Mechanical Systems and Signal Processing 224 (2025) 111949

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

Application of physics-guided deep learning model in tool wear monitoring of high-speed milling

Shenshen Li [a], Jun Li [a], Kunpeng Zhu [a,b,*]

[a] Institute of Precision Manufacturing, School of Machinery and Automation, Wuhan University of Science and Technology, Wuhan 430081, China
[b] Institute of Intelligent Machines, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Changzhou 213164, China

A B S T R A C T

As an important part of intelligent machining system, tool wear monitoring plays a crucial role in ensuring workpiece quality and process safety. At present, the models based on tool wear monitoring mainly include data-driven models and physics-based models. However, data-driven models are limited by physical inconsistency, and physics-based models usually lack accurate description of the machining process for process control. To solve these issues, this study proposes a physics-guided deep learning model for tool wear monitoring. Firstly, a dual scale time series model is established, and physical constraints are added to the model according to the degradation characteristics of tool wear. Secondly, the physics-based loss function is introduced through the physical model of tool wear to constrain the training process of the model. Finally, a model agnostic meta learning algorithm is used to train a pre-weight for the model, so that the model can be quickly applied to different processing conditions. The experimental results show that the physics-guided deep learning model proposed has good accuracy and physical consistency. In the high-speed milling tests, the average MAPE of tool wear prediction is 0.039, and the physical consistency index is 0.035 μm.

1. Introduction

As one of the important technologies in modern precision machining system, high-speed milling is widely applied in the industry of aerospace, mold manufacturing, medical devices and other precision and complex parts [1,2]. Condition monitoring and fault diagnosis during machining are crucial for improving the reliability of machining system operation and reducing the cost of downtime maintenance [3,4]. It has been shown that the downtime of tool faults accounts for 20 % [5] of the whole machine tool downtime in high-speed milling. By on-line monitoring of tool wear condition during processing, the cost of tool can be saved by about 40 % [6]. Therefore, online monitoring of tool wear condition is crucial to the machining process control and optimization. The current tool wear monitoring (TWM) models are mainly divided into physical-based models, data-driven models and hybrid models [7,8].

Physics-based tool wear monitoring models can be divided into empirical formula method and mechanism modeling method [9]. The former establishes tool wear monitoring models through extensive experiments and experience accumulation. For example, Zhang et al. [10] established an empirical model between the multi-channel specific cutting force coefficient (SCFC) and tool wear based on the relationship between milling forces, spindle box vibration, and cutting torque. Bai et al. [11] established a semi-analytical model by experimentally determining two wear coefficients. Zhu and Zhang [12] considered the characteristics of tool wear at different stages and proposed a general tool wear model with adjustable coefficients. The latter is to establish the monitoring model by studying the mechanism of tool wear. For example, Yang et al. [13] used a mixed mechanism wear model based on abrasive wear, adhesive wear, and diffusion wear to predict tool wear values. Das et al. [14] established a tool wear model based on the stress distribution and temperature characteristics on the rake face. Seeholzer et al. [15] proposed a new analysis model based on fiber orientation and tool

* Corresponding author.
E-mail addresses: liss1997@wust.edu.cn (S. Li), jaxsunlee@wust.edu.cn (J. Li), zhukp@iamt.ac.cn (K. Zhu).

https://doi.org/10.1016/j.ymssp.2024.111949
Received 5 May 2024; Received in revised form 24 July 2024; Accepted 11 September 2024
Available online 25 September 2024

structure to predict tool wear in carbon fiber reinforced polymer machining.

Under the background of rapid development of industrial big data and deep learning technology, the data-driven tool wear monitoring model has attracted wide attention [16]. For data-driven models, they predict tool wear through machine learning or deep learning. For example, Zhou et al. [17] proposed an improved multi-scale edge marking graph neural network to predict tool wear conditions by coding multi-dimensional data into a gray recursive graph to extract wear features. Guo et al. [18] proposed a multi-scale pyramid attention network for predicting tool wear conditions based on interpretability. Marei et al. [19] proposed a convolutional neural network based on transfer learning to solve the problem of limited data.

Although the methods of physical models and data models are widely studied today, they are rarely applied to actual production processes. This is because they both have their own limitations or drawbacks. For physical models, they are established under ideal experimental conditions, so physical models usually cannot accurately reflect the actual processing process. And these physical models are not subject to real-time feedback control of the machining process, so there may be error accumulation [20,21]. For data-driven models, firstly, they belong to the black box model with poor interpretability [22]. Secondly, in order to ensure the accuracy of model predictions, these models rely entirely on the learning of a large amount of full life experimental data, which not only requires expensive experimental costs, but also carries the risk of overfitting [23].

Based on the limitations of the aforementioned models, researchers have combined the strengths of both approaches by using physical properties or laws to drive or constrain data models, ultimately achieving the fusion of mechanisms and data [24,25]. In the study of tool condition monitoring, the hybrid models of physics and data can be roughly divided into three categories [26]:

1) The hybrid model based on the combination of data model and physical model, which fuse the results to improve the model performance. Huang et al. [27] proposed a hybrid data-driven physical model framework for predicting tool wear. The physical model establishes the mathematical description of tool wear degradation process. The data driven model is based on multilayer perceptron to predict tool wear. Hanachi et al. [28] proposed a hybrid framework to predict the tool wear condition by combining the results based on the physical model and the data-driven model through the regularization particle filter algorithm. Ma et al. [29] proposed a hybrid-driven probabilistic state space model for tool wear monitoring, which innovatively developed Gaussian processes to integrate data mining and physical models from a probabilistic perspective.
2) The hybrid model based on physical pre-training improves the training and consistency of the model on the basis of constrained model initialization. Jia et al. [30] proposed a physically guided recurrent neural network model. An important aspect of this method lies in its ability to incorporate the knowledge encoded in physics-based models. This allows training the model using very few true observed data while also ensuring high prediction accuracy. Yang et al. [31] used a hybrid model comprised of a physics-based model and a data-driven model to generate power and force signals, followed by Page's cumulative sum test for detecting tool wear on-line using the computer numerical control machine measurements.
3) The hybrid model based on physics-guided loss function, which constructs a regularized loss term through prior knowledge or physical model to improve the physical consistency and generalization ability of the data model. In the process of predicting tool wear conditions, Wang et al. [32] added monotonicity loss function to the proposed end-to-end deep learning model to capture the degradation characteristics in these sequence outputs. Wang et al. [33] proposed a modeling strategy of cross physical data fusion, and added a loss function to the model to eliminate the physical inconsistency in the data-driven model. Li et al. [34] proposed a physically informed loss term based on the monotonic characteristics of tool wear to learn the robust relationship between tool wear rate and force during the life cycle, solving the difficulty of parameter estimation. In the process of predicting fatigue wear life, Zhou et al. [35] constructed a composite loss function with physical constraints. By using a negative log likelihood function to consider fault data and depletion data, they forced the network training process to learn a continuous function to describe the relationship between pressure and life. In predicting the surface roughness of milling processes, Zeng et al. [36] added physical consistency constraints to the loss function, allowing the model to predict surface roughness on limited experimental data. Compared with the best comparison method, the average absolute error on the test set decreased by an average of 3.029 %.

Although the hybrid model based on current tool wear condition monitoring has largely addressed the limitations of physical-based and data-driven models, it also has its own limitations. Firstly, the physical mechanism in the current hybrid model is mainly based on prior knowledge of tool wear, that is, the trend of monotonic increase in tool wear as the process progresses. It does not reflect the deep physical knowledge of the variation of tool wear rate in different wear stages, and lacks accurate description of nonlinear complex systems. Secondly, although pre-training-based hybrid models have partially resolved the issue of data model initialization and reduced the training difficulty of data models, they require significant time and financial costs to be pre-trained for different operating conditions in intelligent machining processes. Consequently, this pre-training-based hybrid model performs poorly in complex scenarios. Finally, many current hybrid models only loosely combine predictions from both sides, and their predictions heavily rely on the prediction errors of the physical and data models. The effectiveness of the hybrid model diminishes when the prediction errors of the physical and data models exceed a certain threshold. Therefore, from the fundamental problem of data models lacking guidance from physical mechanisms has not been effectively solved from a model structure perspective.

To meet there difficulties, this study proposes a physics-guided deep learning model for monitoring tool wear. Based on the characteristics of the force signals and the tool wear process, a dual scale time series model is established. To reduce the physical inconsistency of the data model and minimize the parameter optimization space of the data model, a tool wear mechanism and a physical loss penalty term are introduced to constrain the training process of the data model. To predict tool wear under multiple process conditions, the model parameters are updated using a model agnostic meta learning (MAML) algorithm, which obtains generalized model parameters and improves the model's generalization performance. The organizational structure of this paper is as

follows: Section 2 introduces the framework of tool wear monitoring using a physics-guided deep learning model; Section 3 provides a detailed description of the physics-guided deep learning model for tool wear monitoring; Section 4 evaluates the model's performance using experimental data sets and baseline data sets from milling; and Section 5 presents the conclusions of this study.

2. The framework of tool wear monitoring by physics-guided deep learning model

Compared to traditional milling, high-speed milling is more susceptible to tool wear due to its large cutting speed and high cutting accuracy [37,38]. Therefore, tool wear monitoring plays an important role in ensuring machining quality and efficiency. Milling force, as a direct manifestation of the interaction between the tool and the workpiece, is a highly sensitive signal for tool wear monitoring [39–41]. However, due to the discontinuity of the milling process and the asynchrony of signal acquisition, the milling force signal and tool wear value belong to different time series signal types at different time scales. Their characteristics are as follows:

1) For the milling force signal, it is regarded as the superposition of static force and dynamic force in our research. The static force is the force generated by the discontinuous milling process of the milling cutter, while the dynamic force is mainly the additional force generated by tool wear. Therefore, milling force can be seen as a locally periodic high-frequency time series signal.
2) For the process of tool wear, it can be regarded as a gradual degradation process. The tool wear condition gradually evolves from a slightly wear condition to a severely wear condition, and the tool wear process has a monotonic increasing characteristic.

In order to accurately predict changes in tool wear conditions during the machining process in real-time, a physics-guided deep learning model is constructed in this study. The prediction process of the model is shown in Fig. 1. Firstly, based on the above analysis of the characteristics of milling force and tool wear, a dual scale time series model is constructed. Based on the prior knowledge and physical model of tool wear, monotonicity constraints and physics-based loss functions are added to the structure of the dual scale time series model. This reduces the physical inconsistency of model prediction and makes the training process more stable. Secondly, in the training process of the model, the MAML algorithm is used to train a pre-weight for the model, which enables the model to quickly adapt to different processing conditions, thus improving the generalization ability of the model. Finally, the trained model is used for online tool wear monitoring.

3. The physics-guided deep learning model for tool wear condition monitoring

3.1. The structure of the physics-guided deep learning model

For the research of tool wear condition monitoring, the traditional deep learning method completely relies on the powerful nonlinear fitting capabilities of neural networks to establish the nonlinear mapping relationship between monitoring signals and tool wear conditions, without considering the potential law of tool wear degradation process. Moreover, based on the characteristics of the monitoring signal and the tool wear value described in section II, forcibly relying on the deep learning method to predict the tool wear

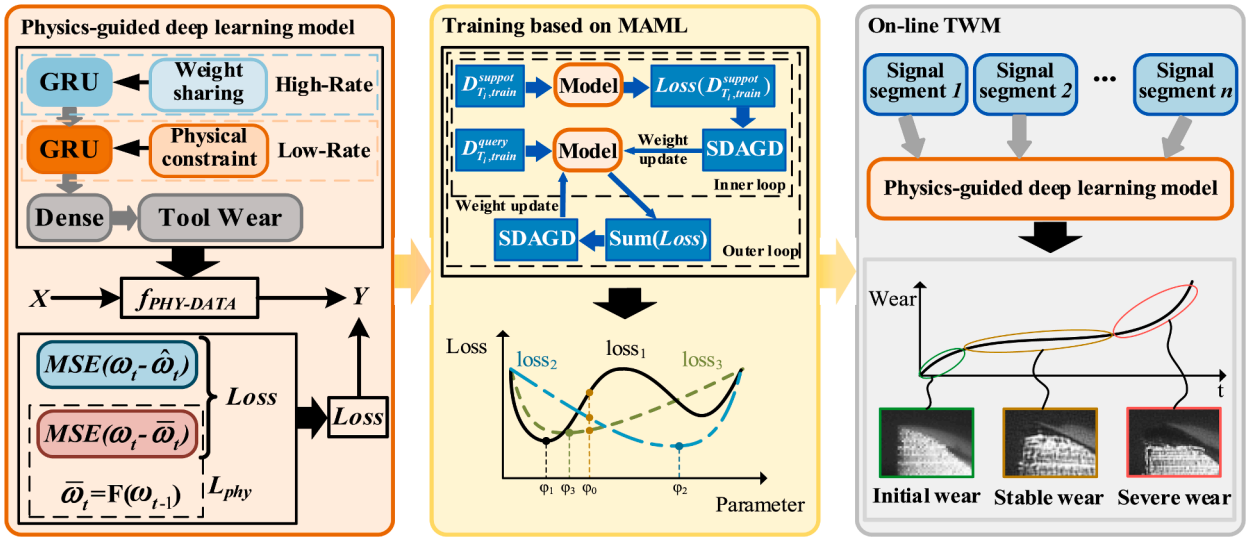


Fig. 1. Framework of the physics-guided deep learning model for tool wear monitoring (High-Rate and Low-Rate represent two different time scales of local and global changes in the signal, respectively; *X* represents the milling force signal collected by the sensor, which is the input of the model; *Y* represents the prediction result of the model; SDAGD refers to Stochastic Diagonal Approximate Gradient Descent used for optimizing the model parameters; φ_0 represents the initial parameters of the model learned based on MAML, and φ_i (i = 1…n) refers to the optimal model parameters under the i-th subtask.).

condition will inevitably lead to too many network layers and prone to overfitting. Based on the above reasons, this study proposes a dual scale time series model based on physical guidance to predict tool wear values. In the selection of the basic time series network, this study takes into consideration that GRU has fewer model parameters compared to LSTM. Therefore, the complexity of the GRU model will be relatively lower, making it more suitable for applications with limited training resources. The specific model architecture is shown in Fig. 2.

It mainly consists of two parts: (1) A gated recurrent neural network layer (GRU) with shared weights, which extracts milling force features from a given milling force signal segment; (2) The physics-guided GRU model (PGA-GRU), which predicts the tool wear condition according to the milling force characteristics extracted from the previous layer. We will describe these two parts in detail below.

3.1.1. The gated recurrent neural network

The milling force signal is a kind of time series signal that can reflect the condition change of machining process. For the processing of time series signal, this paper uses GRU model to extract the features that can reflect the tool wear condition. GRU network is a recurrent neural network, which has a simpler structure than LSTM network. It is proposed to solve the long-term memory problem and the gradient explosion problem in back propagation. Compared with LSTM network, GRU network has only two gates and cancels the memory unit for linear self-updating [42]. This makes the GRU network have fewer parameters and simpler structure, so it is easier to train the model when there are fewer training samples. The forward propagation formula for GRU network is as follows:

$$r_t = \sigma(W_r x_t + R_r h_{t-1} + b_r) \tag{1}$$

$$z_t = \sigma(W_z x_t + R_z h_{t-1} + b_z) \tag{2}$$

$$\overline{h}_t = tanh(W_{\overline{h}} x_t + R_{\overline{h}}(r_t * h_{t-1}) + b_{\overline{h}}) \tag{3}$$

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

Where W and R are the weight matrices of input state x_t and hidden state h_{t-1}; b represents bias term; r, z and $\overline{h}$ represent reset gate, update gate and candidate condition respectively; $\sigma(x) = sigmoid(x)$ represent the activation function of the gate; $tanh(x)$ represents

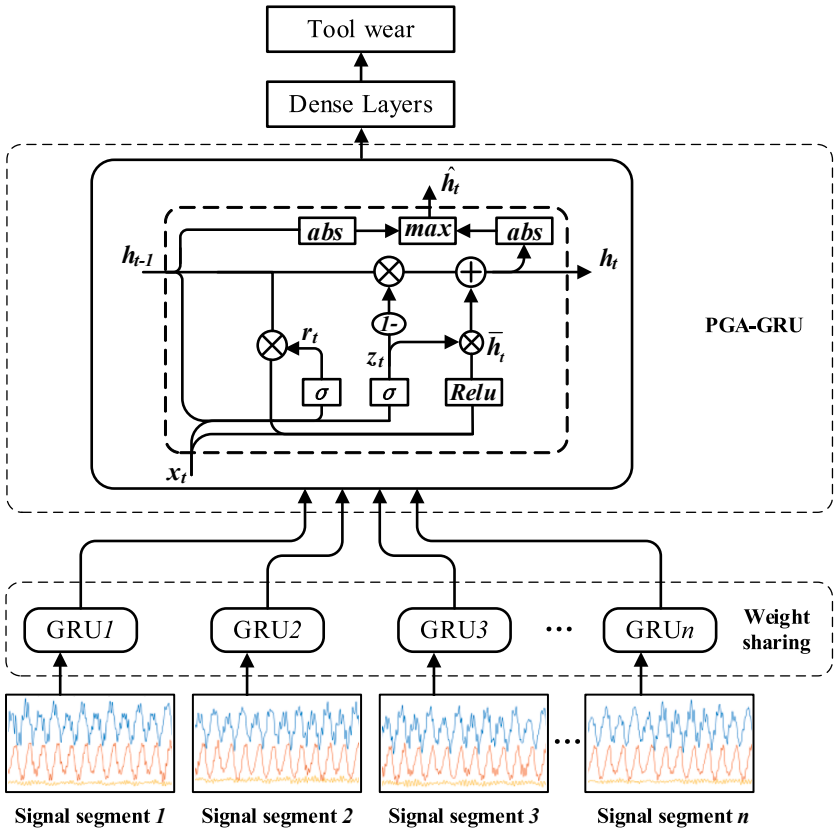


Fig. 2. Physics-guided dual scale time series model.

the activation function from input to output. The structure of GRU network is shown in Fig. 3(a).

3.1.2. The physics-guided GRU model

In the milling process, tool wear has irreversible physical characteristics with processing time. For the traditional deep learning model, it only establishes the relationship between the observed signal and the hidden tool wear condition, that is, it uses the information of the observed signal to infer the degradation process of the hidden tool wear condition, without considering the potential physical relationship of the degradation process of the hidden tool wear condition. Based on the above problems, this study constructs a GRU model based on physical guidance. The framework of this model is shown in Fig. 3(b).

Firstly, compared with the original GRU model, this model adds a monotonic increasing layer, which makes the results predicted by the model more in line with the real change law of tool wear. Secondly, the model also replaces the input–output activation function with the *Relu* activation function. Compared with *tanh* activation function, *Relu* has a wider activation boundary, which can effectively suppress the gradient disappearance problem. Moreover, due to the low computational complexity of *Relu* activation function, its computational speed is faster. The forward propagation process of the physics-guided GRU model proposed in this study is as follows:

$$r_t = \sigma(W_r x_t + R_r h_{t-1} + b_r) \tag{5}$$

$$z_t = \sigma(W_z x_t + R_z h_{t-1} + b_z) \tag{6}$$

$$\overline{h}_t = Relu(W_{\overline{h}} x_t + R_{\overline{h}}(r_t * h_{t-1}) + b_{\overline{h}}) \tag{7}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \overline{h}_t \tag{8}$$

$$\widehat{h}_t = max(abs(h_{t-1}), abs(h_t)) \tag{9}$$

Therefore, by introducing a monotonicity mechanism in the GRU model, it provides a constraint that aligns with our understanding, enabling the model to capture underlying relationships and behaviors according to our intentions. This improves the interpretability of the model and enhances the physical consistency of the predicted results.

3.2. Loss function of the physics-guided deep learning model

3.2.1. The loss function of mean square error

For the data-driven model, the loss function is the mean squared error between the model predicted values and the actual observations, as shown in the following formula:

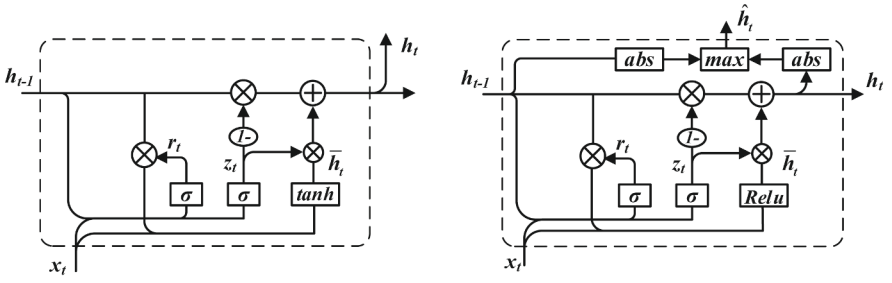
$$\underset{f}{\text{argmin}} Loss_{Data}(\widehat{\omega}, \omega) = \underset{f}{\text{argmin}} \frac{1}{n} \sum_{i=1}^{n} (\omega_i - \widehat{\omega}_i)^2 \tag{10}$$

Where ω is the model predicted values and $\widehat{\omega}$ is the actual observations.

For the data-driven model, the prediction accuracy is affected by the size of the training set and the complexity of the physical model of the prediction problem. When the training set of the model is small or the physical model of the prediction problem is complex, the prediction effect of the data model may not be very good. To solve the above problems, this study improves the prediction accuracy of the model by adding physical constraints to the model loss function.

3.2.2. The loss function of tool wear mechanism

For the construction of the physical loss function for tool wear mechanism, we first fit the parameters of the physical model for tool wear based on the offline measured tool wear values. Then, we establish the mechanistic relationship between adjacent tool wear values based on the physical function of tool wear. Finally, we incorporate the established relationship between adjacent tool wear



(a) Structure diagram of traditional GRU network (b) Structure diagram of physics-guided GRU model

Fig. 3. Structure diagram of GRU model.

values into the neural network to constrain the prediction results of the model. The detailed discussion of this process is shown below.

During the machining process, tool wear is unavoidable due to excessive friction and heat generated on the surface of the tool. The evolution process of tool wear can be divided into three stages: initial wear stage, steady wear stage, and severe wear stage. In order to accurately describe this degradation process, researchers have developed two approaches. On the one hand, researchers have established a tool wear prediction model based on accumulated work experience and experimental data of tool wear, which is called tool wear modeling based on empirical formulas [10–12]. On the other hand, researchers have established a tool wear prediction model by studying the mechanism of tool wear, which is called tool wear modeling based on mechanism [13–15].

Due to the variation of tool wear rate with different wear stages, Zhu et al. [12] established an empirical tool flank wear model based on adjustable model coefficients. According to the physical process of wear, the model uses logarithmic polynomial and high order exponential functions to fit the early wear and late wear of tools. When the model parameters are known, the model is a function of tool wear and milling time. The details are as follows:

$$\omega(t) = A\ln(Bt+1) + Ct^D \tag{11}$$

Where ω represents the flank wear width, t represents the milling time, and A, B, C and D represent the parameters of the model. Due to the adjustability of the model parameters, the model can easily predict the tool wear condition under variable processing conditions.

Although physical models of tool wear have been widely studied, they have inherent shortcomings. The physical model based on ideal conditions often lacks accurate description of the actual machining process, and the models are not subject to real-time feedback control, leading to error accumulation. For these reasons, this paper uses the numerical analysis method to improve the physical model to obtain the relationship between the tool wear condition at the current time and the tool wear condition at the previous time, so as to establish the relationship between the physical model and the machining process. The specific process is as follows:

$$F = \frac{\frac{AB}{Bt+1} + CDt^{D-1}}{A\ln(Bt+1) + Ct^D}\omega \tag{12}$$

$$K_1 = F(t_n, \omega_n) \tag{13}$$

$$K_2 = F(t_n + \frac{1}{2}h, \omega_n + \frac{1}{2}hK_1) \tag{14}$$

$$K_3 = F(t_n + \frac{1}{2}h, \omega_n + \frac{1}{2}hK_2) \tag{15}$$

$$K_4 = F(t_n + h, \omega_n + hK_3) \tag{16}$$

$$\omega_{n+1} = \omega_n + h(1/6K_1 + 2/6K_2 + 2/6K_3 + 1/6K_4) \tag{17}$$

In the equations (12–17) above, equation (12) converts the formula for tool wear as a function of time into an ordinary differential equation, thereby transforming the relationship between tool wear and time into a relationship involving tool wear, tool wear rate, and time. Equation (12) establishes the relationship between the physical model and the machining process through numerical analysis, serving as a prerequisite. In equations (13–16), K_1, K_2, K_3, and K_4 represent the slopes at the starting point, midpoint, 3/4 point, and endpoint, respectively, in the 4-order Runge-Kutta method. These intermediate variables are crucial in iteratively estimating the changes and improving the accuracy of the numerical solution. Equation (17) calculates the weighted average of the intermediate variables (K_1, K_2, K_3, and K_4) obtained from equations (13–16) and adds it to the current position to obtain an approximate value for the next step's position. Here, h represents the time step size.

The physical relationship between adjacent tool wear conditions established by equations (12–17) is used to construct a physical

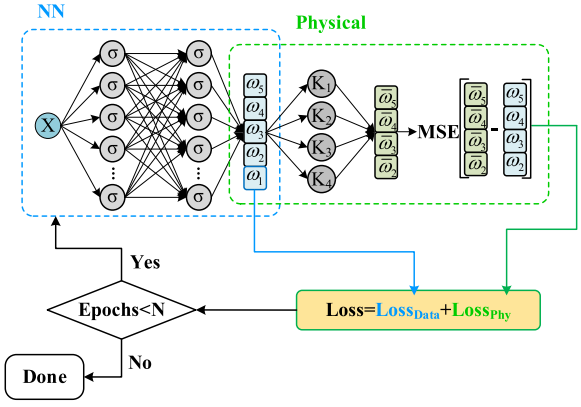


Fig. 4. Loss function structure of tool wear mechanism.

loss term that constrains the relationship between adjacent predicted wear values of the model and reduces physical inconsistency in model predictions. In addition, in actual machining, real wear label values can only be collected offline, which means that monitoring signals do not have enough label matches. Therefore, the physical labels generated by the physical loss term also alleviates the problem of lack of real wear label samples to a certain extent. For the rationality of the physical loss function, firstly, the physical model is an adjustable coefficient tool wear model established based on the wear characteristics of different wear stages, which can accurately reflect the evolution process of wear conditions. Secondly, the physical loss function imposes constraints on the data model. It can be seen as introducing learning biases with clear physical meanings to guide the learning process of the data model, making it tend to converge towards solutions with underlying physical meanings.

The loss function structure of tool wear mechanism is shown in Fig. 4, the physics-based loss function is the mean square error between the labels generated by the physical model and the predicted values of the model. The process is as follows:

$$\overline{\omega}_i = \omega_{i-1} + h(1/6K_1 + 2/6K_2 + 2/6K_3 + 1/6K_4) \tag{18}$$

$$Loss_{Phy}(\omega) = \frac{1}{n}\sum_{i=2}^{n}(\overline{\omega}_i - \omega_i)^2 \tag{19}$$

Where $\overline{\omega}_i$ is the physical prediction. The physics-based loss function establishes the physical relationship between adjacent wear conditions, and the physical wear value in equation (18) is calculated from the model prediction value at the previous time. Therefore, the difference between the predicted values of the physics-based tool wear model and the predicted values of the physical model indirectly reflects the physical inconsistency of the prediction results at the adjacent time of the model. By back propagation of error to update the network weights, the predicted tool wear condition can be more consistent with the physical law of tool wear, avoiding unrealistic predictions that may occur in purely data-driven models. Additionally, the physical loss function based on the physical relationships between adjacent wear conditions can fill in missing wear information or extrapolate from existing wear values, thereby achieving more robust and reliable predictions.

In summary, the structure of the loss function of the physics-guided deep learning model is as follows:

$$\underset{f}{\operatorname{argmin}} Loss(\widehat{\omega}, \omega) = \underset{f}{\operatorname{argmin}}(Loss_{Data}(\widehat{\omega}, \omega) + \lambda_{Phy} Loss_{Phy}(\omega)) \tag{20}$$

Where λ_{Phy} represents the physical loss weight term. The optimization process of the model is realized by gradient derivation and back propagation of the loss value. The selection of the weight of the physical loss term has a great impact on the training process of the model and the final prediction accuracy. Therefore, in different stages of model training, it is necessary to adjust the weight of physical loss term to avoid the influence of invalid loss on model training. For the setting of the weight of the physical loss term, this study considers the following factors:

- The value of the weighting coefficient is determined according to the size relationship between the data loss term and the physical loss term, so as to avoid that the update process of the network is dominated by a single loss term.
- In different training stages, the weighting coefficients of the loss items should also be adjusted to avoid useless loss values updating the network weights so that the network does not converge in the training process. In this study, we used "Epoch" to dynamically adjust the weights between the physical loss and the data loss. During the initial stages of training, with randomly initialized network weights, a lower weight value λ_{Phy} was used for the physical loss. This allowed the model to primarily focus on training based on the data loss. In the later stages of model training, when the descent rate of the data loss function becomes significantly slower, the weight value λ for the physical loss is increased. This further improves the training efficiency of the model by incorporating physical consistency.

Through the above discussion of the physical loss weight λ_{Phy}, the selection of the loss weight λ_{Phy} is related to $Loss_{Data}$ and $Loss_{Phy}$, and also related to the training process of the network. Thus the formula for the weight λ_{Phy} is as follows:

$$\lambda_{Phy} = 1 - (Sigmoid\frac{Loss_{Data}}{Loss_{Phy}})^{epoch} \tag{21}$$

3.3. The model performance evaluation methods

In order to verify the accuracy of the model proposed in this study, this paper uses three evaluation indicators: mean absolute error (*MAE*), mean absolute error percentage (*MAPE*) and physical consistency (*Consistency*) to quantify the prediction effect of different methods on tool wear. *MAE* and *MAPE* are used to characterize the accuracy of model prediction. *Consistency* refers to the degree of violating the irreversibility of tool wear. In this paper, the average absolute error of the wear value predicted by the model in violation of the physical relationship value is taken as the physical consistency index. The formulas are as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\widehat{y}_i - y_i| \tag{22}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{23}$$

$$Consistency = \frac{1}{n}\sum_{i=1}^{n}\begin{cases}|y_{i+1} - y_i| & \text{if} \quad y_{i+1} - y_i \leqslant 0\\ 0 & \text{if} \quad y_{i+1} - y_i > 0\end{cases} \tag{24}$$

4. Model validation and application

4.1. Experimental setup

High-speed milling is generally defined as milling with a small material removal area but fast speed. Compared to traditional milling, high-speed milling provides faster speeds (15000 rpm to 40000 rpm), higher accuracy, and lower tool wear. In our experiment, we used a micro milling cutter with a diameter of 0.5 mm. It typically operates at the micron scale. This micro milling process utilizes high-speed rotating tools to rapidly remove material from the workpiece with small cutting parameters and relatively high feed rates. Therefore, from a process perspective, micro milling can be seen as a special form of high-speed milling that focuses on smaller scales and higher precision.

The tool used in the experiment is a double-edged cemented carbide micro end milling cutter produced by Union, Japan, and the milling cutter coating is UT coating. The workpiece material is steel (40CrNi2Mo). In order to verify the influence of machining conditions on tool wear, Taguchi experiment was used to select the feed rate per tooth, spindle speed and axial cutting depth as the influencing factors. Each influencing factor is divided into 3 levels, and milling experiments under 9 processing conditions are carried out. The parameter settings are shown in Table 1 below:

The milling force signal acquisition device and the tool wear value measurement device are shown in Fig. 5 below. For the acquisition device of the milling force signal, this experiment uses a three-channel dynamometer with a signal amplifier and a data acquisition card for acquisition. The dynamometer model is Kistler 9119A, with a sampling frequency of 24000 Hz. The acquisition process of the force signal involves fixing the force sensor at the bottom of the workpiece for measurement. There is no manual intervention throughout the acquisition process, so it is generally assumed that the measurement process is not affected by human measurement errors. Regarding the instrumental measurement error, the Kistler 9119A triaxial force sensor has a resolution less than 0.002 N and a sensitivity of approximately -13pC/N∼-26pC/N. In our experiment, the minimum change interval of the force signal within one rotation cycle is approximately 0.3 N, which is two orders of magnitude larger than the sensor's resolution (0.3 N $\gg$ 0.002 N). Therefore, the error generated by the instrumental measurement can be considered negligible.

For the collection of tool wear tags, the tool wear observation platform consisting of a coordinate regulator, an industrial camera, a telecentric lens, and a light source collects tool wear values. The resolution of the industrial camera is 2 million pixels. The telecentric lens model is VS-TCH6-65 with a magnification of 6× . The selected light source is a circular vertical light source with adjustable brightness. The offline collection frequency of tool wear tags is observed after the tool milling is fixed length, and the observation process covers the entire tool wear cycle. The measurement value of tool wear is obtained by averaging five measurements of tool wear images captured by an electron microscopy. The errors generated during the measurement process can be categorized into human measurement error (Δ_A) and instrumental measurement error (Δ_B). The human measurement error (Δ_A) is calculated based on five independent measurement values. The instrumental measurement error (Δ_B) is determined using the distortion rate of an electron microscope. Finally, the two uncertainty error components are squared, summed, and combined to obtain the final measurement error (Δ). The final calculation results indicate that the error resulting from the uncertainty of the measurement can be considered negligible compared to the final measured tool wear value. Taking an example of the measurement process for a wear value, the calculation process is as follows:

$$\bar{y} = \frac{1}{5}(0.074182 + 0.074081 + 0.073273 + 0.075371 + 0.073265) = 0.074034mm \tag{25}$$

Table 1
Parameter setting of the experiments.

Test	Spindle n speed(rpm)	Cutting a_p depth(μm)	Feed f_t (μm/tooth)	Number of measurements
T1	18,000	60	2	10
T2	18,000	80	4	10
T3	18,000	100	6	16
T4	24,000	80	6	16
T5	30,000	60	6	15
T6	24,000	60	4	9
T7	24,000	100	2	14
T8	30,000	80	2	10
T9	30,000	100	4	11

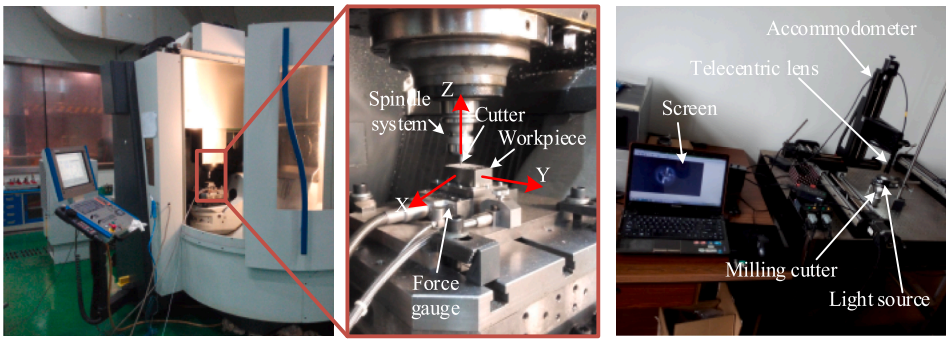


Fig. 5. Experimental device: (a) CNC machine; (b) Milling platform; (c) Tool wear measuring platform.

$$\Delta_A = \frac{1}{\sqrt{5}}\sqrt{\frac{1}{5}\sum_{i=1}^{5}(y_i - \bar{y})^2} = 0.000345mm \tag{26}$$

$$\Delta_B = \frac{\bar{y} \times 0.032\%}{\sqrt{3}} = 0.00001368mm \tag{27}$$

$$\Delta = \sqrt{\Delta_A^2 + \Delta_B^2} = 0.000345271mm \tag{28}$$

$$VB = 0.074034mm \gg 0.000345271mm \tag{29}$$

The observed tool wear values from electron microscopy include three stages of tool wear: initial wear, steady wear, and severe wear, as shown in Fig. 6. The criteria for dividing the tool wear stages are as follows [12]:

$$S_t = \begin{cases} H_1, (\omega'' < 0) \\ H_2, (v_E > v_L, v_E' < v_L') \\ H_3, (v_E < v_L, v_E' < v_L') \end{cases}, \qquad 1 \leqslant t \leqslant T \tag{30}$$

In the above equation, S_t represents the tool wear stage at time t. H_1, H_2, and H_3 correspond to the initial wear, steady wear, and severe wear stages, respectively. v_E and v_L are two intermediate variables constructed to describe the tool wear patterns in the steady wear and accelerated wear stages at time t. v_E' and v_L' are the first derivatives of v_E and v_L. ω' is the second derivative of the tool wear function, reflecting the change in the growth rate of the wear curve. When $\omega' < 0$, it is the initial wear stage (H_1). Both the training set and the test set of the experimental dataset cover the three stages of tool wear mentioned above.

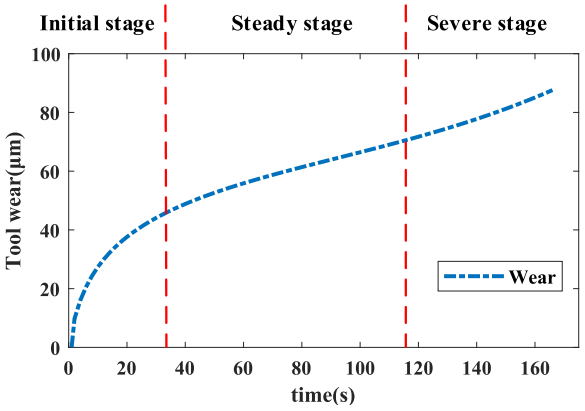


Fig. 6. Schematic diagram of wear stage segmentation.

4.2. Model training

MAML is a method that implements *meta*-learning [43]. Its purpose is to learn the initialization parameters of the model. The learned initialization parameters can be quickly adapted to different sub-tasks, that is, the initialization parameters learned by the MAML algorithm are used for the sub-model, so that the sub-model converges quickly after simple training. Before model training, it is necessary to divide the monitored labeled data into a training set and a test set. The division criteria for the training set (D_{train}) and test set (D_{test}) are to randomly allocate the collected labeled data into training and test sets according to a 2:1 ratio, ensuring that the training set and test set are disjoint and that the model's predictive performance is based on test data that the model has not seen before. For the milling process, the characteristics of tool flank wear are closely related to milling parameters. Therefore, tool wear monitoring under different milling conditions can be regarded as a model of different tasks. Each processing condition represents a type of subtask.

The parameter optimization process of MAML algorithm is shown in Fig. 7. The training samples are divided into different subtask sets $\{T_i\}$ at first, and then each subtask is divided into support set $\{D^{suppot}_{T_i,train}, D^{suppot}_{T_i,test}\}$ and query set $\{D^{query}_{T_i,train}, D^{query}_{T_i,test}\}$. The training process of the model can be divided into two parts: the inner loop and the outer loop. In the inner loop structure, the parameter update process of MAML is as follows:

$$\theta_{T_i} \leftarrow \theta - \alpha \nabla_\theta L(\theta, D^{support}_{T_i,train}) \tag{31}$$

Where $D^{support}_{T_i,train}$ represents the support set of the *i-th* sub-task T_i in the training set, θ represents the initial trainable parameters of the model, α represents the inner loop learning rate, θ_{T_i} represents the optimal parameters obtained on the task, and $\nabla_\theta L(\theta, D^{support}_{T_i,train})$ is the loss gradient. In the outer loop structure, MAML requires rapid adaptation to new tasks, so the outer loop integrates all sub-task training through the support set $D^{query}_{T_i,train}$ of the test data and minimizes its loss to obtain the parameter θ^* of the *meta*-learning training, as follows:

$$\theta^* \leftarrow \theta - \beta \nabla_\theta \sum_{T_i \sim P(T_i)}^{M} L(\theta_{T_i}, D^{query}_{T_i,train}) \tag{32}$$

Where β represents the outer loop learning rate, and θ^* represents the basic parameters of the model that are easy to generalize.

Through the above inner and outer loop training strategy, the model is trained to obtain a general and easy-to-generalization predictor θ^*. The migration prediction of the unknown label data set $D^{support}_{T_i,test}$ is realized by a simple model fine-tuning strategy:

$$\theta^*_{T_i} \leftarrow \theta - \gamma \nabla_\theta L(\theta^*, D^{support}_{T_i,test}) \tag{33}$$

Where γ represents the learning rate of the fine-tuning phase, and $\theta^*_{T_i}$ represents the model parameters that obey the specific sub-task distribution after the fine-tuning.

When optimizing parameters using gradient descent, the convexity of the objective function is a sufficient condition for achieving global optimality. However, the presence of nonlinear mapping relationships can undermine the global convexity of the objective function, resulting in the possibility of suboptimal parameters after optimization. Although the objective function may not satisfy global optimality, incorporating physical constraints into the model construction can ensure local convexity of the objective function within a parameter space that holds practical physical significance.

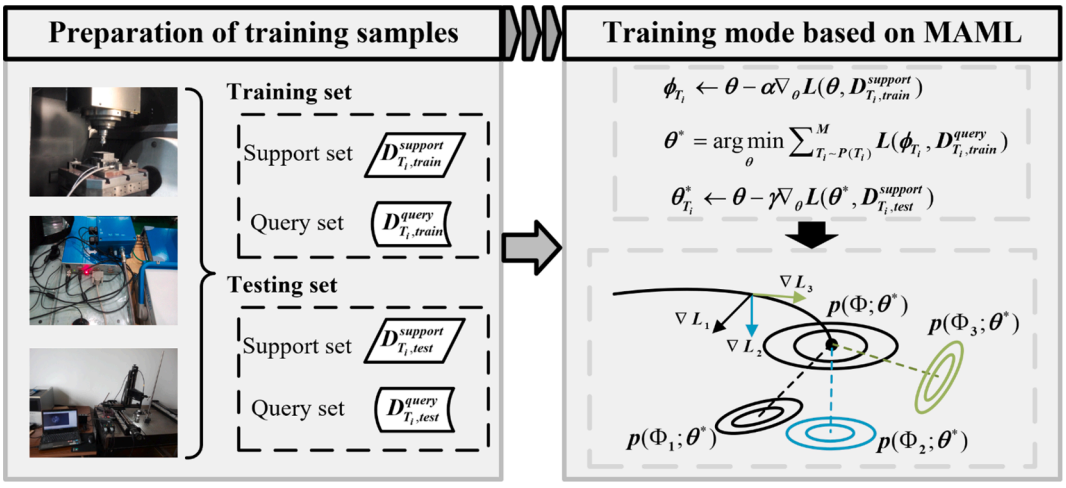


Fig.7. Parameter optimization of the MAML algorithm.

Table 2
MAML algorithm pseudo-code.

Algorithm MAML Algorithm for Tool Wear Monitoring
Input: Distribution over tasks $p(T_i)$, Learning rate hyperparameter α, β;
Output: Network weights θ;
1: Randomly initialize θ;
2: **while** not done **do**
3: Sample batch of tasks: $T_i \sim p(T_i)$;
4: **for** all T_i **do**
5: Evaluate $\nabla_\theta L(\theta, D^{support}_{T_i,train})$;
7: Compute adapted parameters with gradient descent:
$\theta_{T_i} \leftarrow \theta - \alpha \nabla_\theta L(\theta, D^{support}_{T_i,train})$;
8: **end for**
9: Update the meta weight $\theta^* \leftarrow \theta - \beta \nabla_\theta \sum_{T_i \sim P(T_i)}^{M} L(\theta_{T_i}, D^{query}_{T_i,train})$;
10: **end while**

In summary, after introducing the core steps of the MAML algorithm in detail, the pseudo-code of the MAML algorithm is shown in Table 2:

In order to verify the effectiveness of the MAML algorithm, the loss curves of the training set and the validation set of the direct training model are shown in Fig. 8(a), and the loss curves of the training set and the validation set of the training model based on the MAML pre-weight are shown in Fig. 8(b). From the overall trend of the loss value changes in the two graphs, the loss value in Fig. 8(b) converges rapidly, while Fig. 8(a) converges after a period of time, and the loss values of both converge to about 0.5. Comparing the initial loss value, it can be seen that the initial loss value in Fig. 8(b) is significantly smaller than that in Fig. 8(a). This is because the pre-weight for MAML training is the comprehensive optimal weight value of all tasks, which can make the sub tasks converge quickly. For Fig. 8(a), its initial weight is completely random, so its initial loss value is also completely random. Moreover, the training on MAML pre-weight also avoids the model falling into local optimization, making the training of the model easier.

4.3. Results and discussions

In order to verify the effectiveness of the physics-guided deep learning model proposed in this study, the predicted results of the model were compared with those of MLP [44], ResNet [45], LSTM [46], GRU, Physical [12] and Hybrid [29] models, as shown in Fig. 9. To ensure that the comparison models achieve optimal comparative performance, we carefully considered the settings of their hyperparameters. Specifically, for the GRU model, the hyperparameter settings are identical to those of the proposed PGA-GRU model. The only difference between the LSTM and GRU models lies in the type of gate units used, while all other parameters remain the same as those of the GRU model. In the case of the ResNet model, we employed the classic ResNet18 network for the comparative experiments, which has been proven to perform well in practical applications. Similarly, in our comparative experiments, the ResNet model yielded good prediction results. As for the MLP network, we conducted multiple trials to find a set of satisfactory hyperparameters based on the model's prediction performance and training cost. Regarding the physical model, since its hyperparameters are predetermined by the original proposer, we fitted them to the available data. Finally, the hybrid model is constructed based on the combination of the aforementioned physical and data models. Therefore, the parameter settings for the final comparison model are as follows: the structure of the MLP model consists of five full connection layers, and the number of neurons in each layer was set as 8, 16, 32, 16 and 1, respectively. The loss function was MSE. The structure of the ResNet model is ResNet18, and its convolutional kernel size is set to 240×1 and 10×1. The channel is set to 128 and 256, and the loss function is MSE. The structure of the LSTM model consists of three LSTM layers and a full connection layer. The state dimensions of the hidden layer of the LSTM are set to 64, 32, and 16, respectively. The loss function is *logcosh*. The structure of the GRU model consists of three GRU layers and a fully connected layer. The hidden layer unit dimensions of GRU are set to 64, 32, and 16 respectively, and the loss function is *logcosh*. For the above model, the activation function and optimizer are set to *Relu* and *Adam* respectively. For the training process of the model, this study used two-thirds of each dataset from the 9 experimental data sets of T1-T9 for model training, and the remaining one-third for model testing. The predicted results are as follows:

Taking the condition (1) as an example, the prediction effects of different models on tool wear condition are shown in Fig. 9 (a-g). It can be seen that the prediction performance based on MLP and ResNet are the worst, because MLP and ResNet only rely on experimental data to establish the relationship between input and output. The monitoring signal by the sensor is a time series signal, so the current tool wear condition is not only related to the signal monitored by the sensor, but also related to the monitoring signal before the current time. For MLP and ResNet, there are no ability to model signal time series information. In addition, the model itself will have overfitting and underfitting problems, so a large number of data training is required to make the network achieve the ideal prediction effect. LSTM and GRU are currently widely used time series models. They adjust the information flow in the sequence through the gating mechanism to extract the time series information of the signal. The GRU can be seen as a simplified version of the LSTM. Compared with LSTM, GRU only has reset gate and update gate, so it is simpler in structure than LSTM, reducing the training difficulty and overfitting risk of the model. From Fig. 9 (c, d) and Table 3, it can be observed that compared to MLP and ResNet, LSTM and GRU significantly improve the model's prediction accuracy by analyzing the time series information of the signals. Furthermore, due to the fewer training parameters in GRU compared to LSTM, GRU demonstrates better prediction accuracy. This also confirms the reason why

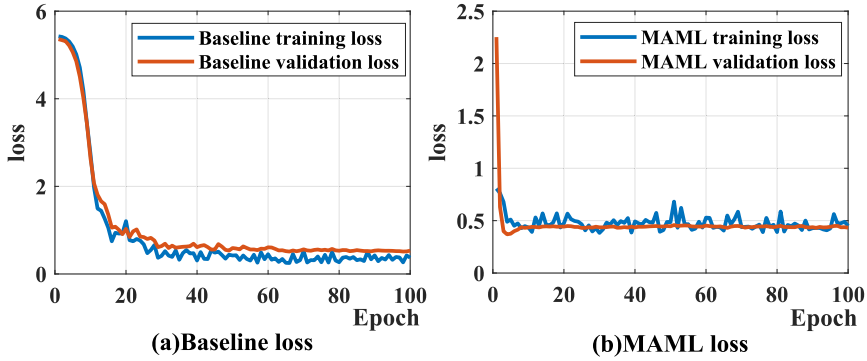


Fig. 8. The influence of MAML algorithm on model training process.

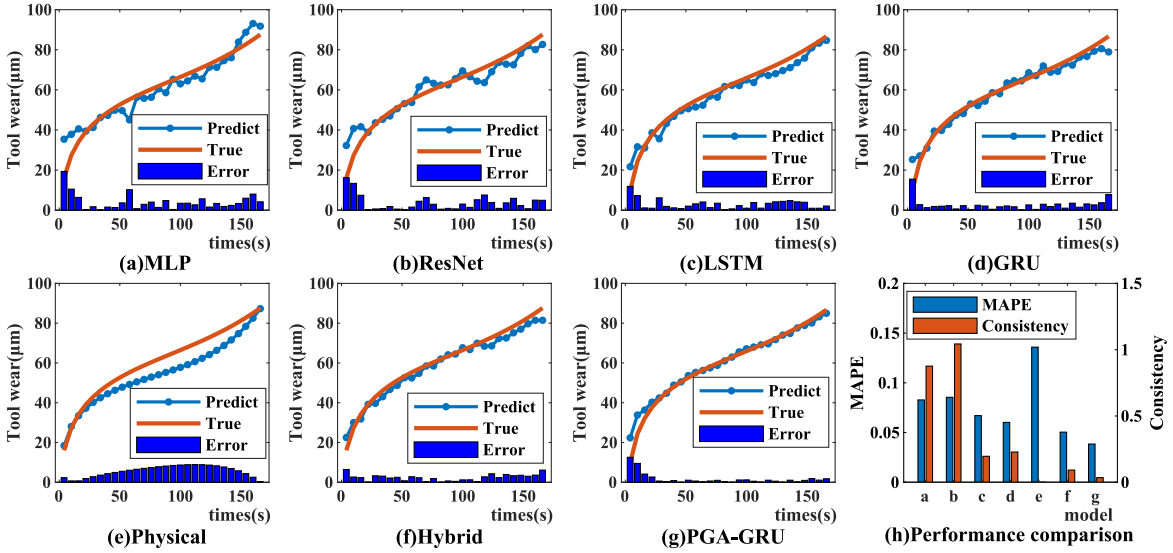


Fig. 9. Comparative analysis of model prediction results.

Table 3
MAE evaluation analysis.

	T1	T2	T 3	T 4	T 5	T 6	T 7	T 8	T 9	Average
MLP	4.118	3.061	2.903	3.798	3.965	5.568	4.143	3.972	2.458	3.776
ResNet	3.677	3.024	3.294	4.432	4.008	4.548	3.774	4.719	3.848	3.925
LSTM	2.974	1.981	2.511	2.934	2.486	3.338	2.674	3.892	2.667	2.829
GRU	2.625	1.963	1.892	2.450	2.377	2.816	3.991	2.404	2.089	2.512
Physical	5.644	7.256	7.220	10.064	7.727	3.372	6.205	5.720	7.417	6.736
Hybrid	2.405	1.527	1.881	2.436	2.085	2.673	3.004	2.195	1.431	2.182
PGA-GRU	**1.646**	**0.882**	**1.550**	**1.239**	**1.094**	**0.955**	**2.631**	**2.005**	**1.770**	**1.530**

we chose GRU in the model structure design. From Fig. 9 (e), it can be observed that the tool wear predicted by the physical model is the most stable among all the models. This is because the physical model is based on empirical knowledge of tool wear mechanisms, which provides strong physical constraints for prediction. However, due to the strict adherence to predetermined physical rules in its predictions, the physical model lacks real-time feedback for handling states. The prediction errors of the physical model cannot be effectively corrected like those of the data model, resulting in poor generalization ability. As for the prediction results of the hybrid model shown in Fig. 9 (f), it combines the results of the physical model and the data model using regularized particle filtering techniques. It treats the prediction results of the data model as observations, while the physical model calculates the transitions between states. This fusion approach improves the accuracy and physical consistency of the prediction results compared to the pure data model. However, due to the significant error accumulation in the first-order Taylor series of the physical model and the large physical inconsistency of the data model without physical constraints, the improvement on the results after particle filtering is not very

significant compared to the method proposed in this study. For Fig. 9(g), the predicted results of the physical guided GRU Model (PGA-GRU) are presented. It can be seen that the PGA-GRU model has the best prediction results. This is because on the basis of the GRU model, considering the progressive degradation characteristics of tool wear, a monotonic mechanism is introduced, so that the tool wear condition at the current moment is greater than the tool wear condition at the previous moment. At the same time, the relationship between adjacent wear values is established by using the physical model of tool wear, so as to introduce the physical constraint loss term to constrain the optimization process of the model. In Fig. 9 (h), a-g correspond to five different prediction models (MLP, ResNet, LSTM, GRU, Physical, Hybrid, PGA-GRU) in sequence. It can be seen that the PGA-GRU model has significantly better prediction accuracy and physical consistency than other models.

The prediction performance of the model under other processing conditions is indirectly expressed by the evaluation indicators shown in Tables 3–5. Table 3 measures the absolute deviation between the real value and the predicted value through the MAE evaluation index. Table 4 measures the relative deviation between the real value and the predicted value through the MAPE evaluation index. Compared with MAE, MAPE uses percentage to measure the size of the error, and does not need to combine the dimension of the real value to give a judgment of the difference. Table 5 measures the physical inconsistency of the model prediction results through the consistency evaluation index. It can be seen from the prediction results of Tables 3–5 that the method proposed in this paper is almost always better than the prediction results of the comparison model under different working conditions.

4.4. The model validation with benchmark data

In order to objectively evaluate the performance of the model proposed in this study, this study evaluates the performance of the model through the "phm2010" public data set [47]. The milling parameters are set as follows: the spindle speed is set to 10360 rpm; The feed rate is set to 1.555 m/min; The milling width is set to 0.125 mm; The milling depth is set to 0.25 mm; The milling tools are three kinds of ball head cemented carbide tools; The workpiece is Inconel 718. The signals in the milling process are collected by multiple sensors, including dynamometers and accelerometers in three different direction channels, and acoustic emission sensors. The sampling frequency of each channel is 50KHz. After cutting one path each time, use LEICA MZ12 microscope to measure the tool wear. For the three groups of experimental data measured, C1 and C4 are used as training sets to train the model, and C6 is used as a test set to verify the effectiveness of the model. For 7-dimensional machining signals, this study uses force sensor signals from different directions to predict tool wear conditions.

In the benchmark dataset, the training process of model parameters is consistent with the training process described in Table 2, which can be mainly divided into: initialization of parameters, fine-tuning of parameters, and evaluation of the model. Firstly, the three sets of experiments under different operating conditions (C1, C4, C6) corresponding to the benchmark dataset are divided into a support set and a query set. The parameters of the model are updated by calculating gradients and accumulating them on the subtasks consisting of C1 and C4 datasets, thereby obtaining the initialization parameters of the model. Secondly, the fine-tuning process of parameters is to update the initial parameters through the support set of the C6 subtask, so that the model parameters applicable to the task can be quickly obtained.

The prediction results of the model proposed in this study on the "phm2010" public dataset are shown in Fig. 10 and Table 6. From the comparison of the prediction results of the seven models in Fig. 10, it can be seen that the prediction effect of the model proposed in this study is significantly better than that of MLP, ResNet, LSTM, GRU, Physical and Hybrid models. In Fig. 10 (h), a-g correspond to five different prediction models (MLP, ResNet, LSTM, GRU, Physical, Hybrid, PGA-GRU) in sequence. It can be observed that the MLP and ResNet models have the lowest prediction accuracy. Although the prediction accuracy of LSTM and GRU models is greatly improved compared with MLP and ResNet models, their physical consistency is also poor. For the physical model, its predictions are the most stable because it is based on the physical rules of tool wear. However, due to the strong constraints imposed by these rules, the prediction process of the model lacks adjustments for the machining process. As for the hybrid model, it relies on particle filtering techniques to improve the prediction results compared to the data model. However, due to the accumulation of errors during the state transition process and the lack of physical constraints on the observed values, the improvement is limited. The PGA-GRU model proposed in this paper not only maintains the prediction accuracy, but also eliminates the physical inconsistency in the prediction process. The significant deviation in the prediction of the above model in the later stage is due to the learning characteristics of deep learning. The wear rate of tools in the initial and severe stages is higher than that in the stable model stage, resulting in uneven distribution of samples, with most samples coming from the stable stage. Compared to the stable stage, there will be significant deviations in the severe stage. The prediction effect of the above model on the benchmark data set is consistent with that of the milling data set in the previous section, so it can be explained that the physical guidance data model established in this paper has good generalization ability under different data sets.

5. Conclusion

Aiming at the monitoring of tool wear condition in machining process, this study proposes a physics-guided deep learning model, which innovatively solves the modeling difficulties caused by scale differences between signals and the difficulties in tool condition monitoring caused by insufficient prior knowledge of data-driven models. The main conclusions are as follows:

1) A dual scale time series model is proposed in this study, which incorporates monotonicity constraints based on the progressive degradation characteristics of tool wear to reduce the physical inconsistency of model predictions. This leads to an accurate mapping between monitoring signals and tool wear, providing an effective way for on-line monitoring of tool wear.

Table 4
MAPE evaluation analysis.

	T 1	T 2	T 3	T 4	T 5	T 6	T 7	T 8	T 9	Average
MLP	0.080	0.081	0.071	0.078	0.087	0.118	0.081	0.088	0.061	0.083
ResNet	0.071	0.073	0.073	0.090	0.090	0.102	0.077	0.102	0.090	0.085
LSTM	0.068	0.052	0.065	0.079	0.055	0.072	0.061	0.086	0.065	0.067
GRU	0.057	0.054	0.042	0.067	0.054	0.070	0.084	0.057	0.056	0.060
Physical	0.103	0.149	0.174	0.197	0.144	0.071	0.115	0.115	0.154	0.136
Hybrid	0.048	0.044	**0.039**	0.057	0.046	0.063	0.063	0.053	0.042	0.050
PGA-GRU	**0.046**	**0.036**	0.044	**0.038**	**0.030**	**0.017**	**0.057**	**0.040**	**0.039**	**0.039**

Table 5
Consistency evaluation analysis.

	T 1	T 2	T 3	T 4	T 5	T 6	T 7	T 8	T 9	Average
MLP	0.495	0.333	0.386	1.027	0.670	0.887	1.354	1.937	0.795	0.876
ResNet	0.532	0.423	1.378	1.448	0.607	1.414	0.931	1.554	1.101	1.043
LSTM	0.234	0.069	0.330	0.033	0.062	0.094	0.330	0.422	0.191	0.196
GRU	0.309	0.120	0.410	0.018	0.090	0.058	0.396	0.365	0.289	0.228
Physical	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
Hybrid	0.092	0.043	0.107	0.000	0.027	0.045	0.226	0.105	0.183	0.092
PGA-GRU	**0.000**	0.003	0.089	**0.000**	**0.000**	0.019	0.064	0.063	0.077	0.035

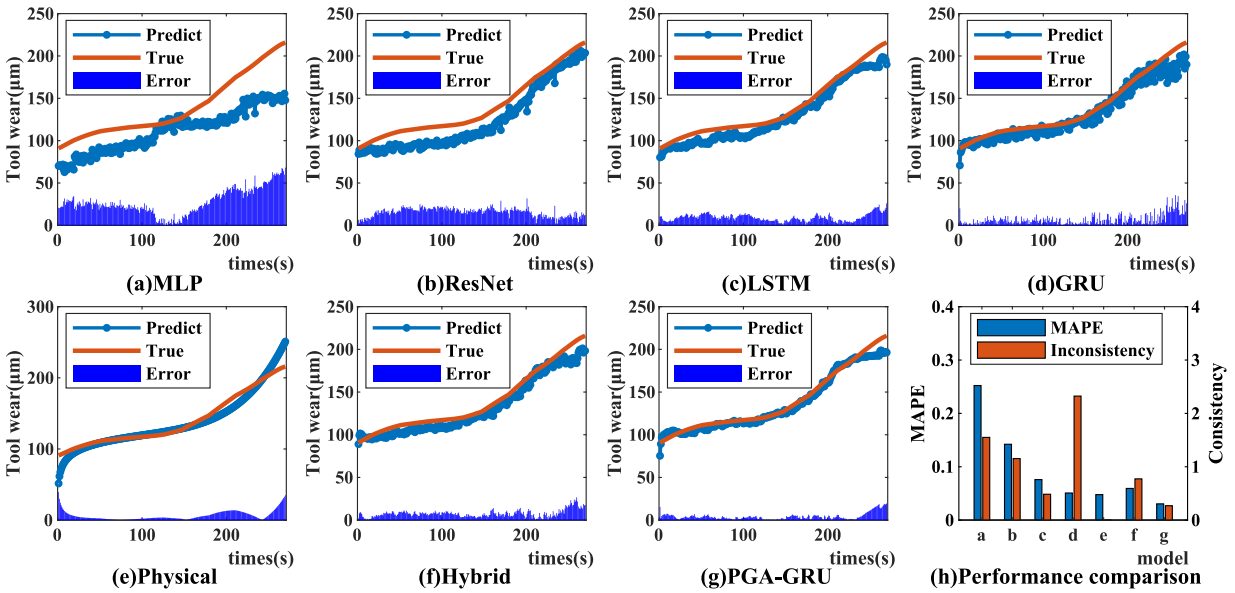


Fig. 10. Comparison of prediction results of the model on benchmark data sets.

Table 6
The performance of various models.

	MAPE	MAE	Inconsistency
MLP	0.252	27.981	1.550
ResNet	0.142	15.719	1.154
LSTM	0.076	9.306	0.485
GRU	0.051	6.931	2.324
Physical	0.0478	6.3627	0
Hybrid	0.0594	7.7488	0.773
PGA-GRU	**0.030**	**4.224**	**0.270**

2) Based on the physical model of tool wear, the physical relationship between adjacent wear values is established, and the physical constraint loss term is introduced. By using physical model labels and real labels to jointly train the model to fully mine the hidden relationship between model input and output.
3) For the problem of tool wear monitoring under various machining conditions, this study uses the MAML algorithm to train the pre-weight of the model to improve the generalization ability of the model under different machining conditions, so that the model is suitable for different machining conditions.

Based on the complexity of machining mechanisms, the guidance of a single wear mechanism on data models is limited. In future research, data-driven methods will be combined with more machining mechanisms, such as machining dynamics models, to improve the generalization of the model under different machining types and conditions.

CRediT authorship contribution statement

Shenshen Li: Writing – original draft, Methodology, Conceptualization. **Jun Li:** Writing – review & editing, Visualization, Validation. **Kunpeng Zhu:** Supervision, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (Grant No.: 52175528), in part by the National Key Research and Development Program of China, the Chinese Ministry of Science and Technology (Grant No.: 2018YFB1703200).

References

[1] S. Zhang, Y. Zhou, H. Zhang, Z. Xiong, S. To, Advances in ultra-precision machining of micro-structured functional surfaces and their typical applications, Int. J. Mach. Tools Manuf. 142 (2019) 16–41.
[2] I. Ullah, S. Zhang, Q. Zhang, R.-Q. Wang, Numerical investigation on serrated chip formation during high-speed milling of Ti-6Al-4V alloy, J. Manuf. Processes 71 (2021) 589–603.
[3] P. Ding, X. Huang, S. Li, C. Zhao, X. Zhang, Real-time reliability analysis of micro-milling processes considering the effects of tool wear, Mech. Syst. Sig. Process. 200 (2023) 110582.
[4] F. Camci, R.B. Chinnam, Health-state estimation and prognostics in machining processes, IEEE T. Autom. Sci. Eng. 7 (3) (July 2010) 581–597.
[5] K. Zhu, Smart machining system: modelling, monitoring and informatics, Springer, 2022.
[6] A.G. Rehorn, J. Jiang, P. Orban, State-of-the-art methods and results in tool condition monitoring: a review, Int. J. Adv. Manuf. Technol. 26 (2005) 693–710.
[7] K.-M. Lee, Y. Huang, J. Ji, C.-Y. Lin, An online tool temperature monitoring method based on physics-guided infrared image features and artificial neural network for dry cutting, IEEE t. Autom. Sci. Eng. 15 (4) (Oct. 2018) 1665–1676.
[8] W. Li, C. Li, N. Wang, J. Li, J. Zhang, Energy saving design optimization of cnc machine tool feed system: a data-model hybrid driven approach, IEEE t. Autom. Sci. Eng. 19 (4) (Oct. 2022) 3809–3820.
[9] H. Luo, D. Zhang, M. Luo, Tool wear and remaining useful life estimation of difficult-to-machine aerospace alloys: a review, China Mechanical Engineering 32 (22) (2021) 2647.
[10] X. Zhang, Y. Gao, Z. Guo, W. Zhang, J. Yin, W. Zhao, Physical model-based tool wear and breakage monitoring in milling process, Mech. Syst. Sig. Process. 184 (2023) 109641.
[11] Y. Bai, F. Wang, R. Fu, J. Hao, L. Si, B. Zhang, W. Liu, J.P. Davim, A semi-analytical model for predicting tool wear progression in drilling CFRP, Wear 486 (2021) 204119.
[12] K. Zhu, Y. Zhang, A generic tool wear model and its application to force modeling and wear monitoring in high speed milling, Mech. Syst. Sig. Process. 115 (2019) 147–161.
[13] S. Yang, G. Zhu, J. Xu, Y. Fu, Tool wear prediction of machining hydrogenated titanium alloy Ti6Al4V with uncoated carbide tools, Int. J. Adv. Manuf. Technol. 68 (2013) 673–682.
[14] R. Das, S.S. Joshi, H.C. Barshilia, Analytical model of progression of flank wear land width in drilling, J. Tribol. 141 (1) (2019) 011601.
[15] L. Seeholzer, T. Krammer, P. Saeedi, K. Wegener, Analytical model for predicting tool wear in orthogonal machining of unidirectional carbon fibre reinforced polymer (CFRP), Int. J. Adv. Manuf. Technol. 119 (11–12) (2022) 7259–7289.
[16] H.V. Dang, H. Tran-Ngoc, T.V. Nguyen, T. Bui-Tien, G. De Roeck, H.X. Nguyen, Data-driven structural health monitoring using feature fusion and hybrid deep learning, IEEE T. Autom. Sci. Eng. 18 (4) (Oct. 2021) 2087–2103.
[17] Y. Zhou, G. Zhi, W. Chen, Q. Qian, D. He, B. Sun, W. Sun, A new tool wear condition monitoring method based on deep learning under small samples, Measurement 189 (2022) 110622.
[18] H. Guo, Y. Zhang, K. Zhu, Interpretable deep learning approach for tool wear monitoring in high-speed milling, Comput. Ind. 138 (2022) 103638.
[19] M. Marei, S. El Zaatari, W. Li, Transfer learning enabled convolutional neural networks for estimating health state of cutting tools, Rob. Comput. Integr. Manuf. 71 (2021) 102145.
[20] R. Gao, L. Wang, R. Teti, D. Dornfeld, S. Kumara, M. Mori, M. Helu, Cloud-enabled prognosis for manufacturing, CIRP Ann. 64 (2) (2015) 749–772.
[21] M. Wang, J. Wang, CHMM for tool condition monitoring and remaining useful life prediction, J. Adv. Manuf. Technol. 59 (2012) 463–471.

[22] X. Li, W. Zhang, Q. Ding, Understanding and improving deep learning-based rolling bearing fault diagnosis with attention mechanism, Signal Process. 161 (2019) 136–154.
[23] S. Siahpour, X. Li, J. Lee, A novel transfer learning approach in remaining useful life prediction for incomplete dataset, IEEE Trans. Instrum. Meas. 71 (2022) 1–11.
[24] T. Xia, Y. Dong, L. Xiao, S. Du, E. Pan, L. Xi, Recent advances in prognostics and health management for advanced manufacturing paradigms, Reliab. Eng. Syst. Saf. 178 (2018) 255–268.
[25] H. Li, Z. Zhang, T. Li, X. Si, A review on physics-informed data-driven remaining useful life prediction: challenges and opportunities, Mech. Syst. Sig. Process. 209 (2024) 111120.
[26] G.E. Karniadakis, I.G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, L. Yang, Physics-informed machine learning, Nat. Rev. Phys. 3 (6) (May 2021) 422–440.
[27] W. Huang, X. Zhang, C. Wu, S. Cao, Q. Zhou, Tool wear prediction in ultrasonic vibration-assisted drilling of CFRP: a hybrid data-driven physics model-based framework, Tribol. Int. 174 (2022) 107755.
[28] H. Hanachi, W. Yu, I.Y. Kim, J. Liu, C.K. Mechefske, Hybrid data-driven physics-based model fusion framework for tool wear prediction, Int. J. Adv. Manuf. Technol. 101 (2019) 2861–2872.
[29] Z. Ma, M. Zhao, X. Dai, Y. Chen, A hybrid-driven probabilistic state space model for tool wear monitoring, Mech. Syst. Sig. Process. 200 (2023) 110599.
[30] X. Jia, J. Willard, A. Karpatne, J.S. Read, J.A. Zwart, M. Steinbach, V. Kumar, Physics-guided machine learning for scientific discovery: An application in simulating lake temperature profiles, ACM/IMS Trans. Data Sci. 2 (3) (2021) 1–26.
[31] Q. Yang, K.R. Pattipati, U. Awasthi, G.M. Bollas, Hybrid data-driven and model-informed online tool wear detection in milling machines, J. Manuf. Syst. 63 (2022) 329–343.
[32] G. Wang, F. Zhang, A sequence-to-sequence model with attention and monotonicity loss for tool wear monitoring and prediction, IEEE Trans. Instrum. Meas. 70 (2021) 1–11.
[33] J. Wang, Y. Li, R. Zhao, R.X. Gao, Physics guided neural network for machining tool wear prediction, J. Manuf. Syst. 57 (2020) 298–310.
[34] Y. Li, J. Wang, Z. Huang, R.X. Gao, Physics-informed meta learning for machining tool wear prediction, J. Manuf. Syst. 62 (2022) 17–27.
[35] T. Zhou, S. Jiang, T. Han, S. Zhu, Y. Cai, A physically consistent framework for fatigue life prediction using probabilistic physics-informed neural network, Int. J. Fatigue 166 (2023) 107234.
[36] S. Zeng, D. Pi, Milling surface roughness prediction based on physics-informed machine learning, Sensors 23 (10) (2023) 4969.
[37] D.H. Tien, T.D. Quy, V.T. Nguyen, N.T. Nguyen, D.T. Do, D. Nguyen, Online monitoring and multi-objective optimisation of technological parameters in high-speed milling process, Int. J. Adv. Manuf. Technol. 112 (2021) 2461–2483.
[38] Y. Li, G.M. Zheng, X. Zhang, X. Cheng, X.H. Yang, R.F. Xu, Cutting force, tool wear and surface roughness in high-speed milling of high-strength steel with coated tools, J. Mech. Sci. Technol. 33 (2019) 5393–5398.
[39] K. Zhu, T. Mei, D. Ye, Online condition monitoring in micromilling: a force waveform shape analysis approach, IEEE Trans. Ind. Electron. 62 (6) (2015) 3806–3813.
[40] T. Mohanraj, S. Shankar, R. Rajasekar, N.R. Sakthivel, A. Pramanik, Tool condition monitoring techniques in milling process—a review, J. Mater. Res. 9 (2020) 1032–1042.
[41] I.N. Tansel, T.T. Arkan, W.Y. Bao, N. Mahendrakar, B. Shisler, D. Smith, M. McCool, Tool wear estimation in micro-machining.: Part I: tool usage–cutting force relationship, Int. J. Mach. Tools Manuf. 40 (2000) 599–608.
[42] J. Chen, H. Jing, Y. Chang, and Q, Liu, Gated recurrent unit based recurrent neural network for remaining useful life prediction of nonlinear deterioration process, Reliab. Eng. Syst. Saf. 185 (2019) 372–382.
[43] C. Finn, P. Abbeel, S. Levine, Model-agnostic meta-learning for fast adaptation of deep networks, Sydney, AU, ICML, 2017, pp. 1126–1135.
[44] A. Pinkus, Approximation theory of the MLP model in neural networks, Acta Numer. 8 (1999) 143–195.
[45] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, Las Vegas, Nevada, USA, CVPR, 2016, pp. 770–778.
[46] Y. Yu, X. Si, C. Hu, J. Zhang, A review of recurrent neural networks: LSTM cells and network architectures, Neural Comput. 31 (7) (2019) 1235–1270.
[47] X. Li, B. Lim, J. Zhou, S. Huang, S. Phua, K. Shaw, M. Er, Fuzzy neural network modelling for tool wear estimation in dry milling operation, PHM Society, San Diego, CA, USA, Annual Conf, 2009.