



# CNC milling cutters condition monitoring based on empirical wavelet packet decomposition

Abir Amar Bouzid<sup>1</sup> · Boualem Merainani<sup>1,2</sup> · Djamel Benazzouz<sup>1</sup>

Received: 20 April 2023 / Accepted: 5 October 2023 / Published online: 3 November 2023  
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2023

## Abstract

Machining is a versatile field in the manufacturing industry. In milling operations, tool wear is considered the most critical factor affecting the surface quality of the milled piece. Furthermore, the gradual tool wear impacts the milling process, leading to significant downtime, which has serious financial consequences. Unavoidably, a sustainable and reliable condition monitoring system must be developed to reduce the risk of downtime and enhance production quality. The deployment of prognostic and health management (PHM) solutions is becoming increasingly important. It is regarded as one of the main levers for monitoring tool wear status. In this paper, a novel methodology is proposed for extracting pertinent health indicators (HIs) that reflect the degradation behavior of a set of milling cutters and estimating their remaining useful lives (RULs). First, a new time-frequency signal-analysis approach, titled empirical wavelet packet decomposition (EWPDP), is proposed to scrutinize the data collected via multi-sensor acquisition. This technique provides a new segmentation of the signal's Fourier spectrum, distributed on levels, to investigate a broader variety of frequency bands and enhance the traditional segmentation structure's performance. Second, a new health indicator is designed based on an innovative selection of the time-domain features computed for each frequency band over each level. Finally, the long short-term memory (LSTM) network is used to estimate the RUL of each cutter. A comparison between the suggested processing method and the wavelet packet transform (WPT) is made to support the hypothesis regarding the effectiveness of the proposed technique. Experimental outcomes seem to be satisfying.

**Keywords** Milling cutter · Condition monitoring · PHM · EWPDP · Feature extraction · RUL

## 1 Introduction

Nowadays, manufacturing industries are undergoing tremendous evolution and ongoing technological growth, with the primary goal of ensuring the reliability of manufacturing systems. In this regard, monitoring tool conditions has turned

into an important task in the field of metal cutting. Although cutting tools account for less than 4% of the total machining cost [1], their failure can result in a machine breakdown of 10 to 40% [2], which increases the overall cost of downtime. Under high temperatures and stress, cutting tools are subjected to severe frictional processes with the working piece. This leads to a critical condition called tool wear, which gradually affects the surface quality of the machined piece until either the milling machine or the workpiece is damaged. An effective prognostic and health management (PHM) system for tool health monitoring is, then, essential to enable informed actions aimed at enhancing cutting tool utilization and preventing catastrophic situations resulting from tool failure. Elattar et al. [3] classified prognostic methods into three main categories: physics-based, statistical reliability-based, and data-driven-based techniques. Physical model-based methods rely on modeling the system dynamics model that reflects its behavior and incorporates the degradation process. However, even if these methods are the most

✉ Abir Amar Bouzid  
a.amarbouzid@univ-boumerdes.dz

Boualem Merainani  
b.merainani@univ-boumerdes.dz;  
boualem.merainani@univ-eiffel.fr

Djamel Benazzouz  
d.benazzouz@univ-boumerdes.dz

<sup>1</sup> Solid Mechanics and Systems Laboratory,  
University of M'hamed Bougara, Frantz Fanon Road,  
35000 Boumerdes, Boumerdes, Algeria

<sup>2</sup> COSYS-SII, Gustave Eiffel University, Allée des Ponts  
et Chaussées, 44344 Bouguenais, Bouguenais, France

well-known in terms of precision and accuracy, the complexity of real-world industrial systems makes the conversion of the physical behavior into an analytical representation highly challenging [4]. Statistical reliability-based approaches, as the name implies, rely on statistical models, such as Weibull and Poisson distributions. They are solely based on historical life-cycle data, which includes the failure rate of a list of components belonging to the same category and specifications as the monitored part. Even though these techniques are simple to implement, they have a major restriction in that they require a large amount of historical data [5], making them inaccurate for newly manufactured components. Data-driven methods are the most widely used and developed prognostic category, owing to their ease of deployment, which does not require an analytical representation of the system's behavior. PHM-data-driven prognostic process is realized in three major steps [4]:

- Data acquisition
- Data processing
- Model learning and RUL prediction

In the data acquisition step, the acquisition system records the numerical data provided by sensors or transducers. Different types of sensors can be used depending on the critical component to be monitored, where, in the milling process, sensor selection is driven by the condition that can be related to tool wear, such as vibrations, cutting forces, acoustic emissions, and power consumption of the spindle [6]. The obtained data is processed, in the next step, to extract critical features with which the degradation level of the cutting tools is estimated and the wear state is evaluated.

Undoubtedly, this is the most challenging step regardless of the type of monitoring sensor. For instance, the vibration and acoustic emission data may contain structural and bearing vibration information. So, the data becomes complex, non-linear, and usually overwhelmed by noise [5].

It is reported, in related works [7, 8], that features can be extracted from time, frequency, and time-frequency domain analysis. Features derived, independently, from the time or/and frequency domains capture specific signal content while potentially losing information about the status of cutting tool degradation due to the non-stationarity of real signals and the sensitivity to the experimental conditions affecting milling cutters [9].

Time-frequency analysis is becoming a research hotspot since the used techniques, such as Hilbert-Huang transform (HHT), wavelet transform (WT), wavelet packet transform WPT, and empirical wavelet transform (EWT), can extract the dynamic health condition of milling cutters [10]. Numerous work has been carried out for this purpose. Benkedjough et al. [9] combined continuous wavelet transform (CWT) with blind source separation (BSS) to estimate the remain-

ing useful life (RUL) of the cutting tools. The CWT was used to separate the signals into coefficients for taking a certain scale of wavelet coefficients for BSS. The energy of the segregated signals is computed to produce the health indicator. Liao et al. [11] worked on monitoring tool wear conditions, proposing a hybrid hidden Markov model (HHMM) based on CWT. To begin, an internal dynamic hidden Markov model (HMM) is created to capture the WT's dynamic dependency at various frequencies with instantaneous time resolution. After that, an external HMM for continuous monitoring is built, which aggregates the WT dependencies and portrays the dynamic worsening of the tool wear state in long term. Shen et al. [12] implemented a strategy for spindle power signal-based online tool condition monitoring, in which a system dedicated to the signal acquisition is constructed, and then, Hilbert-Huang transform (HHT) is applied to extract suitable features that indicate the tool condition. Segregeto et al. [13] employed WPT and machine learning paradigms to estimate tool wear in the turning of Inconel 718. The WPT was applied to vibration, cutting force, and acoustic emission signals to extract appropriate features. The correlation coefficient method is then used to select the most relevant features, and finally, a machine learning model based on an artificial neural network was employed to estimate the flank wear.

Despite the high capabilities, of the aforementioned methods, for signal decomposition, each has its limitation. HHT lacks from mathematical basis. WPT or CWT restricts the extraction of the HIs from a set of frequency bands constructed from a prescribed subdivision scheme of the Fourier axis. By contrast, EWT provides a self-adaptive decomposition of this axis concerning the frequency content of the analyzed signals. However, the contact of the cutting tool with the workpiece will certainly excite the machine structure, creating structural resonant frequencies that will be added to those of the electric motor components. These frequencies are known to have high energy. In this context, as the EWT first detects the local maxima, in the spectrum, to construct filters, the aforementioned frequencies, and the noise will greatly influence the results. Thus, isolating a frequency band in which the cutting tool wear is manifesting is tedious.

With the aim of developing a richer filter bank, a new segmentation model of the Fourier axis is proposed in this paper, denoted as the empirical wavelet packet decomposition (EWPDP), to obtain comprehensive frequency band information, overcoming the traditional segmentation structure issues [14, 15]. Besides, a new health indicator is constructed to represent the degradation state of a collection of CNC milling cutters that rely on a clever selection of the time-domain features from the proposed filter bank. Ultimately, RUL estimation is performed based on the constructed HIs.

For prognostic and model learning, extensive surveys have been devoted to developing an accurate predictive model capable of indicating the RUL of the monitored component. Prognostic uses a variety of machine learning (ML) algorithms, namely artificial neural networks (ANN), support vector machine (SVM), support vector regression (SVR), fuzzy logic (FL), and deep learning (DL) techniques [4, 5, 16]. Due to their outstanding performance, deep learning DL methods, including gated recurrent unit (GRU), convolutional neural network (CNN), and long short-term memory (LSTM) network, have lately become the most advanced techniques. In a study evaluated by Zhao et al. [16], who reviewed the recent research on machine health monitoring using deep learning models, in which they employed a variety of ML and DL techniques for monitoring cutting tool wear conditions, they acknowledged that, even though DL approaches can automatically learn from raw data, the lack of data samples, the complexity of the operational environment, and the presence of noise in data necessitate feature extraction prior to a deep learning model. Aghazadeh et al. [17] introduced a methodology for tool condition monitoring that uses WPT for the processing stage. Later, spectral subtraction is carried out to remove the noise, and then several health indicators are generated. Finally, CNN is used to predict the tool wear state. Zhou et al.'s [18] research focused on predicting a tool's remaining useful life under variant working conditions. First, relevant wear characteristics were extracted from the data using HHT. Next, the working conditions and the features were combined to create an input matrix, which captured the spatiotemporal relationship under various working conditions. Finally, the input matrix was loaded into an LSTM predictive model to predict the tool's remaining useful life. Wu et al. [19] introduced a tool wear prediction model that employs a two-step approach: singular value decomposition (SVD) for feature extraction and bidirectional long short-term memory (Bi-LSTM) for time series analysis. SVD was used initially to reduce the noise in the raw signal, thereby shortening the input signal length for the Bi-LSTM network and simplifying its complexity. The processed data was then fed into the Bi-LSTM network to capture time series variations across current and previous sampling periods in order to predict tool wear. Shah et al. [20] worked on tool wear prediction in face milling using a singular generative adversarial network (SinGAN) and an LSTM network. Firstly, the Morlets wavelets were used to construct the scalogram from the acoustic emissions and vibration signals. Next, the appropriate wavelet functions were defined using the relative wavelet energy (RWE). To extract the feature vector, SinGAN was employed to generate more scalograms, and then image quality parameters were extracted to construct the

vector. Finally, the extracted feature vector was used to train the LSTM model to predict tool wear. Different LSTM models were used for the sake of comparison: vanilla, stacked, and Bi-LSTM.

Long short-term memory (LSTM) networks have gained popularity as a powerful deep learning architecture for handling sequential data in several industrial engineering applications [21–23]. This algorithm is adopted in this work as a prediction model to estimate the RUL of each cutter, where the extracted HIs are provided to the LSTM network as inputs. Experimental results show unequivocally the efficacy of the proposed approach.

The remainder of this essay is organized as follows: Sect. 2 provides an overview of the EWT theory along with a description of the proposed EWPD, Sect. 3 summarizes the long short-term memory network principal, Sect. 4 introduces the 2010PHM data set, Sect. 5 describes the proposed procedure and discusses the acquired results, and finally, Sect. 6 brings this work to a close.

## 2 Empirical wavelet packet decomposition EWPD

This section presents the developed EWPD method. Here, the raw signal spectrum is segmented based on a new segmentation model to acquire a series of demodulation frequency bands using the theory of EWT.

### 2.1 Empirical wavelet transform

In 2013, a fully adaptive signal-analysis approach called empirical wavelet transform (EWT) has been introduced by Jerome Gilles [24], within the context of non-stationary signal decomposition. The main objective of this method is to decompose a signal into several modes to extract the most useful information that the signal contains. These modes represent the amplitude modulation-frequency modulation (AM-FM) components obtained traditionally after filtering the signal with an adaptive wavelet filter bank. Traditionally, the wavelet group is built by dividing the signal's Fourier spectrum  $[0, \pi]$  into  $N$  continuous segments by identifying local maxima and then taking support boundaries  $\omega_n$  at the intermediate points between subsequent maxima. An example of the filters is shown in Fig. 1 [25].

$\Lambda_n = [\omega_{n-1}, \omega_n]$  denotes each segment, which clearly demonstrates that  $\bigcup_{n=1}^N \Lambda_n = [0, \pi]$ . A transition phase  $T_n$  with a width of  $2\tau_n$  is centered around each  $\omega_n$  in the Fourier axis (Fig. 1 (blue dotted areas)).

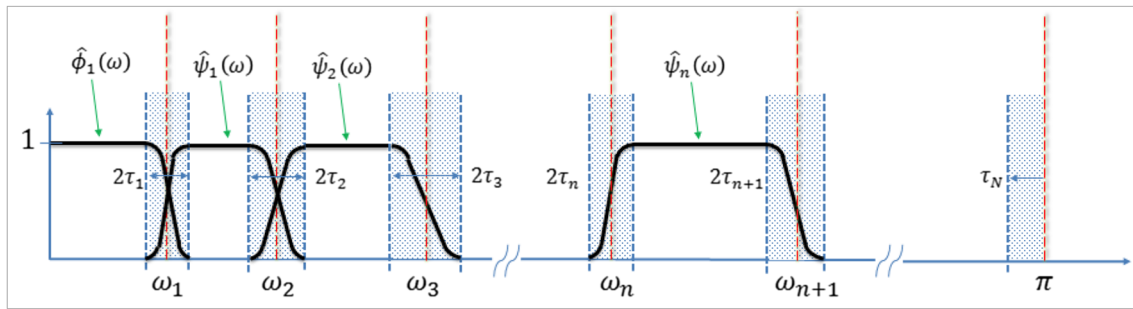


Fig. 1 Empirical wavelet filter bank [25]

The empirical wavelet scaling function and the empirical wavelets are defined as low-pass and band-pass filters by Eqs. 1 and 2 [26], respectively.

$$\hat{\phi}_n(\omega) = \begin{cases} 1 & \text{if } |\omega| \leq \omega_n - \tau_n \\ \cos \left[ \frac{\pi}{2} \beta \left( \frac{1}{2\tau_n} (\omega - \omega_n + \tau_n) \right) \right] & \text{if } \omega_n - \tau_n \leq |\omega| \leq \omega_n + \tau_n \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$$\hat{\psi}_n(\omega) = \begin{cases} 1 & \text{if } \omega_n + \tau_n \leq |\omega| \leq \omega_{n+1} - \tau_{n+1} \\ \cos \left[ \frac{\pi}{2} \beta \left( \frac{1}{2\tau_{n+1}} (\omega - \omega_{n+1} + \tau_{n+1}) \right) \right] & \text{if } \omega_{n+1} - \tau_{n+1} \leq |\omega| \leq \omega_{n+1} + \tau_{n+1} \\ \sin \left[ \frac{\pi}{2} \beta \left( \frac{1}{2\tau_n} (\omega - \omega_n + \tau_n) \right) \right] & \text{if } \omega_n - \tau_n \leq |\omega| \leq \omega_n + \tau_n \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

To ensure that the aforementioned functions form a tight frame, the right choice of  $\tau_n$  is important. The simplest way to choose this parameter is to make it proportional to  $\omega_n$  :  $\tau_n = \gamma \omega_n$  where  $0 < \gamma < 1$ .

$\beta(x)$  is an arbitrary function defined as follows [26]:

$$\beta(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \beta(x) + \beta(1-x) = 1 \quad \forall x \in [0, 1] \\ 1 & \text{if } x \geq 1 \end{cases} \quad (3)$$

The most commonly used function that satisfies these properties is given by the following:

$$\beta(x) = x^4 (35 - 84x + 70x^2 - 20x^3) \quad (4)$$

Equations 5 and 6 define the approximation and detail coefficients that are obtained from the inner product of the processed signal with the scaling and empirical wavelet functions.

$$\mathcal{W}_f^e(0, t) = \langle f, \phi_1 \rangle = \int f(\tau) \overline{\phi_1(\tau - t)} d\tau \quad (5)$$

$$\mathcal{W}_f^e(n, t) = \langle f, \psi_n \rangle = \int f(\tau) \overline{\psi_n(\tau - t)} d\tau = \left( \hat{f}(\omega) \overline{\hat{\psi}_n(\omega)} \right)^\vee \quad (6)$$

By inverting the EWT, the signal is reconstructed as follows:

$$f(t) = \mathcal{W}_f^e(0, t) * \phi_1(t) + \sum_{n=1}^N \mathcal{W}_f^e(n, t) * \psi_n(t) \quad (7)$$

Equation 7 shows that the input signal is decomposed into empirical modes,  $f_k$ , which are given by the following:

$$f_0(t) = \mathcal{W}_f^e(0, t) * \phi_1(t) \quad (8)$$

$$f_k(t) = \mathcal{W}_f^e(k, t) * \psi_k(t) \quad (9)$$

### 2.2 Empirical wavelet packet decomposition

Conventional segmentation models like binary tree [14] and 1/3 binary tree proposed by Antoni [15], shown in Fig. 2 a and b respectively, have an insufficient segmentation accuracy. Many frequency bands, particularly those around the red lines in the prior figures, cannot be explored due to subdivision constraints. Consequently, the signal characteristics are not fully considered. To overcome this shortcoming, we propose the EWPD that provides a new segmentation model of the Fourier axis. This is represented in Fig. 2c.

The EWPD can be explained through the following steps:

1. Introduce the two parameters, the scale  $h$ , and the initial boundary values IBV.  $h \in \mathbb{N}^+$  controls the number of wavelet filters constructed by Eqs. 1 and 2 using the  $IBV$ , given in Eq. 10, to fix the limit of each segment  $\omega_n$ .

$$IBV_h^j = \left( \frac{j}{h+1} \right) \pi \left( \frac{L_s}{2\pi} \right) \quad (10)$$

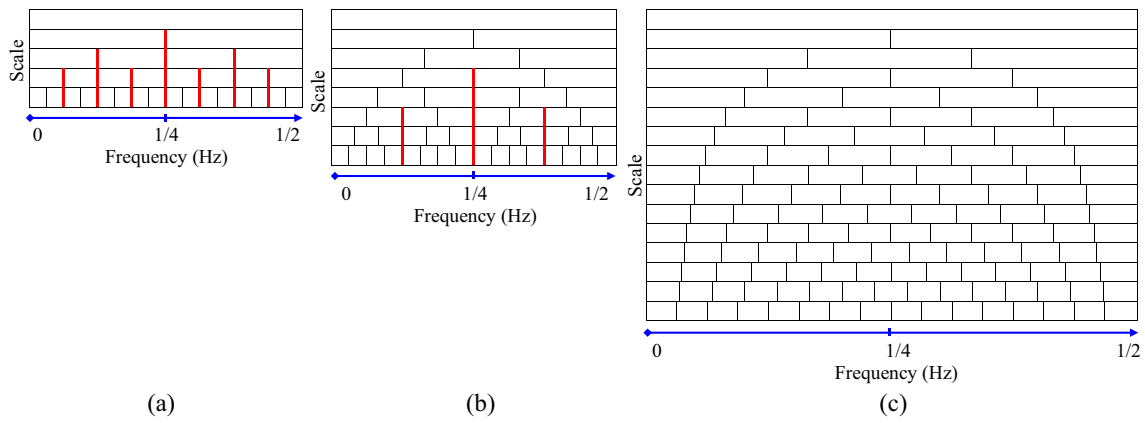


Fig. 2 Binary tree structure (a), 1/3 binary tree structure (b), and EWPD structure (c)

where  $j$  is the boundary index. It is given as  $j = [1, 2, 3, \dots, h]$ , and  $L_s$  is the signal length.

It is worth noting that for a given scale value, let us take  $h = 10$ , a set of  $IBV$  is calculated for every  $j \in (1, 2, \dots, 10)$

2. Adapt the initial set of boundaries to the analyzed signal by computing some neighborhood, as  $[IBV_h^j - \epsilon, IBV_h^j + \epsilon]$ . The objective behind this action is to preserve the information contained in the signal, i.e., by avoiding amplitude attenuation of harmonics in cases where the computed  $IBV$  matches with their frequencies. The neighborhood is chosen to be 20 times the frequency step, defined as  $20 * \frac{L_s}{F_s}$ , where  $F_s$  is the sampling frequency.
3. Detect the global minima in each neighborhood as  $min(FFT(f(IBV_h^j - \epsilon, IBV_h^j + \epsilon)))$ , where  $f$  is the processed signal. Then, redefine the boundary  $\omega_n$  at the corresponding frequency.

The segmentation prescribed in the above steps is illustrated in Fig. 3.

4. Build the wavelet filter bank based on the new Fourier segments. This operation is capable of extracting a predefined number of modes  $f_h^k(t)$  from the signal  $f(t)$  in each level, and thus,

$$f(t) = \sum_{k=1}^{N_m} f_h^k(t) \tag{11}$$

$k$  represents the mode reference, and  $N_m$  is the total number of modes in each level, where  $N_m = h + 1$ .

### 3 Long short-term memory network

Researchers have been exploring deep learning DL techniques recently, owing to their superior performance and

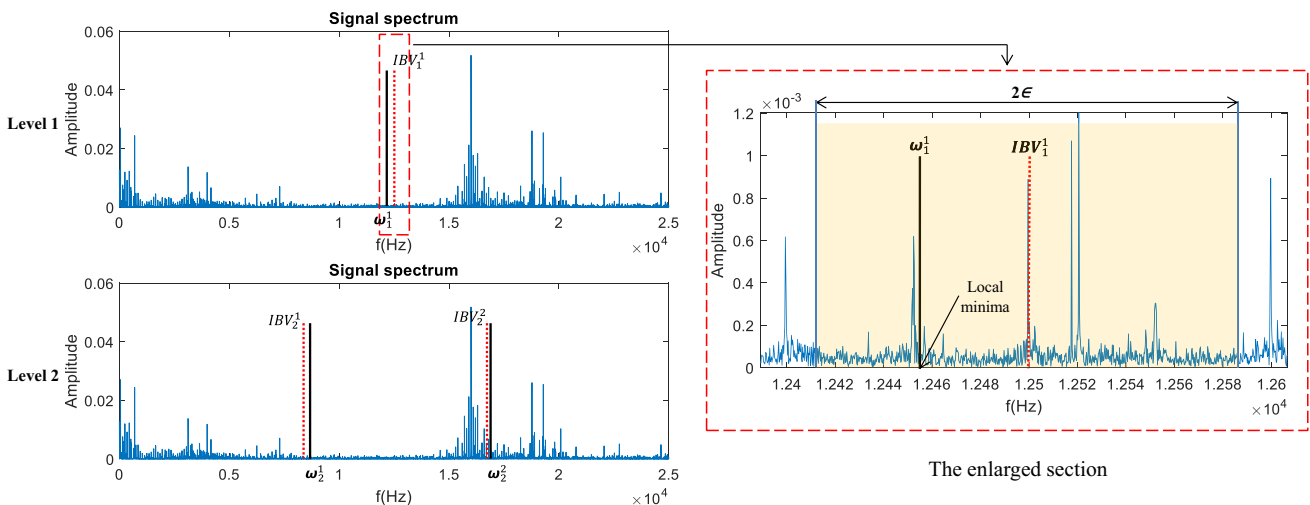


Fig. 3 EWPD segmentation for level 1 and level 2



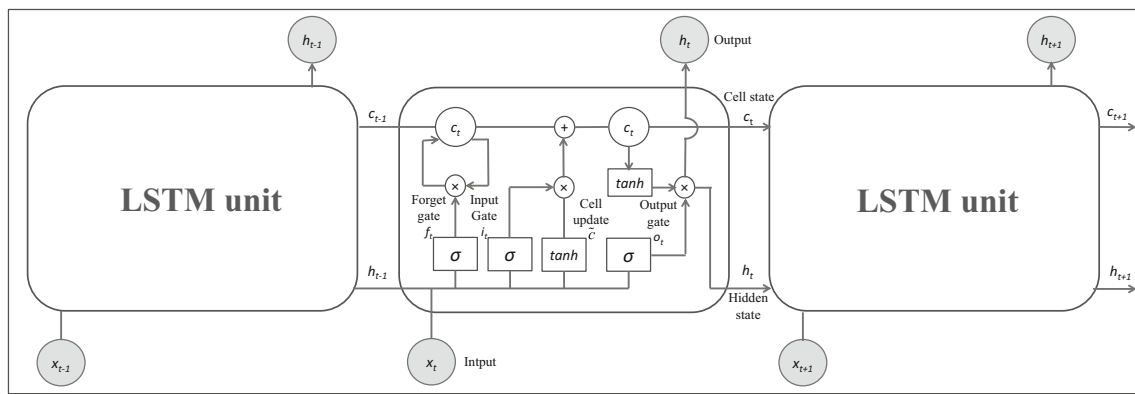


Fig. 4 Schematic of the LSTM network [28]

efficiency in a variety of fields. Recurrent neural network (RNN) is one of the deep learning techniques that is specifically designed for addressing sequential problems, where the primary objective is to memorize long-term dependencies. However, throughout long time series, RNNs fail to achieve their main goal due to the vanishing or exploding gradient problem. The RNN-LSTM network [27] was developed to overcome this problem. Unlike classic RNNs, the LSTM unit is fitted with a cell and a gating system as illustrated in Fig. 4 [28], which control the information flow (cell state, output), enabling the network to learn from large sequences of data. The gating unit has three gates: the forget gate, the input gate, and the output gate.

Based on the input  $x_t$  and the previous cell's hidden state  $h_{t-1}$ , the forget gate refreshes the memory cell by altering the weight of the self-loop cell state in the following way:

$$f_t = \sigma(\mathcal{W}_f x_t + \mathcal{U}_f h_{t-1} + b_f) \quad (12)$$

As depicted in the equation below, the input gate is in charge of commanding the information intended for the memory cell:

$$i_t = \sigma(\mathcal{W}_i x_t + \mathcal{U}_i h_{t-1} + b_i) \quad (13)$$

The output gate determines the weight of the LSTM cell output:

$$o_t = \sigma(\mathcal{W}_o x_t + \mathcal{U}_o h_{t-1} + b_o) \quad (14)$$

The input activation vector updates the weight of the memory cell as follows:

$$\tilde{c}_t = \tanh(\mathcal{W}_c x_t + \mathcal{U}_c h_{t-1} + b_c) \quad (15)$$

Finally, the cell state of the LSTM unit is updated as established in Eq. 16, then the unit output is given by Eq. 17:

$$c_t = f_t c_{t-1} + i_t \tilde{c}_t \quad (16)$$

$$h_t = o_t \tanh(c_t) \quad (17)$$

where  $\mathcal{W}$  and  $\mathcal{U}$  are the input and recurrent weight matrices for the gating unit and the cell respectively.  $b$  denotes the biased vectors of each gate and the cell as well.  $f_t$ ,  $i_t$ , and  $o_t$  represent the forget, input, and output gates activation vectors.  $c_t$ ,  $\tilde{c}_t$ , and  $h_t$  indicate the internal cell state which is the long-term state [29], the cell update, and the current output which is the short-term state [29], respectively.  $\sigma$  and  $\tanh$  are the sigmoid and the tangent hyperbolic activation functions.

The LSTM network's previously described gating architecture provided it with strong learning potential for sequential problems, making the model adaptable for processing, forecasting, predicting, and classifying time series data.

## 4 Experimental setup

This section includes a presentation of the dataset used in this study as well as the experimental configuration for the proposed method.

### 4.1 Dataset details

The high-speed CNC milling machine platform used to evaluate the effectiveness of the proposed method is illustrated in Fig. 5 [30]. The data is presented on the "prognostic data challenge 2010" database [31]. In this experiment, a 3-mm ball nose tungsten carbide end mill (stainless steel, HRC52) [32] with three flutes was tested. Table 1 lists the cutting parameters.

The database is constructed using three sensor types. A Kistler quartz three-axis dynamometer was mounted in the machining table, to measure the forces induced by the cutting tool in  $X$ ,  $Y$ , and  $Z$  directions. Three Kistler piezo accelerometers were installed in the workpiece to get the three-axis vibrations generated by the milling process. Last,

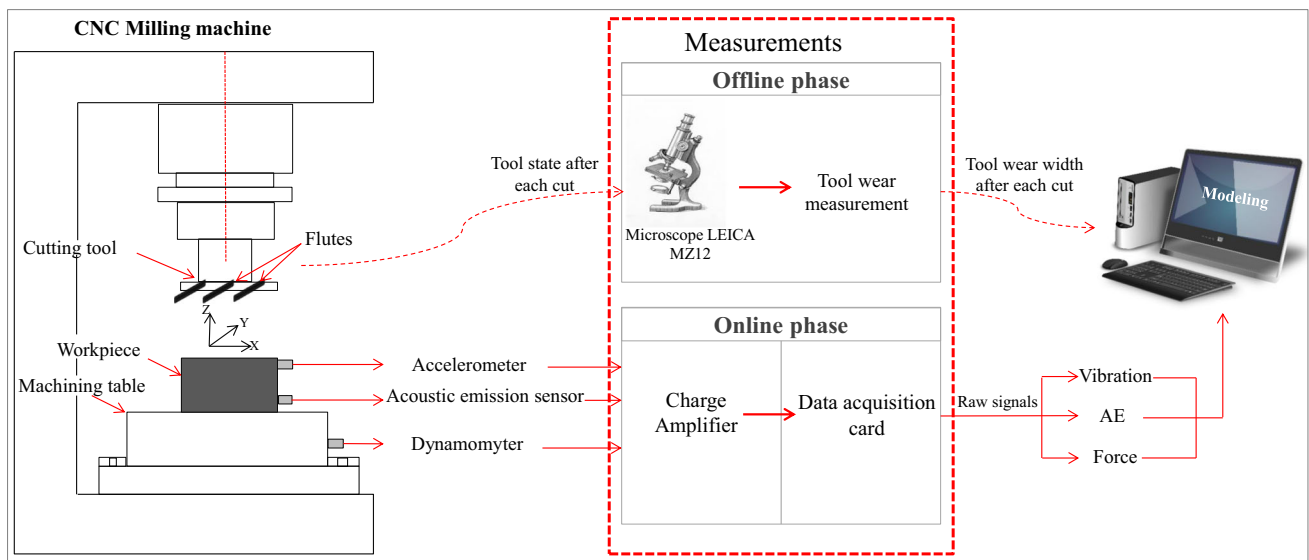


Fig. 5 Dataset collection process for the 2010PHM data challenge [30]

a Kistler acoustic emission (AE) sensor was fixed as well in the working piece, to get the high-frequency stress wave, which represents the surface movement of the working piece in the machining process [32]. Besides, the actual tool’s flank wears were measured, offline, using a microscope after each cut cycle. This information serves as a benchmark to evaluate the predicted RUL. For more details, please refer to [31].

### 4.2 Experimental configuration

Six identical cutters, whose characteristics are mentioned previously, submitted to invariant milling conditions were employed in the experiment. The data set collected is divided into six records (given as C1 to C6), each of which contains 315 files, indicating the cut cycles for each cutting tool. Each of these data acquisition files is in the (.csv) format and contains seven channels. Each channel represents a time series related to a single sensor acquisition for over 200 thousand acquisition points (different lengths for each cut). Table 2 defines the channels.

As regards the real cutting tool wear measured by the Microscope, they are available on the experiment platform [31] for only three of the six cutters used throughout the milling process which are C1, C4, and C6. Subsequently, only their corresponding recordings were considered for analysis. Indeed, the wear value is given for each flute after finishing each surface. As the quality of the machined surface

is important, the maximum wear across the flutes is determined as the wear limit one could safely achieve for any flute. Figure 6 displays the wear after each cut of C1’s flutes, with the maximum wear highlighted.

## 5 Methodology, results, and discussion

This section outlines the strategy employed for determining the cutting tool’s degradation behavior and remaining useful life. It gives, also, a discussion of the obtained results.

Figure 7 illustrates the flowchart of the proposed method, serving as a visual guide for discussing the subsequent steps: feature extraction, feature evaluation, selection, and RUL estimation.

### 5.1 Feature extraction

Our method’s primary purpose is to monitor the cutting tool’s degradation in terms of flank wear. To do that properly, the data acquisition stage is the first focus, as it is crucial in the execution of the monitoring. Therefore, the data measured by the three sensor types was first considered and processed by the EWPD. For each cutter, i.e., C1, C4, and C6, the measured signals are decomposed into several modes for a total of 29 levels. An example of the decomposition is presented in Fig. 8. Although the level number is user-defined, it is rec-

Table 1 Operating conditions

Spindle speed of cutter	Feed rate	Radial depth of cut	Axial depth of cut	Sampling frequency
10,400 rpm	1555 mm/min	0.125 mm	0.2 mm	50 kHz/channel

**Table 2** Organization of the data acquisition file

Channel	Channel1	Channel2	Channel3	Channel4	Channel5	Channel6	Channel7
Sensor	Force(N) in X-axis	Force(N) in Y-axis	Force(N) in Z-axis	Vibration (g) in X-axis	Vibration (g) in Y-axis	Vibration (g) in Z-axis	AE-RMS (V)

Note:  $X$ ,  $Y$ , and  $Z$  represent the axis of the first, second, and third dimensions in the space, respectively

ommended to set it proportional to the sampling frequency. The higher it is, the more a higher level number can be chosen. In this way, incomplete information coming from the very narrow-band filters is abandoned.

Following the decomposition, the focus shifts to feature extraction, aimed at acquiring valuable and meaningful information about the cutter's degradation. In this paper, a new feature extraction scheme is developed and detailed in Algorithm 1.

The foundational aspects of the algorithm rely on statistical measures, commonly called time-domain features (TDFs), calculated for each mode along the 29 levels. Certainly, using all TDFs may seem impractical for use. Therefore, only root mean square, standard deviation, and variance are chosen whose equations are given in Table 3. In short, starting from the premise that monotonic HIs reflect more of the cutting tool degradation and especially help the deep learning model learn quickly and efficiently. The computed features, from the empirical modes, for each cutting cycle, are stacked in a vector, defined as  $TF^i$ , from which we proceed to a clever selection of features of interest. This is achieved by simply tracking the minimum value in the vector  $TF^i$  that is just slightly greater than that in the vector  $TF^{i-1}$ . Figure 9 presents the processing results of the acoustic emission, vibration, and force signals. Figure 9 a–c depicts the RMS-EWPD-derived features. The STD-EWPD-derived features are represented in Fig. 9 d–f, while the VAR-EWPD-

### Algorithm 1 Feature extraction.

---

**Require:**  $N_s$ : Number of signals  
 $N_l$ : Number of levels  
**Ensure:** Health Indicator ( $HI$ )

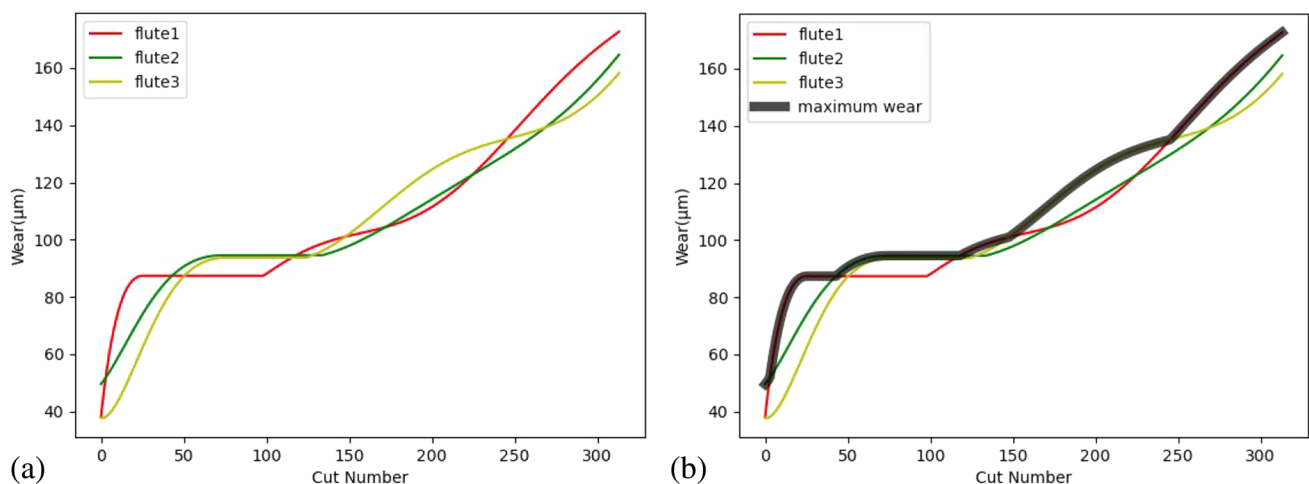
```

1: for  $i \leftarrow 1$  to  $N_s$  do
2:   Initialize the temporal feature vector:  $TF^i$ 
3:   for  $h \leftarrow 1$  to  $N_l$  do
4:     for  $j \leftarrow 1$  to  $h$  do
5:       Get  $\omega_h^j$  ▷ as indicated in Section 2.2
6:     end for
7:     Filter the signal  $f_i$  using EWT filter bank
8:     Precise the number of modes as:  $N_m \leftarrow h + 1$ 
9:     for  $k \leftarrow 1$  to  $N_m$  do
10:      Calculate temporal feature:  $A_h^k$ 
11:      Concatenate the result as  $[TF^i, A_h^k]: TF^i$ 
12:    end for
13:  end for
14:  if  $i \leq 1$  then
15:     $HI(i) \leftarrow \min(TF^i)$ 
16:  else
17:    Try:  $HI(i) \leftarrow \min(TF^i(TF^i > TF^{i-1}))$ 
18:    Except:  $HI(i) \leftarrow \max(TF^i)$ 
19:  end if
20: end for

```

---

derived features are drawn in Fig. 9 g–i. It can be seen that the resulting HIs show no random fluctuations and misleading trends. They exhibit a monotonically increasing behavior throughout the cutting tools' lifetime that is easy to model by LSTM for RUL estimation.



**Fig. 6** Wear across cutter 1's flutes (a). Maximum wear across cutter 1's flutes (b)



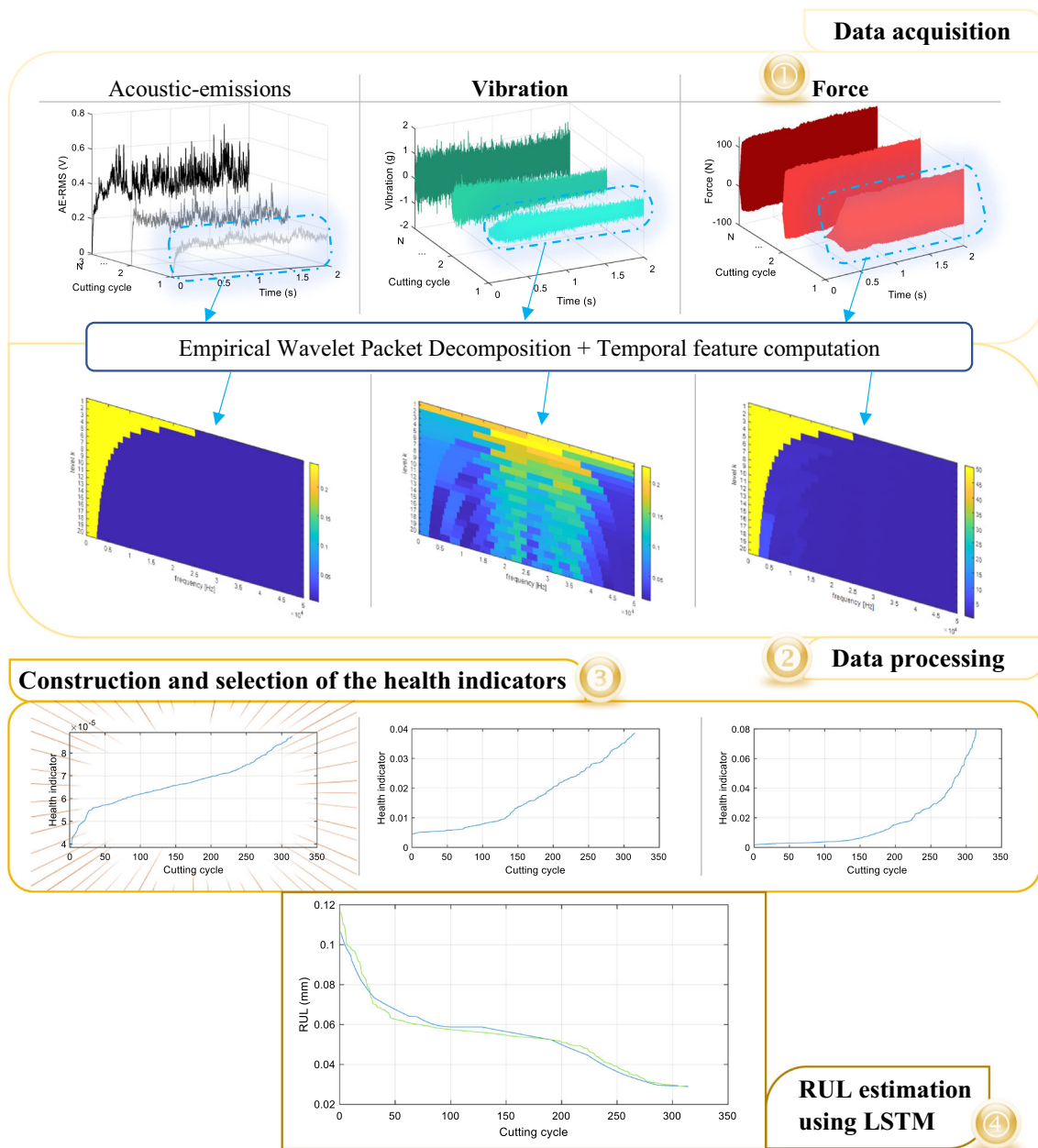


Fig. 7 Flowchart of the proposed method

### 5.2 Feature evaluation and selection

When creating a predictive model, it is necessary to reduce the number of input variables, which helps lower the modeling’s computing cost and boosts the model’s performance. Thus, this step is intended to evaluate the obtained features and quantify their performance, to select the most appropriate sensor results that provide the best presentation of the cutter’s degradation. To do so, Coble J. B. [33] has proposed fundamental metrics defining an ideal prognostic parameter, such as monotonicity  $M$  and trendability  $T$ . The following equations [34] formalize these

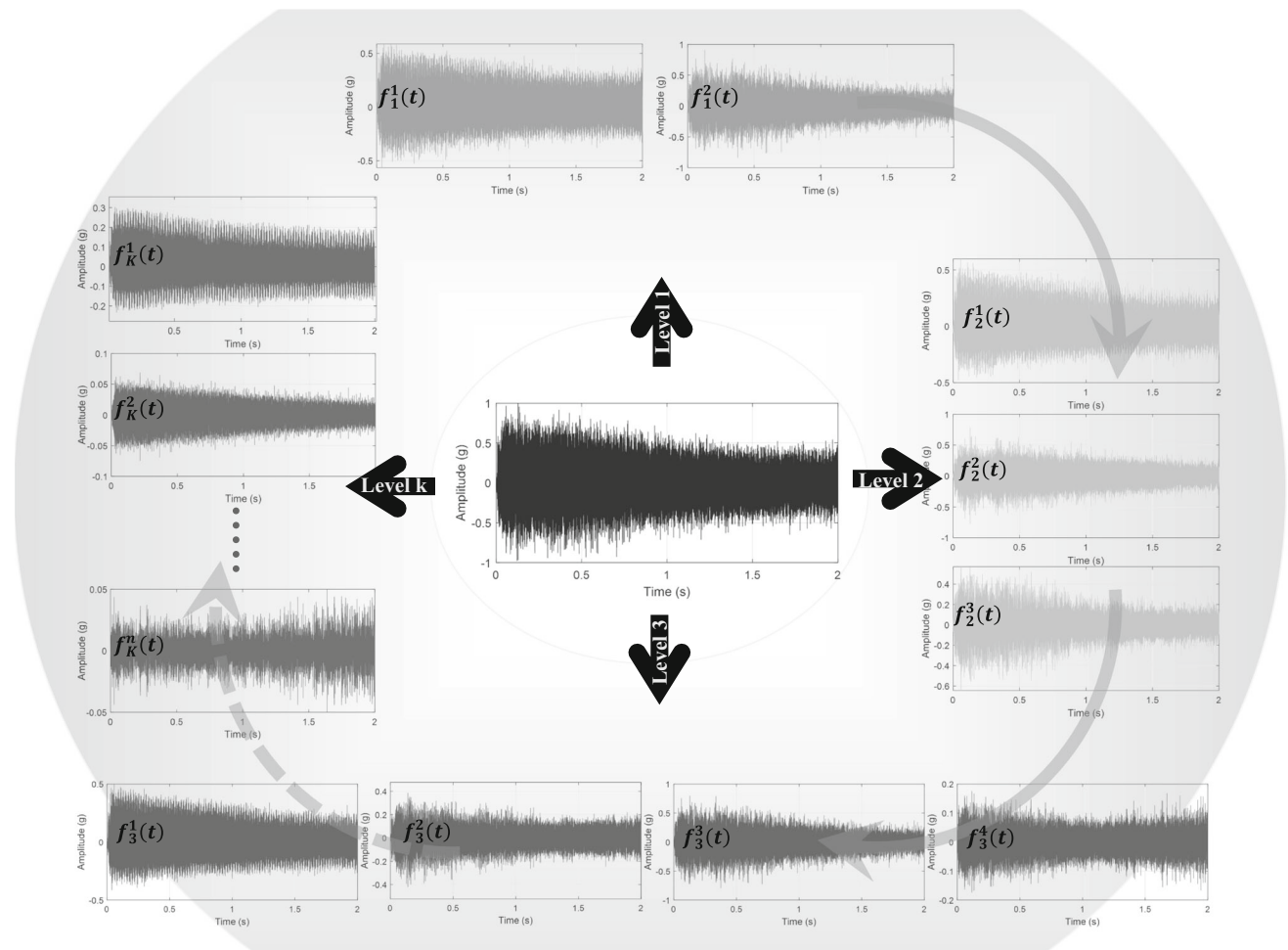
metrics:

$$M = \text{mean} \left( \left| \frac{\text{positive}(\text{diff}(y_i)) - \text{negative}(\text{diff}(y_i))}{n - 1} \right| \right) \tag{18}$$

Where,  $i = 1, 2, 3, \dots, n$ .  $n$  is the length of the indicator  $y$ .

$$T = \min (|\text{corrcoef}(y_i, y_j)|) \tag{19}$$

where  $i, j = 1, 2, 3, \dots, m$ .  $m$  is the total number of monitored systems.



**Fig. 8** Empirical wavelet packet decomposition process

The term “monotonicity” refers to the indicator’s primary positive or negative trend [34]. This metric is considered one of the most precise measures for assessing the efficacy of the health indicator since the cutting tool’s degradation is related to wear progression, an irreversible phenomenon. A suitable condition indicator typically has a monotonic trend as a system gets closer to failure [35].

Trendability is defined as the degree to which a population of systems’ parameters have the same underlying shape and can be characterized by the same functional form [35].

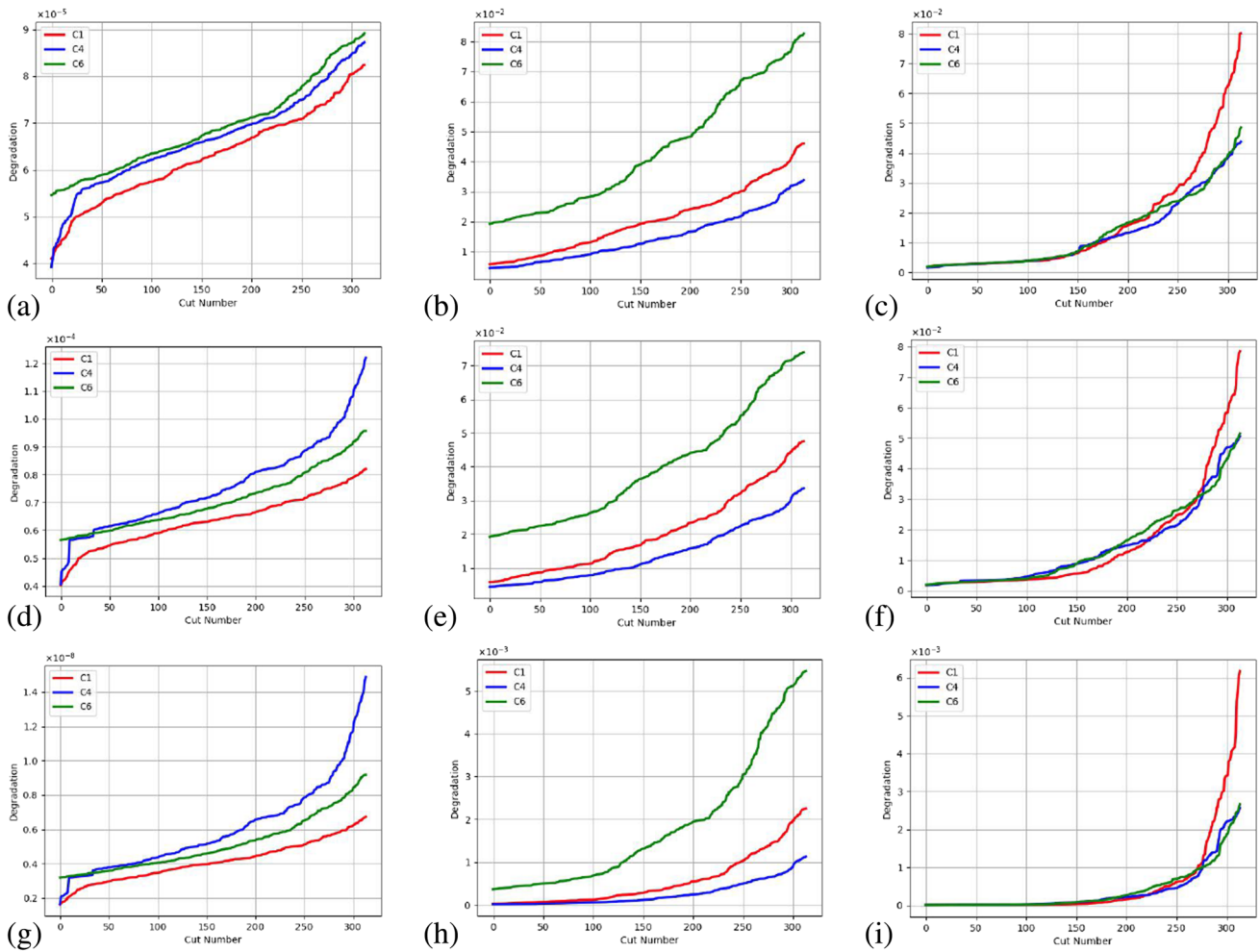
Each of these measures would have a value ranging from 0 to 1, with 1 representing an optimal score on that statistic and 0 denoting that the condition indicator is inappropriate for the depiction of the component degradation [35].

The results in Tables 4, 5, and 6 reveal that all the extracted degradation indicators exhibit a fully monotonic trend ( $M = 1$ ), making them appropriate for representing the degradation phenomenon. Besides that, trendability results clearly illustrate that acoustic-emissions-sensor-derived features are more trendable in all cases, indicating that they are better correlated with the progression of the flank wear. Consequently, acoustic emissions sensor-based results are selected to be employed for the remainder of this work.

Next, for comparison purposes, the same data was processed by the WPT method, and naturally, the resulting modes are analyzed with the same statistical measures. The WPT-derived features, represented in Fig. 10, exhibit spurious fluctuations. The trends are very difficult to model, so poor RUL estimates are expected. The average values of monotonicity and trendability for both methods’ health indi-

**Table 3** Time-domain features

Time-domain feature	Equation
Root mean square (RMS)	$\sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$
Standard deviation (SD)	$\sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x}_i)^2}$
Variance (VAR)	$\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x}_i)^2$



**Fig. 9** RMS-EWPD features derived from acoustic emission, vibration, and force sensors (a–c), SD-EWPD features (d–f), and VAR-EWPD features (g–i)

caters are shown in Table 7. Pearson’s correlation coefficient (r) between the actual wear and HIs is also applied to each method to indicate the most correlated features with tool wear that are not inflated by other cutting circumstances. The average value of Pearson’s correlation is presented in Table 7. The results show that, regardless of cutting parameters, the recommended features have the strongest correlation with wear conditions.

**Table 4** RMS-EWPD features fitness

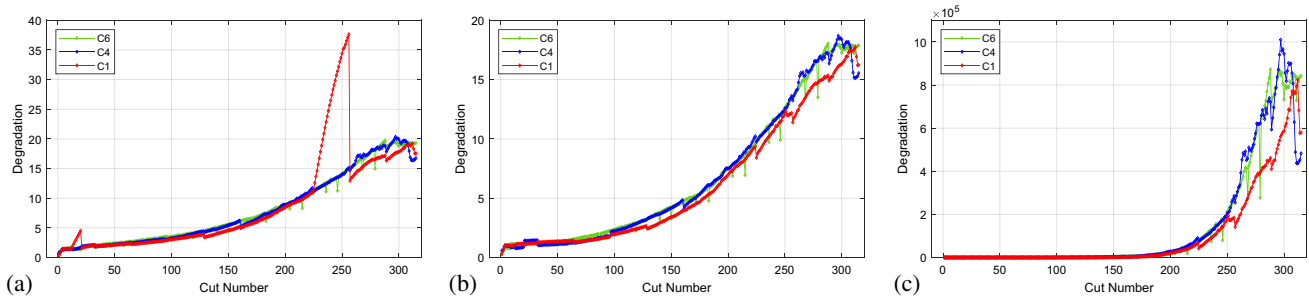
Sensor	C1		C4		C6	
	M	T	M	T	M	T
Acoustic emission	1	0.9894	1	0.9754	1	0.9808
Vibration	1	0.9836	1	0.9732	1	0.9810
Force	1	0.8564	1	0.9156	1	0.9294

**Table 5** SD-EWPD features fitness

Sensor	C1		C4		C6	
	M	T	M	T	M	T
Acoustic emission	1	0.9842	1	0.9614	1	0.9739
Vibration	1	0.9701	1	0.9654	1	0.9729
Force	1	0.8283	1	0.8991	1	0.9276

**Table 6** VAR-EWPD features fitness

Sensor	C1		C4		C6	
	M	T	M	T	M	T
Acoustic emission	1	0.9862	1	0.9282	1	0.9560
Vibration	1	0.8959	1	0.8909	1	0.9254
Force	1	0.6525	1	0.7436	1	0.7965



**Fig. 10** RMS-WPT-derived features (a), SD-WPT-derived features (b), and VAR-WPT-derived features (c)

### 5.3 Remaining useful life estimation

As the RUL estimation is the major target in this paper, this section is intended to explain the adopted strategy to determine the cutting tools RULs using the LSTM network.

Many studies claim to do prognostics by feeding directly the raw data, usually with a high dimension, into the deep learning model. But working with such data is difficult because it can be computationally expensive. In contrast, feature extraction may help to reduce the complexity of the data, eliminate irrelevant and redundant information, and make the model focus on the most important information that can contribute to a better RUL prediction.

As shown in Fig. 11, RUL estimation can be summarized in two main steps:

- Preparing input data for regression using LSTM
- Building the LSTM network

Concerning the input data, six combinations were explored to validate the model's performance and ensure its ability to generalize to different data distributions. They are explained in Table 8. Before training, three HIs for each cutter, i.e., RMS-EWPD, SD-EWPD, and VAR-EWPD-derived indicators, are stacked into an input matrix of the shape of (315, 3). These data are normalized with the min-max scaler, i.e., they are transformed so that they lie in the range [0, 1]. Normalizing data is a pre-processing step that is commonly used in deep learning. It helps improve the model's convergence

during training, reduces the risk of overfitting, and improves the model's generalization ability. In the case of large data, it helps to ensure that the model is not biased towards any particular feature and can learn the relationships between the features and the target more efficiently [36]. The min-max scaler is given in Eq. 20:

$$HI_{scaled} = \frac{HI - \min(HI)}{\max(HI) - \min(HI)} \quad (20)$$

The LSTM architecture is designed to receive input data in a three-dimensional shape, i.e., the number of samples, time steps, and features [37]. So that the network can analyze the input data in the proper sequence and identify patterns in the data.

The normalized data are reshaped to three dimensions, specifically (samples, time steps, features), to match the algorithm's requirements. As a result, the input matrix is reshaped to 315, 3, and 1, where the number of rows in the original input matrix is treated as the sample number, the number of columns in the same matrix is treated as the number of time steps, and they are stacked in one feature.

In the second step, the model is built, with comprehensive architectural specifications provided in Table 9. The weights of the cell update and the gating unit are adjusted in the training state using Adam optimizer [38]. For iteratively updating network weights, this algorithm is employed as a substitute to typical stochastic gradient descent such as AdaGrad and RMSProp that keep all weight updates at the same learning rate [39]. This internal optimization algorithm is known to be reliable in the selection of hyperparameters [40].

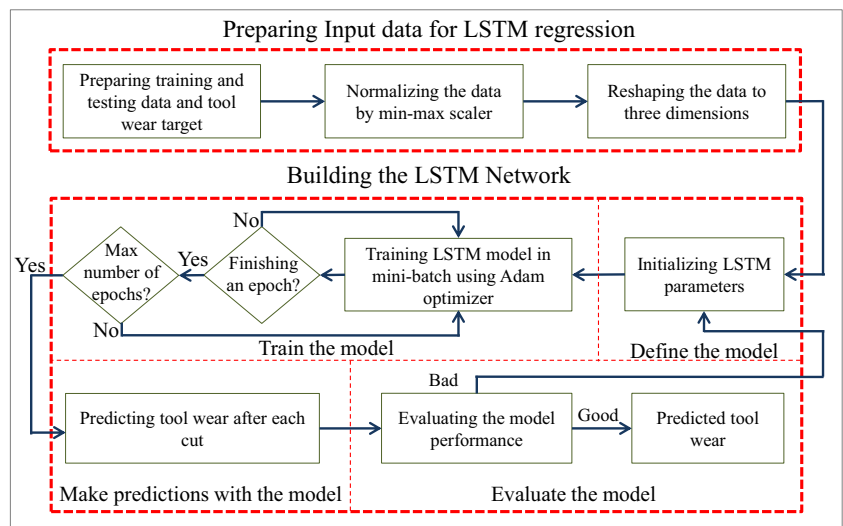
A loss evaluation function is performed after each training epoch using root mean square error (RMSE). This error function is typically reserved for regression prediction models [37], and it is defined as follows:

$$RMSE = \sqrt{\frac{1}{K} \sum_{i=1}^K (R_i - P_i)^2} \quad (21)$$

**Table 7** Performance comparison metrics for WPT-derived features and EWPD-derived features

Method	Average value		
	$M$	$T$	$r$
WPT-derived features	0.6826	0.9446	0.9133
EWPD-derived features	1	0.9706	0.9731

**Fig. 11** Diagram of RUL prediction by LSTM



where  $R_i$  and  $P_i$  are the real and predicted values respectively.

The RMSE decreases over the training until obtaining the best possible fitted results.

The predicted remaining useful lives of the cutters C1, C4, and C6 by the LSTM network, trained by the proposed HIs, together with the real RULs are represented in Fig. 12. It can be seen that the predictions follow a similar pattern as ground truths and have close correspondence. These results are confirmed by comparison with the RULs, estimated by LSTM and trained with WPT-derived features. The accuracy (Acc) (Eq. 23) metric, the mean absolute error (MAE) (Eq. 22), and the root mean square error RMSE were used for error quantification.

As shown in Table 10, for each cutter, both RMSE and MAE values appear to be lower by using the proposed methodology. Since lower values indicate a better fitting of the ground truth, we can infer that the proposed methodology delivers the best results. The accuracy results confirm our method’s suitability for RUL prediction with results close to 1.

$$MAE = \frac{1}{K} \sum_{i=1}^K |R_i - P_i| \tag{22}$$

**Table 8** Training and testing sets configuration

Testing set	C1	C4	C6			
Training set	C4	C6	C1	C6	C1	C4

and

$$A_{cc} = \frac{1}{K} \sum_{i=1}^K \exp\left(-\frac{|R_i - P_i|}{R_i}\right) \tag{23}$$

where  $R_i$  and  $P_i$  are the real RUL and the predicted RUL respectively.

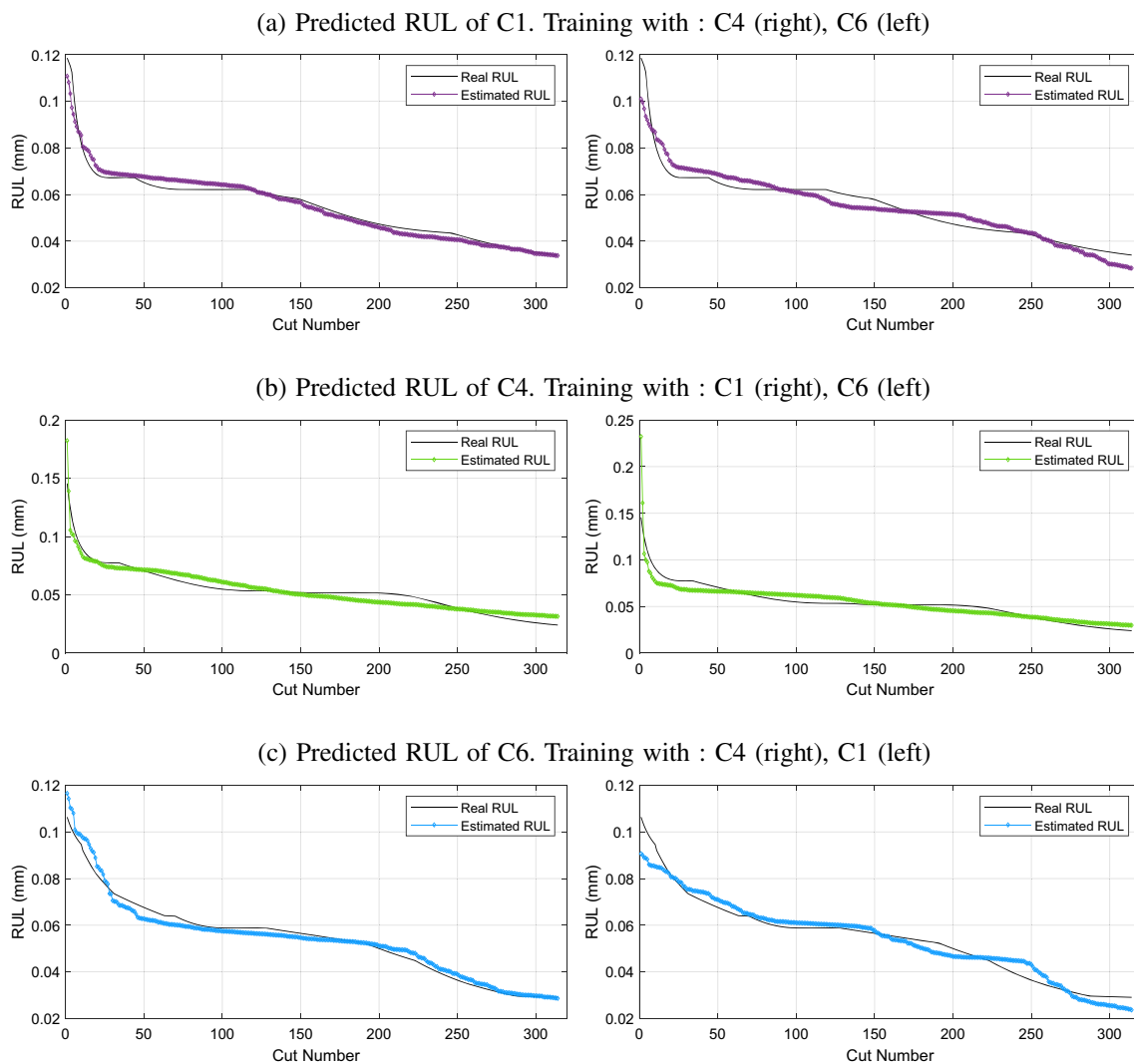
## 6 Conclusion

In this paper, a signal processing approach was introduced, complemented by artificial intelligence for tracking cutting tools’ degradation and estimating their remaining useful lives. The approach, termed empirical wavelet packet decomposition, provides a novel segmentation technique of the signal’s Fourier spectrum, enabling a more comprehensive exploration of frequency bands. Leveraging these segmented

**Table 9** Model architecture details

Layer (type)	Parameters
Input layer	Input shape (315, 3)
LSTM	256 hidden units
Batch normalization	momentum = 0.99, epsilon = 0.001
Activation function	Relu
LSTM	64 hidden units
Batch normalization	momentum = 0.99, epsilon = 0.001
Activation function	Tanh
Dropout	Dropout rate of 0.2
Dense	1 unit





**Fig. 12** RUL prediction of cutting tools

sub-bands, a novel and robust health indicator is constructed. Furthermore, a widespread reference method called WPT was used for comparison. The observations of this study can be summarized as follows:

- The proposed EWPD exhibits superior segmentation performance when compared to traditional binary trees and 1/3 binary tree structures, allowing for a more precise highlighting of the signal's unique characteristics and properties.
- The proposed feature extraction scheme, designed to exploit richer filter banks, yields features that consistently exhibit a monotonically increasing trend throughout the entire lifespan of the cutting tools.
- The acoustic-emissions-sensor-derived and selected features were found to be better correlated with the pro-

gression of the cutting tool flank wear; thus, they are considered the HIs.

- The LSTM network was used to learn the HIs and estimate the RULs. The predicted RULs accurately follow the ground truth patterns, proving the high performance of the proposed method.

The developments in this paper show that a thorough focus on feature extraction and selection is important to ensure the construction of an effective health indicator that is easy to model. While the proposed method exhibits promising results, some limitations should be addressed, for example, the decomposition level in EWPD needs to be adaptively determined. Moreover, there is a need for conducting additional experiments across diverse datasets to validate its generalization.

**Table 10** Comparison of prediction performance using two different inputs for the LSTM network

Method	RMSE		MAE		$A_{ec}$													
	C1	C4	C1	C4	C1	C4												
LSTM-based WPT	0.0077	0.0070	0.0095	0.01	0.0072	0.0081	0.0639	0.0658	0.0807	0.0858	0.0689	0.0749	0.93	0.92	0.88	0.87	0.91	0.90
LSTM-based EWP	0.0039	0.0025	0.0064	0.0054	0.0024	0.0036	0.0561	0.0442	0.0711	0.0669	0.0468	0.0533	0.94	0.97	0.91	0.91	0.96	0.94

**Author contribution** The preparation and execution of this work involved input from all of the authors; AAB, BM, and DB collected and analyzed the data. All writers provided feedback on earlier iterations of the manuscript after AAB wrote the first draft. The final text has been reviewed and approved by all authors.

## Declarations

**Competing interests** The authors declare no competing interests.

## References

- Sun H, Liu Y, Pan J, Zhang J, Ji W (2020) Enhancing cutting tool sustainability based on remaining useful life prediction. *J Clean Prod* 244:118794. <https://doi.org/10.1016/j.jclepro.2019.118794>
- Liu C, Li Y, Hua J, Lu N, Mou W (2018) Real-time cutting tool state recognition approach based on machining features in NC machining process of complex structural parts. *J Adv Manuf Technol* 97(1):229–241. <https://doi.org/10.1007/s00170-018-1916-y>
- Elattar HM, Elminir HK, Riad A (2016) Prognostics: a literature review. *Complex Intell Syst* 2(2):125–154. <https://doi.org/10.1007/s40747-016-0019-3>
- Gouriveau R, Medjaher K, Zerhouni N (2016) From prognostics and health systems management to predictive maintenance 1: monitoring and prognostics. John Wiley and Sons, New Jersey
- Laddada S, Si-Chaib MO, Benkedjough T, Draï R (2020) Tool wear condition monitoring based on wavelet transform and improved extreme learning machine. *Proc Inst Mech Eng Part C J Mech Eng Sci* 234(5):1057–1068. <https://doi.org/10.1177/0954406219888544>
- Liu Y, Wang X, Lin J, Zhao W (2016) Early chatter detection in gear grinding process using servo feed motor current. *J Adv Manuf Technol* 83(9):1801–1810. <https://doi.org/10.1007/s00170-015-7687-9>
- Wang C, Gan M, Zhu C (2018) Fault feature extraction of rolling element bearings based on wavelet packet transform and sparse representation theory. *J Intell Manuf* 29(4):937–951. <https://doi.org/10.1007/s10845-015-1153-2>
- Ji Y, Wang X, Liu Z, Yan Z, Jiao L, Wang D, Wang J (2017) EEMD-based online milling chatter detection by fractal dimension and power spectral entropy. *J Adv Manuf Technol* 92(1):1185–1200. <https://doi.org/10.1007/s00170-017-0183-7>
- Benkedjough T, Zerhouni N, Rechak S (2018) Tool wear condition monitoring based on continuous wavelet transform and blind source separation. *J Adv Manuf Technol* 97(9):3311–3323. <https://doi.org/10.1007/s00170-018-2018-6>
- Gougam F, Chemseddine R, Benazzouz D, Benagoune K, Zerhouni N (2021) Fault prognostics of rolling element bearing based on feature extraction and supervised machine learning: application to shaft wind turbine gearbox using vibration signal. *Proc Inst Mech Eng Part C J Mech Eng Sci* 235(20):5186–5197. <https://doi.org/10.1177/0954406220976154>
- Liao Z, Gao D, Lu Y, Lv Z (2016) Multi-scale hybrid HMM for tool wear condition monitoring. *J Adv Manuf Technol* 84(9):2437–2448. <https://doi.org/10.1007/s00170-015-7895-3>
- Shen B, Gui Y, Chen B, Lin Z, Liu Q, Liu Q (2020) Application of spindle power signals in tool condition monitoring based on HHT algorithm. *J Adv Manuf Technol* 106(3):1385–1395. <https://doi.org/10.1007/s00170-019-04684-0>
- Segreto T, D'Addona D, Teti R (2020) Tool wear estimation in turning of Inconel 718 based on wavelet sensor signal analysis and machine learning paradigms. *Prod Eng* 14(5):693–705. <https://doi.org/10.1007/s11740-020-00989-2>
- Gao RX, Yan R (2011) Wavelet packet transform. Wavelets: theory and applications for manufacturing. Springer, Boston, pp 69–81. [https://doi.org/10.1007/978-1-4419-1545-0\\_5](https://doi.org/10.1007/978-1-4419-1545-0_5)
- Antoni J (2007) Fast computation of the kurtogram for the detection of transient faults. *Mech Syst Signal Process* 21(1):108–124. <https://doi.org/10.1016/j.ymssp.2005.12.002>
- Zhao R, Yan R, Chen Z, Mao K, Wang P, Gao RX (2019) Deep learning and its applications to machine health monitoring. *Mech Syst Signal Process* 115:213–237. <https://doi.org/10.1016/j.ymssp.2018.05.050>
- Aghazadeh F, Tahan A, Thomas M (2018) Tool condition monitoring using spectral subtraction and convolutional neural networks in milling process. *J Adv Manuf Technol* 98(9):3217–3227. <https://doi.org/10.1007/s00170-018-2420-0>
- Zhou J-T, Zhao X, Gao J (2019) Tool remaining useful life prediction method based on LSTM under variable working conditions. *J Adv Manuf Technol* 104(9):4715–4726. <https://doi.org/10.1007/s00170-019-04349-y>
- Wu X, Li J, Jin Y, Zheng S (2020) Modeling and analysis of tool wear prediction based on SVD and BiLSTM. *J Adv Manuf Technol* 106:4391–4399. <https://doi.org/10.1007/s00170-019-04916-3>
- Shah M, Vakharia V, Chaudhari R, Vora J, Pimenov DY, Giasin K (2022) Tool wear prediction in face milling of stainless steel using singular generative adversarial network and LSTM deep learning models. *J Adv Manuf Technol* 121(1–2):723–736. <https://doi.org/10.1007/s00170-022-09356-0>
- Wu X, Li J, Jin Y, Zheng S (2020) Modeling and analysis of tool wear prediction based on SVD and BiLSTM. *J Adv Manuf Technol* 106(9):4391–4399. <https://doi.org/10.1007/s00170-019-04916-3>
- Marani M, Zeinali M, Songmene V, Mechefske CK (2021) Tool wear prediction in high-speed turning of a steel alloy using long short-term memory modelling. *Measurement* 177:109329. <https://doi.org/10.1016/j.measurement.2021.109329>
- Chen X, Zhang B, Gao D (2021) Bearing fault diagnosis base on multiscale CNN and LSTM model. *J Intell Manuf* 32(4):971–987. <https://doi.org/10.1007/s10845-020-01600-2>
- Gilles J (2013) Empirical wavelet transform. *IEEE Trans Signal Process* 61(16):3999–4010. <https://doi.org/10.1109/TSP.2013.2265222>
- Merainani B, Rahmoune C, Benazzouz D, Ould-Bouamama B (2018) A novel gearbox fault feature extraction and classification using Hilbert empirical wavelet transform, singular value decomposition, and SOM neural network. *J Vib Control* 24(12):2512–2531. <https://doi.org/10.1177/1077546316688991>
- Merainani B, Bouzid AA, Ratni A, Benazzouz D (2020) Detection of bearing fault using empirical wavelet transform and S transform methods. Paper presented at the 2020 1st International Conference on Communications, Control Systems and Signal Processing (CCSSP). Algeria, University of Echahid Hamma Lakhdar-El-Oued, pp 453
- Graves A (2012) Long short-term memory. In: Supervised sequence labelling with recurrent neural networks. Studies in Computational Intelligence, vol. 385. Berlin, Heidelberg, Springer, pp 37–45. [https://doi.org/10.1007/978-3-642-24797-2\\_4](https://doi.org/10.1007/978-3-642-24797-2_4)
- Zhang J, Wang P, Yan R, Gao RX (2018) Long short-term memory for machine remaining life prediction. *J Manuf Syst* 48:78–86. <https://doi.org/10.1016/j.jmsy.2018.05.011>
- Sun H, Zhang J, Mo R, Zhang X (2020) In-process tool condition forecasting based on a deep learning method. *Robot Comput Integr Manuf* 64:101924. <https://doi.org/10.1016/j.rcim.2019.101924>
- Wang J, Xie J, Zhao R, Zhang L, Duan L (2017) Multisensory fusion based virtual tool wear sensing for ubiquitous manufacturing. *Robot Comput Integr Manuf* 45:47–58. <https://doi.org/10.1016/j.rcim.2016.05.010>

31. Bird J (2010) 2010 PHM society conference data challenge. <https://www.phmsociety.org/competition/phm/10>
32. Cai W, Zhang W, Hu X, Liu Y (2020) A hybrid information model based on long short-term memory network for tool condition monitoring. *J Intell Manuf* 31(6):1497–1510. <https://doi.org/10.1007/s10845-019-01526-4>
33. Coble JB (2010) Merging data sources to predict remaining useful life—an automated method to identify prognostic parameters. Dissertation. Knoxville: The University of Tennessee
34. Saidi L, Ali JB, Bechhoefer E, Benbouzid M (2017) Wind turbine highspeed shaft bearings health prognosis through a spectral kurtosis-derived indices and SVR. *Appl Acoust* 120:1–8. <https://doi.org/10.1016/j.apacoust.2017.01.005>
35. Coble J, Hines JW (2009) Identifying optimal prognostic parameters from data: a genetic algorithms approach. Paper presented at the Annual Conference of the PHM Society. University of Tennessee
36. Serafeim L (2021) Everything you need to know about min-max normalization in Python. <https://towardsdatascience.com/everything-you-need-to-know-about-min-max-normalization-in-python-b79592732b79/>
37. Brownlee J (2017) Long short-term memory networks with python: develop sequence prediction models with deep learning. Machine Learning Mastery, Austin
38. Kingma DP, Ba J (2015) Adam: a method for stochastic optimization. Paper presented at the 3rd International Conference for Learning Representations, San Diego, 2015
39. Brownlee J (2017) Gentle introduction to the Adam optimization algorithm for deep learning - machine learning mastery. <https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/>
40. Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT Press, Cambridge

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.