Heterogeneous Sensors-based Feature Optimisation and Deep Learning for Tool Wear Prediction

Xiaoyang Zhang[1], Sheng Wang[1], Weidong Li[1,2*], Xin Lu[3]

[1] Faculty of Engineering, Environment and Computing, Coventry University, UK

2 School of Logistics Engineering, Wuhan University of Technology, China

3 Faculty of Science & Technology, Bournemouth University, UK

* Corresponding author: weidong.li@coventry.ac.uk

Abstract

During machining processes, accurate prediction of cutting tool wear is prominent to prevent ineffective tool utilisation and significant resource waste. Tool wear conditions and progression involve complex physical mechanisms, and a promising approach is to deploy heterogeneous sensors and design a deep learning algorithm to conduct real-time tool wear monitoring and precious prediction. To tackle the challenge of deep learning algorithms in processing complex signals from heterogeneous sensors, in this paper, a systematic methodology is designed to combine signal de-noising, feature extraction, feature optimisation and deep learning-based prediction. In more details, the methodology is comprised of the following three steps: (i) signal de-noising is carried out by a designed Hampel filter-based method to eradicate random spikes and outliers in the signals for raw data quality enhancement; (ii) features extracted from heterogeneous sensors in the time and frequency domains are optimised using designed recursive feature elimination and cross-validation (RFECV)-based and Isomap-based methods; (iii) a convolutional neural networks (CNN) algorithm is devised to process the optimised features to implement tool wear prediction. In this paper, a case study showed that 80% features were reduced from the originally extracted features and 86% prediction accuracy was achieved based on the developed methodology. The presented methodology was benchmarked with several main-stream methodologies, and the superior performance of the methodology over those comparative methodologies in terms of prediction accuracy was exhibited.

Keywords: Tool wear prediction, feature optimisation, convolutional neural networks (CNN), heterogeneous sensors

1. Introduction

According to statistics [1], 7%-20% downtime for a computerised numerical control (CNC) machine tool system has been caused by cutting tool failures, 3%-12% of machining time has been wasted on replacing tools, and only 50%-80% of a tool life has been effectively used. An essential root cause of those phenomena is that manufacturing companies usually do not have the capability of accurately predicting tool wear conditions so that tools are either over-used or under-used. It is essential to develop an effective tool wear prediction so that the entire lifespans of tools can be utilised.

With regard to tool failure, there are several forms like flank wear, crater wear, plastic deformation and fractures. Flank wear and crater wear are two main forms of tool wear [2]. The flank wear is caused by abrasion and adhesion, and it occurs at the contact face between the tool flank and the workpiece surface [3]. The crater wear happens at the rake face close to the tool edge, which is caused by the high temperature chips flow. The crater wear is not a severe deterioration of the cutting tool, as it needs long-term accumulation to cause failure and usually show on ductile materials [4]. Thus, flank wear is an important factor to characterise the tool wear and commonly used for the tool wear monitoring [5].

Physics-based modelling has been actively investigated to predict various tool wears. However, due to the dynamic operation environments of machine tool systems, it is challenging to developing the relevant accurate physical mechanisms. For the research of flank wear prediction, data-based modelling (e.g., by leveraging the latest deep learning technologies) has been actively explored as an increasingly popular solution for flank wear prediction [7]. In such a data-based modelling, signals (e.g., vibration, acoustic emission (AE) or electricity consumption) from sensors mounted on a CNC machine tool system for tool condition monitoring are mined to estimate tool wear conditions and progression. However, there are still challenges for the data-based modelling:

- In view of increasingly complex machining processes and dynamic machining models, factors that affect flank wear are not limited to processing parameters. Uncertain parameters such as tool runout, lubrication and residual stress will also affect flank wear greatly. Therefore, research on the more advanced data-based modelling is imperative to better quantify those uncertainty parameters;
- Most of the developed research for the flank wear prediction has low scalability in dealing with various machining circumstance. For data-based modelling, the sensitivity of each sensor signal to the flank wear may vary with different machining conditions, predicting flank wear using solidified signal-wear patterns will result in reduced accuracy [6];
- Considering that the material and geometry of a tool exhibit non-uniform characteristics throughout the tool lifespan, a single type of sensors is not always sensitive and effective.

Thus, it is a promising approach to deploying multiple types of sensors (called heterogenous sensors in this research) onto a CNC machine tool system to provide more comprehensive condition monitoring information of cutting tools along the entire tool lifecycles. Based on sensor signals, artificial intelligence algorithms have been designed to extract features from the signals to perform tool wear analysis. Conventional machine learning algorithms, including the artificial neural network (ANN), fuzzy inference system and support vector machine (SVM), have been applied in this research area. An important characteristic of the conventional algorithms is that features are explicitly defined and extracted to support following reasoning processes. Nevertheless, the performances of the algorithms are incompetent in processing complex signals. In recent years, deep learning algorithms, such as convolutional neural networks (CNN), recurrent neural networks (RNN), or long short-term memory (LSTM), have been actively explored to various engineering applications including tool wear prediction

[8-11]. The distinguishing advantage of deep learning algorithms is that features can be intelligently extracted from highly non-linear and complex signals via a cascade of multiple (deep) layers and leaning mechanisms. However, it is challenging for deep learning algorithms to evaluate signals from heterogenous sensors. Signals from heterogenous sensors contain some redundant and irrelevant data, which will hugely prolong the training time of deep learning algorithms. Moreover, the redundant and irrelevant data also result in highly correlated features extracted from the signals, leading to over-fitting to particular features in modelling and poor accuracy of tool wear prediction. For features extracted from the signals, their importance contributions or relevance to tool wear are not the same, and it is not necessary to engage all of features for analysis. Thus, for those features extracted from signals of heterogenous sensors, it is critical to pinpoint and eliminate correlated features and less important (irrelevant) features in order to maintain essential information to better support deep learning algorithms.

To address the above challenging issues, in this paper, a novel systematic methodology for tool wear prediction is designed. The methodology is innovative in that a deep learning algorithm is hybridised with signal de-noising and feature optimisation methods to process complex signals from heterogenous sensors for facilitating tool wear prediction. The characteristics and functions of the methodology are below:

- It is essential to de-noise signals of heterogenous sensors in order to provide high-quality raw data for tool wear prediction. In this research, a Hampel filter-based method is devised to clean signals by eliminating random spikes and outliers in the signals;
- Based on features extracted from signals of heterogenous sensors in the time and frequency domains, recursive feature elimination, cross-validation (RFECV)-based and Isomap-based methods are designed to systematically identify the redundancy and relevance of the features to tool wear conditions for feature optimisation;
- A CNN model is developed to process optimised features for tool wear prediction with achieved high computational efficiency and prediction accuracy;
- A case study was employed for the validation of the research methodology. The study achieved a 80% condensation of correlated or less relevant features from the originally extracted features and 86% prediction accuracy on tool wear. Moreover, the methodology was benchmarked with several mainstream algorithms. Results showed that the methodology achieved a better performance over the comparative algorithms in terms of prediction accuracy.

The rest of the paper is organised as follows. Some relevant literatures are reviewed in Section 2. In Section 3, the research methodology is explained in detail. Section 4 elaborates the evaluation processes of the methodology and the benchmarking results with comparative algorithms. Conclusions are drawn in Section 5.

2. Literature Review

During a CNC machining process, the wear progression of a cutting tool can be reflected in the fluctuation of the sensor signals used to monitor cutting tool conditions. However, these signals are usually enormous in data generation and multi-dimensional. Features, which are extracted from sensor signals to create a smaller set of important information in the time and frequency domains, can reveal the inherent characteristics of the signals and be easier to interpret tool conditions and predict tool lives [12]. For instance, the skewness feature characterises the degree of asymmetry around the mean of the distribution of signals, which is affected by tool conditions and tool wear; the maximum and minimum feature values of signals can display the fluctuation of the signal amplitude that is closely related to tool wear [13].

Furthermore, optimisation on signal features has been designed to facilitate tool wear prediction [14-15], which is to keep only the critical features with obvious physical and statistical significance to tool wear conditions and progression by minimising redundant or irrelevant features. Feature optimisation exhibits advantages in the following aspects: (i) the required data storage space for extracted features can be reduced; (ii) the computational efficiency for tool wear prediction can be optimised; (iii) the correlations between features are alleviated to avoid overfitting and boost the prediction accuracy. Developed methods of feature optimisation are generally categorised as filter and wrapper methods [16], their cons and pros are summarised in Table 1.

Table 1: Pros and cons of the filter methods and the wrapper methods.

Method	Advantage	Disadvantage
Filter method	• A relatively simpler and faster computing process	• Feature dependencies not considered • Low accuracy
Wrapper method	• Feature evaluation with cross-validation for ensuring the method reliability • Higher accuracy	• Extensive computation

The filtering methods select representative features through a threshold set manually [17]. The weakness of the filtering methods is that they only consider individual features and ignore the possible interactions and correlations between features. Consequently, the methods may result in the acquisition of highly relevant features, which will make the performance of a diagnosis task poor in terms of algorithm training efficiency and accuracy [18].

The wrapper methods embed an intelligent learning estimator to evaluate feature importance sequentially to obtain the best features ultimately. The higher accuracy of the intelligent learning estimator, the better performance of the wrapper methods in identifying optimal features [19]. In the work of [20], a recursive feature elimination (RFE) method was designed to delete the worst features according to the importance ranking of multi-dimensional features. Nevertheless, in REF, a challenging issue is how to select a suitable classifier to distinguish features [21-23]. In the work of [24], several classifiers, namely t-test filtering, the sparse-constrained dimensionality reduction model (SCDRM), and SVM, were tested for feature optimisation in an application of the image classification. SVM showed a better performance among the comparative methods. In addition, a similar study on feature optimisation of breast cancer data based on SVM was conducted in [25]. However, for the above research works, there are two issues to be further resolved: (i) the methods are ineffective for processing heterogenous sensor signals; (ii) there still exist a number of correlated features unidentified and unremoved yet by using REF and SVM.

Some research works for the above two types of feature optimisation methods are summarised in Table 2.

Table 2: Summaries of some research works for feature optimisation.

Method	Algorithm	Sensor signal	Reference
Filter method	Pearson correlation coefficient (PCC)	Vibration	[8]
		Current	[26]
	mRMR	Cutting force	[27]
	ANOVA	Vibration, Cutting force	[28]
	Fisher's discriminant ratio (FDR)	Sound, Voltage	[29]
		Vibration, Cutting force	[30]
Wrapper method	Genetic algorithm	Cutting force, Vibration, AE	[1]
		Cutting force, Vibration, AE	[31]
	Stepwise selection	Vibration	[32]
	Decision tree	Vibration	[13]

In recent years, deep learning algorithms have been actively applied for tool wear prediction. Nevertheless, deep algorithms are unable to highlight the correlations of features and their contributions to an application problem explicitly. Besides, deep learning algorithms are ineffective in processing heterogenous sensor signals, which hinder the algorithm accuracy and computational efficiency [33]. Feature optimisation can be considered as an effective data processing method to be incorporated with deep learning algorithms to facilitate decision making [34-36]. Some research works developed feature optimisation on deep learning algorithms for biomedical applications. For instance, in a study of the cancer image classifications, the impacts of feature optimisation on the performance of three deep learning algorithms, i.e., CNN, deep belief network (DBN) and recursive neural network (RNN), were evaluated [37]. Experimental results verified that feature optimisation improved the training accuracy of the deep learning algorithms. There are also some related research works developed for machine diagnostics applications. In the work [38], three feature optimisation methods, i.e., random forest (RF), linear SVM, and radial basis function SVM were combined with DNN be applied to three datasets: (i) acoustic signals collected in reciprocating air compressors; (ii) vibration signals collected from deep groove ball bearings; (iii) signals for steel plate faults. Prediction results showed that significant improvements were achieved.

Nevertheless, though various research works were conducted, majority of the previous research works focused on feature optimisation from the signals of homogenous sensors. To the best of our knowledge, systematic processes to optimise features extracted from signals of heterogenous sensors, which are more effective in monitoring and analysis of tool wear conditions, are still rare. It is imperative to conduct research to systematically assess the redundancy and relevance of features related to tool wear prediction based on heterogenous sensors. Meanwhile, it is also foremost to validate the proposed research methodology on its effectiveness in tool wear prediction using a complex case study under an industrial environment.

3. Methodology

3.1 Overall framework of the methodology

The flow of the methodology, which is depicted in Fig. 1, consists of the following steps:

- Signal pre-processing: Signals from heterogenous sensors mounted on a CNC machine tool system are pre-processed using an Hampel filter-based method to diminish noises in the signals and improve the quality of the signals;
- Feature extraction: A series of features in the time and frequency domains are extracted from the pre-processed signals;
- Feature optimisation: To expedite the computational efficiency and improve the accuracy of tool wear prediction, it is necessary to optimise (minimise) the number of features from the above extracted features. For this purpose, in this research, two methods are developed: (i) a RFECV-based method designed to determine optimal features and remove less important (relevant) features; (ii) an Isomap-based method designed to further optimise feature selection;
- Tool wear prediction: A CNN model is trained and applied to accomplish tool wear prediction.

The rational and justifications of using Hampel filter, REFCV and Isomap in this research will be explained in the following subsections.

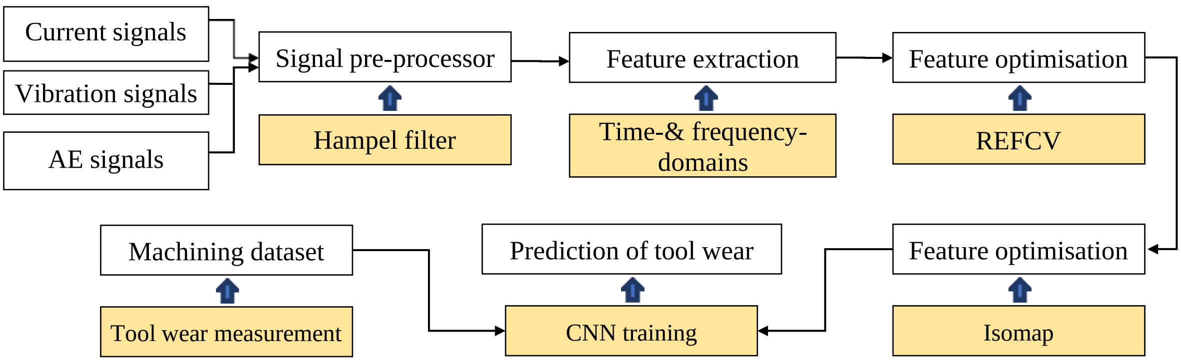


Fig. 1: The framework of the methodology developed in this research.

In this research, a heterogenous sensors-based dataset for cutting tool monitoring from the NASA Ames and UC Berkeley [39] was used as a case study to elaborate and validate the proposed approach. In the case study, accelerometers (the ENDEVCO model 7201-50), AE sensors (the WD model 925), and current sensors, were deployed on a CNC machine tool system (the Matsuura machining centre MC-510V) to monitor tool conditions. In more details, two different sensors were used to monitor the AC (the OMRON K3TB-A1015 current converter) and DC (the CTA 213 current sensor) current of the machine tool system. Two accelerometers and two AE sensors were installed on the working table and the spindle of the machine tool system respectively, to monitor vibration and acoustic signals. The specifications of these sensors are summarised in Table 3.

Table 3: The specification of the applied sensors.

Sensor model	Frequency	Measurement Range	Sensitivity
ENDEVCO model 7201-50 [40]	13KHz	0~2000gpk	50 pC/g
WD model 925 [41]	2MHz	125~2MHz	56 V/(m/s)
OMRON K3TB-A1015[42]	5KHz	64~160 A	-
CTA 213[43]	1KHz	0-600A	-

During the machining process, a 70mm face milling tool with 6 inserts (KC710) was adopted. The inserts are coated with multiple layers of titanium carbide, titanium carbonitride, and titanium nitride for roughing. The geometrical parameters and technical data of KC710 are given in Table 4.

Table 4. The geometrical parameters and technical data of KC710 [44].

Inscribed circle size (mm)	Thickness (mm)	Cutting edge length (mm)	Corner radius (mm)	Cutting edges per insert
9.5250	3.1750	9.5250	0.7940	4

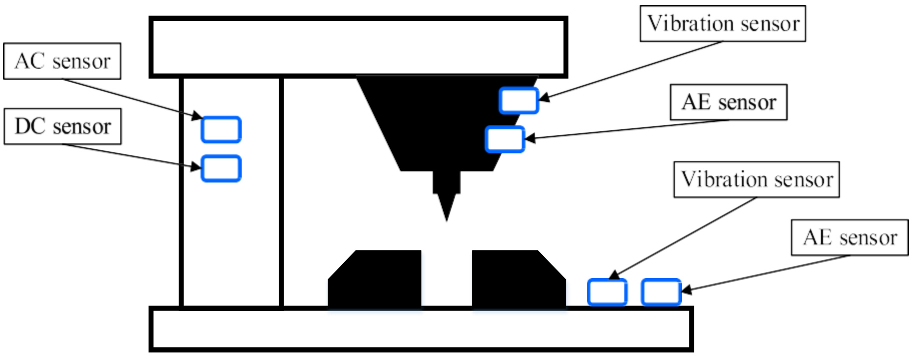


Fig. 1: The deployment of sensors on the machine tool system.

Fig. 2 shows the deployment of the sensors in the machine tool system. In the following subsections, the signals from the sensors are denoted as V_S (vibration signals from the spindle), V_T (vibration signals from the table), AE_S (AE signals from the spindle), AE_T (AE signals from the table), AC (AC signals) and DC (DC signals).

The entire dataset of the case study is divided into fifteen machining setups, and each setup is defined as a run-to-fail machining process based on the set of parameters (cutting speed, depth of cut, feed and workpiece material). According to the guidance of the industrial applicability and manufacturer, which is indicated in the third-party dataset used in this research [39], the cutting speed was set to 200m/min, corresponds to the 826 rev/min of the spindle speed, and two depth of cut 1.5mm and 0.75mm, two feed 0.5mm/rev and 0.25mm/rev were employed. The fifteen setups and machining parameters are detailed in Table 5. In the table, the specifications of materials used are defined. The definitions are represented in standard nomenclature.

Table 5: Fifteen machining setups in the case study.

Setup	Cutting speed (m/min)	Depth of cut (mm)	Feed （mm/rev）	Workpiece material
1	200	1.5	0.5	Cast iron
2	200	0.75	0.5	Cast iron
3	200	0.75	0.25	Cast iron
4	200	1.5	0.25	Cast iron
5	200	1.5	0.5	stainless steel J45
6	200	0.75	0.25	stainless steel J45
7	200	0.75	0.5	stainless steel J45
8	200	1.5	0.5	Cast iron
9	200	1.5	0.25	Cast iron
10	200	0.75	0.25	Cast iron
11	200	0.75	0.5	Cast iron
12	200	0.75	0.25	stainless steel J45
13	200	0.75	0.5	stainless steel J45
14	200	1.5	0.25	stainless steel J45
15	200	1.5	0.5	stainless steel J45

During the entire machining process for each setup, intermediate measurements on tool wear conditions were arranged. The period between two consecutive measurements in the machining process of a setup is defined as a run. In the experiment, for the fifteen setups, a total of 164 runs were recorded. After each run, the flank wear of a tool was measured using a microscope. The flank wears of the experiment are shown in Table 6. Moreover, according to the ISO3685:1993, in a metal cutting, the standard threshold for a uniform tool flank wear is usually set as 0.4mm. In the subsequent analysis, the threshold of the flank wear is set to be 0.4mm. Based on this value, the wear status of a tool is judged to be either unworn or worn. For simplicity, in the rest of the paper, the flank wear is called wear.

Table 6: Measured wears for the experiment of the case study (the worn cases are highlighted).

	Setup														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Run	Wear (mm)														
1	0	0.08	0	0.08	0	0	0	0	0	0	0.05	0	0	0.08	0.05
2	0.04	0.14	0.13	0.13	0.16	0.09	0.18	0.1	0.04	0.04	0.08	0.05	0.09	0.15	0.13

3	0.07	0.14	0.13	0.2	0.29	0.13	0.3	0.14	0.08	0.07	0.1	0.10	0.17	0.28	0.24
4	0.11	0.14	0.17	0.31	**0.44**	0.22	0.36	0.19	0.16	0.07	0.12	0.13	0.24	0.37	0.31
5	0.16	0.15	0.19	0.35	**0.53**	0.24	**0.44**	0.27	0.25	0.08	0.17	0.17	0.30	**0.48**	**0.40**
6	0.20	0.16	0.20	**0.40**		0.34	**0.62**	0.38	0.36	0.09	0.20	0.32	0.35	**0.56**	**0.62**
7	0.24	0.18	0.23	**0.49**		**0.46**		**0.47**	**0.43**	0.10	0.24	0.38	**0.60**	**0.70**	
8	0.29	0.22	0.23			**0.53**		**0.64**	**0.47**	0.12	0.32	**0.49**	**0.81**		
9	0.28	0.26	0.26					**0.81**	**0.53**	0.16	0.36	**0.56**	**1.14**		
10	0.32	0.31	0.28						**0.70**	0.18	**0.40**	**0.68**			
11	0.38	0.38	0.33							0.2	**0.45**	**0.83**			
12	**0.40**	**0.43**	0.36							0.23	**0.49**	**0.92**			
13	**0.43**	**0.48**	**0.44**							0.26	**0.58**	**1.07**			
14	**0.45**	**0.55**	**0.55**							0.29	**0.65**	**1.30**			
15	**0.5**									0.31		**1.53**			
16	**0.53**									0.37					
17	**0.54**									**0.40**					
18										**0.42**					
19										**0.47**					
20										**0.57**					
21										**0.65**					
22										**0.68**					
23										**0.76**					

The purpose of this research is to prove the feasibility of the methodology of heterogeneous sensors-based feature optimisation and deep learning for tool wear prediction. On the other hand, tool wear is a complicated physical phenomenon and more comprehensive investigations on the mechanism of tool wear are necessary. For instance, it is known that the cutting speed is an important influential parameter in tool wear. Due to the limitation of the current available data, the relevant research was not carried out in this paper. In the future work, more experiences will be conducted to evaluate the impact of varying cutting speeds on tool wear and the underlying mechanism.

3.2 Signal pre-processing

During the process of tool condition monitoring, sensor signals may contain random spikes or outliers (e.g., caused by erroneous values of sensors, environment-related disturbance, etc.), which can adversely affect the prediction accuracy of tool wear. To alleviate the issue, in this research, a signal de-noising method enabled by the Hampel filter is adopted. The Hampel filter was proved to have a good performance of removing spikes or outliers without affecting the entire signals [45]. The processes of the method are below.

The median and the median absolute deviation (MAD) are two important estimators of the Hampel filter. For the signals from a sensor, they are represented as $A = \{x_1, x_2, \ldots, x_n\}$. The signals are divided into multiple subsets by a moving window with a fixed size (2K+1). The median of the signals under a moving window is below:

$$m_{K,j} = median\{x_{j,} \dots , x_{j+2K}\} \tag{1}$$

where $m_{K,j}$ is the median of the subset j of A; x_j represents the j-th signal in A; K is the half-width of this subset.

For instance, K is chosen to be 2 in this paper, and the median of the first subset of A can be obtained as $m_{2,1} = median\{x_1,x_2,x_3,x_4,x_5\}$

The scale estimator MAD of the subset can be calculated by:

$$\text{MAD} = median\{|x_j - m_{K,j}|, \dots , |x_{j+2K} - m_{K,j}|\} \tag{2}$$

which is used to estimates the window's standard deviation $\sigma_{K,j}$ below:

$$\sigma_{K,j} = \alpha \cdot \text{MAD} \tag{3}$$

where α is the unbiased estimator of a Gaussian distribution, and $\alpha \approx 1.4826$ [46].

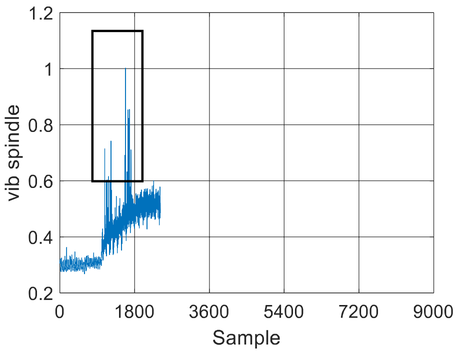
Thereby, for the above example, $\sigma_{2,1} = \alpha \cdot median \left\{ \begin{matrix} |x_1 - m_{2,1}|, |x_2 - m_{2,1}|, |x_3 - m_{2,1}|, \\ |x_4 - m_{2,1}|, |x_5 - m_{2,1}| \end{matrix} \right\}$

Next, spikes or outliers in signals can be judged based on the following Equation (4). That is, if the difference between a signal x_j and the median $m_{K,j}$ is greater than $t \cdot \sigma_{K,j}$, x_j is judged to be a spikeor outlier signal and it will be replaced with the median $m_{K,j}$. Otherwise, the signal remains the same.

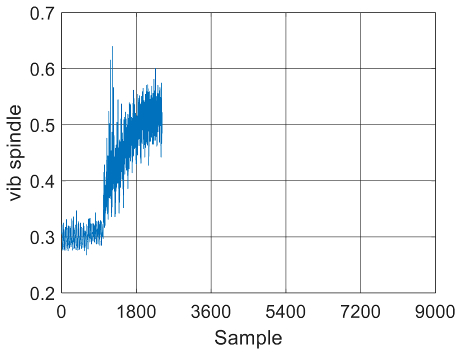
$$H_{K,j}(x_j) = \begin{cases} x_j, & |x_j - m_{K,j}| \le t \cdot \sigma_{K,j} \\ m_{K,j}, & |x_j - m_{K,j}| > t\sigma_{K,j} \end{cases} \tag{4}$$

where $H_{K,j}$ is the Hampel filter; t is the scale factor, which equals to 3 in general [47]. For the first and last K number of samples, the filter algorithm prepends or appends the frame with zeros respectively to have a complete window.

Fig. 3 shows the result of applying the Hampel filter-based method to the vibration signals in the case study of this research. In the original vibration signals of Run 6 of Setup 1 (shown in Fig. 3(a)), some spikes can be observed. These spikes would degrade the prediction accuracy of tool wear. After the application of the Hampel filter-based method, signal spikes are eliminated from the original signals. The processed signals are depicted in Fig. 3(b).



(a) Original signals with spikes



(b) Processed signals

Fig. 3: Signal processing using the Hampel filter-based method (t = 3, K = 2, vibration denoted as vib).

3.3 Feature extraction

Feature extraction is an effective means to reduce the complexity of the prediction process. In this research, features of signals from the three types of heterogenous sensors are defined and extracted in the time domain and frequency domains. Some considerations of the feature extraction are below:

A feature in the time domain can be either dimensional or dimensionless. A dimensional feature refers to one with a measurement unit, such as the means of signals. A dimensionless feature denotes a product or a ratio of dimensional features. For instance, the crest factor, which is the ratio of the peak to the root mean square (RMS) of signals to detect whether there is an obvious peak in the signals, is a dimensionless feature. A dimensional feature could have an intuitive physical meaning, and it can reveal the inherent attributes of the signals. However, under complex machining situations, these dimensional features may be influenced by various working loads and machining factors. In contrast, dimensionless features could be more stable in representing tool wear conditions from certain perspectives, and they can be complementary to dimensional features. Hence, in this research, dimensional and dimensionless features in the time domain are extracted to minimise the negative influence of dynamic factors throughout machining.

Due to the complexity of machining conditions and working environments, acquired sensor signals, especially vibration and AE signals, may contain various noisy information generated during a machining process. Even after the de-noising process on signals using the Hampel filter-based method, it could be still difficult to extract all the essential information (i.e., features) in the time domain. Instead, features in the frequency domain can be complimentary to those features in the time domain to provide more perspectives for comprehensive analysis. Therefore, features from the signals of the current, vibration and AE sensors are extracted in both the time and frequency domains. The diagram of the signal feature extraction is shown in Fig. 4.

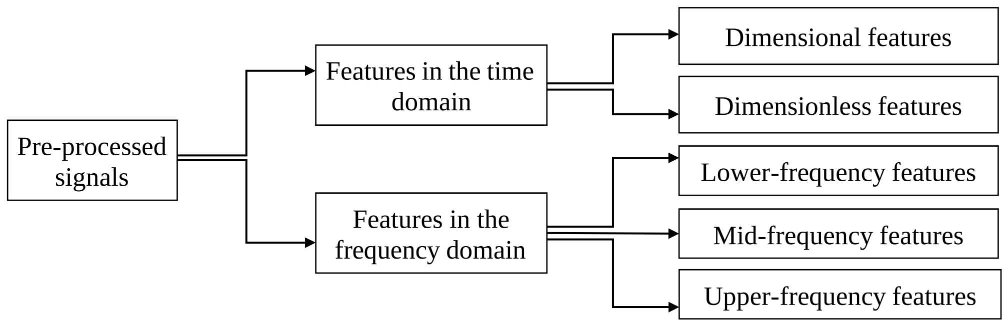


Fig. 4: The process of signal feature extraction.

<u>Feature extraction in the time domain</u>

The time domain refers to the change of the signal amplitude along time. Signals used in this research are collected in the time domain. Taking some runs of Setup 10 as examples, the signals of the current, AE and vibration in the time domain are shown in Fig. 5.

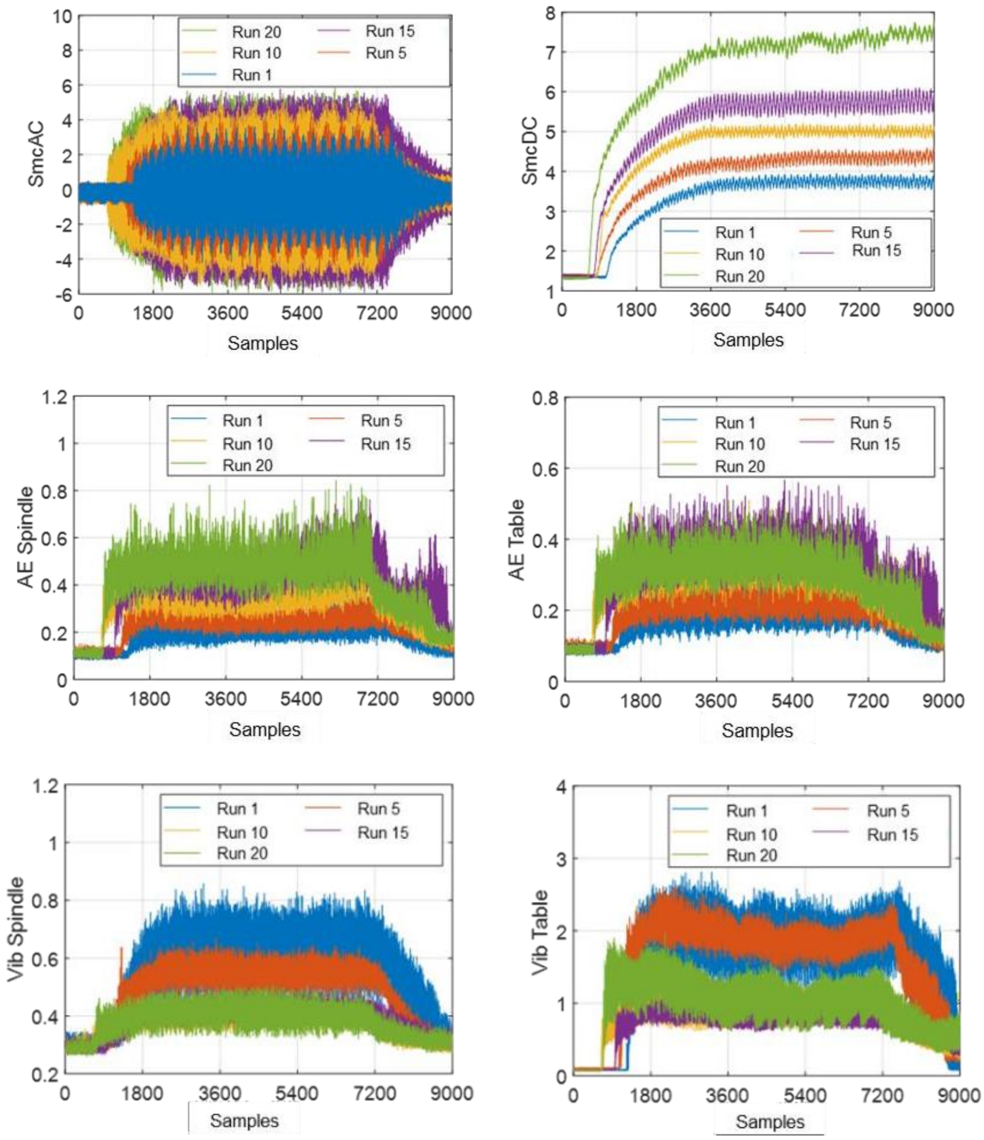


Fig. 5: Signals of some runs of Setup 10 in the time domain.

It can be observed that, with the increased number of cuts, the amplitudes of the signals of current and AE rise. However, the vibration signals do not follow the same trend. The tool gradually becomes dull from the initial sharpness, and the contact area between the tool and the workpiece is increased during the process. The chatter of the cutting tool gets smaller, and thereby, the vibration amplitude drops [47]. A similar conclusion from the perspective of materials and cutting mechanisms was drawn. As machining proceeds, the processing temperature and plastic deformation of the workpiece increase, resulting in reduced cutting forces that will eventually reduce the amplitude of the vibration signals. Meanwhile, plastic deformation is difficult to be detected by vibration sensors in some cases, and the AE sensor is a better option. This is also deemed to be a reason that signals from multiple sensors are indispensable to reflect the overall tool wear trend in a more comprehensive means. Signals for the rest of the cases exhibit a similar trend as presented above.

In this research, eleven dimensional features and six dimensionless features in the time domain are extracted from the signals of each sensor, respectively (Table 7). In total, there are 102 features in the time domain for all the sensor signals (i.e., 17 features for each of the six sensors respectively).

Table 7: Features in the time domain

<table>
<tr><th colspan="4">Time domain</th></tr>
<tr><th colspan="2">Dimensional feature</th><th colspan="2">Dimensionless feature</th></tr>
<tr><th>Feature</th><th>Formula</th><th>Feature</th><th>Formula</th></tr>
<tr><td>Mean</td><td>$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$</td><td>Impulse factor</td><td>$T_{if} = \frac{|T_{max}|}{\mu}$</td></tr>
<tr><td>Max</td><td>$T_{max} = \max(x_i)$</td><td rowspan="2">Kurtosis factor</td><td rowspan="2">$T_{kf} = \frac{T_{kur}}{{T_{rms}}^4}$</td></tr>
<tr><td>Min</td><td>$T_{min} = \min(x_i)$</td></tr>
<tr><td>Standard deviation</td><td>$T_{std} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2}$</td><td rowspan="2">Margin factor</td><td rowspan="2">$T_{if} = \frac{|T_{max}|}{T_r}$</td></tr>
<tr><td>Peak to peak</td><td>$T_{peak} = |T_{max}| - |T_{min}|$</td></tr>
<tr><td>RMS</td><td>$T_{rms} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} {x_i}^2}$</td><td rowspan="2">Shape factor</td><td rowspan="2">$T_{sf} = \frac{T_{rms}}{\mu}$</td></tr>
<tr><td>Skewness</td><td>$T_{ske} = \frac{\sum_{i=1}^{n}(x_i - \mu)^3}{(n-1){T_{std}}^3}$</td></tr>
<tr><td>Kurtosis</td><td>$T_{kur} = \frac{\sum_{i=1}^{n}(x_i - \mu)^4}{(n-1){T_{std}}^4}$</td><td rowspan="2">Crest factor</td><td rowspan="2">$T_{cf} = \frac{|T_{max}|}{T_{rms}}$</td></tr>
<tr><td>Mean absolute deviation</td><td>$T_{mad} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)$</td></tr>
<tr><td>Median</td><td>$T_{med} = \begin{cases} X[\frac{n+1}{2}], \mathit{if\ n\ is\ odd} \\ \frac{X[\frac{n}{2}] + X[\frac{n}{2}+1]}{2}, \mathit{if\ no\ is\ even} \end{cases}$
$X\ \mathit{is\ the\ ordered\ list\ of\ dataset}$</td><td rowspan="2">Skewness factor</td><td rowspan="2">$T_{skef} = \frac{T_{kur}}{{T_{rms}}^3}$</td></tr>
<tr><td>Variance</td><td>$T_{var} = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}$</td></tr>
</table>

In the above table, $T_r = (\frac{1}{n}\sum_{i=1}^{n}\sqrt{|x_i|})^2$ is the root value;; x_i is a signal sample; i=1, 2, ..., n; n is the total sample number of signals.

In the following subsections, the features in the time domain are named as the acronym of the features appended with the acronym of the signals. For instance, mean_AE_S stands for the mean feature of the AE signals obtained from the spindle of the machining system.

Frequency domain processing

Various uncertain factors may distort features in the time domain even after the Hampel filter-based de-noising process. To improve the prediction accuracy of tool wear, features are also extracted in the frequency domain using the power spectral density (PSD)-based method. PSD is a Fourier transformation and describes the powers of signals at different frequencies [48]. The advantages of using PSD to analyse the frequency domain include: (i) the irregularity of signals by the wavelength and amplitude can be displayed; (ii) frequency distributions hidden in signals with noises can be revealed; (iii) signal powers caused by random changes can be minimised; (iv) PSD can help reduce spectrum leakage, avoiding signals being not taken into account or signals being mistakenly considered as a repeated period.

PSD can be obtained below:

$$F(f) = \sum_{n=0}^{N-1} x_n e^{-i2\pi fn/N} \tag{5}$$

$$P(f) = \lim_{T\to\infty} \frac{|F(f)|^2}{T} \tag{6}$$

where x_n is the n^{th} sample of the time-series signal; $n = \{0, \cdots, N-1\}$; i is the imaginary unit and $i = \sqrt{-1}$; $\overline{T}$ is the time period; f is the spatial frequency; $F(f)$ is the Fast Fourier transformation; $P(f)$ is PSD.

Abnormal behaviours of an object usually occur close to the natural frequency of the object. For a cutting tool, it will be more effective to capture features near the natural frequency of the tool for tool wear prediction. That is, features related to tool wear could be found in the frequency bands with the larger amplitude of PSD. Thus, the sensor signals are converted to the frequency domain using PSD for analysis.

The effect of using PSD can be illustrated in Fig. 6, where PSD for the AE_S signals for Setup 11 is displayed. Fig. 6 is composed of three axes, i.e., frequency, tool wear and PSD. It shows that dominant frequencies appear in the 0-40 Hz and 40-80 Hz frequency bands. Moreover, Fig. 6 presents that the amplitude of each frequency band increases with the process of tool wear exacerbation. In particular, the PSD increment around 80Hz is more intensive as the tool deterioration than that in the 0-40 Hz frequency band. It may attribute to the fact that the frequency of 80 Hz embodies the natural frequency of the cutting tool. PSD reflects the increasingly severe friction between the tool and workpiece.

Based on the above results, changes in tool wear conditions can be characterised by features extracted from a specific frequency band in the spectrum, rather than from the entire bands [49]. Thus, the frequencies of sensor signals are divided into a lower-frequency band (0 Hz - 40 Hz), a middle-frequency band (40 Hz - 90 Hz) and an upper frequency band (90 Hz -125 Hz) in order to choose a suitable band to reveal the critical features for tool wear conditions.

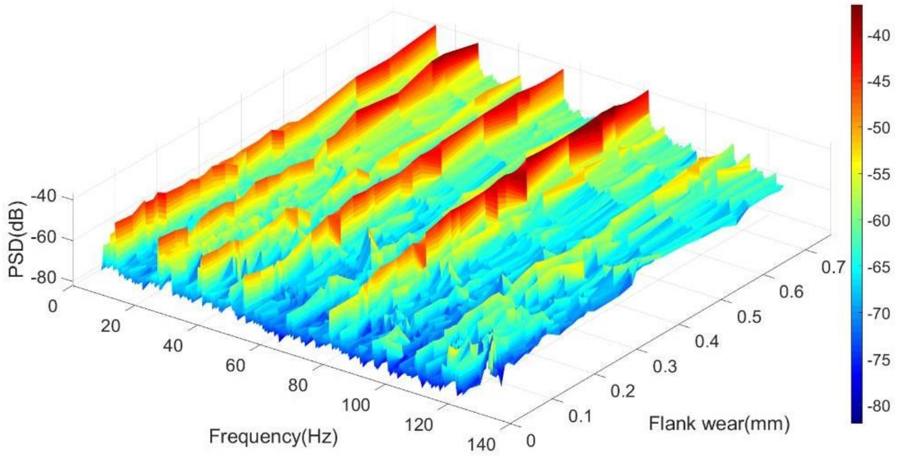


Fig. 6: PSD of the AE_S signals for Setup 11.

In this research, six features for each frequency band are extracted from the signals of each sensor. In total, 108 features are extracted for the signals from the six sensors at the three frequency bands. The related information is shown in Table 8.

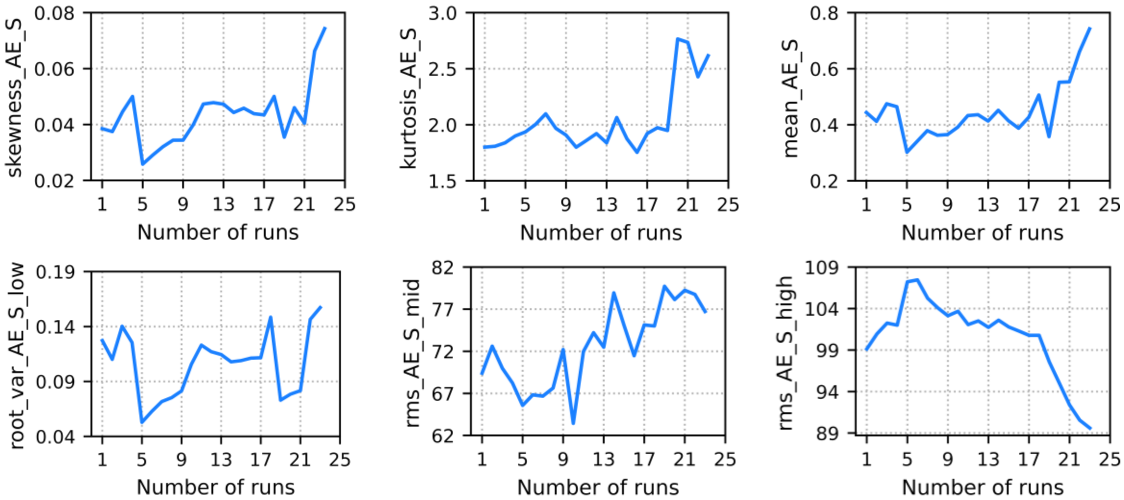
Table 8: Features under the frequency domain.

Frequency domain	
Feature	Formula
Mean	$f_m = \frac{1}{n}\sum_{i=1}^{n} p_i$
Max	$f_{max} = \max(p_i)$
Min	$f_{min} = \min(p_i)$
Root mean square	$f_{rms} = \sqrt{\frac{\sum_{i=1}^{n} f_i^2 p_i}{\sum_{i=1}^{n} p_i}}$

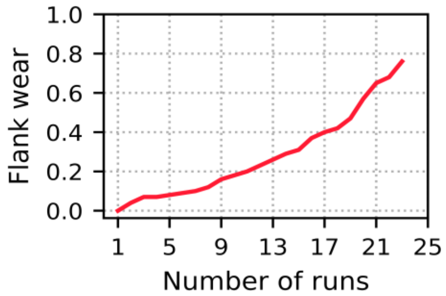
Frequency centre	$f_{fc} = \frac{\sum_{i=1}^{n} f_i p_i}{\sum_{i=1}^{n} p_i}$
Root variance frequency	$f_{rv} = \sqrt{\frac{\sum_{i=1}^{n} (f_i - f_m)^2 p_i}{\sum_{i=1}^{n} p_i}}$

In the table, f_i is the frequency value with i=1, 2, …, n denotes the spectrum lines, p_i is the power spectrum density at f_i.

In the following subsections, the features in the frequency domain are named as the acronym of the features appended with the acronym of the signals and the frequency band. For instance, rms_AE_S_low stands for the rms feature of the AE signals obtained from the spindle of the machining system at the lower frequency band.



(a) Extracted features of the AE_S signals



(b) Trend of tool wear

Fig. 7: Extracted features vs. the trend of tool wear.

By performing feature extraction in the time and frequency domains, a total of 210 features are obtained for the signals of vibration, AE and current. Taking the features of the AE_S signals for Setup 10 as an example, some extracted features and measurement values of tool wear are shown in Fig. 7. It can be observed that some extracted features, such as mean and kurtosis in the time domain and RMS at the middle-frequency band are more consistent with the trend of tool wear, while other features do not follow this trend. It means that features deviated from the trend of tool wear need to be identified and removed to improve the prediction accuracy of tool wear.

Furthermore, Pearson correlation coefficient (PCC) is used to analyse the relationship between extracted features. PCC is a simple and effective statistical method to evaluate the relationship between variables based on the covariance matrix [50], and it has been widely employed in some research related to the feature selection [51-52]. To optimise the computational efficiency and alleviate over-fitting for tool wear prediction, the correlation between selected features should be small. PCC is a value between [-1, 1] and it can be calculated below:

$$P_{a,b} = \frac{Cov(a,b)}{\sigma_a \sigma_b} = \frac{\sum_{i=1}^{n}(a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_{i=1}^{n}(a_i - \bar{a})^2} \cdot \sqrt{\sum_{i=1}^{n}(b_i - \bar{b})^2}} \quad (7)$$

where, $P_{a,b}$ denotes PCC of two features a and b; $Cov(a,b)$ denotes the covariance of the features; σ_a and σ_b denote the standard deviations of the features respectively; a_i and b_i denote the samples indexed with i in individual features respectively; $\bar{a}$ and $\bar{b}$ denote the mean values of individual features respectively.

According to PCC, the association between two features presents the positive or negative state. A greater value means a high correlation. The correlation matrix based on PCC is shown in Fig. 8.

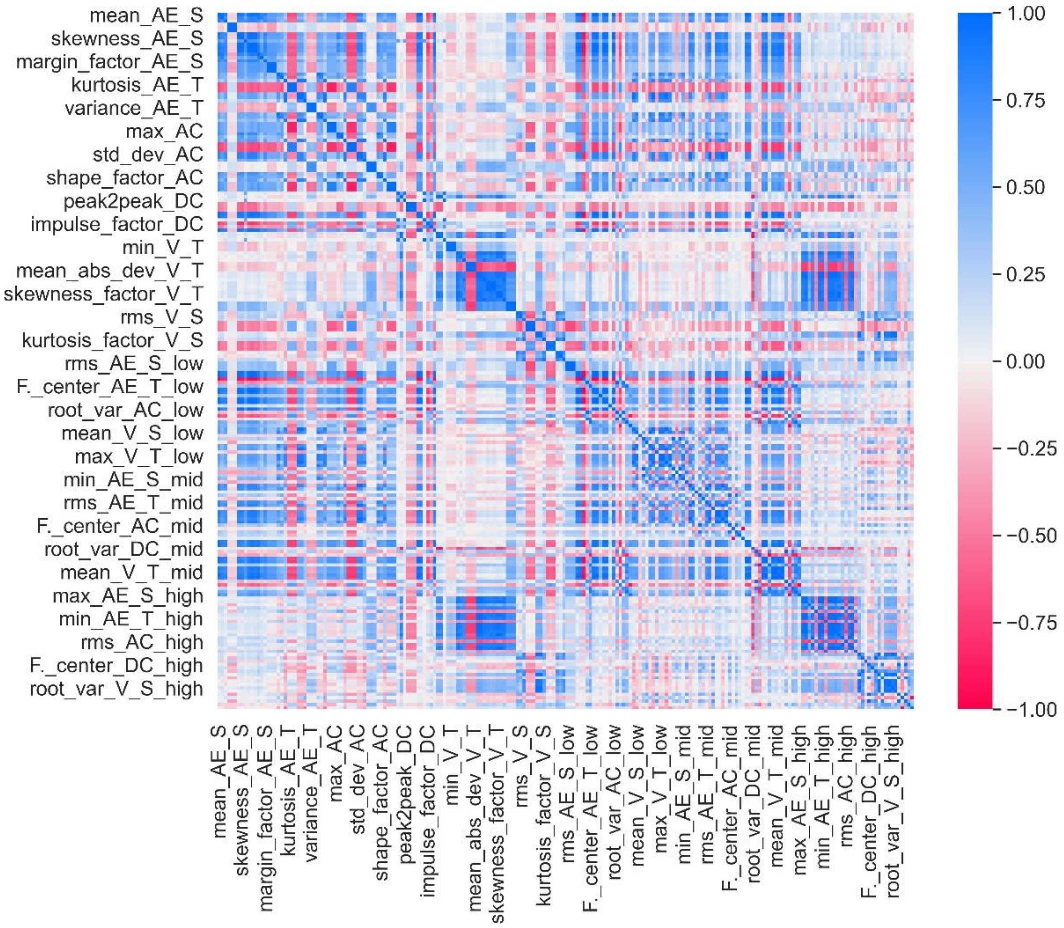


Fig. 8: The correlation matrix of extracted features.

Fig. 8 expresses the PCC of the extracted 210 features (102 features in the time domain and 108 features in the frequency domain) by the shade of the colour. Based on the analysis, it is clearly shown that it is necessary to eliminate redundant or less relevant features in order to obtain more efficient results for tool wear prediction.

3.4 Feature optimisation based on REFCV

For the extracted features, some could be less important (irrelevant) to tool wear prediction. It is worth identifying and removing those features to maintain the highest prediction accuracy and computational efficiency.

In this research, a RFECV-based method is designed to select optimal features from the initially extracted features recursively. The flowchart of the method is shown in Fig. 9. Some major steps of the method are depicted below:

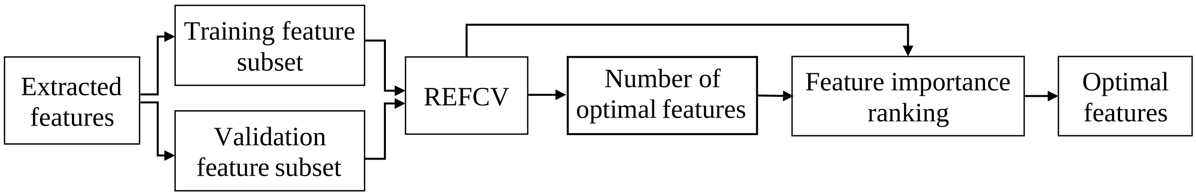


Fig. 9: The flowchart of the RFECV-based method for feature optimisation.

1. For all types of sensors, the set of features extracted in both time and frequency domain under the i^{th} run is modelled as a vector f_i ($i = 1,2, \dots, 164$, 164 is the total number of runs), and each feature in the set is denoted as $f_i(j)$ (j =1, 2, …, m, m is the total number of features for all sensor signal).

2. In view of the fact that the feature value of a vector could be in different ranges, a normalisation process is performed based on the Nadir and Utopia points. The Utopia point z^U and Nadir point z^N provide the lower bound and upper bound of the value of features, respectively. The normalisation process of each feature in a vector f_i is as below (the normalised feature is f'_i):

$$f'_i(j) = (f_i(j) - z^U)/(z^N - z^U) \tag{8}$$

3. The set of f_i is segmented into a training subset and a validation subset randomly according to an approximate 7/3 ratio for M times. Each randomly generated group is denoted as G_k(k=1, …, M) with the indices of runs denoted as $l = (k_1, \dots, k_L)$, L is the total number of runs in each training dataset. Fig. 10 shows a schematic diagram of this step.

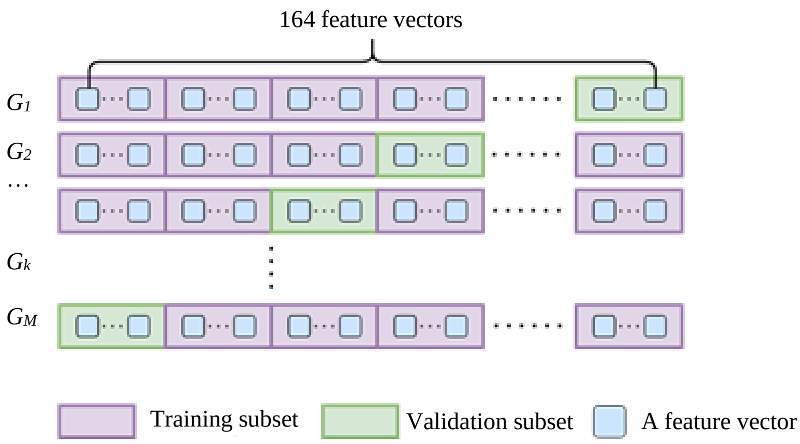


Fig. 10: A schematic illustration for Step 2.

4. For each group G_k, an SVM classifier is used to conduct a binary classification process for tool wear (unworn or worn) based on the training and validation subsets of feature vectors according to the following procedures.

In the process, the SVM classifier distinguishes the features for the unworn and worn status of a tool in the training subset through a hyperplane, the training stops until an optimal hyperplane is achieved. The maximum distance between features and the optimal hyperplane implies the lowest classification error [53]. In general, a hyperplane could be defined below:

$$y_k = \sum_{l=1}^{L} \sum_{j=1}^{m} w_{k_l,j} \cdot f'_{k_l}(j) + b \tag{9}$$

where y_k represents the state of tool wear for the G_k group ($y_{,k} = 1$ means the tool is worn, and $y_k = -1$ means the tool is unworn); f'_{k_l} is the normalised feature vector with an index of k_l in the training set of G_k group; $w_{k_l,j}$ denotes a weight for the feature j in f'_{k_l}; b is the bias.

Based on the dataset in the case study, the SVM classifier is trained using the training subset to obtain w_j. After the validation of the SVM classifier, the classification accuracy is assessed by a confusion matrix. There are four possible results generated, i.e., true positives (*TP*), true negatives (*TN*), false positives (*FP*) and false negatives (*FN*). These four outcomes can be defined below:

- *TP*: The feature indicates the tool is worn, and the tool is actually worn;
- *TN*: The feature indicates the tool is unworn, and the tool is actually unworn;
- *FP*: The feature indicates the tool is worn, and the tool is actually unworn;
- *FN*: The feature indicates the tool is unworn, and the tool is actually worn.

The classification accuracy of the validation subsets can be assessed using Equation (10), which represents the proportion of the correct classification in the total classification.

$$\text{Accuracy}(y_k) = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

5. After each group of G_k (k=1, …, M) is processed according to the above steps, the average classification accuracy for each feature is calculated, i.e., $\frac{1}{M}\sum_{k=1}^{M}(\mathrm{Accuracy}(y_k)$. Moreover, the importance ranking criterion of features can be calculated as $C_j = \sum_{l=1}^{L} w_{kl,j}^2$ (C_j denotes the importance score of the feature j in a vector). Then, the importance of each feature in the group can be ranked.

6. The feature with the least importance is eliminated, so that the number of features in the vector becomes m-1. The above steps are recursively repeated until all features are processed. During the process, the sets of features are denoted as S_m, S_{m-1}, …, S_1.

7. For the set of S_m, S_{m-1}, …, S_1, their average classification accuracies are ranked. The number of features in the set with the highest classification accuracy is chosen as the optimal number of features. According to the importance score and the optimal number of the features, optimal features are selected.

To illustrate the REFCV-based method, the classification accuracy of all extracted features is shown in Fig. 11.

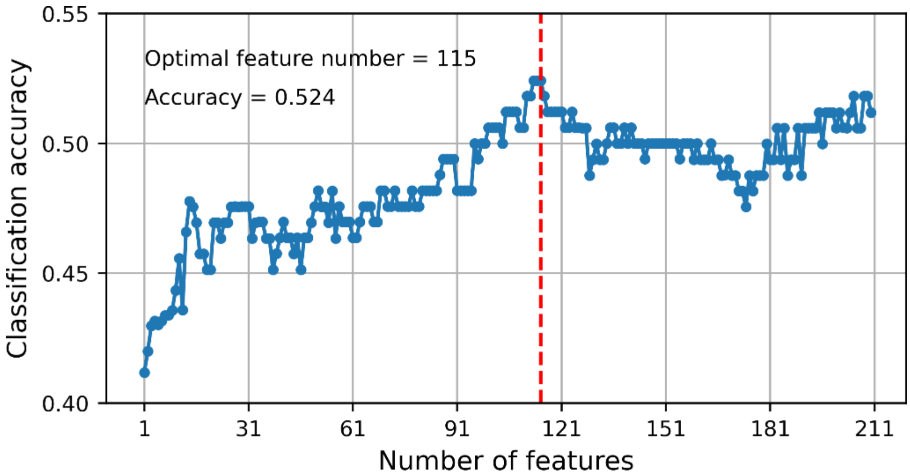


Fig. 11: The classification accuracy of features.

According to Fig. 11, the highest accuracy for all signal is achieved with 115 features in both the time and frequency domain.

Furthermore, the contribution degree (score) of each feature to tool wear is shown in Fig. 12. Based on the classification accuracy result of different feature number and the importance ranking of each feature, the most valuable features corresponding to the optimal number can be selected. After the application of the RFECV-based method, the optimal results are 44 features in the time domain and 71 feathers in the frequency domain (in total 115 optimal features). The results are shown in Table 9 and Table 10.

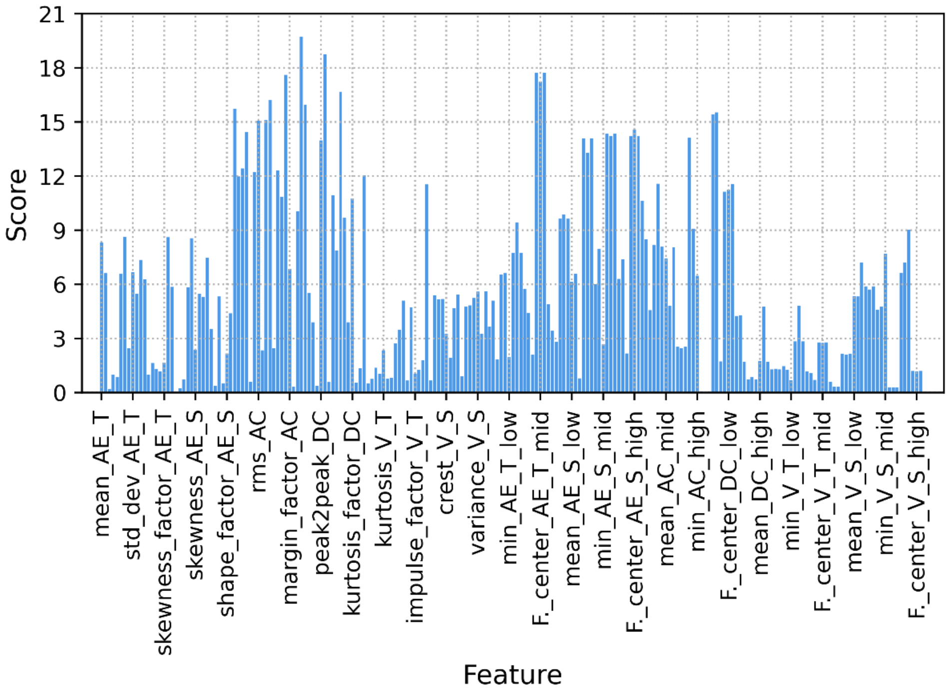


Fig. 12: The importance ranking of all extracted features.

Table 9: Optimal features in the time domain.

Time domain		
Signal	Feature number	Feature
AE_T	6	RMS, Mean, Median, Standard deviation, Maximum, Variance
AE_S	5	Max, Mean, RMS, Median, Peak to peak
AC	10	Kurtosis factor, Impulse factor, Mean, Variance, Standard deviation, RMS, Crest factor, Minimum, Peak to peak, Mean absolute deviation
DC	10	Kurtosis factor, Skewness factor, Mean, Variance, Standard deviation, RMS, Median, Maximum, Minimum, Peak to peak
V_T	7	Shape factor, Kurtosis, Variance, Standard deviation, Skewness, Mean absolute deviation, Margin factor
V_S	6	Kurtosis factor, Variance, RMS, Mean, Median, Shape factor

Table 10: Optimal features under the frequency domain.

Frequency domain						
	Lower band		Mid band		Upper band	
Signals	Feature number	Feature	Feature number	Feature	Feature number	Feature
AE_T	2	Max. Frequency centre	3	RMS, Mean Root variance	4	Max., Root variance, Frequency centre, RMS

AE_S	3	RMS, Min. Frequency centre	3	Root variance, Frequency centre, RMS	6	RMS, Mean, Max., Root variance, Min., Frequency centre
AC	4	RMS, Mean, Max., Frequency centre	4	RMS, Mean, Max., Min.	6	RMS, Mean, Max., Root variance, Min., Frequency centre
DC	4	RMS, Mean, Max., Frequency centre	5	RMS, Mean, Max., Frequency centre, Root variance	4	RMS, Root variance, Max., Frequency centre
V_T	3	RMS, Root variance, Frequency centre	3	Root variance, Frequency centre, RMS	3	RMS, Root variance, Frequency centre
V_S	5	RMS, Mean, Max., Root variance, Frequency centre	4	Mean, Max., Root variance, Min.	5	RMS, Mean, Max., Root variance, Frequency centre

After the above process, PCC on optimal features is evaluated again. Fig. 13 displays the correlation matrix of the 115 features. The results show that, compared with Fig. 8, the correlation of features is reduced, i.e., the average correlation now is 0.23, which is much smaller than 0.47 in Fig. 8. Nevertheless, features from different sensor signals are prone to contain duplicated information, so that there could be still correlations between some features that are not eliminated in the above optimisation process. To further improve the performance of tool wear prediction, features can be further optimised by an Isomap-based method described in the following subsection.

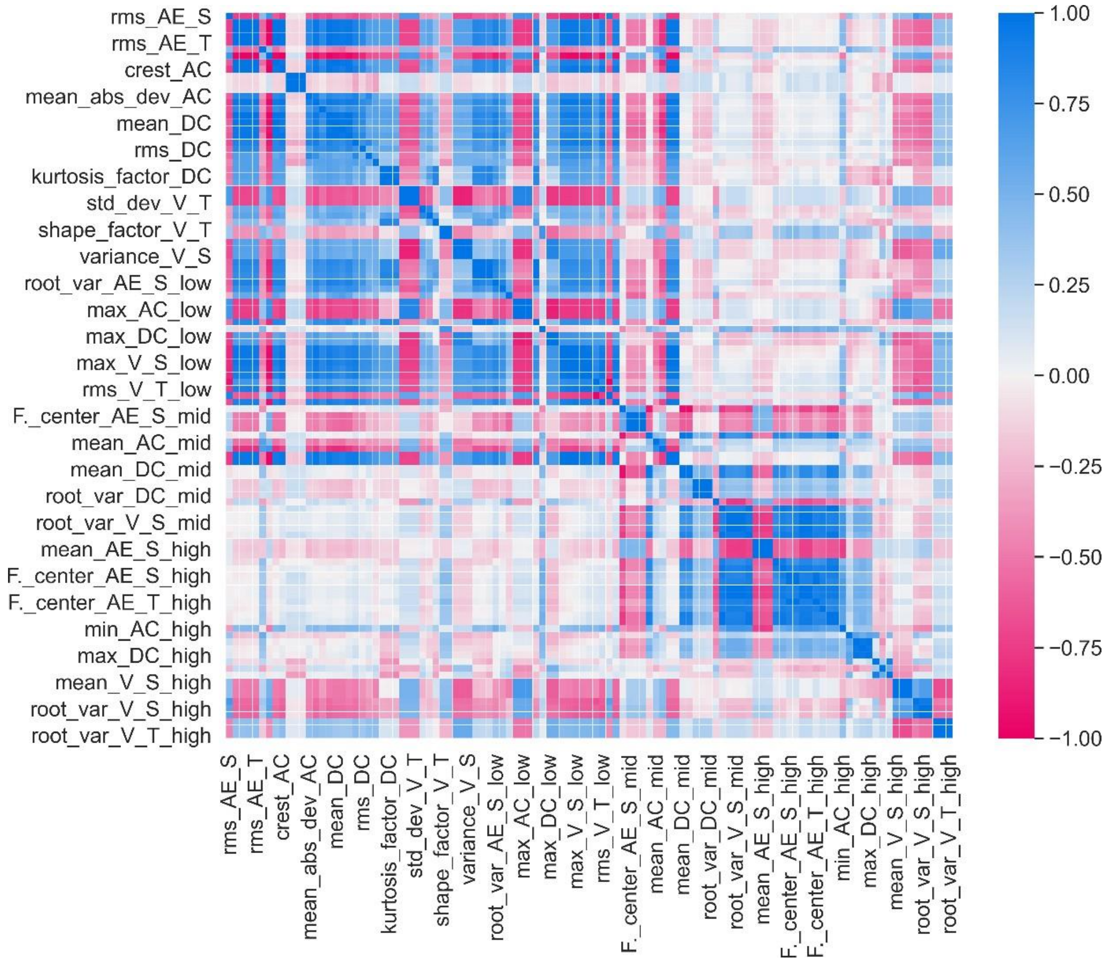
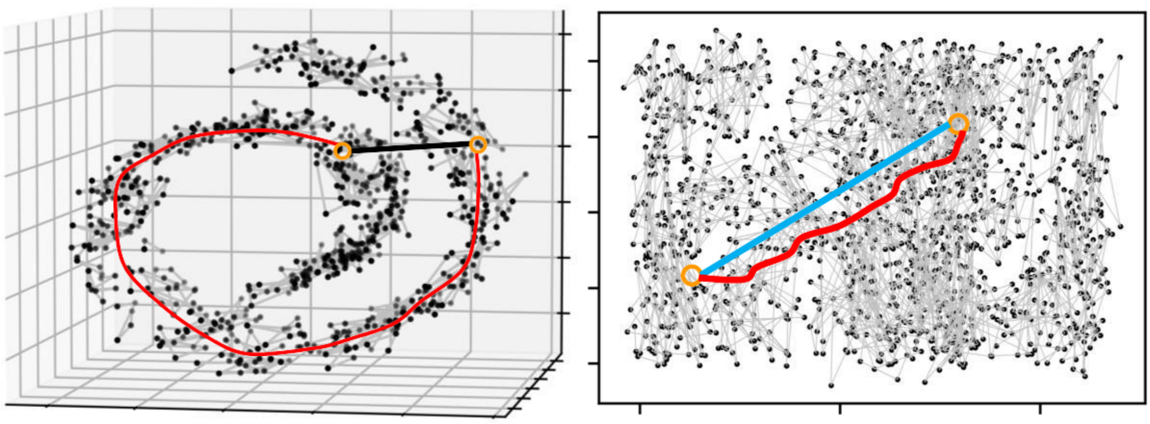


Fig. 13: The correlation matrix of optimal features.

3.5 Feature optimisation based on Isomap

In this subsection, selected optimal features will be further condensed to improve the computational efficiency and accuracy of tool wear prediction. The features can be mapped into a multi-dimensional space for analysis. Generally, the principal component analysis (PCA) is one of the most prevalent methods used for data dimensionality reduction of a multi-dimensional space. However, PCA is not suitable for processing data with non-linear correlations [54]. Isomap was introduced which can measure the scale of a multi-dimensional space with non-linear data effectively to enhance computing efficiency and maintain data accuracy [55]. As shown in Fig. 14(a), the Euclidean distance between two points (black line) in the high dimensional space (A 3D example is used for visualisation purpose) cannot represent the actual distance along the manifold (geodesic distance) which is indicated in red line. By constructing a neighbourhood graph based on the high-dimensional data then mapping it to a low-dimensional space using Isomap, as illustrated in In Fig. 14(b), the geodesic distance (red line) and the Euclidean distance (blue line) in the lower dimensional space are close, so the Euclidean distance can be used as an alternative and the conventional data dimensionality reduction methods can be applied to further optimise the feature selection.



(a) The neighbourhood graph in the high dimensional space (b) Lower dimensional embedding

Fig. 14: The illustration of a neighbourhood graph and 2D embedding based on Isomap.

1. The distance matrix $D = [d_{ij}]_{M\times M}$ is calculated, where the shortest distance between a feature pair $d_{ij} = \min\{d(f_i, f_j)\}$ is obtained by the Dijkstra's algorithm and Floyd–Warshall algorithm [56], and d is the geodesic distance between two features;

2. According to the principle of the multi-dimensional scaling (MDS), the mapped coordinate matrix Z in a lower-dimension space can be derived by eigenvalue decomposition from the inner product matrix B that is computed by applying the central matrix to the squared matrix of D:

$$B = [b_{ij}]_{M\times M} \quad b_{ij} = -\frac{1}{2}\left(d_{ij}^2 - \frac{1}{M}\sum_{i=1}^{M} d_{ij}^2 - \frac{1}{M}\sum_{j=1}^{M} d_{ij}^2 - \frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M} d_{ij}^2\right) \tag{11}$$

In order to remain the information of the original feature set, the predefined number of the desired dimension of the lower dimensional space $N' = N - 1$. N denotes the dimensions of original high-dimensional space. Based on the N' largest eigenvalues and corresponding eigenvectors of B, the matrix Z can be obtained:

$$Z = [\mathrm{V}\Lambda^{1/2}]^T \tag{12}$$

where V is the matrix of N' eigenvectors of B; Λ is the diagonal matrix of the N' eigenvaluesof B.

The new representative components of features are generated as the columns in the matrix Z. As the dimension of the matrix Z is $M\times N'$, the number of the components equals to the number of original features, which is 110 in this paper.

The optimisation of feature selection is conducted based on the new representative components of features. The cumulative variance sequentially accumulates the variances of each component to evaluate the proportion of the original information contained in the features in different dimensions. As shown in Fig. 15, after the cumulative variance calculation, the first 10 components contain about 90% of the variance. That is, 90% of the raw information is preserved. When the number of components reaches 40, it can be used to describe nearly 100% of the information. Therefore, 40 components (features) are finally kept as the optimised features.

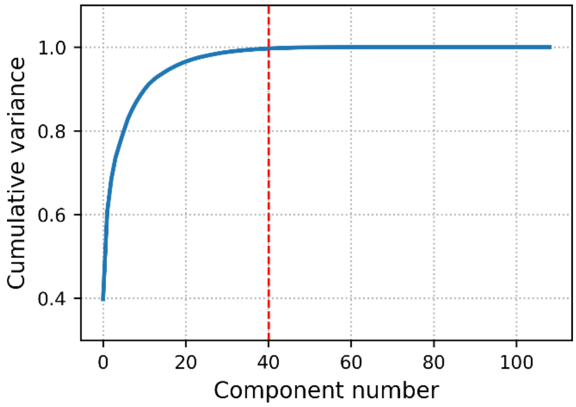


Fig. 15: The cumulative variance based on Isomap.

In addition, with the determined representative components, Fig. 16 shows the correlation matrix of these 40 components (features). Obviously, the correlations between these components are greatly reduced. The average correlation coefficient is now 0.025.

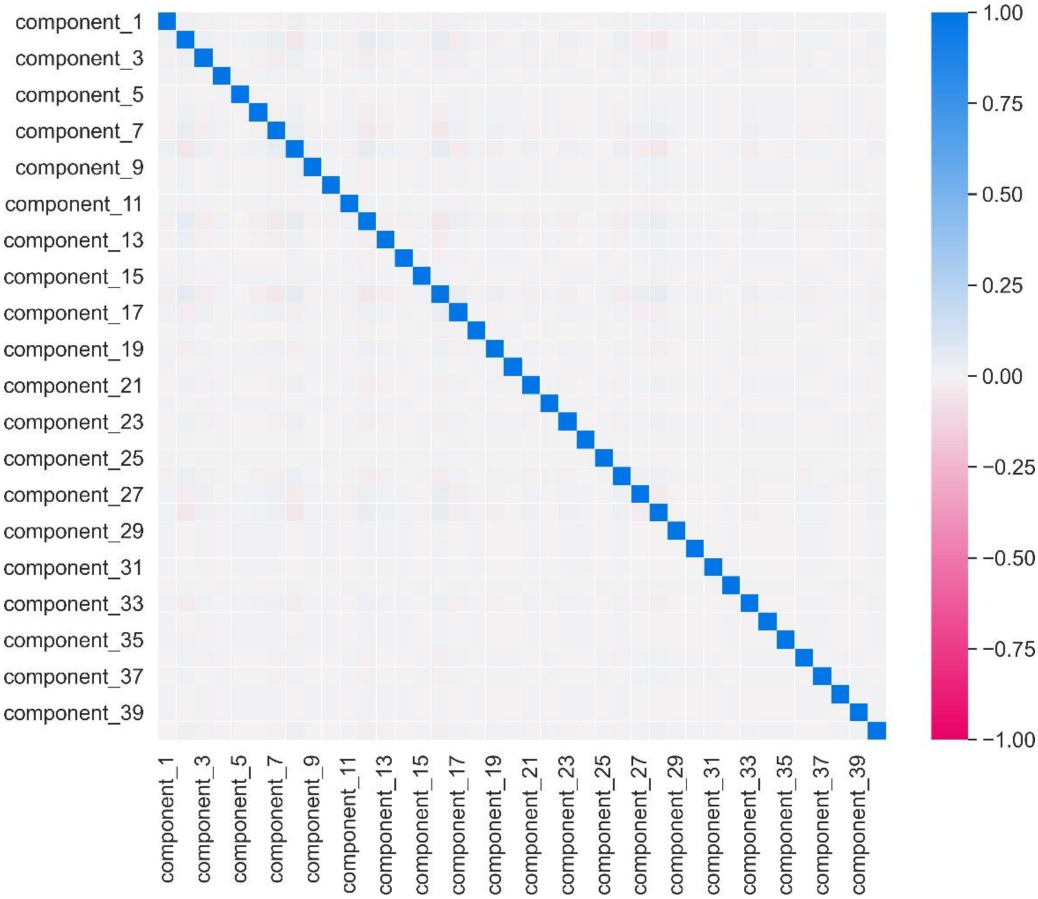


Fig. 16: The correlation matrix of Isomap-based components.

4. Method Performance Evaluation

4.1 Assessment based on different features sets

In this research, a CNN model is adopted for tool wear prediction. An optimised architecture and parameters of the CNN model for the tool wear prediction were determined through trials and error comparisons. The following four feature sets were used for the training, validation and comparative analysis of the designed CNN model:

- Feature set 1: the set of 210 features (originally extracted features),
- Feature set 2: the set of 115 features (after feature optimisation by RFECV),
- Feature set 3: the set of 40 features (after feature optimisation by Isomap),
- Feature set 4: the set of 40 features from Feature set 3 added by three machining parameters, which are depth of cut, feed and workpiece material (considering numerous studies showed that machining parameters have significant influences on tool conditions.

The CNN model is shown in Fig. 17 (using Feature Set 4 as an exemplary input). The model structure for four feature sets are summarised in Table 11. In this research, Keras and Tensorflow were utilised to establish the CNN model. The CNN model was executed on the Apache spark, which is a cluster-computing framework.

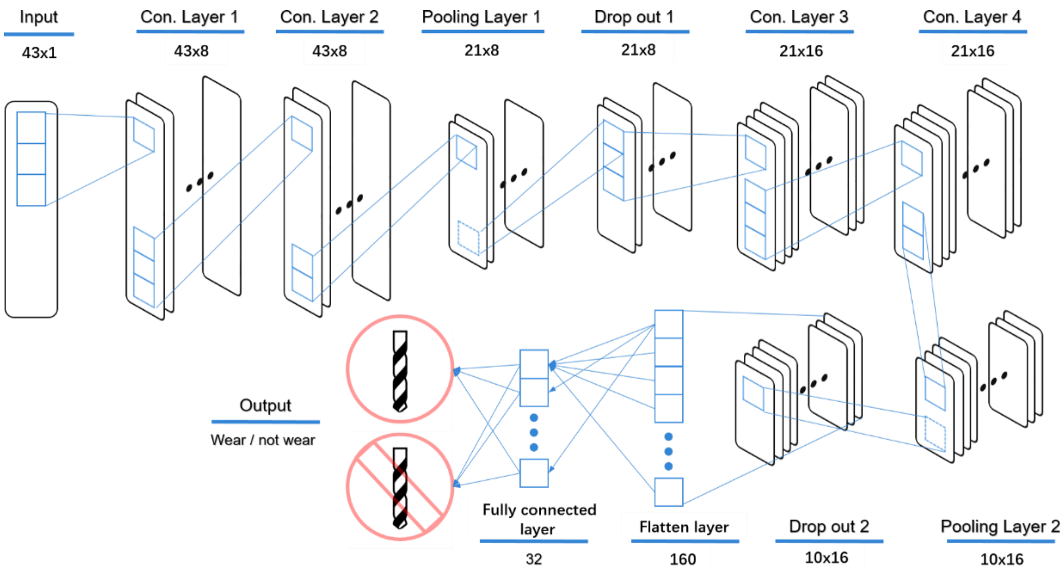


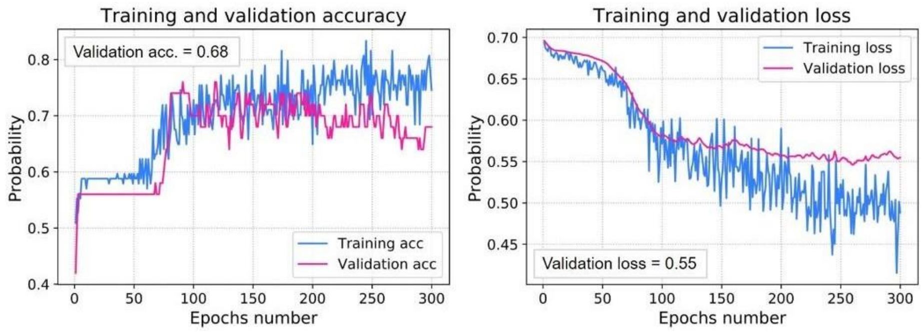
Fig. 17: The architecture of the CNN model.

Table 11: Parameter of the CNN model for different feature sets.

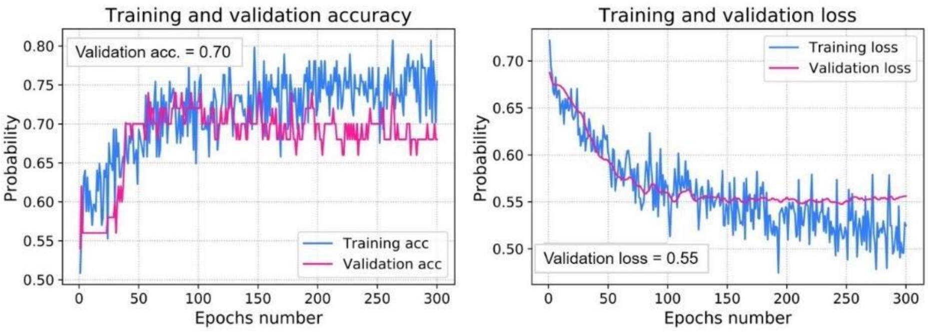
Layer	Set 1	Set 2	Set 3	Set 4
Input	210x1	115x1	40x1	43x1
Convolution layer 1	210x8	115x8	40x8	43x8
Convolution layer 2	210x8	115x8	40x8	43x8
Pooling layer 1	105x8	57x8	20x8	21x8
Drop out 1	105x8	57x8	20x8	21x8
Convolution layer 3	105x16	57x16	20x16	21x16

Convolution layer 4	105x16	57x16	20x16	21x16
Pooling layer 2	52x16	28x16	10x16	10x16
Drop out 2	52x16	28x16	10x16	10x16
Flatten layer	832	448	160	160
Fully connected layer	32	32	32	32

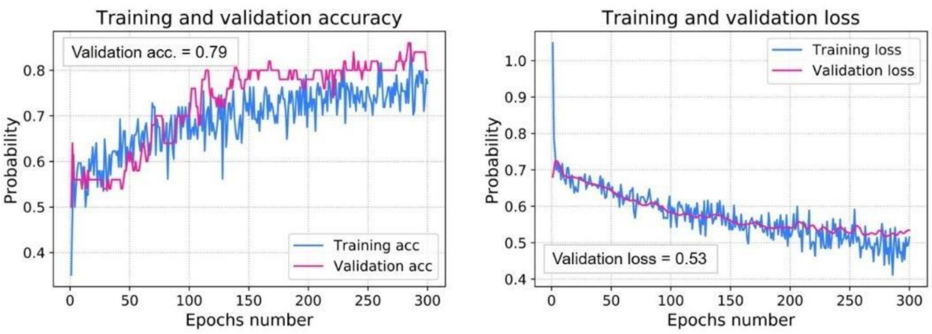
The results of the training and validation accuracy, training and validation loss for the four feature sets are displayed in Fig. 18. The validation accuracies of the four feature sets are summarised in Table 12. It shows that the validation accuracy increases along with the increase of the training accuracy until it reaches optimal values. However, the accuracy of feature sets 1 and 2 is relatively low. The reason is that, the accuracies of validation on new-in data is poor and overfitting occurs, as correlated features in the feature sets are retained. For Feature sets 3 and 4, correlated features in the originally extracted features are removed through RFECV and Isomap, so that overfitting issues are eliminated, and accuracies are improved significantly. Feature set 4 achieved the highest validation accuracy of 86%, which also proves that a feature set enhanced by machining parameters is an effective strategy.



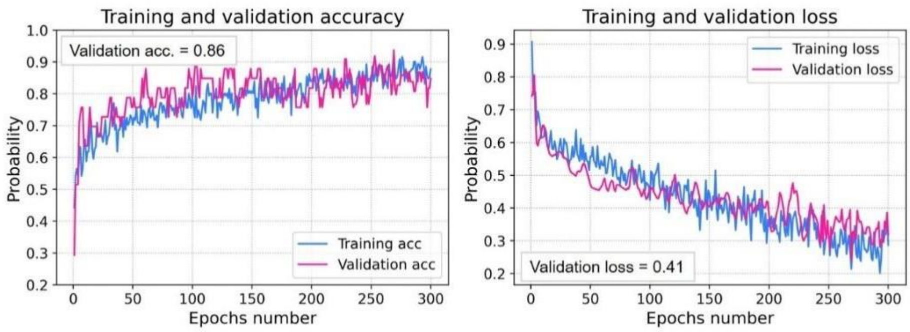
(a) Training and validation for Feature set (1)



(b) Training and validation for Feature set (2)



(c) Training and validation for Feature set (3)



(d) Training and validation for Feature set (4)

Fig. 18: The prediction accuracy and loss of the CNN model.

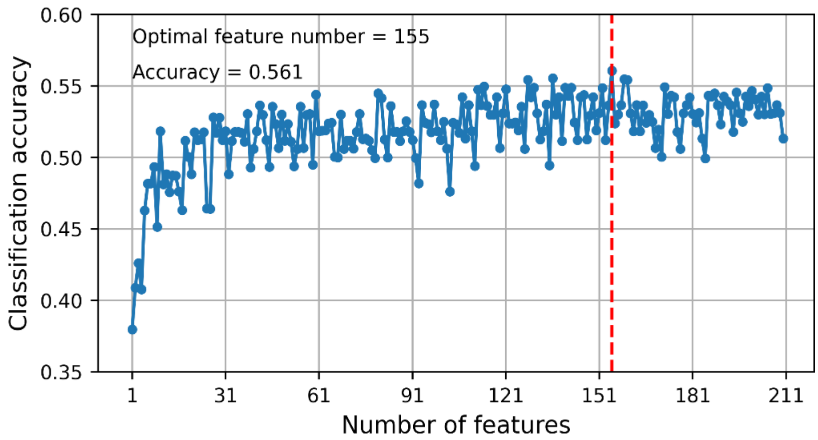
Table 12: The data sizes and validation accuracies of the different feature sets for the CNN model.

Input Data	Data size	Validation accuracy
Feature set 1 (Fig. 18(a), (b))	164×210	68%
Feature set 2 (Fig. 18(c), (d))	164×115	70%
Feature set 3 (Fig. 18(e), (f))	164×40	79%
Feature set 4 (Fig. 18(g), (h))	164×43	86%

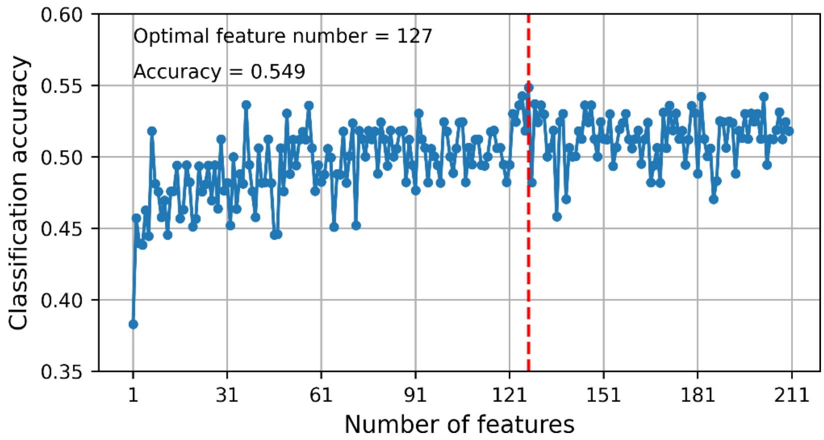
4.2 Comparison of RFECV with other methods for feature optimisation

To assess the performance of RFECV, comparative analyses with three prevalent machine learning algorithms, i.e., random forest (RF), extra tree (ET) and logistic regression (LR), were conducted. Both of RF and ET consist of a large number of decision trees, where the final decision is obtained taking into account the prediction of every tree. In RF, a bagging method is used to train sample data (features) and assign weights to each of them to establish a regression model [57]. ET adopts a method different from RF: it trains all the features on each decision tree, and then randomly divides these features at the nodes of a single decision tree to construct a complete classification model [58]. In addition, LR utilises the sigmoid function to estimate the classification probability between the feature and dependent variables to build a regression model [59].

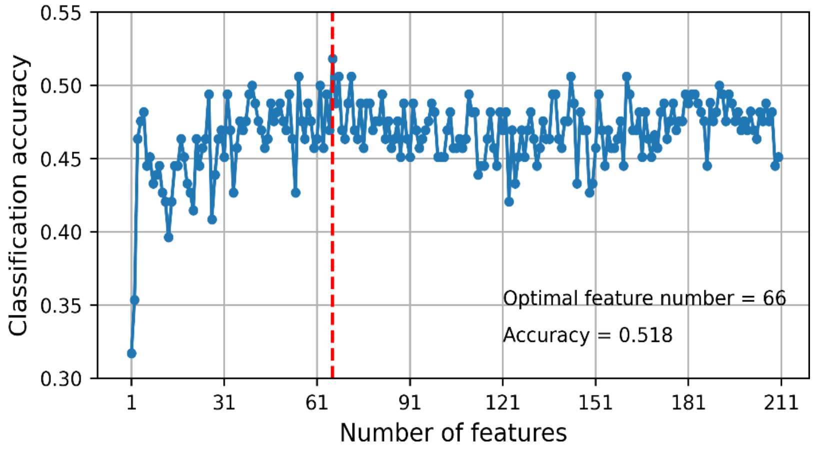
Based on the three algorithms, the optimal number of features, the importance of each feature and corresponding optimal features were determined successively from the originally extracted 210 features. The classification accuracy of the feature set in different sizes are shown in Fig. 19, the importance ranking of features are shown in Fig. 20. The results of optimised features using RF, ET and LR are summarised in Table 13. Finally, the number reductions of features are 26%, 40% and 68% by using RF, ET and LR respectively, in comparison with the 45% by using REFCV.



(a) RF model

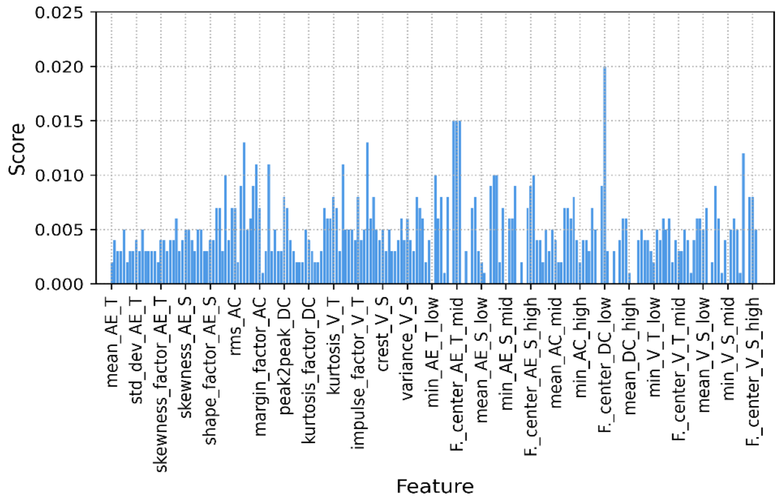


(b) ET model

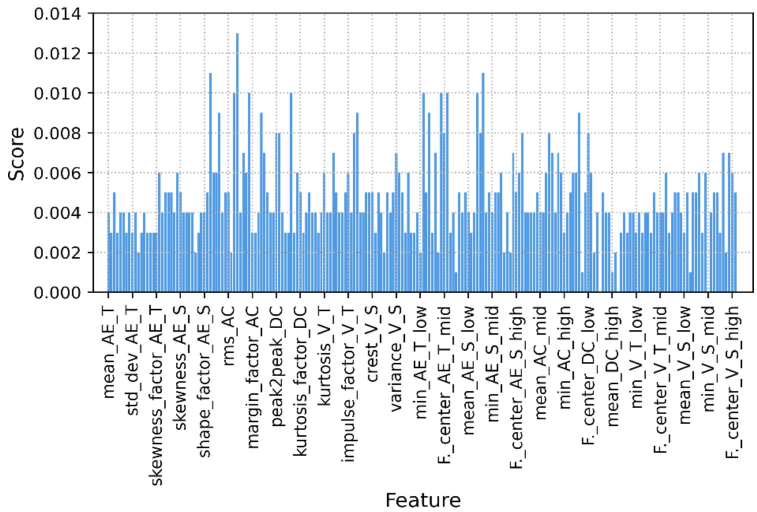


(c) LR model

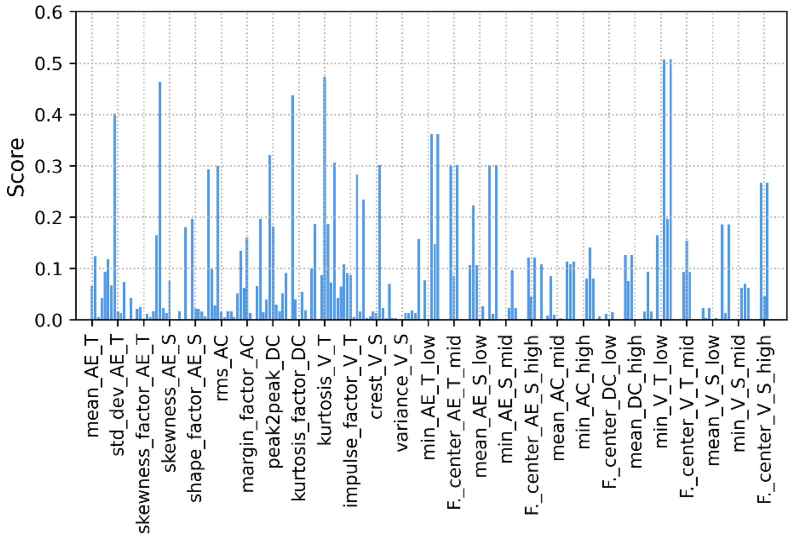
Fig. 19: The classification accuracies of all feature sets determined by different intelligent models.



(a) RF model



(b) ET model



(c) LR model

Fig. 20: Importance ranking of all features by different intelligent models.

Table 13: The number of optimal features by different intelligent models.

Signals	Random forest				Extra tree				Logistic regression			
	Time domain	Frequency domain			Time domain	Frequency domain			Time domain	Frequency domain		
		low	mid	high		low	mid	high		low	mid	high
AE_S	14	5	4	5	7	4	4	3	2	3	3	1
AE_T	12	5	4	4	8	3	3	5	5	2	2	1
AC	11	6	4	6	10	4	5	6	6	4	1	1
DC	10	4	5	4	11	4	5	4	7	4	1	1
V_S	12	5	4	5	9	5	4	5	6	1	3	1
V_T	11	4	6	5	8	4	3	3	6	1	3	1
Total	155				127				66			

Isomap-based optimisation on feature sets generated by RF, ET and LR is performed. Fig. 21 displays the cumulative variances of the Isomap-based results achieved by the three methods.

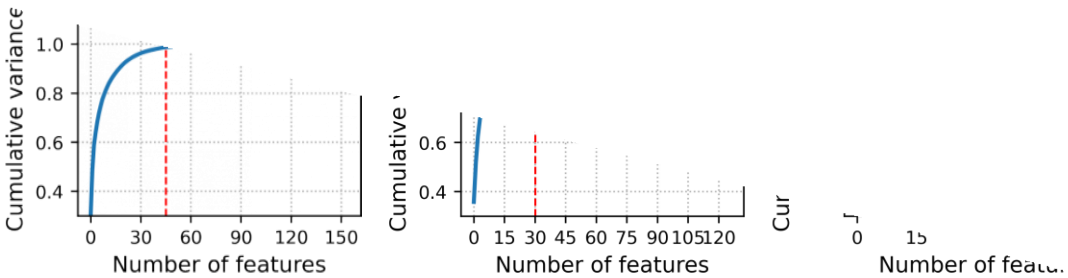
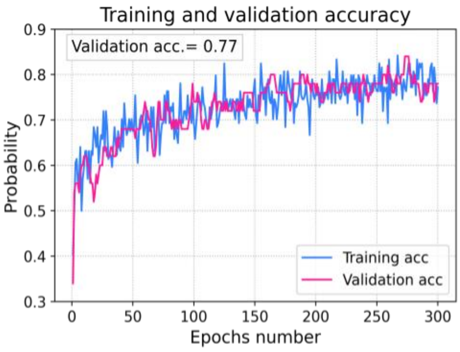
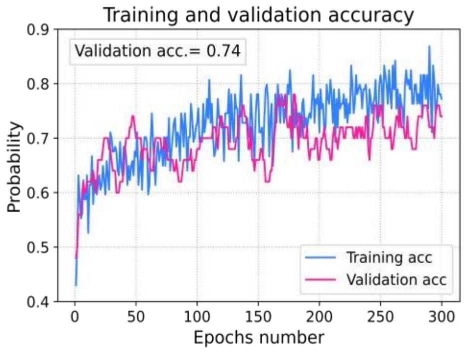


Fig. 21: The cumulative variances of Isomap for the features achieved by RF, ET and LR.

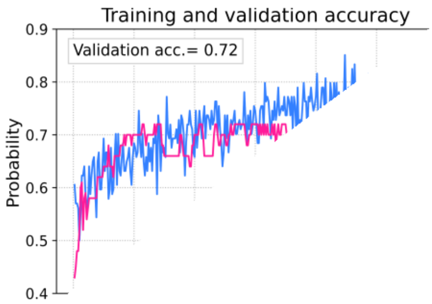
According to the cumulative variance, for RF, 45 features (representative components) can effectively represent the source information of the optimal features (the cumulative variance is close to 100%). To reach the same percentage, both ET and LR need 30 features. After the application of Isomap, on the basis of the extracted features (210), the feature reduction rates of using RF, ET and LR reach 78%, 85% and 85%, respectively, in comparison with 81% by REFCV. Fig. 22 shows the validation accuracy of the CNN model on the optimised feature sets achieved based on RF, ET, LR and REFCV. The validation accuracy of RFECV is the highest, which is 86%, in comparison with 77% by RF, 74% by ET and 72% by LR. As the validation accuracy is a more critical criterion, it can be concluded that REFCV outperforms RF, ET and LR.



(a) RF model



(b) ET model



(c) LR model

Fig. 22: The validation accuracy of the CNN model based on features achieved by intelligent models.

5. Conclusions

Tool wear prediction is one of the essential research issues to effectively implement sustainability management of CNC machine tool systems. Tool wear progression is governed by complex underlining physics of tools and influenced by various machining factors. To process complex signals from heterogenous sensors for accurate tool wear prediction, in this paper, a novel methodology of incorporating signal de-nosing, feature optimisation and a CNN model is presented. The characteristics of the methodology include:

- In this research, signal de-noising is conducted by a Hampel filter-based method to provide high-quality data for tool wear prediction;

- The quality and relevance of features extracted from heterogenous sensors are optimised based on designed REFCV-based and Isomap-based methods, which improve the computational efficiency and tool wear prediction accuracy of the CNN model;
- A case study with signals of heterogenous sensors from a CNC machine tool system was used to validate the research methodology. Based on the case study, 80% of features originally extracted from the signals were reduced after feature optimisation and 86% of prediction accuracy was achieved by the CNN model. The proposed methodology was benchmarked with comparative algorithms to demonstrate its better performance in terms of prediction accuracy.

In the future research, more complex case studies and datasets will be explored to validate the proposed methodology, and the developed methodology will be explored to be applied to crater wear and other forms of tool failure to develop a more comprehensive toolkit. Meanwhile, it also needs to investigate how to handle incomplete or biased datasets using designing appropriate deep transfer learning models.

References:

[1] Y. Zhou, W. Xue (2018), A multisensor fusion method for tool condition monitoring in milling, *Sensors*, 18(11), pp. 3866.

[2] A. A. Thakre, A. V. Lad, and K. Mala (2019), Measurements of Tool Wear Parameters Using Machine Vision System, *Modelling and Simulation in Engineering*, 2019(9), pp. 1–9.

[3] N. T. Alagan, P. Hoier, P. Zeman, U. Klement, T. Beno, and A. Wretland (2019), Effects of high-pressure cooling in the flank and rake faces of WC tool on the tool wear mechanism and process conditions in turning of alloy 718, *Wear*, 434–435, p. 102922.

[4] V. D. Patel and A. H. Gandhi (2016), Analytical and Empirical Modeling of Wear and Forces of CBN Tool in Hard Turning - A Review, *Journal of The Institution of Engineers (India): Series C*, 98(4), pp. 507–513.

[5] A. Singh, S. Ghosh, and S. Aravindan (2020), Flank wear and rake wear studies for arc enhanced HiPIMS coated AlTiN tools during high speed machining of nickel-based superalloy, *Surface and Coatings Technology*, 381, p. 125190.

[6] M. Kuntoğlu, A. Aslan, H. Sağlam, D. Y. Pimenov, K. Giasin, and T. Mikolajczyk (2020), Optimization and Analysis of Surface Roughness, Flank Wear and 5 Different Sensorial Data via Tool Condition Monitoring System in Turning of AISI 5140, *Sensors*, 20(16), p. 4377.

[7] G. Serin, B. Sener, A. M. Ozbayoglu, and H. O. Unver (2020), Review of tool condition monitoring in machining and opportunities for deep learning, *International Journal of Advanced Manufacturing Technology*, 109 (3–4), pp. 953–974.

[8] C. Zhang, X. Yao, J. Zhang, H. Jin (2016), Tool condition monitoring and remaining useful life prognostic based on a wireless sensor in dry milling operations, *Sensors*, 16(6), pp. 795.

[9] A. Kothuru, S. P. Nooka, R. Liu (2019), Application of deep visualization in CNN-based tool condition monitoring for end milling, *Procedia Manufacturing*, 34, pp. 995–1004.

[10] A.P. Kene, S.K. Choudhury (2019), Analytical modeling of tool health monitoring system using multiple sensor data fusion approach in hard machining, *Measurement*, 145, pp. 118–129.

[11] K. L. de Calle, S. Ferreiro, A. Arnaiz, B. Sierra (2019), Dynamic condition monitoring method based on dimensionality reduction techniques for data-limited industrial environments, *Computers in Industry*, 112, pp. 103114.

[12] L. Feng, C. Zhao, B. Huang (2019), A slow independent component analysis algorithm for time series feature extraction with the concurrent consideration of high-order statistic and slowness, *Journal of Process Control*, 84, pp. 1–12.

[13] C.K. Madhusudana, H. Kumar, S. Narendranath (2019), Vibration-based fault diagnosis of a face milling tool using empirical mode decomposition features and artificial neural network, *International Journal of Condition Monitoring*, 9(2), pp. 25–34.

[14] T. Mohanraj, S. Shankar, R. Rajasekar, R. Deivasigamani, P.M. Arunkumar (2019), Tool condition monitoring in the milling process with vegetable based cutting fluids using vibration signatures, *Materials Testing*, 61(3), pp. 282–288.

[15] E. G. Plaza, P.J.N. López, E.M.B. González (2019), Efficiency of vibration signal feature extraction for surface finish monitoring in CNC machining, *Journal of Manufacturing Processes*, 44, pp. 145–157.

[16] A. Bommert, X. D. Sun, B. Bischl, J. Rahnenführer, M. Lang (2020), Benchmark for filter methods for feature selection in high-dimensional classification data, *Computational Statistics & Data Analysis*, 143, pp. 106839.

[17] Z.M. Hira, D.F. Gillies (2015), A review of feature selection and feature extraction methods applied on microarray data, *Advances in Bioinformatics*, 2015, pp. 1-13.

[18] W. Chang et al. (2019), A machine learning-based prediction method for hypertension outcomes based on medical data, *Diagnostics*, 9(4), pp. 178.

[19] Y. Mao, Y. Yang (2019), A wrapper feature subset selection method based on randomized search and multilayer structure, *BioMed Research International*, 2019, pp. 1-9.

[20] S. Chatterjee, D. Dey, S. Munshi (2019), Integration of morphological pre-processing and fractal based feature extraction with recursive feature elimination for skin lesion types classification, *Computer Methods and Programs in Biomedicine*, 178, pp. 201–218.

[21] B. Pes (2019), Ensemble feature selection for high-dimensional data: a stability analysis across multiple domains, *Neural Computing and Applications*, 32(10), pp. 5951–5973.

[22] B. Niu, J. Sun, B. Yang (2020), Multisensory based tool wear monitoring for practical applications in milling of titanium alloy, *Materials Today: Proceedings*, 22, pp. 1209–1217.

[23] C. Zhou, K. Guo, B. Yang, H. Wang, J. Sun, L. Lu (2020), Singularity analysis of cutting force and vibration for tool condition monitoring in milling, *IEEE Access*, 7, pp. 134113–134124.

[24] C. Lai, S. Guo, L. Cheng, W. Wang (2017), A comparative study of feature selection methods for the discriminative analysis of temporal lobe epilepsy, *Frontiers in Neurology*, 8, pp. 633.

[25] W. Wu, S. Faisal (2019), A data-driven principal component analysis-support vector machine approach for breast cancer diagnosis: Comparison and application, *Transactions of the Institute of Measurement and Control*, 42(7), pp. 1301–1312.

[26] B. Neef, J. Bartels, S. Thiede (2018), Tool wear and surface quality monitoring using high frequency CNC machine tool current signature, *Proceedings of the 16th IEEE International Conference on Industrial Informatics (INDIN)*, Porto, 2018, pp. 1045-1050.

[27] G. Wang, Y. Yang, Z. Li (2014), Force sensor based tool condition monitoring using a heterogeneous ensemble learning model, *Sensors*, 14(11), pp. 21588–21602.

[28] R. Mali, M.T. Telsang, T.V.K. Gupta (2017), Real time tool wear condition monitoring in hard turning of Inconel 718 using sensor fusion system, *Materials Today: Proceedings*, 4(8), pp. 8605–8612.

[29] Z. Zhang, H. Chen, Y. Xu, J. Zhong, N. Lv, S. Chen (2015), Multisensor-based real-time quality monitoring by means of feature extraction, selection and modeling for Al alloy in arc welding, *Mechanical Systems and Signal Processing*, 60–61, pp. 151–165.

[30] Y. Chen, H. Li, L. Hou, J. Wang, X. Bu (2018), An intelligent chatter detection method based on EEMD and feature selection with multi-channel vibration signals, *Measurement*, 127, pp. 356–365.

[31] V. Pandiyan, W. Caesarendra, T. Tjahjowidodo, H.H. Tan (2018), In-process tool condition monitoring in compliant abrasive belt grinding process using support vector machine and genetic algorithm, *Journal of Manufacturing Processes*, 31, pp. 199–213.

[32] G.D. Simon, R. Deivanathan (2019), Early detection of drilling tool wear by vibration data acquisition and classification, *Manufacturing Letters*, 21, pp. 60–65.

[33] J. Lee, J.Y. Jeong, C.H. Jun (2020), Markov blanket-based universal feature selection for classification and regression of mixed-type data, *Expert Systems with Applications*, 158, pp. 113398.

[34] Y. Li *et al.* (2018), DEEPre: sequence-based enzyme EC number prediction by deep learning, *Bioinformatics*, 34(5), pp. 760–769.

[35] F. Luo, M. Wang, Y. Liu, X.-M. Zhao, A. Li (2019), DeepPhos: prediction of protein phosphorylation sites with deep learning, *Bioinformatics*, 35(16), pp. 2766–2773.

[36] J. Ngarambe, A. Irakoze, G. Y. Yun, G. Kim (2020), Comparative Performance of Machine Learning Algorithms in the Prediction of Indoor Daylight Illuminances, *Sustainability*, 12(11), pp. 4471.

[37] Y. Chen, H. Li, L. Hou, X. Bu (2019), Feature extraction using dominant frequency bands and time-frequency image analysis for chatter detection in milling, *Precision Engineering*, 56, pp. 235–245.

[38] S. Maurya, V. Singh, N.K. Verma (2020), Condition monitoring of machines using fused features from EMD-based local energy with DNN, *Journal of IEEE Sensors*, 20(15), pp. 8316–8327.

[39] A. Agogino, K. Goebel (2007), Milling Data Set - NASA Ames Prognostics Data Repository, Available: http://ti.arc.nasa.gov/project/prognostic-data-repository (last accessed on 6 December 2020)

[40] Endevco, Endevco Model 7201-50, buy.endevco.com. https://buy.endevco.com/accelerometer/7201-accelerometer-5 (accessed Mar. 07, 2021).

[41] Physical Acoustics (2011), WD - 100-900 kHz Wideband Differential AE Sensor, Acoustic Emission Systems and NDT Products by PHYSICAL ACOUSTICS, https://www.physicalacoustics.com/by-product/sensors/WD-100-900-kHz-Wideband-Differential-AE-Sensor (accessed Mar. 07, 2021).

[42] Electronics datasheets (2011), SET-3B from Omron, www.electronicsdatasheets.com, https://www.electronicsdatasheets.com/manufacturers/omron/parts/set-3b#datasheet (accessed Mar. 07, 2021).

[43] Flex-core, Hall Effect DC Current Transducer w/ Signal Conditioner, FLEX-CORE®. https://www.flex-core.com/products/transducers/current-transducers/dc-current-transducers/ctl-cta-dc-current-transducer/ (accessed Mar. 07, 2021).

[44] Kennametal, "KICR - Indexable Inserts," www.kennametal.com. https://www.kennametal.com/us/en/products/p.inserts-for-kicr-speb-inch.1157649.html#tad (accessed Mar. 07, 2021).

[45] Z. Yao, J. Xie, Y. Tian, Q. Huang (2019), Using Hampel identifier to eliminate profile-isolated outliers in laser vision measurement, *Sensors*, 2019, pp. 1-12.

[46] R. K. Pearson, Y. Neuvo, J. Astola, and M. Gabbouj (2016), Generalized Hampel Filters, *EURASIP Journal on Advances in Signal Processing*, 2016(1), pp. 87.

[47] M. A. F. Ahmad, M. Z. Nuawi, S. Abdullah, Z. Wahid, Z. Karim, and M. Dirhamsyah (2015), Development of Tool Wear Machining Monitoring Using Novel Statistical Analysis Method, I-kazTM, *Procedia Engineering*, 101, pp. 355–362.

[48] B.K. Pappachan, W. Caesarendra, T. Tjahjowidodo, T. Wijaya (2017), Frequency domain analysis of sensor data for event classification in real-time robot assisted deburring, *Sensors*, 17(6), pp. 1247.

[49] P. Krishnakumar, K. Rameshkumar, K. I. Ramachandran (2015), Tool wear condition prediction using vibration signals in high speed machining (HSM) of Titanium (Ti-6Al-4V) alloy, *Procedia Computer Science*, 50, pp. 270–275.

[50] Y. Mu, X. Liu, L. Wang (2018), A Pearson's correlation coefficient-based decision tree and its parallel implementation, *Information Sciences*, 435, pp. 40–58.

[51] F. Gottwalt, E. Chang, T. Dillon (2019), CorrCorr: A feature selection method for multivariate correlation network anomaly detection techniques, *Computers & Security*, 83, pp. 234–245.

[52] B. Kalaiselvi, M. Thangamani (2020), An efficient Pearson correlation based improved random forest classification for protein structure prediction techniques, *Measurement*, 162, pp. 107885.

[53] M. A. Nanda, K. B. Seminar, D. Nandika, and A. Maddu (2018), A Comparison Study of Kernel Functions in the Support Vector Machine and Its Application for Termite Detection, *Information*, 9(1), pp. 5.

[54] A. Caggiano, R. Angelone, F. Napolitano, L. Nele, R. Teti (2018), Dimensionality reduction of sensorial features by principal component analysis for ANN machine learning in tool condition monitoring of CFRP drilling, *Procedia CIRP*, 78, pp. 307–312.

[55] J. Wang, J. Xie, R. Zhao, L. Zhang, L. Duan (2017), Multisensory fusion based virtual tool wear sensing for ubiquitous manufacturing, *Robotics and Computer-Integrated Manufacturing*, 45, pp. 47–58.

[56] Q. Liu, Y. Cai, H. Jiang, J. Lu, L. Chen (2018), Traffic state prediction using ISOMAP manifold learning, *Physica A: Statistical Mechanics and its Applications*, 506, pp. 532–541.

[57] D. Degenhardt, S. Seifert, and S. Szymczak (2019), Evaluation of variable selection methods for random forests and omics data sets, *Briefings in Bioinformatics*, 20(2), pp. 492–503.

[58] A. Sharaff, H. Gupta (2019), Extra-Tree Classifier with Metaheuristics Approach for Email Classification, *Advances in Intelligent Systems and Computing*, 924, pp. 189–197.

[59] M. A. Kahya, S. A. Altamir, Z. Y. Algamal (2019), Improving firefly algorithm-based logistic regression for feature selection, *Journal of Interdisciplinary Mathematics*, 22(8), pp. 1577–1581.