

## Article

# Tool Wear State Recognition Based on One-Dimensional Convolutional Channel Attention

Zhongling Xue<sup>1,2</sup>, Liang Li<sup>2</sup>, Ni Chen<sup>1,2,\*</sup>, Wentao Wu<sup>2</sup>, Yuhang Zou<sup>2</sup> and Nan Yu<sup>3</sup>

<sup>1</sup> Key Laboratory of Advanced Manufacturing and Intelligent Technology, Ministry of Education, Harbin University of Science and Technology, Harbin 150080, China

<sup>2</sup> College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics & Astronautics, Nanjing 210016, China

<sup>3</sup> Institute of Materials and Processes, School of Engineering, University of Edinburgh, Edinburgh EH8 9BT, UK

\* Correspondence: ni.chen@nuaa.edu.cn

**Abstract:** Tool wear state recognition is an important part of tool condition monitoring (TCM). Online tool wear monitoring can avoid wasteful early tool changes and degraded workpiece quality due to later tool changes. This study incorporated an attention mechanism implemented by one-dimensional convolution in a convolutional neural network for improving the performance of the tool wear recognition model (1DCCA-CNN). The raw multichannel cutting signals were first preprocessed and three time-domain features were extracted to form a new time-domain sequence. CNN was used for deep feature extraction of temporal sequences. A novel 1DCNN-based channel attention mechanism was proposed to weigh the channel dimensions of deep features to enhance important feature channels and capture key features. Compared with the traditional squeeze excitation attention mechanism, 1DCNN can enhance the information interaction between channels. The performance of the model was validated on the PHM2010 public cutting dataset. The excellent performance of the proposed 1DCCA-CNN was verified by the improvement of 4% and 5% compared to the highest level of existing research results on T1 and T3 datasets, respectively.

**Keywords:** tool condition monitoring; tool wear state recognition; 1D Convolution; channel attention



**Citation:** Xue, Z.; Li, L.; Chen, N.; Wu, W.; Zou, Y.; Yu, N. Tool Wear State Recognition Based on One-Dimensional Convolutional Channel Attention. *Micromachines* **2023**, *14*, 1983. <https://doi.org/10.3390/mi14111983>

Academic Editors: Benny C. F. Cheung and Chenyang Zhao

Received: 30 September 2023

Revised: 19 October 2023

Accepted: 21 October 2023

Published: 26 October 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The high-speed contact between the cutting edge and the workpiece surface during the cutting process leads to increasing tool wear, which can affect the machining accuracy [1]. In particular, excessive tool wear can lead to waste due to unmet manufacturing requirements for the workpiece [2]. Tool wear state recognition allows real-time monitoring of tool wear for tool replacement and maintenance to avoid production interruptions and cost losses due to severe tool wear [3]. Therefore, it is important to realize the tool wear state recognition.

Tool condition monitoring (TCM) tasks are mainly categorized into direct and indirect methods. The direct method directly observes the wear area of the tool and measures the tool wear through the microscope, which provides more intuitive and accurate results. With the advancement of intelligent algorithms, many researchers are currently realizing the segmentation of the tool wear area by means of machine vision to achieve the measurement of tool wear [4–6]. However, the direct method requires stoppage for tool observation, which reduces the efficiency of machining. Moreover, chips and cutting fluids affect the observation of tool wear areas. Therefore, the direct method is not suitable for actual machining. The indirect method is realized by monitoring cutting signals such as force signals [7], vibration signals [8], and acoustic emission signals [9] to monitor the tool condition. These signals are collected in real time by sensors mounted on the workpiece or spindle, etc., which have less influence on the machining process and are used in more application scenarios. Therefore, it is important to study the mapping algorithm of cutting signal and tool state for TCM system.

With the development of artificial intelligence technology, more and more intelligent algorithms are applied to tool wear monitoring tasks [10]. Qin et al. [11] used stacked sparse self-coding networks for tool wear monitoring. Duan et al. [12] converted cutting signals into time-frequency maps by short-time Fourier transform, then feature extraction by PCANet and GA-SPP, and finally tool wear prediction by SVM. Wei et al. [13] screened the sensitive low-dimensional features of force signals by whitening variational mode decomposition (WVMD) and Joint information entropy (JIE), and optimal-path forest (OPF) was used as a classifier to realize the tool wear state recognition. Chan et al. [14] achieved tool wear state monitoring by extracting overall and local features of the signal through LSTM. Zhou et al. [15] utilized graph neural networks to achieve tool wear state monitoring with small samples. Hou et al. [16] used lightweight networks to achieve tool wear state monitoring after augmenting and balancing unbalanced data via WGAN-GP. These intelligent algorithms achieved excellent results in the task of tool wear monitoring.

Attentional mechanisms can enhance the perceptual ability, adaptability and interpretability of neural networks [17]. With the use of a large number of sensors, the channel dimensions of cutting signals are increasing, and the attention mechanism can effectively combine data from different channels to improve the performance and interpretability of the model. Therefore, more and more researchers are applying attention mechanisms to tool wear monitoring tasks. Li et al. [18] transformed the force signal into a time-frequency map by continuous wavelet transform and established a channel space attention mechanism to realize tool wear state monitoring. Zeng et al. [19] fused and converted the multi-sensing data into images, and selected the information in the channel and spatial domains through the attention mechanism to realize the deep feature extraction of tool wear. Hou et al. [20] combined channel attention with multiscale convolution to extract multiscale spatial-temporal features in cutting signals for tool wear monitoring. Zhou et al. [21] proposed Dual Attention Mechanism Network to learn pixel feature dependency and inter-channel correlation respectively. He et al. [22] proposed a cross-domain adaptation network based on attention mechanism to realize the tool wear state recognition and prediction. Dong et al. [23] combined the channel attention mechanisms of CaAt1 and CaAt5 with ResNet18 for tool wear monitoring. Guo et al. [24] focused on the important parts of the sequence information through an attention mechanism to achieve a multi-step tool wear prediction. Lai et al. [25] used the attention mechanism for weighted fusion of frequency and spatial features and performed an interpretability analysis of tool wear state recognition algorithms. Feng et al. [26] captured the complex spatial-temporal relationship between tool wear and features by weighting features in both the spatial and temporal dimensions through an attentional mechanism. Huang et al. [27] fused different scale features extracted by CNN through an attention mechanism to achieve tool wear prediction from multi-sensor data.

Three attention modules were mentioned in the above study. The first attention module type is a weighted fusion of the results of different layers of networks. By fusing different networks it is true that the recognition accuracy can be improved, but it also increases the overall computational cost of the network. The second type of attention module is the conversion of a temporal signal into a 2D time-frequency map by time-frequency transformations, and to perform feature extraction by channel and spatial attention mechanism in the field of image classification. However, the data structure of images is more complex compared to sequences and requires a network with larger number of parameters. The last attention module type is a channel attention mechanism for multidimensional sequences. The weighted representation of features is achieved by learning the weights of each channel in the feature map. However, current researchers mainly realize the channel attention by squeezing excitation on the channels of multidimensional sequences, which is inefficient in capturing the dependencies between the channels in this mechanism [28].

The attention operation is realized with 1D convolution instead of fully connected squeezing excitation. The coverage of cross-channel interactions is controlled by the size of

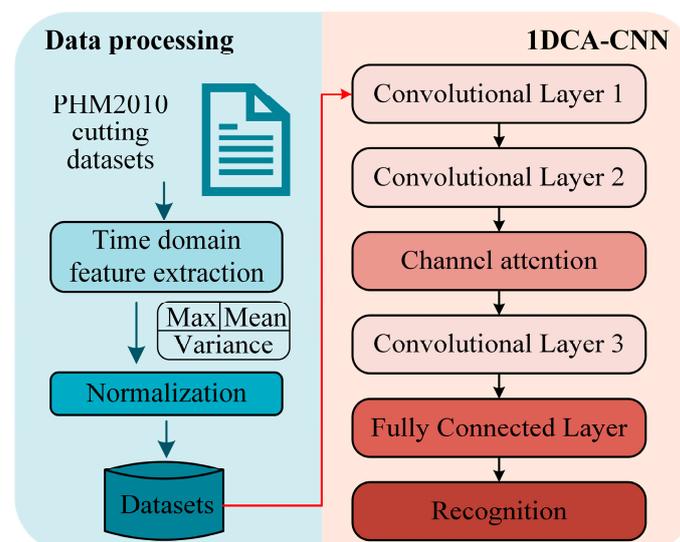
the 1D convolutional kernel, which improves the inter-channel dependencies. The main contributions of this paper are as follows:

- (1) A 1DCCA-CNN model was proposed to realize the tool wear state recognition. The features of the cutting signal were extracted by a one-dimensional convolutional neural network. A novel channel attention was proposed. The inter-channel weight relationships are learned by one-dimensional convolution rather than squeezed excitation of fully connected layers, which can improve the interaction ability between different channels, and extract the features strongly related to the tool wear to improve model performance.
- (2) Validation was performed on the PHM2010 public cutting dataset. Cross-validation datasets with different groups were designed. The proposed model's performance on the cross-validation dataset was evaluated by confusion matrix, accuracy, precision and recall. The superiority of the proposed model was verified by comparing with other models.

## 2. Structure of the Proposed Model

### 2.1. Overall Framework

The overall framework consists of data processing and 1DCCA-CNN model (Figure 1). The PHM2010 public cutting dataset [29] was used for model training and validation. The time-domain features (max, mean and variance) of the raw signals in the PHM2010 cutting dataset were extracted to form the training dataset and test dataset. The proposed 1DCCA-CNN composed of three convolutional layers, a channel attention layer and fully connected layers. The first convolutional layer was used to recognize low-level features and the second convolutional layer was used to recognize mid-level features. At this time, the channel dimension of the feature map was large, and a channel attention layer needed to be accessed. This channel attention layer adopted 1DCNN to realize the calculation of channel weights, which can improve the information interaction between channels. The third convolutional layer was used for extract high-level features. Finally, the feature map was mapped to the category dimension by a fully connected layer to realize the tool wear state recognition.

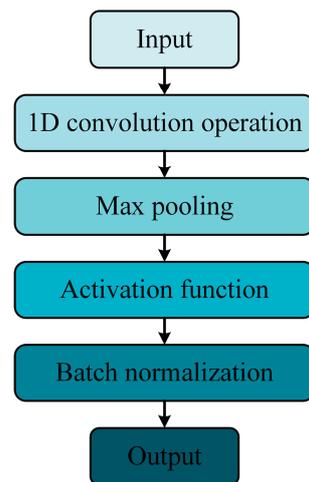


**Figure 1.** The overall structure of the proposed model.

### 2.2. 1D Convolutional Neural Network Layer

1D Convolutional Neural Network (1D CNN) is a neural network architecture for processing one-dimensional sequential data in deep learning, which has the advantages of localized feature extraction, translational invariance and hierarchical feature extraction. In the tool wear state recognition task, some of the key features would be distributed in the

local area, therefore, the superior local feature extraction capability of convolutional neural can effectively capture these features. The translational invariance of the convolutional layer can also automatically learn the tool wear characteristics at different positions. Meanwhile, the bottom convolutional layer of the convolutional neural network can capture low-level features, and the high-level convolutional layer can capture deeper abstract features, thus improving tool wear state recognition accuracy. There are a total of three 1D CNN layers in this study which are all composed of convolution operation, activation function, pooling layer and batch normalization (Figure 2).



**Figure 2.** The structure of a 1D convolutional layer.

1D convolutional operations are used to capture localized features in the input sequence and are applicable to a variety of sequence data such as text, audio, and time series. The convolution kernel is slid over the input sequence and the dot product of the convolution kernel with the input is computed to extract features at different locations (Figure 3). This allows the model to automatically capture localized features in the input sequence without explicitly specifying the location of the features. The formula for the convolution operation is shown in Equation (1).

$$y_i = \sum_{j=0}^{M-1} x_{i+j} \times h_j \quad (1)$$

where  $y_i$  denotes the  $i$ th element of the output sequence of the convolution operation,  $i = 0, 1, \dots, (N - M)$ ,  $M$  is the convolution kernel size, and  $N$  is the length of the input sequence;  $x_{i+j}$  is the  $i + j$ th element in the input sequence  $x$ ; and  $h_j$  is the  $j$ th weight in the convolution kernel  $h$ .

The pooling operation downsamples the feature map to reduce the size. The pooling operation also extracts the most salient features, reduces the number of parameters in the model, and helps prevent overfitting. The maximum pooling layer was used in this study as shown in Equation (2):

$$y_j = \max(a[j : j + K]) \quad (2)$$

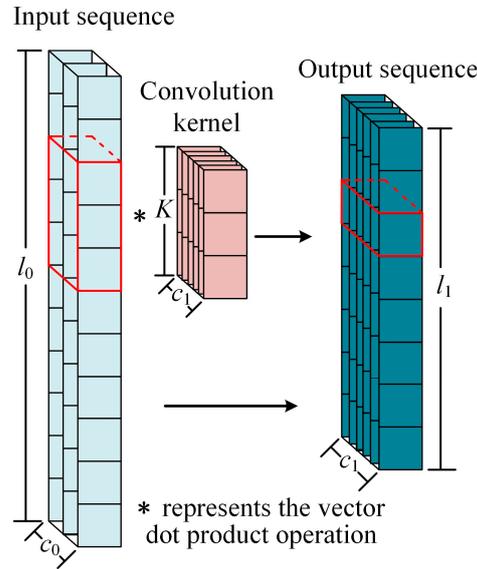
where  $y_i$  denotes the  $i$ th element of the output sequence of the pooling operation,  $i = 0, 1, \dots, \frac{N}{K}$ ,  $K$  is pooling kernel size,  $N$  is the length of the input sequence;  $a[j : j + K]$  is the  $K$  consecutive elements from the input sequence  $a$ ,  $j = 0, 1, \dots, N - K$ .

The activation function introduces nonlinear properties that allow neural networks to learn and represent more complex functional relationships. Among them, RELU is widely used in convolutional neural networks due to its high computational efficiency

while mitigating the gradient vanishing problem and its contribution to the enhancement of the generalization ability of the model. The formula of RELU is shown in Equation (3).

$$\text{RELU}(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} \quad (3)$$

where  $x$  is a point in the feature sequence.



**Figure 3.** The operation of one-dimensional convolution.

The main role of Batch Normalization (BN) is to accelerate the training of neural networks, to improve the stability of the training, as well as to reduce the gradient vanishing problem. The BN contributes to the training of the network by normalizing each small batch of data to adjust the distribution of the input data to a standard normal distribution with a mean of zero and a standard deviation of one. For input data with batch size  $m$  and channel size  $C$ , the calculation process of BN is as follows:

Step 1: Calculate the sample mean for each channel  $c$

$$\mu_c = \frac{1}{m} \sum_{i=1}^m x_i^c \quad (4)$$

where  $x_i^c$  is the  $c$ th channel of the  $i$ th sample.

Step 2: Calculate the standard deviation of each channel  $c$

$$\sigma_c = \sqrt{\frac{1}{m} \sum_{i=1}^m (x_i^c - \mu_c)^2 + \epsilon} \quad (5)$$

where  $x_i^c$  is the  $c$ th channel of the  $i$ th sample;  $\epsilon$  is the smoothing term, which a very small positive number that prevents the divisor from going to zero, and is here set to  $10^{-5}$ .

Step 3: Normalize for each channel  $c$

$$\hat{x}_i^c = \frac{x_i^c - \mu_c}{\sigma_c} \quad (6)$$

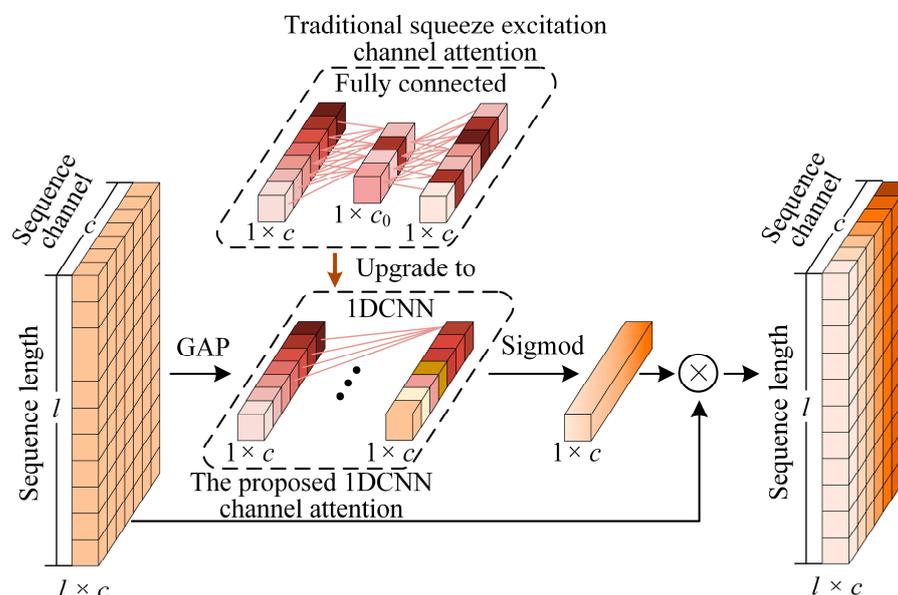
Step 4: Perform a linear transformation on the normalized values

$$y_i^c = \gamma \hat{x}_i^c + \beta \quad (7)$$

where  $y_i^c$  is the normalized result of the final output,  $\gamma$  is the learnable scaling factor, and  $\beta$  is the learnable translation factor.

### 2.3. 1D Channel Attention

The main significance of the channel attention is to enhance the model’s adaptability to different channel features. By automatically adjusting the weight of each channel, the model can better capture key features, reduce redundant information, and improve model performance and generalization ability. The channel attention mechanism traditionally applied to multidimensional sequential tasks is realized by squeezing excitation through a fully connected layer (Figure 4). This study takes advantage of the good cross-channel interaction capability that convolution has, and proposes to use 1DCNN instead of the commonly used squeezing excitation of the fully connected layer to realize the channel attention mechanism (Figure 4), which can enhance the information interaction capability between channels. For the input sequence of  $L \times C$ , the implementation process of the 1DCNN attentional mechanism is as follows:



**Figure 4.** Traditional squeeze excitation channel attention and the proposed structure of a channel attention mechanism based on 1DCNN.

Step 1: Perform one-dimensional global average pooling on the input sequence

$$GAP_c = \frac{1}{L} \sum_{i=1}^L x_{i,c} \tag{8}$$

where  $x_{i,c}$  is the eigenvalue of channel  $c$  of the input sequence at position  $i$ ,  $c = 0, 1, \dots, C$ ,  $C$  is the channel size of the input sequence, and  $L$  is the length of the input sequence.

Step 2: Perform one-dimensional convolution operation on the features in the channel dimension

$$y_i = \sum_{j=0}^{M-1} c_{i+j} \times h_j \tag{9}$$

where  $y_i$  denotes the  $i$ th element of the output sequence of the convolution operation,  $i = 0, 1, \dots, C$ ,  $C$  is the channel size of the input features,  $M$  is convolution kernel size;  $c_{i+j}$  is the eigenvalue on the  $i + j$ th channel in the feature  $c$ ;  $h_j$  is the  $j$ th weight in the convolution kernel  $h$ .

Step 3: Sigmoid calculations are performed on the results of the one-dimensional convolution to obtain the weights of each channel

$$\sigma(y) = \frac{1}{1 + e^{-y}} \quad (10)$$

where  $y$  is the sequence of features after one-dimensional convolution.

Step 4: The complete channel attention is realized after loading the obtained weights on each channel.

#### 2.4. Full Connected Layer

The fully connected layer integrates the feature information extracted from the previous convolutional layers in the convolutional neural network into a global feature vector and is used in the final classification or regression task. The neurons in the fully connected layer are connected to all the neurons in the previous layer, and by learning the weight parameters, the nonlinear relationships in the data can be captured and the expressive power of the network can be improved. In this study the fully connected layer consists of two linear mapping layers, a RELU activation function and a Sigmoid (Figure 5). Linear mapping layer 1 maps global feature vectors to lower dimensions. The RELU activation layer implements the nonlinearization. Linear mapping layer 2 maps global feature vectors to the classification dimension. Sigmoid maps the feature values between 0 and 1 to get a score for each category to realize the tool wear state recognition. The specific process is as follows:

$$Linear1(x) = xw_1 + b_1 \quad (11)$$

$$Linear2(x) = RELU(Linear1(x))w_2 + b_2 \quad (12)$$

$$Out(x) = Sigmoid(Linear2(x)) \quad (13)$$

where  $w$  is the weight matrix;  $b$  is the bias;  $x$  is the input vector; the formula for *RELU* is shown in Equation (3) and the formula for *Sigmoid* is shown in Equation (10).

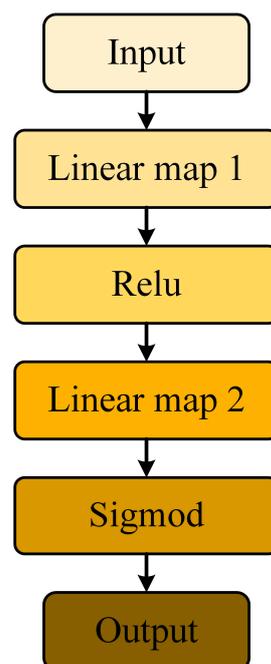


Figure 5. The structure of the fully connected layer.

### 3. Experiment

#### 3.1. Experiment Set

The PHM2010 public cutting dataset was used to validate the performance of the model. There were six sets of cutting experiments in the PHM2010 public cutting dataset for the full life cycle of the tool, but only three sets, C1, C4, and C6, gave the values of the tool wear after each tool travel. Therefore, only these three sets of experimental data were used. The cutting condition is shown in Figure 6. Roder Tech RFM769 high speed CNC machine was used for the experiment. The cutting work piece was stainless steel. The tool used was 3-flute ball cutters. The cutting mode was dry milling and side milling. The length of each pass was 108 mm. The cutting parameters are shown in Table 1.

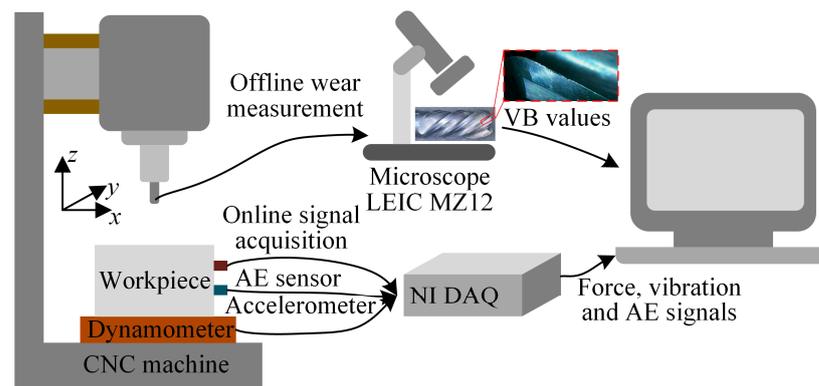


Figure 6. PHM2010 public cutting dataset working conditions.

Table 1. Experimental parameters for the PHM2010 public cutting dataset.

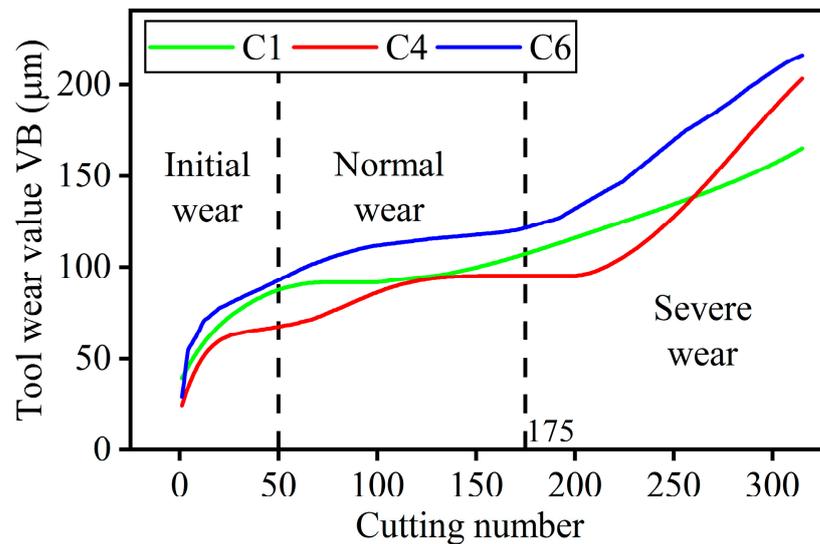
Experimental Parameters	Size
Spindle speed (RPM)	10,400
Feed rate (mm/min)	1555
Cutting depths (mm)	0.2
Cutting width (mm)	0.125
Sampling frequency	50 kHz

Cutting forces were measured in three directions during cutting using a Kistler quartz 3-component platform dynamometer. Three Kistler piezoelectric accelerometers were also deployed to measure triaxial vibration of the workpiece. The Kistler acoustic emission sensor was used to measure the high-frequency stress waves generated during the cutting process. The data collected by these three sensors were unified by the NI data acquisition card to form a 7-channel signal, which were  $x$ -direction cutting force,  $y$ -direction cutting force,  $z$ -direction cutting force,  $x$ -direction vibration,  $y$ -direction vibration,  $z$ -direction vibration and acoustic emission signal. After each pass, the average wear bandwidth of the flank face VB of each insert was measured using the LEIC MZ12 Microscope as the wear value of the insert. In order to intuitively illustrate the areas of tool wear measured in the dataset, The tool image taken at other factory was added to the Off-line measurement section of Figure 6. The average wear value of the three inserts was used as the wear value of the whole tool. The full life cycle tool wear for the three sets of experiments is shown in Figure 7.

#### 3.2. Data Preparation

Tool wear can be divided into initial wear, normal wear and severe wear according to the rate of change of wear. In the early stage of tool use, the surface of the newly sharpened tool is relatively rough. At this time, the contact stress is high and the wear is faster, this stage is defined as the initial wear of the tool. With the increase in cutting time, the cutting process tends to stabilize, the wear is flat, this stage is defined as the normal wear stage of

the tool. When the tool becomes blunt with the increase of use time, the cutting performance of the tool decreases sharply and the wear accelerates again, this stage is defined as the severe molding stage of the tool. According to the slope of the wear curve of the full life cycle of the tool in Figure 7, the tool was divided into three wear stages. The 1st pass to the 50th pass was defined as initial wear, the 51st pass to the 175th pass was defined as normal wear, and the 176th pass to the 315th pass was defined as severe wear.



**Figure 7.** Tool Full life cycle wear curve for in PHM2010 dataset.

As shown in Figure 8, we preprocessed the raw signals collected by the sensors. The raw signal consisted of a total of seven channels. A non-overlapping window was divided for each channel, and then time-domain features were extracted for the signals in each window, which were the maximum value, the mean value and the variance. The three extracted features were then connected in the temporal direction to form a new input sequence of dimension  $\mathbb{R}^{100 \times 21}$ . Due to the different units and value ranges of data collected by different sensors, Min-Max normalization was used to eliminate the difference in magnitude between different channels in order to avoid the model being more sensitive to certain features and to accelerate the convergence of the model. The labels of the sequences are defined for training based on the fact that they are in the wear stage. Due to the uneven amount of data under the three types of tool wear, especially the normal and severe wear stages were much larger than the initial wear, which would cause a reduction in the generalization ability of the model. Therefore, the raw signals of the initial wear stage were upsampled by repartitioning the window randomly. The normal and severe wear stage samples were randomly selected for downsampling. Finally, the sample size of each wear stage for each of the three sets of experiments, C1, C4 and C6, was one thousand, totaling three thousand samples for each set of experiments. To better validate the performance of the proposed model, three sets of crossover datasets divided by experimental groups were designed. The three sets of crossover datasets are shown in Table 2.

**Table 2.** Three cross-validation datasets.

Datasets	Training Datasets	Test Datasets
T1	C1 + C4	C6
T2	C4 + C6	C1
T3	C6 + C1	C4

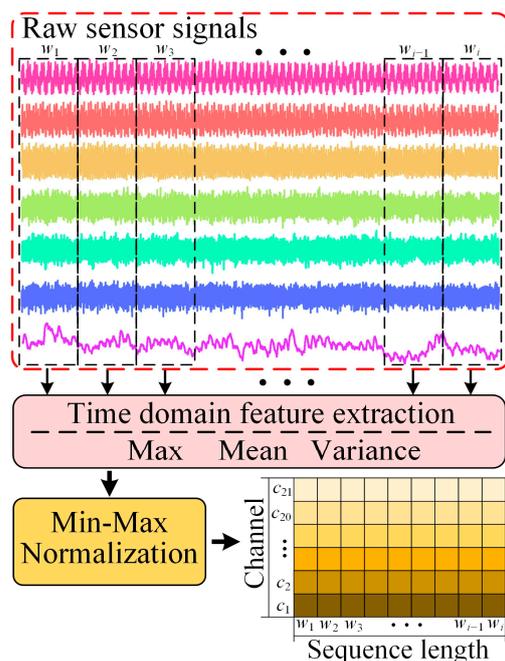


Figure 8. The flow of raw sensor data processing.

3.3. Model Parameters and Training Parameter Settings

Training was performed in Windows 10 platform using GPU. The graphics card was RTX 3090 and the CPU was AMD EPYC 7624. The deep learning network was built by Pytorch 1.11.0 framework. Python version was 3.8. Cuda version was 11.3. Adam was used as the optimizer for training, and improved the generalization of the model by setting weight\_decay to prevent overfitting in model training. The loss function was calculated using the cross-entropy function. The parameters of each layer of the proposed model are shown in Table 3. The hyperparameters for model training are shown in Table 4. The training dataset loss curve and test dataset accuracy curve of the proposed model during training are shown in Figure 9. It was observed that the model started to converge gradually at the 40th round of training, and the accuracy of the test dataset quickly reached a high level.

Table 3. The parameters of each layer of the proposed model.

Layer Name	Description	Parameters Setting
Convolutional layer 1	1D Convolution	filters = 42, kernel size = 3, padding = 0, stride = 1, activation = RELU, out size = 98 × 42
	1D Max pooling	pool size = 2, padding = 0, stride = 2, out size = 49 × 42
	Batch normalization	feature number = 42, out size = 49 × 42
Convolutional layer 2	1D Convolution	filters = 84, kernel size = 3, padding = 0, stride = 1, activation = RELU, out size = 47 × 84
	Max pooling	pool size = 2, padding = 0, stride = 2, out size = 23 × 84
	Batch normalization	feature number = 84, out size = 23 × 84
Channel attention	GMP	out size = 1 × 84
	1D Convolution	filters = 1, kernel size = 3, padding = 1, stride = 1, out size = 1 × 84
	Sigmoid	out size = 1 × 84
	Weighted multiplication	out size = 23 × 84
Convolutional layer 3	1D Convolution	filters = 168, kernel size = 3, padding = 0, stride = 1, activation = RELU, out size = 21 × 42
	1D Max pooling	pool size = 2, padding = 0, stride = 2, out size = 10 × 168
	Batch normalization	feature number = 168, out size = 10 × 168

Table 3. Cont.

Layer Name	Description	Parameters Setting
Fully connected layer	Flatten	out size = 1 × 1680
	Linear 1	out size = 1 × 64
	RELU	out size = 1 × 64
	Linear 2	out size = 1 × 3
	Sigmoid	out size = 1 × 3

Table 4. Hyperparameters for model training.

Hyperparameters	Size
Learning rate	0.0001
Batch size	256
Epoch	200
Weight_decay	0.05

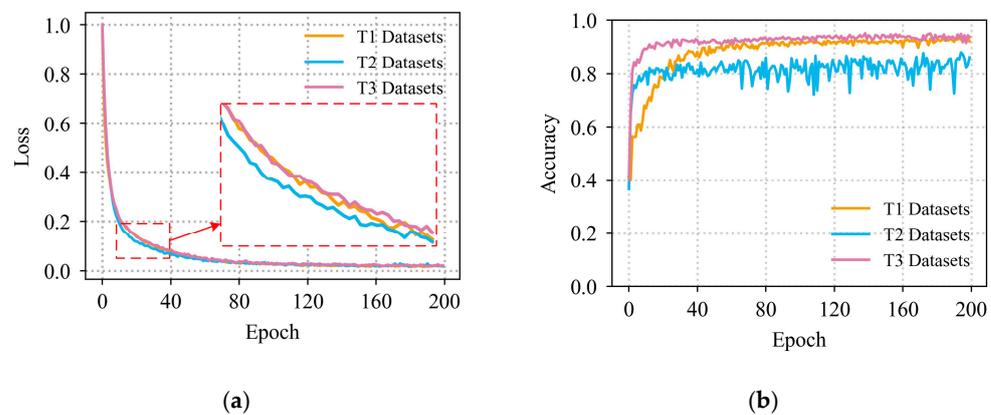


Figure 9. The proposed model training situation (a) Loss curves on the training dataset (b); Accuracy curve on the test dataset.

### 3.4. Analysis of Experimental Results

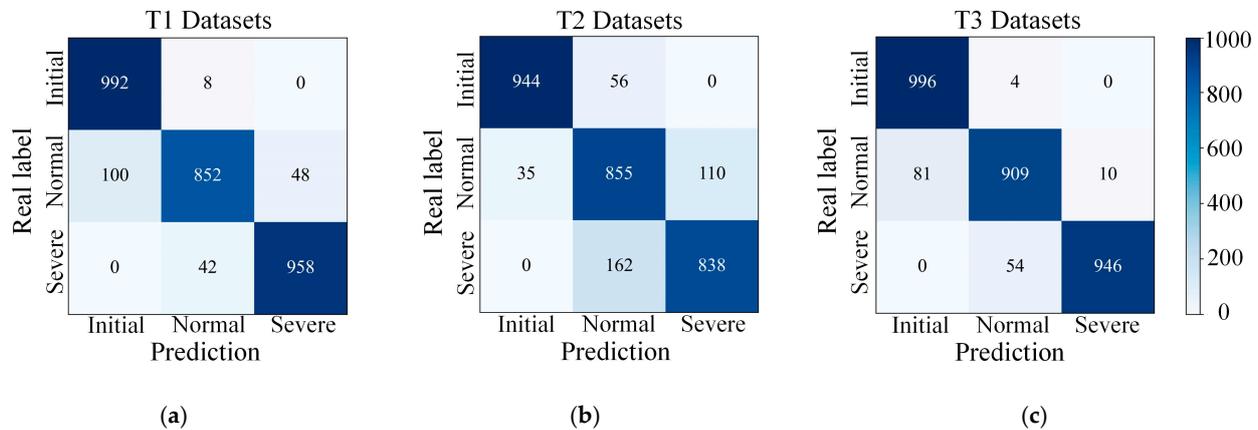
The confusion matrix of the recognition results of the proposed 1DCCA-CNN model on the three datasets is shown in Figure 10. To further evaluate the model’s performance, Accuracy, Precision, and Recall were used as metrics. The formulas for calculating the three metrics are as shown in Equations (13)–(15), respectively. Accuracy measures the accuracy of the model on the overall dataset. Precision focuses on the accuracy of the positive case prediction. Recall focuses on the ability of the model to recognize positive cases. In the tool wear recognition task, the initial and normal wear stages should avoid false alarms leading to early tool change and waste, while in the severe wear stage should try to recognize the severe wear state in time. Therefore, more attention should be paid to the model’s precision in the initial and normal wear phases, and more attention should be paid to the model’s recall in the severe wear phase. The proposed model’s evaluation metrics for the three datasets are presented in Table 5.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

$$Precision = \frac{TP}{TP + FP} \tag{15}$$

$$Recall = \frac{TP}{TP + FN} \tag{16}$$

where *TP* (*True Positives*) represents the number of positive samples properly classified as positive by the model; *TN* (*True Negatives*) represents the number of negative samples properly classified as negative by the model; *FP* (*False Positives*) represents the number of negative samples incorrectly classified as positive by the model; *FN* (*False Negatives*) represents the number of positive samples incorrectly classified as negative by the model.



**Figure 10.** The confusion matrix of the proposed model on the (a) T1 test datasets, (b) T2 test datasets, and (c) T3 test datasets.

**Table 5.** Evaluation metrics for the proposed model in the three datasets.

Numerical (%)		Evaluation Metrics		
		Precision	Recall	Accuracy
T1 datasets	Initial wear	90.8	99.2	93.4
	Normal wear	94.5	85.2	
	Severe wear	95.2	95.8	
T2 datasets	Initial wear	96.4	94.4	87.9
	Normal wear	79.7	85.5	
	Severe wear	88.4	83.8	
T3 datasets	Initial wear	92.5	99.6	95.0
	Normal wear	94.0	90.9	
	Severe wear	99.0	94.6	

To better validate the properties of the model, two ablation models were set up for comparative validation. The first ablation model (CNN) removed the attention mechanism and only preserved the one-dimensional convolutional module compared to the proposed model. The second ablation model (SECA-CNN) uses squeeze excitation channel attention mechanism instead of 1DCNN channel attention in the proposed model. It was also compared with existing tool wear state recognition algorithms that were also validated using the T dataset. Dong et al. [23] used the attention mechanism in CaAt1 and CaAt5 for tool wear state recognition in ResNet-1d. Yin et al. [30] implemented 1D-CNN with DGCCA for tool wear state recognition. Li et al. [31] combined GBDT with H-ClassRBM for tool wear state recognition. The comparison results of the recognition accuracy of different models are shown in Table 6.

It can be seen that the recognition accuracies of the proposed model under the three datasets were improved by 3.8%, 6.7%, and 4.7%, respectively. Compared with the CNN without channel attention, which verified the significance of the attention mechanism in optimizing the quality of the extracted features and improving the performance of the tool wear state recognition model. Compared with SECA-CNN, the recognition accuracy of the proposed 1DCCA-CNN improved by 1.2%, 2.1%, and 1.6% on the three datasets, respectively, which was due to the poor ability of the conventional squeeze excitation

channel attention mechanism to capture channel dependencies, and resulted in the model's generalization capability decrease. Using 1D CNN instead of squeeze excitation can effectively achieve inter-channel interaction. Compared to existing research findings, the proposed model demonstrated a significant lead in recognition accuracy on the T1 and T3 datasets. The recognition accuracy on the T1 dataset improved by 4% compared to the highest level achieved in existing research, while on the T3 dataset, the recognition accuracy improved by 5% compared to the highest level achieved in existing research.

**Table 6.** The recognition accuracies of different comparative models.

Model	T1 Datasets	T2 Datasets	T3 Datasets
CNN	89.7	81.2	90.3
SECA-CNN	92.3	85.8	93.4
CaAt-ResNet-1d [23]	86.0	88.0	89.2
1D-CNN-DGCCA [30]	89.5	88.4	90.0
GHCRBM [31]	89.2	90.8	58.4
The proposed model	93.5	87.9	95.0

#### 4. Conclusions and Future Work

This paper proposed a tool wear status recognition algorithm based on 1DCNN channel attention mechanism. The low and intermediate features in the signal were extracted by two one-dimensional convolution layers, and the channels of the features were weighted by the proposed channel attention mechanism to enhance the important channels and capture the key features. Then the high-level features were fetched by the last one-dimensional convolutional layer, and finally mapped to the classification layer through the fully connected layer. The main contributions of this study are as follows:

- (1) The channel attention mechanism was proposed by using 1DCNN instead of the traditional squeeze excitation. The good cross-channel information acquisition ability of convolutional was used to improve the information interaction between channels, so as to effectively capture the dependency between channels.
- (2) The model performance was verified on PHM2010 public dataset. Compared with CNN without attention mechanism, the recognition accuracy of the proposed 1DCCA-CNN improved 3.8%, 6.7%, and 4.7% respectively under the three datasets, which verified the importance of applying channel attention to model performance improvement. Compared with SECA-CNN, which used the traditional squeeze excitation attention mechanism, the recognition accuracy of the proposed 1DCCA-CNN improved 1.2%, 2.1%, and 1.6%, respectively, under the three datasets, which verified the superior performance of the proposed 1DCNN attention mechanism.
- (3) The proposed model had high accuracy, precision and recall, which verified the recognition performance of the model. Compared with the existing research results, the proposed model performed well on the T1 and T3 datasets, increasing by 4% and 5% respectively compared with the highest level of the existing research results, which verified the superior properties of the model.

The proposed methodology still has some limitations. First of all, the recognition accuracy of the model on T2 datasets is still slightly lower than the highest level of existing research results, and further improved performance will be considered in data preprocessing and network depth in the future. Secondly, this study is only applicable to a single working condition, and the tool wear state recognition algorithm under multiple working conditions should be studied later. Finally, the explainability of the proposed model is poor, and the visualization of the model should be further improved in the future to improve the explainability.

**Author Contributions:** Z.X.: Conceptualization, Methodology, Software, Algorithm, Validation, Data curation, Writing—original draft, Writing—review & editing. L.L.: Conceptualization, Methodology, Supervision, Writing instruction, Funding acquisition. N.C.: Conceptualization, Methodology, Super-

vision, Writing instruction, Funding acquisition. W.W.: Investigation, Validation. Y.Z.: Investigation, Validation. N.Y.: Conceptualization, Methodology, Supervision, Writing instruction. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study supported by Opening Project of the Key Laboratory of Advanced Manufacturing and Intelligent Technology (Ministry of Education), Harbin University of Science and Technology (No. KFKT2022), the National Natural Science Foundation of China (NSFC) [No. 51975288] and the Fundamental Research Funds for the Central Universities [No. NT2022014] for their support.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Pimenov, D.Y.; Gupta, M.K.; Da Silva, L.R.R.; Kiran, M.; Khanna, M.; Krolczyk, G.M. Application of measurement systems in tool condition monitoring of Milling: A review of measurement science approach. *Measurement* **2022**, *199*, 111503. [\[CrossRef\]](#)
- Pimenov, D.Y.; Bustillo, A.; Wojciechowski, S.; Sharma, V.S.; Gupta, M.K.; Kuntoglu, M. Artificial intelligence systems for tool condition monitoring in machining: Analysis and critical review. *J. Intell. Manuf.* **2023**, *34*, 2079–2121. [\[CrossRef\]](#)
- Li, X.B.; Liu, X.L.; Yue, C.X.; Liang, S.Y.; Wang, L.H. Systematic review on tool breakage monitoring techniques in machining operations. *Int. J. Mach. Tools Manuf.* **2022**, *176*, 103882. [\[CrossRef\]](#)
- Banda, T.; Farid, A.A.; Li, C.; Jauw, V.L.; Lim, C.S. Application of machine vision for tool condition monitoring and tool performance optimization—A review. *Int. J. Adv. Manuf. Technol.* **2022**, *121*, 7057–7086. [\[CrossRef\]](#)
- Yu, J.B.; Cheng, X.; Lu, L.; Wu, B. A machine vision method for measurement of machining tool wear. *Measurement* **2021**, *182*, 109683. [\[CrossRef\]](#)
- Shang, H.B.; Sun, C.; Liu, J.X.; Chen, X.F.; Yan, R.Q. Defect-aware transformer network for intelligent visual surface defect detection. *Adv. Eng. Inform.* **2023**, *55*, 101882. [\[CrossRef\]](#)
- Zhang, Y.P.; Qi, X.Z.; Wang, T.; He, Y.H. Tool Wear Condition Monitoring Method Based on Deep Learning with Force Signals. *Sensors* **2023**, *23*, 4595. [\[CrossRef\]](#)
- Wang, L.Q.; Li, X.; Shi, B.; Munochiveyi, M. Analysis and selection of eigenvalues of vibration signals in cutting tool milling. *Adv. Mech. Eng.* **2022**, *14*, 1079553789. [\[CrossRef\]](#)
- Liu, M.K.; Tseng, Y.H.; Tran, M.Q. Tool wear monitoring and prediction based on sound signal. *Int. J. Adv. Manuf. Technol.* **2019**, *103*, 3361–3373. [\[CrossRef\]](#)
- Nasir, V.; Sassani, F. A review on deep learning in machining and tool monitoring: Methods, opportunities, and challenges. *Int. J. Adv. Manuf. Technol.* **2021**, *115*, 2683–2709. [\[CrossRef\]](#)
- Qin, Y.Y.; Liu, X.L.; Yue, C.X.; Zhao, M.W.; Wei, X.D.; Wang, L.H. Tool wear identification and prediction method based on stack sparse self-coding net-work. *J. Manuf. Syst.* **2023**, *68*, 72–84. [\[CrossRef\]](#)
- Duan, J.; Hu, C.; Zhan, X.B.; Zhou, H.D.; Liao, G.L.; Shi, T.L. MS-SSPCANet: A powerful deep learning framework for tool wear prediction. *Robot. Comput.-Integr. Manuf.* **2022**, *78*, 102391. [\[CrossRef\]](#)
- Wei, X.D.; Liu, X.L.; Yue, C.X.; Wang, L.H.; Liang, S.Y.; Qin, Y.Y. Tool wear state recognition based on feature selection method with whitening variational mode decomposition. *Robot. Comput.-Integr. Manuf.* **2022**, *77*, 102344. [\[CrossRef\]](#)
- Chan, Y.W.; Kang, T.C.; Yang, C.T.; Chang, C.H.; Huang, S.M.; Tsai, Y.T. Tool wear prediction using convolutional bidirectional LSTM networks. *J. Supercomput.* **2022**, *78*, 810–832. [\[CrossRef\]](#)
- Zhou, Y.Q.; Zhi, G.F.; Chen, W.; Qian, Q.J.; He, D.D.; Sun, B.T.; Sun, W.F. A new tool wear condition monitoring method based on deep learning under small samples. *Measurement* **2022**, *189*, 110622. [\[CrossRef\]](#)
- Hou, W.; Guo, H.; Yan, B.N.; Xu, Z.; Yuan, C.; Mao, Y. Tool wear state recognition under imbalanced data based on WGAN-GP and light-weight neural network ShuffleNet. *J. Mech. Sci. Technol.* **2022**, *36*, 4993–5009. [\[CrossRef\]](#)
- Niu, Z.Y.; Zhong, G.Q.; Yu, H. A review on the attention mechanism of deep learning. *Neurocomputing* **2021**, *452*, 48–62. [\[CrossRef\]](#)
- Li, R.Y.; Wei, P.N.; Liu, X.L.; Li, C.L.; Ni, J.; Zhao, W.K.; Zhao, L.B.; Hou, K.L. Cutting tool wear state recognition based on a channel-space attention mechanism. *J. Manuf. Syst.* **2023**, *69*, 135–149. [\[CrossRef\]](#)
- Zeng, Y.F.; Liu, R.L.; Liu, X.F. A novel approach to tool condition monitoring based on multi-sensor data fusion imaging and an attention mechanism. *Meas. Sci. Technol.* **2021**, *32*, 55601. [\[CrossRef\]](#)
- Hou, W.; Guo, H.; Luo, L.; Jin, M.J. Tool wear prediction based on domain adversarial adaptation and channel attention multiscale convolutional long short-term memory network. *J. Manuf. Process.* **2022**, *84*, 1339–1361. [\[CrossRef\]](#)
- Zhou, J.Q.; Yue, C.X.; Liu, X.L.; Xia, W.; Wei, X.D.; Qu, J.X.; Liang, S.Y.; Wang, L.H. Classification of Tool Wear State based on Dual Attention Mechanism Network. *Robot. Comput.-Integr. Manuf.* **2023**, *83*, 102575. [\[CrossRef\]](#)
- He, J.L.; Sun, Y.X.; Yin, C.; He, Y.; Wang, Y.L. Cross-domain adaptation network based on attention mechanism for tool wear prediction. *J. Intell. Manuf.* **2022**, *34*, 3365–3387. [\[CrossRef\]](#)
- Dong, L.; Wang, C.S.; Yang, G.; Huang, Z.Y.; Zhang, Z.Y.; Li, C. An Improved ResNet-1d with Channel Attention for Tool Wear Monitor in Smart Manufacturing. *Sensors* **2023**, *23*, 1240. [\[CrossRef\]](#) [\[PubMed\]](#)

24. Guo, B.S.; Zhang, Q.; Peng, Q.J.; Zhuang, J.C.; Wu, F.H.; Zhang, Q. Tool health monitoring and prediction via attention-based encoder-decoder with a multi-step mechanism. *Int. J. Adv. Manuf. Technol.* **2022**, *122*, 685–695. [[CrossRef](#)]
25. Lai, X.W.; Zhang, K.; Zheng, Q.; Li, Z.X.; Ding, G.F.; Ding, K. A frequency-spatial hybrid attention mechanism improved tool wear state recognition method guided by structure and process parameters. *Measurement* **2023**, *214*, 112833. [[CrossRef](#)]
26. Feng, T.T.; Guo, L.; Gao, H.L.; Chen, T.; Yu, Y.X.; Li, C.G. A new time-space attention mechanism driven multi-feature fusion method for tool wear monitoring. *Int. J. Adv. Manuf. Technol.* **2022**, *120*, 5633–5648. [[CrossRef](#)]
27. Huang, Q.Q.; Wu, D.; Huang, H.; Zhang, Y.; Han, Y. Tool Wear Prediction Based on a Multi-Scale Convolutional Neural Network with Attention Fusion. *Information* **2022**, *13*, 504. [[CrossRef](#)]
28. Wang, Q.L.; Wu, B.G.; Zhu, P.F.; Li, P.H. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
29. PHM Society. PHM Society Conference Data Challenge. 2010. Available online: <https://www.phmsociety.org/competition/phm/10> (accessed on 20 December 2022).
30. Yin, Y.; Wang, S.X.; Zhou, J. Multisensor-based tool wear diagnosis using 1D-CNN and DGCCA. *Appl. Intell.* **2023**, *53*, 4448–4461. [[CrossRef](#)]
31. Li, G.F.; Wang, Y.B.; He, J.L.; Hao, Q.B.; Yang, H.J.; Wei, J.F. Tool wear state recognition based on gradient boosting decision tree and hybrid classification RBM. *Int. J. Adv. Manuf. Technol.* **2020**, *110*, 511–522. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.