

Distributed Deep Learning Enabled Prediction on Cutting Tool Wear and Remaining Useful Life

Weidong Li^{1*}, Xiaoyang Zhang², Sheng Wang³, Xin Lu⁴, Zhiwen Huang¹

¹ School of Mechanical Engineering, University of Shanghai for Science and Technology, China

² Laboratory of Science and Technology on Millimetre-wave, Beijing Institute of Remote-Sensing
Equipment, China

³ Data Analytics and Machine Learning Group, Jaguar Land Rover Inc., U.K.

⁴ Faculty of Science and Technology, Bournemouth University, U.K.

* Corresponding author: weidongli@usst.edu.cn

Abstract

To optimise the utilisation cost of cutting tools, it is imperative to develop an online system to efficiently and accurately predict tool wear conditions and remaining useful lives (RULs). With this aim, a novel system is proposed based on deep learning algorithms distributed over an edge-cloud computing architecture. The system is innovative in the following aspects: (i) a lightweight convolutional neural network-random forest (CNN-RF) model is designed to be executed on an edge device to assess tool wear conditions efficiently, which supports severe tool resilience and tool replacement when necessary; (ii) a convolutional neural network-long short-term memory (CNN-LSTM) model is designed to be executed on a cloud to process long-term signals to predict the RUL of the cutting tool, which supports fine-tuning tool parameters dynamically; (iii) a signal compression mechanism is developed to condense the signals of tooling conditions into 2D images so the signal volumes transferred over the network are minimised and signal security is improved. Experiments were performed in a real-world machining workshop for research methodology validation. It showed that the prediction accuracies for tool wear and RUL achieved 90.6% and 93.2%, respectively, and the volume of signals transferred over the network was reduced by 89.0%. The experiments and benchmarks with comparative algorithms demonstrated that the system and its methodology exhibited great potential to reinforce cutting tool optimisation for real-world applications.

Keywords: Cutting tool wear, Remaining useful life, Edge-cloud computing, Deep learning

1. Introduction

Severe wear conditions of cutting tools will deteriorate product quality during machining processes. Effective prediction of the wear conditions and remaining useful lives (RULs) for cutting tools is a prominent requirement to achieve machining optimisation [1]. Tool condition monitoring, which could be conducted based on vibration, electrical, and/or acoustic emission signals from cutting tools collected on the fly, is a valuable enabler for developing such prediction functions. Commercial cloud-based tool condition monitoring systems have been provided on the market, including Montronix, Nordmann, Promotec, Marposs, Brankamp, Kistler, etc. In addition, various models for cutting tool condition predictions have been developed based on those cloud-based monitoring systems and machine learning algorithm-enabled data analytical approaches [2]. Nevertheless, surveys showed that over 74% of company executives and IT stakeholders did not prefer pure cloud computing solutions [3]. Major gaps between the developed systems/models and industrial requirements include the low efficiency/accuracy of tool wear condition predictions, and low efficiency and insecurity of data transfer in clouds. It is imperative to develop solutions to (i) improve prediction efficiency and accuracy in the wear conditions and RULs of cutting tools; (ii) minimise data volumes and ensure data security during sharing over industrial networks to support effective decision-making for severe tool resilience and tool replacement.

To bridge the gaps, this paper presents an innovative system of incorporating deep learning algorithms and edge-cloud computing paradigms to realise efficient prediction of wear conditions of cutting tools and accurate prediction of RULs of the tools. The research exemplifies that an effective synergy of deep learning algorithms and a distributed computing architecture is beneficial to addressing the versatile requirements arising from manufacturing processes. Research innovations include the following aspects:

- An edge-cloud computing system is designed as a “divide-and-conquer” infrastructure to support the prediction of cutting tools’ wear conditions and RULs. A novel encryption mechanism is realised to ensure the volume reduction and security assurance of tool condition monitoring signals exchanged over the industrial network;

- A convolutional neural network-random forest (CNN-RF) model is designed to implement efficient tool wear condition prediction by leveraging the limited memory and computational capacities of edge devices in the infrastructure. A convolutional neural network-long short-term memory (CNN-LSTM) model for the RUL predictions of cutting tools, which requires interpreting time-series information in longer-term accumulated signals, is designed to execute on a cloud server by taking more intensively computational powers;
- Real-world experiments were conducted to validate the system and research methodology. Benchmarks with some typical intelligent algorithms were carried out to evidence the effectiveness and advantages of this research in terms of computational accuracy, computational efficiency and data reduction rate transferred over the network.

2. Literature Review

Edge computing (fog computing) has been designed to accelerate computational efficiency by processing data close to data originating sources. As a promising technology, edge computing has been adopted by a variety of applications. An intelligent factory management system was arranged on an edge service (e.g., Raspberry Pi) to optimise customised production lines [4]. An evolutionary optimisation algorithm was executed on the edge service to efficiently pursue a balanced optimisation between the energy consumption and the workload of production equipment during scheduling/rescheduling. A convolutional neural network (CNN)-based system to classify defective products on an assembly line was developed [5]. The CNN-based system was distributed into an edge-cloud architecture for system optimisation. In [6], the shallow layer of the CNN-based system was arranged on an edge device to pre-process product data in a more efficient means. The refined data was further dispatched to the deep layer of the CNN-based system at a cloud server to conduct detailed classifications of defective products. An edge gateway based on Raspberry Pi was installed in a workshop, and a Kalman filter was set up on the edge device to filter manufacturing resource signals. The research was validated using case studies from the workshop. Nevertheless, the relevant research is still preliminary, and it is imperative to develop more appropriate functions for edge/cloud computing infrastructures to accelerate the related decision-making processes.

To predict the wear conditions and RULs of cutting tools, force signals, power signals, sound signals, vibration signals, and acoustic emission signals have been widely collected. Machine learning algorithms, in particular, deep learning algorithms, have been increasingly developed to support relevant decision-making based on the signals [7]. An approach for tool wear prediction was designed based on a genetic algorithm and a backpropagation (BP) neural network [8]. An approach integrating a CNN model and a stacked bi-directional/ unidirectional LSTM model was designed to predict the wear condition and RUL of a cutting tool [9]. In this approach, the CNN model was designed to extract features from monitoring signals for dimensionality reduction, and the LSTM model was applied to train these features to achieve tool wear and RUL prediction. A data-based approach to predict cutting tool wear and RUL for micro-milling was developed [10]. Manufacturing parameters and tool condition images captured during the process were used to support a deep belief network (DBN) model to perform prediction. A transfer learning strategy was designed to conduct cutting tool prognostics for customised machining processes [11]. In [12], the advantages, disadvantages, and prospects of using sensors, including force sensors, vibration sensors, acoustic emission sensors, current and power sensors, image sensors, and thermal sensors, for milling operations were discussed. Trends in the application of advanced sensor systems for milling was summarised.

Though the aforementioned research has achieved some successes in tool wear condition and RUL prediction, there is great potential for further improvements. Based on the facts that multiple types of sensors for tool condition monitoring are increasingly adopted and a large volume of signals is generated for processing, the current research requirements include (i) designing a well-performed data processing mechanism to fuse multi-types of data to facilitate the predictions; (ii) developing appropriate intelligent algorithms that are executable on an edge-cloud computing architecture efficiently to enhance the performance of the predictions.

3. Research Methodology

3.1 System Functions

The system consists of the following components (more technical details are given in sub-sections):

(1) An edge-cloud computing infrastructure: the infrastructure consists of sensors, edge devices and a cloud server to support intelligent function implementation;

(2) On CNC machines: power sensors, vibration sensors and edge devices are mounted in CNC machines to monitor and process the working conditions of cutting tools throughout the machining process lifecycle in a real-time means;

(3) On the edge devices: (i) a CNN-RF model is designed to execute on the edge devices to predict tool wear conditions on the fly, which supports tool severity resilience and tool changes when necessary; (ii) a 2D image-based signal compression algorithm is devised to enhance the privacy of signals that are transferred to the cloud server for predicting the cutting tool's RUL;

(4) On the cloud server: a CNN-LSTM model is developed on the cloud to predict the cutting tool's RUL to fine-tune machining parameters dynamically.

The edge-cloud infrastructure is designed based on the infrastructure of our previous research, which effectiveness was proved in machining process lifecycle optimisation [4]. The effectiveness of the CNN-RF model and the CNN-LSTM model will be justified based on experiments and comparative algorithms presented in Section 4.

3.2 Cutting Tool Wear Prediction on Edge Devices

The CNN model is effective in extracting features autonomously [2], but its training process is computationally expensive. Instead, conventional machine learning models, such as the random forest (RF) model, can be more efficient in training but they need to take pre-extracted features. RF is a classifier that contains some decision trees trained using subsets of a given dataset and takes the average to enhance the predicted accuracy and training efficiency [13]. In this research, the strengths of the CNN model and the RF model are combined to design a new CNN-RF model to support cutting tool wear prediction on edge devices. The advantage of the CNN-RF model is that it is lightweight in structure and only a modest amount of training data is required to achieve high predictions.

Two CNN models are used as the base models to process power and vibration signals for feature extraction, respectively. Among the developed CNN models, LeNet-5 is a lightweight model and can achieve a relatively accurate outcome [14]. In this research, LeNet-5 is adopted to design the base model.

The RF model is then used to fuse the intermediate prediction results (features) extracted from the two base models to conduct cutting tool wear prediction. In the RF model, bagging and random selection mechanisms are developed to avoid the domination of a single decision tree to generate deviated prediction results. The designed RF model is illustrated in Fig. 1. Its detailed process is given below.

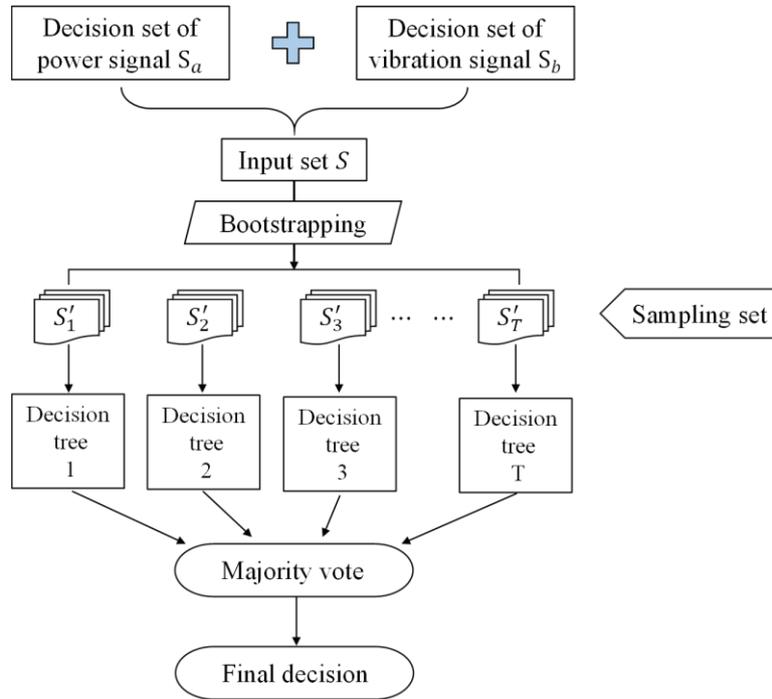


Fig. 1: The flow chart of the RF model.

Step 1: Intermediate features extracted by the two CNN models, i.e., $S_a = \{X_{a1}, X_{a2}, \dots, X_{an}\}$ and $S_b = \{X_{b1}, X_{b2}, \dots, X_{bn}\}$ that correspond to the power and vibration signals, are combined into a new set $S = \{S_a, S_b\}$ as the input of the RF model. The sample number of S is $2n$. Here, X denotes the probability of the correct prediction on tool wear, and n denotes the number of the input signals;

Step 2: Based on S , a bootstrapping method is adopted to randomly select a sample from S as an element of the sampling set. This sample is put back into S and the selection is performed again. After $2n$ times of repeated random selection, a sampling set S' containing $2n$ samples is obtained;

Step 3: The sampling set obtained in Step 2 is used as the root node of a decision tree. To perform the sample classification, calculation on the information gain is conducted for node split. From the root node, m ($m \ll 2n$) samples are randomly selected as sample of the lower branch node. Information Gain (G) provides the entropy difference between the root node and branch node, thus judging an optimal

sample that should be split as a branch node. Given there are two prediction labels (i.e., severe worn and unworn) in this research, the node split operation is executed twice. The entropy of the root node can be represented as:

$$E(A) = -\sum_{i=1}^m p_i \log_2 p_i \quad (1)$$

where p_i denotes the correct prediction probability of a sample.

The entropy of the lower node can be represented as:

$$E(B) = -\sum_{i=1}^v \frac{D_i}{D} \cdot p_i \log_2 p_i \quad (2)$$

where D represents the number of samples in the training dataset; D_i represent the number of samples that are separated in the training dataset.

Based on the above formulas, Information Gain is calculated below:

$$G = E(A) - E(B) \quad (3)$$

Step 4: According to the evaluation of Information Gain, a decision tree with a labelled prediction capability can be established. For the purpose of sufficient robustness and generalisation of the results, Steps 2 and 3 are repeated $2n$ times to establish $2n$ decision trees to compose a RF;

Step 5: Through the above-described steps, the desired RF model is established. The majority vote of all the decision trees can be used to represent the prediction result, which is defined as:

$$Accuracy = \sum_{i=1}^m p_i \sum_{j \neq i}^m p_j = \sum_{i=1}^m P_i(1 - p_i) = 1 - \sum_{i=1}^m p_i^2 \quad (4)$$

where $\sum_{i=1}^m p_i$ and $\sum_{j \neq i}^m p_j$ represent the probability of the correct and incorrect prediction respectively.

3.3 Compression on Cutting Tools' Signals

For the purpose of signal security and reduction, it is necessary to convert the signals of cutting tool wear conditions into a 2D image form (the process is illustrated in Fig. 2).

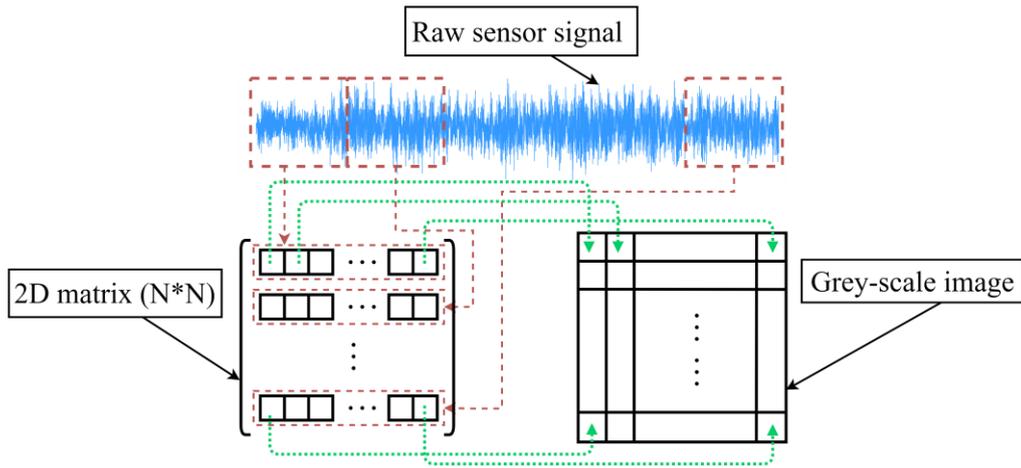


Fig. 2: Illustration of the signal compression process.

The technical details are depicted below:

Step 1: During a machining process, it is normal to record standby periods in signals that come from repositioning or resetting a cutting tool. For the purpose of optimisation, signals in those standby periods will be eliminated;

Step 2: The optimised signals are equally segmented into N sub-sessions. The lengths of the sub-sessions are the same to ensure a packed matrix structure from those sub-sessions meeting the input structure requirement by intelligent algorithms developed in this research;

Step 3: The above sub-sessions are organised as a 2D matrix in the dimension of $N \times N$. In the matrix, each element represents the pixel of the grey-level image. The process of establishing the matrix is illustrated in Fig. 3;

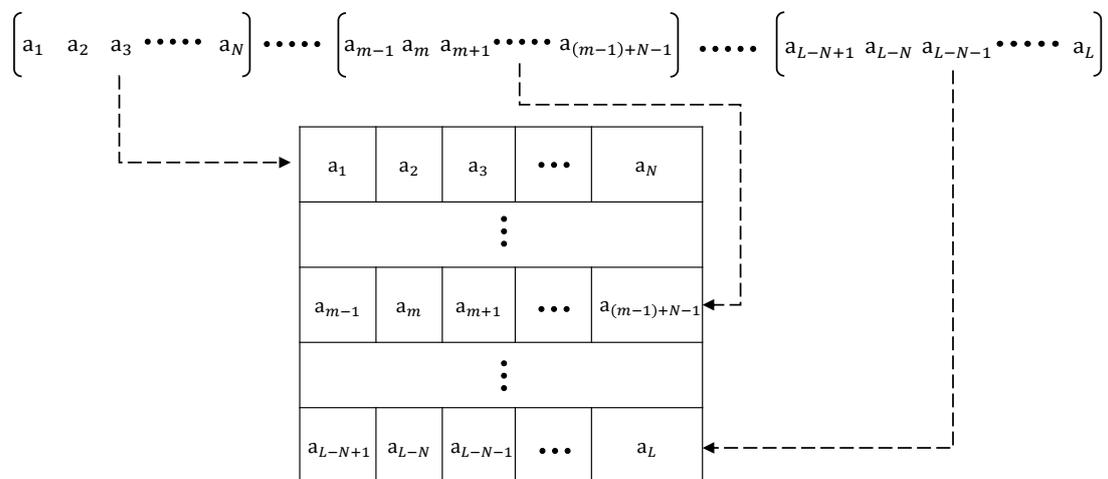


Fig. 3: The illustration of the compression process of signals in a matrix.

Step 4: The matrix will be converted into greyscale pixels. The process is conducted through normalising values in the matrix so the pixel intensity ranges from 0 to 255. The normalisation process is carried out based on the following formula:

$$P(i, j) = \text{round} \left(\frac{S(i, j) - \text{Min}(S)}{\text{Max}(S) - \text{Min}(S)} \times 255 \right) \quad (5)$$

where $i = 1, 2, \dots, N$, $j = 1, 2, \dots, N$; $S(i, j)$ represents the values of the signals; $P(i, j)$ represents the pixel intensity; $\text{round}(\cdot)$ represents a function of returning an integer as an input value.

3.4 RUL Prediction of Cutting Tool on Cloud

Prediction on a cutting tool's RUL is a computationally intensive process. A cloud server has the capability to process the volume of signals so it is more appropriate to conduct the RUL prediction of a cutting tool. Two CNN models are adopted at the cloud side as base models to extract features from signals. Moreover, extracted features will be further considered as the input of a subsequent LSTM model for RUL prediction. LSTM is a special kind of the recurrent neural network (RNN) model that is capable of learning long-term dependencies in time-series data to conduct prediction tasks [2, 10]. The CNN-LSTM model benefits from the combined strengths of CNN and LSTM in feature extraction and time-series information learning to support the research for the RUL predictions of cutting tools.

As a more powerful CNN architecture, AlexNet is capable of processing a large volume of signals and executing the process on a cloud server [15]. Thus, in this research, AlexNet is adopted as the base model. LSTM is superior in processing time-series data owing to a memory cell design to retain temporal features. The structure of the LSTM model in this research is shown in Fig. 4.

The extracted feature array by the AlexNet model is represented as $X_m = [x_m]$, where x is for the feature value and m is for the number of features. The array is used as the input of the LSTM model. In the LSTM model, the memory cell state of each unit in the hidden layer is used to recursively update the prediction result h_t of the current time-step t based on h_{t-1} of the previous time-step $t - 1$. In this process, the relevant information is filtered by the forget gate f_t and combined with a new feature x_t . The intermediate result h_t at the time-step t can be calculated under the control of the intermediate gate o_t using the following formula:

$$h_t = o_t \odot \tanh c_t \quad (6)$$

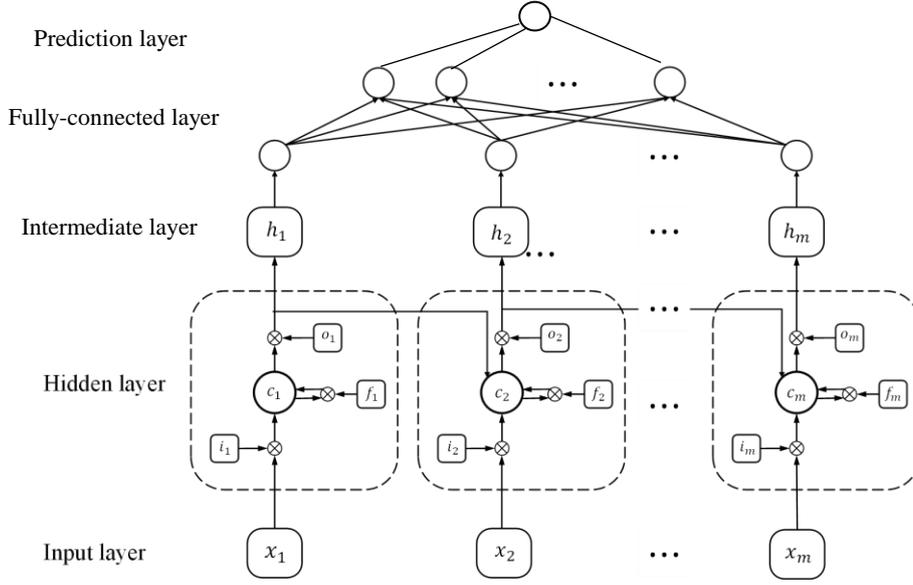


Fig. 4: The structure of the LSTM model.

The fully-connected layer and linear regression layer in the LSTM model are connected with the output gate of each LSTM unit sequentially. Generated features are input into the fully-connected layer to produce the output vector $H = \{h_1, h_2, \dots, h_l\}$. That is, the output of the fully-connected layer is calculated below:

$$F = \sum_{i=1} ReLU(w_i \cdot h_i + b_i) \quad (7)$$

where $ReLU$ denotes the Rectified Linear Unit's activation functions; w_i is the weight; b_i is the bias.

The RUL prediction of a cutting tool is realised in the prediction layer below:

$$P = \sum_{i=1} (w_i \cdot F_i) \quad (8)$$

The mean absolute error (MAE) is designed to evaluate the predictive model [12]:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\tilde{y} - y| \quad (9)$$

where n is the number of the training dataset; \tilde{y} is the prediction value; y is the ground-truth value.

4. Experiments and Analyses

4.1 Experimental Configuration

Experiments for research validation were carried out based on the teeth machining process of a ring synchroniser in a seven-speed dual-clutch transmission gearbox [16]. In the experiments, the JCMT JCS30 CNC machine and a cutting tool TNEX 1604 XF were utilised (the specifications of the machine, the machining parameters, and tooling are given in Tables 1-2, respectively). The material for the ring synchroniser is 16MnCr5 (EN 10084: 2008). The chemical composition and mechanical properties of the material are given in Table 3. The ring synchroniser and the insert of the cutting tool are illustrated in Fig. 5(a). Fig. 5(b) shows the deployment arrangement of the predictive system into the machine tooling system. Three YHDC SCT013 current sensors and an MMA7361 three-axis accelerometer were mounted on the three-phase power supply and the tool holder of the machine. Power and vibration signals during experimental machining processes were collected.

Table 1: Specifications of the JCMT JCS30 CNC machine.

Specifications	Maximum machining diameter (mm)	Maximum machining modulus (mm)	Maximum tooth width (mm)	Maximum speed of tool spindle (rpm)	Maximum driving power of tool spindle (kw)	Inclination angle of tool shaft (degree)	Maximum speed of workpiece spindle (rpm)
Values	300	3	100	6000	23.1	25	29.1

Table 2: Tooling specifications and machining conditions.

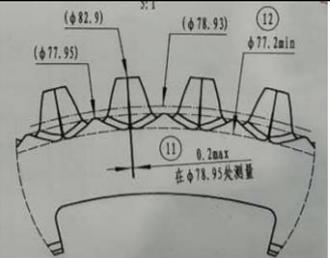
	<p>Machining conditions: continuous machining Work material for the ring synchroniser: 16MnCr5 (EN 10084: 2008) Hardness of the work material: HRC58-62 Cutting tool: TNEX 1604 XF Cutting parameters: S=2400 rpm, F₁=25 mm/min (roughing), F₂=15 mm/min (semi-finishing), F₃=5 mm/min (finishing), V_c=170 mm/min (in this experiment, roughing machining was used) Cutting mode: dry cutting</p>
---	--

Table 3: The chemical composition and mechanical properties of the work material (ring synchroniser).

Chemical composition	C	Si	Mn	S	P	Cr
Values	0.14 ~ 0.19	≤0.40	1.00 ~ 1.30	0.010 ~ 0.035	≤0.035	0.80 ~ 1.10
Mechanical property	Tensile strength (MPa)	Yield strength (MPa)	Elongation (%)	Reduction of area (%)	Impact toughness value (J/cm ²)	Hardness (HB)
Values	1373	1187	13	57	72	≤297

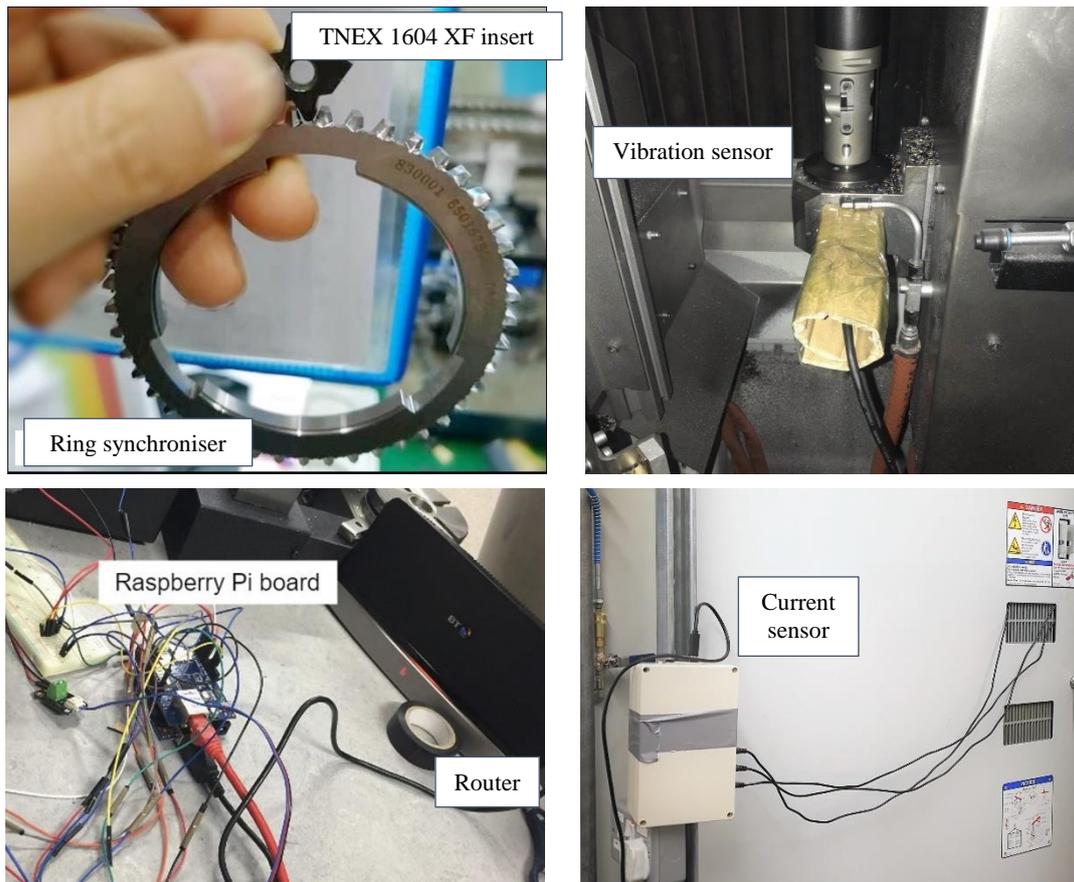


Fig. 5: (a) The ring synchroniser and the cutting tool's insert; (b) The system deployed into the machine.

To minimise the randomness of experimental results, three groups of machining processes were carried out using the same machining parameters (the parameters are given in Table 2). The number of machined components is 37, 29 and 39 in each group, respectively. A new cutting tool was used at the beginning of each machining group. Table 4 shows the total number of machined components and the component number machined by a worn cutting tool.

Table 4: Information for machined components.

The machining group number	The quantity of machined components	The machined components using a worn tool
1	37	26 to 37
2	29	14 to 29
3	39	11 to 39

To alleviate the negative impact of random factors, in the experiments, three groups of machining processes under the same machining parameters and settings were conducted. In each group, a brand-

new cutting tool was used so the experiments were arranged on the same basis. Tool wear measurements used the parameters specified in ISO 3685:1993, particularly the variables V_B (average flank wear width) [12]. During the machining processes in the experiments, the flank wear conditions of the cutting tool insert were measured. Fig. 6 illustrates the wear conditions when the No. 30 component was machined in the machining group 1. The width of the tool wear is 0.357mm, which is greater than the threshold of V_B (0.150mm) pre-set by the company. The EDS analysis data of the tool are given in Fig. 7 (the chemical composition of the surface for the cutting tool is: N-62.86%, Al-20.00%, Ti-17.14%).

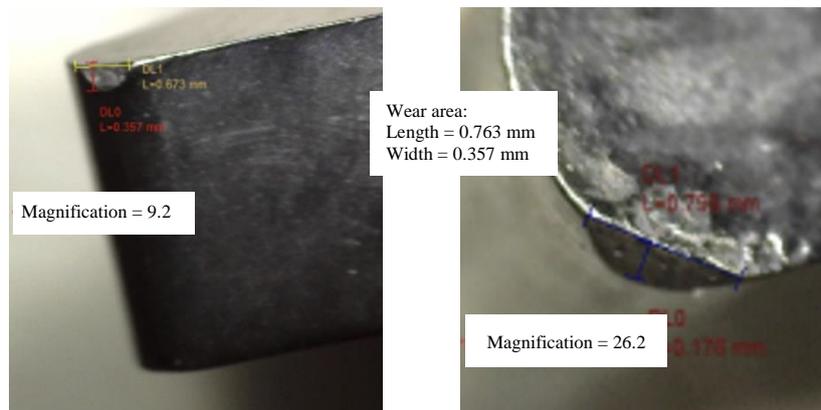


Fig. 6: The SEM (Scanning Electron Microscope) photographs of a tool's flank wear condition.

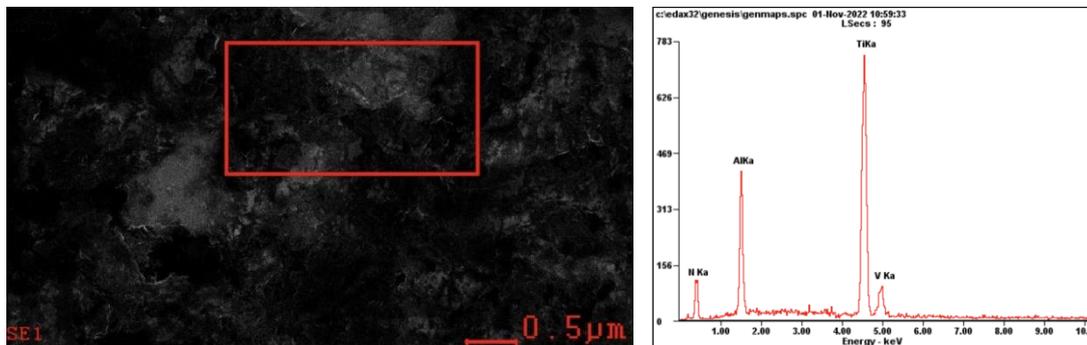


Fig. 7: The EDS analysis of the cutting tool.

The acquisition frequencies of vibration and power signals during the machining processes were set at 4,000Hz and 20Hz, respectively. The set frequencies are sufficient to collect time-domain and frequency-domain changes in the machining conditions during the experiments. Some of the acquired power and vibration signals are illustrated in Fig. 8.

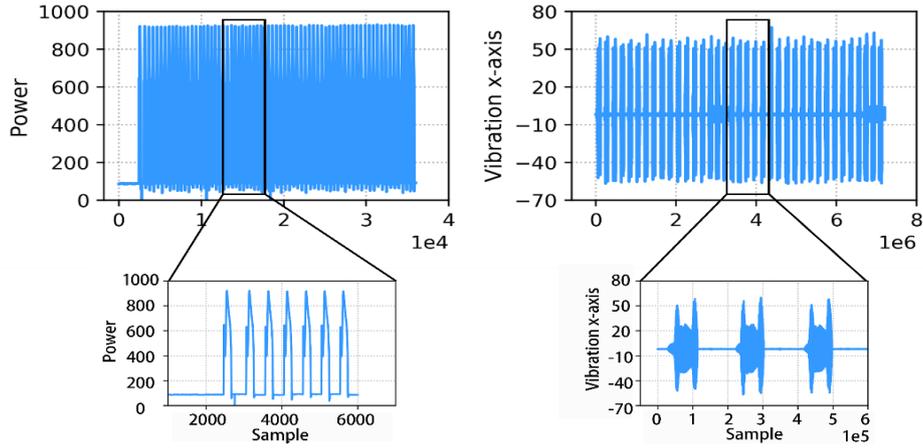


Fig. 8: Acquired power and vibration signals.

4.2 Cutting Tool Wear Prediction

Tool wear condition prediction using the CNN-RF model was executed on the edge device (Raspberry Pi). In the model, 2/3 experimental results were used for model training, and 1/3 experimental results were used for validation. Fig. 9 shows the performance results of two CNNs in the CNN-RF model for processing power and vibration signals in terms of prediction accuracy and loss. In Fig. 9, the accuracies and losses are for the training and validation datasets, respectively.

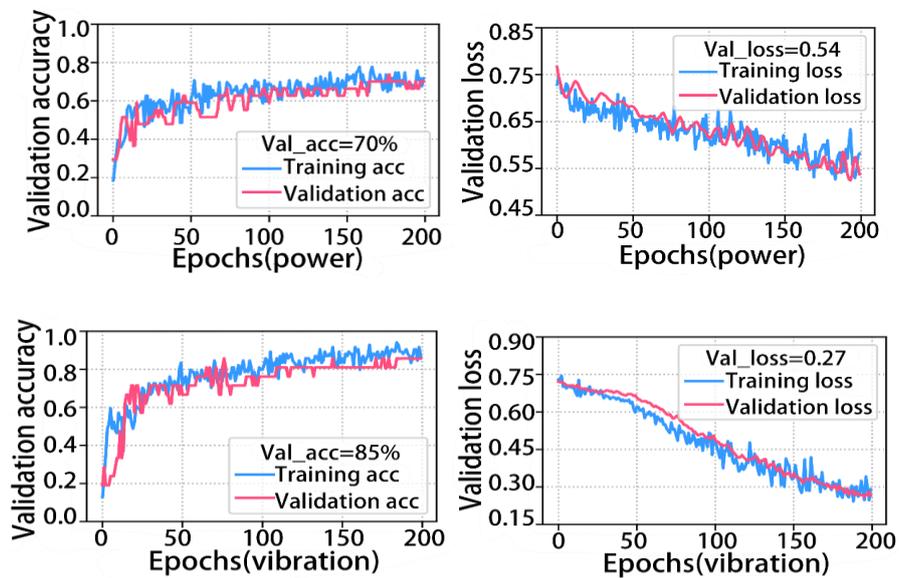


Fig. 9: Accuracies and losses of the CNNs for power and vibration signals.

The accuracies of the cutting tool wear prediction for the power and vibration signals were 70.0% and 85.0%, respectively. The losses for the signals were 0.54 and 0.27, respectively. Considering that the power and vibration signals exhibited different effects on the prediction accuracy, extracted features

from the signals using the two CNNs were fused to enhance the prediction performance. With the introduction of the RF model, the prediction accuracy was further improved to reach 90.6% (the performance of the RF model was indicated using accuracy, precision, recall and F1-score shown in Table 5). Industrial surveys conducted proved that the prediction accuracy of the CNN-RF model met the general requirement of industrial practices [16].

Table 5: The evaluation result of the RF model.

Accuracy	Precision	Recall	F1-score
90.6%	94.7%	90.0%	92.3%

The memory utilisation rate of the CNN-RF model was monitored. In the entire monitoring process, the maximum usage was 334.3MiB (0.55GB), which was within the capacity of Raspberry Pi. It showed that the CNN-RF model was executable and effective on the edge end. To further demonstrate the performance and superiority of the CNN-RF model, four mainstream machine learning models were used for benchmark, i.e., support vector machine (SVM), multilayered perceptron (MLP), k-nearest neighbours (KNN) and naive bayes (NB). Assessment results in terms of accuracy, precision, recall and F1-score are shown in Table 6. For accuracy, precision and F1-score, the CNN-RF model achieved the best results. For recall, the CNN-RF model achieved the second-best result. In summary, the CNN-RF model performed stably and its results were satisfactory in general.

Table 6: Evaluation results of the machine learning algorithms.

Model	Evaluation criteria			
	Accuracy	Precision	Recall	F1-score
CNN-RF	93.26%	94.74%	90.00%	92.31%
SVM	87.50%	94.44%	85.00%	89.47%
MLP	87.50%	86.36%	95.00%	90.48%
KNN	84.38%	85.71%	90.00%	87.80%
NB	81.25%	93.73%	75.00%	83.33%

Moreover, the computational time for the above five models was monitored. The results are shown in Fig. 10. It showed that MLP took the longest time, i.e., 1.8s, while SVM, KNN and NB took 1.06s, 0.6s and 0.5s, respectively. The computational time for the CNN-RF model was moderate. That is, CNN-RF was 15.5% faster than KNN, 29.5% faster than NB, but 60.6% less than MLP and 33.0% less

than SVM. It can be concluded that the CNN-RF model presents a more balanced performance in prediction accuracy and computational efficiency.

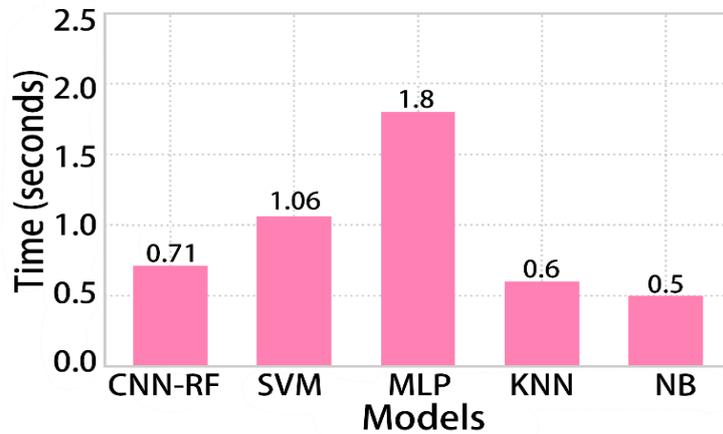


Fig. 10: The computational time by the models.

4.3 Signal Compression

To facilitate the processing of the system presented in this research, the continuously collected power and vibration signals were segmented according to the three groups of experiments. That is, 37 subsets for group 1, 29 subsets for group 2, and 39 subsets for group 3, were created. Segmented signals were normalised to the same size by removing invalid values in the signals, followed by the compression process of converting the signals into the image format. Statistically, the compression rate for the signals of the three groups via image compression process reached 89.0% on average. In Fig. 11, the comparative dataset sizes of the signals and the compressed images for the three groups are shown. The compression rates of the three groups are summarised in Table 7.

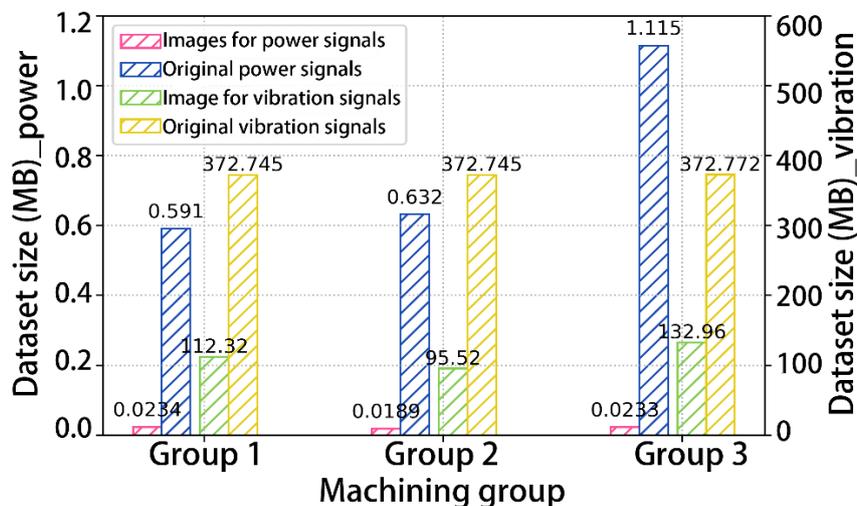


Fig. 11: The dataset sizes of three groups before and after the image compression process.

Table 7: The compression rates of the power and vibration signals.

Signal	Group 1	Group 2	Group 3
Power signals	96.0%	97.0%	97.9%
Vibration signals	69.8%	74.3%	64.3%

The type of Raspberry Pi used in this system is Raspberry Pi zero, and its RAM capacity is 1.0G. To verify whether Raspberry Pi has an enough RAM to process data, the memory usage during the entire life of the experiments was recorded. The maximum RAMs required for processing the power and vibration signals were 0.27GB and 0.55GB, respectively, which were within the allowable capacity of the Raspberry Pi zero.

4.4 Cutting Tool RUL Prediction

The CNN-LSTM model was used to predict the RUL of a cutting tool on the cloud. In this research, three datasets were adopted to validate this research. Moreover, as the most ordinarily used tool life criteria, the flank wear is adopted to express the cutting tool RUL in this work. In general, the threshold of the cutting tool RUL should be determined by the actual application scenario and demand. Based on the dataset employed, the actual RUL of each cutting tool is defined as the remaining processable cycle, 37, 29 and 39, respectively. Based on the predicted tool flank wear, the remaining processable cycle can be calculated below:

$$RUL_p^i = \frac{w_p^i \cdot RUL_a^i}{w_a^i} \quad (10)$$

where RUL_a denotes the actual remaining processable cycle, which is $\{37, 36, \dots, 1\}$ for example, w_p denotes the predicted flank wear, w_a denotes the actual measured flank wear, i denotes the number of the cuts.

In addition, to emphasise that the developed method is feasible to predict the RUL in a wide range of practical applications, a polynomial regression fitting model is constructed based on the remaining processable cycle obtained by Eq. (10) and the predicted flank wear. The regression function can be described below:

$$F = b_0 + b_1 \cdot w_1 + b_2 \cdot w_2^2 + \dots + b_k \cdot w_k^k = \sum_{j=0}^k b_j \cdot w^k \quad (11)$$

where F denotes the RUL of a cutting tool, w denotes the predicted flank wear for the cutting tool, b_j denotes the regression coefficients, k denotes the number of the cuts.

The RUL prediction results for the three validation sets are shown in Fig. 12. It displays that the model fitted the flank wear and the RUL effectively. To evaluate the accuracy of the prediction, the mean absolute error (MAE) ($MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y} - y|$, where n is the training sample size, \hat{y} is the predicted RUL and y is the actual RUL), was used to quantify the absolute error between prediction and actual values. MAEs of the three validation sets were 0.88, 0.89 and 0.90, respectively. As the values of the three R^2 were close to 1, it demonstrates that the results generated by the CNN-LSTM model fitted the ground-truth RUL values well. Thus, it can conclude that the CNN-LSTM model has a good performance in predicting the RUL of a cutting tool.

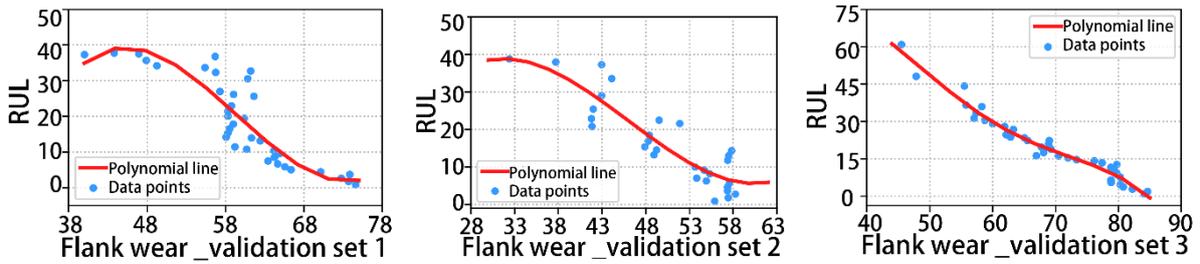


Fig. 12: The predicted RULs of cutting tools based on three validation datasets.

5. Conclusions

To minimise tooling waste and improve machining quality, it is paramount to carry out an efficient prediction of tool wear conditions and an accurate prediction of tool RUL during CNC machining processes. With this aim, in this research, a novel system for cutting tool wear and RUL prediction is developed. The major functions of the system and the system validation are from the following aspects:

- A lightweight CNN-RF model is designed and deployed on an edge device to efficiently estimate cutting tool wear conditions to support tool severity resilience;
- Furthermore, a CNN-LSTM model with a comprehensive analytic capability is designed and deployed at a cloud to estimate cutting tool's RUL based on longer-term accumulated signals to optimise tool parameters on the fly;

- This designed system was evaluated using real-world manufacturing experiments. Experimental results show that the accuracy of the tool wear prediction by the CNN-RF model was 90.6%, the accuracy of the cutting tool's RUL prediction by the CNN-LSTM model was 93.2%, and the volume reduction of cutting tool condition signals by the signal compression mechanism achieved 89.0%;
- Benchmarks on several mainstream machine learning algorithms were carried out to demonstrate the effectiveness and superiority of the developed research.

In future, the approach developed in this research can be extended to provide a more comprehensive assessment and prediction of tool wear conditions (including flank face wear, rake face wear, etc.) based on more types of signals.

Acknowledgement:

This research was sponsored by the National Natural Science Foundation of China (Project No. 51975444). This research was also partially supported by the SABOT project from the Innovative UK funder (UK). The authors would acknowledge Dr Jinsheng Wang and Mr Chaowei Hao from the Grindocor Ltd. for assisting in experimental design.

References:

- [1] Bernini L., Albertelli P., Monno M., Mill condition monitoring based on instantaneous identification of specific force coefficients under variable cutting conditions, *Mechanical Systems and Signal Processing*, 185: 109820, 2023.
- [2] Li W.D., Liang Y.C., Wang S., *Data Driven Smart Manufacturing Technologies and Applications*, Springer, 2021.
- [3] Kong C.P., Wei L., Zhou X.H., Niu Q., and Jiang J.G., A study on a general cyber machine tools monitoring system in smart factories, *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture* 235(14): 2250–61, 2021.
- [4] Liang Y.C., Li W.D., Lu X., Wang S., Fog computing and convolutional neural network enabled prognosis for machining process optimization, *Journal of Manufacturing Systems*, 52: 32-42, 2019.

- [5] Yeh C.S., Chen S.L., Li I.C., Implementation of MQTT protocol based network architecture for smart factory, Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 235(13):2132-2142, 2021.
- [6] Li L., Ota K., Dong M., Deep learning for smart industry: Efficient manufacture inspection system with fog computing, IEEE Transactions on Industrial Informatics, 14: 4665-4673, 2018.
- [7] Xu L., Huang C.Z., Li C.W., Wang J., Liu H.L., Wang X.D., Prediction of tool wear width size and optimization of cutting parameters in milling process using novel ANFIS-PSO method, Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 236(1-2): 111-122, 2022.
- [8] Zhao H., Liu H., Multiple classifiers fusion and CNN feature extraction for handwritten digits recognition, Granular Computing, 5: 411-418, 2019.
- [9] Wei W., Cong R., Li Y., Abraham A.D., Yang C., Chen Z., Prediction of tool wear based on GA-BP neural network, Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 236(12): 1564-1573, 2022.
- [10] 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, 154: 107461, 2020.
- [11] Bagri S., Manwar A., Varghese A., Mujumdar S., Joshi S.S., Tool wear and remaining useful life prediction in micro-milling along complex tool paths using neural networks, Journal of Manufacturing Processes, 71: 679-698, 2021.
- [12] Yu D., Gupta P.M., da Silva L.R.R., Kiran M., Khanna N., Krolczyk G.M., Application of measurement systems in tool condition monitoring of milling: A review of measurement science approach, Measurement, 199: 111503, 2022.
- [13] Wu D.Z., Jennings C., Terpenney J., Gao R.X., Kumara S., A comparative study on machine learning algorithms for smart manufacturing: Tool wear prediction using random forests, Transactions of the ASME, Journal of Manufacturing Science and Engineering, 139(7): 071018, 2017.

- [14] LeCun Y., Bottou L., Bengio Y., Haffner P., Gradient-based learning applied to document recognition, Proceedings of the IEEE. 86: 2278-2324, 1998.
- [15] Krizhevsky A., Sutskever I., Hinton G. ImageNet classification with deep convolutional neural networks, Communications of the ACM, 60: 84-90, 2017.
- [16] Zhang X.Y. Deep Learning Driven Tool Wear Identification and Remaining Useful Life Prediction, PhD Dissertation, Coventry University, 2021.