

Research on Tool Wear State Recognition Method Based on Multi-Scale Feature Extraction and Deep Residual Network Fusion

Erliang Liu¹, Cong Liu¹, Yuhang Du¹, Baiwei Zhu^{1*}, Limin Shi²

¹School of Mechanical Engineering and Railway Transit, Changzhou University, 213164, Changzhou, China

²Mechano-Electronic and Automotive Engineering College, Tianshui Normal University, 741001, Tianshui, China

Abstract: To enhance the intelligence of machining processes, accurate recognition of tool wear states has become a key issue in the manufacturing field. However, due to the non-stationary and high-dimensional nature of cutting signals, traditional methods face significant challenges in feature extraction and state classification. In the context of cutting processes, challenges such as difficulty in identifying tool wear states and the complex composition of monitoring information features persist. To address these issues, this paper proposes a deep learning model that integrates multi-scale feature extraction with a residual connection network (Multi-scale ResNet). Specifically, cutting vibration signals are processed using continuous wavelet transform (CWT), which enables the conversion of time-frequency information into images. The proposed deep learning model is then used for feature extraction and state identification. The proposed model is validated through cutting experiments conducted on γ -TiAl alloys. Experimental results show that the Multi-scale ResNet model achieves higher recognition accuracy than traditional models such as convolutional neural networks – support vector machines (CNN–SVM), Transformer, and ResNet in the initial and normal wear stages. It effectively mitigates misjudgments associated with initial and normal wear, achieving a prediction accuracy of 93.8 %, a recall rate of 94.2 %, and an F1 score of 94 %. This model offers a novel and effective approach for tool wear state monitoring, contributing to improved cutting processing efficiency and increased intelligence in production.

Keywords: tool wear, multi-scale features, deep residual network, cutting vibration, continuous wavelet transform

1. INTRODUCTION

During the machining process, tool wear is unavoidable. Recognizing tool wear conditions is one of the most important technologies for increasing cutting efficiency and advancing the intelligent transformation of manufacturing processes. There are two types of tool condition monitoring methods: direct and indirect. The direct method uses high-speed cameras and optical microscopes to determine the tool wear level. In contrast, the indirect method derives tool wear conditions from sensor signals collected during machining, such as machine tool current, power, noise emission signals, or vibration signals. Unlike direct monitoring, indirect approaches are less intrusive to the machining process and are therefore more widely used [1]. Currently, the most widely applied machine learning methods for tool condition monitoring include artificial neural networks (ANN), ensemble learning, support vector machines (SVM), Bayesian network classifiers, and Hidden Markov Models (HMM) [2], [3]. In addition, convolutional neural networks (CNN) and long short-term memory (LSTM) networks are among the most commonly used deep learning models for tool wear

prediction [4], [5]. However, a major limitation of these approaches is that they typically use shallow architectures, which are inadequate for capturing deep features from large-scale datasets [6]. Furthermore, manual feature extraction may result in the loss of critical information contained in raw signals [7], ultimately affecting both the training efficiency and recognition accuracy of the model. Deep learning methods, by eliminating the need for handcrafted features, overcome the bottlenecks of traditional neural networks, such as gradient vanishing and local minima, and demonstrate significant potential in condition monitoring tasks [8]. Liu et al. [9] established the relationship between acoustic signals and tool wear under various cutting conditions, using regression analysis and ANN to predict the degree of tool wear, thereby eliminating the dependency on specific cutting parameters. Guo et al. [10] analyzed the correlation between tool wear and the fluctuation trends of milling signals and developed a model linking tool wear with multifractal parameters, enabling tool wear condition monitoring. Mohanraj T et al. [11] used vibration signals as training data, extracting wavelet coefficients and statistical features, and

applied various algorithms to classify tool wear states. Shi C et al. [12] proposed a novel deep learning-based data-driven modeling framework, in which multiple parallel feature spaces were constructed to train the model. By integrating high-level features with tool wear-related characteristics, their method achieved a classification accuracy exceeding 96 %. Ma M et al. [13] fused vibration and acoustic signals and introduced a Deep Coupled Restricted Boltzmann Machine (DCRBM) model. The proposed symmetric DCRBM outperformed other fusion strategies and demonstrated excellent performance in tool condition evaluation. Zhang et al. [14] proposed a compact CNN to address the overfitting problem caused by the limited number of fault samples.

The residual neural network (ResNet) model, developed by Kaiming He, was designed to construct ultra-deep neural networks that are not affected by the vanishing gradient problem [15]. ResNet is a conventional feedforward network architecture with residual connections, where the output of the L^{th} layer is defined based on the output of the $(L-1)^{\text{th}}$ layer and the output after performing various operations. These operations typically include convolution with filters of different sizes, batch normalization (BN), and the application of activation functions (e.g., ReLU) to the $(L-1)^{\text{th}}$ layer's output [16], [17]. ResNet architectures have been developed with varying depths, such as ResNet-18, ResNet-34, ResNet-50, ResNet-101, ResNet-152, and even ResNet-1202. The widely used ResNet-18 consists of 17 convolutional layers and one fully connected layer at the end of the network.

As indicated by the aforementioned literature, tool wear state recognition in the field of deep learning is primarily achieved through CNNs and LSTM networks. However, CNNs are inherently limited in their ability to fully capture temporal dependencies within the data, necessitating various auxiliary processing steps. Conversely, while LSTM networks can model time-series information, they still suffer from attenuation of long-range dependencies and are intrinsically more complex than CNNs, which negatively impacts computational efficiency. Due to their inadequate utilization of multi-scale features, conventional neural network models struggle to accurately capture and distinguish the subtle variations among different wear regions. This limitation results in significant misclassifications and ambiguous boundaries when identifying the various stages of tool wear. Therefore, this paper proposes a deep residual connection network model based on stacked convolutional structures and multi-scale feature extraction modules (Multi-scale ResNet), aiming to enhance the perception and extraction capabilities for multi-frequency and multi-scale features in cutting vibration signals. By incorporating multi-scale convolutional structures and residual connection mechanisms, the model maintains the expressive power of deep-layer features while effectively mitigating the information loss problem encountered in traditional networks as depth increases. The proposed model was validated

through cutting experiments on γ -TiAl alloys. The results demonstrate that Multi-scale ResNet achieves higher accuracy in tool wear state recognition tasks, particularly in the initial wear and normal wear stages, significantly reducing misjudgment rates and exhibiting stronger robustness and practical application value.

2. TOOL WEAR SIGNAL PROCESSING AND FEATURE EXTRACTING

During the machining process, as cutting time increases, tool wear progressively intensifies. This wear progression is typically illustrated by a tool wear curve, as shown in Fig. 1. The tool wear process can be divided into three distinct stages: the initial wear stage, the normal wear stage, and the severe wear stage. Due to highly variable cutting parameters and complex working conditions in real-world manufacturing environments, accurately identifying the tool wear state has become crucial for ensuring machining stability and maintaining product quality. If the wear condition is not correctly identified, improper timing of tool replacement may occur: replacing the tool too early leads to underutilization and increased production costs, while replacing it too late may result in decreased machining quality, tool breakage, or even damage to the machine itself. Such issues significantly compromise overall processing efficiency and product quality. Therefore, there is an urgent need to establish an efficient and reliable tool wear state monitoring system to achieve real-time and accurate assessment of tool condition during the machining process. To this end, this paper proposes a deep neural network structure (Multi-scale ResNet) based on deep residual connections and stacked multi-scale feature extraction modules, aiming to fully exploit the multi-scale and multi-frequency feature information in cutting vibration signals, thereby improving the accuracy and response speed of tool wear state recognition. The model effectively alleviates information attenuation and gradient vanishing problems typically found in traditional deep networks while ensuring deep feature extraction, providing a feasible technical approach and theoretical support for real-time tool wear state monitoring.

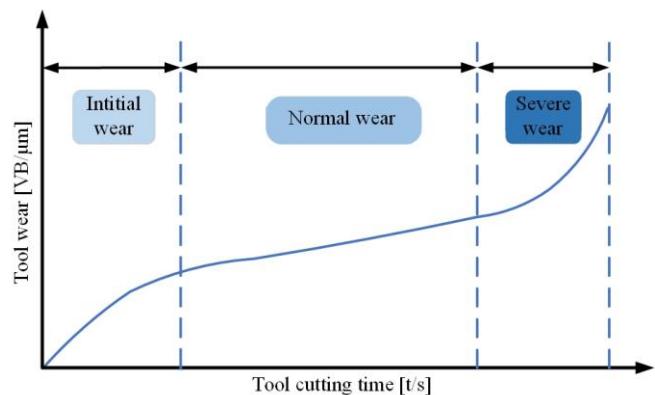


Fig. 1. Tool wear curve.

In this study, cutting tool wear experiments on γ -TiAl alloy were conducted using Mitsubishi carbide inserts CNMG120408-MJ (carbide grade RT9010) on a CKA6150 lathe. The machining parameters were set as follows: cutting speed v_c was 25 m/min, feed rate f was 0.15 mm/r, and depth of cut a_p was 0.15 mm/r. Vibration signals during the cutting process were collected using a vibration acquisition system from Shenzhen Jilanding Intelligent Technology Co., Ltd. The flank wear of the tool was measured and recorded using a VHX-700C super-depth 3D microscopy system (Japan), which is equipped with a 20-2000 \times zoom optical lens, providing a maximum image resolution of 0.5 μm , and employs automatic focusing and extended depth-of-field (EDF) techniques to obtain the flank wear measurements of the tool. The experimental equipment used is shown in Fig. 2.

Tool wear was quantitatively measured using an EDF 3D microscopic imaging system. The acquired wear data were subsequently classified into distinct wear states using the Expectation-Maximization (EM) algorithm. The classification results, along with the corresponding time intervals for each wear stage, are presented in Table 1. This table provides the measured wear values and associated machining durations for each stage. In the severe wear stage, tool failure is indicated by a wear value of $+\infty$.

Table 1. Tool wear status classification.

Tool wear [mm]	Tool wear status	Time [min]
[0, 0.0675)	Initial wear	[0, 9)
[0.0675, 0.245)	Normal wear	[9, 23.5)
[0.245, $+\infty$)	Severe wear	[23.5, 29]

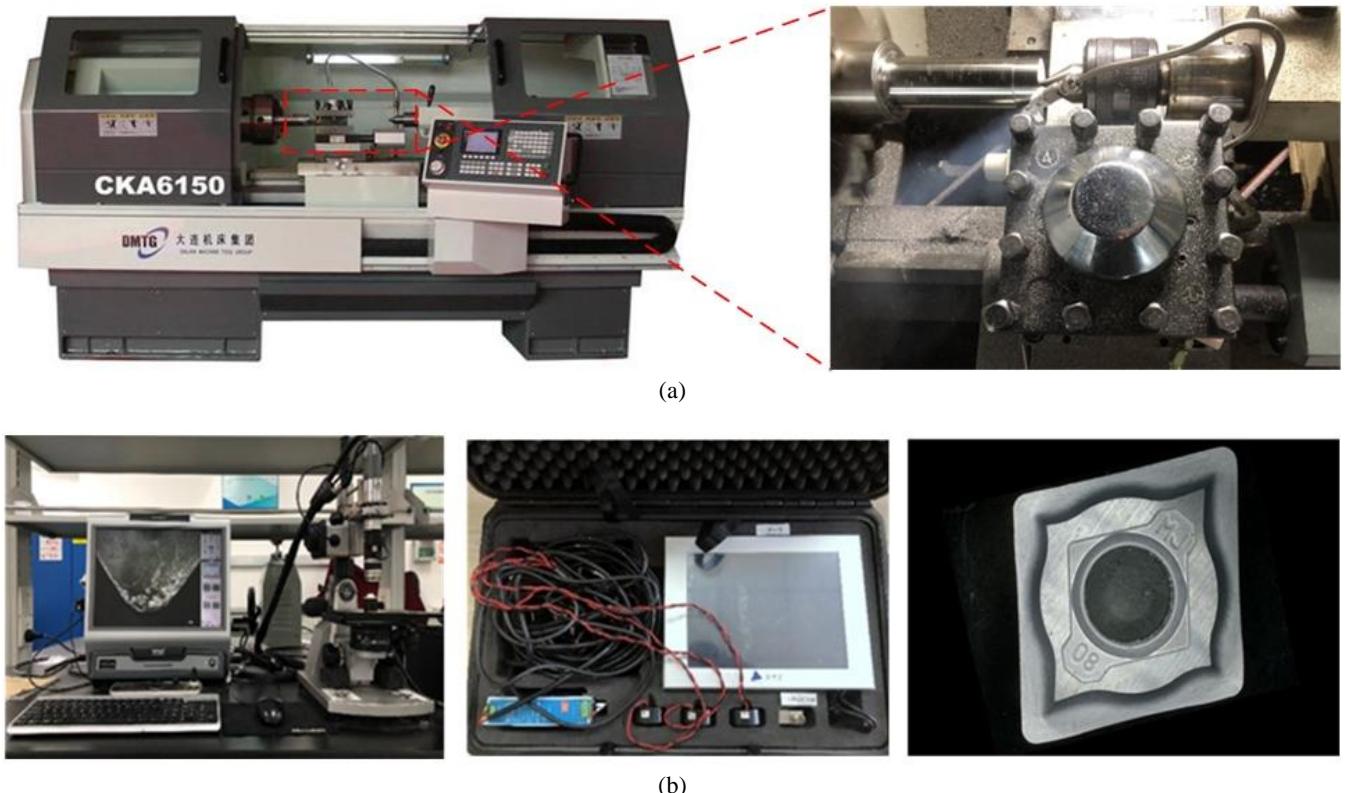


Fig. 2. Experimental and signal acquisition equipment. (a) Experimental machine tool and cutting test site; (b) Super depth-of-field microscope, vibration signal acquisition equipment, and test tool.

A. Data preprocessing

During the cutting process, a large volume of raw signal data is collected from sensors. This data contains information relevant to tool wear as well as noise and other irrelevant components. If used directly for recognition by deep learning models, it may result in excessive computational load and reduced recognition accuracy. Therefore, it is necessary to preprocess the low-quality raw data. During cutting, the tool does not engage with the workpiece during tool entry and

exit, so no wear occurs at these stages. Thus, in the data processing stage, signals corresponding to tool entry and exit should be removed to prevent interference with the monitoring results. Specifically, the first and last 2.5 % of the raw data are considered invalid and are excluded, with the remaining data regarded as effective cutting signals. From these effective cutting signals, the relatively stable segments of the cutting process are further extracted as the final valid data input for the model. The raw signal data processing procedure is illustrated in Fig. 3.

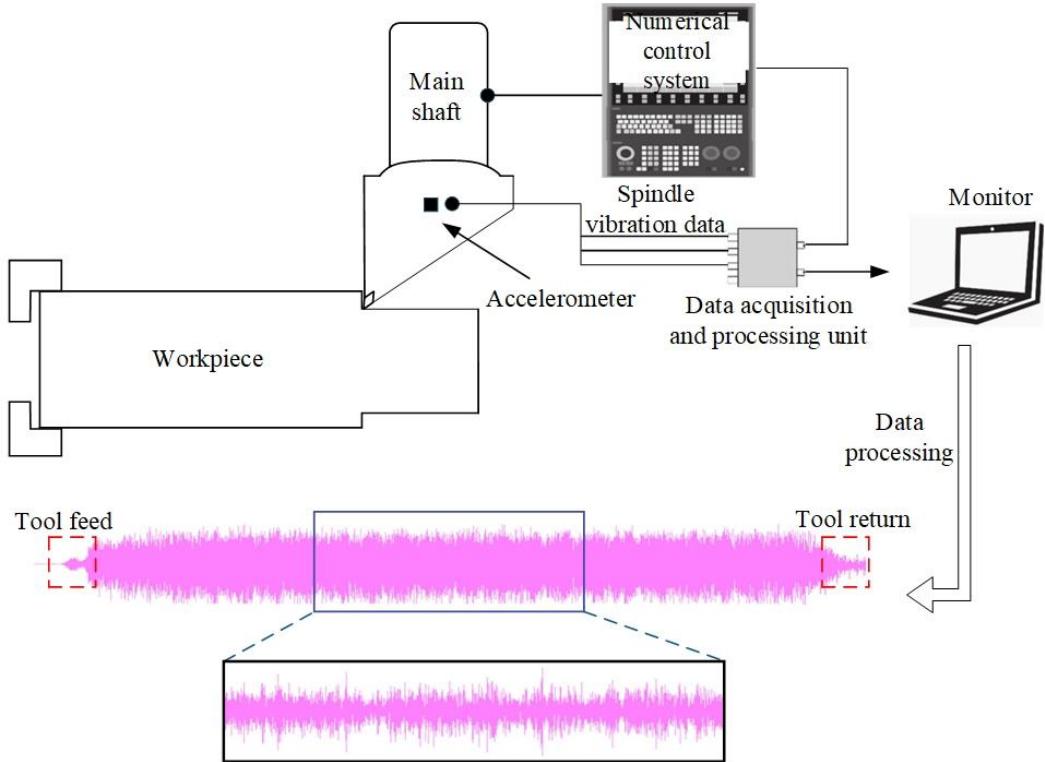


Fig. 3. Data collection and preprocessing.

B. Time-frequency domain feature extraction

Time-domain or frequency-domain analysis methods are suitable for processing stationary signals, but they can extract feature information only from either the time or frequency domain. However, the signals obtained during the cutting process are often non-stationary. Therefore, this study adopts an EDF 3D microscopic imaging system to transform the time-series data into images, enabling a joint representation of both time and frequency characteristics. These images effectively capture the dynamic patterns of the signal over time. Subsequently, a deep learning model is constructed to extract features from the images and achieve accurate recognition of tool wear states.

To ensure that the time-frequency representations capture all relevant dynamic components of the cutting process, the vibration signals were sampled at 20 kHz. This sampling rate has been widely adopted in previous studies. According to the Nyquist criterion, the sampling rate should be at least twice the highest frequency present in the signal to avoid aliasing. Preliminary experiments were also conducted to compare different sampling rates (10 kHz, 20 kHz, and 40 kHz), and the results showed that 20 kHz adequately preserves the frequency components related to tool wear while keeping the data volume and computational load manageable.

Common time-frequency analysis (TFA) methods include the Fourier Transform [18], continuous wavelet transform (CWT) [19], and Wigner–Ville Distribution [20]. These methods can generate time-frequency representations of signals. However, when analyzing signals with multiple frequency components, the Wigner–Ville Distribution, as a quadratic transform, may experience cross-term interference. The Fourier Transform, on the other hand, uses

constant sampling intervals in both the time and frequency domains, which limits its ability to adjust the transformation window size according to frequency variations. In contrast, the CWT overcomes the resolution limitations of the Fourier Transform by providing multi-resolution representations, making it especially suitable for analyzing complex time-series signals. Therefore, in this section, the CWT is used to convert the preprocessed vibration signals into $22 \times 224 \times 3$ time-frequency images, as shown in Fig. 4.

Let the signal collected during the cutting process be $x(t)$. The CWT of $x(t)$ is expressed as:

$$WT(\tau, s) = \int_{-\infty}^{\infty} x(t) \psi_{\tau,s}^*(t) dt = \frac{1}{\sqrt{s}} \int_{-\infty}^{\infty} x(t) \psi_{\tau,s}^*\left(\frac{t-\tau}{s}\right) dt \quad (1)$$

where $\psi(t)$ is the wavelet basis function, and τ, s are the translation and scaling coefficients.

On the time axis, the wavelet is shifted by the translation coefficient τ and convolved with the target signal, enabling extraction of the signal's temporal dependencies. In terms of frequency, the length and frequency of the wavelet are adjusted by the parameters. The window length changes synchronously with the wavelet length, giving wavelet analysis multi-resolution properties. This approach allows for finer capture of the signal's temporal variations when processing high-frequency components and more accurate distinction of the signal's frequency characteristics when handling low-frequency components. Compared to the short-time Fourier Transform (STFT), this method is better suited for capturing transient changes in non-stationary signals.

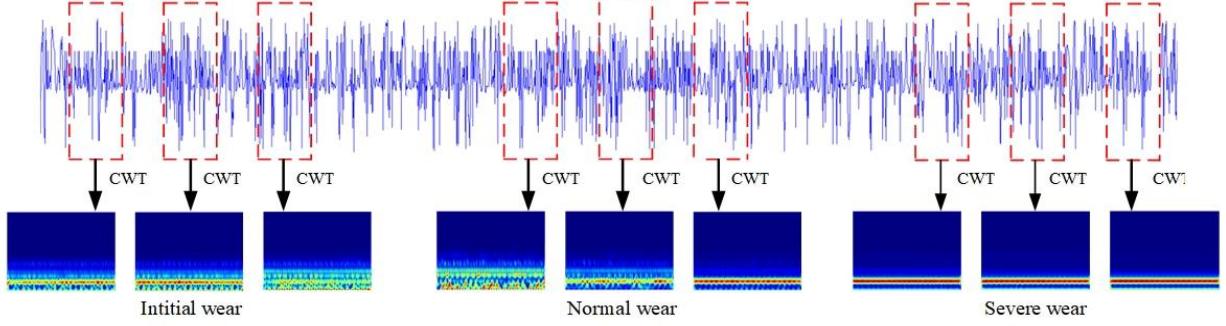


Fig. 4. The vibration signals and their CWT at different stages of tool wear.

3. DEEP LEARNING-BASED TOOL WEAR CONDITION MONITORING

The concept of multi-scale feature extraction originates from the human visual system's ability to distinguish objects at various scales. In image processing, multi-scale convolution helps capture both local and global information. In time-frequency images, different scale convolution kernels can detect vibration patterns at different frequency bands. On the other hand, the introduction of ResNet addresses the gradient degradation problem in deep networks and allows information to be transmitted across different layers, enabling the construction of more complex nonlinear mappings. Therefore, this paper embeds the multi-scale convolution structure into the residual blocks of ResNet to enhance the model's feature perception range and improve the network's ability to distinguish differences in vibration images under different wear states.

The ResNet network, with its designed residual connections, effectively addresses the problems of gradient vanishing and gradient explosion in deep learning models [15]. Residual connections enable the network to transmit information directly across multiple layers, helping to capture features at different levels. This cross-layer feature transmission enhances the network's perception and representation capabilities. Therefore, ResNet is selected as

the base model for deep learning. However, the residual structure of ResNet focuses only on feature extraction at a single scale, which may cause the model to overly focus on certain parts of the sensor signal while neglecting other important information. To address this, this paper incorporates a multi-scale feature extraction module based on ResNet to improve the model's ability to extract multi-scale features. Experimental results show that the proposed model outperforms traditional methods.

A. Residual block network

ResNet is a significant milestone in the development of convolutional neural networks. In ResNet, a residual block contains a skip connection, allowing the neural network to directly learn the residual, i.e., the difference between the input and the desired output, as shown in Fig. 5. This design facilitates easier information flow within the network and effectively addresses the gradient vanishing and degradation issues that occur as network depth increases. The calculation formula for the residual block is as follows:

$$y = F(x, \{W_i\}) + x \quad (2)$$

where x and y represent the input and output vectors, and $F(x, \{W_i\})$ denotes the residual mapping to be learned.

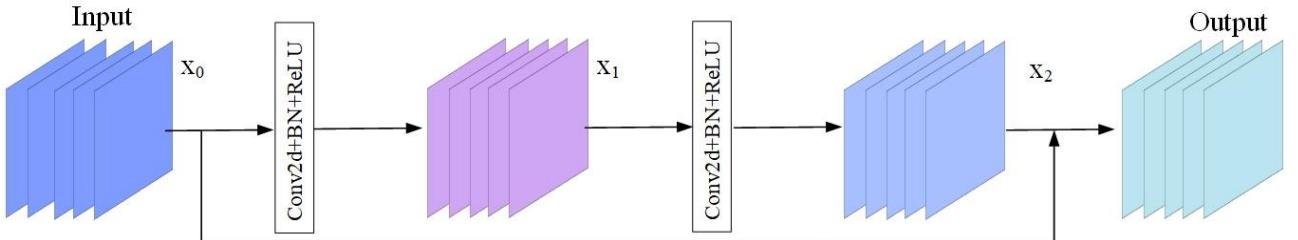


Fig. 5. Basic block in ResNet network.

B. Multi-scale feature extraction module

The original residual structure is insufficient for effectively extracting the multi-scale features of vibration signals. In contrast, the multi-scale feature extraction module enhances the representation of multi-scale features by constructing

hierarchical residual connections within a single residual block, thereby expanding the receptive field of each network layer. This significantly improves the model's ability to capture complex signals in time-frequency images and better extract features across different frequencies and time

domains. The multi-scale feature extraction module replaces a single 3×3 filter group with multiple smaller filter groups to achieve this effect. This module first applies a 1×1 convolution to the input feature map, uniformly dividing it into s feature subsets, denoted as x_i , where $i \in \{1, 2, \dots, s\}$. Each feature subset x_i has the same spatial resolution as the original feature map but only $1/s$ of the channels. Except for x_1 , each x_i is associated with a 3×3 convolution kernel, denoted as $K_i()$, with its output represented as y_i . Before being input to $K_i()$, each x_i is added to the output of the previous layer y_{i-1} to enable cross-layer feature fusion. To increase the number of feature subsets while reducing the number of parameters, the 3×3 convolution for x_1 is omitted. Therefore, the output y_i can be expressed as:

$$y_i = \begin{cases} x_i & i = 1; \\ K_i(x_i) & i = 2; \\ K_i(x_i + y_{i-1}) & 2 < i \leq s; \end{cases} \quad (3)$$

C. Framework for tool wear state recognition

First, the sensor vibration signal data is preprocessed to remove outliers and eliminate segments corresponding to tool entry and exit. The processed vibration signals are then transformed into time-frequency representations using CWT, enabling the extraction of time-domain features. The resulting time-frequency images are subsequently divided into training and validation sets in a 7:3 ratio. Finally, the constructed convolutional neural network, Multi-scale ResNet, is used to recognize tool wear conditions, as shown in Fig. 6.

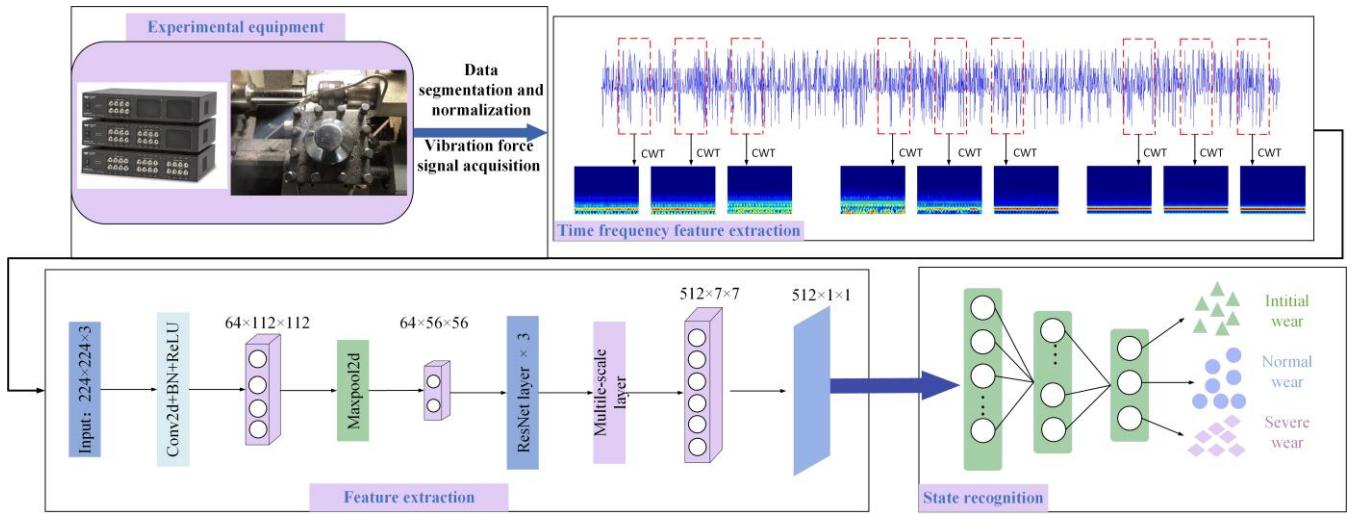


Fig. 6. Overall framework of the proposed methodology.

D. A Multi-scale resnet-based approach for tool wear condition recognition

The time-frequency images obtained through CWT are first uniformly resized to RGB images with dimensions of $224 \times 224 \times 3$, where 224×224 denotes the spatial resolution and 3 represents the three color channels. These images are then input into the model, sequentially passing through the Stem layer, the ResNet module, the multi-scale feature extraction module (Multi-scale ResNet), and the feature reduction layer (Reduction Layer). In the Stem layer, the input image first passes through a 7×7 convolutional layer, followed by BN and a ReLU activation function for initial feature extraction. This is followed by a max pooling layer to reduce the spatial dimensions of the feature maps. Next, the feature maps are fed into the backbone ResNet network, which consists of multiple residual units (Basic Blocks), which further extract high-level semantic features. Each residual unit contains two 3×3 convolutional layers

connected by a shortcut to enable residual learning, enhancing training stability and generalization performance. After the third residual stage (Layer 3), a multi-scale feature fusion module (Multi-scale Layer) is introduced. This module uses a parallel architecture to extract features at multiple receptive field scales. It integrates local and global information through multiple convolutional paths, including 1×1 convolutions, 3×3 convolutions, and dilated convolutions. The outputs from these branches are concatenated along the channel dimension to fuse multi-scale features, effectively enhancing the model's ability to represent the diverse characteristics of tool wear. After feature fusion, a global average pooling layer compresses the 2D feature maps into a 1D vector, which is further mapped into a 128-dimensional feature representation via a dimensionality reduction layer. Finally, this feature vector is fed into a fully connected layer and classified into different tool wear categories using the Softmax activation function, as shown in Fig. 7.

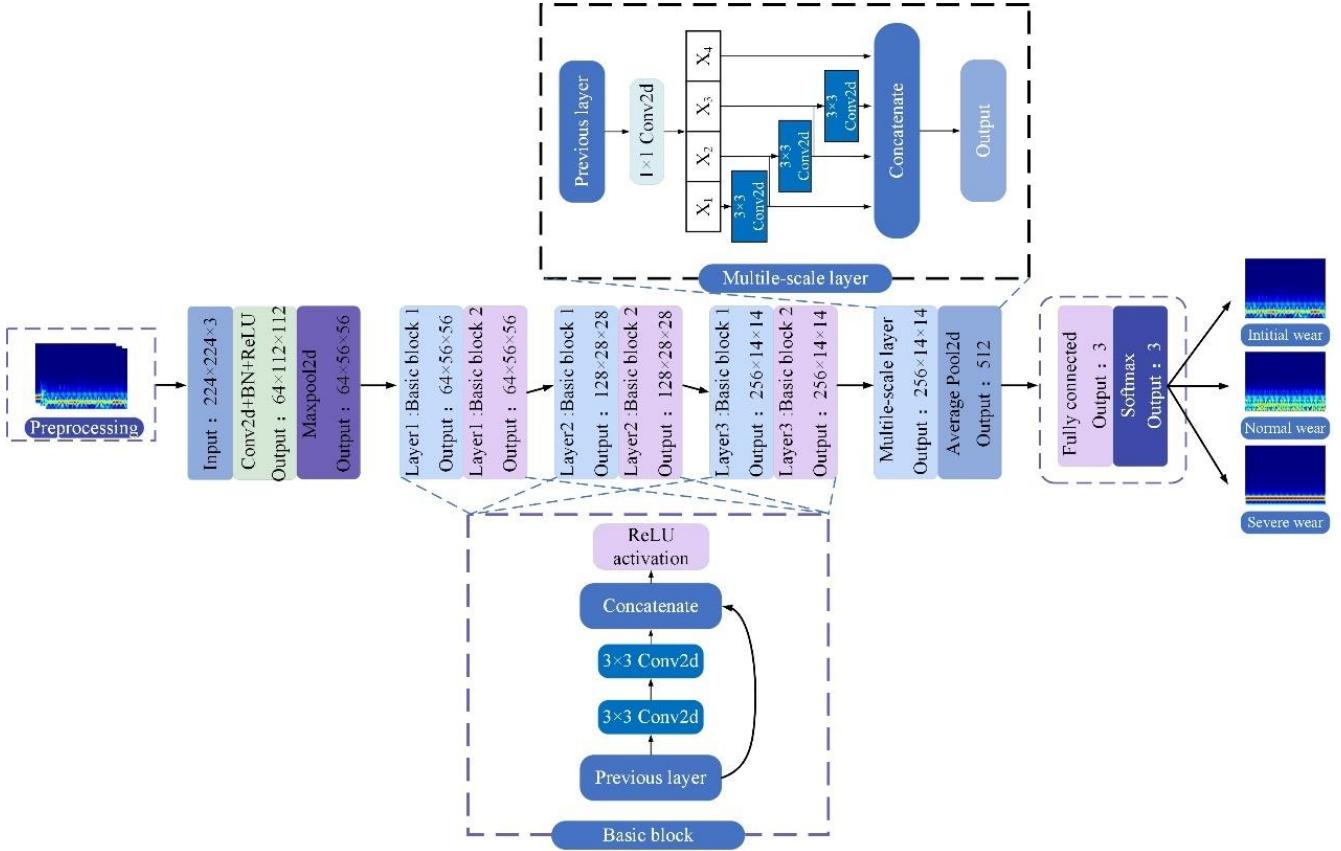


Fig. 7. Overview of a Multi-scale ResNet-based model for tool wear condition recognition.

4. MODEL TRAINING AND DATA VISUALIZATION

In this study, model training and validation were performed on a workstation equipped with an Nvidia GeForce A6000 GPU and running Windows 11. The training environment used PyCharm 2023.3.1 and PyTorch 2.1.0+cu118. A mini-batch training strategy was adopted, and the learning rate was dynamically adjusted during training using a Warmup schedule followed by a cosine annealing strategy. The AdamW optimizer was used for iterative optimization of the model parameters, with a weight decay term included to mitigate overfitting. The learning rate was set to 0.0002, with cosine annealing applied for learning rate scheduling, iterating every 50 epochs. The Adam optimizer was used to update the loss function during training. The maximum number of iterations for model training was set to 100, with a batch size of 128. The input to the Multi-scale ResNet model consists of time-frequency images representing tool wear features, obtained through CWT. The model output corresponds to three tool wear states, so the Softmax function was used for multi-class classification, with the following expression:

$$Softmax = \frac{e^{v_i}}{\sum_{j=1}^k e^{v_j}} \quad (4)$$

where k denotes the number of neural network outputs; v is the output vector; i denotes the category of cutting tool wear; v_j is the j^{th} category value in the output vector v .

To evaluate the performance difference between Multi-scale ResNet and other models, the prediction results of the classification model were compared with the true results by calculating the numbers of true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN). Commonly used evaluation metrics such as Confusion matrix, Recall, Precision, F1 score, and Accuracy were calculated using these values, as shown in (5-8):

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$F1score = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (7)$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (8)$$

A. Results and discussion

The model was trained using 70 % of the dataset as the training set, while 30 % was used as the validation set to assess the model's effectiveness. In this study, model performance was evaluated using accuracy, recall, precision, and F1 scores, as shown in Table 2.

Table 2 presents the performance evaluation metrics of the model after training, based on 10 training runs. These metrics include accuracy, recall, and precision for different stages of tool wear. The data show that the Multi-scale ResNet model achieves significant improvements in both the Initial wear and Normal wear stages. Data visualization indicates that the average accuracy, recall, and F1 score of the Multi-scale ResNet model are 2.9 %, 1.6 %, and 2.4 % higher, respectively, than those of the CNN–SVM model. Compared to the Transformer model, the Multi-scale ResNet model outperforms it by 6.1 %, 6.3 %, and 6.2 %, and by 3.4 %, 4.4 %, and 4.0 % compared to the ResNet model.

From Table 2 and the Confusion matrix shown in Fig. 8, it is clear that most misclassifications occur in the initial wear and normal wear stages. This is because the feature maps of normal wear include characteristics from both the initial wear and rapid wear stages, which significantly impact the model's recognition accuracy. In most machining scenarios, it is

recommended to replace the cutting tool before severe wear occurs. This requires the model to have higher accuracy in recognizing rapid tool wear. The Multi-scale ResNet model, through its multi-scale feature extraction module, effectively leverages vibration signal features related to tool wear, helping the model better distinguish between initial wear and normal wear.

Fig. 9 shows the classification results of four models: Multi-scale ResNet, CNN–SVM, Transformer, and ResNet. In the figure, the green line represents initial wear, the yellow line represents normal wear, the red line represents severe wear, and the blue line represents the model's predicted results. By observing the chart, it is evident that all four models perform well in distinguishing severe wear. However, when identifying initial wear and normal wear, they often make misclassifications. In contrast, the Multi-scale ResNet model significantly reduces misclassifications between these two stages, improving the model's recognition accuracy.

Table 2. Comparison of experimental results with other models.

Models		Accuracy [%]	Recall [%]	Precision [%]	F1 score [%]
CNN–SVM ^[5]	Initial wear	--	90.9	83.3	87.0
	Normal wear	--	86.8	93.9	90.2
	Severe wear	--	100.0	95.5	97.7
	Average	90.6	92.6	90.9	91.6
Transformer ^[6]	Initial wear	--	78.8	76.4	77.6
	Normal wear	--	84.9	86.5	85.7
	Severe wear	--	100.0	100.0	100.0
	Average	86.0	87.9	87.7	87.8
ResNet	Initial wear	--	78.8	83.8	81.3
	Normal wear	--	90.6	87.3	88.9
	Severe wear	--	100.0	100.0	100.0
	Average	88.8	89.8	90.4	90.0
Multi-scale ResNet	Initial wear	--	90.0	87.1	88.5
	Normal wear	--	92.4	94.2	93.3
	Severe wear	--	100.0	100.0	100.0
	Average	93.3	94.2	93.8	94.0

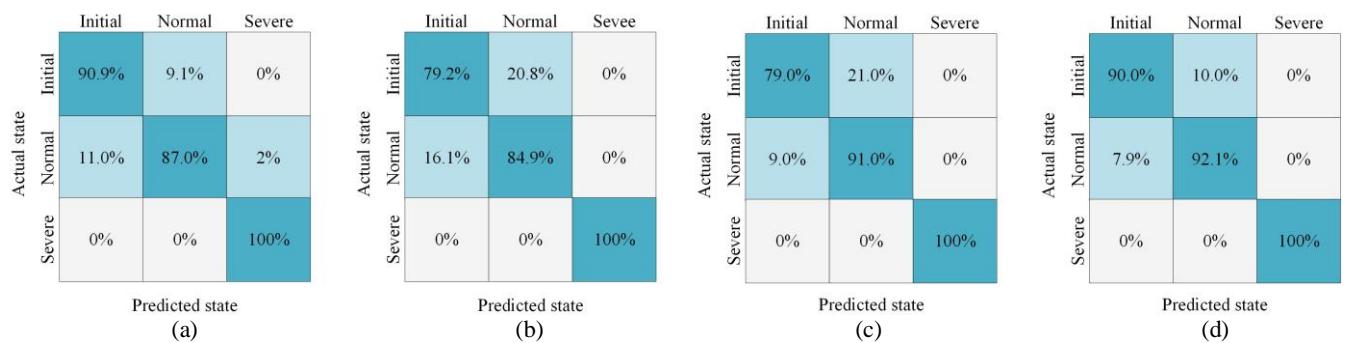


Fig. 8. Confusion matrix (a) CNN–SVM; (b) Transformer; (c) ResNet; (d) Multi-scale ResNet.

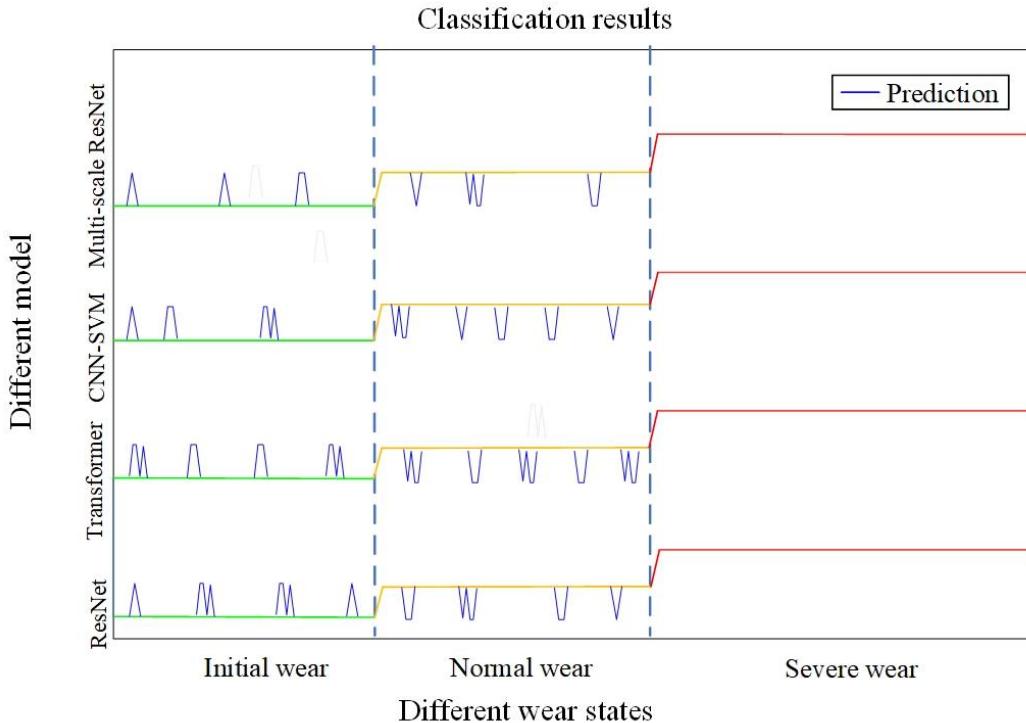


Fig. 9. Recognition results of the four models.

5. CONCLUSION

In this study, cutting vibration signals related to tool wear were obtained through γ -TiAl alloy cutting experiments, and a deep learning model was constructed to identify and predict tool wear states. The main conclusions are as follows:

1. This paper proposes a tool wear state recognition model based on Multi-scale ResNet, which more accurately captures multi-scale features in vibration signals, thereby improving the accuracy of tool wear state identification.
2. The proposed model integrates a multi-scale feature extraction module with a deep residual connection network, effectively addressing issues such as vanishing gradients, exploding gradients, and insufficient feature extraction capabilities in deep learning models.
3. Based on the γ -TiAl alloy cutting experiment results, the Multi-scale ResNet model proposed in this paper outperforms traditional CNN-SVM, Transformer, and ResNet models in recognition accuracy during the initial wear and normal wear stages. The model achieves a prediction accuracy of 93.8 %, a recall of 94.2 %, and an F1 score of 94 %, with an average accuracy of 93.3 %. This model provides an effective solution for real-time monitoring and intelligent recognition of tool wear states, helping to improve the efficiency and quality of cutting processes and promoting the intelligent development of production processes.
4. Future research will focus on enhancing the model's generalization capability across multiple cutting conditions and various tool types, incorporating transfer learning and related techniques to ensure reliable

recognition under diverse machining scenarios. To meet the requirements of real-time monitoring and embedded deployment, efforts will be made to advance model lightweighting and edge-side adaptation through network architecture optimization, pruning, and quantization, aiming to achieve higher generalizability, stronger robustness, and improved real-time performance, thereby expanding the model's practical applicability in industrial production.

AUTHOR CONTRIBUTIONS

E.L.: Conceptualization, Methodology, Writing — original draft, Supervision; C.L. and Y.D.: Formal analysis, Investigation, Writing — original draft; B.Z.: Writing — review & editing; L.S.: Resources, Funding acquisition. All authors have read and agreed to the published version of the manuscript.

FUNDING

This research was funded by the National Natural Science Foundation of China (52265059, 52075135) and the Natural Science Foundation of Jiangsu Province (BK20230632).

ACKNOWLEDGMENT

The authors would like to thank the National Natural Science Foundation of China (52265059, 52075135) and the Natural Science Foundation of Jiangsu Province (BK20230632). The authors greatly appreciate the reviews, the suggestions from reviewers, and the editor's encouragement.

REFERENCES

[1] Pimenov, D. Y., Gupta, M. K., da Silva, L. R. R., Kiran, M., Khanna, N., Krolczyk, G. M. (2022). Application of measurement systems in tool condition monitoring of Milling: A review of measurement science approach. *Measurement*, 199, 111503. <https://doi.org/10.1016/j.measurement.2022.111503>

[2] Teti, R., Jemielniak, K., O'Donnell, G., Dornfeld, D. (2010). Advanced monitoring of machining operations. *CIRP Annals*, 59 (2), 717-739. <https://doi.org/10.1016/j.cirp.2010.05.010>

[3] Abellán-Nebot, J. V., Subirón, F. R. (2010). A review of machining monitoring systems based on artificial intelligence process models. *The International Journal of Advanced Manufacturing Technology*, 47 (1), 237-257. <https://doi.org/10.1007/s00170-009-2191-8>

[4] Cai, W., Zhang, W., Hu, X., Liu, Y. (2020). A hybrid information model based on long short-term memory network for tool condition monitoring. *Journal of Intelligent Manufacturing*, 31, 1497-1510. <https://doi.org/10.1007/s10845-019-01526-4>

[5] Wu, X., Li, J., Jin, Y., Zheng, S. (2020). Modeling and analysis of tool wear prediction based on SVD and BiLSTM. *The International Journal of Advanced Manufacturing Technology*, 106, 4391-4399. <https://doi.org/10.1007/s00170-019-04916-3>

[6] Wang, J., Ma, Y., Zhang, L., Gao, R. X., Wu, D. (2018). Deep learning for smart manufacturing: Methods and applications. *Journal of Manufacturing Systems*, 48, 144-156. <https://doi.org/10.1016/j.jmsy.2018.01.003>

[7] Nasir, V., Sassani F. (2021). A review on deep learning in machining and tool monitoring: Methods, opportunities, and challenges. *The International Journal of Advanced Manufacturing Technology*, 115, 2683-2709. <https://doi.org/10.1007/s00170-021-07325-7>

[8] Cheng, M., Jiao, L., Shi, X., Wang, X., Yan, P., Li, Y. (2020). An intelligent prediction model of the tool wear based on machine learning in turning high strength steel. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 234 (13), 1580-1597. <http://dx.doi.org/10.1177/0954405420935787>

[9] Liu, M.-K., Tseng, Y.-H., Tran, M.-Q. (2019). Tool wear monitoring and prediction based on sound signal. *The International Journal of Advanced Manufacturing Technology*, 103, 3361-3373. <https://doi.org/10.1007/s00170-019-03686-2>

[10] Guo, J., Li, A., Zhang, R. (2020). Tool condition monitoring in milling process using multifractal detrended fluctuation analysis and support vector machine. *The International Journal of Advanced Manufacturing Technology*, 110, 1445-1456. <https://doi.org/10.1007/s00170-020-05931-5>

[11] Mohanraj, T., Yerchuru, J., Krishnan, H., Nithin Aravind, R. S., Yameni, R. (2021). Development of tool condition monitoring system in end milling process using wavelet features and Hölder's exponent with machine learning algorithms. *Measurement*, 173, 108671. <https://doi.org/10.1016/j.measurement.2020.108671>

[12] Shi, C., Panoutsos, G., Luo, B., Liu, H., Li, B., Liu, X. (2019). Using multiple-feature-spaces-based deep learning for tool condition monitoring in ultraprecision manufacturing. *IEEE Transactions on Industrial Electronics*, 66 (5), 3794-3803. <https://doi.org/10.1109/TIE.2018.2856193>

[13] Ma, M., Sun, C., Chen, X., Zhang, X., Yan, R. (2019). A deep coupled network for health state assessment of cutting tools based on fusion of multisensory signals. *IEEE Transactions on Industrial Informatics*, 15 (12), 6415-6424. <https://doi.org/10.1109/TII.2019.2912428>

[14] Zhang, K., Chen, J., Zhang, T., Zhou, Z. (2020). A compact convolutional neural network augmented with multiscale feature extraction of acquired monitoring data for mechanical intelligent fault diagnosis. *Journal of Manufacturing Systems*, 55, 273-284. <https://doi.org/10.1016/j.jmsy.2020.04.016>

[15] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 770-778. <https://doi.org/10.1109/CVPR.2016.90>

[16] He, K., Zhang, X., Ren, S., Sun, J. (2016). Identity mappings in deep residual networks. In *Computer Vision – ECCV 2016*. Springer, LNIP 9908, 630-645. https://doi.org/10.1007/978-3-319-46493-0_38

[17] Zhang, K., Tang, B., Deng, L., Liu, X. (2021). A hybrid attention improved ResNet based fault diagnosis method of wind turbines gearbox. *Measurement*, 179, 109491. <https://doi.org/10.1016/j.measurement.2021.109491>

[18] Gabor, D. (1946). Theory of communication. Part 1: The analysis of information. *Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering*, 93 (26), 429-441. <https://doi.org/10.1049/ji-3-2.1946.0074>

[19] Burrus, C., Gopinath, R., Guo, H. (1997). *Introduction to Wavelets and Wavelet Transforms: A Primer*. Pearson, ISBN 978-0134896007.

[20] Claasen, T. A. C. M., Mecklenbrauker, W. F. G. (1980). The Wigner distribution - a tool for time-frequency signal analysis. Part II: Discrete time signals. *Philips Journal of Research*, 35 (4), 276-300.

Received May 19, 2025
Accepted November 12, 2025