validation are employed as final performance indicators. **Figure 4** displays the training dataset comparative analysis, and **Figure 5** presents the testing dataset comparison outcomes.

As demonstrated in **Figure 4**, the proposed method exhibits significant performance advantages across all five evaluation metrics: accuracy, precision, recall, F1-score, and specificity. Most notably, recall shows the most substantial

improvement, increasing dramatically from the baseline method's 93.76% to 98.31%, representing an enhancement of 4.55 percentage points. Improvements in accuracy and F1-score follow closely, advancing from 93.78% and 93.80% to 98.25% and 98.28% respectively, indicating increases of 4.47 and 4.48 percentage points. In terms of precision, the proposed method achieves 98.25%, surpassing the baseline method's 94.02% by 4.23 percentage points, demonstrating significantly enhanced reliability of prediction results. Although the improvement in specificity is comparatively modest, it still increases from 96.90% to 98.63%, showing a gain of 1.73 percentage points, further confirming the model's robustness in handling negative samples.

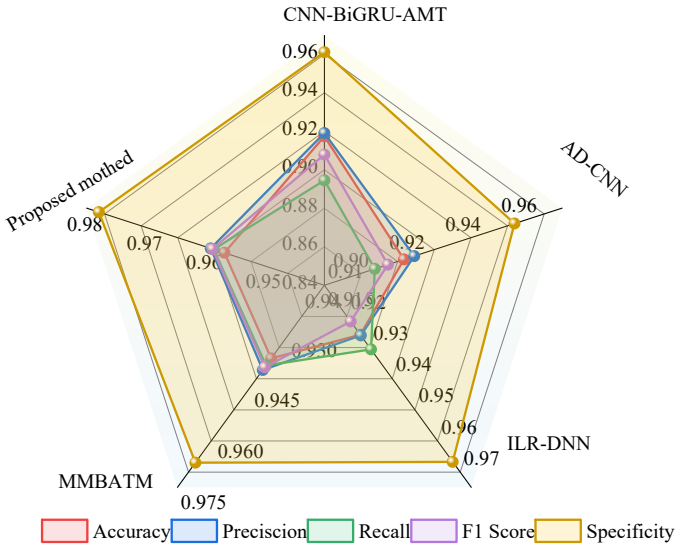

Figure 5. Test set detection results.

As illustrated in **Figure 5**, the proposed method demonstrates significant improvements across all five key evaluation metrics compared to the baseline approaches, highlighting its distinctive advantages. The method not only achieves optimal performance but also exhibits exceptional stability, with minimal fluctuation among various indicators. Specifically, the classification accuracy attains 95.08%, indicating a performance gain of up to 3.34 percentage points over existing methods. The precision escalates to 95.91%, surpassing comparative methods by up to 4.00 percentage points. Notably, the method exhibits remarkable performance in recall, achieving 95.86%, which constitutes an enhancement of up to 6.41 percentage points over comparative approaches. Furthermore, the F1-score and specificity metrics attain 95.88% and 97.79%, respectively, marking improvements of 5.09 and 1.68 percentage points compared to alternative methods. These quantitative results comprehensively validate the superior performance of the proposed method across all evaluation metrics.

The training time and prediction time of different techniques are illustrated in **Table 3**.

Table 3 reveals that the proposed algorithm exhibits moderately higher computational time (104.2 s) than alternative methodologies (78.93 s for CNN-BiGRU-MSA). However, this increased computational cost is considered acceptable given the significant performance improvement (7.6% increase in accuracy). Notably, during the prediction phase, the proposed method requires only 76.17 ms to process

40 samples. This processing speed fully satisfies real-time monitoring requirements. These results validate that the proposed method not only achieves superior detection performance in practical applications but also maintains adequate computational efficiency.

Table 3. Calculation time for different methods.

Time	CNN-BiGRU-MSA	ISR-CNN	ILR-DNN	MMBATM	Proposed mothed
Training (S)	78.93	85.76	83.43	100.5	104.2
Prediction10 (ms)	19.23	14.39	13.17	24.54	24.78
Prediction40 (ms)	62.45	45.76	40.11	75.64	76.17

5.2. Comparative analysis of data augmentation methods

For validating the efficacy of the presented data expansion method, this investigation implements comparison studies with four prominent augmentation algorithms: SMOTE, GAN, WGAN, and WGAN-GP. The experimental investigation is based on 100 initial samples, with various methods employe to expand each state sample to 200 instances. To facilitate comprehensive assessment of the augmentation performance, analyses are performed in both time and frequency domains. **Figure 6** illustrates the comparative results using the absolute mean value of time-domain features as the evaluation metric. **Figure 7** presents the comparison results using the mean square frequency of spectral features in the frequency domain.

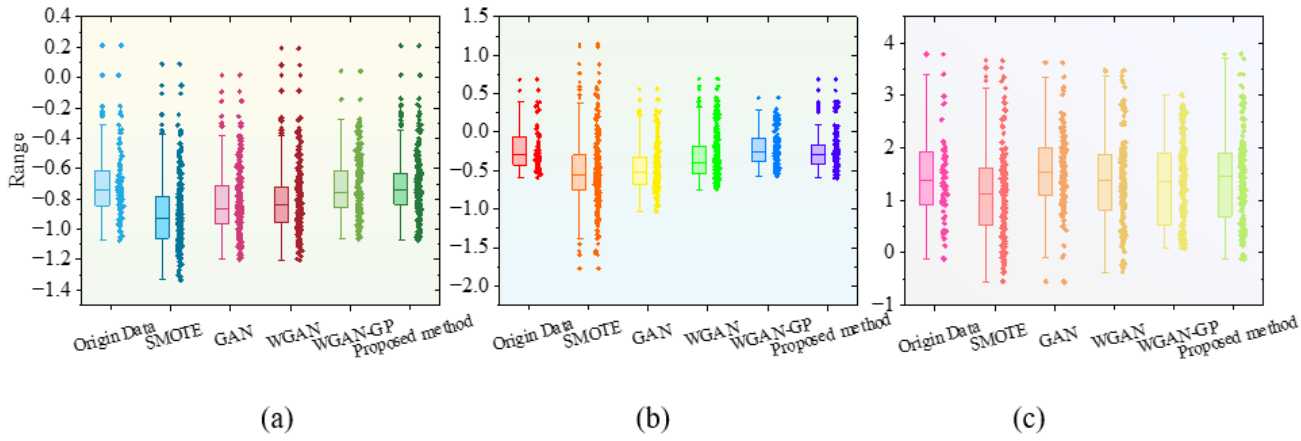

Figure 6. Comparison of time-domain feature data generated by different methods: **(a)** Stable cutting; **(b)** slight chatter; **(c)** severe chatter.

Comparative analysis of **Figures 6** and **7** demonstrates that the traditional SMOTE method exhibits significant limitations across all operating conditions in both feature sets, not only generating numerous outliers but also producing distribution ranges that substantially exceed the reasonable intervals of the original data. These limitations are particularly evident in the slight chatter condition in **Figure 6** (with outliers reaching [−2.0, 1.0]) and the severe chatter condition in **Figure 7** (exhibiting severe distribution distortion). Optimization methods based on the GAN framework demonstrate progressive improvements. GAN and WGAN achieve certain advances in distribution convergence through enhanced generation mechanisms and loss function designs. However, both methods show insufficient capability in maintaining

distribution characteristics across both feature sets. WGAN-GP, through the incorporation of gradient penalty mechanisms, significantly enhances distribution characteristic restoration capability, showing favorable performance in stable cutting and slight chatter conditions in both **Figures 6** and **7**.

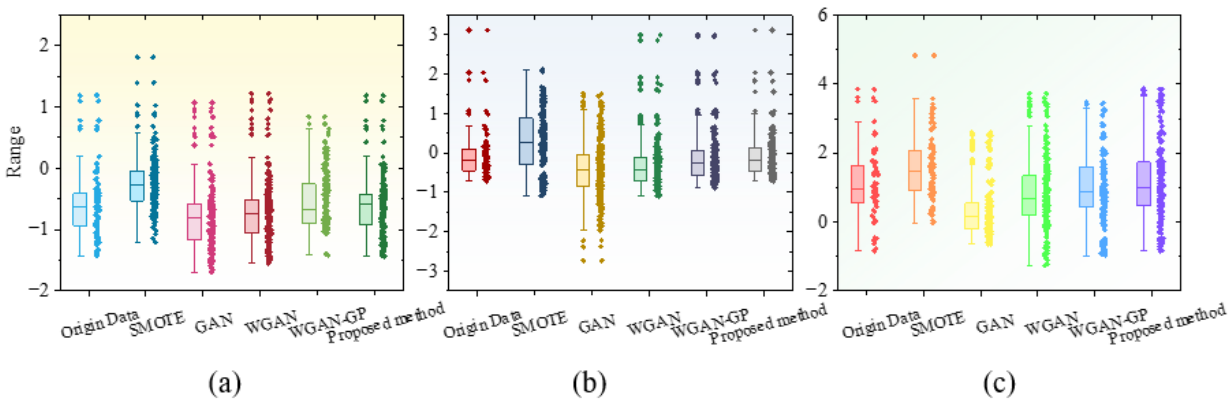

Figure 7. Comparison of frequency-domain feature data generated by different methods: **(a)** Stable cutting; **(b)** Slight chatter; **(c)** Severe chatter.

Notably, the Proposed method demonstrates superior distribution characteristic preservation capability across all conditions in both feature sets. This superiority is manifested not only in the consistency of statistical indicators (including box plot characteristics and quartile ranges) but also in its precise capture of distribution details. Specifically, under slight chatter conditions in **Figure 6**, the method achieves near-perfect replication of the Original Data's distribution characteristics. Under severe chatter conditions in **Figure 7**, it maintains optimal distribution morphology restoration. The cross-validation of these experimental results comprehensively validates the excellent generalization capability and robustness of the proposed method, while highlighting its stability advantages across different feature quantities and operating conditions.

The quality of generated samples from the five methods is evaluated using Wasserstein distance (WD) and Maximum Mean Discrepancy (MMD) metrics. The comparative results across different methods are presented in **Table 4**.

Table 4. Quality evaluation results of samples generated by different methods.

Category	SMOTE	GAN	WGAN	WGAN-GP	Proposed method
WD	1.72	1.16	0.84	0.73	0.65
MMD	0.54	0.36	0.17	0.12	0.08

As shown in **Table 4**, the traditional SMOTE method exhibits WD and MMD metrics of 1.72 and 0.54 respectively, indicating significant deviation between generated samples and the original distribution. Optimization methods based on the GAN framework demonstrate progressive improvements: GAN reduces these metrics to 1.16 and 0.36, WGAN further optimizes them to 0.84 and 0.17 through the Wasserstein distance loss function, and WGAN-GP improves the metrics to 0.73 and 0.12 through gradient penalty strategy. Notably, the Proposed method demonstrates superior performance, reducing WD and MMD to 0.65 and 0.08 respectively,

significantly outperforming other comparative methods. The consistent trend across both evaluation metrics comprehensively validates the superiority of the proposed method in preserving the distribution characteristics of the original data.

The generated data from various methods are input into the pre-trained MNBGA model described in Section 5.1. The mean detection results from ten-fold cross-validation on the test set are illustrated in **Figure 8**.

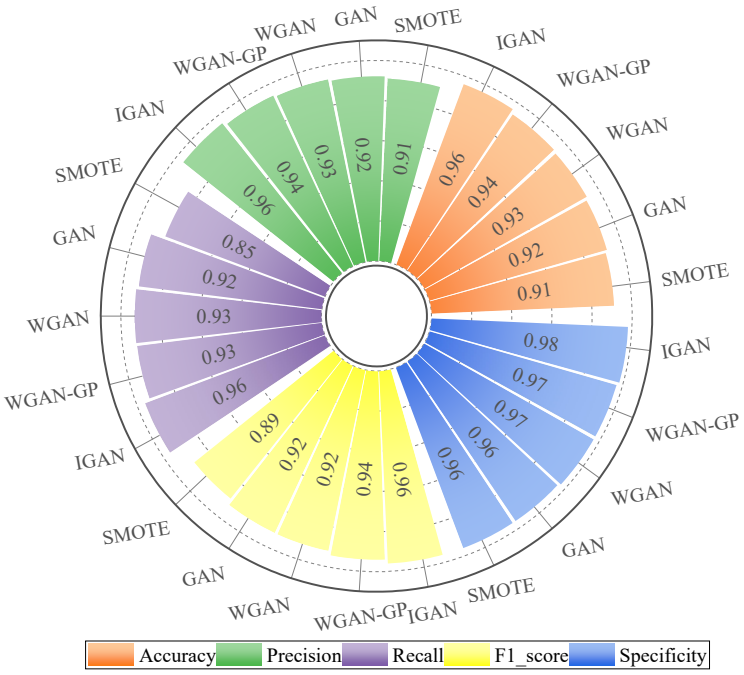

Figure 8. Detection results of different data augmentation methods.

As evident from **Figure 8**, comparative analysis of the five models reveals a significant progressive improvement in performance from SMOTE to IGAN. IGAN demonstrates superior comprehensive performance, achieving exceptional results exceeding 95% across all evaluation metrics. WGAN-GP ranks second, maintaining stable performance within the 93%–96% range across all indicators. WGAN and GAN rank third and fourth respectively, maintaining satisfactory performance levels above 90% despite their relatively lower rankings. SMOTE, representing traditional method, exhibits adequate performance in Specificity but demonstrates comparative weakness in metrics such as Recall, highlighting the distinct advantages of generative adversarial network-based approaches over conventional methods. This performance progression effectively illustrates the improvements achieved through model architecture optimization.

Analysis across the five evaluation metrics reveals differentiated performance characteristics among the models. Specificity emerges as the most stable metric across all models, consistently exceeding 96% with minimal inter-model variation. Accuracy and Precision demonstrate the next highest level of consistency, with all models maintaining performance above 90%. However, more pronounced inter-model variations are observed in Recall and F1-score metrics, with SMOTE notably underperforming in these two indicators relative to its performance in other metrics.

Different data augmentation methods are employed to expand each fault sample

to 150 instances across varying initial sample sizes, followed by training and testing the optimized chatter detection model to evaluate metrics including accuracy. The mean detection results from ten-fold cross-validation on the test set are presented in **Table 5**.

Table 5. Processing state detection results using different data augmentation methods.

Sample size	Methods	Accuracy	Precision	Recall	F1-score	Specificity
20	SMOTE	0.58	0.64	0.63	0.55	0.80
	GAN	0.61	0.66	0.61	0.58	0.81
	WGAN	0.63	0.65	0.62	0.57	0.81
	WGAN-GP	0.67	0.78	0.67	0.65	0.84
	IGAN	0.79	0.80	0.79	0.79	0.89
30	SMOTE	0.65	0.81	0.64	0.55	0.84
	GAN	0.70	0.74	0.70	0.67	0.85
	WGAN	0.70	0.74	0.70	0.66	0.84
	WGAN-GP	0.76	0.75	0.76	0.76	0.88
	IGAN	0.85	0.88	0.85	0.86	0.92
40	SMOTE	0.73	0.78	0.73	0.71	0.87
	GAN	0.78	0.76	0.82	0.78	0.89
	WGAN	0.82	0.82	0.82	0.82	0.91
	WGAN-GP	0.84	0.85	0.84	0.84	0.92
	IGAN	0.90	0.88	0.88	0.88	0.95
50	SMOTE	0.81	0.81	0.81	0.81	0.90
	GAN	0.83	0.89	0.78	0.79	0.92
	WGAN	0.84	0.83	0.84	0.83	0.92
	WGAN-GP	0.90	0.90	0.90	0.90	0.95
	IGAN	0.93	0.93	0.93	0.93	0.96
60	SMOTE	0.81	0.81	0.82	0.81	0.91
	GAN	0.85	0.86	0.85	0.85	0.93
	WGAN	0.88	0.88	0.88	0.88	0.94
	WGAN-GP	0.92	0.92	0.92	0.92	0.96
	IGAN	0.94	0.94	0.94	0.94	0.97

All methods demonstrate performance improvements as the sample size increases from 20 to 60. Taking IGAN as an exemplar, its accuracy progressively increases from 78% at 20 samples to 94% at 60 samples, exhibiting a consistent performance enhancement trajectory. IGAN outperforms alternative methods across all five evaluation metrics (accuracy, precision, recall, F1-score, and specificity) at every sample size. Notably, under limited sample conditions (20 samples), IGAN achieves 78% accuracy, whereas SMOTE attains only 58%, highlighting IGAN's significant advantages in small-sample scenarios.

When the training set contains 40 or more initial samples, the IGAN method enables the model to achieve diagnostic accuracy exceeding 90%, meeting engineering

practical requirements. The method maintains high F1-scores across all sample sizes, indicating effective balance between precision and recall. Specificity consistently remains at elevated levels (reaching 97% at 60 samples), suggesting low false positive rates in practical applications.

SMOTE demonstrates the lowest detection accuracy under severe sample deficiency conditions, achieving only 58%. While detection accuracy improves with increased sample size, the improvement margin remains modest. GAN and WGAN methods demonstrate capability to enhance state detection accuracy to some extent, though their improvement margins show minimal differentiation. WGAN-GP exhibits detection accuracy most comparable to IGAN. The performance gap between these two methods gradually narrows as sample size increases, decreasing from 11.26% (accuracy) at 20 samples to 2.43% at 60 samples.

During the machining process, the collected chatter data is relatively scarce, and the significant disparity between stable and chatter data leads to imbalanced training data for constructing detection models, which affects the accuracy of detection models. To comprehensively evaluate the performance of the improved GAN in data imbalance scenarios, we design specific comparative experiments to validate the effectiveness of the proposed method. In this study, imbalanced datasets are defined as data collections with significantly uneven sample distributions among three milling states (stable cutting, slight chatter, and severe chatter).

Table 6. Unbalanced test dataset.

Imbalance ratio	Stable real data	Slight chatter real data	Severely chatter real data	Generated data	Test real data
4:1:1	120	30	30	90/90	40
6:1:1	120	20	20	100/100	40
12:1:1	120	10	10	110/110	40
24:1:1	120	5	5	115/115	40

A standardized ratio format is employed to precisely quantify the degree of dataset imbalance, specifically including four imbalance configurations: 4:1:1, 6:1:1, 12:1:1, and 24:1:1. In this representation, the first value represents the number of dominant stable cutting samples, while the latter two values correspond to the minority class samples of slight chatter and severe chatter, respectively. This design reflects the data distribution characteristics in actual industrial environments, where stable machining state data is relatively abundant while abnormal chatter state data is relatively scarce. The experimental data configuration is presented in **Table 6**. Each imbalanced dataset maintains consistent overall data scale to ensure fair experimental comparison. The detection results obtained through different methodologies are illustrated in **Figure 9**.

As evidenced by **Figure 9**, when the imbalance ratio decreases from 24:1:1 to 4:1:1, all methods exhibit a marked upward trend in classification accuracy, accompanied by a corresponding reduction in error values. This validates the substantial impact of dataset balance on classification performance. Among all test scenarios, the IGAN method consistently demonstrates superior performance, maintaining classification accuracy between 84% and 94% across various imbalance

ratios, while achieving the minimal error range (0.02–0.05), thus exhibiting exceptional stability. Notably, under highly imbalanced conditions (24:1:1), the IGAN method (accuracy: 84%) shows an improvement of approximately 518 percentage points compared to the original imbalanced data (accuracy: 33%), substantiating the significant advantages of the improved generative adversarial network in addressing class imbalance issues.

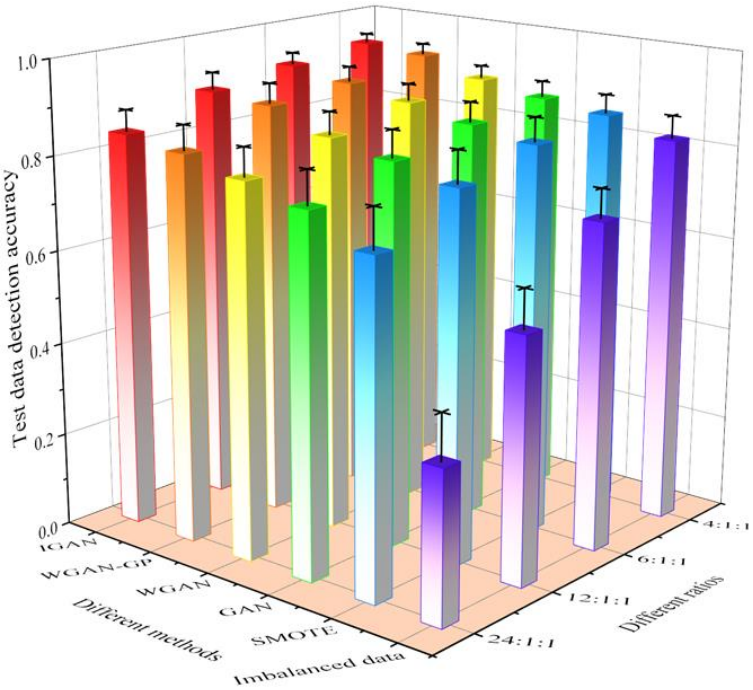

Figure 9. Detection accuracy results of different methods.

6. Conclusions

This thesis proposes an IGAN incorporating Wasserstein distance, cycle consistency, and channel attention mechanisms for data augmentation in small-sample milling chatter detection. A deep learning model integrating MSCNN, BiGRU, and MSA is constructed for milling chatter detection. The Ivy algorithm is employed for hyperparameter optimization of the detection model. The advantages of the proposed method are validated through evaluation metrics including accuracy. The specific conclusions are as follows.

1) The proposed IGAN method demonstrates significant advantages in data augmentation, outperforming baseline methods in preserving original data distribution characteristics. The WD evaluation result is 0.65, representing a 62.2% reduction compared to traditional SMOTE and a 43.9% reduction compared to GAN. The MMD value of 0.08 shows reductions of 85.2% and 77.8% compared to SMOTE and GAN methods, respectively. This substantial improvement in distribution similarity indicates that IGAN-generated samples maintain higher consistency with the original distribution, effectively preserving critical feature information.
2) Under small-sample conditions, the IGAN-MNBGA method exhibits excellent detection performance. With only 20 initial samples, the method achieves 78%

accuracy, while traditional SMOTE achieves only 58% under identical conditions. As sample size increases beyond 40, detection accuracy exceeds 90%, meeting engineering application requirements and demonstrating the method's effectiveness in addressing small-sample chatter detection.

3) The proposed method demonstrates robust performance under imbalanced data conditions. Under extreme imbalance ratios (24:1:1), the IGAN method maintains 84.32% detection accuracy, representing a 51.81 percentage point improvement over original imbalanced data (32.51%) and a 13.76 percentage point improvement over SMOTE (70.56%), indicating significant advantages in handling class imbalance issues.
4) Hyperparameter optimization through the Ivy algorithm significantly enhances model performance. The optimized MNBGA model shows accuracy improvements of 4.69, 4.94, and 4.76 percentage points in training, validation, and test sets, reaching 98.54%, 96.67%, and 95.0% respectively, demonstrating the crucial impact of parameter optimization.
5) Compared to baseline methods, the proposed approach demonstrates superior performance across all five evaluation metrics: accuracy, precision, recall, F1-score, and specificity. The F1-score improvement on the test set reaches 5.09 percentage points.
6) Computational efficiency analysis indicates that while the proposed method requires slightly longer training time (104.2 s) compared to other methods, it processes 40 samples in just 76.17 ms during prediction, fully meeting real-time monitoring requirements. The ratio between computational overhead and performance improvement demonstrates a reasonable balance between computational efficiency and detection performance.

The current research is primarily validated using TC4 titanium alloy and fixed tool parameters, lacking generalization performance evaluation under diverse materials and complex processing conditions. Sensitivity analysis reveals strong dependence of model performance on hyperparameter selection, potentially necessitating time-consuming parameter optimization for different application scenarios. To address these limitations, future research will focus on developing lightweight model architectures, reducing computational resource requirements through knowledge distillation and model compression techniques, and constructing meta-learning-based rapid adaptation frameworks to enhance model transferability across different operating conditions.

Author contributions: Methodology, HG; software, XX; validation, HG, XX and HS; formal analysis, YY; investigation, HG; data curation, HS; writing—original draft preparation, HG, XX; writing—review and editing, HG, HS, YY, YW; visualization, YW. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by the Henan Province Young Backbone Teachers Support Program in Higher Education (grant number 2023GGJS157), the Science and Technology Planning Project in Henan Province (grant number 252102220089, 242102240132), Program for Innovative Research Team (in Science and Technology) in University of Henan Province (grant number

24IRTSTHN020) and the key research and development Program of Henan Province (grant number 241111111600).

Conflict of interest: The authors declare no conflict of interest.

References

1. Wang R, Song Q, Liu Z, et al. Multi-condition identification in milling Ti-6Al-4V thin-walled parts based on sensor fusion. Mechanical Systems and Signal Processing. 2022; 164: 108264. doi: 10.1016/j.ymssp.2021.108264
2. Gao H, Shen H, Liu X, et al. Mechanics and dynamics research considering the tool radial runout effect in plunge milling. The International Journal of Advanced Manufacturing Technology. 2019; 106(5-6): 2391-2402. doi: 10.1007/s00170-019-04780-1
3. Cao H, Lei Y, He Z. Chatter identification in end milling process using wavelet packets and Hilbert–Huang transform. International Journal of Machine Tools and Manufacture. 2013; 69: 11-19. doi: 10.1016/j.ijmachtools.2013.02.007
4. Liu Y, Wang X, Lin J, et al. An adaptive grinding chatter detection method considering the chatter frequency shift characteristic. Mechanical Systems and Signal Processing. 2020; 142: 106672. doi: 10.1016/j.ymssp.2020.106672
5. Wan S, Liu S, Li X, et al. Milling chatter detection based on information entropy of interval frequency. Measurement. 2023; 220: 113328. doi: 10.1016/j.measurement.2023.113328
6. Chen D, Zhang X, Zhao H, et al. Development of a novel online chatter monitoring system for flexible milling process. Mechanical Systems and Signal Processing. 2021; 159: 107799. doi: 10.1016/j.ymssp.2021.107799
7. Wang C, Zhang Y, Hu J. Adaptive milling chatter identification based on sparse dictionary considering noise estimation and critical bandwidth analysis. Journal of Manufacturing Processes. 2023; 106: 328-337. doi: 10.1016/j.jmapro.2023.10.012
8. Wan M, Wang WK, Zhang WH, et al. Chatter detection for micro milling considering environment noises without the requirement of dominant frequency. Mechanical Systems and Signal Processing. 2023; 199: 110451. doi: 10.1016/j.ymssp.2023.110451
9. Lu Y, Ma H, Sun Y, et al. An interpretable anti-noise convolutional neural network for online chatter detection in thin-walled parts milling. Mechanical Systems and Signal Processing. 2024; 206: 110885. doi: 10.1016/j.ymssp.2023.110885
10. Xiao C, Wang L, Yu J. Order-difference of normalized square envelope spectrum and its applications in early chatter detection of roll grinder. Mechanical Systems and Signal Processing. 2025; 224: 112067. doi: 10.1016/j.ymssp.2024.112067
11. Zhang X, Zheng L, Fan W, et al. Multi-domain data-driven chatter detection in robotic milling under varied robot poses based on directional attention mechanism. Mechanical Systems and Signal Processing. 2025; 227: 112406. doi: 10.1016/j.ymssp.2025.112406
12. Deng K, Yang L, Lu Y, et al. Multitype chatter detection via multichannelinternal and external signals in robotic milling. Measurement. 2024; 229: 114417. doi: 10.1016/j.measurement.2024.114417
13. Gao H, Wang H, Shen H, et al. Multi-modal denoised data-driven milling chatter detection using an optimized hybrid neural network architecture. Scientific Reports. 2025; 15(1). doi: 10.1038/s41598-025-88242-7
14. Albertelli P, Braghieri L, Torta M, et al. Development of a generalized chatter detection methodology for variable speed machining. Mechanical Systems and Signal Processing. 2019; 123: 26-42. doi: 10.1016/j.ymssp.2019.01.002
15. Zhao Y, Adjallah KH, Sava A, et al. Incipient chatter fast and reliable detection method in high-speed milling process based on cumulative strategy. ISA Transactions. 2022; 131: 397-414. doi: 10.1016/j.isatra.2022.05.039
16. Matthew D E, Shi J, Hou M, et al. Improved STFT analysis using time-frequency masking for chatter detection in the milling process. Measurement, 2024, 225: 113899. doi: 10.1016/j.measurement.2023.113899.
17. Yang B, Guo K, Zhou Q, et al. Early chatter detection in robotic milling under variable robot postures and cutting parameters. Mechanical Systems and Signal Processing. 2023; 186: 109860. doi: 10.1016/j.ymssp.2022.109860
18. Shi J, Matthew DE, Tian W, et al. Adaptive threshold discrimination and synchronous squeezing transform for high-speed milling chatter detection. Journal of Manufacturing Processes. 2024; 131: 619-640. doi: 10.1016/j.jmapro.2024.09.030
19. Wan S, Li X, Yin Y, et al. Milling chatter detection by multi-feature fusion and Adaboost-SVM. Mechanical Systems and Signal Processing. 2021; 156: 107671. doi: 10.1016/j.ymssp.2021.107671
20. Zhang P, Gao D, Hong D, et al. Improving generalisation and accuracy of on-line milling chatter detection via a novel hybrid deep convolutional neural network. Mechanical Systems and Signal Processing. 2023; 193: 110241. doi: 10.1016/j.ymssp.2023.110241

21. Sun Y, He J, Ma H, et al. Online chatter detection considering beat effect based on Inception and LSTM neural networks. Mechanical Systems and Signal Processing. 2023; 184: 109723. doi: 10.1016/j.ymssp.2022.109723
22. Li R, Wei P, Liu X, et al. Cutting tool wear state recognition based on a channel-space attention mechanism. Journal of Manufacturing Systems. 2023; 69: 135-149. doi: 10.1016/j.jmsy.2023.06.010
23. Wei P, Li R, Liu X, et al. Research on tool wear state identification method driven by multi-source information fusion and multi-dimension attention mechanism. Robotics and Computer-Integrated Manufacturing. 2024; 88: 102741. doi: 10.1016/j.rcim.2024.102741
24. Gao H, Shen H, Yue C, et al. A monitoring method of milling chatter based on optimized hybrid neural network with attention mechanism. Facta Universitatis; 2024.
25. Jauhari K, Rahman AZ, Huda MA, et al. A hybrid deep learning-based approach for on-line chatter detection in milling using deep stem-inception networks and residual channel-spatial attention mechanisms. Mechanical Systems and Signal Processing. 2025; 226: 112357. doi: 10.1016/j.ymssp.2025.112357
26. Wang Y, Gao X, Wang P, et al. Multi-scale enhancement and chatter prediction of milling vibration signals from thin-walled parts based on attentional representation enhancement network. Mechanical Systems and Signal Processing. 2025; 225: 112302. doi: 10.1016/j.ymssp.2025.112302
27. Sun H, Jin H, Zhuo Y, et al. Investigation on a chatter detection method based on meta learning for machining multiple types of workpieces. Journal of Manufacturing Processes. 2024; 131: 1815-1832. doi: 10.1016/j.jmapro.2024.09.091
28. Liu C, Xu X, Wu J, et al. Deep transfer learning-based damage detection of composite structures by fusing monitoring data with physical mechanism. Engineering Applications of Artificial Intelligence. 2023; 123: 106245. doi: 10.1016/j.engappai.2023.106245
29. Zhang X, Shi B, Feng B, et al. A hybrid method for cutting tool RUL prediction based on CNN and multistage Wiener process using small sample data. Measurement. 2023; 213: 112739. doi: 10.1016/j.measurement.2023.112739
30. Zhao S, Zhang H, Peng F, et al. United optimization strategy of ultrasonic vibration assisted process and multiple parameters for machining deformation reduction. Journal of Manufacturing Processes. 2024; 131: 1942-1958. doi: 10.1016/j.jmapro.2024.09.111
31. Zhao S, Peng F, Sun H, et al. Physical multi-factor driven nonlinear superposition for machining deformation reconstruction. International Journal of Mechanical Sciences. 2024; 262: 108723. doi: 10.1016/j.ijmecsci.2023.108723
32. Yin C, Wang Y, Ko JH, et al. Attention-driven transfer learning framework for dynamic model guided time domain chatter detection. Journal of Intelligent Manufacturing. 2023; 35(4): 1867-1885. doi: 10.1007/s10845-023-02133-0
33. Jiang P, Fan J, Li L, et al. A hybrid approach combining mechanism-guided data augmentation and machine learning for biomass pyrolysis. Chemical Engineering Science. 2024; 296: 120227. doi: 10.1016/j.ces.2024.120227
34. Kuo PH, Luan PC, Tseng YR, et al. Machine tool chattering monitoring by Chen-Lee chaotic system-based deep convolutional generative adversarial nets. Structural Health Monitoring. 2023; 22(6): 3891-3907. doi: 10.1177/14759217231159865
35. Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks. International Conference on Machine Learning. 2017.
36. Gulrajani I, Ahmed F, Arjovsky M, et al. Improved training of Wasserstein Gans. In: Proceedings of the 31st International Conference on Neural Information Processing Systems; 2017.
37. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. In: Proceedings of the 28th International Conference on Neural Information Processing Systems; 2014.
38. Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE. 1998; 86(11): 2278-2324. doi: 10.1109/5.726791
39. Ma J, Luo D, Liao X, et al. Tool wear mechanism and prediction in milling TC18 titanium alloy using deep learning. Measurement. 2021; 173: 108554. doi: 10.1016/j.measurement.2020.108554
40. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Advances in neural information processing systems. arXiv. 2017.
41. Ghasemi M, Zare M, Trojovský P, et al. Optimization based on the smart behavior of plants with its engineering applications: Ivy algorithm. Knowledge-Based Systems. 2024; 295: 111850. doi: 10.1016/j.knosys.2024.111850
42. Powers DMW. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv. 2020.

43. Gao H, Shen H, Yu L, et al. Milling chatter detection system based on multi-sensor signal fusion. IEEE Sensors Journal. 2021; 21(22): 25243-25251. doi: 10.1109/jsen.2021.3058258
44. Zhou J, Yue C, Qu J, et al. BDTM-Net: A tool wear monitoring framework based on semantic segmentation module. Journal of Manufacturing Systems. 2024; 77: 576-590. doi: 10.1016/j.jmsy.2024.10.012