Journal of
Manufacturing and Materials Processing

MDPI

Article

A Real-Time Anomaly Detection Model of Nomex Honeycomb Composites Disc Tool

Xuanlin Wang [1,†], Peihao Tang [1,†], Jie Xu [1,†], Xueping Liu [1,*] and Peng Mou [2,*]

[1] Institute of Data and Information, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China; wang-xl23@mails.tsinghua.edu.cn (X.W.); tph22@mails.tsinghua.edu.cn (P.T.); xu-j22@mails.tsinghua.edu.cn (J.X.)
[2] Department of Mechanical Engineering, Tsinghua University, Beijing 100084, China
* Correspondence: liuxp@sz.tsinghua.edu.cn (X.L.); moupeng@tsinghua.edu.cn (P.M.)
† These authors contributed equally to this work.

Abstract

Nomex honeycomb composites (NHCs) are highly sensitive to the abnormal wear state of disc tools during cutting, leading to poor product quality. This paper proposes a real-time anomaly detection method combining a novel CNN–GRU–Attention (CGA) deep learning model with an Exponentially Weighted Moving Average (EWMA) control chart to monitor sensor data from the disc tool. The CGA model integrates an improved CNN layer to extract multidimensional local features, a GRU layer to capture long-term temporal dependencies, and a multi-head attention mechanism to highlight key information and reduce error accumulation. Trained solely on normal operation data to address the scarcity of abnormal samples, the model predicts cutting force time series with an RMSE of 0.5012, MAE of 0.3942, and R^2 of 0.9128, outperforming mainstream time series data prediction models. The EWMA control chart applied to the prediction residuals detects abnormal tool wear trends promptly and accurately. Experiments on real NHC cutting datasets demonstrate that the proposed method effectively identifies abnormal machining conditions, enabling timely tool replacement and significantly enhancing product quality assurance.

Keywords: anomaly detection; machine learning; disc tool; CNN–GRU–Attention; EWMA control chart

check for updates

Academic Editor: Steven Y. Liang

Received: 8 July 2025
Revised: 12 August 2025
Accepted: 14 August 2025
Published: 15 August 2025

Citation: Wang, X.; Tang, P.; Xu, J.; Liu, X.; Mou, P. A Real-Time Anomaly Detection Model of Nomex Honeycomb Composites Disc Tool. *J. Manuf. Mater. Process.* **2025**, *9*, 281. https://doi.org/10.3390/jmmp9080281

Copyright: © 2025 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

Nomex honeycomb composites (NHCs) are prepared by imitating the hexagonal array structure of honeycomb using Nomex paper, which has low density and high axial compressive strength [1]. At the same time, materials with this structure usually have good bending and twisting properties [2,3]. Therefore, NHCs are commonly used in the manufacturing of aircraft components [4].

High-speed milling is a common processing method for NHCs. However, there are several differences compared to the traditional metal milling process [5]. First, due to the difference between NHCs and metal, coolant cannot be used for heat dissipation during the machining process, and NHCs have poor thermal conductivity [6]. Second, disc tools made of high-speed steel are commonly used because of their impact resistance and vibration tolerance [7]. However, these tools also suffer from poor thermal conductivity, resulting in accelerated wear in the cutting of NHCs. Third, the disc cutter used in NHC processing is very thin. Even if the cutter is slightly damaged or worn, the surface quality of the workpiece will deteriorate rapidly [8]. All these reasons make the surface processing

quality of NHCs highly sensitive to tool wear. As shown in Figure 1, tool wear can cause defects such as tears and cell collapse on the surface of NHCs, which fail to meet the requirements of machining surface quality [9].Therefore, the real-time anomaly detection of tool wear state and rapid response are critical to ensuring NHC machining quality.

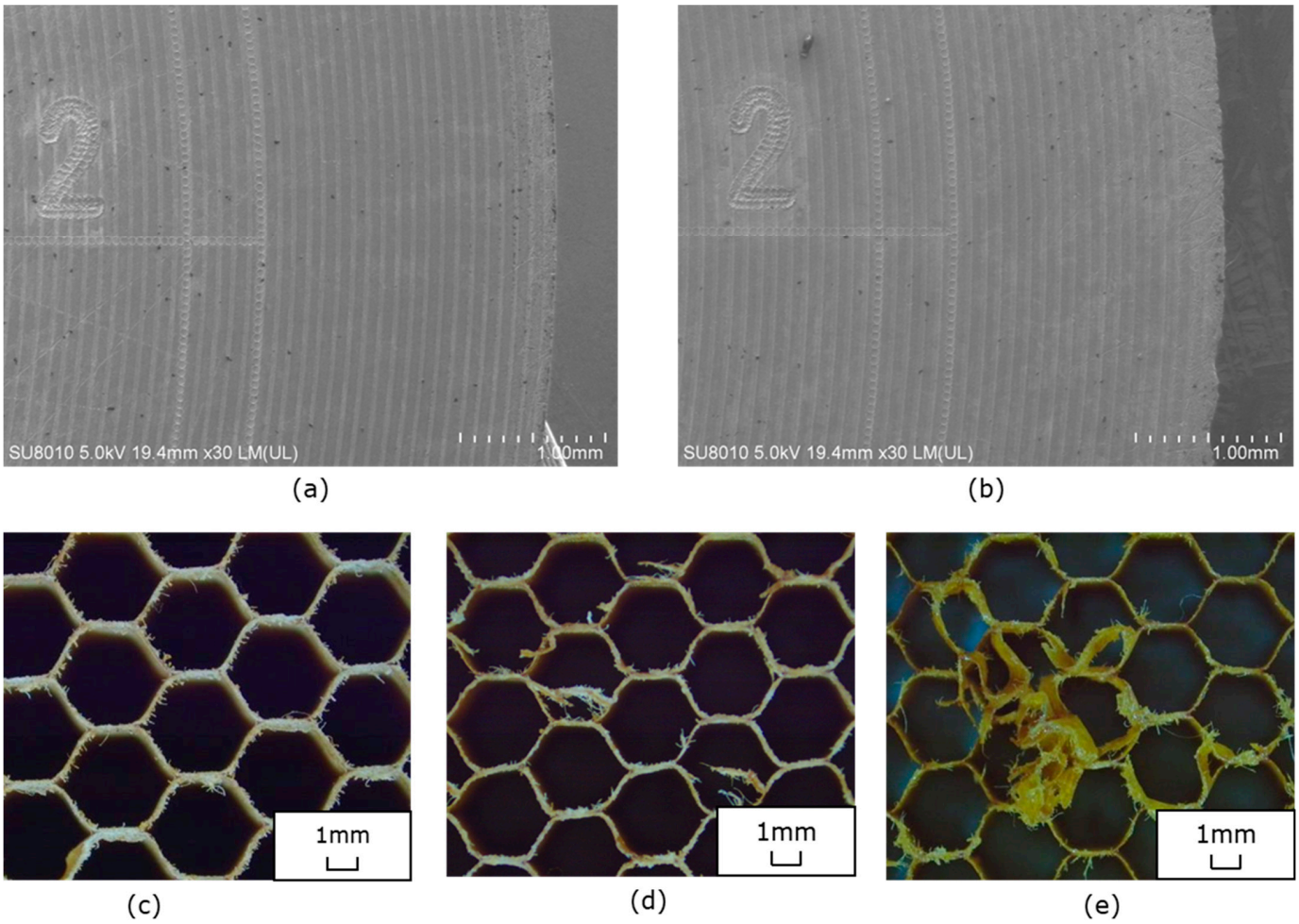


Figure 1. Pictures of (**a**) normal disc tool, (**b**) broken disc tool, (**c**) normal surface of the sample, and (**d**) unqualified surface: tears, (**e**) unqualified surface: cell collapse.

Currently, based on Internet of Things technology, analyzing the abnormal forms in time series data collected by sensors is an effective solution for the real-time anomaly detection of tool wear status [10]. Several mainstream methods, including mathematical statistics, a machine learning model, and a deep learning neural network, were used to detect anomalies in time series data.

Among them, statistical methods such as statistical process control (SPC) charts based on the 3-sigma principle can be used to detect time series data anomalies [11,12]. Statistical methods offer interpretability and require limited training data but struggle to establish a universal theoretical framework and identify relevant features in the data. Machine learning algorithms can effectively identify correlated features [13], which compensates for the shortcomings of statistical methods to some extent. Erfani et al. [14] used deep belief networks to extract general underlying features and then differentiated abnormal patterns using a supervised support vector machine (SVM) classifier. Wang et al. [15] used a comprehensive feature and parameter-optimized isolation forest model to detect abnormal data segments in rolling bearing vibration signals. Additionally, methods based on isolation forest [16], density [17], dimensionality reduction [18], and regression [19]

have also been used for anomaly detection in time series data. However, machine learning methods struggle to accurately capture change patterns or account for causal relationships in time series data, especially when feature interactions are complex.

Deep learning performs well when data feature patterns are complex, solving the problem of normal models being unable to accurately mine data-changing patterns. Therefore, deep learning has gradually become the primary choice for anomaly detection [20]. Oshida et al. [21] proposed a stacked LSTM encoder–decoder model for tool wear anomaly detection while turning difficult-to-cut materials, judging the current tool wear state through the training of audio signals. Munir, Mohsin et al. [22] used a convolutional neural network (CNN) as the prediction model to process data and compared the predicted value with the actual value to achieve anomaly detection, which is capable of detecting a wide range of anomalies, such as contextual anomalies and discords. Wang et al. [23] employed RNN-based encoder–decoder models for time series prediction and anomaly detection, allocating weights to time series data using an attention model. These deep learning methods have improved prediction accuracy to a certain extent, but a single network is limited when encountering more complex problems. Therefore, some scholars have combined different networks to achieve better performance. Zhang et al. [24] combined an improved CNN model with an RNN model to predict the lifespan of mechanical equipment by analyzing sensor signals. Zhang et al. [25] also combined the CNN model with the LSTM model to predict the trend of the air quality index, which effectively utilized CNN's powerful feature extraction function and LSTM's time series processing ability. Kong et al. [26] used a CNN-GRU network to predict data for the supervisory control and data acquisition (SCADA) system during state monitoring and plotted an EWMA control chart for anomaly detection. The above deep learning model based on RNN can retrieve previous data information, but this also leads to the accumulation of reconstruction errors, which will affect the accuracy of the anomaly detection. In addition, when the structure of the model is too simple, it cannot fully recognize the high-dimensional features. However, directly increasing the number of network layers will significantly increase the number of model computation parameters, resulting in excessively long model running times.

In summary, existing research often faces significant challenges: traditional methods, such as statistical process control and traditional machine learning, lack the ability to effectively capture the complex time dependencies and multidimensional feature interactions in cutting force data. Meanwhile, many deep learning methods rely on a single network structure or simple combinations, making it difficult to address issues such as cumulative reconstruction errors and limited feature extraction capabilities, especially under real-time constraints during the machining process. Additionally, the scarcity of abnormal tool wear samples poses a critical limitation for supervised learning models, restricting their application in real-world industrial scenarios.

To address these gaps, focusing on the requirements of accurate and rapid response for real-time tool anomaly detection in NHC processing, a CNN–GRU–Attention (CGA) combined with an EWMA control chart is proposed. The key contributions include the following: (1) The design of an improved parallel dilated convolution layer enables multidimensional local feature extraction without excessive computational overhead, overcoming the inability of the traditional CNN model to fully capture diverse feature scales; (2) A GRU layer is used to capture long-term temporal dependencies in time series; (3) A multi-head attention mechanism highlights critical information while suppressing noise, mitigating the problem of reconstruction error accumulation common in RNN model; (4) The model is trained on normal data only, and abnormality detection is achieved by monitoring the changes in the residual sequence based on the EWMA control chart, circumventing the scarcity of abnormal samples.

2. Methodology

2.1. Overview of the Method

The anomaly detection model based on CNN–GRU–Attention and the EWMA control chart proposed in this paper can effectively detect the anomaly of the time series data generated in the tool machining process. The structure of this anomaly detection model is shown in Figure 2.

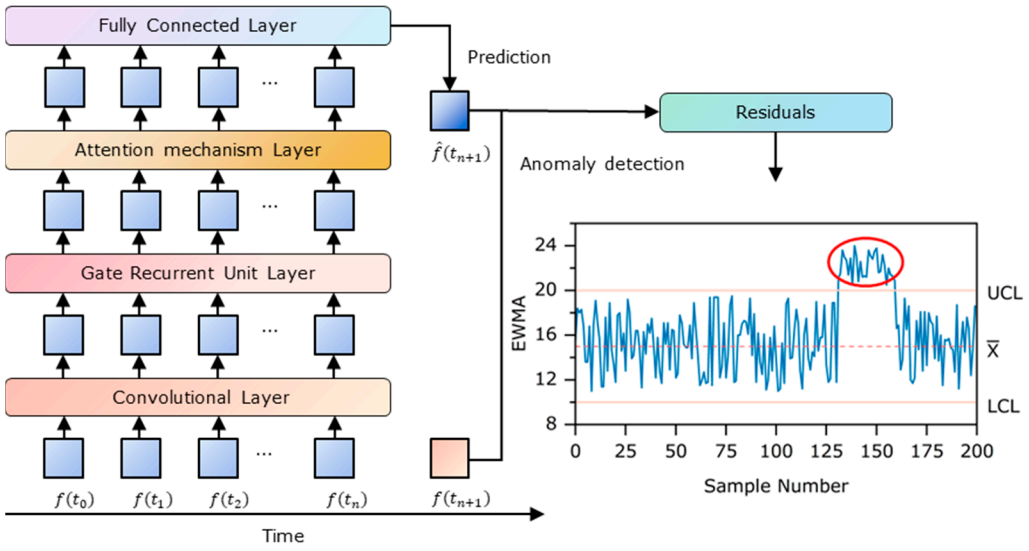


Figure 2. The structure of the anomaly detection model.

The whole model can be divided into two parts: the prediction stage and the anomaly detection stage. In the prediction stage of time series data, the dataset collected during the process of tool machining is first obtained. After preprocessing and normalizing, the dataset is put into the CNN layer for feature extraction. After that, the dimension is reduced by the pooling layer to prevent over-fitting. After inputting the pooled data into the GRU layer to extract the time series characteristics of the data, the data is imported into the attention mechanism layer to extract important information, and finally, the prediction results are output through a fully connected layer.

In the anomaly detection stage of time series data, the CGA model trained in the prediction stage is used to predict and fit using existing sensor data. After finding the residual sequence between the fitting result and the real data, the EWMA control chart is used to judge whether the change trend of the residual sequence deviates from the normal mode to realize the anomaly detection of sensor time series data as marked with red circle in Figure 2.

2.2. Improved CNN Model

A convolutional neural network (CNN) is a deep learning model that can effectively extract feature information from data sequences [27]. The CNN model usually consists of input, convolutional, pooling, and fully connected layers. The input layer of the CNN is responsible for importing raw data into the network model. The convolutional layer is the core part of the CNN, where the convolutional kernels slide forward on the input data to extract features. After the data is convolved, a non-linear mapping layer, which includes an activation function, can be set to enhance the model's trainability. LeakyReLU is used as the computational function for gating units. It solves the issue of neuron death by setting a minimal slope k in the negative range of the ReLU function [28] while inheriting the advantages of the ReLU function [29]. By applying a pooling operation to the output of

the convolutional layer, the dimension of the feature can be reduced, the computational complexity of the model can be decreased, and overfitting can be prevented. Common pooling operations include max pooling, average pooling, etc. The fully connected layer expands the output data of the layer into a one-dimensional vector and connects it to the output layer through a weight matrix, resulting in the final output.

A CNN can be divided into one-dimensional CNN models and multidimensional CNN models. In this case, a one-dimensional CNN model was chosen to process the sensor time series data generated during the tool processing to extract the temporal features. The temporal data patterns generated during the tool cutting process are complex, and multiple-dimensional feature correlations exist among each data point. When using a simple one-dimensional convolutional network to extract features from the tool processing, it is difficult to fully extract the features of different dimensions in the temporal data. However, adding too many layers to the network may lead to overfitting. To solve this problem, a parallel dilated convolution model is used for convolutional operations. The dilated convolution model uses dilated kernels as parallel convolution units. Compared to regular convolution, dilated convolution increases the receptive field by adding holes in the kernels, and different dilated rates result in different receptive fields, allowing the model to capture multi-scale data information. The structure of the dilated convolution model is shown in Figure 3a. The dilated rate (d) of 1 corresponds to standard convolution, and when $d > 1$, the dilated convolution kernel expands the receptive field.

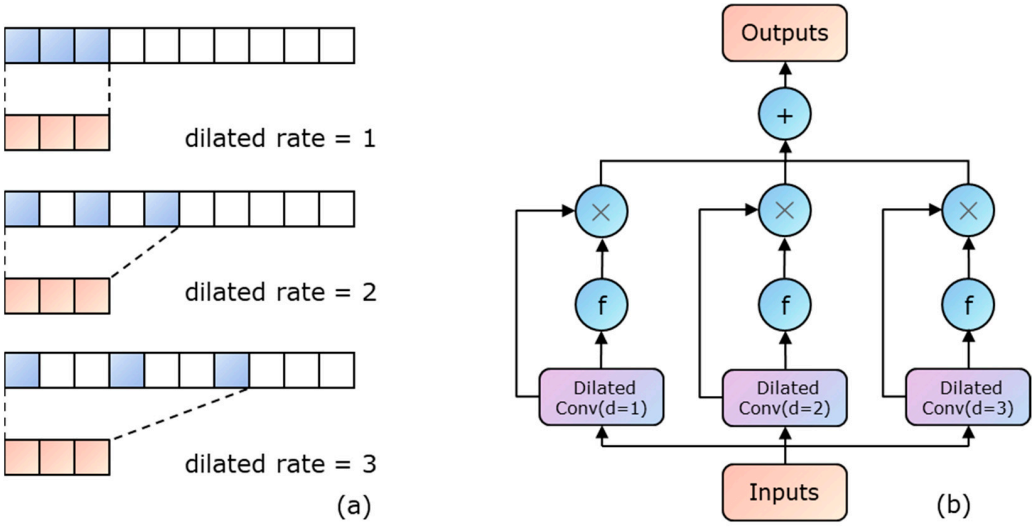


Figure 3. (**a**) The structure of the dilated convolution model and (**b**) the structure of the parallel dilated convolutional model.

The structure of the parallel dilated convolutional model we used is shown in Figure 3b. It extracts data features of different dimensions by constructing three dilated convolutional kernels with different dilated rates. After convolving the input data, a gate unit is set for each convolutional output to calculate the weights. The weights obtained from the LeakyReLU function are multiplied by the convolutional results, allowing each of the three convolutional kernels to extract data features of different dimensions. Finally, the three outputs are added together to obtain the final output of the convolutional layer.

2.3. GRU Model

The long short-term memory (LSTM) network is an evolution of the recurrent neural network (RNN), which resolves the issues of long-term dependencies and exploding gradients in the RNN and enables better extraction of non-linear relationships in time

series data [30,31]. The gated recurrent unit (GRU) is a variant of LSTM, which is another commonly used model for recurrent neural networks [32]. The data collected by the sensor during the tooling process is a type of time series data, which exhibits specific trend characteristics with the change of time. By analyzing the correlation and trend of time series data, it is possible to predict the future fluctuations of the data. Therefore, using the GRU model to analyze the data after CNN feature extraction can better capture the impact of preceding time series data on the current data. The structure of LSTM and the GRU is shown in Figure 4.

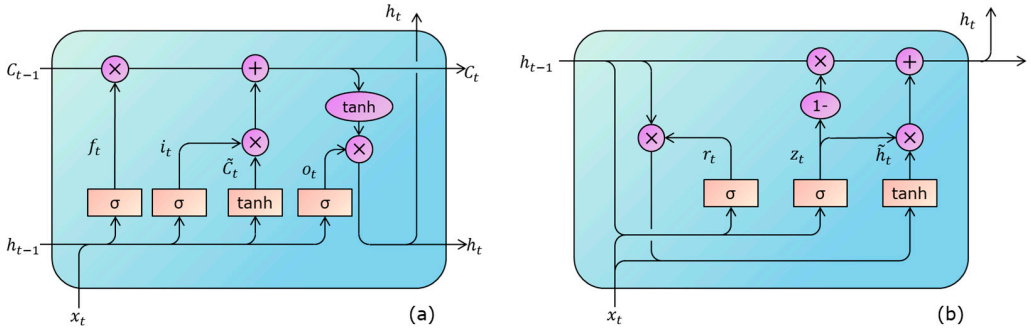


Figure 4. (**a**) The structure of LSTM and (**b**) the structure of the GRU.

LSTM is composed of three gate units: the forget gate, input gate, and output gate. The GRU simplifies LSTM by combining the forget gate and the input gate into an update gate. This reduces the number of parameters in the model, improves training speed, and lowers the risk of overfitting. Therefore, the GRU was used as the training model in this case. The GRU updates the current hidden state for model training by incorporating the previous hidden state and current input information through the update and output gates. The calculation formulas are defined as [32]:

$$r_t = \sigma(W_{rh}h_{t-1} + W_{rx}x_t + b_r) \tag{1}$$

$$z_t = \sigma(W_{zh}h_{t-1} + W_{zx}x_t + b_z) \tag{2}$$

$$\widetilde{h}_t = tanh[W_{hh}(r_t \bigodot h_{t-1}) + W_{xh}x_t + b_h] \tag{3}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \widetilde{h}_t \tag{4}$$

where x_t represents the t-th input in the sequence, and h_t and h_{t-1} represent the hidden layer matrix at time t and $t-1$. $\widetilde{h}_t$ represents the temporary hidden layer matrix at time t. r_t represents the output of the reset gate. z_t represents the output of the update gate, determining the proportion of h_{t-1} and $\widetilde{h}_t$ in the hidden state h_t at the current moment. W_{zh} and W_{zx} represent the weights of the update gate, while W_{rh} and W_{rx} represent the weights of the reset gate. σ represents the sigmoid function, which scales variables between 0 and 1.

2.4. Attention Mechanism Model

The attention mechanism was initially proposed by the Bengio team in 2014 [33]. It can select the most crucial information for the current task from a large amount of information and allocate limited resources to the more essential parts. When using a GRU on cutting-tool time series, long-range dependencies require information to pass through many time steps, which weakens input correlations. Meanwhile, reconstruction errors accumulate backward in the RNN, degrading recognition accuracy. To better extract

the feature dependency relationship between the time series data of tool processing, a multi-head self-attention layer is introduced to further train the output of the GRU layer. It assigns weights to the relevance of the current data in predicting the data at the next time step, filters out important information in the time series data, and improves the model's prediction accuracy. The structure of the self-attention mechanism and multi-head self-attention is shown in Figure 5.

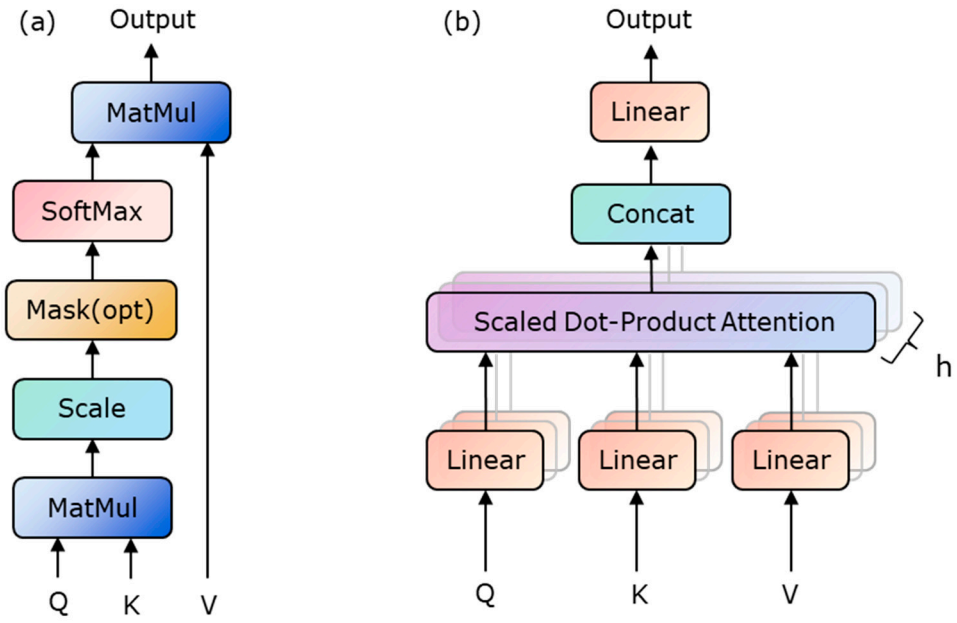


Figure 5. (**a**) The structure of the self-attention mechanism and (**b**) the structure of the multi-head self-attention mechanism.

The structure of the self-attention mechanism is shown in Figure 5a. The self-attention mechanism requires three input parameters—query, key, and value—where query represents the objective function, key represents the correlation between input data, and value represents the specific numerical values of the input data. Firstly, calculate the correlation between query and key to obtain the raw score, which is commonly achieved using methods such as dot product and cosine similarity. Scale the raw score and normalize it using the SoftMax function, which maps the score to a range between 0 and 1, to obtain the weight coefficient. The attention value is obtained by the weighted summation of value according to the weight coefficient, defined as [33]:

$$A_v = SoftMax(F(Q,K))\cdot V \tag{5}$$

where A_v represents the output of the attention mechanism. Q, K, V, respectively, represent query, key, and value. *SoftMax* represents the normalization exponential function.

The structure of the multi-head self-attention mechanism is shown in Figure 5b. Firstly, several self-attention mechanism models are trained independently, and each attention head gets an output separately. All outputs are spliced into a whole and multiplied by the output matrix to obtain the final output result of multi-head self-attention model. The multi-head self-attention mechanism enables the model to pay attention to multiple critical areas of time series data at the same time and enhances the feature expression ability of the model. The formula of multi-head self-attention is defined as [34]:

$$MultiHead(Q,K,V) = Concat(head_1,\ldots,head_h)W^o \tag{6}$$

where $head_i$ represents the output of attention head, and W^o represents the weight matrix.

2.5. Exponentially Weighted Moving Average Control Chart

After using a deep learning model to predict time series data, the prediction results are subtracted from the actual results to obtain a reconstruction error sequence. If the time series data is normal, the reconstruction error will randomly vary within a small range. Typically, data points that exhibit sudden, drastic changes are considered criteria for anomaly detection. However, some abnormal situations may occur without obvious changes in the data, making it difficult to detect the occurrence of anomalies solely based on the analysis of individual data points, such as the partial wear of cutting tools. When abnormal situations occur, not only will the values of the reconstruction error change but also the change trend of the reconstruction error will be altered. Therefore, analyzing the change trend of the reconstruction error sequence can more accurately determine the occurrence of anomalies. Moreover, to meet the real-time requirement of abnormal detection in NHC processing, a time-saving method should be chosen to identify abnormal patterns.

Using an exponentially weighted moving average (EWMA) control chart [35] for anomaly diagnosis of the changing trend in the reconstruction error sequence can reduce the influence of noise by calculating the weighted average of historical observations, resulting in a smoother dataset [36]. Additionally, compared with using complex machine learning algorithms to detect the anomaly of the residual sequence, the EWMA control chart can quickly find out the abnormal information of the residual sequence by using statistical rules, which meets the real-time requirements of the actual machining process. The calculation formula for the standard EWMA is defined as [35]:

$$EWMA_t = \lambda \cdot e_t + (1 - \lambda) \cdot EWMA_{t-1} \tag{7}$$

where $EWMA_t$ is the calculation result of EWMA, t represents the timestamp, and e_t is the average output. λ is a constant between 0 and 1, representing the influence of historical data.

The calculation formulas for the upper confidence limit (UCL) and the lower confidence limit (LCL) of the EWMA control chart can be described as

$$UCL(t) = \mu_e + K\sigma_e \sqrt{\frac{\lambda}{2-\lambda}[1 - (1-\lambda)^{2t}]} \tag{8}$$

$$LCL(t) = \mu_e - K\sigma_e \sqrt{\frac{\lambda}{2-\lambda}[1 - (1-\lambda)^{2t}]} \tag{9}$$

where μ_e and σ_e represent the mean and standard deviation of the residual sequence, and K and λ are parameters that affect the positions of the UCL and LCL. Based on existing experience [37], K is set to 3, and λ is set to 0.2. Under normal circumstances, the data points should randomly fluctuate within the control limits. If consecutive data points exceed the upper or lower confidence limits, or if the data exhibits a non-random pattern, it can be considered that the current manufacturing process is in an abnormal state.

3. Experiment

In this paper, the actual cutting force dataset from the disc tool when processing honeycomb composite materials was used to verify the effect of the proposed anomaly detection model. In the prediction part, the comparison results between the proposed model, the traditional model, and other deep learning models are shown. The anomaly detection part shows that the EWMA control chart model can detect the abnormal situation corresponding to the cutting force sensor data when the tool has abnormal wear. The proposed model was implemented using Python 3.9.1 and PyTorch 2.6.0+cu118, and the experiments were run on a laptop with an Intel® Core™ i5-9300H Processor (Santa

Clara, CA, USA) and 16 GB DDR4-2666 RAM (Micron Technology®, headquartered in Boise, ID, USA).

3.1. Data Source

The dataset used in this paper comes from the real-time data collected by sensors when processing NHCs. As shown in Figure 6, the disc cutter was installed on the ultrasonic cutter frame connected with the spindle, and the NHCs were fixed on the machine tool table via specially designed fixtures. The parameters of NHCs are detailed in our previous study [7]. The specifications of the blades are as follows: 27 mm in diameter, 14° wedge angle, 0.9 mm thickness, 3500 rpm spindle speed, and 5 m/min feed speed. Three pre-treatments (PTs) were applied to M2 HSSS to prepare the disc tool samples. PT1 is the control group. PT2 is for hardness enhancement. PT3 is for toughness enhancement. For specific preparation processes and performance parameters, please refer to our previous research work [38]. The sample preparation fulfilled the ASTM E3(2011) standards. The disc tool was driven by an ultrasonic controller (Tsingding® model UMINT-20-300, Shenzhen, China) with a rated power of 900 W and an ultrasonic amplitude of 25 μm during the off-site test. Due to its intermittent cutting characteristics, ultrasonic vibration machining can effectively reduce cutting forces. In addition, it can also effectively improve surface quality. In our setup, the NHC workpiece was rigidly mounted to the dynamometer (Kistler® model 9119 A, Winterthur, Switzerland), which was further fixed to the machine tool through a vise, with locating pins to prevent slip and bypass load paths and ensure a single, stiff load path. The sampling frequency of the dynamometer was 20 kHz. The dynamometer was connected with the computer through a data acquisition board (Kistler® model 5697 A, Winterthur, Switzerland) and a charge amplifier (Kistler® model 5080A100804, Winterthur, Switzerland). The dynamometer data was processed by the computer.

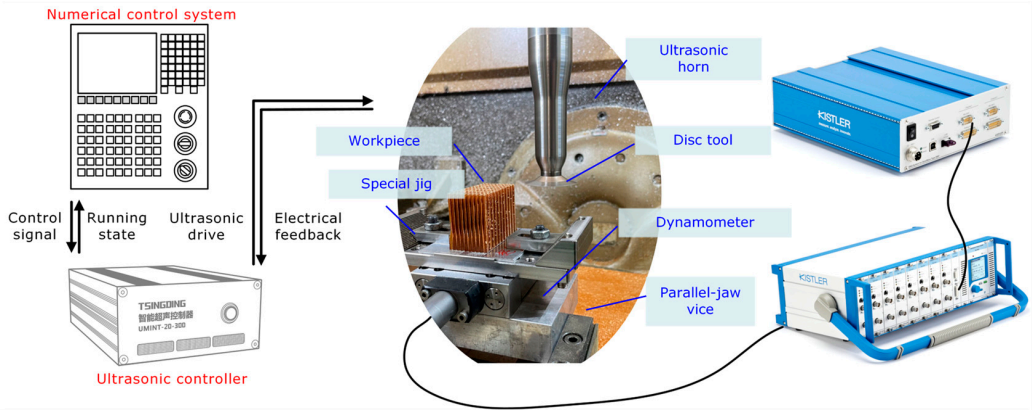


Figure 6. Schematic diagram of ultrasonic machining system and sensor acquisition device.

The time series data of the cutting force in the feed direction of the disc tool during the normal cutting process is shown in Figure 7. It shows the superposition of three periodic characteristics: ultrasonic vibration, tool rotation, and honeycomb structure. The ultrasonic vibration period manifests as high-frequency fluctuations, reflecting the microscopic impact of vibration on the cutting force; the tool rotation period causes medium-frequency fluctuations, reflecting the periodic influence of tool rotation; and the honeycomb period manifests as low-frequency fluctuations, reflecting the macroscopic influence of material structure on the force. The intertwining of these three factors forms complex fluctuations that comprehensively affect the stability and quality of the cutting process. Since the disc tool mainly moves along a horizontal direction on a flat surface in the cutting process, the longitudinal force component has little effect on the cutting process and is not sensitive

to it. The main wear on the tool and fluctuations in cutting force are primarily caused by the feed direction force component. Therefore, only the cutting force in the cutting direction was considered, which can more accurately reflect the wear status of the tool and the characteristics of the cutting process. In order to reduce the unnecessary data processing calculation, the data processing frequency of the cutting force was set to 2 kHz by down-sampling. In addition, since the spindle speed is 3500 rpm, the downsampled frequency is sufficient to cover each rotation cycle. In the data preprocessing stage, the central cutting process part of the time series data was intercepted for subsequent model training. The data of 800 time steps was intercepted, and the time interval of each step was 0.5 ms.

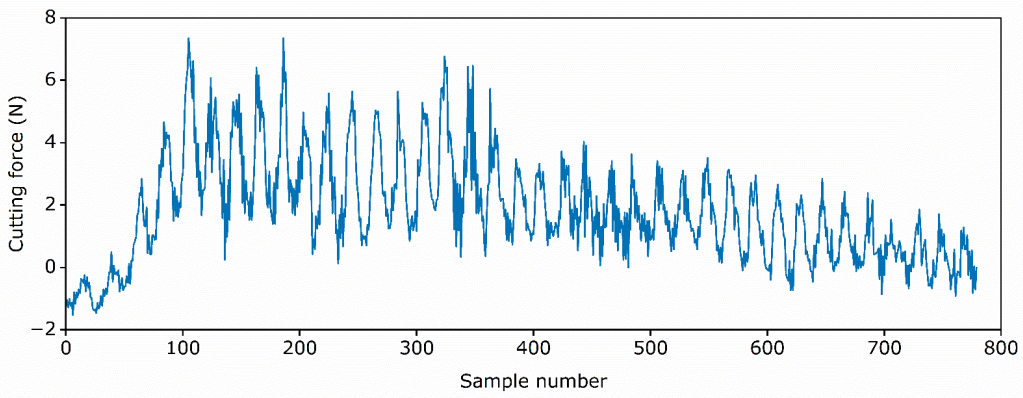


Figure 7. The time series data of the cutting force in the cutting process.

3.2. Model Evaluation Index

In the prediction model, the model proposed in this paper was compared with other models to show the performance of the model. Root mean square error (RMSE), mean absolute error (MAE), and the determining coefficient R^2 score were used as indicators to evaluate the accuracy of different forecasting models. MSE is the mean sum of the squares of the predicted and original value errors, and RMSE is the square root of MSE. MAE is the average absolute deviation of all individual observed values from the arithmetic mean. It can reflect the average difference between the real value and the predicted value. The R^2 score describes the deviation of the predicted data from the original data, which can measure the degree of the model fitting to the observed value.

3.3. Model Parameter Setting

The deep learning model for prediction consists of four layers: the CNN, the GRU, an attention layer, and a fully connected (FC) layer. In the model, the weight calculation function of the CNN layer is LeakyReLU, the number of convolution layers is 1, and the number of the CNN output channel was set to 128. The activation function used between the fully connected layers is ReLU. The learning rate was set to 0.00022, and the iteration number of the model was 200.

Through several groups of comparative experiments, the optimal parameters of time step and batch size, the number of GRU layers, and the size of the convolution kernel in the CNN were verified. As shown in Table 1, when the GRU layer number is set to 1, the prediction model has the best result on this dataset. As shown in Table 2, when the time step is set to 20 and the batch size is set to 64, the prediction result of the model is closest to the real value. As shown in Table 3, when the size of the convolution kernel in the CNN is 5, the prediction model has the smallest prediction error. The downward arrow (↓) in the table indicates that smaller values are better, while the upward arrow (↓) indicates the opposite.

Table 1. The prediction results of the model under different GRU layer number settings.

GRU Layers	RMSE↓	MAE↓	R^2
1	0.501205	0.394193	0.912846
2	0.516358	0.403856	0.907496
3	0.555385	0.442965	0.892985

Table 2. The prediction results of the model under different time step and batch size settings.

Time Step	Batch Size	RMSE↓	MAE↓	R^2
10	64	1.711630	1.322358	−0.001070
10	128	1.640167	1.280547	0.080778
20	64	0.501205	0.394193	0.912846
20	128	0.567782	0.443489	0.888154
30	64	0.549895	0.422061	0.892133
30	128	0.534528	0.416167	0.898077

Table 3. The prediction results of the model under different convolution kernel size settings.

Kernel Size	RMSE↓	MAE↓	R^2
3	0.602966	0.485827	0.873863
5	0.501205	0.394193	0.912846
7	0.565571	0.455293	0.889023

3.4. Prediction Result Analysis

In order to verify the performance of the CNN–GRU–Attention model proposed in this paper, ablation experiments were conducted to determine whether the prediction results of the combined model improved.

(1) CNN: Extract the characteristic relationship in the dataset for prediction.
(2) LSTM: Analyze the time series characteristics in the dataset.
(3) GRU: a variant of LSTM.
(4) CNN-GRU: Based on the model proposed in this paper, the attention model is removed, the CNN is used to extract the characteristic relationship of data, and the GRU is used to extract time series characteristics.
(5) Normal CNN–GRU–Attention: Based on the model proposed in this paper, the parallel dilated convolutional layer is replaced by the ordinary convolution layer.

The prediction results of our CNN–GRU–Attention model and the above comparison prediction models for the cutting force data of the disc tool are shown in Table 4, which shows that our proposed CNN–GRU–Attention model achieves the best prediction accuracy. To demonstrate the predictive performance of the models more intuitively, a comparison between the predicted values and the valid values of 780 data points at the same positions from each model's prediction data is shown in Figure 8. In these figures, the blue curve represents the true values of the cutting force data, while the yellow curve represents the predicted values. The *x*-axis represents the corresponding time of the data, and the *y*-axis represents the numerical value of the cutting force.

Table 4. The prediction results of ablation experiments.

Model	RMSE↓	MAE↓	R^2	Items per Second↑
CNN [27]	0.667312	0.504933	0.845505	337.80 it/s
LSTM [31]	2.192965	1.689043	−0.643266	233.09 it/s
GRU [32]	2.426704	1.900223	−1.043104	258.89 it/s
CNN-GRU	0.579525	0.434325	0.883480	220.10 it/s
normal CNN–GRU–Attention	0.584356	0.450506	0.881529	201.17 it/s
our CNN–GRU–Attention	0.501205	0.394193	0.912846	197.40 it/s

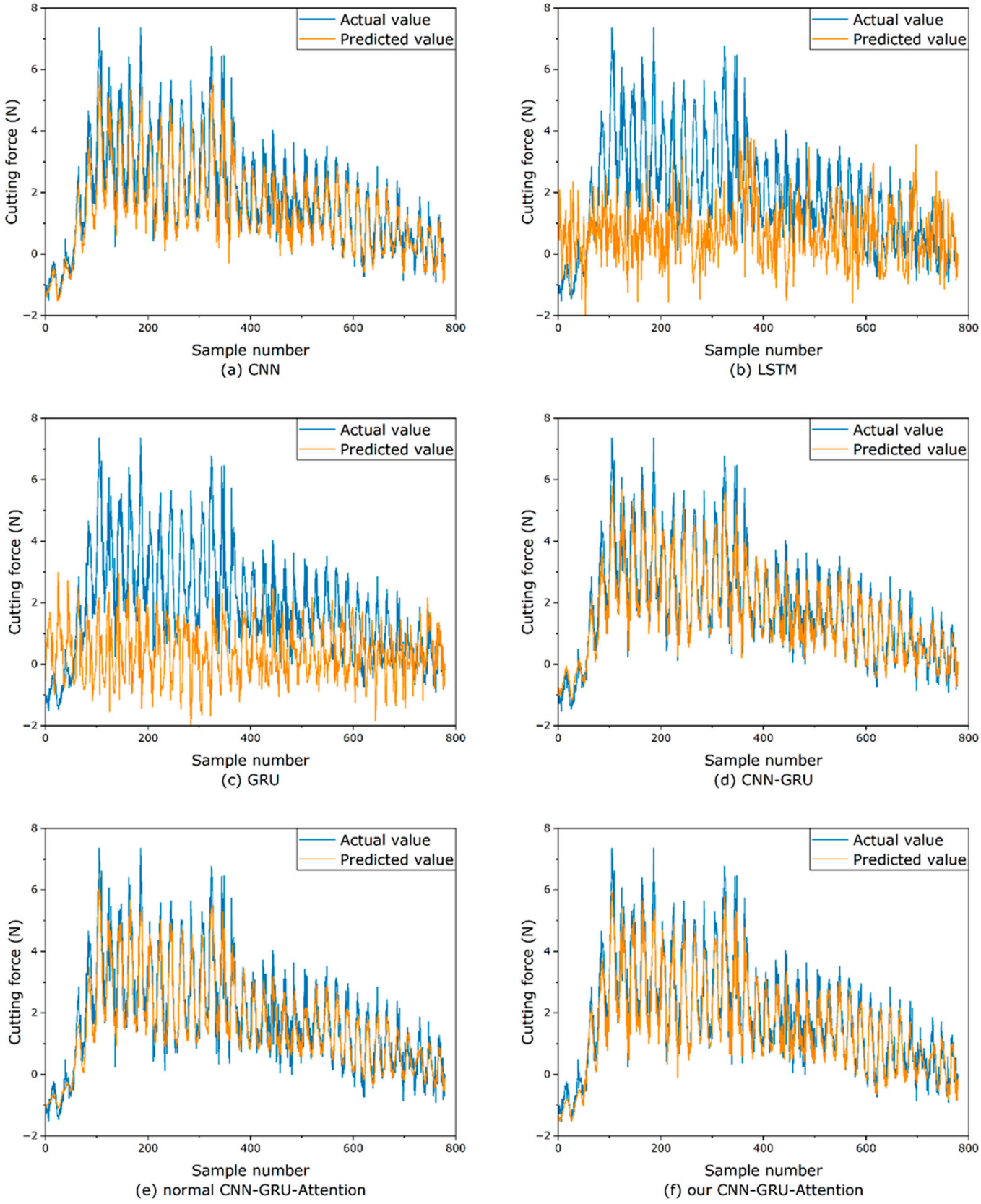


Figure 8. The prediction images of different models in the ablation experiments. (**a**) CNN model; (**b**) LSTM model; (**c**) GRU model; (**d**) CNN-GRU model; (**e**) normal CNN-GRU-Attention model; (**f**) our CNN-GRU-Attention model.

The CNN model [27] only extracts the local features of data, without considering the characteristics of data as a time series, and cannot analyze the correlation between distant data well. As an RNN model, the LSTM model [31] can extract long-distance feature correlation between time series data, but it cannot extract local correlation features between data well in the absence of the convolution layer. The GRU model [32] is a variant of the LSTM model, which simplifies the calculation steps and has similar calculation effects to the LSTM model. Comparing the prediction results of the RNN model with those of the CNN model, the values of RMSE and MAE increased by more than 200%, while the value of R^2 was negative, which means that the prediction effect of the model is very poor. Combining the CNN model with the GRU model, the local correlation characteristics and long-term time characteristics of data can be effectively extracted and analyzed, and the prediction effect is better. Compared with the results of the CNN model, RMSE and MAE decreased by more than 13%, which means that the prediction error is obviously reduced. At the same time, the value of R^2 increased by 3.8%. The CNN–GRU–Attention model introduces the attention mechanism on the basis of the CNN-GRU model, which gives different weights to the eigenvalues through the calculation of attention scores, so that the eigenvalues that have a significant influence on the prediction results can be extracted, and the prediction effect of the model is better. In order to improve the prediction performance of the model, this paper further optimizes the structure of the CNN model by using the Parallel Dilated Convective Model. Compared with the normal CGA model, the values of RMSE and MAE are reduced by 14.2% and 12.5%, respectively, and the R^2 is increased to 0.91. Items per second is 197.40 it/s, which is only a very slight decrease.

3.5. Anomaly Detection Result Analysis

As shown in Figure 9, normal analysis was performed on the residuals predicted by the CGA model. It can be seen that the normal cutting residual data conforms to a normal distribution, and an EWMA control chart can be used for anomaly detection. Due to tool wear during the processing, the cutting force of the tool deviates from the normal variation pattern after a certain amount of processing. The EWMA control chart model is used to monitor the changes in the residual sequence of cutting forces. The UCL and LCL of the EWMA control chart are determined by calculations based on the normal tool processing process data. As there may be many kinds of anomalies in the time series data of disc cutters, there is no unified quantitative index for the judgment rules of the control chart, so it is not compared with other anomaly detection models. The residual sequence determines whether the judgment rule of the EWMA control chart is effective, so it can be considered that the key factor affecting the accuracy of anomaly detection is the accuracy of the prediction model.

The EWMA control charts of cutting force residual sequences, plotted based on the machining process of disc tools with different processing lengths, are shown in Figure 10. Figure 10a shows the EWMA control chart of a disk tool after processing for 160 m. In this situation, most of the data points fall within the control limits of the EWMA statistical chart. Figure 10b shows the EWMA control chart of a disk tool after processing for 1320 m. At the beginning of the cutting process, the residual sequence experiences significant fluctuations as the first red circle marked, indicating the occurrence of the chatter phenomenon during the current cutting process and leading to noticeable high-frequency fluctuations in cutting forces. Subsequently, after entering a stable cutting process, the residual sequence consistently exceeds the control limits of the control chart for multiple consecutive values as the second red circle marked. This signifies that the cutting force of the current tool is far beyond the range of normal cutting forces. By analyzing the aforementioned changes, the occurrence of abnormal tool wear can be successfully detected. Figure 10c shows the

EWMA control chart of a disk tool after processing 1800 m. Similar to the situation in Figure 10b, the control chart can also recognize the abnormal state of the current tool based on the changing characteristics of the data. Timely detection of cutting forces enables manufacturers to promptly identify the abnormal state of the tool and replace it, reducing unnecessary material waste and ensuring the quality of the machined products. EWMA control charts need to assume that the residual sequence obeys a normal distribution, and the possible skewed distribution of data will lead to performance degradation. At the same time, their performance depends on the choice of the smoothing parameter (λ), and exploring the adaptive adjustment method of parameters may improve the flexibility of the model.

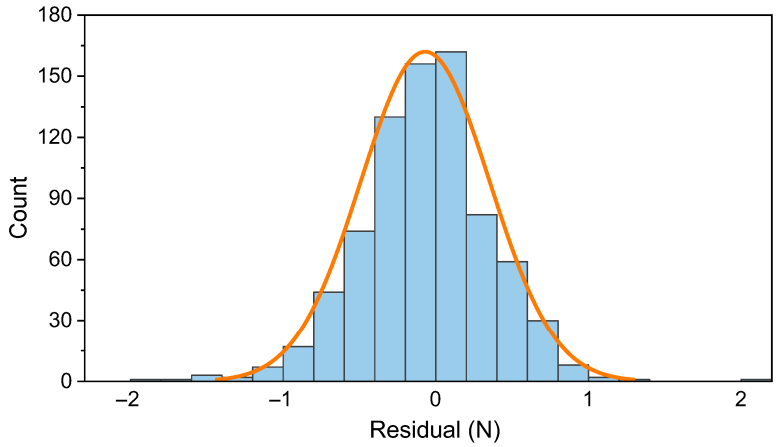


Figure 9. Residual normal distribution fitting results.

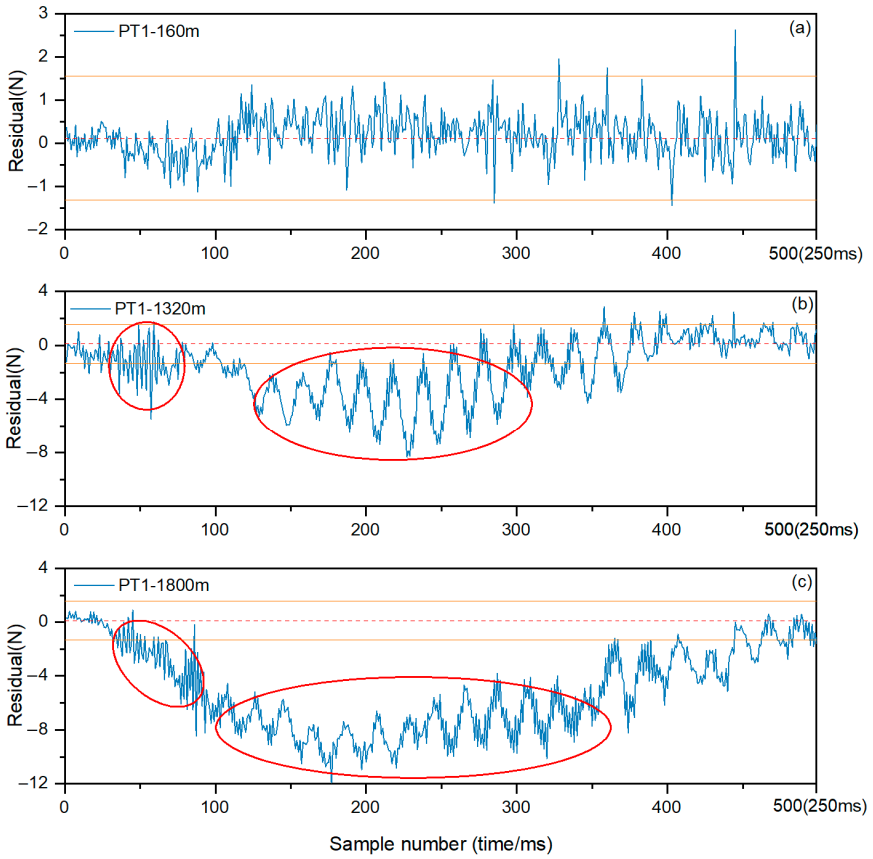


Figure 10. The images of EWMA control charts with different processing lengths (Figure (**a**), (**b**), (**c**) are 160 m, 1320 m and 1800 m respectively) with PT1.

We conducted anomaly detection on three different tools subjected to three distinct pre-treatment processes, measuring the machining distance at which significant anomalies occurred, as shown in Figure 11. PT1 served as the control group and exhibited anomalies first. PT2 underwent hardness enhancement and exhibited anomalies at approximately 1320 m of machining distance. PT3 underwent toughness enhancement and exhibited anomalies last at 2320 m. The above anomaly detection results are consistent with the conclusions from our previous research [38], indicating that the anomaly detection model has good performance. By examining the corresponding tool images, it can be observed that PT1 has low hardness and poor wear resistance, with continuous wear on the cutting edge. PT2 has high hardness but a large grain size and low toughness, resulting in large material defects. PT3 has good toughness and a uniformly fine grain size, superior wear resistance compared to PT1, and no large material defects, demonstrating better wear resistance performance.

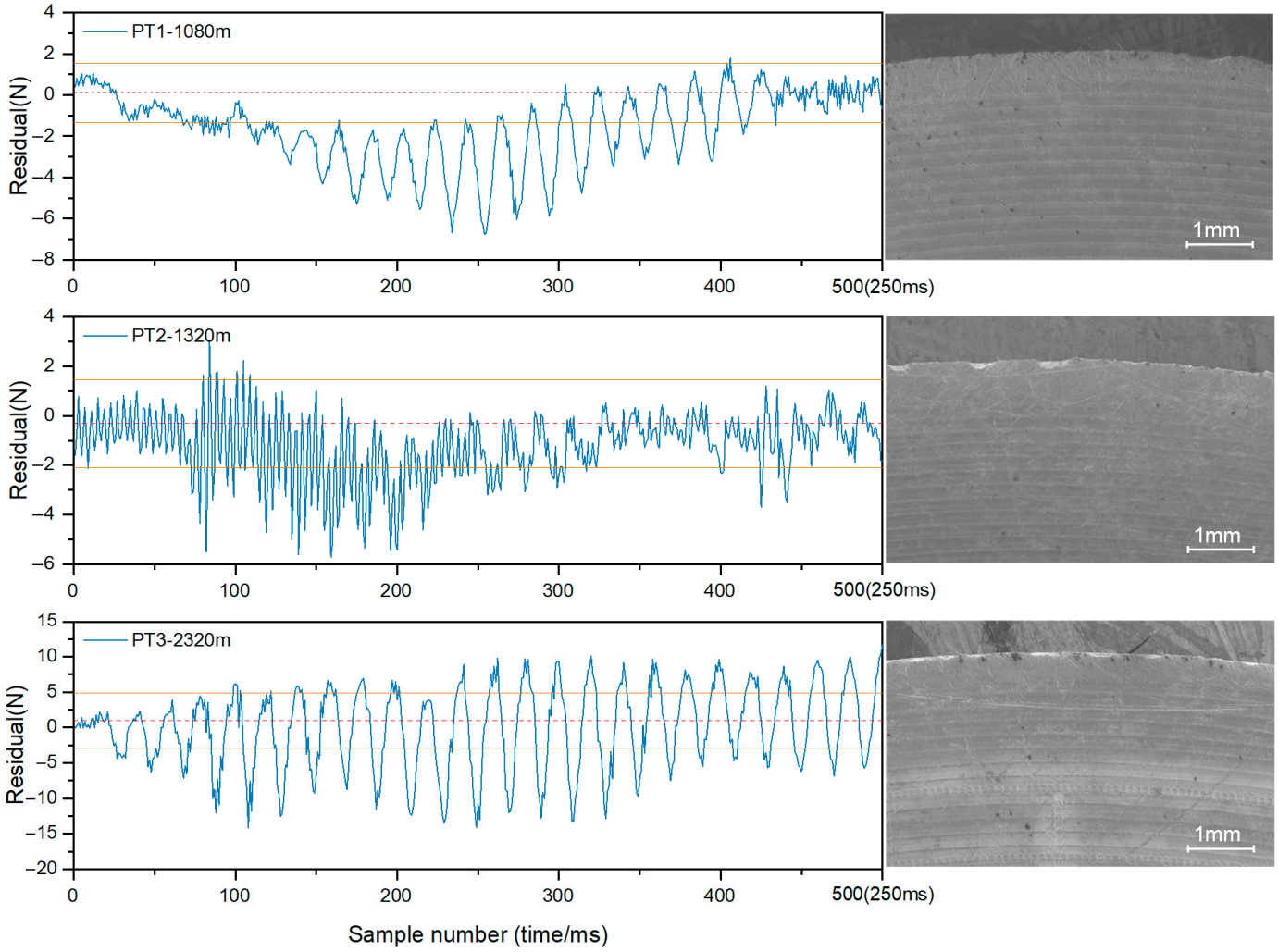


Figure 11. The images of the EWMA control charts with different tools and corresponding tool wear images.

4. Conclusions

This study proposes a real-time anomaly detection method for disc tool wear during the processing of Nomex honeycomb composites (NHCs). By combining a CNN–GRU–Attention deep learning model with an EWMA control chart, the method effectively moni-

tors the cutting force sensor data to identify abnormal tool states promptly, thus ensuring product quality and reducing material waste.

The key contributions of this work are as follows:

(1) Accurate time series prediction: The CNN–GRU–Attention model leverages a parallel dilated convolution to capture multidimensional local features, a GRU network to extract temporal dependencies, and an attention mechanism to highlight critical information, enabling highly accurate predictions of cutting force variations during the tool machining process.
(2) Robust anomaly detection: The EWMA control chart analyzes the trend of residual errors between predicted and actual cutting force data, quickly detecting subtle and gradual abnormalities caused by tool wear, which traditional point-based detection methods might miss.
(3) Practical applicability: The model is trained only with normal data, addressing the challenge of imbalanced datasets common in machining processes, and achieves real-time detection performance suitable for manufacturing environments.

Overall, the proposed model provides a concrete, effective, and computationally efficient solution to improve tool wear monitoring in NHC processing, supporting more reliable and high-quality manufacturing. However, due to limitations in experimental equipment, we were unable to conduct experiments with more processing materials and process settings. In fact, future work may focus on expanding the generalization ability of the model using more instances through methods such as boosting ensembles and optimizing model parameters automatically. The setting of hyperparameters for the deep learning prediction model proposed in this study is mainly based on experience and experiments, so that further improvements can be made to these parameters through methods such as Bayesian optimization. Moreover, higher-dimensional data can be used for anomaly detection to better judge the abnormal state of the disc tool in processing NHCs.

Author Contributions: Conceptualization, X.L.; Data curation, J.X.; Funding acquisition, P.M.; Methodology, P.T. and X.W.; Project administration, X.L. and P.M.; Resources, J.X., X.L. and P.M.; Software, X.W. and P.T.; Validation, P.T., J.X. and X.W.; Visualization, X.W. and P.T.; Writing—original draft, X.W. and P.T.; Writing—review and editing, J.X., X.L. and P.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Key Research and Development Program of China, grant number 2022YFC3901704. And The APC was funded by National Key Research and Development Program of China, grant number 2022YFC3901704.

Data Availability Statement: Data will be made available on request.

Acknowledgments: This work was supported by the National Key Research and Development Program of China (grant number 2022YFC3901704).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

NHCs	Nomex honeycomb composites
SPC	Statistical process control
SVM	Support vector machine
LSTM	Long short-term memory
CNN	Convolutional neural network
RNN	Recurrent neural network
SCADA	Supervisory control and data acquisition

EWMA	Exponentially weighted moving average
GRU	Gated recurrent unit
ReLU	Rectifier linear unit
UCL	Upper confidence limit
LCL	Lower confidence limit
RMSE	Root mean square error
MAE	Mean absolute error
R2	R-Square
TCN	Temporal convolutional network

References

1. Jaafar, M.; Makich, H.; Nouari, M. A new criterion to evaluate the machined surface quality of the Nomex® honeycomb materials. *J. Manuf. Process.* **2021**, *69*, 567–582. [CrossRef]
2. Nasri, M.R.; Salari, E.; Salari, A.; Vanini, S.A.S. Nonlinear bending and buckling analysis of 3D-printed meta-sandwich curved beam with auxetic honeycomb core. *Aerosp. Sci. Technol.* **2024**, *152*, 109339. [CrossRef]
3. Ghasemi, F.; Salari, E.; Rastgoo, A.; Li, D.; Deng, J. Nonlinear vibration analysis of pre/post-buckled 3D-printed tubular metastructures. *Eng. Anal. Bound. Elem.* **2024**, *165*, 105777. [CrossRef]
4. Mughal, K.H.; Qureshi, M.A.M.; Jamil, M.F.; Ahmad, S.; Ahmad Khalid, F.; Qaiser, A.A.; Maqbool, A.; Raza, S.F.; Zhang, J. Investigation of hybrid ultrasonic machining process of Nomex honeycomb composite using a toothed disc cutter. *Ultrasonics* **2024**, *141*, 107343. [CrossRef]
5. Xu, J.; Feng, P.; Gong, Y.; Wang, J.; Yang, H.; Feng, F. Exploiting damage for inhibiting damage: A counterintuitive reasoning out of in-situ orthogonal cutting for brittle fiber composite. *J. Mater. Process. Technol.* **2025**, *343*, 118961. [CrossRef]
6. Ahmad, S.; Zhang, J.; Feng, P.; Yu, D.; Wu, Z.; Ke, M. Processing technologies for Nomex honeycomb composites (NHCs): A critical review. *Compos. Struct.* **2020**, *250*, 112545. [CrossRef]
7. Xia, Y.; Zhang, J.; Wu, Z.; Feng, P.; Yu, D. Study on the design of cutting disc in ultrasonic-assisted machining of honeycomb composites. In Proceedings of the IOP Conference Series: Materials Science and Engineering, Kazimierz Dolny, Poland, 21–23 November 2019; Volume 611, p. 012032. [CrossRef]
8. Ortega, N.; Martynenko, V.; Perez, D.; Martinez Krahmer, D.; López de Lacalle, L.N.; Ukar, E. Abrasive Disc Performance in Dry-Cutting of Medium-Carbon Steel. *Metals* **2020**, *10*, 538. [CrossRef]
9. Xu, J.; Zhang, K.; Zha, H.; Liu, J.; Yuan, X.; Cai, X.; Xu, C.; Ma, Y.; Feng, P.; Feng, F. Surface integrity of Nomex honeycomb composites after ultrasonic vibration machining by using disc cutters. *J. Manuf. Process.* **2023**, *102*, 1010–1022. [CrossRef]
10. Ryalat, M.; ElMoaqet, H.; AlFaouri, M. Design of a smart factory based on cyber-physical systems and Internet of Things towards Industry 4.0. *Appl. Sci.* **2023**, *13*, 2156. [CrossRef]
11. Blázquez-García, A.; Conde, A.; Mori, U.; Lozano, J.A. A review on outlier/anomaly detection in time series data. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–33. [CrossRef]
12. Shi, W.; Chen, F.; Qu, F.; Zhang, R.; Lu, C. Abnormality diagnosis method for manufacturing process based on Bayesian network. *J. Xi'an Jiaotong Univ.* **2018**, *52*, 9–14. [CrossRef]
13. Niu, L.; Wang, Q.; Chen, B.; Zhao, Y.; Zhang, Y. Eco-driving decision making based on V2X communication and spatio-temporal prediction of pedestrians. *IEEE Trans. Intell. Transp. Syst.* **2025**, *11*, 11905–11915. [CrossRef]
14. Erfani, S.M.; Rajasegarar, S.; Karunasekera, S.; Leckie, C. High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning. *Pattern Recognit.* **2016**, *58*, 121–134. [CrossRef]
15. Wang, H.; Li, Q.; Liu, Y.; Yang, S. Anomaly data detection of rolling element bearings vibration signal based on parameter optimization isolation forest. *Machines* **2022**, *10*, 459. [CrossRef]
16. Xu, H.; Pang, G.; Wang, Y.; Wang, Y. Deep isolation forest for anomaly detection. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 12591–12604. [CrossRef]
17. Liu, B.; Tan, P.N.; Zhou, J. Unsupervised anomaly detection by robust density estimation. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 22 February–1 March 2022; Volume 36, pp. 4101–4108. [CrossRef]
18. Soleimani-Babakamali, M.H.; Soleimani-Babakamali, R.; Sarlo, R.; Farghally, M.F.; Lourentzou, I. On the effectiveness of dimensionality reduction for unsupervised structural health monitoring anomaly detection. *Mech. Syst. Signal Process.* **2023**, *187*, 109910. [CrossRef]
19. Aldekoa, I.; del Olmo, A.; Sastoque-Pinilla, L.; Sendino-Mouliet, S.; López-Novoa, U.; de Lacalle, L.N. Early detection of tool wear in electromechanical broaching machines by monitoring main stroke servomotors. *Mech. Syst. Signal Process.* **2023**, *204*, 110773. [CrossRef]

20. Pang, G.; Shen, C.; Cao, L.; van den Hengel, A. Deep learning for anomaly detection: A review. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–38. [CrossRef]
21. Oshida, T.; Murakoshi, T.; Zhou, L.; Ojima, H.; Kaneko, K.; Onuki, T.; Shimizu, J. Development and implementation of real-time anomaly detection on tool wear based on stacked LSTM encoder-decoder model. *Int. J. Adv. Manuf. Technol.* **2023**, *127*, 263–278. [CrossRef]
22. Munir, M.; Siddiqui, S.A.; Dengel, A.; Ahmed, S. DeepAnT: A deep learning approach for unsupervised anomaly detection in time series. *IEEE Access* **2018**, *7*, 1991–2005. [CrossRef]
23. Wang, Y.; Perry, M.; Whitlock, D.; Sutherland, J.W. Detecting anomalies in time series data from a manufacturing system using recurrent neural networks. *J. Manuf. Syst.* **2022**, *62*, 823–834. [CrossRef]
24. Zhang, L.; Wang, B.; Yuan, X.; Liang, P. Remaining useful life prediction via improved CNN, GRU and residual attention mechanism with soft thresholding. *IEEE Sens. J.* **2022**, *22*, 15178–15190. [CrossRef]
25. Zhang, J.; Li, S. Air quality index forecast in Beijing based on CNN-LSTM multi-model. *Chemosphere* **2022**, *308*, 136180. [CrossRef]
26. Kong, Z.; Tang, B.; Deng, L.; Liu, W.; Han, Y. Condition monitoring of wind turbines based on spatio-temporal fusion of SCADA data by convolutional neural networks and gated recurrent units. *Renew. Energy* **2020**, *146*, 760–768. [CrossRef]
27. Li, Z.; Liu, F.; Yang, W.; Peng, S.; Zhou, J. A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 6999–7019. [CrossRef]
28. Glorot, X.; Bordes, A.; Bengio, Y. Deep sparse rectifier neural networks. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, Fort Lauderdale, FL, USA, 11–13 April 2011; pp. 315–323.
29. Xu, B. Empirical evaluation of rectified activations in convolutional network. *arXiv* **2015**, arXiv:1505.00853. [CrossRef]
30. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
31. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
32. Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv* **2014**, arXiv:1406.1078. [CrossRef]
33. Bahdanau, D.; Cho, K.; Bengio, Y. Neural machine translation by jointly learning to align and translate. *arXiv* **2014**, arXiv:1409.0473.
34. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), Long Beach, CA, USA, 4–9 December 2017; pp. 6000–6010.
35. Roberts, S. Control chart tests based on geometric moving averages. *Technometrics* **2000**, *42*, 97–101. [CrossRef]
36. Sukparungsee, S.; Areepong, Y.; Taboran, R. Exponentially weighted moving average—Moving average charts for monitoring the process mean. *PLoS ONE* **2020**, *15*, e0228208. [CrossRef]
37. Saleh, N.A.; Aly, A.A.; Mahmoud, M.A. Effect of contaminated phase I data on the phase II–EWMA control chart performance under non-normality. *Commun. Stat. Simul. Comput.* **2024**, *53*, 4430–4448. [CrossRef]
38. Xu, J.; Yue, Q.; Zha, H.; Yuan, X.; Cai, X.; Xu, C.; Ma, Y.; Feng, P.; Feng, F. Wear reduction by toughness enhancement of disc tool in Nomex honeycomb composites machining. *Tribol. Int.* **2023**, *185*, 108475. [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.