

Article

A Machining State-Based Approach to Tool Remaining Useful Life Adaptive Prediction

Yiming Li *, Xiangmin Meng , Zhongchao Zhang and Guiqiu Song

School of Mechanical Engineering & Automation, Northeastern University, Shenyang 110819, China; m.ether0330@gmail.com (X.M.); Bruno Zhang@yahoo.com (Z.Z.); gqsong@mail.neu.edu.cn (G.S.)

* Correspondence: liyiming@mail.neu.edu.cn; Tel.: +86-024-8367-1699

Received: 31 October 2020; Accepted: 4 December 2020; Published: 6 December 2020



Abstract: The traditional predictive model for remaining useful life predictions cannot achieve adaptiveness, which is one of the main problems of said predictions. This paper proposes a LightGBM-based Remaining useful life (RUL) prediction method which considers the process and machining state. Firstly, a multi-information fusion strategy that can effectively reduce the model error and improve the generalization ability of the model is proposed. Secondly, a preprocessing method for improving the time precision and small-time granularity of feature extraction while avoiding dimensional explosion is proposed. Thirdly, an importance coefficient and a custom loss function related to the process and machining state are proposed. Finally, using the processing data of actual tool life cycle, through five evaluation indexes and 25 sets of contrast experiments, the superiority and effectiveness of the proposed method are verified.

Keywords: remaining useful life; LightGBM; loss function; curse of dimensionality; multi-information fusion

1. Introduction

Remaining useful life (RUL) prediction is an important research direction in the field of prognostics and health management (PHM). Especially in modern Computer numerical control (CNC) machining, this tool is a key component of CNC machine tools, and its life management plays an important role in the efficiency of CNC machine tools, workpiece quality and cost control. Therefore, RUL prediction of the tools in the early stage is of great significance.

In recent years, with the development of sensor technology and signal processing technology, many RUL prediction methods have been put forward one after another. These papers have contributed to sensor technologies, signal processing and decision-making strategies for process monitoring [1–3].

The estimation of RUL can be divided into two categories—i.e., physical model-based methods and data-driven methods [4]. Usually, the physical method-based model is a formula derived from failure physics to predict the theoretical damage evolution—e.g., ref [5] proposed a Taylor tool life formula-based method to estimate RUL of the tools. In ref [6,7], the Paris-Erodogan model was used to predict the crack path and crack size of bearings and gearbox pinions, respectively. The authors of [8] proposed a Paris-Erodogan model combined with the finite element model to represent the time evolution of tooth cracks in the gear. In ref [9], a method to extend the processing equation based on Taylor speed to predict the RUL of the tool is proposed.

However, when modeling the degradation process, many practical factors, such as cutting parameters and processing steps, are easily overlooked [10]. This is one of the limitations of Taylor's tool life formula. Additionally, most of the coefficients involved in the physical model are determined experimentally, which makes it difficult to match the complex industrial production environment.

To solve the above problems, the data-driven method is proposed as an alternative approach. The data-driven methods build models by mining historical data of machinery—e.g., the artificial neural network (ANN) model trained by historical data was used in [11] to predict the RUL of machinery. Luo [12] used multiple sensors to collect information for collaborative merging to provide more accurate perceptions and make further optimal decisions. In ref [13], Luo et al. also proposed that the decision-making process of automated mechanical and electrical systems relies on multiple-sensor data, and the information from distributed sensors needs to be fused in a robust way. In [14], the RUL of CNC machining lathes and the associated confidence limits are estimated by using a dynamic Bayesian network. The authors of [15] proposed a three-stage method for assessing machine health degradation and using Cox's proportional hazard model and a support vector machine (SVM) to predict RUL based on vibration signal fusion. Loutas [16] used a support vector regression (SVR) to estimate the RUL of rolling bearings and establish a data-driven approach.

Soualhi [17] proposed a method combining Hilbert–Huang Transform (HHT), SVM and SVR to detect bearings and obtain an estimation of RUL by a further time series prediction based on SVR. In ref. [18], a tool wear detection system based on Relevance Vector Machine (RVM) classifiers is constructed, which realized the multi-classification of the tool wear state during cutting. In reference [19], the SVR model is used to predict tool wear. In ref [20], the neuro-fuzzy network is adopted to predict the tool wear and RUL. Lei [21] proposed a model-based method for predicting mechanical RUL. This method used the vibration signals from the accelerated degradation testing dataset of the rolling bearings to prove the accuracy of the proposed model. Wang [22] used an adaptive Expectation-Maximum (EM) algorithm and Bayesian update algorithm to update the parameters of the data-driven model and proved the effectiveness of the proposed method in improving RUL estimation accuracy and shortening convergence time.

In addition to the abovementioned studies, in ref. [23–32], there are also many scholars who have conducted relevant research. However, the existing data-driven methods still have some limitations, and the problems that need to be solved can be summarized as follows:

- (1) Traditional predictive models cannot achieve self-adaptiveness in different complex processes. A conservative protection strategy could cause excessive tool wear and lead to a rapid increase in cutting force, affecting the processing quality of the workpiece and reducing the yield of the qualified workpiece; excessive protection strategies could waste the RUL of the tool, increase unnecessary downtime and lead to a decrease in production efficiency and an increase in manufacturing costs. Finding RUL prediction methods related to the process will effectively improve the quality of workpiece, increase production efficiency and optimize workpiece costs.
- (2) Traditional data sources rely on a single type of data and are mostly single dimensions. Such a prediction model will lack the coupling nonlinear influence factors under a different process, resulting in the reduction in credibility in the prediction process, reduction in the confidence interval, the generalization ability of the model not being strong and the actual working conditions not being able to be accurately described. It is especially important to choose the right data dimensions and combinations.
- (3) When extracting small-time granularity features in the traditional way, the features extracted by the quadratic features are directly added to the previously extracted features, which causes the inconsistency of the sample sparsity, and reduces the generalization ability of the model, the over-fitting of the model, “curse of dimensionality” and other issues. It is necessary to find a preprocessing method to solve the dimensional explosion problem.

In order to solve the above problems, this paper proposes a RUL prediction method that considers the related process and processing state. The contributions of this paper include:

- (1) A multi-information fusion strategy that can effectively reduce the model error and improve the generalization ability of the model is proposed.

- (2) A preprocessing method for improving the time precision and time granularity of feature extraction while avoiding dimensional explosion is proposed.
- (3) An importance coefficient and a custom loss function related to process and machining state are proposed. The new prediction model can realize the adaptive prediction of RUL under different processes.

The rest of this paper is organized as follows. Section 2 is a description of the proposed method. Then, the effectiveness of the proposed method is discussed in Section 3. Finally, Section 4 concludes the present work.

2. Units Proposed Method

2.1. Architecture of the Proposed Method

This paper presents a LightGBM-based approach to RUL prediction related to process. Figure 1 shows the architecture of the RUL prediction method based on LightGBM, which consists of the following three parts: data preprocessing, model training and model evaluation. In part 1, load data, a combination of different data types of vibration data and current data, verifying the superiority of the multi-information fusion combination method, are used; using the sliding windows method-clustering algorithm to expand features and avoid “curse of dimensionality”, the extracted feature matrix and the original feature matrix are merged in the model training phase. In part 2, secondary correction of residuals is made to adjust the distribution to meet the expected requirements of industry, choosing the LightGBM model that meets the fast and timely needs of the industry as the original model. The importance model and the custom loss function associated with the custom coefficient are proposed to adjust the original model. In part 3, different test sample sets are used to verify the new prediction model. (1) Visual analysis by using different quantitative indicators to ensure a more comprehensive assessment of the model’s error, fit and generalization ability. (2) Draw a histogram of the residual distribution, visually observe and evaluate the damage at both sides of the risk from the perspective of engineering significance.

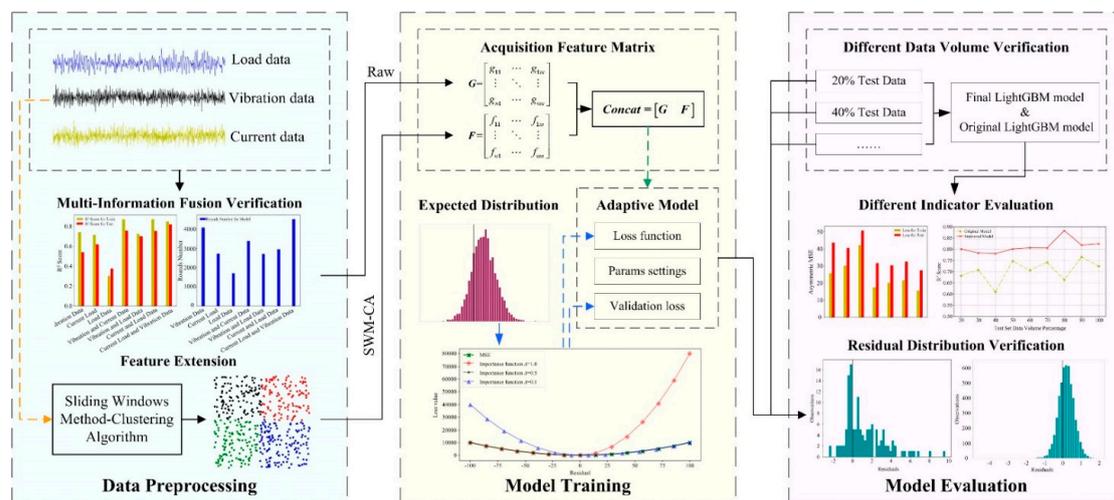


Figure 1. Architecture of remaining useful life (RUL) prediction method based on LightGBM.

2.2. SWM-CA

In the traditional RUL prediction research, the existing literature mostly performs feature extraction in the time domain and frequency domain in a given time interval. However, some transient mutation characteristics of the data within a given time interval may be masked due to larger time being fine-grained, resulting in the inability to capture the characteristics of the abrupt signal, causing imperfect or under-fitting of the model. However, the sliding windows method will generate

a large number of subdata segments. The new features extracted from a large number of subdata segments will cause inconsistency in sample sparsity, reduction in model generalization ability and model over-fitting such as the curse of dimensionality problem. Therefore, this paper proposes using the clustering algorithm (CA) to perform unsupervised analysis on all subsamples after sliding windows method (SWM) segmentation and to prevent feature redundancy caused by improved resolution while ensuring small-time granularity, preventing feature redundancy due to increased resolution and making the features perform well in the model.

Unsupervised learning is used to classify the intrinsic properties and laws of data. The clustering algorithm divides the sample set $D = \{x_1, x_2, \dots, x_m\}$ into a number of disjoint subsets, namely the sample cluster $C = \{C_1, C_2, \dots, C_k\}$, and minimizes the squared error of the divided sample clusters [33].

$$E = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|_2^2 \quad (1)$$

where μ_i is the mean vector of the sample cluster C_i , $\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$.

The effect of clustering is weighed by introducing the intra-cluster similarity DB index and the inter-cluster similarity Jaccard coefficient.

The Davies–Bouldin Index (DB index) measures internal indicators for clustering performance,

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{avg(C_i) + avg(C_j)}{d_{cen}(C_i, C_j)} \right) \quad (2)$$

where $d_{cen}()$ is used to calculate the distance between two samples, $d_{cen}(C_i, C_j) = dist(\mu_i, \mu_j)$; $avg(C)$ represents the average distance between samples within cluster C .

The Jaccard Coefficient is an external indicator for clustering performance measurement.

$$JC = \frac{a}{a + b + c} \quad (3)$$

$$\begin{cases} a = |SS|, SS = \{(x_i, x_j) | \lambda_i = \lambda_j, \lambda_i^* = \lambda_j^*, i < j\} \\ b = |SD|, SD = \{(x_i, x_j) | \lambda_i = \lambda_j, \lambda_i^* \neq \lambda_j^*, i < j\} \\ c = |DS|, DS = \{(x_i, x_j) | \lambda_i \neq \lambda_j, \lambda_i^* = \lambda_j^*, i < j\} \\ d = |DD|, DD = \{(x_i, x_j) | \lambda_i \neq \lambda_j, \lambda_i^* \neq \lambda_j^*, i < j\} \end{cases} \quad (4)$$

where set a represents a sample pair that belongs to the same cluster in C and belongs to the same cluster in C^* ; set b represents a sample pair that belongs to the same cluster in C and is not affiliated with the same cluster in C^* ; set c indicates that the sample pairs that are not affiliated to the same cluster in C and belong to the same cluster in C^* ; set d represents a sample pair that is not affiliated with the same cluster in C and is not affiliated with the same cluster in C^* .

The sliding windows method clustering algorithm (SWM-CA) has the following advantages: by using the SWM-CA to process data, it can achieve fine-grained features of small-time coverage in small-time, and it can also eliminate a large number of redundant features in extended features, so as to alleviate the feature loss and redundancy caused by the improvement of time resolution.

2.3. Loss Function of p -LightGBM

(1) The importance coefficient

When solving the actual problem, if our predicted RUL is shorter than the actual RUL, then the tool will be replaced before it expires, wasting the service life of the tool and increasing the using cost;

if the predicted RUL is longer than the actual RUL, then the tool will continue to work in the failed state for a considerable period of time, causing the workpiece to fail or even be directly scrapped. The positive and negative values of the residual between the true value of the RUL and the predicted value are often not equivalent. Ideally, we hope that the prediction model can accurately predict the RUL, but in reality, the error is unavoidable. Therefore, the residual distribution needs to be developed in the direction of actual expectation, and the risk is secondarily controlled by correcting the residual distribution. Therefore, in the face of this kind of unequal risk, this paper proposes the importance coefficient p , which evaluates the workpieces and tools under different processes and adjusts and distinguishes the punishments on both sides of the risk to achieve model self-adaptively and predictions of the RUL under different processes.

First, in order to quantify the importance of the process steps, this paper proposes the importance coefficient p related to the process and machining state. The importance coefficient p can be calculated not only by the method mentioned in this article, but also by other methods. Different importance determination strategies can be used to quantify the importance, and a simple way is used in this paper.

$$p = \frac{T_{sp}}{T_{sp} + \Delta} \quad (5)$$

where p is importance coefficient; T_{sp} is the sequence of the process in the whole process and the value is the actual processing process scaled to the interval $[0, 1]$; Δ is the machining state coefficient, and its value is the reciprocal of the importance factor of the previous process or the current process—the value is in the range of $(0, 1)$.

In order to more intuitively observe the influential rule of the processing state coefficient on the importance coefficient, several representative machining state values were selected to visually describe them, and the machining state coefficients $\Delta = 0.1$, $\Delta = 0.5$ and $\Delta = 1.0$ were used, respectively. The importance of the process in different cases, in turn, indicates the three cases in which the preorder or current machining process is important, general and unimportant. Figure 2 shows the importance coefficient curve of different processing state coefficients. It can be seen that if the importance of the preprocessing process is high, that is, Δ has a small value, the importance coefficient p can ensure that a large weight is given in the premachining state, and can make sure that harsher penalties are imposed on the negative residuals when the preorder or current processing steps are highly important.

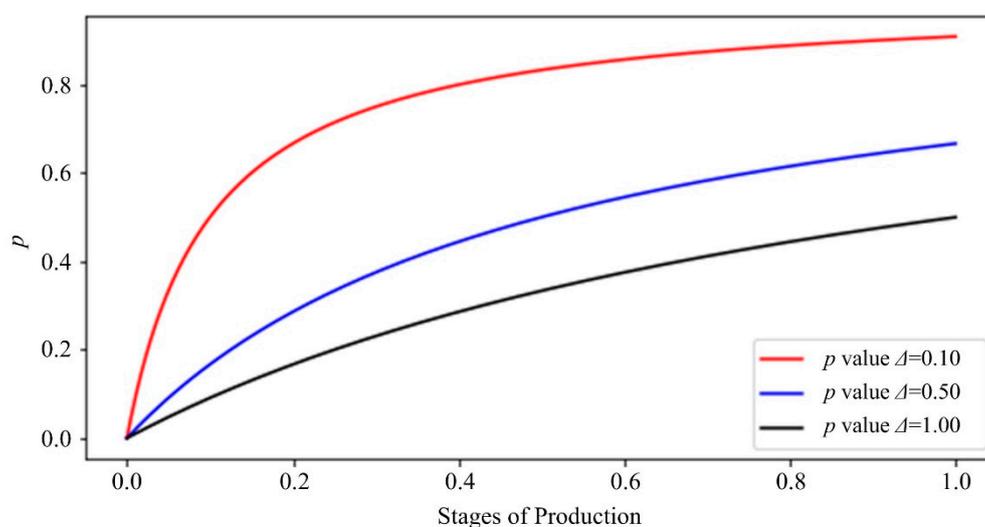


Figure 2. Importance coefficient curve of different processing state coefficients.

(2) Custom loss function

In order to achieve adaptive adjustment of the unbalanced sides of the risk, and to match the industrial production practice, this paper proposes a custom loss function (CLF) considering the importance coefficient p . We take the CLF as an example and transform it adaptively so that the model can be coupled with the actual situation of the industry.

$$\begin{cases} \frac{\delta}{1+e^{-1/(p-1/2)}}(|\alpha| - \frac{1}{2}\delta), & \alpha \geq \delta \\ \frac{\delta}{1+e^{-1/(p-1/2)}}\alpha^2, & 0 \leq \alpha < \delta \\ (1 - \frac{\delta}{1+e^{-1/(p-1/2)}})\alpha^2, & -\delta < \alpha < 0 \\ (\delta - \frac{\delta}{1+e^{-1/(p-1/2)}})(|\alpha| - \frac{1}{2}\delta), & \alpha \leq -\delta \end{cases} \quad (6)$$

where α is the residual; δ is a parameter of CLF used to enhance the robustness of the squared error loss to outliers. That is, when the residual is smaller than δ , the square error is used, and when the predicted value is larger than δ , the linear error is used.

In order to visually verify and reflect the penalty effect of CLF in different situations on both sides of the risk, CLF corresponding to different importance degrees was compared with the Mean Square Error (MSE). The figure below was used so that when we set different importance parameters and the residuals are in different positive and negative directions, the loss function imposes different degrees of punishment on the model, as shown in Figure 3.

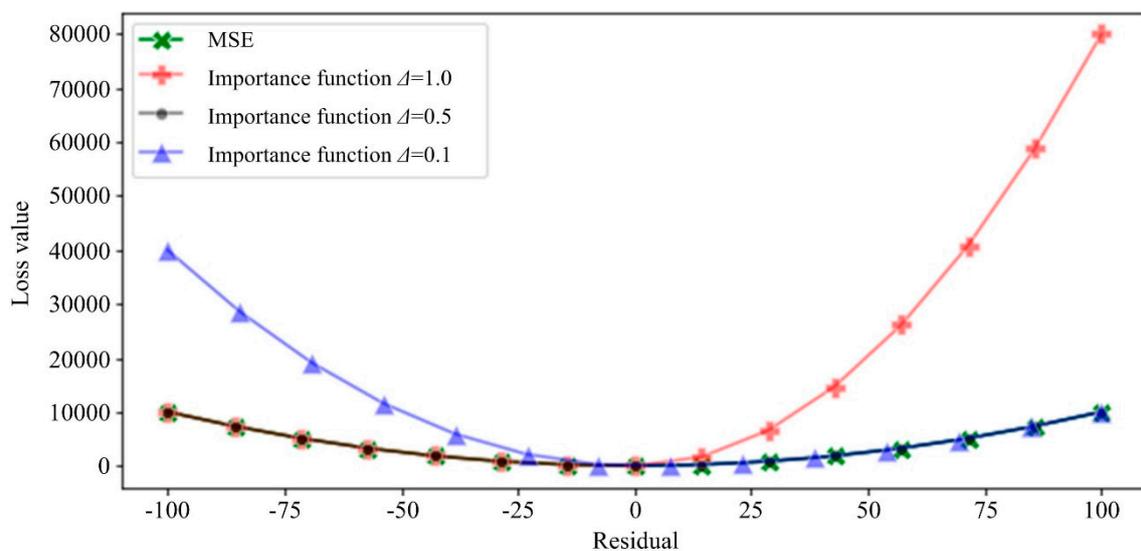


Figure 3. Comparison of importance function and MSE function of different processing state coefficients when $T_{sp} = 0.5$.

The level of punishment for the positive and negative residuals was judged by observing the magnitude of the first derivative of the loss function on both sides of the positive and negative residuals. In Figures 2 and 3, it can be seen that when the current process is in the middle of the overall process ($T_{sp} = 0.5$), the preorder steps are important ($\Delta = 0.1$), general ($\Delta = 0.5$) and not important ($\Delta = 1.0$). In the three cases, the corresponding importance of p are 0.83, 0.5 and 0.33, respectively. When $p = 0.5$, the custom function degenerates to the equivalent of the penalty on both sides, and the effect is the same as the normal loss function. In the case of $p = 0.5$, we applied different levels of punishment to the positive and negative residuals, and the effect is in line with the expected hypothesis.

(3) Verify the engineering practice effect of CLF

In order to verify the self-adaptiveness of the proposed CLF, five groups of comparative engineering practice are conducted in this section. We used the LightGBM model for training [34]. LightGBM is a Gradient Boosting Regression Tree (GBDT) model based on the histogram algorithm in the Boosting framework. LightGBM can speed up and optimize computing memory and efficiency. LightGBM uses two ways to achieve its fast accuracy, namely gradient-based one side sampling (Goss) and exclusive feature bundling (EFB). Because such advantages are extremely important for large-scale data, we chose this model as the target model. Figure 4 shows the flowchart of verification. Table 1 is the model fit conditions under different comparisons.

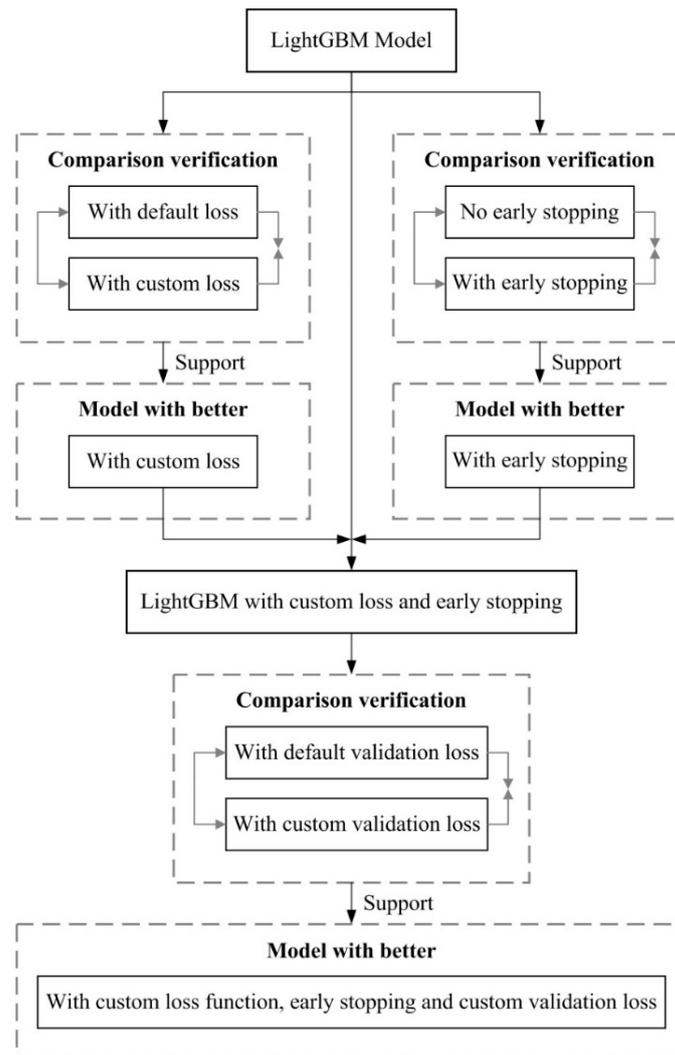


Figure 4. The flowchart of verification comparison.

Table 1. Model fitting condition under different comparisons.

| Model Setting | Boosting Rounds | MSE (Train) | Asymmetric Loss (Train) | Asymmetric Loss (Test) |
|--|-----------------|-------------|-------------------------|------------------------|
| LightGBM default | 100 | 0.236246 | 0.628296 | 1.31852 |
| LightGBM with custom loss | 100 | 0.330155 | 0.27638 | 0.819872 |
| LightGBM with early stopping | 780 | 0.137639 | 0.0531724 | 0.783725 |
| LightGBM with early stopping and custom loss | 1848 | 0.162248 | 0.0136494 | 0.868132 |
| LightGBM with early stopping, custom loss and custom validation loss | 241 | 0.22839 | 0.13002 | 0.740384 |

The comparison validation in Table 1 was designed to verify the effectiveness using the following three aspects: the CLF, the early stopping hyperparameter and an external validation function which matches the CLF. First, under the condition of fixed Boosting Rounds, the asymmetric loss (train) values of default LightGBM and LightGBM with custom loss were 0.628296 and 0.27638, LightGBM with custom loss exhibited a better performance. At the same time, LightGBM with custom loss performs equally well in more important test sets, effectively improving the prediction accuracy of the model. Second, LightGBM with an early stop had better loss convergence and a better generalization ability than default LightGBM. Its test set had a 38% reduction in asymmetric loss and was particularly effective. Using LightGBM with early stop and custom loss can also optimize the purpose of improvement. Last, the model was optimized using a custom external validation function. The effect is more obvious when compared to functions that do not use custom external validation. The model using the custom validation function reaches the optimal value when training 241 rounds, and the one without using it reaches the optimal value in the 1848 round. In the case where both are optimal at the same time, the model generalization ability using the custom verification loss was better, and its performance in the test set improved by 12%. The result shows that: (1) LightGBM with custom loss has a smaller error of convergence value; (2) using the custom loss function and matching the loss verification function as the external verification loss, the obtained model has better adaptability and robustness, and has faster error convergence speed. To further illustrate this, we observed the residual histogram (Figure 5) for more details.

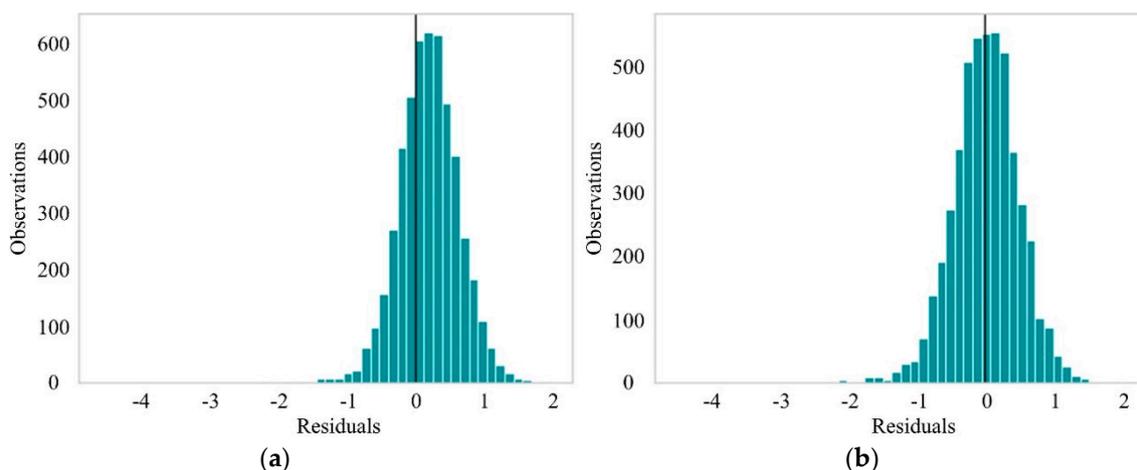


Figure 5. Error histograms of predictions from different loss functions ($\Delta = 0.1$). (a) LightGBM with early stopping, custom loss and custom validation loss. (b) default LightGBM.

We selected the training samples with the preorder process importance $\Delta = 0.1$ for training, and according to the previous theory, we must have more severe penalties for dealing with negative residuals. As can be seen from Figure 5a, the LightGBM model considering the custom loss made more predictions on the right side of the error histograms, and the residual shifted to the right—namely, the actual value was greater than the predicted value. It is further proved that using the custom loss function proposed in this paper can effectively reduce the loss in the opposite direction to the industrial expectation and make a second correction to the loss in order to obtain a lower hazard in practice.

3. Experiments and Discussion

In this section, the proposed method is verified by five evaluation indexes and 25 groups of comparative experiments. Experiments in Sections 3.2–3.4 were conducted to verify the effectiveness of multi-information fusion strategy and data type combination, SWM-CA and RUL prediction models.

3.1. Process Description

The data used in the experiments were collected from a CNC machining process. This process is a full-life cycle cutting experiment using a ball-end mill. Since the experimental process is to collect the complete life cycle of multiple sets of tools as soon as possible, we simplified the processing process. Although the simplified process will be different from the real industrial production, the data obtained in this way can still be used for data analysis and theoretical verification. Throughout the experiment, the data collects controller signals, process information and sensor data during processing according to the Cyber Physical Systems (CPS) framework. Altogether 745 min, 24.78 GB and 149 valid data samples were collected. Using the full-life data of the nine ball-end mills as the basis for building the dataset. We separated the data from the two ball-end milling cutters as part of the generalization ability of the final test model. This part of the data does not participate in any training of the model. In the experiment, we collected three kinds of signals: load signal, vibration signal and current signal. In addition, in the vibration signal and current signal, we collected three-dimensional vibration signal data, which are x, y, z direction and three-phase current data. Figure 6 shows the main components and sensor positions of the CNC.

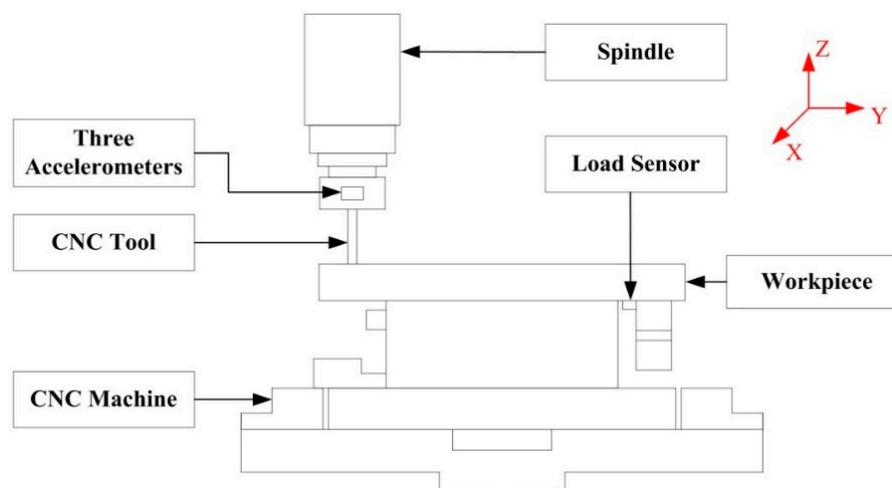


Figure 6. The main components and sensor positions of the Computer numerical control (CNC).

3.2. Verifying the Validity Multi-Information Fusion Strategy and Data Type Combination

Ref. [35–37] proved that the degradation of the tool can be estimated by the current signal/electric power of the machine tool; ref. [38,39] used the spindle load as a data source to predict the RUL of the tool; ref. [40–45] used vibration signals as the basis of the data-driven model to achieve RUL prediction. The above studies are based on a single data type. In order to verify the superiority of the proposed multi-information fusion strategy and the validity of the data type combination, this paper used the LightGBM model to compare the prediction results of engineering practice with seven different types and different combinations of data types. Since this paper is trying to emphasize the idea of feature extraction in the time dimension and the integration of the model with engineering practice, it has not expanded and extended more features. Therefore, we selected some representative features as basic features and input them into the model. Feature sets including time domain, frequency domain and time–frequency features were considered, as shown in Table 2.

Table 2. Compare the feature set used for verification.

| Features | Equations |
|---------------------------------|---|
| Root mean square x_{rms} | $x_{rms} = \left(\frac{1}{N} \sum_{i=1}^N x_i^2\right)^{1/2}$ |
| Square root amplitude x_{sra} | $x_{sra} = \left(\frac{1}{N} \sum_{i=1}^N \sqrt{ x_i }\right)^2$ |
| Kurtosis value x_{kv} | $x_{kv} = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{\sigma}\right)^4$ |
| Skewness value x_{sv} | $x_{sv} = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{\sigma}\right)^3$ |
| Peak to peak value x_{ppv} | $x_{ppv} = \max(x_i) - \min(x_i)$ |
| Crest factor x_{cf} | $x_{cf} = \max(x_i) / \left(\frac{1}{N} \sum_{i=1}^N x_i^2\right)^{1/2}$ |
| Impulse factor x_{if} | $x_{if} = \max(x_i) / \frac{1}{N} \sum_{i=1}^N x_i $ |
| Clearance factor x_{CF} | $x_{CF} = \max(x_i) / \left(\frac{1}{N} \sum_{i=1}^N \sqrt{ x_i }\right)^2$ |
| Center of gravity frequency | $FC = \int_0^{+\infty} fS(f)df / \int_0^{+\infty} S(f)df$ |
| Mean square frequency | $MSF = \int_0^{+\infty} f^2S(f)df / \int_0^{+\infty} S(f)df$ |
| Root mean square frequency | $RMSF = \sqrt{MSF}$ |
| Variance of frequency | $VF = \int_0^{+\infty} (f - FC)^2S(f)df / \int_0^{+\infty} S(f)df$ |

This focuses on asymmetric loss to compare the error sizes and generalization capabilities of the model. The R^2 Score in statistics, which is the goodness of fit, was introduced for normalization comparison. Table 3 is the prediction result of different types and different combinations of data types.

Table 3. Predictions for different types and different combinations data types.

| Model Setting | Boosting Rounds | Asymmetric Loss (Test) | Asymmetric Loss (Train) | R^2 Score (Test) | R^2 Score (Train) |
|--|-----------------|------------------------|-------------------------|--------------------|---------------------|
| LightGBM with Vibration Data | 4083 | 43.616 | 25.629 | 0.5407 | 0.7411 |
| LightGBM with Current Data | 2713 | 40.591 | 30.194 | 0.6184 | 0.7146 |
| LightGBM with Load Data | 1707 | 50.859 | 42.304 | 0.3748 | 0.3030 |
| LightGBM with Vibration and Current Data | 3384 | 31.744 | 17.579 | 0.7561 | 0.8718 |
| LightGBM with Vibration and Load Data | 2699 | 30.360 | 20.254 | 0.6987 | 0.7237 |
| LightGBM with Current and Load Data | 2940 | 32.607 | 21.716 | 0.7545 | 0.8732 |
| LightGBM with Current, Load and Vibration Data | 4522 | 27.472 | 15.615 | 0.8176 | 0.8465 |

In Table 3, it can be seen that asymmetric loss does not show good results on both the test set and the training set when using the single data type, and the best fit of the single type of data is only 0.6184. In the case of using two data types, we can clearly see that the combination of any data type is greatly improved compared to the previous single data type. By contrast, when using three data types, the current signal, load signal and vibration signal can be effectively combined to better characterize the change of cutting force. Although the convergence speed of the model slowed down,

there was a significant improvement in the generalization ability of the model (27.472 in asymmetric loss (test) and 0.8176 in R^2 Score (test)). The result shows that the multi-information fusion technology implemented by three types of data combination can effectively reduce the error size of the model and improve the generalization ability of the model and better predict the target value. For data combinations, we wanted to perform common feature extraction on different data signals and we did not want to spend too much time at this step. This step focuses more on the effect changes brought about by the use of data types, rather than the application of data feature extraction methods. In order to more intuitively compare the error size and generalization ability of the model, Figure 7 shows a comparison chart of different evaluations.

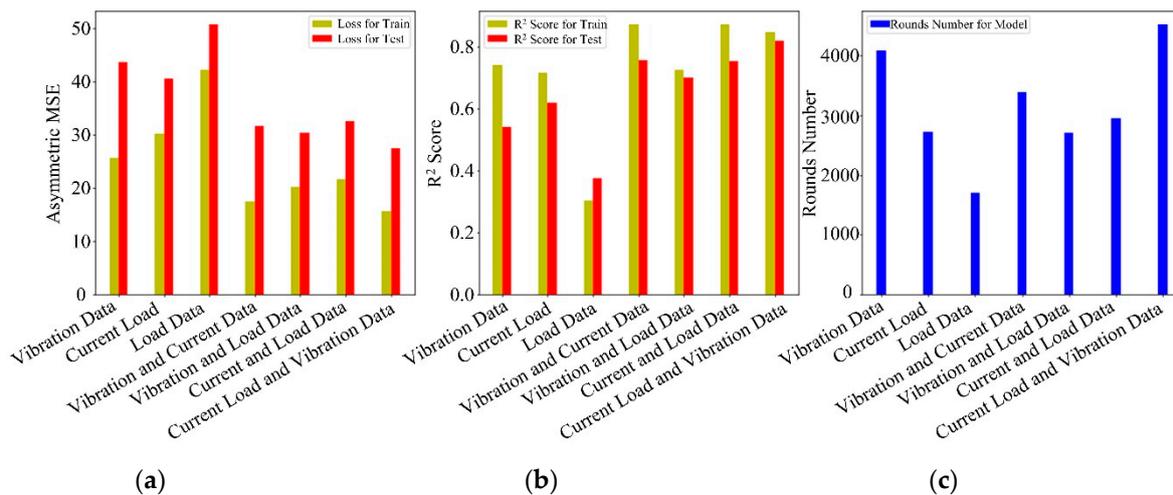


Figure 7. Comparison chart of different evaluations. (a) Asymmetric Loss. (b) R^2 Score. (c) Boosting Rounds.

3.3. The Effectiveness of SWM-CA

Feature extraction in Section 3.2 was performed over the entire time interval, which will cause some of the local key features to be ignored due to excessive time granularity. In order to solve this problem, this paper used the sliding windows method to perform small-time fine-grained interception and increment raw data in units of unit time and extracted the feature, and then was taken as a feature vector into the clustering algorithm. The Silhouette coefficient determines the optimal number of clusters.

$$s(i) = \frac{\frac{a}{a+b+c} - \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{\text{avg}(C_i) + \text{avg}(C_j)}{d_{\text{cen}}(C_i, C_j)} \right)}{\max \left\{ \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{\text{avg}(C_i) + \text{avg}(C_j)}{d_{\text{cen}}(C_i, C_j)} \right), \frac{a}{a+b+c} \right\}} \quad (7)$$

where C_i, C_j, k is Davies–Bouldin index params; a, b, c is Jaccard coefficient params.

The closer the distance between the samples are in the cluster, the farther the distance between the samples is, namely the smaller the DB index is, the larger the Jaccard is, the larger the average contour coefficient is, and the better the clustering effect is, and so a best cluster number $k = 4$ is obtained. The next step it to set the number of classification categories, and then obtain the subdata groups of the four groups of categories, and finally add all the subdata segment features to the original feature matrix. This is transformed into a small-time fine-grained feature representation which uses the density center data of these four types of data groups as a substitute. Figure 8 shows the flowchart of SWM-CA.

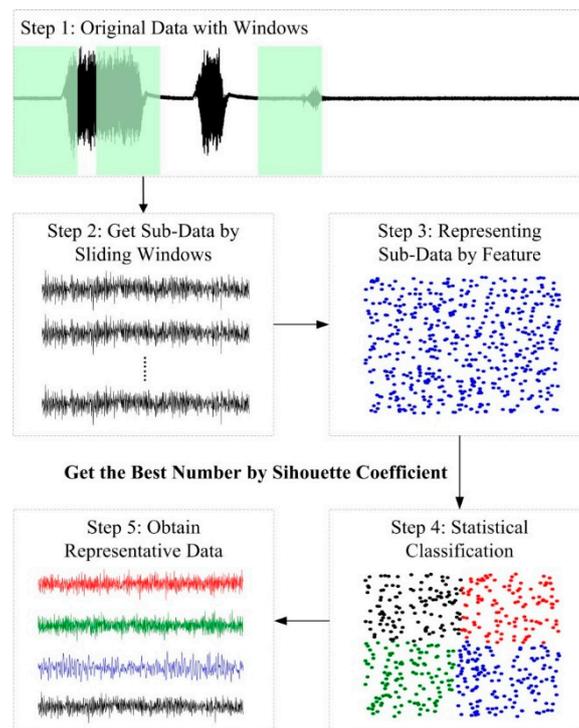


Figure 8. The flowchart of sliding windows method clustering algorithm (SWM-CA).

In order to verify effect of the proposed SWM-CA, two groups of comparative experiments were conducted. That is, under the condition of using the same model, the residual distribution was compared between the model using SWM-CA and the model without SWM-CA, and the distribution range and distribution law were observed.

Figure 9 shows the result of the comparison. Since we imposed penalties on negative residuals, the data distribution is unequal on both sides. The residuals used in the original data features have a wide range of distributions. By contrast, the residual distribution range of the secondary extraction feature using SWM-CA is narrowed, and the residual is more concentrated on the positive side, which ensures that the absolute error is reduced, further reducing the harm in real production. The result shows that: (1) the secondary feature extraction using SWM-CA can effectively cover small-time fine-grained features; (2) SWM-CA can effectively expand features and enhance the generalization ability of the model.

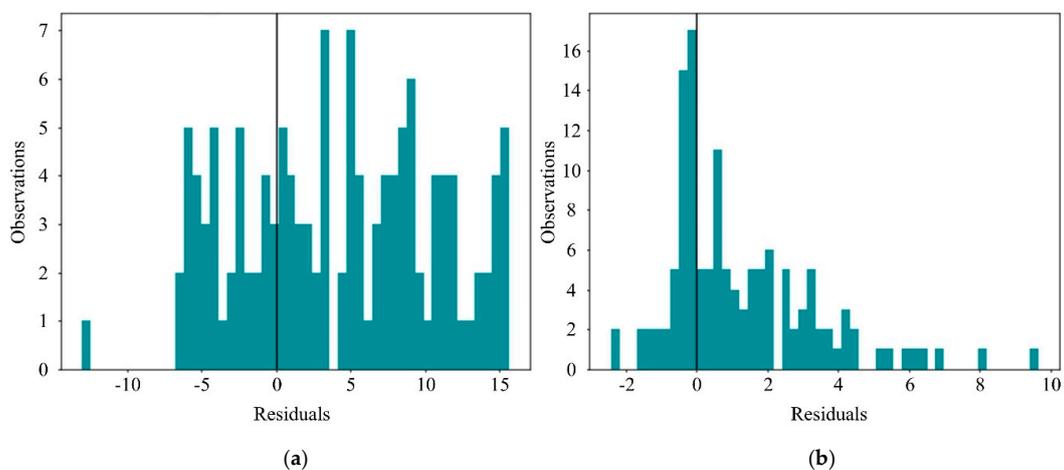


Figure 9. Error histograms of predictions from different group. (a) LightGBM with raw feature. (b) LightGBM with secondary extraction feature.

3.4. Comparative Experiments of the RUL Predictive Model

In order to verify the effect of the improved LightGBM, comparisons were made using an improved machine learning model (p -LightGBM) and an unimproved machine learning model. Table 4 shows the result of the comparison. First of all, for the data part, we used the dataset that had not been used by the model—that is, the full-life data collected by the two ball-end mills. Through such dataset verification, we could verify the generalization ability of the model because, compared to our previous training set, this part of the data and the data used may have the same distribution as the previous dataset. However, since we are not using this dataset, we can use it to generalize the validation model. The improved model exhibits a better performance, although there is a slight loss in the fit and loss on the training set, but overall, it has better adaptability to unknown data. It has a faster convergence speed and is more in line with the actual industrial demand forecast results.

Table 4. Predictions for different types and different combinations data types.

| Model Setting | Boosting Rounds | Asymmetric Loss (Test) | Asymmetric Loss (Train) | R2 Score (Test) | R2 Score (Train) |
|-------------------------------------|-----------------|------------------------|-------------------------|-----------------|------------------|
| LightGBM with Default Data Feature | 4522 | 27.472 | 15.615 | 0.81760 | 0.8465 |
| p -LightGBM with New Data Feature | 4227 | 24.383 | 16.542 | 0.8268 | 0.8357 |

In order to more comprehensively observe the performance of the improved model before and after the improvement, explain how well the model performs in sample intervals of different sizes, comparisons were made by using sample test sets of different sizes (20% sample and 100% sample).

Figure 10 shows the result of the comparison. With the improved LightGBM, the error and accuracy will fluctuate slightly as the sample size changes. However, overall, it is still better than the state before the improvement. The result shows that, considering the strategy of multi-information fusion, combining different data types, using the SWM-CA to unsupervise the data and extract the features twice, and adopting the improved LightGBM model, can maintain better generalization ability and prediction accuracy.

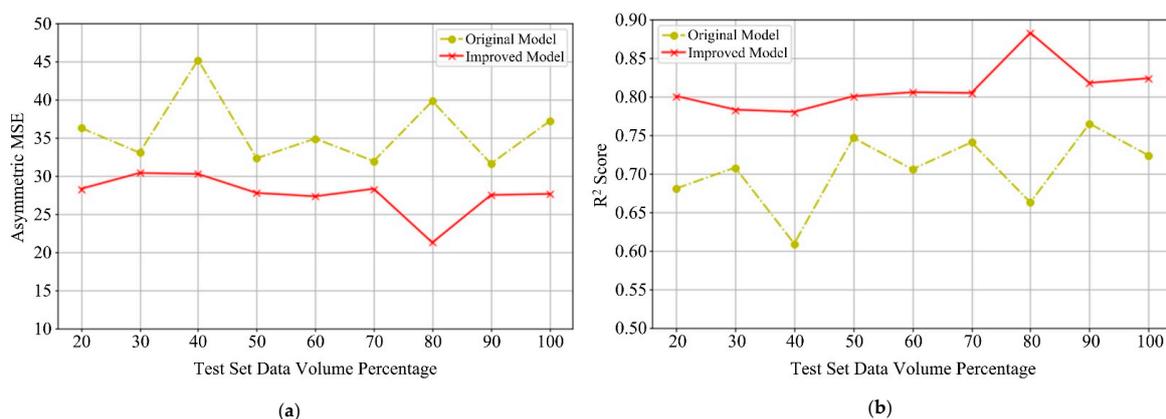


Figure 10. Comparison results of different evaluations under different sample. (a) Asymmetric Loss. (b) R² Score.

4. Conclusions

This paper presents a LightGBM-based approach to RUL prediction related to process and machining state. We establish a CLF considering the importance coefficient p to realize the self-adaptive adjustment of the imbalance between the two sides of the risk, reduce the loss in the opposite direction to the industrial expectation, achieve a secondary correction of the loss and match the industrial production practice. In the data preprocessing stage, a multi-information fusion strategy was adopted

to reduce the error of the model, improving the confidence of features and improving the generalization ability of the model. At the same time, the unsupervised analysis of the subsamples after sliding windows method segmentation was performed by a clustering algorithm, which not only takes the small-time fine-grained features into account but also avoids the curse of dimensionality of features. The effectiveness of the proposed method was verified by an experimental comparison evaluation. Based on data evaluation results, it can be observed that considering the multi-information fusion strategy and the secondary extraction feature using the sliding windows method and clustering algorithm, the improved LightGBM model was used for RUL prediction, which is more in line with industrial needs and has better generalization ability and prediction accuracy.

For further research in the future, we will not be limited to the GBDT model, but will focus more on the application of attention technology. We will find a way to combine the actual industrial pain points and technical advantages for further exploration. There is no doubt that this method can not only be applied to cutting tools, but can also be transplanted to other objects. Therefore, future research will carry out further experiments and research on the universality of this method.

Author Contributions: This research was designed by Y.L. and X.M.; Formal analysis was completed by Y.L. and X.M.; the first version manuscript was prepared by Y.L. and X.M.; this version of the manuscript was read, approved and substantially contributed to by Y.L., X.M., Z.Z. and G.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (51805077), the Natural Science Foundation of Liaoning Province (2019-MS-118) and the Fundamental Research Funds for the Central Universities of China (N2024002/17).

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

References

1. Byrne, G.; Dornfeld, D.; Inasaki, I.; Ketteler, G.; König, W.; Teti, R. Tool Condition Monitoring (TCM)—The Status of Research and Industrial Application. *CIRP Ann.* **1995**, *44*, 541–567. [\[CrossRef\]](#)
2. Teti, R.; Jemielniak, K.; O'Donnell, G.; Dornfeld, D. Advanced monitoring of machining operations. *CIRP Ann.-Manuf. Technol.* **2010**, *59*, 717–739. [\[CrossRef\]](#)
3. Caggiano, A.; Rimpault, X.; Teti, R.; Balazinski, M.; Chatelain, J.F.; Nele, L. Machine learning approach based on fractal analysis for optimal tool life exploitation in CFRP composite drilling for aeronautical assembly. *CIRP Ann.-Manuf. Technol.* **2018**, *67*, 483–486. [\[CrossRef\]](#)
4. Sikorska, J.; Hodkiewicz, M.; Ma, L. Prognostic modelling options for remaining useful life estimation by industry. *Mech. Syst. Signal Process.* **2011**, *25*, 1803–1836. [\[CrossRef\]](#)
5. Silva, R.; Reuben, R.; Baker, K.; Wilcox, S. Tool wear monitoring of turning operations by neural network and expert system classification of a feature set generated from multiple sensors. *Mech. Syst. Signal Process.* **1998**, *12*, 319–332. [\[CrossRef\]](#)
6. Li, Y.; Billington, S.; Zhang, C.; Kurfess, T.; Danyluk, S.; Liang, S. Adaptive prognostics for rolling element bearing condition. *Mech. Syst. Signal Process.* **1999**, *13*, 103–113. [\[CrossRef\]](#)
7. Kacprzyński, G.J.; Sarlashkar, A.; Roemer, M.J.; Hess, A.; Hardman, B. Predicting remaining life by fusing the physics of failure modeling with diagnostics. *JOM* **2004**, *56*, 29–35. [\[CrossRef\]](#)
8. Li, C.J.; Lee, H. Gear fatigue crack prognosis using embedded model, gear dynamic model and fracture mechanics. *Mech. Syst. Signal Process.* **2005**, *19*, 836–846. [\[CrossRef\]](#)
9. Marksberry, P.; Jawahir, I. A comprehensive tool-wear/tool-life performance model in the evaluation of NDM (near dry machining) for sustainable manufacturing. *Int. J. Mach. Tools Manuf.* **2008**, *48*, 878–886. [\[CrossRef\]](#)
10. Wang, P.; Gao, R.X. Adaptive resampling-based particle filtering for tool life prediction. *J. Manuf. Syst.* **2015**, *37*, 528–534. [\[CrossRef\]](#)
11. Tian, Z. *An Artificial Neural Network Approach for Remaining Useful Life Prediction of Equipments Subject to Condition Monitoring*; IEEE: New York, NY, USA, 2009; pp. 143–148.
12. Luo, R.C.; Chang, C.-C.; Lai, C.C. Multisensor Fusion and Integration: Theories, Applications, and its Perspectives. *IEEE Sens. J.* **2011**, *11*, 3122–3138. [\[CrossRef\]](#)

13. Luo, R.C.; Chang, C.-C. Multisensor Fusion and Integration: A Review on Approaches and Its Applications in Mechatronics. *IEEE Trans. Ind. Inform.* **2011**, *8*, 49–60. [[CrossRef](#)]
14. Tobon-Mejia, D.A.; Medjaher, K.; Zerhouni, N. CNC machine tool's wear diagnostic and prognostic by using dynamic Bayesian networks. *Mech. Syst. Signal Process.* **2012**, *28*, 167–182. [[CrossRef](#)]
15. Tran, V.T.; Pham, H.T.; Yang, B.-S.; Nguyen, T.T. Machine performance degradation assessment and remaining useful life prediction using proportional hazard model and support vector machine. *Mech. Syst. Signal Process.* **2012**, *32*, 320–330. [[CrossRef](#)]
16. Loutas, T.H.; Roulias, D.; Georgoulas, G. Remaining Useful Life Estimation in Rolling Bearings Utilizing Data-Driven Probabilistic E-Support Vectors Regression. *IEEE Trans. Reliab.* **2013**, *62*, 821–832. [[CrossRef](#)]
17. Soualhi, A.; Medjaher, K.; Zerhouni, N. Bearing Health Monitoring Based on Hilbert-Huang Transform, Support Vector Machine, and Regression. *IEEE Trans. Instrum. Meas.* **2015**, *64*, 52–62. [[CrossRef](#)]
18. Wang, G.; Yang, Y.; Xie, Q.; Zhang, Y. Force based tool wear monitoring system for milling process based on relevance vector machine. *Adv. Eng. Softw.* **2014**, *71*, 46–51. [[CrossRef](#)]
19. Zhang, K.-F.; Yuan, H.-Q.; Nie, P. A method for tool condition monitoring based on sensor fusion. *J. Intell. Manuf.* **2015**, *26*, 1011–1026. [[CrossRef](#)]
20. Zhang, C.; Yao, X.-F.; Zhang, J.; Jin, H. Tool Condition Monitoring and Remaining Useful Life Prognostic Based on a Wireless Sensor in Dry Milling Operations. *Sensors* **2016**, *16*, 795. [[CrossRef](#)]
21. Lei, Y.; Li, N.; Gontarz, S.; Lin, J.; Radkowski, S.; Dybala, J. A Model-Based Method for Remaining Useful Life Prediction of Machinery. *IEEE Trans. Reliab.* **2016**, *65*, 1314–1326. [[CrossRef](#)]
22. Wang, Y.; Yizhen, P.; Zi, Y.; Jin, X.; Tsui, K.-L. A Two-Stage Data-Driven-Based Prognostic Approach for Bearing Degradation Problem. *IEEE Trans. Ind. Inform.* **2016**, *12*, 924–932. [[CrossRef](#)]
23. Pedregal, D.J.; Carnero, M.C. State space models for condition monitoring: A case study. *Reliab. Eng. Syst. Saf.* **2006**, *91*, 171–180. [[CrossRef](#)]
24. Zhang, J.; Starly, B.; Cai, Y.; Cohen, P.H.; Lee, Y.-S. Particle learning in online tool wear diagnosis and prognosis. *J. Manuf. Process.* **2017**, *28*, 457–463. [[CrossRef](#)]
25. Liu, X.; Song, P.; Yang, C.; Hao, C.; Peng, W. Prognostics and Health Management of Bearings Based on Logarithmic Linear Recursive Least-Squares and Recursive Maximum Likelihood Estimation. *IEEE Trans. Ind. Electron.* **2017**, *65*, 1549–1558. [[CrossRef](#)]
26. Wu, J.; Su, Y.; Cheng, Y.; Shao, X.; Deng, C.; Liu, C. Multi-sensor information fusion for remaining useful life prediction of machining tools by adaptive network based fuzzy inference system. *Appl. Soft Comput.* **2018**, *68*, 13–23. [[CrossRef](#)]
27. Li, X.; Ding, Q.; Sun, J.-Q. Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliab. Eng. Syst. Saf.* **2018**, *172*, 1–11. [[CrossRef](#)]
28. Zhang, C.; Yao, X.; Tan, W.; Zhang, Y.; Zhang, F. Proactive Scheduling for Job-Shop Based on Abnormal Event Monitoring of Workpieces and Remaining Useful Life Prediction of Tools in Wisdom Manufacturing Workshop. *Sensors* **2019**, *19*, 5254. [[CrossRef](#)]
29. Yang, Y.; Guo, Y.; Huang, Z.; Chen, N.; Li, L.; Jiang, Y.; He, N. Research on the milling tool wear and life prediction by establishing an integrated predictive model. *Measurement* **2019**, *145*, 178–189. [[CrossRef](#)]
30. Guo, L.; Lei, Y.; Xing, S.; Yan, T.; Li, N. Deep Convolutional Transfer Learning Network: A New Method for Intelligent Fault Diagnosis of Machines With Unlabeled Data. *IEEE Trans. Ind. Electron.* **2019**, *66*, 7316–7325. [[CrossRef](#)]
31. Kumar, A.; Chinnam, R.B.; Tseng, F. An HMM and polynomial regression based approach for remaining useful life and health state estimation of cutting tools. *Comput. Ind. Eng.* **2019**, *128*, 1008–1014. [[CrossRef](#)]
32. An, Q.; Tao, Z.; Xu, X.; El Mansori, M.; Chen, M. A data-driven model for milling tool remaining useful life prediction with convolutional and stacked LSTM network. *Measurement* **2020**, *154*, 107461. [[CrossRef](#)]
33. Selim, S.Z.; Ismail, M.A. K-Means-Type Algorithms: A Generalized Convergence Theorem and Characterization of Local Optimality. *IEEE Trans. Pattern Anal. Mach. Intell.* **1984**, *6*, 81–87. [[CrossRef](#)] [[PubMed](#)]
34. Li, Y.; Meng, X.; Zhang, Z.; Song, G. A Remaining Useful Life Prediction Method Considering the Dimension Optimization and the Iterative Speed. *IEEE Access* **2019**, *7*, 180383–180394. [[CrossRef](#)]
35. Al-Sulaiman, F.A.; Baseer, M.A.; Sheikh, A.K. Use of electrical power for online monitoring of tool condition. *J. Mater. Process. Technol.* **2005**, *166*, 364–371. [[CrossRef](#)]

36. Liu, H.; Lee, B.; Tarn, Y. Monitoring of drill fracture from the current measurement of a three-phase induction motor. *Int. J. Mach. Tools Manuf.* **1996**, *36*, 729–738. [[CrossRef](#)]
37. Li, X.; Tso, S. Drill wear monitoring based on current signals. *Wear* **1999**, *231*, 172–178. [[CrossRef](#)]
38. Zhu, K.; Zhang, Y. A generic tool wear model and its application to force modeling and wear monitoring in high speed milling. *Mech. Syst. Signal Process.* **2019**, *115*, 147–161. [[CrossRef](#)]
39. Nouri, M.; Fussell, B.K.; Ziniti, B.L.; Linder, E. Real-time tool wear monitoring in milling using a cutting condition independent method. *Int. J. Mach. Tools Manuf.* **2015**, *89*, 1–13. [[CrossRef](#)]
40. Chen, S.-L.; Jen, Y. Data fusion neural network for tool condition monitoring in CNC milling machining. *Int. J. Mach. Tools Manuf.* **2000**, *40*, 381–400. [[CrossRef](#)]
41. Mahajan, A.; Wang, K.; Ray, P. Multisensor integration and fusion model that uses a fuzzy inference system. *IEEE/ASME Trans. Mechatron.* **2001**, *6*, 188–196. [[CrossRef](#)]
42. Banerjee, T.P.; Das, S. Multi-sensor data fusion using support vector machine for motor fault detection. *Inf. Sci.* **2012**, *217*, 96–107. [[CrossRef](#)]
43. Drouillet, C.; Karandikar, J.; Nath, C.; Journeaux, A.-C.; El Mansori, M.; Kurfess, T. Tool life predictions in milling using spindle power with the neural network technique. *J. Manuf. Process.* **2016**, *22*, 161–168. [[CrossRef](#)]
44. Ghorbani, S.; Kopilov, V.V.; Polushin, N.I.; Rogov, V.A. Experimental and analytical research on relationship between tool life and vibration in cutting process. *Arch. Civ. Mech. Eng.* **2018**, *18*, 844–862. [[CrossRef](#)]
45. Huang, M.; Liu, Z. Research on Mechanical Fault Prediction Method Based on Multifeature Fusion of Vibration Sensing Data. *Sensors* **2019**, *20*, 6. [[CrossRef](#)] [[PubMed](#)]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).