International Journal of Network Dynamics and Intelligence

Scilight

Article

Multi-Task Network based Health Status Assessment of Cutting Tools

Xiaoyu Yin [1], Xiaohua Li [2], Yong Zhang [1], Wei Zhou [1,*], and Zhenxing Liu [1]

[1] School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan 430081, China
[2] Chengdu Aircraft Industrial (Group) Co., Ltd, Chengdu 610073, China
[*] Correspondence: chow_ww@163.com

Received: 5 March 2024
Accepted: 7 March 2025
Published: 27 June 2025

Abstract: In modern industrial machining processes, cutting tools play an important role, and their health directly affects the quality of machined parts. However, changing cutting tools based on experience not only increases costs but also reduces productivity. Therefore, predicting the future health of the cutting tools in advance can enable cutting tools changes to be carried out at the right time. The existing health status assessment of a cutting tools consists of three main areas: cutting tools wear prediction, health stages division, and reliability assessment. But traditional deep learning models usually process these three tasks separately, ignoring the correlation that exists between the three tasks. In order to solve this problem, this paper proposes a multi-task model based on temporal convolutional network (TCN) and progressive layered extraction (PLE) network. Firstly, the pre-processed data are subjected to feature extraction and feature selection and the redundant information between the features is reduced by using an autoencoder. Secondly, the TCN network module is used to extract the correlation feature information of multiple tasks in time, the PLE network module learns the difference information between each task, and finally, the prediction output is made by each sub-task module for each sub-task. Finally, the dynamic weight average method is used to adjust the weights of the loss function, which avoids manual adjustment of the weights. The experimental results show that the method not only achieves the prediction of multiple tasks but also has high prediction accuracy.

Keywords: multi-task learning; cutting tools wear prediction; health stages division; reliability assessment

1. Introduction

Cutting tools play an important role in the production of parts for the automotive, mold, and aerospace industries. Inevitably, tool wear occurs during the process of cutting parts [1]. When the degree of wear of the tool is more serious, it will lead to the production of poor quality parts, seriously affecting the machining accuracy and efficiency [2]. Replacing cutting tools early in the machining process increases costs and replacing them only after damage occurs can result in damage to the workpiece [3]. According to statistics, tool wear causes approximately 20% of total equipment downtime in actual production processes [4]. By monitoring the health of tools during production, downtime can be reduced by approximately 75% [5]. Therefore, changing tools before they reach a certain level of wear can increase equipment productivity. However, in order to reduce the probability of failures during the production process, tools are usually replaced early before damage occurs, often using only 50-80% of their expected life [6]. This not only increases production costs, but also leads to reduced productivity. Therefore, it is very necessary to conduct tool health status assessment, which can provide a scientific and accurate basis for tool replacement. By evaluating the health status of the tool, the tool can be fully utilised, saving the downtime of the machining process and the cost of tool replacement.

The health status assessment of cutting tools includes prediction of cutting tools wear (TW), division of health stages (HS), and assessment of reliability (R). Where the TW reflects the cutting quality of the cutting tools, the HS reflects the current state of health and wear of the cutting tools, and the R of the cutting tools affects the efficiency and stability of the machining process. Existing health status assessment methods are mainly classified into model-based and data-driven approaches [7, 8]. Model-based methods involve modeling the entire cutting tools working system, which often requires a lot of expert experience and professional knowledge. In addition, real work scenarios are complex, and it is difficult to build a model that fully reflects the actual cutting tools working conditions [9]. Compared with model-based methods, data-driven methods only require data collection and do not require much expert experience and professional knowledge [10], making them more suitable for complex scenarios. Therefore, data-driven methods are increasingly important.

Liu et al. [11] proposes a meta-invariant feature space learning method that constructs an invariant feature space for the pairing task to approximate the marginal distribution, whose natural laws under crossover conditions are learnt through meta-learning, and which can be adapted to and accurately predict tool wear under crossover conditions with a small number of new samples. Zhou et al. [12] proposed a milling tool condition monitoring (TCM) method based on two-tier angular kernel limit learning machine (TAKELM) and binary difference evolution (BDE). Among them, TAKELM enhances intrinsic feature extraction from microarray data without the need to manually preset the kernel function or optimize its hyperparameters. The BDE algorithm is applied to search for the optimal combination of feature parameters among alternative feature parameters in multiple domains in order to achieve the minimum number of feature parameters that satisfy very small prediction errors. Qin et al. [13] proposed a tool wear identification and prediction method based on stack sparse self-coding network, which can simplify the process of monitoring model establishment, accurately monitor the tool wear state and wear amount, and has certain reference value for efficient tool replacement in the actual metal cutting process. Wei et al. [14] proposed to perform spectral analysis of milling force to obtain signal frequencies that reflect the degree of tool wear. Then, a force signal decomposition model based on whitened variational mode decomposition was developed for filtering out the sensitive signals and effectively avoiding mode mixing. Next, the joint information entropy is used to select the force signal features. Compared with other dimensionality reduction algorithms, the optimal subset of features obtained by the JIE method can reflect the strong correlation between features and wear states. Finally, the identification of tool wear states is achieved by the optimal path forest algorithm. Liu et al. [15] proposed a novel TWM model that firstly uses a parallel residual network to adaptively extract multi-scale local features of sensor signals, and then uses a stacked bi-directional long and short-term memory network to acquire time-series features related to tool wear characteristics. Finally, the predicted tool wear values are outputted through a fully connected network and a smoothing correction method is applied to improve the prediction accuracy.

HS division reflects the short-term state of the cutting tools, while TW prediction represents the state of the entire life cycle of the cutting tools, and R assessment embodies the stability of the entire life cycle of the cutting tools. These three tasks reflect different aspects of the current health information of cutting tools from different perspectives. However, the existing research ignores the intrinsic connection between the three, which helps to improve the prediction accuracy of the three tasks. Therefore, this paper establishes a multi-task deep learning model to achieve simultaneous prediction of tool HS division, TW prediction, and R evaluation. It provides a more comprehensive and scientific basis for tool change, which is conducive to helping actual industrial production save time and economic costs.

Multi-task learning (MTL) has been applied to various fields such as natural language processing, speech recognition, computer vision [16], and health management. Wang et al. [17] proposed a multi-task learning model based on DBN for simultaneously learning the wear state of cutting tools and the surface quality of parts. Miao et al. [18] constructed a dual-task network based on LSTM to predict both HS and RUL simultaneously. Kim et al. [19] used a multi-task learning network based on CNN to complete the joint prediction of HS and RUL. Although these models have achieved relatively good results, they all ignore one issue, that is, HS is a classification task while RUL is a regression task. The label of the two tasks is continuous and discrete values with different units. Additionally, traditional multi-task network loss functions usually use the same weights or manual adjustment methods. Improving model performance through manual weight adjustment usually requires multiple attempts, which consumes a lot of time.

To achieve simultaneous prediction of TW, HS division, and R assessment without manual weight adjustment, this paper proposes a multi-task network model based on the progressive layered extraction (PLE) and temporal convolutional network (TCN). As a typical multi-task learning model, the PLE network separates task-specific and shared parameters by adopting a progressive layering approach through the perspective of joint representation and information routing, and successfully realizes the mutual integration and transfer of knowledge between multiple tasks, thus reducing the problem of poor prediction results due to the weak correlation between multiple tasks. However, the health status assessment of a tool is essentially a time-series problem, and PLE was originally designed to be used in recommender systems. TCN by combining the null causal convolution and residual cross-layer connectivity approach, which allows TCN to obtain a larger sensory field with fewer number of networks layers, and is more conducive to processing time-series data with long-term historical dependencies. At the same time, the cross-layer connection approach can inhibit the emergence of gradient vanishing or explosion problems in the process of deepening the network model and can better learn the features in tool data. Therefore, to cope with the demand for information mining of tool data, which is time-series data, the TCN network is added before the PLE network to be able to extract the time-series information in the tool data, to better achieve accurate prediction of the three tasks in the tool health status assessment. In addition, the dynamic weight average method is used to adapt the weights of multi-task loss during model training. The main contributions of this paper can be summarized as follows:

- Multi-task learning is introduced into the field of cutting tools health status assessment for multifaceted cutting tools health monitoring.
- The PLE structure behind the shared TCN layer is adopted to improve the generalization ability of the multi-task network model to address the problem of multiple tasks with different label scales resulting in less correlation.
- A loss function based on dynamic weight average is utilized to automatically learn the weight values between multiple losses, which avoids the difficulties process of manual weight tuning in multi-task scenarios.

The organization structure of this paper is as follows: Section 2 introduces the proposed multi-task network framework. Section 3 covers data processing, feature extraction, and label construction. Section 4 presents the experimental results and performance comparison. Section 5 summarizes the paper and provides prospects.

2. Methodology

This paper aims to achieve three main objectives: TW prediction, HS division, and R assessment of the cutting tools. First, the collected tool data are preprocessed, including steps to remove noisy data and abnormal data. Next, we construct reliability labels and health stages labels using logistic regression and Gaussian mixture models. Then, the training set data is input into the TCN-PLE model for training in order to optimize the whole model. Finally, the trained model is used to predict the test set. In this way, we can obtain accurate prediction results for the state of the tools in the test set. The overall framework of this paper is shown in Figure 1. This section will mainly introduce TCN and PLE and elaborate on the entire multi-task network based on these basic modules.

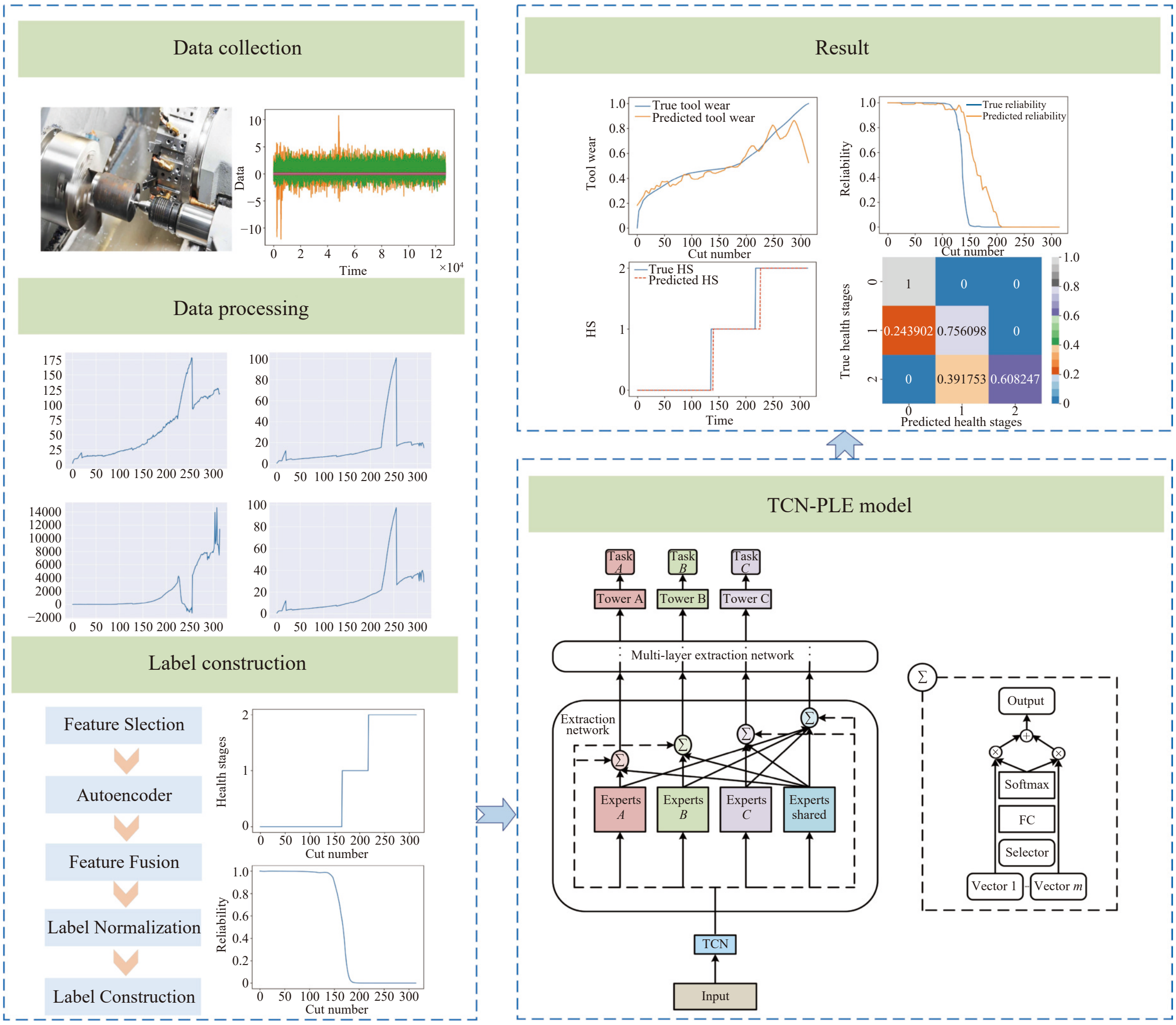


Figure 1. The overall flowchart of the proposed methodology.

2.1. Temporal Convolutional Network

TCN [20] is a time-series prediction model based on causal convolution and dilated convolution, with the addition of a residual module to solve gradient vanishing problems. The main idea of TCN is to share weights locally, which helps capture long-term dependencies in the input sequence.

1) *Causal Convolution*: Compared to ordinary convolution, causal convolution only moves the convolution kernel at the front or back of the sequence, thus ensuring that the output results are consistent with the temporal relationship of the input sequence. For one-dimensional time series input $X = (x_0, x_1, \cdots, x_t, \cdots, x_T)$, the network's output y_t at time t only depends on the current time x_t and some past inputs (i.e., $x_{t-1}, x_{t-2}, x_{t-3}$) but not on any future inputs (i.e., $x_{t+1}, x_{t+2}, x_{t+3}, \cdots, x_T$). This means that the network's output is only affected by past input information, thereby avoiding the problem of information leakage.

2) *Dilated Convolution*: The dilated convolution, compared to ordinary convolution, introduces the parameter d to represent the size of dilation. With the introduction of the dilation rate, the receptive field of the convolutional kernel is increased. For a time series input $X = (x_0, x_1, \cdots, x_t, \cdots, x_T)$, when the convolutional kernel is $K = (k_1, k_2, \cdots, k_M)$, the dilation convolution operation $D(\cdot)$ is:

$$D(T) = \sum_{i=1}^{M} K(i) \cdot x_{T-d \cdot i} \tag{1}$$

where M is the size of the convolutional kernel, d is the dilation rate, $K(i)$ indicates the filtering operation. and $x_{T-d \cdot i}$ is the past data. By increasing the values of M and d, TCN can accept a wider range of input information. Additionally, due to its parallel processing approach, the model has high computational efficiency.

3) *Residual Block:* In addition to using the dilation convolution approach, TCN can also increase its receptive field by increasing the number of hidden layers. However, as the number of hidden layers increases, the model may suffer from gradient vanishing. To address this issue, TCN employs residual blocks. If the fitting function of TCN is $F(x)$, then for an input x, the final output can be described as:

$$o = \text{Activation}(F(x) + x) \tag{2}$$

2.2. Progressive Layered Extraction

PLE [21] is a classic multi-task network. Compared to other multi-task networks that share a single expert network among multiple tasks, it proposes the method of giving each task a specific expert network and cascading multiple layers of expert networks, which reduces the impact of multi-task learning due to weak or complex correlation between multiple tasks. PLE has good performance in many multi-task learning scenarios. In PLE, a customized gate control model is proposed to separate shared network layers from specific network layers. Customized gate control is equivalent to the single-layer version of PLE.

The customized gate control mainly consists of a bottom expert module and a top tower network, where each expert module is composed of multiple expert sub-networks. The expert modules are divided into shared networks and specific networks, where the shared network module learns features that are shared by multiple tasks, while the specific network learns features corresponding to a single task. Finally, the shared and specific task experts are selectively fused through the gate network. For task k, the output of the gate network can be expressed as:

$$g^k(x) = w^k(x) S^k(x) \tag{3}$$

where x denotes the input. $w^k(x)$ is the weighting function used to compute the weight vector of task k after linear transformation and SoftMax layer:

$$w^k(x) = Softmax\left(W_g^k x\right) \tag{4}$$

where $W_g^k \in R^{(c_k+c_s)\times d}$ is the parameter matrix after linear transformation of the input x, c_s, and c_k are the number of shared and task K's specific networks, respectively, and d is the input dimension. $S^k(x)$ is the matrix composed of the shared network and the task k's specific network. The $S^k(x)$ can be represented as:

$$S^k(x) = \left[E_{(k,1)}^T, E_{(k,2)}^T, \cdots, E_{(k,m_k)}^T, E_{(s,1)}^T, E_{(s,2)}^T, \cdots, E_{(s,m_s)}^T\right]^T \tag{5}$$

where $E_{(l,i)}^T$ is selected vector.

2.3. Multi-task Network Structure

The proposed multi-task networks based on TCN and PLE can achieve cutting tools HS division, R assessment, and TW prediction. The network structure is shown in Figure 2. It consists mainly of four parts: the input layer, the shared bottom of TCN, the PLE layer, and the output layer. TCN module is primarily used for extracting related features across multiple tasks, while the PLE module is employed to learn from different tasks and simultaneously output results from multiple tasks. The outputs of the three tasks are represented as follows:

$$y_{r,r,c} = t_{r,r,c}\left(f_{r,r,c}(x)\right) \tag{6}$$

$$f_{r,r,c}(x) = \sum_{i=1}^{n} g_{i;r,r,c}(h(x)) \tag{7}$$

$$g_{i;r,r,c} = W_g\left(g_{i-1;r,r,c}(x)\right) S(x) \tag{8}$$

where $y_{r,r,c}$ is the prediction results for R, TW, and HS. $t_{r,r,c}$ is the tower network, $f_{r,r,c}(x)$ is gating network output, $h(x)$ is the output of TCN, $g_{i;r,r,c}$ is gating network, W_g is the weighting function when the input is $g_{i-1;r,r,c}(x)$, $S(x)$ is the selected matrix.

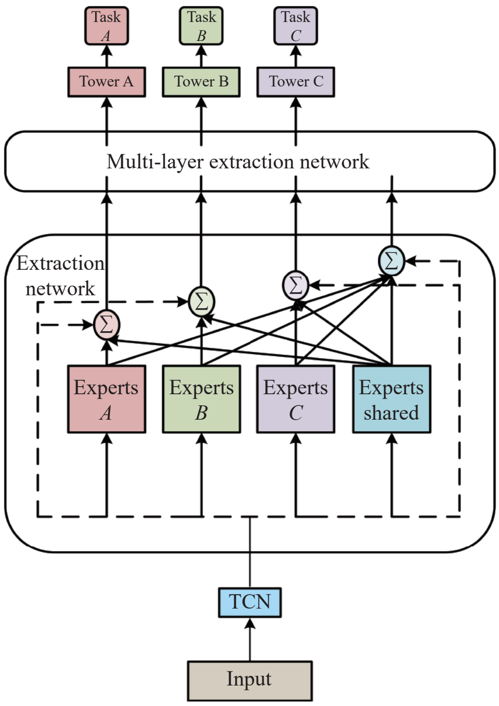


Figure 2. Multi-task learning model based on TCN and PLE.

2.4. Multi-task Loss Function Based on Dynamic Weight Average

In multi-task learning, the weight setting of the loss function has a significant impact on the effectiveness of the model. Therefore, this paper adopts a dynamic weight average [22] method to dynamically adjust the weight values in the model. In the multi-task learning model, the dynamic weight average determines the difficulty of each task by calculating the ratio of the loss values of each task for the first two times and then sets the weight value according to the difficulty.

In this study, HS division is a classification task and therefore uses the cross-entropy loss function. R assessment and TW prediction are regression tasks and therefore use the mean squared error loss function. The final loss function can be expressed as:

$$L(x_k, s_k, w_k, r_k) = \lambda_1 L_{HS}(x_k, h_k) + \lambda_2 L_{TW}(x_k, t_k) + \lambda_3 L_R(x_k, r_k) \tag{9}$$

$$L_{HS}(x_k, h_k) = -\sum_{k=1}^{M} h_k \ln(p_k) \tag{10}$$

$$L_{TW}(x_k, t_k) = -\frac{1}{K}\sum_{k=1}^{K}(t_k - f_{TW}(x_k))^2 \tag{11}$$

$$L_R(x_k, r_k) = -\frac{1}{K}\sum_{k=1}^{K}(r_k - f_R(x_k))^2 \tag{12}$$

where, h_k is the actual HS category, p_k is the predicted probability of the kth category, M is the number of HS categories, K is the total number of samples, t_k is the actual TW value, and r_k is the actual R value. The $\lambda_k$is the weight coefficient of each loss function. $f_{HS}(*)$is the HS prediction function, $f_{TW}(*)$ is the TW prediction function, and $f_R(*)$ is the R prediction function. The weight λk for the kth task can be written as follows:

$$\lambda_k(t) = \frac{K\exp\left(\frac{w_k(t-1)}{T}\right)}{\sum_i \exp\left(\frac{w_i(t-1)}{T}\right)} \tag{13}$$

$$w_k(t-1) = \frac{L_k(t-1)}{L_k(t-2)} \tag{14}$$

where $w_k(*)$is used to calculate the decline rate of loss, t is the number of iterations, T is a hyperparameter, and the larger T is, the smaller the weight difference between tasks will be. Algorithm 1 shows how to determine the parameters in the loss function based on dynamic weight average.

Algorithm 1: The proposed model optimization by using dynamic weight average loss function

Input: Training featurex_k, regression label $w_k, r_k$and classification label s_k ; Model initialization of network parameters θ and loss function parameters $\lambda 1, \lambda 2, \lambda 3$

Output: Loss function parameters $\lambda 1, \lambda 2, \lambda 3$

1: $t \leftarrow 1$;

2: while $t \leqslant T$ **do**

3: for *each batch B* **do**

4: Feeding training feature x_k and label $w_k, r_k, s_k$to the model;

5: Calculate the predicted values $fHS(xk), fHS(xk), fHS(xk)$ through forward propagation;

6: Update the model parameters $\theta, \lambda 1, \lambda 2, \lambda 3$ with Adam optimizer;

7: end

8: $t \leftarrow t+1$;

9: end

10: return $\theta, \lambda 1, \lambda 2, \lambda 3$

3. Dataset and Label Construction

This section will first introduce the PHM2010 dataset and construct two labels for reliability and health stages based on this dataset. As the dataset contains a label for cutting tools wear, it will not be discussed in this section.

3.1. Dataset Introduction

The PHM2010 dataset was collected by collecting vibration signals and cutting force signals data in the X, Y, and Z directions during the processing of 6mm alloy file cutting stainless steel workpieces as well as the root mean square value of acoustic emission signals. A total of seven sets of feature data were obtained. Through six repetition experiments, six test datasets (C1, C2, C3, C4, C5, and C6) were obtained. Among the data of C1, C4, and C6, the wear amount of three cutting edges after each cutting is also included. The average value of these three wear amounts is used as the cutting tools wear value.

3.2. Data Processing

During the cutting process of workpieces, there is a period of idle time for the cutting tools. The vibration signals generated during this idle time are also collected by sensors. These signals belong to invalid signals that can affect the effectiveness of the model, so they need to be deleted. In this paper, we use the truncation method to remove the invalid data. In addition, due to external factors such as the environment, noise is contained in the acquired data during signal collection. To clean up noisy data, we use wavelet threshold denoising. The comparison between data before and after cleaning is shown in Figure3, with blue representing processed data and yellow representing noisy data.

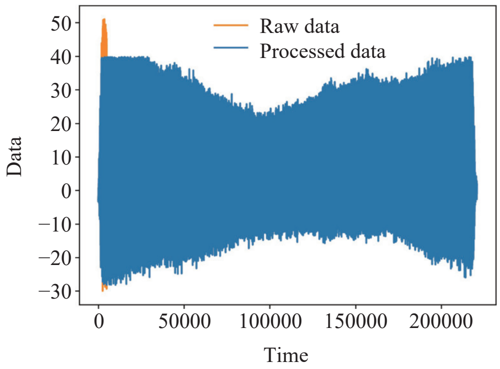


Figure 3. Data processing comparison diagram.

3.3. Feature Extraction and Selection

During the process of cutting tools machining, data is sampled at a high frequency, resulting in a large amount of data. To address this, we can extract features from the high-dimensional collected data. In this paper, we extracted 16 features from both time and frequency domains. As shown in Table 1, after feature extraction, the dimensionality of the data was reduced. However, not all of these features were effective for subsequent HSA. To address this issue, we performed feature selection using monotonicity (Mon) [23] and trendiness (Tre) [24]. Mathematically expressed as:

$$Mon = \left| \frac{df_i > 0}{K-1} - \frac{df_i < 0}{K-1} \right| \tag{15}$$

$$Tre = \frac{\left| \sum_{j=1}^{k} \left(f_i^j - \bar{f_i} \right) \left(t_j - \bar{t} \right) \right|}{\sqrt{\sum_{j=1}^{k} \left(f_i^j - \bar{f_i} \right)^2 \sum_{j=1}^{k} \left(t_j - \bar{t} \right)^2}} \tag{16}$$

$$Cie = \frac{\eta_1 Mon + \eta_2 Tre}{2} \tag{17}$$

where df_i denotes the differential between f_i^{j-1} and f_i^j, K denotes the number of samples, f_i^j denotes the jth value of feature f_i, and $\bar{f_i}$ denotes the mean of f_i, t_j is the working cycles, $\bar{t}$ is the average of all work cycles.

Table 1 Statistical feature

Feature	Formula	Feature	Formula
Maximum	$Max(x)$	Minimum	$Min(x)$
Mean	$\frac{1}{n}\sum_{i=1}^{n} x_i$	Peak	$Max(x) - Min(x)$
Absolute Mean	$\frac{1}{n}\sum_{i=1}^{n} \lvert x_i \rvert$	Rms Value	$\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$
Root Amplitude	$\left(\frac{1}{n}\sqrt{\sum_{i=1}^{n} \lvert x_i \rvert}\right)^2$	variance	$\sqrt{\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2}$
Kurtosis	$\frac{\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^4}{\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2}$	Skewness	$\frac{\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^4}{\left(\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2\right)^{\frac{3}{2}}}$
Peak Indicator	$\frac{Max(x)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}}$	Margin Indicator	$\frac{Max(x_i)}{\left(\frac{1}{n}\sqrt{\sum_{i=1}^{n} \lvert x_i \rvert}\right)^2}$
Impulse Indicator	$\frac{Max(x)}{\frac{1}{n}\sum_{i=1}^{n} \lvert x_i \rvert}$	Waveform Indicator	$\frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}}{\frac{1}{n}\sum_{i=1}^{n} \lvert x_i \rvert}$
Frequency Center	$\frac{\sum_{k=1}^{K} f_k s(k)}{\sum_{k=1}^{K} s(k)}$	Average Frequency	$\frac{\sum_{k=1}^{K} f_k s(k)}{K}$

Among them, Mon and Tre are the coefficients of monotonicity and trendiness respectively. Cie is an evaluation coefficient for features, and the larger its value is, the more beneficial it is to the model. And weight coefficients are given by η_1 and η_2, where $\eta_1 + \eta_2 = 1$. Considering that the monotonicity of features can better reflect the process of cutting tools degradation, a large value is set for η_1, with a value of 0.7, and η_2 is set at 0.3. After feature selection, redundant information still exists among high-dimensional features. To address this issue, we use a convolutional autoencoder for feature fusion.

3.4. Multi-task Learning

Multi-task learning aims to improve the performance of models by simultaneously learning multiple related tasks. Considering that TCN can mine long-term dependencies in data, PLE can effectively solve the problem of imbalanced performance among multiple tasks. Therefore, this paper proposes a multi-task network built on the basis of TCN and PLE. TCN is mainly used to discover long-term feature information between multiple tasks from input data, while PLE learns the correlation information between different tasks to improve the efficiency of learning each task and thus achieve effective prediction for multiple tasks.

3.5. HS Label Construction

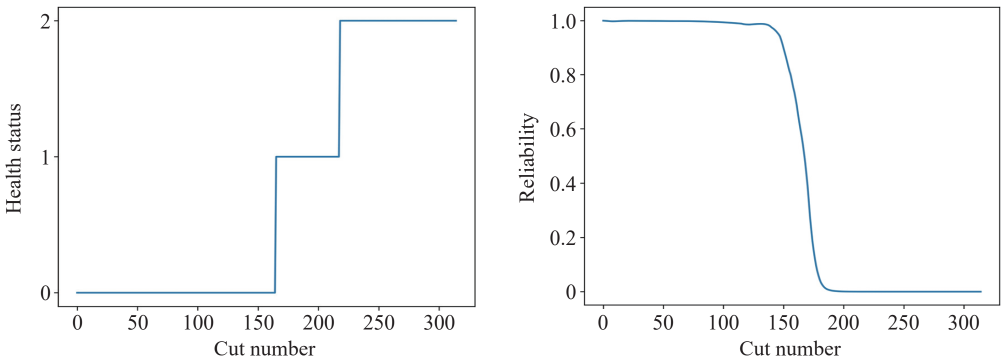
The state of cutting tools wear is usually divided into three stages: initial wear, normal wear, and severe wear [13]. If subjectively dividing these three stages, it may lead to errors. To address this issue, the paper adopts Gaussian Mixture Model (GMM) to classify the state of cutting tools wear and construct a HS label. GMM is a common clustering algorithm with soft cluster properties, making it possible to group similar patterns flexibly. GMM can be represented by the weighted sum of N components of Gaussian density:

$$P(x|\{x,\mu_m,\sum m\}) = \sum_{i=1}^{N} w_m g\left(x|\mu_m,\sum m\right) \tag{18}$$

where w_m is the mixture weight, μ_m, and $\sum m$ are the mean vector and covariance matrix respectively. Each Gaussian component can be represented as:

$$g(x|\mu_m,\sum m) = \frac{1}{(2\pi)^{D/2}\sqrt{\sum m}} e^{\left(-\frac{1}{2}(x-\mu_m)\sum_m^{-1}(x-\mu_m)\right)} \tag{19}$$

where D is the data dimension. The label constructed is shown in Figure 4(a), and since there are three states of cutting tools wear, the health stages is divided into three stages.



(a) **(b)**

Figure 4. **(a)** HS label. **(b)** R label.

3.6. R Label Construction

Logistic regression is a common method for reliability assessment [25], which can estimate probabilities through the analysis of feature parameters and predict the probability of events using the sigmoid function. According to the definition of reliability, we assume that the ith feature of time t is $X_i(t) = (x_1(t), x_2(t), \ldots, x_i(t))$. When the cutting tools is running well at time t, $y(t) = 1$; when the cutting tools is heavily worn, $y(t) = 0$. The cutting tools reliability can be expressed as:

$$R(t|X_i) = P(y_i = 1|X_i) = \frac{\exp(\alpha_0 + \alpha_1 x_1(t) + \alpha_2 x_2(t) + \cdots + \alpha_i x_i(t))}{1 + \exp(\alpha_0 + \alpha_1 x_1(t) + \alpha_2 x_2(t) + \cdots + \alpha_i x_i(t))} \tag{20}$$

where $\alpha_0, \alpha_1, \cdots, \alpha_i$ are the model coefficients. The model coefficients are solved using maximum likelihood estimation. The constructed label is shown in Figure 4(b). It reflects that as the number of uses of the cutting tools increases, the cutting tools wear intensifies and the reliability of the cutting tools gradually decreases.

4. Experiment and Analysis

This section will validate the proposed method on the PHM2010 dataset and compare it with the performance of other deep learning models.

4.1. Data Preprocessing

Due to the high dimensionality of the original data for the cutting tools, it is not conducive to subsequent data analysis. After feature extraction and feature selection, the data dimension has been reduced, but there is still data redundancy. To reduce data redundancy, a convolutional autoencoder is adopted. Since different features have different units of measure, this dataset has been preprocessed using min–max normalization:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{21}$$

where X is the original data, X' is the standardized data, X_{max} and X_{min} are the maximum and minimum values of x, respectively. Figure 5 show the normalized data.

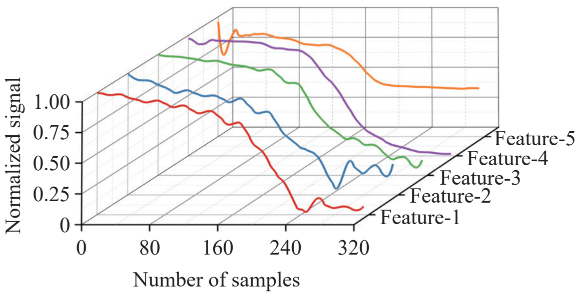


Figure 5. Normalized data.

4.2. Error Measures

HS division in tool health status assessment is a classification task, and TW prediction and R assessment are regression tasks. For HS division accuracy (ACC) and MacroF1 ($F1_{mac}$) are used as evaluation metrics, and larger values of ACC and MacroF1 indicate better model classification. ACC and $F1_{mac}$ are defined as follows:

$$ACC = \frac{n_c}{n} \times 100\% \tag{22}$$

$$F1_{mac} = \frac{2 * P_{mac} * R_{mac}}{P_{mac} + R_{mac}} \tag{23}$$

where n_c is the number of correctly predicted classifications, n is the total number of points predicted over the entire life cycle of a tool, $P_{mac} = \frac{1}{K}\sum_{k=0}^{K-1} \frac{TP_k}{TP_k + FP_k}$, $R_{mac} = \frac{1}{K}\sum_{k=0}^{K-1} \frac{TP_k}{TP_k + FN_k}$. where K denotes the number of categories of HS, TP_k denotes the number of positive samples predicted as positive samples, FP_k denotes the number of negative samples predicted as positive samples, and FN_k denotes the number of positive samples predicted as negative samples.

TW prediction and R assessment use root mean square error (RMSE), mean absolute error (MAE) and R-squared (R^2) as evaluation metrics; the smaller the values of RMSE and MAE, and the closer R^2 is to 1, the better the model's prediction is. The three evaluation metrics are defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2} \tag{24}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \left|\left(y_i - \hat{y}_i\right)\right| \tag{25}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2} \tag{26}$$

where y_i represents the true value, $\hat{y}_i$ represents the predicted value, and $\bar{y}_i$ represents the average of all predicted values.

4.3. Experimental Setting

Since only C1, C4, and C6 cutting tools in the dataset provided wear value labels, we used C1 and C4 as the training set and C6 as the test set. In terms of network architecture, a TCN with a depth of three layers will be employed. As there are three tasks, four expert modules, and three tower networks will be used in PLE. Each expert module consists of linear layers, while each tower network is composed of linear layers and ReLU activation functions, and dropout is used to reduce overfitting. The learning rate is 0.0007, the batch size is 32, and the Adam optimizer is utilized.

4.4. Ablation Study

The ablation study in this section compares the following three network structures: 1) PLE; 2) TCN-PLE denotes a multi-task network without the use of a dynamic weight average; 3) TCN-PLE-DWA denotes a multi-task network with the use of dynamic weight average.

From Table 2, it can be seen that TCN-PLE-DWA outperforms TCN-PLE in both reliability and cutting tools wear value prediction tasks, and has 6.4% higher accuracy in the classification task of health stages classification. The results show that weight adjustment of the loss function by dynamic weight average effectively solves the imbalance of effects in multi-tasks.

Table 2 Result of ablation experiments on C6

Model	R|TW RMSE	R|TW MAE	R|TW R^2	HS ACC	HS MarcoF1
PLE	**0.222**|0.116	**0.096**|0.077	**0.767**|0.615	0.831	0.815
TCN-PLE	0.241|0.096	0.118|0.072	0.739|0.7633	0.857	0.876
TCN-PLE-DWA	0.231|**0.069**	0.106|**0.052**	0.741|**0.889**	**0.921**	**0.911**

4.5. Comparative Study

In order to validate the effectiveness of the TCN-PLE model, this section compares the proposed model with models such as Multi-task CNN, Multi-task LSTM [18], Multi-task GRU and other models as well as typical multi-tasking models such as MMOE [26] and PLE. The evaluation results of different models on C6 are shown in Table 3, Figure 6 and Figure 7 present the prediction effects of different models on TW and R tasks. Finally, Figure 8 shows the accuracy and Macro-F1 of different models in dividing HS.

Table 3 Model performance comparison on C6

Model	R|TW RMSE	R|TW MAE	R|TW R^2	HS ACC	HS MarcoF1
CNN (mtl)	0.409|0.307	0.412|0.275	−11.79|−6.22	0.809	0.763
GRU (mtl)	0.358|0.337	0.349|0.315	−4.201|−3.606	0.888	0.876
LSTM (mtl)	0.369|0.309	0.355|0.294	−4.305|−2.711	0.879	0.872
MMOE	0.249|0.071	0.116|0.058	0.714|0.878	0.844	0.821
PLE	**0.222**|0.116	**0.096**|0.077	**0.767**|0.615	0.831	0.815
TCN-PLE-DWA	0.231|**0.069**	0.106|**0.052**	0.741|**0.889**	**0.921**	**0.911**

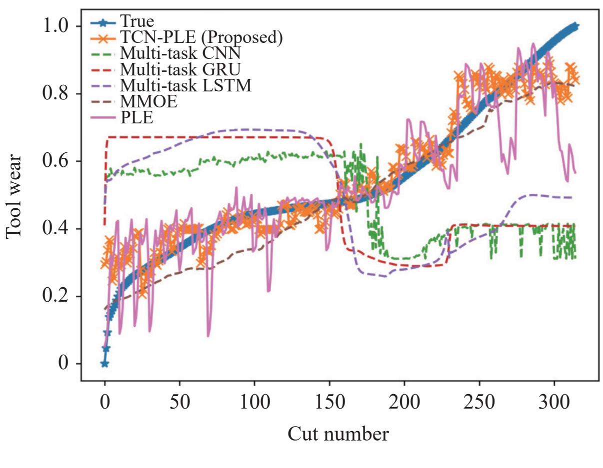


Figure 6. Cutting tools wear prediction.

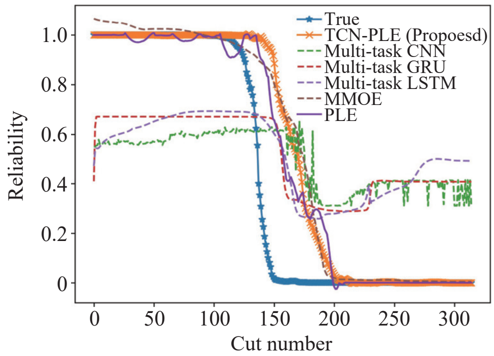


Figure 7. Reliability prediction.

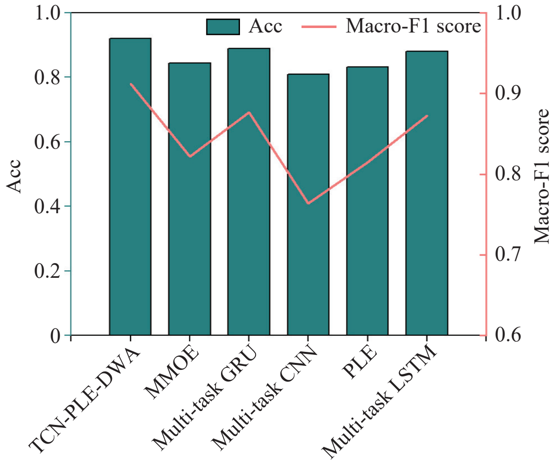


Figure 8. The performance of different models on HS.

As shown in Figure 6 and Figure 7, the classic networks were not effective for multi-task problems, as their predictions did not reflect the trends of TW and R. However, the results predicted by MMOE, PLE, and TCN-PLE networks could generally reflect the trends of TW and R. Figure 8 shows the performance of different models on the HS division. Unlike the previous ones, LSTM and GRU, two classic networks, even outperformed MMOE and PLE in this task. However, this also indicates that using classic networks for multi-tasking will sacrifice the effect of one task to enhance the effect of another task. Although they had good performance in HS division, they performed poorly in the other two prediction tasks. In contrast, TCN-PLE performed well in all three tasks without significant differences in performance among them. Tab. III. reflects the performance of different models in three tasks, where it is easy to see that TCN-PLE outperforms PLE in TW prediction and HS division. In conclusion, TCN-PLE can effectively complete the HSA of the cutting tools.

5. Conclusion

In this paper, a multi-task prediction network model based on TCN-PLE is proposed for cutting tools health status assessment method. Compared with the previous method that only studies the cutting tools wear value as a single task, this method can realize the simultaneous prediction of three tasks, HS, R and TW. The model consists of a TCN module and a PLE module. The TCN is used to extract deeper potential shared information among multiple tasks from the input data, and the PLE module is used to learn the difference information of each sub-task to realize the prediction of results for multiple tasks. The experimental results show that the TCN-PLE model has more balanced results on multiple tasks and improved results compared to multi-task networks such as MMOE and PLE than using networks such as CNN, LSTM, and GRU for multi-task learning. In the future, we plan to combine federated learning with multi-task learning to address the challenge of insufficient data on the full life cycle of a tool due to data silo problems in real industrial production processes. With the federated learning approach, we can increase the amount of data required for the tool health status assessment task by integrating and sharing data from different plants or equipment while protecting data privacy. In this way, we can further improve the accuracy of the model.

Author Contributions: Xiaoyu Yin: investigation, writing, methodology; **Xiaohua Li:** validation, visualization; **Yong Zhang:** conceptualization, funding acquisition, project administration; **Wei Zhou:** data collation, software;

Zhenxing Liu: supervision, investigation. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the National Nature Science Foundation of China (Grant 62273264).

Data Availability Statement: Not applicable.

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

References

1. Hou, W.; Guo, H.; Luo, L.; *et al*. Tool wear prediction based on domain adversarial adaptation and channel attention multiscale convolutional long short-term memory network. J. Manuf. Process., **2022**, *84*: 1339−1361. doi: 10.1016/j.jmapro.2022.11.017
2. Guo, H.; Lin, X.; Zhu, K.P. Pyramid LSTM network for tool condition monitoring. IEEE Trans. Instrum. Meas., **2022**, *71*: 2509511. doi: 10.1109/TIM.2022.3173278
3. You, Z.C.; Li, S.C.; Li, C.G.; *et al*. A novel evaluation metric based on dispersion of wear distance for in situ tool condition monitoring. IEEE Trans. Instrum. Meas., **2023**, *72*: 3501210. doi: 10.1109/TIM.2022.3224995
4. Kuntoğlu, M.; Sağlam, H. Investigation of progressive tool wear for determining of optimized machining parameters in turning. Measurement, **2019**, *140*: 427−436. doi: 10.1016/j.measurement.2019.04.022
5. Li, D.Q.; Li, Y.; Liu, C. Gaussian process regression model incorporated with tool wear mechanism. Chin. J. Aeronaut., **2021**, *35*: 393−400. doi: 10.1016/j.cja.2021.08.009
6. Salonitis, K.; Kolios, A. Reliability assessment of cutting tool life based on surrogate approximation methods. Int. J. Adv. Manuf. Technol., **2014**, *71*: 1197−1208. doi: 10.1007/s00170-013-5560-2
7. Zhang, Y.; Xin, Y.Q.; Liu, Z.W.; *et al*. Health status assessment and remaining useful life prediction of aero-engine based on BiGRU and MMoE. Reliab. Eng. Syst. Saf., **2022**, *220*: 108263. doi: 10.1016/j.ress.2021.108263
8. Zhang, Y.; Sun, J.H.; Zhang, J.; *et al*. Health state assessment of bearing with feature enhancement and prediction error compensation strategy. Mech. Syst. Signal Process., **2023**, *182*: 109573. doi: 10.1016/j.ymssp.2022.109573
9. Wang, G.; Zhang, F. A sequence-to-sequence model with attention and monotonicity loss for tool wear monitoring and prediction. IEEE Trans. Instrum. Meas., **2021**, *70*: 3525611. doi: 10.1109/TIM.2021.3117082
10. Wang, T.; Li, X.; Wang, W.; *et al*. A spatiotemporal feature learning-based RUL estimation method for predictive maintenance. Measurement, **2023**, *214*: 112824. doi: 10.1016/j.measurement.2023.112824
11. Liu, C.Q.; Li, Y.G.; Li, J.J.; *et al*. A meta-invariant feature space method for accurate tool wear prediction under cross conditions. IEEE Trans. Industr. Inform., **2022**, *18*: 922−931. doi: 10.1109/TII.2021.3070109
12. Zhou, Y.Q.; Sun, B.T.; Sun, W.F. A tool condition monitoring method based on two-layer angle kernel extreme learning machine and binary differential evolution for milling. Measurement, **2020**, *166*: 108186. doi: 10.1016/j.measurement.2020.108186
13. Qin, Y.Y.; Liu, X.L.; Yue, C.X.; *et al*. Tool wear identification and prediction method based on stack sparse self-coding network. J. Manuf. Syst., **2023**, *68*: 72−84. doi: 10.1016/j.jmsy.2023.02.006
14. Wei, X.D.; Liu, X.L.; Yue, C.X.; *et al*. Tool wear state recognition based on feature selection method with whitening variational mode decomposition. Robot. Comput. Integr. Manuf., **2022**, *77*: 102344. doi: 10.1016/j.rcim.2022.102344
15. Liu, X.L.; Liu, S.Y.; Li, X.B.; *et al*. Intelligent tool wear monitoring based on parallel residual and stacked bidirectional long short-term memory network. J. Manuf. Syst., **2021**, *60*: 608−619. doi: 10.1016/j.jmsy.2021.06.006
16. Ruder, S. An overview of multi-task learning in deep neural networks. arXiv: 1706.05098, 2017. doi:10.48550/arXiv.1706.05098.
17. Wang, Y.Q.; Qin, B.; Liu, K.; *et al*. A new multitask learning method for tool wear condition and part surface quality prediction. IEEE Trans. Industr. Inform., **2021**, *17*: 6023−6033. doi: 10.1109/TII.2020.3040285
18. Miao, H.H.; Li, B.; Sun, C.; *et al*. Joint learning of degradation assessment and RUL prediction for aeroengines via dual-task deep LSTM networks. IEEE Trans. Industr. Inform., **2019**, *15*: 5023−5032. doi: 10.1109/TII.2019.2900295
19. Kim, T.S.; Sohn, S. Y. Multitask learning for health condition identification and remaining useful life prediction: Deep convolutional neural network approach. J. Intell. Manuf., **2021**, *32*: 2169−2179. doi: 10.1007/s10845-020-01630-w
20. Bai, S.J.; Kolter, J.Z.; Koltun, V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv: 1803.01271, 2018. doi:10.48550/arXiv.1803.01271.
21. Tang, H.Y.; Liu, J.N.; Zhao, M.; *et al*. Progressive layered extraction (PLE): A novel multi-task learning (MTL) model for personalized recommendations. In *Proceedings of the 14th ACM Conference on Recommender Systems, 22–26 September 2020*; ACM: New York, 2020; pp. 269–278. doi:10.1145/3383313.3412236
22. Liu, S.K.; Johns, E.; Davison, A.J. End-to-end multi-task learning with attention. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), *Long Beach, CA, USA, 15–20 June 2019*; IEEE: New York, 2018; pp. 1871–1880. doi:10.1109/CVPR.2019.00197
23. Javed, K.; Gouriveau, R.; Zerhouni, N.; *et al*. Enabling health monitoring approach based on vibration data for accurate prognostics. IEEE Trans. Ind. Electron., **2015**, *62*: 647−656. doi: 10.1109/TIE.2014.2327917
24. Lei, Y.G.; Li, N.P.; Guo, L.; *et al*. Machinery health prognostics: A systematic review from data acquisition to RUL prediction. Mech. Syst. Signal Process., **2018**, *104*: 799−834. doi: 10.1016/j.ymssp.2017.11.016
25. Chen, B.J.; Chen, X.F.; Li, B.; *et al*. Reliability estimation for cutting tools based on logistic regression model using vibration signals. Mech. Syst. Signal Process., **2011**, *25*: 2526−2537. doi: 10.1016/j.ymssp.2011.03.001
26. Ma, J.Q.; Zhao, Z.; Yi, X.Y.; *et al*. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018*; ACM: New York, 2018; pp. 1930–1939. doi:10.1145/3219819.3220007

Citation: Yin, X.; Li, X.; Zhang, Y.; *et al*. Multi-Task Network based Health Status Assessment of Cutting Tools. *International Journal of Network Dynamics and Intelligence*. 2025, 4(2), 100008. doi: 10.53941/ijndi.2025.100008