Journal of Manufacturing Systems 75 (2024) 163–173

Contents lists available at ScienceDirect

Journal of Manufacturing Systems

journal homepage: www.elsevier.com/locate/jmansys

Technical paper

Digital twin-based anomaly detection for real-time tool condition monitoring in machining

Zepeng Liu [a], Zi-Qiang Lang [b,*], Yufei Gui [b], Yun-Peng Zhu [c], Hatim Laalej [d]

[a] School of Engineering, Newcastle University, NE1 7RU, UK
[b] Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, S10 2TN, UK
[c] School of Engineering and Materials Science, Queen Mary University of London, London, E1 4NS, UK
[d] Advanced Manufacturing Research Centre, University of Sheffield, Sheffield, S10 2TN, UK

ARTICLE INFO

Keywords:
Digital twin
Tool condition monitoring (TCM)
Fault diagnosis
Nonlinear system identification
Nonlinear output frequency response functions (NOFRFs)

ABSTRACT

Real-time tool condition monitoring (TCM) has been emerging as a key technology for smart manufacturing. TCM can improve the dimensional accuracy of products, minimize machine tool downtime, and eliminate scraps and re-work costs. Digital twins offer new opportunities for real-time monitoring of machining processes, which can, in principle, take into account changes in machining processes and operating environments, help understand mechanisms of cutting tool wear, and improve the anomaly detection accuracy and fault diagnosis results. The present study exploits these potential advantages of digital twins and proposes a new digital twin-based anomaly detection framework for real-time TCM in machining. The framework of the digital twin consists of three parts: the physical product, the virtual product and data flow connections. Within this framework of the digital twin, the "physical product" represents the machining processes. The "virtual product" includes a real-time data-driven model representing the dynamic relationship between vibration data measured from machining processes as well as the model frequency features (MFFs)-based diagnostics for cutting tool anomaly detection. The "data flow connections" involve real-time measured vibration data and machine tool numerical controller (NC) signals providing real-time information on machine tool dynamics and various machining processes. The novelty is associated with an innovative integration of real-time data-driven modeling, MFFs extraction, and MFFs and machine tool NC signal-based tool wear diagnostics. This, for the first time, enables the concept of digital twins to be potentially applied to the TCM for complicated dynamic machining processes which, as far as we are aware of, has never been achieved before. Comprehensive field studies have demonstrated the effectiveness of the proposed digital twin-based TCM framework and its potential industrial applications.

1. Introduction

Smart manufacturing, a broad category of manufacturing that employs digital information technology, flexible skilled workforce training and computer-integrated manufacturing, is the central driver of the new industrial revolution and will remain the primary pathway for transforming and upgrading manufacturing in the coming decades [1]. Cutting tools or cutters are one of the keys and essential components for cutting, shaping and removing materials from workpieces in manufacturing. However, cutting tools are often operated in excessive force, high stresses and high-temperature environments, which can increase their failure rates [2]. The failure of cutting tools usually damages the workpieces and entails production disruptions or even production shutdowns, causing economic loss which usually amounts to plenty of times the materials cost [2]. Therefore, tool condition monitoring (TCM) is of critical importance for the reduction of machine tool downtime and the increase of production reliability.

An extensive review of the literature shows that the techniques for TCM generally have two categories, which are direct and indirect approaches. In the direct approach, the machine has to be stopped; and the actual tool wear can be directly measured by deploying the Tool Makers Microscope or Optical Microscope which often causes unnecessary downtime during machining [3]. Conversely, indirect approaches use indirect measurements to deduce tool wear via an empirically determined correlation between the tool wear and measurements [3]. In contrast to direct approaches, indirect approaches may be less accurate, but they also have lower complexity and are more suitable for real-time TCM [3]. The most widely reported indirect TCM approach is the signal analysis techniques shown in Fig. 1(a), which extract

* Corresponding author.
E-mail address: z.lang@sheffield.ac.uk (Z.-Q. Lang).

https://doi.org/10.1016/j.jmsy.2024.06.004
Received 9 September 2023; Received in revised form 17 February 2024; Accepted 11 June 2024
Available online 27 June 2024

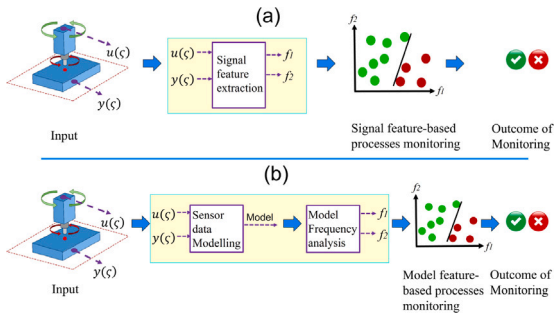

Fig. 1. TCM techniques. (a) signal feature-based TCM and (b) model feature-based TCM [9].

significant signal features from deployed sensing devices, e.g., acoustic emission, cutting force, vibration, sound and power, in order to continuously monitor the actual health status of cutting tools [4]. Various signal feature extraction methods, such as time-domain analysis [5], frequency-domain analysis [6], wavelet packet decomposition [7] and AI-based methods [8], have been extensively researched for application in fault diagnosis of cutting tools, but only a few of them have been implemented in real-world applications. This is because signal feature-based methods have two fundamental drawbacks and limitations. First, signal features have low adaptability to overcome the impact of varying working environments, such as tool materials or workpiece variations, coolant concentration variations and ambient noises. The other drawback is that the design of the signal feature extraction is complicated because the number of candidate signal features is theoretically infinite, which significantly increases the difficulty of the feature selection [9].

In order to address the challenges faced by the signal feature-based method, a model feature-based TCM methods are proposed [9]. As can be seen in Fig. 1(b), the model feature-based method firstly builds a dynamic process model from the collected signals via sensor data modeling rather than directly using the data collected from sensors. After that, the frequency response characteristics of the model are extracted and used as the features to conduct TCM, which are referred to as model features. Experimental studies verify that the model feature-based method has higher adaptability and lower complexity when compared with the signal feature-based method.

Whereas, as cited in [9], the model features-based method has only been validated under simple machining process conditions, such as straight line cutting. It has not been applied to more complex dynamic machining processes for real-time tool wear prediction. However, to the best of the authors' knowledge, there is still no publication and industrial standard on real-time TCM for complicated dynamic machining processes. The very promising concept of digital twin and its potential applications in advanced manufacturing have provided great opportunities to solve this more fundamental challenge.

The first appearance of "digital twin" dates back to 2003 when Grieves first introduced the term at the University of Michigan Executive Course on Product Lifecycle Management (PLM) [10]. Although the concept was not yet concrete enough at the time, Grieves proposed a preliminary form of the digital twin, consisting of three parts: the physical product, the virtual product, and their data flow connections. However, due to some technical limitations, the growth of digital twins was relatively slow at that time [11]. In 2012, the National Aeronautics and Space Administration (NASA) and United States Air Force (USAF) revisited the concept of the digital twin. It is an ultra-fidelity simulation process that integrates multi-physics, multi-scale, and probability to reflect the state of the corresponding twin in real-time, based on historical data, real-time sensor data, and physical models [12]. Since then, digital twins become a popular research topic due to the great development of sensor detection, big data analysis, Internet of Things (IoT), deep learning, etc [13].

In recent years, digital twins are being more and more used in the field of condition monitoring and fault diagnosis (CMFD). Moghadam et al. [14] present a digital twin-based CMFD approach for offshore wind turbines drivetrain systems where the digital twin in the study includes a torsional dynamic model, online measurements and fatigue damage estimation. The remaining useful life of the drivetrain can be estimated by means of the digital twin. Wang et al. [15] proposed a digital twin model including a geometric model, physical model, behavior model and rule model to conduct fault prediction of the autoclave. Tao et al. [16] proposed a five-dimension digital twin model for wind turbines in order to improve the accuracy of prognosis. Xia et al. [17] proposed a digital twin model for machinery fault diagnosis where the digital twin is built by establishing the simulation model which can be updated through the real-time data collected from the physical asset. The proposed digital twin is validated through a case study of triplex pump fault diagnosis. Qiao et al. [18] presents a data-driven based digital twin and a deep learning model to conduct TCM in machining. To realize a more reliable TCM approach, Luo et al. [19] proposed a hybrid digital twin model that consists of model-based digital twins and data-driven digital twins to take into consideration the environmental variations in the life cycle of the tool. As can be seen in the literature, a digital twin can be used to conduct CMFD in real-time, which offers a significant approach to the real-time implementation of the TCM of complicated machining process.

Motivated by the needs to fundamentally address challenges with TCM in advanced manufacturing, in the present study, an innovative digital twin-based anomaly detection framework for real-time TCM is proposed. The framework systematically integrates the advantages of the novel model feature-based TCM in [9] and the promising concept of digital twin aiming to comprehensively resolve adaptivity, complexity, and real-time implementation issues with existing TCM technologies.

In this innovative framework of the digital twin, the "physical product" consists of the complex dynamics of the machining process, which covers the dynamics of the spindle, cutting tool, workpiece, milling fixture, and the interactions between the cutting tool and workpiece. The "virtual product" includes an on-line updated data-driven model describing, in real-time, the overall behaviors of the "physical product", the frequency domain analysis results of the data-driven model, as well as a diagnostic algorithm that performs TCM based on the data-driven model's frequency domain characteristics. The "data flow connection" consists of real-time machine tool vibration data and machine tool numerical controller (NC) information required to facilitate the "virtual product" updating, analysis, and TCM implementation.

This novel integration of the innovative model sensor-based TCM and the promising concept of digital twin would, for the first time, enable TCM to be literally applied to complicated dynamic machining processes, which have so far never been achieved. The proposed new digital twin-based anomaly detection approach is comprehensively evaluated by conducting a range of *in-situ* manufacturing experiments. Very promising results have been achieved demonstrating that the new digital twin framework for TCM has significant potential to be adopted and applied in industrial applications.

2. Digital twin for real-time tool condition monitoring (TCM)

2.1. Basic structures

The concept of a digital twin was first proposed by Grieves in 2003 [10]. Subsequently, the University of Sheffield Advanced Manufacturing Research Centre (AMRC) provided a more comprehensive and solid definition, building upon the original concept: "The digital twin is a live digital coupling of the state of a physical asset or process to a virtual representation with a functional output [20]". Fig. 2(a) presents a flowchart to illustrate this definition where six highlighted parts which are Live, Digital coupling, State, Physical product or process, Virtual product and Functional output, are showed. Additional explanations of these parts are provided below [20]:

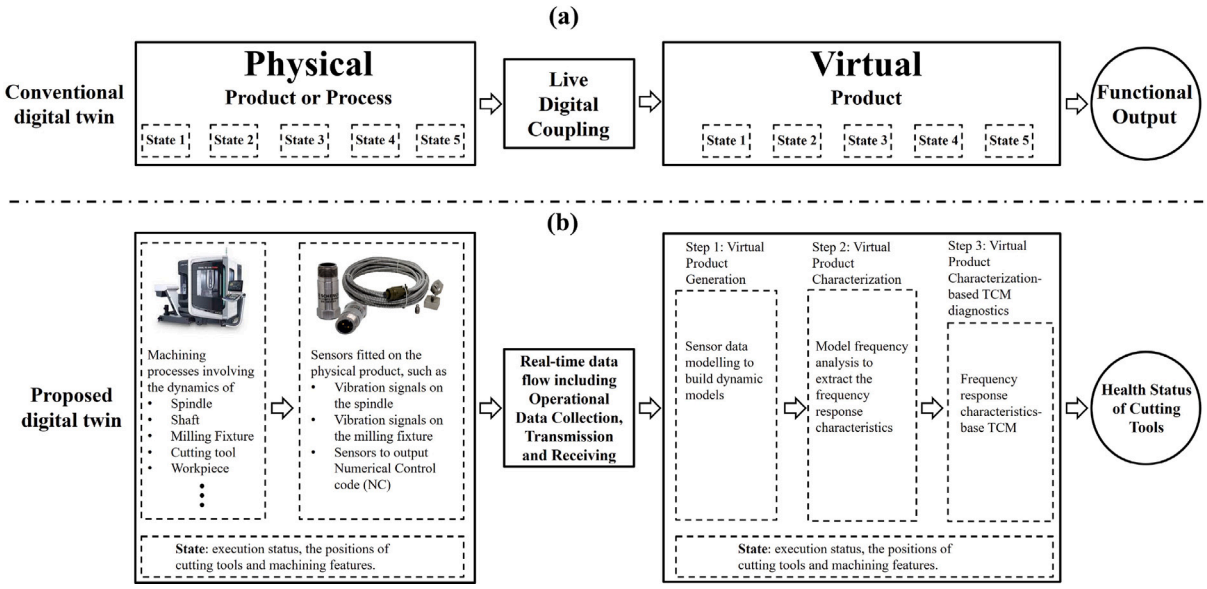

Fig. 2. Schematic of the data connections of digital twins, (a) conventional digital twin and (b) proposed new digital twin for real-time TCM.

(1) *Live:* In the context of a digital twin, the data being utilized must be live, which implies that there is no appreciable difference between the state of the physical product or process and its virtual counterpart at any given moment. There are two key factors for live which are latency and sampling rate. Latency refers to the delay between the occurrence of an event and the receipt of information, while sampling rate relates to the regularity of updated information being transmitted. Both factors are key considerations for a live connection and highly dependent on the use-case.

(2) *Digital coupling:* Digital coupling refers to the connection from the physical product or process to its virtual counterpart, and it involves the data collection, data transmission and data receiving. Security is the paramount consideration for the coupling mechanism because failing to protect the data can be disastrous for a business. Added security can increase the safety of the system, but might lose practicality because of the complexity of the hardware and software development.

(3) *State:* State refers to the specific condition or situation of a physical product or process at a given time. Similar to live, the state description must be measured to an acceptable level of accuracy in order to avoid making incorrect future decisions. Furthermore, a state model needs to be accompanied by supporting descriptions, often provided through metadata, that provide context for the live state information.

(4) *Physical product or process:* A digital twin requires a physical part which can either be a product or process. The physical product is a tangible unique item and the physical process is a specific workflow that has been established. Both of these have economic, social, or commercial value since they can be owned and traded. The characteristic of a physical product or process displaying its behavior in the real world and generating state information is the starting point for building a basic digital twin.

(5) *Virtual product:* The virtual product is the utilization of the information from the physical product in its model. Qi et al. [21] defined that the virtual product of digital twin comprises five types of models which are the geometric model, physical model, behavioral model, and rule model. A brief summary of these models is provided below.

- The geometric model describes the geometric shape, embodiment and appearance of a physical counterpart in terms of appropriate data structures.
- The physical model adds information such as accuracy information (e.g., dimensional tolerance, shape tolerances, etc.) and assembly information (e.g., mating relationship and assembly order.) to describe the attribute and constraint of the physical counterpart.
- The behavioral model describes various behaviors of a physical counterpart to achieve functionality, respond to changes, adjust internal operations, maintain health, etc.
- The rule model describes the rules extracted from historical data, expert knowledge, and predefined logic.
- Finally, model verification, validation and accreditation (VV&A) is used to analyze if and to what extent the correctness, tolerance, availability, and operation results meet the requirements, with the aim of improving the accuracy of the model and the confidence of the simulation.

(6) *Functional output:* For functional output to occur, the digital twin must provide actionable information to an external system or a human observer. Information can be considered actionable if it facilitates an informed decision that generates economic, social, or commercial benefits for one or more stakeholders.

2.2. Existing challenges

The current digital twin framework presented in Fig. 2(a) constitutes a foundational aspect of a digital twin, where geometric and physical models are established through the incorporation of various modeling assumptions and expert input. However, the application of this framework in the context of TCM is challenging due to the complexity of machine tools and the limited understanding of the mechanisms of tool degradation. As a result, the geometric and physical model-based digital twin is only capable of detecting significant tool damage, such as tool breakage. The monitoring of tool wear, particularly in extreme and complicated working environments, poses challenges for the conventional digital twin framework [19,22].

2.3. Proposed digital twin framework

In order to fundamentally address the challenges faced by conventional digital twin-based TCM methods and improve the monitoring accuracy of cutting tools, this study introduces a novel digital twin framework. This approach abandons the dependence on geometric and

physical models and instead constructs digital twin models based on the dynamic relationship between vibration data measured from machining processes. The fundamental principle is that cutting tool anomalies in machining can introduce additional dynamic effects. By analyzing the changes in these dynamics, the digital twin model can evaluate the health status of cutting tools. Fig. 2(b) illustrates the framework of the proposed digital twin, with further elaborations on the novel approach provided below.

(1) *Live:* In the context of the proposed digital twin, live represents the "real-time" operation data collected from sensors mounted on the machine tool, characterized by specific sampling rates and latency. For example, at AMRC, the real-time data can be sampled at a minimum rate of 1 Hz, with an approximate latency of 100 ms when transmitted to a shop-floor network [20].
(2) *Digital coupling:* Digital coupling incorporates aspects of data collection, transmission, and receiving. At AMRC, the collected data can be sent over a shop floor integration layer, such as an Open Platform Communications United Architecture (OPC UA) server [20].
(3) *State:* State in the proposed digital twin refers to a description of the execution status, such as in machining and shutdown, as well as the positions of cutting tools and machining features. In real-time machining, state information will be provided in numerical control (NC) code.
(4) *Physical product or process:* The physical product or process comprises two parts.

- The first part consists of sensors installed on the machine tool, such as those measuring vibration signals on the spindle and milling fixture, as well as sensors providing NC signals.
- The second part encompasses the machining process, which includes the dynamics of the spindle, cutting tool, workpiece, milling fixture, and the interactions between the cutting tool and workpiece. Cutting tool anomalies can give rise to additional dynamic effects within the physical product, subsequently leading to further changes in the product's behavior.

(5) *Virtual product:* As can be seen in Fig. 2(b), the virtual product has the following three steps:

- The first step is the virtual product generation where the vibration signals measured from the spindle and the milling fixture, respectively, are used to build the dynamic model to represent the dynamics between the cutting tool supported by the spindle and the workpiece clamped on the milling fixture. This process is termed as sensor data modeling.
- The second step involves the virtual product characterization which is to conduct model frequency analysis to extract the frequency response characteristics of the dynamic model. It is well known that the extracted frequency response characteristics can represent unique and physically meaningful features of a dynamic process [9,23].
- The last step is the virtual product characteristic-based TCM where the extracted frequency response characteristics are used as the features for evaluating the health conditions of cutting tools.

More explanations of the proposed virtual product will be provided in Section 3.

(6) *Functional output:* The functional output of the proposed digital twin is the health status of the cutting tool, which serves to indicate whether the tool is in a healthy condition, indicating that the tool can still be used, or in a worn state, indicating the tool needs replacement.

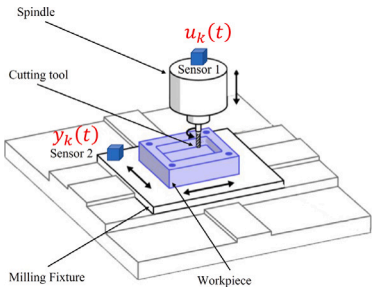

Fig. 3. Machining process illustration.

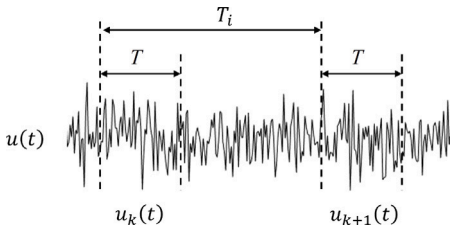

Fig. 4. Short-time vibration signal of $u_k(t)$.

3. Sensor data modeling and model frequency analysis

3.1. Sensor data modeling for virtual product generation

Virtual product generation in Fig. 2(b) uses sensor data modeling to model the dynamics between the cutting tool supported by the spindle and the workpiece clamped on the milling fixture in real-time. As depicted in Fig. 3, let $u_k(t)$ and $y_k(t)$ represent the vibration signals of time length T collected from the spindle and milling fixture at the moment of the kth data collection, respectively, i.e. $u_k(t) = u[t+(k-1)T_i]$ and $y_k(t) = y[t+(k-1)T_i]$ with $t = 1, \ldots, T$ and $k = 1, \ldots, K$. The notation T_i represents the monitoring period of the digital twin such that the vibration signals will be transferred from the physical product to the virtual product at every T_i time interval. An explanation of $u_k(t)$ is shown schematically in Fig. 4.

Since $u_k(t)$ and $y_k(t)$ are two vibration responses of the machining process, a general model representing the dynamics of the machining process can be represented as

$$u_k(t) = \mathbb{D}_{u,k}\left[P_k(t)\right], k = 1, \ldots, K \tag{1}$$

$$y_k(t) = \mathbb{D}_{y,k}\left[P_k(t)\right], k = 1, \ldots, K \tag{2}$$

where $P_k(t)$ represents the input power applied to the machining process at the moment of the kth data collection. $\mathbb{D}_{u,k}$ denotes the dynamic relationship between $u_k(t)$ and $P_k(t)$, and $\mathbb{D}_{y,k}$ denotes the dynamic relationship between $y_k(t)$ and $P_k(t)$. Under the condition that there exist unique inverse operations $\mathbb{D}_{u,k}^{-1}$ and $\mathbb{D}_{y,k}^{-1}$ for $\mathbb{D}_{u,k}$ and $\mathbb{D}_{y,k}$,

$$P_k(t) = \mathbb{D}_{u,k}^{-1}\left[u_k(t)\right] \tag{3}$$

$$P_k(t) = \mathbb{D}_{y,k}^{-1}\left[y_k(t)\right] \tag{4}$$

Therefore,

$$\mathbb{D}_{y,k}^{-1}\left[y_k(t)\right] = \mathbb{D}_{u,k}^{-1}\left[u_k(t)\right] \tag{5}$$

which implies

$$y_k(t) = \mathbb{D}_{y,k}\left\{\mathbb{D}_{u,k}^{-1}\left[u_k(t)\right]\right\} = \mathbb{D}_k\left[u_k(t)\right] \tag{6}$$

where $\mathbb{D}_k$ represents the dynamic relationship between $u_k(t)$ and $y_k(t)$.

For sensor data modeling, it can incorporate additional dynamic effects from cutting tool anomalies into the model $\mathbb{D}_k$. In particular, sensor data modeling is a data-driven technique. An effective modeling strategy, as discussed in Section 3.4, can explicitly represent the state of the tool while mitigating the adverse effects of external environmental factors on fault detection.

3.2. Model frequency analysis for virtual product characterization

After the relationship between $u_k(t)$ and $y_k(t)$ is modeled, the second step is to perform virtual product characterization. This is achieved by conducting model frequency analysis where the frequency response properties of the model $\mathbb{D}_k$ are evaluated. These properties can represent unique and physically meaningful features of milling processes [23].

If the dynamics of the machining process shown in (6) is a linear system, data-driven modeling techniques will produce a linear model to represent the dynamics $\mathbb{D}_k$. In such a situation, the frequency domain properties of model (6) can be expressed as

$$\frac{Y_k(j\omega)}{U_k(j\omega)} = \frac{\mathcal{F}[y_k(t)]}{\mathcal{F}(u_k(t))} = \frac{\mathcal{F}\{\mathbb{D}_k[u(t)]\}}{\mathcal{F}(u_k(t))} = H_k(j\omega) \tag{7}$$

where $\mathcal{F}(\cdot)$ denotes the Fourier transform. $U_k(j\omega)$ and $Y_k(j\omega)$ are, respectively, the Fourier transform of $u_k(t)$ and $y_k(t)$. $H_k(j\omega)$ indicates the frequency response function (FRF) of model (6) at the moment of the kth data collection, and the notation ω is the frequency variable.

Nonetheless, in real-world situations, the dynamics of the machining process are often nonlinear; so model (6) is, in general, a nonlinear model, which makes data-driven modeling and the associated model frequency analysis be a complicated task. To solve this issue, nonlinear system modeling and nonlinear model frequency analysis were proposed by the authors in [9] where the Nonlinear Output Frequency Response Functions (NOFRFs) were used to extend the FRF to the nonlinear cases to fulfill more complicated TCM tasks.

The definition of NOFRFs is that if a nonlinear system is asymptotically stable at the zero equilibrium, the model output response can be determined by a Volterra series as [23]

$$\begin{aligned} y_k(t) &\approx \sum_{n=1}^{N} y_{n,k}(t) \\ &= \sum_{n=1}^{N} \sum_{\tau_1=0}^{\infty} \cdots \sum_{\tau_n=0}^{\infty} h_{n,k}(\tau_1, \ldots \tau_n) \prod_{i=1}^{n} u_k(t-\tau_i) \end{aligned} \tag{8}$$

where the notation N indicates the maximum order of the system nonlinearity. $y_{n,k}(t)$ denotes the nth order output at the moment of the kth data collection, and $h_{n,k}(\tau_1, \ldots, \tau_n)$ is the nth order Volterra kernel of the system output at the moment of the kth data collection. According to [23,24], the output spectrum $y_k(t)$ can be represented as

$$\mathcal{F}(y_k(t)) = Y_k(j\omega) \approx \sum_{n=1}^{N} Y_{n,k}(j\omega) = \sum_{n=1}^{N} G_{n,k}(j\omega) U_{n,k}(j\omega) \tag{9}$$

where $Y_{n,k}(j\omega)$ and $U_{n,k}(j\omega)$ are the nth order output frequency spectrum and input frequency spectrum, respectively.

$$G_{n,k}(j\omega) = \frac{Y_{n,k}(j\omega)}{U_{n,k}(j\omega)} \tag{10}$$

is defined as the nth order NOFRF at the moment of the kth data collection with $\omega \in \Omega_n$ and $n = 1, \ldots, N$ with $U_{n,k}(j\omega) \neq 0$ [25]. Ω_n is the frequency support of $U_{n,k}(j\omega)$ where $U_{n,k}(j\omega) \neq 0$. As can be seen, (10) is expressed in a manner which is similar to the FRF. When $n = N = 1$, $G_{n,k}(j\omega) = G_{1,k}(j\omega) = H_k(j\omega)$ reduces to the FRF of a linear system shown in (7).

The evaluation of the NOFRFs can be carried out using the equations shown in Appendix and the evaluated NOFRFs can be expressed as $G^*_{n,k}(j\omega)$ with

$$\mathbf{f}_{n,k} = [|G^*_{n,k}(j\omega_n^1)|, \ldots, |G^*_{n,k}(j\omega_n^{L_n})|] \tag{11}$$

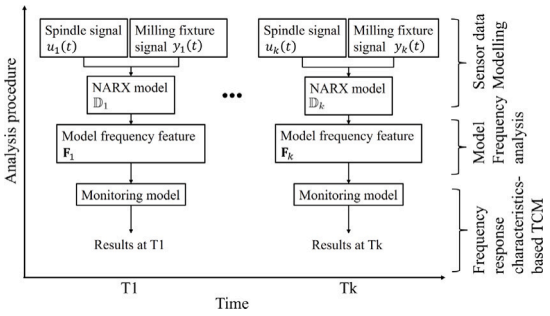

Fig. 5. An illustration of how the proposed new digital twin works for TCM.

where $\mathbf{f}_{n,k}$ denotes the nth order NOFRFs-based features at the moment of the kth data collection. Here L_n represents the number of model features of the nth order, and $\omega_n^1, \ldots, \omega_n^{L_n} \in \bar{\Omega}_n$, $\bar{\Omega}_n \in \Omega_n$. As a result, the overall NOFRFs-based features at the moment of the kth data collection can be obtained as

$$\mathbf{F}_k = [\mathbf{f}_{1,k}, \ldots, \mathbf{f}_{n,k}], n = 1, \ldots, N, k = 1, \ldots, K \tag{12}$$

where $\mathbf{F}_k$ is defined as the model frequency features (MFFs) having the potential to reflect and monitor the process of tool wear from the 1-st to the Kth data collection period.

3.3. Frequency response characteristics-based TCM

As can be seen in Fig. 5, at time T1, the machine tool dynamics represented by the dynamic relationship between the spindle vibration signal $u_1(t)$ and the milling fixture vibration signal $y_1(t)$ are modeled using a nonlinear system identification approach and the measured data of $u_1(t)$ and $y_1(t)$, producing a data-driven model (known as NARX model in the paper). Then, from $\mathbb{D}_1$, the frequency domain features (known as NOFRFs) of the identified NARX model are extracted, producing MFF $\mathbf{F}_1$ for monitoring the cutting tool conditions at time T1. At a subsequent time Tk, following the same procedure, the data-driven NARX model is updated to $\mathbb{D}_k$, and then the corresponding MFF is updated to $\mathbf{F}_k$ and used to monitor the cutting tool conditions at time Tk.

The advantages of the frequency response characteristics-based TCM are that MFFs describe exactly the state of the cutter over time. This is due to the fact that the model developed via sensor data modeling can often be uniquely determined by the cutter status; and the frequency analysis of the model can effectively reveal unique and physically meaningful features of the health conditions of the cutter [9]. The details of the MFFs-based diagnostics will be introduced in Section 4.

3.4. Implementation algorithm

3.4.1. Sensor data modeling

Sensor data modeling is a data-driven modeling technique aiming to find the dynamic relationship between the $u_k(t)$ and $y_k(t)$ collected from the vibration sensors mounted on the spindle and milling fixture. The relationship at the moment of the kth data collection can be described by using the Nonlinear AutoRegressive with eXogenous input (NARX) model

$$\begin{aligned} y_k(t) = &F^{\ell}[y_k(t-1), \cdots, y_k(t-\delta_y), \\ &u_k(t-1), \cdots, u_k(t-\delta_u)] + e_k(t) \end{aligned} \tag{13}$$

where F^{ℓ} indicates the polynomial function and the notation ℓ is the maximum polynomial degree with $\ell \in \mathbb{Z}^+$. $e(t)$ indicates the unmodeled dynamics and noise. δ_y and δ_u are the maximum lags for the

milling fixture and spindle vibration signals, respectively. To identify the polynomial function F^{ℓ}, (13) is presented as follows

$$y_k(t) = \sum_{m=1}^{M} \theta_{m,k} d_{m,k}(t) + e_k(t) \tag{14}$$

where $\theta_{m,k}$ with $m = 1, \ldots, M$ are the model parameters at the moment of the kth collection, and the notation M indicates the total number of the model terms. $d_{m,k}(t)$ are the regressors composing $y_k(t-1), \cdots, y_k(t-\delta_y), u_k(t-1), \cdots, u_k(t-\delta_u)$. The polynomial model (14) can be further written in the matrix form

$$\mathbf{y}_k = \mathbf{D}_k \boldsymbol{\Theta}_k + \boldsymbol{\Xi}_k \tag{15}$$

where $\mathbf{y}_k = [y_k(1), \ldots, y_k(T)]^{\mathrm{T}}$ with $k = 1, \ldots, K$, and $\boldsymbol{\Theta}_k = [\theta_{1,k}, \ldots, \theta_{M,k}]^{\mathrm{T}}$, and $\boldsymbol{\Xi}_k = [e_k(1), \ldots, e_k(T)]^{\mathrm{T}}$. The dictionary $\mathbf{D}_k$ is a matrix of $\mathbf{D}_k = \{\mathbf{d}_{1,k}, \ldots, \mathbf{d}_{M,k}\}$ and $\mathbf{d}_{m,k} = [d_{m,k}(1), \ldots, d_{m,k}(T)]$.

In practical data-driven modeling, the parsimonious principle is the fundamental principle which ensures the smallest possible model is used to explain the data [26]. In this study, the orthogonal least squares (OLS) algorithm is used because it proved to be an efficient learning procedure for building parsimonious models in smart manufacturing-related TCM [9]. After OLS is applied to select the model structure, the final model can be written as

$$\mathbf{y}_k = \mathbf{W}_k \boldsymbol{\alpha}_k + \boldsymbol{\Xi}_k \tag{16}$$

where $\mathbf{W}_k$ is the OLS selected model structure which can be formed as $\mathbf{W}_k = \{\mathbf{d}_{1,k}, \ldots, \mathbf{d}_{M_o,k}\}$ and $M_o \ll M$. The notation $\boldsymbol{\alpha}_k$ is an OLS parameter vector which can be evaluated by solving the following l_2-norm optimization problem.

$$\begin{aligned} \boldsymbol{\alpha}_k &= \arg\min_{\boldsymbol{\alpha}} \{ \left\| \mathbf{W}_k \boldsymbol{\alpha}_k - \mathbf{y}_k \right\|_2^2 + \lambda_k \left\| \boldsymbol{\alpha}_k \right\|_2 \} \\ &= (\mathbf{W}_k{}^{\mathrm{T}} \mathbf{W}_k + \lambda_k \boldsymbol{I})^{-1} \mathbf{W}_k{}^{\mathrm{T}} \mathbf{y}_k \end{aligned} \tag{17}$$

where $\boldsymbol{I}$ is an $M_0 \times M_0$ identity matrix; and the notation λ_k denotes the penalty parameter which is one of the key factors to obtain a satisfactory NARX model. The discussion of penalty parameter λ_k tuning has been well defined in [9].

3.4.2. Model frequency analysis

To evaluate the NOFRFs of the model, the Generalized Associated Linear Equations (GALEs) algorithm is used to decompose the built NARX model (13) into a series of linear difference equations. Depending on the decomposed equations, NOFRFs can be evaluated from low order to an arbitrarily high order [27]. The evaluated NOFRFs can be expressed as $G^*_{n,k}(j\omega)$ and the equations of NOFRFs evaluation are presented in Appendix.

4. Algorithm of model frequency features (MFFs) and machine tool numerical controller (NC) signal-based diagnostics

4.1. Status index modeling

Suppose that the machine tool is in operation and $H(k)$ represents the status index (SI) reflecting the tool wear status of the cutter at the moment of the kth data collection. $H(k)$ can be defined as a moving average model shown below

$$H(k) = \begin{cases} \frac{Y(k-W+1)+,\ldots,+Y(k)}{W}, k \geq W \\ 0, k < W \end{cases} \tag{18}$$

where the notation W is the window length, and $Y(k) \in \{0, 1\}$ is defined as the binary prediction result shown below:

$$\begin{aligned} Y(k) =& \mathbb{C}\left[\mathbf{F}_k, \boldsymbol{\varphi}_k\right], k = 1, \ldots, K \\ =& \begin{cases} \mathbb{C}_1\left[\mathbf{F}_k, \boldsymbol{\varphi}_k\right], \boldsymbol{\varphi}_k \in \left\{ \boldsymbol{\varphi}_{k_1^1}, \ldots, \boldsymbol{\varphi}_{k_1^{\zeta_1}} \right\} \\ \vdots \\ \mathbb{C}_q\left[\mathbf{F}_k, \boldsymbol{\varphi}_k\right], \boldsymbol{\varphi}_k \in \left\{ \boldsymbol{\varphi}_{k_q^1}, \ldots, \boldsymbol{\varphi}_{k_q^{\zeta_q}} \right\} \end{cases} \end{aligned} \tag{19}$$

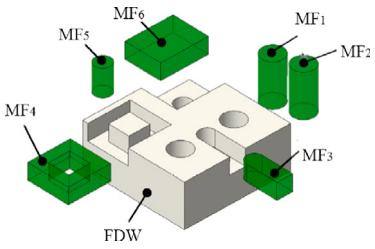

Fig. 6. An example of a final designed workpiece [28].

with

$$\boldsymbol{\mathcal{N}} = [\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_k]^{\mathrm{T}} \tag{20}$$

$$\bar{\boldsymbol{\mathcal{N}}} = [\boldsymbol{\varphi}_{k_1^1}, \ldots, \boldsymbol{\varphi}_{k_1^{\zeta_1}}, \ldots, \boldsymbol{\varphi}_{k_q^1}, \ldots, \boldsymbol{\varphi}_{k_q^{\zeta_q}}]^{\mathrm{T}} \tag{21}$$

$$\boldsymbol{\varphi}_k = [\varphi_1, \ldots, \varphi_r], r = 1, \ldots, R \tag{22}$$

$$\mathbb{C} = \left\{ \mathbb{C}_1, \ldots, \mathbb{C}_q \right\}, q = 1, \ldots, Q \tag{23}$$

The notation $\boldsymbol{\mathcal{N}}$ is the collected NC signal, and $\bar{\boldsymbol{\mathcal{N}}}$ represents the NC program employed to control the manufacturing operations, enclosing the entire machining process. $\boldsymbol{\varphi}_k$ indicates the NC signal of the machining process at the moment of the kth data collection; and $\boldsymbol{\varphi}_{k_q^{\zeta_q}}$ represents the ζ_q-th NC signal at k_q-th set. $\mathbb{C}_q[.]$ is a sub-classification model that is a function of both $\mathbf{F}_k$ and $\boldsymbol{\varphi}_k$. $\mathbb{C}[.]$ is a binary classification model that composes the Q sub-classification models to produce binary prediction results. As can be seen in (19), the sub-classification model $\mathbb{C}_q[.]$ to be updated is determined by the set to which $\boldsymbol{\varphi}_k$ is assigned.

For example, Fig. 6 shows a final designed workpiece (FDW) where six machining features (MF1 to MF6 shown in Fig. 6) need to be removed, and the designed machining processes to remove these features are recorded in the NC program $\bar{\boldsymbol{\mathcal{N}}}$. Assuming that the machining process for these six features are different, $\bar{\boldsymbol{\mathcal{N}}}$ can be partitioned into six distinct sets, each correlating to a specific machining feature. The binary classification model can be set to $\mathbb{C} = \left\{ \mathbb{C}_1, \ldots, \mathbb{C}_q \right\}, q = 1, \ldots, 6$, and the collected NC signals $\boldsymbol{\varphi}_k$ will update the sub-classification models $\mathbb{C}_q[.]$ to output the prediction result $Y(k)$.

4.2. Prediction threshold determination

In real-time machining, it is necessary to monitor the SI for TCM. If the SI value exceeds the prediction threshold T_p, the tool is classified as worn. Otherwise, the tool can be deemed to be in a healthy condition. The determination of T_p is in the training stage which is based on the following polynomial curve fitting.

$$\begin{aligned} T_p &= H(k) - e_p(k) \\ &= H_{\mathrm{fit}}(k) \\ &= \vartheta_1 k^{\psi} + \vartheta_2 k^{\psi-1} +, \ldots, +\vartheta_{\psi} k + \vartheta_{\psi+1}, k = T_m \end{aligned} \tag{24}$$

where ψ indicates the maximum degree of the polynomial function, and $\boldsymbol{\vartheta} = [\vartheta_1, \vartheta_2, \ldots, \vartheta_\kappa, \vartheta_{\kappa+1}]^{\mathrm{T}}$ indicates the coefficients of the function which can be determined using the least square method. $H_{\mathrm{fit}}(k)$ indicates the polynomial fitted curve and $e_p(k)$ represents the residual. The notation T_m is defined as the True Moment indicating the moment when the tool wear reaches the predetermined tool wear threshold. Therefore, this tool wear threshold is directly linked to the moment T_m and can be expressed as a function of T_m, denoted as $W(T_m)$.

Consequently, as illustrated in Fig. 7, the MMFs and NC signal-based diagnostics incorporate two AI models: the moving average model, which outputs the SI, and the classification model, which predicts binary results. These two models work cooperatively to conduct TCM, thus ensuring real-time and accurate monitoring of the tool condition.

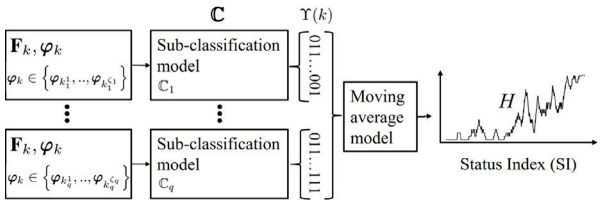

Fig. 7. Flowchart of the MMFs and NC signal-based diagnostics.

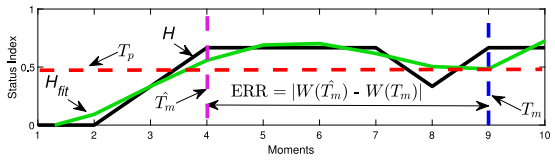

Fig. 8. 2-D plot of the status index. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5. A new digital twin-based anomaly detection technique for real-time tool condition monitoring

The proposed digital twin-based anomaly detection technique integrates the aforementioned real-time sensor data modeling and model frequency analysis (introduced in Section 3), and MFFs and machine tool NC signal-based diagnostics (described in Section 4), aiming to monitor the cutting tool status in real-time and diagnose cutting tool anomalies when tool wear reaches a certain threshold. The technique involves two parts: off-line training and on-line monitoring.

5.1. Off-line training

Step 1: Off-line MFFs evaluation

MFFs are evaluated based on off-line collected spindle and milling fixture vibration signals. The details of the MFFs evaluation have been described in Section 3.

Step 2: AI labeling

To label the MFFs, the condition of the tool is firstly defined. If the tool wear is within a predetermined wear threshold $W(T_m)$, the tool is classified as healthy. Otherwise, the tool is classified as worn. If the MFFs are extracted from the machining process under healthy tool conditions, the labeling is the binary number 0 (health). Alternatively, if the tool is worn, the label is the binary number 1 (worn).

Step 3: Machining features definition

The categories of machining features are defined based on the geometric shapes and topology structures. In accordance with the feature definition methods outlined in previous literature [29], Q different machining features are defined, necessitating the development of Q sub-classification models to predict the health status of the cutting tool.

Step 4: Data partitioning

The identified machining features enable partitioning of the NC program $\tilde{\mathcal{N}}$ into Q different sets.

Then, the collected NC signals $\mathcal{N}$ can be divided into train NC signals $\mathcal{N}_{\text{train}}$ and test NC signals $\mathcal{N}_{\text{test}}$, and each dataset can be expressed as

$$\mathcal{N}_{\text{train}} = [\boldsymbol{\varphi}_1, \dots, \boldsymbol{\varphi}_{k_1}]^{\text{T}} \tag{25}$$

$$\mathcal{N}_{\text{test}} = [\boldsymbol{\varphi}_2, \dots, \boldsymbol{\varphi}_{k_2}]^{\text{T}} \tag{26}$$

In parallel, the off-line evaluated MFFs are also divided into two datasets which are train MFFs datasets, denoted as $\mathcal{F}_{\text{train}}$, and test MFFs datasets, denoted as $\mathcal{F}_{\text{test}}$.

$$\mathcal{F}_{\text{train}} = [\mathbf{F}_1, \dots, \mathbf{F}_{k_1}]^{\text{T}} \tag{27}$$

$$\mathcal{F}_{\text{test}} = [\mathbf{F}_2, \dots, \mathbf{F}_{k_2}]^{\text{T}} \tag{28}$$

Step 5: MFFs and NC signal-based AI model training

The objective of this step is to train the AI models, including classification and moving average models introduced in Section 4, for TCM. Specific procedures are as follows:

(a) The binary classification model $\mathbb{C}[.]$ is trained off-line using the labeled $\mathcal{N}_{\text{train}}$ and $\mathcal{F}_{\text{train}}$ where $\mathbb{C}[.]$ encloses Q sub-classification models.

(b) Substituting $\mathcal{N}_{\text{train}}$ and $\mathcal{F}_{\text{train}}$ into (18) and (19), the SI of the train dataset can be obtained as $H_{\text{train}}(k)$ which can further determine the prediction threshold T_p based on (24).

Step 6: Model testing

By substituting $\mathcal{N}_{\text{test}}$ and $\mathcal{F}_{\text{test}}$ into (18) and (19), the SI of the test dataset can be obtained as $H_{\text{test}}(k)$. If $H_{\text{test}}(k)$ exceeds the prediction threshold T_p, the tool is regarded as worn. The moment when SI reaches T_p is defined as the Prediction Moment $\hat{T}_m$, and the tool wear at $\hat{T}_m$ moment is defined as the prediction tool wear $W(\hat{T}_m)$. The difference between the prediction tool wear $W(\hat{T}_m)$ and the predetermined tool wear threshold $W(T_m)$ is

$$\text{ERR} = \left| W(\hat{T}_m) - W(T_m) \right| \tag{29}$$

where $W(\hat{T}_m)$ and $W(T_m)$ indicate the prediction tool wear and predetermined tool wear which are the function of $\hat{T}_m$ and T_m, respectively. ERR represents the prediction error, indicating the precision of the diagnostics. If the prediction error ERR is greater than a predefined error threshold T_e, e.g. 20 μm, the off-line training is deemed invalid, necessitating a reset of the relevant tunable parameters in the modeling stage for another training round. It is worth mentioning that, to prevent mistaken alarms, if the $H_{\text{test}}(k)$ curve surpasses the prediction threshold T_p persistently (e.g., for more than three moments), this indicates that the tool is definitively worn. Otherwise, the tool is considered to be in good condition and can remain in use.

For example, Fig. 8 displays the 2-D plot of SI, where the black curve represents $H(k) = [0, 0, 0.33, 0.67, 0.67, 0.67, 0.67, 0.33, 0.67, 0.67]$, and the green curve represents the polynomial fitted curve $H_{\text{fit}}(k)$. The red dash line is defined as prediction threshold T_p. The purple dash line symbolizes the prediction moment $\hat{T}_m$ where the tool wear at this moment is $W(\hat{T}_m)$. The blue dash line indicates the true moment T_m, indicating when tool wear meets the predetermined tool wear threshold $W(T_m)$. ERR indicates the differences between the predicted tool wear $W(\hat{T}_m)$ at the moment $\hat{T}_m$ and the predetermined tool wear threshold $W(T_m)$ at the moment T_m.

5.2. On-line monitoring

The second part involves online monitoring using the online extracted MFFs and off-line trained AI models for real-time TCM. The online monitoring can be implemented through the following steps.

Step 7: On-line MFFs evaluation

The on-line MFFs are evaluated based on the real-time collected vibration signals in order to describe the state of the cutter over time.

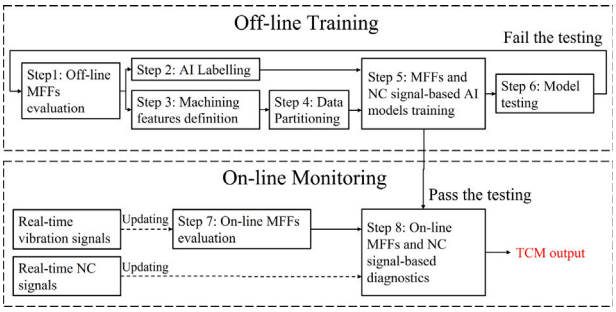

Fig. 9. Flowchart of the proposed digital twin-based anomaly detection technique enclosing off-line training and on-line monitoring.

Step 8: On-line MFFs and NC signal-based diagnostics

The evaluated MFFs and online collected NC signals are input to the off-line trained moving average models and classification models for real-time TCM. If the on-line predicted SI, denoted as $H_{\text{online}}(k)$, surpasses the off-line trained prediction threshold T_p persistently, the tool is considered to be worn.

5.3. Discussion of parameter tuning

With all the steps introduced earlier, the schematic flowchart of the digital twin-based anomaly detection technique enclosing off-line training and on-line monitoring is depicted in Fig. 9. In regard to the off-line training, modeling parameters in step 1, such as time delay δ_u, δ_y and polynomial degree ℓ, play a vital role in optimizing the TCM results. When tuning these modeling parameters, the prediction error ERR can be compared. The parameters that yield the smallest ERR in the test dataset are selected for use.

The second prominent factor is the window length of the moving average model shown in (18). A smaller window length for TCM will make the model more responsive to changes in tool wear. Conversely, a larger window length results in less sensitivity in the prediction. However, there is scant information on how to preset this window length. The most effective method to identify the optimal length is to try a number of different window lengths until the one that is appropriate for the TCM strategy is found.

6. Field experiment

6.1. Design of experiment

To validate the effectiveness of the newly proposed digital twin-based anomaly detection framework for real-time TCM in machining, a run-to-failure experiment was carried out at the Advanced Manufacturing Research Centre (AMRC), The University of Sheffield, Sheffield, U.K., as shown in Fig. 10. The machine is branded as a 5-axis DMU 40 evo machine center, and Sandvik 4-flute carbide end mills were used to cut the Ti-6AL-4V workpieces. For the machining parameters, the rotation speed and feed rate were set to 2586.27 r/min and 1055.20 mm/min, respectively. Furthermore, the cutting width and depth were 2 mm and 20 mm, respectively. As can be seen in Fig. 10, vibration signals of the spindle and platform were acquired from two PCB-type accelerometers mounted on the spindle and milling fixture. The sensitivities for the spindle accelerometer and milling fixture accelerometer are 10 mV/g and 500 mV/g, respectively; and the frequency range for both accelerometers is from 1 Hz to 5 kHz. In this study, the sampling rate for both accelerometers was set to 51.2 kHz.

During the machining process, two cutting tools, termed T1 and T2, were used to cut the workpiece. The milling strategy for the experiment is dynamic milling dividing the workpiece into four layers, each of which can be divided into six different machining features with the

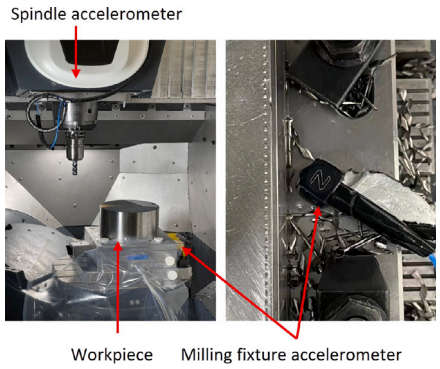

Fig. 10. Test rig of the milling experiment.

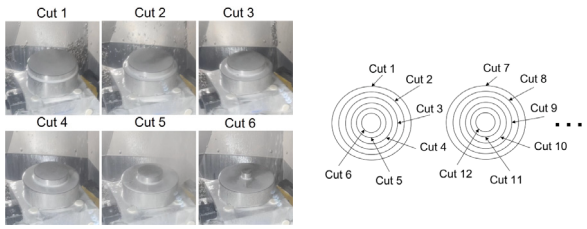

Fig. 11. Image of (a) the workpiece after each round cut and (b) the experiment process.

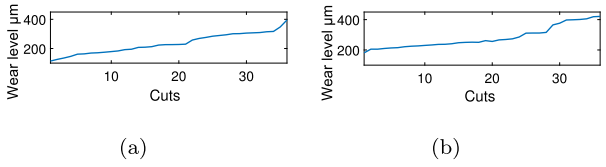

Fig. 12. History of the most severe tool wear among 4 flute cutters measured from microscope (a) Tool 1 and (b) Tool 2.

shape of a ring round (see Fig. 11(a)). After the last round cut of the layer is completed, the surface of the workpiece will be cleaned, which will be used for the other new six round-cuts (see Fig. 11(b)). Every time a round cut is completed, the machine will stop to check the tool wear via the microscope. At the end of the experiment, both tools have cut 6 layers, for a total of 36 round cuts. Fig. 12 shows the history of the most severe tool wear among 4 flute cutters measured from a microscope where the tool wear of both tools changes from around 150 µm to over 400 µm.

6.2. Off-line training

The dataset collected from T1 was firstly used to conduct off-line training. Based on Fig. 9, the off-line training process for T1 involves the evaluation of MFFs. This evaluation consists of sensor data modeling and model frequency analysis. Sensor data modeling involves setting the signal length T to 1 s and the monitoring period of the digital twin T_i to 5 s. This configuration dictates that the digital twin will update every 5 s, and a 1-second snapshot of the signal is collected for each modeling analysis. For model frequency analysis, given the tooth passing frequency of 172.41 Hz, an input excitation ranging from 167 Hz to 177 Hz, $u^*(t) = \frac{3}{2\pi}\frac{\sin(2\times177\times\pi\times t)-\sin(2\times167\times\pi\times t)}{t}$ with $-1 < t < 1$, was designed to investigate the nonlinear characteristics at this frequency range and to evaluate the corresponding NOFRFs of each identified NARX model.

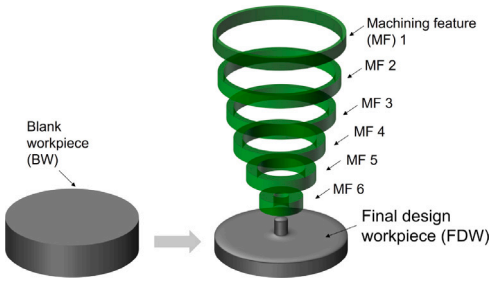

Fig. 13. Machining features of the designed experiment.

The experiment of T1 conducted 36 round cuts, and the final MFFs over the life cycle of the tool totally has 2160 MFFs. To label the MFFs, the predetermined tool wear threshold $W(T_m)$ is set to 280 μm. MFFs derived from machining processes where the tool wear is below 280 μm were labeled as 0, signifying a healthy tool status. MFFs associated with tool wear exceeding the tool wear threshold were labeled as 1, indicative of a worn tool state. As a result, the initial 24 round cuts, which span the first four layers of the machining process, yielded MFFs assigned the label 0; and the last 12 round cuts collected from the last 2 layers were labeled as 1.

As can be seen in Fig. 13, the design of the experiment enclosed six unique machining features. Each of the round cuts can be associated with a distinct machining feature due to the varying rates of cutting force changes across different rounds, which consequently lead to alterations in the dynamics of the machining process. As a result, in response to these six distinct machining features, six different sub-classification models, denoted as R1 to R6, were necessarily constructed to address each machining feature. The NC program can thus be separated into six different sets, each corresponding to a specific machining feature. The collected NC signals and off-line evaluated MFFs are then partitioned into training datasets and test datasets, respectively. Each dataset comprises 1080 entities, including NC signals and NOFRFs-based features.

In MFFs and NC signal-based AI model training, the Support Vector Machine (SVM) was employed as the classification model to generate binary prediction results because it is a powerful classification technique in relevant applications. As can be seen in Fig. 14(a), the Area Under the Curve (AUC) from six sub-classification models (R1 to R6) varies from 0.78 to 0.88, and the Distance to Top Left Corner (DTLCs) [9] changes from 0.24 to 0.36, indicating a comparable level of reliability across these sub-classification models.

After inputting the train MFFs and NC signals into the trained classification models and moving average models, the SI of the train dataset, denoted as H_{train}, can be received shown in Fig. 14(b). This in turn allows for the calculation of the prediction threshold T_p based on (24), as presented by the red dash line in Fig. 14(b).

The final step in the off-line training involves testing the performance of the trained AI models by feeding in the test MFFs and NC datasets. As presented in Fig. 14(c), the SI of the test dataset, denoted as H_{test}, was computed. From the plot, H_{test} surpasses the prediction threshold T_p [red dash line in Fig. 14(c)] at the 644-th moments and remains above it for more than three moments. Therefore, the prediction moment $\hat{T}_m$ [indicated by the purple dash line in Fig. 14(c)] was set at the 644-th moment which corresponds to $W(\hat{T}_m) = 268.95$ μm. The difference between $W(\hat{T}_m)$ and $W(T_m)$ is 11.05 (= |268.95 − 280|) μm which falls within the predefined error threshold T_e of 20 μm, thus validating the efficacy of the off-line training process.

6.3. On-line monitoring

After successfully testing the AI models trained off-line, on-line monitoring was carried out. In this stage, the monitoring period of the digital twin T_i was set to 10 s. This setup was based on the consideration that our computational system (equipped with an Intel(R) Core(TM) i9-12900K CPU and 64-GB RAM) typically requires approximately 5 to 8 s to evaluate a single NOFRFs-based feature. To ensure effective monitoring, it is crucial that this computational duration remains shorter than T_i.

The on-line evaluation process involved the input of the on-line evaluated MFFs and the collected NC signals into the trained AI models to obtain the on-line SI result H_{online}. As can be seen in Fig. 14(d), this on-line diagnostics indicates that the prediction moment $\hat{T}_m$, when the tool begins to wear out, is at the 845-th moment, corresponding to a tool wear $W(\hat{T}_m)$ of 285.73 μm. The difference between the on-line predicted tool wear $W(\hat{T}_m)$ and the predetermined tool wear threshold $W(T_m)$ is 5.73 (= |285.73 − 280|) μm. This falls within the predefined error threshold T_e of 20 μm, thus affirming the reliability and effectiveness of the proposed digital twin-based TCM method.

6.4. Validation results

To further validate the feasibility of our proposed digital twin-based anomaly detection framework, we changed the order of the tests where the data collected from T2 was applied to off-line training and the data from T1 was used for on-line monitoring. Fig. 15(a) shows the SI of the off-line test results where the ERR is 6.65 μm. For the online monitoring, as can be seen in Fig. 15(b), the ERR is 11.05 μm. Both ERRs, including off-line and on-line, are within the error threshold T_e of 20 μm proving that our proposed digital twin-based anomaly detection framework is effective and accurate for TCM in machining.

6.5. Comparative study

To evaluate the performance of the proposed method, a comparative study using the conventional signal feature-based method was conducted, in which the collected signals are directly used for feature extraction, and the extracted feature is the amplitude in the frequency domain of the tooth passing frequency. More discussions of the conventional signal feature-based method can be seen in [30].

First, the dataset from T1 was used to conduct off-line training; and the candidate signal features included mean, kurtosis, entropy, variance, skewness, standard deviation, and median. After using some feature selection methods, the dominant signal features were selected [9]. Then, the selected features were used to train the AI model for TCM.

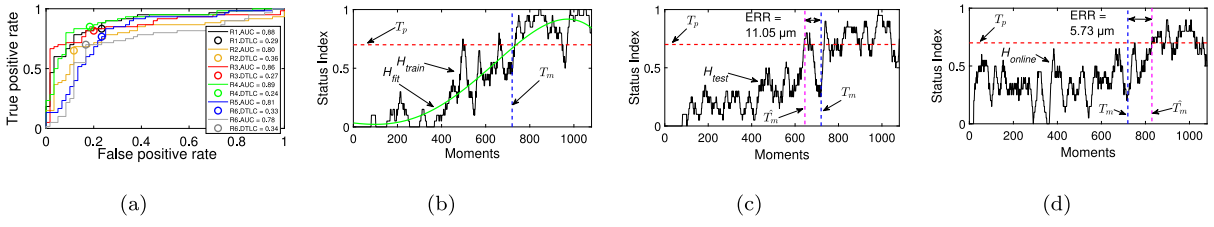

Fig. 14. The result of T1 off-line training and T2 on-line testing. (a) The ROC of the six sub-diagnostic models, (b) the SI of the train datasets, (c) the SI of the off-line test datasets and (d) the SI of the on-line monitoring.

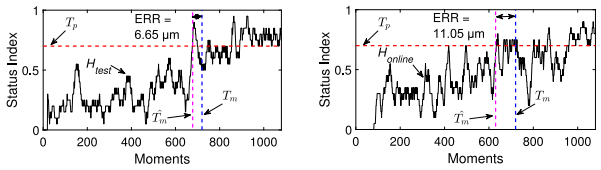

Fig. 15. The result of T2 off-line training and T1 on-line testing (a) the SI of the off-line test dataset and (b) the SI of the on-line monitoring.

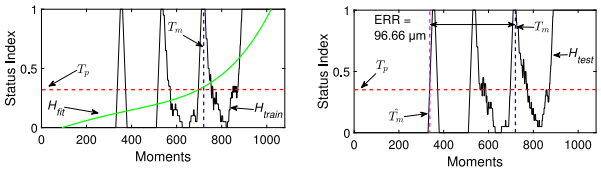

Fig. 16. The result of T1 off-line training using signal feature-based methods (a) the SI of the train dataset and (b) the SI of the off-line test dataset.

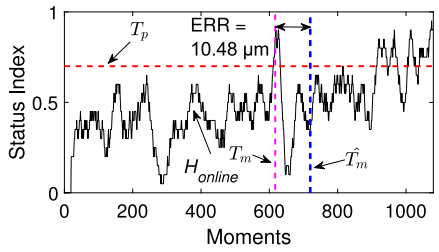

Fig. 17. The SI of the on-line monitoring.

Fig. 16(a) displays the SI of the train dataset. It can be seen that the generated SI randomly produced large fluctuations. This is because machining processes are complicated, and signal features have low robustness and reliability to overcome the variations during machining. Finally, the trained AI model was applied to the test dataset with the aim of validating the performance of the trained model. As presented in Fig. 16(b), the model starts to trigger wear at a very early stage which corresponds to 183.34 μm, resulting in an ERR of 96.66 μm, exceeding the set error threshold T_e of 20 μm. This comparative study underscores the robustness and reliability of NOFRFs-based features, and demonstrates the effectiveness of the proposed digital twin-based anomaly detection framework in real-time TCM applications.

Furthermore, the evaluated model features were used to train the logistic regression classifier instead of the SVM classifier. Fig. 17 shows the SI of online monitoring results where the ERR is 10.48 μm which falls within the predefined error threshold T_e of 20 μm. This results further indicate the robustness and reliability of NOFRFs-based features.

7. Conclusion

In this article, a new digital twin-based anomaly detection framework is proposed to provide a solution for real-time tool condition monitoring (TCM) in machining. The "physical product" of the proposed digital twin consists of the complex dynamics of the machining process, incorporating the dynamics of the spindle, cutting tool, workpiece, milling fixture, and the interactions between the cutting tool and workpiece. The "virtual product" includes an on-line updated data-driven modeling that describes the overall behaviors of the "physical product" in real time, as well as the frequency domain characteristics of the data-driven model, and a diagnostic algorithm that performs TCM based on the frequency domain characteristics of the data-driven model. The "data flow connection" includes measured vibration signals and machine tool numerical controller (NC) signals to facilitate the " virtual product" updating, analysis, and implementation of TCM.

The principal contributions of this novel digital twin framework lie in its high robustness and reliability, which can overcome variations in difficult parts with different machining features in machining and can harmonize the diagnostic results generated by the different machining processes. Experimental studies have demonstrated the effectiveness of the proposed method, especially for complicated machining processes.

Our near future work will be devoted to the application study on more complicated machining parts and extending it to remaining useful life estimation. In the course of this work, research will also be made in the areas of (i) improving the modeling speed of sensor data modeling and model frequency analysis, and (ii) application to other rotating components in manufacturing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Engineering and Physical Sciences Research Council, United Kingdom under Grant No. (EP/T024291/1).

Appendix. Generalized associated linear equations

Consider rewriting the built NARX model (13) into the following discrete-time equation,

$$y_k(t) = \sum_{j=1}^{J}\sum_{p=0}^{j}\sum_{l_1,\ldots,l_{p+q}=1}^{L} c_{p,q,k}(l_1,\ldots,l_{p+q}) \times \prod_{i=1}^{p} y_k(t-l_i)\prod_{i=p+1}^{p+q} u_k(t-l_i) \tag{A.1}$$

where J and L are integers, $p+q=j$, and $c_{p,q,k}(l_1,\ldots,l_{p+q})$ denotes the coefficients of the model at the moment of the kth data collection [31]. The Generalized Associated Linear Equations (GALEs) of the NARX model (A.1) are decomposed as

$$\begin{aligned} y_{n,k}(t) &= \sum_{l_1=1}^{L} c_{1,0,k}(l_1)y_{n,k}(t-l_1) + \sum_{l_1,l_n=1}^{L} c_{0,n,k}(l_1,.,l_n) \\ &\times \prod_{i=1}^{n} u_k(t-l_i) \\ &+ \sum_{q=1}^{n-1}\sum_{p=1}^{n-q}\sum_{l_1,l_{p+q}=1}^{L} c_{p,q,k}(l_1,\ldots,l_{p+q})y^{\mathbf{L}}_{n-q,p,k}(t) \\ &\times \prod_{i=p+1}^{p+q} u_k(t-l_i) + \sum_{p=2}^{n}\sum_{l_1,l_p=1}^{L} c_{p,0,k}(l_1,\ldots,l_p)y^{\mathbf{L}}_{n,p,k}(t) \end{aligned} \tag{A.2}$$

where $n=1,\ldots,N$, $\mathbf{L}=[l_1,\ldots,l_n]$ and

$$\begin{cases} y^{\mathbf{L}}_{n,p,k}(t) = \sum_{i=1}^{n-(p-1)} y_{i,k}\left(t-l_p\right) y^{\mathbf{L}_p}_{n-i,p-1,k}(t) \\ y^{\mathbf{L}}_{n,1,k}(t) = y_{n,k}\left(t-l_1\right) \end{cases} \tag{A.3}$$

Suppose $u^*_k(t)$ is the input excitation signal for NOFRFs evaluation, so the model output $y^*_{n,k}(t)$ can be evaluated using (A.2). Substitute $u^*_{n,k}(t)$ and $y^*_{n,k}(t)$ into (10), the NOFRFs, $G^*_{n,k}(j\omega)$, can be determined as

$$G^*_{n,k}(j\omega) = \frac{\mathcal{F}[y^*_{n,k}(t)]}{\mathcal{F}[u^*_{n,k}(t)]} = \frac{\mathcal{F}[y^*_{n,k}(t)]}{\mathcal{F}\{[u^*_k(t)]^n\}} = \frac{Y^*_{n,k}(j\omega)}{U^*_{n,k}(j\omega)} \tag{A.4}$$

where $U^*_{n,k}(j\bar{\omega}) \neq 0$ and $n=1,\ldots,N$ representing the nth order NOFRFs.

References

[1] Lu Y, Xu X, Wang L. Smart manufacturing process and system automation - A critical review of the standards and envisioned scenarios. J Manuf Syst 2020;56:312–25.

[2] Jung I, Lubich V, Wieland H. Tool failure - causes and prevention. In: 6th international tooling conference. Sweden; 2002, p. 1343–62.

[3] Teti R, Jemielniak K, O'Donnell G, Dornfeld D. Advanced monitoring of machining operations. CIRP Ann 2010;59(2):717–39.

[4] Mohanraj T, Shankar S, Rajasekar R, Sakthivel N, Pramanik A. Tool condition monitoring techniques in milling process - a review. J Mater Res Technol 2020;9(1):1032–42.

[5] Aralikatti SS, Ravikumar K, Kumar H. Fault diagnosis of single-point cutting tool using vibration signal by rotation forest algorithm. SN Appl Sci 2019;1(9):1–8.

[6] Huchel Ł, Krause TC, Ługowski T, Leeb SB, Helsen J. Chasing the cut: A measurement approach for machine tool condition monitoring. IEEE Trans Instrum Meas 2020;70:1–10.

[7] Chen B, Chen X, Li B, He Z, Cao H, Cai G. Reliability estimation for cutting tools based on logistic regression model using vibration signals. Mech Syst Signal Process 2011;25(7):2526–37.

[8] Sun C, Ma M, Zhao Z, Tian S, Yan R, Chen X. Deep transfer learning based on sparse autoencoder for remaining useful life prediction of tool in manufacturing. IEEE Trans Ind Inform 2018;15(4):2416–25.

[9] Liu Z, Lang Z-Q, Zhu Y-P, Gui Y, Laalej H, Stammers J. Sensor data modeling and model frequency analysis for detecting cutting tool anomalies in machining. IEEE Trans Syst Man, Cybern: Syst 2023;53(5):2641–53.

[10] Grieves M. Digital twin: manufacturing excellence through virtual factory replication. White Pap 2014;1(2014):1–7.

[11] Qin Y, Wu X, Luo J. Data-model combined driven digital twin of life-cycle rolling bearing. IEEE Trans Ind Inf 2021;18(3):1530–40.

[12] Glaessgen E, Stargel D. The digital twin paradigm for future NASA and US air force vehicles. In: 53rd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference 20th AIAA/ASME/AHS adaptive structures conference 14th AIAA. 2012, p. 1818.

[13] Tao F, Zhang H, Liu A, Nee AY. Digital twin in industry: State-of-the-art. IEEE Trans Ind Inform 2018;15(4):2405–15.

[14] Moghadam FK, Nejad AR. Online condition monitoring of floating wind turbines drivetrain by means of digital twin. Mech Syst Signal Process 2022;162:108087.

[15] Wang Y, Tao F, Zhang M, Wang L, Zuo Y. Digital twin enhanced fault prediction for the autoclave with insufficient data. J Manuf Syst 2021;60:350–9.

[16] Tao F, Zhang M, Liu Y, Nee AY. Digital twin driven prognostics and health management for complex equipment. Cirp Ann 2018;67(1):169–72.

[17] Xia M, Shao H, Williams D, Lu S, Shu L, de Silva CW. Intelligent fault diagnosis of machinery using digital twin-assisted deep transfer learning. Reliab Eng Syst Saf 2021;215:107938.

[18] Qianzhe Q, Wang J, Ye L, Gao R. Digital twin for machining tool condition prediction. Procedia CIRP 2019;81:1388–93.

[19] Luo W, Hu T, Ye Y, Zhang C, Wei Y. A hybrid predictive maintenance approach for CNC machine tool driven by digital twin. Robot Comput-Integr Manuf 2020;65:101974.

[20] Scott R. Untangling the requirements of a digital twin. The University of Sheffield AMRC; 2020.

[21] Qi Q, Tao F, Hu T, Anwer N, Liu A, Wei Y, Wang L, Nee A. Enabling technologies and tools for digital twin. J Manuf Syst 2021;58:3–21, Digital Twin towards Smart Manufacturing and Industry 4.0.

[22] Cai Y, Starly B, Cohen P, Lee Y-S. Sensor data and information fusion to construct digital-twins virtual machine tools for cyber-physical manufacturing. Procedia Manuf 2017;10:1031–42.

[23] Billings SA. Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains. John Wiley & Sons; 2013.

[24] Lang Z, Billings S. Energy transfer properties of non-linear systems in the frequency domain. Internat J Control 2005;78(5):345–62.

[25] Peng Z, Lang ZQ, Billings SA. Non-linear output frequency response functions for multi-input non-linear Volterra systems. Internat J Control 2007;80(6):843–55.

[26] Chen S. Local regularization assisted orthogonal least squares regression. Neurocomputing 2006;69(4):559–85.

[27] Zhu Y, Lang Z, Mao H, Laalej H. Nonlinear output frequency response functions: A new evaluation approach and applications to railway and manufacturing systems' condition monitoring. Mech Syst Signal Process 2022;163:108179.

[28] Wen X, Liu J, Du C, Qu P, Sheng S, Liu X, Zhou H, Kang C, Chen Y, Dong L. The key technologies of machining process design: a review. Int J Adv Manuf Technol 2022;1–19.

[29] Liu C, Li Y, Li Z. A machining feature definition approach by using two-times unsupervised clustering based on historical data for process knowledge reuse. J Manuf Syst 2018;49:16–24.

[30] Qin X, Huang W, Wang X, Tang Z, Liu Z. Real-time remaining useful life prediction of cutting tools using sparse augmented Lagrangian analysis and Gaussian process regression. Sensors 2022;23(1):413.

[31] Billings S, Peyton Jones J. Mapping non-linear integro-differential equations into the frequency domain. Internat J Control 1990;52(4):863–79.