

Hybrid Data-driven and Model-informed Online Tool Wear Detection in Milling Machines

Qian Yang^{a,*}, Krishina Pattipati^a, Utsav Awasthi^b, George M. Bolas^b

^a*Department of Electrical & Computer Engineering, University of Connecticut, 371
Fairfield Way, Storrs, CT 06269, USA*

^b*Department of Chemical & Biomolecular Engineering, University of Connecticut, 159
Discovery Dr., Storrs, CT 06269, USA*

Abstract

Precision machining tool wear is responsible for low product throughput and quality. Monitoring the tool wear online is vital to prevent degradation in machining quality. However, direct real-time tool wear measurement is not practical. This paper presents residual-based anomaly detection models, combining a hybrid model comprised of a physics-based model and a data-driven model (a decision tree or a neural network) to predict signals of interest (e.g., power or forces) under nominal conditions, followed by Page's cumulative sum test for detecting tool wear on-line using the computer numerical control machine measurements. The most informative features are ranked using dynamic programming and its approximation variants from real-time measurements and machine settings, such as the width of cut, depth of cut, feed rate and spindle speed, that serve as inputs to the predictive models. The baseline nominal model is incrementally updated with experimental data via a gradient boosted adaptation model to generate the residuals that account for discrepancies between the actual machine data under normal conditions and the baseline nominal model predictions. The hybrid model is validated against 20 Mazak milling machine experimental tests and one Haas run-to-failure experiment. The proposed anomaly detector is applied to synthetic data from simulations of the physics-based model at different operating conditions, measurement noise levels, and tool wear levels, and the methods were able to achieve an overall 92% accuracy in data with 1% noise. The anomaly detection methods based on hybrid model reduced the false alarms of either the data-driven or physical-based models alone, and are found to be capable of good online detection of tool wear.

Keywords: Online anomaly detection, Hybrid Physics-based and Machine Learning Models, Decision tree, Neural network, Cumulative sum test

Nomenclature

*Corresponding author

Email address: qian.4.yang@uconn.edu (Qian Yang)

Preprint submitted to Journal of Manufacturing Systems

March 30, 2022

35			K_s	Force coefficient (N/mm ²)
	$\bar{h}$	Mean chip thickness (mm)	K_t	Tangential force component coefficient (N/mm ²)
	$\bar{p}_i$	Idle-running power of a spindle motor (kW)	65 M	Torque (N-m)
	μ	Friction coefficient	N_t	Number of teeth
40	Φ	Tooth pitch (degree)	P	Mean cutting power (kW)
	ϕ	Tool rotation angle (degree)	r	Radius of the tool (mm)
	ϕ_c	Angle between the net cutting force (degree)	v_s	Spindle speed (RPM)
	ϕ_e	Exit angle (degree)	70 W	Tool wear (mm)
45	ϕ_r	Rake angle (degree)	ANN	Artificial neural networks
	ϕ_s	Entry angle (degree)	CNC	Computer Numerical Control
	A	Chip area (mm ²)	CNN	Convolutional Neural Networks
	a_e	Width of cut (mm)	CUSUM	cumulative Sum
	a_p	Depth of cut (mm)	75 DCNN	Deep convolutional neural network
50	c_1-c_8	Regression coefficients	DP	Dynamic programming
	D	Diameter of the tool (mm)	DT	Decision tree
	f	Feed rate (mm/min)	LSTM	Long Short-Term Memory
	F_n	Normal force component (N)	MC	Monte Carlo
	F_t	Tangential force component (N)	80 MCA	Morphological component analysis
55	f_t	Feed rate per tooth (mm/min)	MLP	Multi-layer perceptron
	F_x	x - direction force component (N)	MSE	Mean squared error
	F_y	y - direction force component (N)	MVR	Multivariate regression
	H	Material hardness (HRC)	NN	Neural network
	h	Instantaneous chip thickness (mm)	85 SVM	Support vector machine
60	K_n	Normal force component coefficient (N/mm ²)	TCM	Tool condition monitoring

1. Introduction

1.1. Motivation

Computer Numerical Control (CNC) machining plays an important role in modern digital manufacturing. CNC machining is used for shaping metals or other rigid materials, usually by cutting, boring, grinding, shearing, or other forms of subtractive deformation [1]. Currently, milling is the second most common method for machining custom parts to precise tolerances [2]. Face milling, due to its ability to quickly cut and shape large objects or large surfaces, is widely used to cut a flat surface, and can be performed using a wide range of tools. The cutting tool in a milling machine is regularly subjected to high stress and temperature gradients near its surface due to metal-to-metal contact with the workpiece. Tool wear is a gradual process due to regular machining operations, and can be of different types: abrasion, adhesion, diffusion, chemical wear and fatigue. Forms of tool wear include flank and crater wear, thermal cracking and tool fracture, thermal and mechanical load wear, edge chipping and entry or exit failures. The average width of the flank wear land is the most common variable used to evaluate the tool wear status. Blunt tools impact the product quality and machining efficiency [3]. For example, unwanted vibrations of blunt tools lead to poor quality in surface finish and inaccurate dimensions [4]. In a modern manufacturing industry, nearly 79.6% of the downtime of a machine tool is caused by these mechanical failures [5]. Consequently, incipient detection of tool wear in milling machines is salient for minimizing the cost of scrap and for reducing down-time [6]. Modern milling machines are able to measure operating signals, such as machine power, cutting force, feed rate, spindle speed, width and depth of cut [3], which are used to characterize the state of the tool and for early detection of anomalies in machine's state of operation [4]. Among the sensors mentioned above, the power sensor is cheap, non-intrusive, and is convenient to acquire [7][8][9]. Fusion of these data with domain expert knowledge is, however, at its infancy in the manufacturing industry and is the focus of this work.

1.2. Previous Work

There are two distinct approaches to monitor a milling machine tool condition: Direct and Indirect. Direct approaches measure the geometric parameters of the cutting tool or the work piece, while the Indirect approaches monitor and translate more commonly available machine sensor signals into a tool wear status estimate [10]. Few image-based direct online tool condition monitoring (TCM) methods that involve measuring the flank wear and machined parts have been investigated [11]. T. Pfeifer *et al.* [12][13] presented a CCD-array camera and a vision system to monitor the tool flank wear from the processed image. Wang *et al.* [14] used a neural network, together with a computer vision system, to estimate flank wear. Specifically, a morphological component analysis (MCA)-based tool wear monitoring system was proposed to deal with noise, blurred boundary, and misalignment of captured images [15]. A novel shape-based descriptor B-ORCHIZ was introduced [16] and a support vector machine

130 (SVM) was used to discriminate among two (low-high) and three (low-medium-high) different tool wear levels. However, it is usually hard to measure the geometric parameters instantaneously and the inconsistency due to variation in illumination, translation and rotation of the captured images prevent the direct methods from being implemented in the industry [4].

135 Indirect monitoring approaches estimate tool wear based on sensed measurements, such as the power consumed, cutting forces in the x - y - z directions, vibrations, acoustic emissions, spindle motor and feed currents [1][4][16]. Liu *et al.* [17] evaluated the impact of different levels of cutting tool wear on energy consumption and built a predictive model to describe the relationship between
140 the machine's energy consumption and the process and environmental variables. Spindle power data analyzed and evaluated by an artificial intelligence system to monitor the tool wear in real time [9]. Cutting forces are reliable indicators for online tool condition monitoring and were used in [11], where the cutting force signatures in both time and frequency domain were extracted and changes
145 in the variance and the first harmonic component of the cutting forces were related to the flank wear. Nouri *et al.* [4] introduced a real-time monitoring method to detect the milling tool wear by tracking the coefficients of a model of the cutting force and demonstrated its effectiveness on multiple tool wear experiments using different cutting conditions and materials. Bustillo *et al.* [18]
150 presented Random Forest ensembles combined with a Synthetic Minority Over-sampling Technique for the prediction of the flatness deviation caused by the tool wear in face milling. Zhou *et al.* [19] proposed a two-layer network for tool condition monitoring using acoustic sensor signals. The method used two-layer angle kernel functions without hyperparameters and avoided the complications
155 in conventional kernel functions.

Multi-sensor fusion improves tool condition monitoring over a single sensor approach. Ghosh *et al.* [20] fused features extracted from cutting forces, spindle vibration, spindle current, and sound pressure level to estimate the average flank wear using a neural network. A multi-sensor tool wear prediction system
160 utilizing a third order regression model was investigated [10]. Acceleration and electrical current signals were fused for tool wear prediction. Han *et al.* [21] fused the audio and vibration signals to monitor the machine state using three methods, including K-nearest neighbor, convolutional neural networks and support vector machines. Nowadays, embedded sensors are used to monitor tool
165 wear in CNC milling machines. An artificial neural network (ANN) was trained with data from a built-in vibration sensor to classify the tool wear state [3]. Palanisamy *et al.* [22] presented regression and ANN models to predict the tool flank wear in milling machines and Six Sigma software was used to determine the input features for experimentation. It was claimed that the ANN model
170 was much more robust and accurate in estimating tool wear when compared to traditional multivariate regression (MVR) models. Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) are the most popular deep learning models for tool wear prediction. Huang *et al.* [23] presented a deep convolutional neural network (DCNN)-based tool wear prediction method by
175 fusing multi-domain features, which was shown to improve prediction accuracy.

Cai *et al.* [24] presented a hybrid information system, which utilized a stacked long short-term memory network (LSTM) for tool wear prediction.

Most of the existing methods are based on the stationarity assumption (i.e., constant speed and load) and, as such, are only suitable for fault detection under constant operating conditions [25]. In production machining processes, however, the working conditions are affected by many factors, including dissimilar cutting tools, diverse work pieces, varying cutting parameters, different machining loads, time varying speeds, to name a few [26]. This variability results in non-stationary signals, requiring context-dependent models. Bustillo *et al.* [27] studied the modelling of machining repetitions using regression trees, kNNs, artificial neural networks, bagging and Random Forests. This work predicted tool-life in face-turning operations under different machining conditions and showed that the use of raw experimental data gave higher accuracy when compared to their averaged values. Pimenov *et al.* [28] presented a study to predict the surface roughness using main drive power and monitoring the wear on face milling teeth online. Meta-learning [29] and transfer learning [30] methods became popular in solving the tool condition monitoring problem by demonstrating their ability to learn the hidden rules behind a variety of experiments with small samples. A tool wear prediction method based on meta learning was proposed by Li *et al.* [31], which adapted to new machining tasks of different cutting conditions and predicted the tool wear status with enhanced accuracy. A transfer learning approach, combined with Convolutional Neural Networks, was proposed by Mohamed *et al.* [30] to predict the health state of cutting tools.

It should be noted that the physics-based models are subject to modeling uncertainty, while the data-driven prediction models are affected by measurement noise. To improve the diagnostic and prognostic accuracy, hybrid techniques, which fuse data-driven and physics-based models, were presented in recent works. Hanachi *et al.* [32] combined particle filters and a numerical inference model, and reduced the tool wear prediction errors by almost half. To predict the tool wear, a physics-guided neural network model was presented by Wang *et al.* [33], and eliminated the inconsistency in conventional data-driven models. Compared to the data-driven or physics-based approaches in isolation, a hybrid model-based fusion framework is a relatively unexplored area in machining process monitoring. The framework presented here augments the physics-based model predictions with an incremental data-driven model predictions, and thereby demonstrates a significant improvement in spindle power estimation and a consequent reduction in false alarms.

1.3. Contributions and Organization of the Paper

This paper presents a hybrid monitoring methodology, combining data-driven and model-informed methods, using power consumption and cutting force as sensed variables (we focus on power consumption for illustrative purposes in the paper). The methodology presented can be applied to different modes of operation and cutting conditions. The objective is to determine the level of the geometrically averaged flank tool wear based on the spindle power or force signal

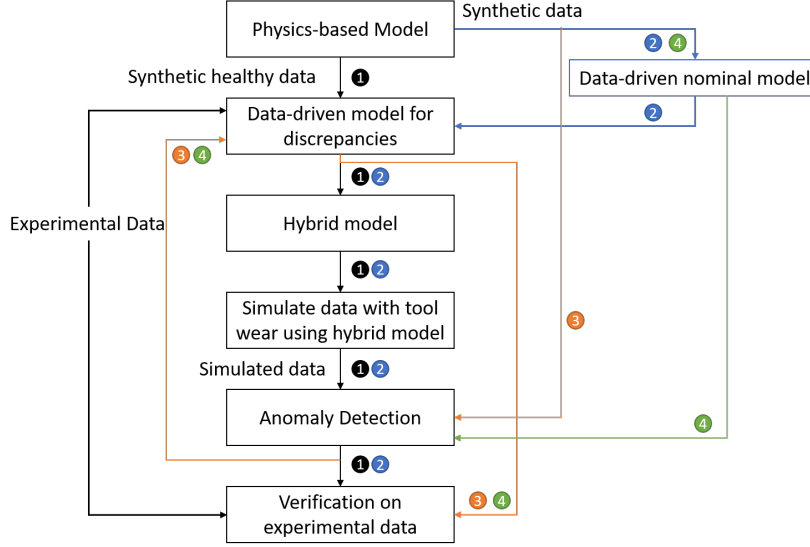


Figure 1: Potential implementations of the hybrid model

residuals. Monitoring of the tool condition is accomplished by using a digital twin of machining process in the form of a hybrid model that predicts power and force measurements. Although extensive research efforts have been made on analytical expressions of the milling process, complete and accurate physics-based models that capture all aspects of milling operations do not exist. The main difficulty is that there are too many uncontrollable or immeasurable internal and external variables that would have to be monitored and considered. Moreover, process parameters and workpiece and cutting tool condition vary with time. On the other hand, purely data-driven models miss key physics and do not capture what is known and well-understood about the physics of milling. A hybrid model promises to merge the powers of these modeling approaches and is the key focus of this work. The hybrid model, proposed here, can be potentially implemented in the four ways shown in Fig. 1. We will present the first and second approaches of the following in this paper: Hybrid Physics-based and Incremental Data-driven Model Validated on Healthy Data; Combined Data-driven and Incremental Data-driven Model Validated on Healthy Data.

The rest of this paper is organized as follows: Section 2 describes the Physics-based model used to generate the synthetic nominal data; it briefly presents the key equations of the analytical model. In Section 3, our modeling and anomaly detection process are outlined and details of the proposed residual-based method for incipient anomaly detection are briefly introduced. In Section 4, the performance of the residual-based anomaly detector is evaluated using simulation and experimental data. Finally, conclusions and future research directions are presented in Section 5.

245 2. Physics-based Model

In this section, we briefly present the analytical model for the prediction of power and force in milling as a function of inputs (spindle speed, feed rate, width of cut and depth of cut) and machine parameters (number of teeth, diameter of cutter, offset distance between workpiece and tool, workpiece dimensions, etc.).
 250 More details on this semi-empirical model, along with an extensive validation against actual machining data, can be found in [34].

The face milling model is derived from the original equations presented in [35]. The model calculates the power consumption and directional force components. Milling depends on the radial immersion of the tool in the workpiece, i.e., the overlap between tool and workpiece, and on the diameter of the tool, 255 D , and the width of cut, a_e . The tool rotation angle, ϕ , varies from 0° to 360° , depending on the tooth pitch Φ ($= 360/N_t$, where N_t is the total number of teeth in the tool) and the immersion of the tool. The type of milling considered for data generation is Up milling. The entry or start angle, $\phi_s = 0^\circ$ and the exit angle $\phi_e = \cos^{-1}((r - a_e)/r)$ depend on the tool immersion.
 260

The net cutting force can be approximated as the product of chip area, A , and the specific force, K_s , applied, as shown in (1):

$$F = K_s A = K_s a_p h \quad (1)$$

where h is the instantaneous chip thickness and is calculated as $h = f_t \sin(\phi)$, and a_p is the depth of cut. The feed rate per tooth, f_t , can be computed by the feed rate, f , spindle speed, v_s , and number of teeth, N_t , as $f_t = f(v_s N_t)^{-1}$. In orthogonal cutting, the tool is perpendicular to the direction of motion and hence the cutting force is divided into the normal force component, F_n , and the tangential force component, F_t , as shown in (2) and (3):

$$F_n = F \cos(\phi_c) = K_s a_p h \cos(\phi_c) = K_n a_p h \quad (2)$$

$$F_t = F \sin(\phi_c) = K_s a_p h \sin(\phi_c) = K_t a_p h \quad (3)$$

where ϕ_c is the angle between the net cutting force, F , and the normal force component F_n .

In face milling, because of the forces involved in cutting, the tool gets worn out. This wear can result in the breaking of the tool or improper surface finish of the work piece; hence, tool wear can be considered a fault. To capture the effect of tool wear in the model, the force coefficient, K_n , is computed using the empirical correlation shown in (4), as proposed by [36]:

$$K_n = c_1 [1 + c_2 (W - 0.1)] \bar{h}^{-c_3} N_t^{c_4} \quad (4)$$

where c_1 , c_2 , c_3 , c_4 are regression coefficients, N_t is the number of teeth, $\bar{h}$ is the mean chip thickness and W is the tool wear. Similarly, the tangential force component coefficient, K_t , is computed using the empirical correlation as proposed by [37], shown in (5) :

$$\ln K_t = c_5 + c_6 \ln \bar{h} + c_7 \ln v_c + c_8 \sin(\phi_r) \quad (5)$$

where c_5 , c_6 , c_7 and c_8 are regression coefficients, v_c is the cutting speed, and ϕ_r is the rake angle, which is the angle the cutting surface of the tool makes with the perpendicular to the cutting edge. The tangential force equation (3), is also updated to capture the effect of tool wear [37], as shown in (6):

$$F_t = K_t A + \mu H W a_p \quad (6)$$

where H is the material hardness. The mean chip thickness, $\bar{h}$, can be calculated by integrating the instantaneous chip thickness over the immersion of tool in the workpiece [34].

In the case of a tool with multiple teeth, depending on tool immersion and tooth pitch, multiple teeth can perform cutting simultaneously. If the rotation angle of the first tooth is ϕ , then the angle of every subsequent tooth j , is $\phi_j = \phi + j\Phi$, for all $2 \leq j \leq N_t$. The normal cutting force and tangential cutting force are calculated when the cutting angle ϕ_j of each tooth lies between the start angle, ϕ_s , and exit angle, ϕ_e , and is zero otherwise. The conditional equations in (7) calculate the normal force and the tangential force for each tooth j :

$$\begin{aligned} F_{n,j} &= K_n a_p f_t \sin(\phi_j) & \phi_s \leq \phi_j \leq \phi_e \\ F_{t,j} &= K_t a_p f_t \sin(\phi_j) + \mu H W a_p & \phi_s \leq \phi_j \leq \phi_e \\ F_{t,j} &= 0, F_{n,j} = 0; & \text{otherwise} \end{aligned} \quad (7)$$

A dynamometer can measure the forces in the x and y directions; hence, the tangential and normal cutting forces are expressed in a fixed frame of reference. The directional force components in this frame of reference are obtained by summing the projection of tangential and normal cutting forces over every single tooth cutting at an instant. Equations (8) and (9) below show the cutting forces in the x and y directions:

$$F_x = \sum_{j=1}^{N_t} F_{t,j} \cos(\phi_j) + F_{n,j} \sin(\phi_j) \quad (8)$$

$$F_y = \sum_{j=1}^{N_t} F_{t,j} \sin(\phi_j) - F_{n,j} \cos(\phi_j) \quad (9)$$

The mean cutting power can be calculated by adding the idle running power or power consumed in air cutting, $\bar{p}_i$, to the product of torque, M , and rotational spindle speed, v_s :

$$P = \frac{2\pi M v_s}{60} + \bar{p}_i \quad (10)$$

where the torque, M , is calculated as the product of tangential force and the radius of the cutter, $D/2$, as shown in (11):

$$M = \sum_{j=1}^{N_t} (D/2) F_{t,j} \quad (11)$$

Table 1: System inputs, outputs and parameters

Inputs	
v_s	Spindle speed (RPM)
f	Feed rate (mm/min)
a_e	Width of cut (mm)
a_p	Depth of cut (mm)
Outputs	
$\bar{P}$	Mean cutting power (kWatt)
F_x	x - direction force component (N)
F_y	y - direction force component (N)
Parameters	
N_t	Number of teeth
D	Diameter of milling cutter (mm)
$\bar{p}_i$	Idle-running power of a spindle motor (kWatt)
μ	Friction coefficient
H	Material hardness (HRC)
$c_1 - c_8$	Regression coefficients
Faults	
W	Tool wear (mm)

For the cutting to take place, the spindle must provide sufficient torque against the tangential cutting force. Similar to the force components, the torque is summed over the number of teeth engaged in the cut based on the cutting angle, ϕ_j . The inputs, outputs, parameters and faults (in the form of model parameters) of the physics-based model of precision machining are summarized in Table 1. This physics based model was validated against experimental data of power consumption and used parameter estimation to determine the regression coefficients in (4) and (5), [34].

3. Methodology

The proposed framework for anomaly detection enables us to answer two salient architectural questions: (a) what is the optimal ranking of input features in terms of their ability to distinguish between healthy and faulty tools? (b) what is the minimal tool wear that is detectable by the anomaly detection method at different sensor noise levels? Our tool wear monitoring method associated with tool wear detection is as follows:

- **Synthetic data generation:** Devise a database for the variability and operating envelope of the system of interest. Specifically, define the input space (discrete or continuous), the known or anticipated variance of system (or model) parameters and boundaries, and the noise in system outputs. Use the physics-based model to simulate the power consumption and cutting force for various combinations of feed rate, depth of cut, width of cut and cutting speed that cover the input space. After injecting sensor noise into the resulting data, we form the synthetic database of healthy machine outputs.

- **Train nominal machine model:** Using the synthetic data, train a decision tree (DT) or a neural network (NN) model to predict the power consumption and cutting force. If the physics-based nominal model is computationally feasible to simulate in real-time (as is the case here), it can be used to predict the healthy data as well.
- **Incremental model update to match the healthy experimental data:** Generate experimental data under healthy conditions covering the operating envelope. Train an incremental data-driven DT or NN model to account for discrepancies between the experimental data and model predictions (physics-based or data-driven model of the previous step).
- **Generate tool wear data:** Generate a new synthetic database using the hybrid model (physics-based + incremental data-driven model) that captures how tool wear propagates through the machining system to be evidenced in outputs.
- **Feature ranking:** Use the Kullback-Leibler divergence measure between the healthy and faulty residuals of the aforementioned database or feature subsets in a dynamic programming algorithm or its approximate variants to rank order the features for anomaly detection.
- **Anomaly detection:** Use the cumulative Sum (CUSUM) test [38][39] on residuals of the synthetic database to detect the tool wear degradation.
- **Performance evaluation and model selection for deployment:** Evaluate the performance of CUSUM test under various tool wear conditions to quantify the minimal detectable tool wear. Select the best prediction model (physics-based nominal model + incremental data-driven model or nominal data-driven model + incremental data-driven model) by comparing their anomaly detection performance in terms of missed detections and false alarms.

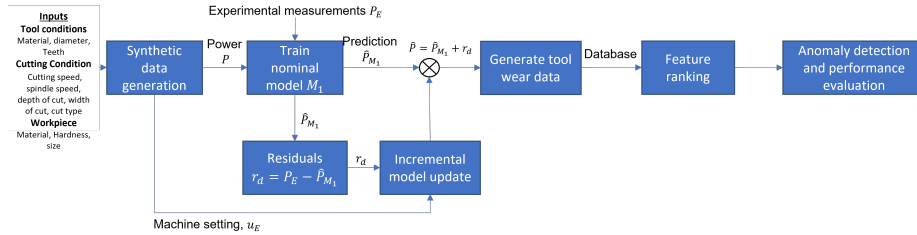


Figure 2: The proposed framework for anomaly detection.

3.1. Synthetic data generation

Monte Carlo (MC) simulation was used to generate the nominal and anomalous synthetic data. The efficiency of MC simulation stems from its use of a large

320 number of pseudo-random numbers to simulate a complex system with uncertain input variables chosen from user-defined probability distributions [40][41]. Thus, to conduct a Monte Carlo simulation, one needs a model of the system and user-selected probability distributions for the model inputs.

325 Using the theoretical milling force model described in Section 2, combined with an incremental data-driven model to account for discrepancies between the nominal physics-based model predictions and the nominal (healthy) experimental data (see below), distributions of the spindle power and forces in the x and y directions were calculated for the cutting conditions listed in Table 2. For the MC simulation, different geometric parameters and process parameters, such as 330 the cutting speed, v_s , feed rate, f , width of cut, a_e and depth of cut, a_p , on the power and force signals were propagated thorough the model of Section 2 and the incremental data-driven model to calculate their impact on power and forces.

Table 2: Machining conditions studied

Depth of cut (mm)	$a_p \sim U(0.01, 10)$
Feed rate (mm/rev)	$f \sim U(200, 800)$
Spindle speed (rpm)	$v_s \sim U(1000, 2000)$
Width of cut (mm)	$a_e \sim U(5, 63.6)$
Offset distance (mm)	$\delta = 0$
Tool wear (mm)	$W = [0, 0.01, 0.02, 0.03, \dots, 0.1]$

335 One thousand Monte Carlo simulations were run at each of the 11 different levels of tool wear of Table 2. For the healthy condition, tool wear is zero; for the faulty conditions, tool wear was set from 0.01 mm to 0.1 mm, for a total of 10 distinct tool wear values. Each simulation was run for a machining operation of 100 seconds and 500 samples of power consumption (P) and forces in x (F_x), y (F_y) directions were collected (sampling interval 0.2 sec). Thus, we 340 collected a total of 550,000 data points over all simulation runs: 100 (runs) $\times$ 500 (number of samples of the signal) $\times$ 11 (values of tool wear) W . A sample time sequence from a single Monte Carlo run of the model is shown in Fig. 3a.

345 Anomaly detection methods are often sensitive to sensor noise. The impact of sensor noise on the method proposed here was studied with a zero mean Gaussian noise with a standard deviation of 0.1% to 3% of the magnitude of the signals injected into the synthetic data generated with the model. A sample of the impact of noise on the power measurement for the same simulation of Fig. 3a of tool wear $W = 0$ mm shown in Fig. 3b. Here we have illustrated the impact of 0.1%, 1% and 3% noise also to show their effects on the power signal.

350

3.2. Train the nominal machine model

In the proposed anomaly detection algorithm, a model built from nominal data is used to predict the outputs of the fault-free milling processes. A model with lower prediction error is preferred. Here, we employ decision trees and

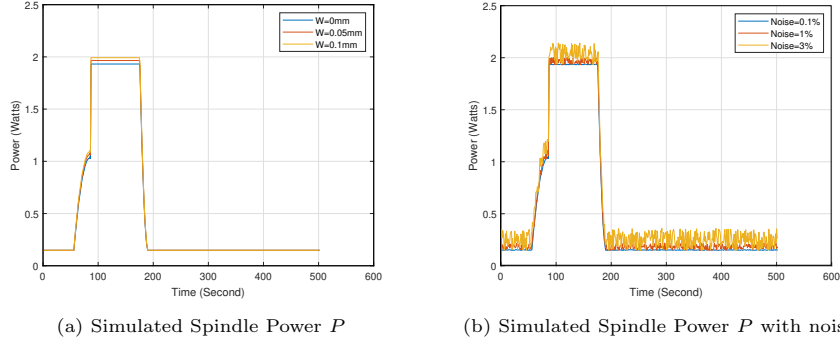


Figure 3: Power signal of simulation from the physics-based model for validation; Run # 1: $a_p = 1.5875$ mm, $a_e = 31.75$ mm, $v_s = 1299$ RPM, $f = 647.7$ mm/min

neural networks to predict the spindle power or the discrepancy between the experimentally measured spindle power and the nominal model predictions from the physics-based or the data-driven model based on depth of cut, width of cut, feed rate, and cutting speed as input features. The residuals of spindle power, which are the deviation of the predictions from the actual measurements, are generated and are used as indicators for anomaly detection.

3.2.1. Decision Tree-based Regression Model

DTs are non-parametric supervised learning methods, and are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to achieve a goal, but are also popular in machine learning. Decision trees have advantages over other regression models, because they can handle large datasets and are invariant to scaling of data. They are intuitive and are able to handle missing variables through surrogate splits. They also do not require data pre-processing to remove redundant variables, and result in explainable models [42].

Suppose that we have a scalar output (also termed dependent or response) variable Y , and a p -vector of explanatory (also termed independent or feature or parameter) variables, X . Assume $Y \in A$. The regression tree partitions A into disjoint regions A_k and provides a fitted value, usually the mean value, $E(Y | X \in A_k)$ within each feature region A_k indexed by node k . To determine how to split region A_k , the weighted mean squared error (MSE) of the responses in region A_k is calculated using (12):

$$\varepsilon_k = \sum_{j \in K} w_j (y_j - \bar{y}_k)^2 \quad (12)$$

where w_j is the weight of observation j , and K is the set of all observation indices in region A_k . Here, the default $w_j = 1/n$ is used, where n is the sample

size at the root node. The probability that an observation is in region A_k is

$$P(A_k) = \sum_{j \in K} w_j \quad (13)$$

370 Model complexity and accuracy, which are competing characteristics of the DT model, need to be simultaneously considered as stopping rules [43]. Since continuous attributes have noise, the model could fit the noise in addition to the structure in the model. A complex model would cause over-fitting and is less reliable when predicting testing data. One can prevent over-fitting by: (i) 375 stopping early (e.g., by constraining the leaf nodes to have a minimum number of records or by restricting the depth of any leaf node from the root node); (ii) pruning back the full tree to an appropriate size using cost-complexity pruning or error-based pruning [44]; or (iii) using ensemble trees to reduce variability in predictions. The latter include bagging, random forests, gradient boosting 380 and adaptive boosting [45][46]. Although decision trees have high accuracy in prediction, there are some limitations in the functions they are able to approximate due to the simple models used in their leaves. Nevertheless, they provide explainable models that have low computational complexity in terms of running time and storage [47].

385 3.2.2. Neural Network Models

To investigate the effect of different regression models on the anomaly detection accuracy, a multi-layer perceptron (MLP) model is developed. MLP is a feed-forward network with an input layer, one or more hidden layers and a fully connected output layer. MLPs can ideally approximate any function with 390 desired accuracy, given enough hidden neurons and data [48]. Due to their ability to learn complex functional relationships among multiple variables, neural networks have performed well in applications without prior knowledge on the relationship between inputs and outputs [49].

The network includes a large number of interconnected neurons. Each neuron receives one or more inputs and predicts an output signal through the mapping function. The mapping function g is unknown, but is approximated by a MLP. In layer j , the input vector x_j is an output of the previous layer ($j - 1$), and is multiplied by a weighted vector w_{ij} . Then the output z_i of neuron i of layer j is given by an activation function Ψ as in (14):

$$z_i = \Psi\left(\sum_{j=1}^n x_j w_{ij}\right) \quad (14)$$

The weight parameters w_{ij} are optimized to minimize the mean square error 395 between the observed output and model predictions via a stochastic gradient descent method or their accelerated versions [50][51][52]. The gradient is computed via the back propagation method [52]. The number of hidden layers applied is determined by estimating the generalization error of each network. For the current application to power consumption and force signals of the precision 400 machining process, it was found that one hidden layer is adequate.

3.3. Incremental model update to match healthy experimental data

3.3.1. Healthy experimental data

The physics-based model was tuned and validated on a series of face milling tests conducted using a Mazak Variaxis 630-5X II T 5-axis legacy CNC machine in dry milling, with the tool properties shown in Table 3. Flat plate stocks ($254 \times 254 \times 38.1 \text{ mm}^3$) of AISI 4340 material with hardness HRC20 and a five-teeth cutting tool from KENNAMETAL were used. To eliminate z -direction forces, 90° insert angle was used. The Mazak has an inbuilt controller to control the milling or turning operation and acquires in-process data for machine power, spindle power, spindle vibration. Extra sensors, such as a Dynamometer and infrared sensor, are added to measure the force and the temperature. The experimental database was collected for different machine settings, including repeatability tests. This resulted in 20 data sets of single-pass linear cuts (see Table 4). Machine data were acquired at a sampling frequency of 50 Hz. All the 20 experimental tests were recorded with a healthy tool. More information about the machining tests and test process design can be found in [34].

Table 3: Tool properties for Mazak experimental tests.

Cutting Tool	Diameter	63.5 mm
	Number of teeth	5
	Insert angle	90°
	Max. cutting power	40 hp
	Max. spindle speed	8000 RPM
Material AISI 4340	Composition	0.4% C, 0.8% Cr, 0.25% Mo and 1.8% Ni
	Hardness	20 HRC
	Dimensions	$254 \times 254 \times 38.1 \text{ mm}^3$

Table 4: Settings of machining experiments with a healthy tool

No.	Cutting speed v_s (RPM)	Feed rate f (mm/min)	Depth of cut a_p (mm)	Width of cut a_e (mm)
1-6	1299	647.7	1.5875	31.75
7-12	1299	647.7	2.54	31.75
13	2000	200	2.54	34.3
14	1500	800	2.54	5
15	1500	200	2.54	63.5
16	1000	500	2.54	5
17	1000	500	2.54	63.5
18	1500	500	2.54	34.3
19	2000	800	2.54	34.3
20	1500	200	2.54	5

The experimental data was used to train an incremental data-driven model that augments the predictions of an existing data-driven model or to augment the predictions of a physics-based model. Fig. 4 shows a sample of the Mazak spindle power, its prediction by the physics-based model and the hybrid model (physics-based model + incremental data-driven model). Overall, the spindle power agreement between models and machine data was very good.

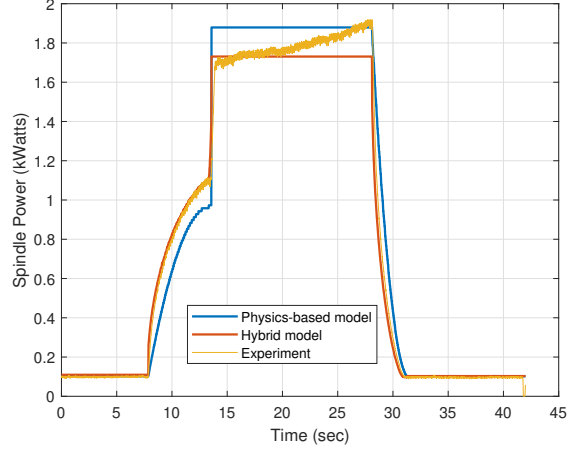


Figure 4: Sample spindle power measurement from the Mazak machine and physics-based and hybrid model predictions; Run #1: $a_p = 1.5875$ mm, $a_e = 31.75$ mm, $v_s = 1299$ RPM, $f = 647.7$ mm/min.

3.3.2. Incremental model update

Due to errors and simplifications in the physics-based model of the machine dynamics and process noise, the physics-based model does not describe the cutting behavior well. Our approach is to adapt these models by adding a data-driven model component to account for the errors between model predictions and the actual experimental data. Fig. 5 shows the general structure of the hybrid model development.

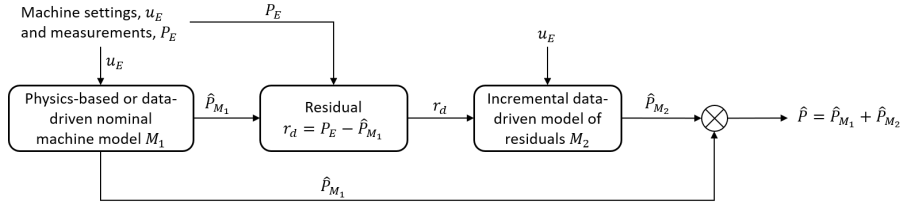


Figure 5: Architecture of the hybrid anomaly detector

The hybrid approach includes three stages: 1) Utilize the physics-based or data-driven model M_1 to predict spindle power $\hat{P}_{M_1}$ and calculate the residual r_d between the experimental data and predictions from model M_1 ; 2) Randomly select 80% of the experimental data (16 tests out of the 20 from Table 4) to build model M_2 of the residuals r_d between M_1 and the healthy experimental data, and test the model predictions on the remaining 20% of the experimental data (remaining 4 tests); 3) Combine the physics-based model (or data-driven model) predictions and the incremental data-driven model predictions to obtain the hybrid model prediction $\hat{P}$.

3.4. Feature Ranking using Dynamic programming (DP)

3.4.1. Dynamic programming (DP)

Many machine learning algorithms can be viewed as performing estimation of the outputs given a set of input features. To avoid over-fitting the model and reduce its complexity, a small set of truly relevant and measurable features is needed. Feature ranking is performed to solve this sensor (feature) optimization problem. In developing regression models, we considered the width of cut, a_e , depth of cut, a_p , feed rate, f , and spindle speed, v_s as the input features available and spindle power P and cutting force F as the output variables.

Dynamic programming (DP) [53][54] is used to select a ranked subset of k -optimal ($k \leq N$) features from a given set of N features. The dynamic programming method splits a global optimization problem into a series of sub-problems in a recursive manner. The functional form of DP enables one to choose a ranked subset of features of specified cardinality with maximum effectiveness (e.g., minimum mean square error, MSE, between nominal data and nominal model predictions or maximum Kullback–Leibler divergence, D_{KL} , between the healthy and faulty condition).

The Kullback–Leibler divergence of (15), D_{KL} (also called relative entropy) is a measure of discrimination between two probability density functions: $p(x)$ (the distribution of residuals of physics-based, hybrid model, under the healthy condition) and $q(x)$ (the distribution of residuals under a faulty condition) on space $\mathcal{X}$ [55]:

$$D_{KL}(p \parallel q) = \sum_{x \in \mathcal{X}} p(x) \log\left(\frac{p(x)}{q(x)}\right) \quad (15)$$

Larger D_{KL} values signify better discrimination between the two distributions and, thus, separation of the evidence of the fault scenarios of interest. Here, we use D_{KL} between the residuals of the healthy data with $W = 0$ mm and faulty data $W = 0.01$ mm as the criterion for ranking features.

Suppose a set of N features $F = (f_1, f_2, \dots, f_N)$ is available, and a subset of k best features is needed among the entire set F . The dynamic programming considers the sub-problems of selecting i -best features, $i = (1, 2, \dots, k)$, and the subset is named as X_i . Let $J_{D_{KL}}(X_i)$ be the performance criterion of subset X_i , which is assumed to be a monotonic function of i . The forward DP recursion for selecting the i -best features based on D_{KL} between the healthy and faulty condition proceeds for $i = 1, 2, \dots, k$ with the initial subset as $X_0 = \emptyset$ proceeds as follows:

$$\begin{aligned} J_{D_{KL}}(X_i^*) &= \max_{X_i \in F} J_{D_{KL}}(X_i) \\ &= \max_{X_{i-1} \in F} \max_{\substack{f \in F \setminus X_{i-1}: \\ X_{i-1} \cup f = X_i}} J_{D_{KL}}(X_{i-1} \cup f) \end{aligned} \quad (16)$$

The optimal feature subsets (X_i^*) according to D_{KL} are given as:

$$\begin{aligned} X_i^* &= \arg \max_{X_i \in F} J_{D_{KL}}(X_i) \\ f_i^* &= X_i^* \setminus X_{i-1}^* \end{aligned} \quad (17)$$

DP can also be implemented as a backward recursion for $i = N - 1, N - 2, \dots, k$ with:

$$\begin{aligned} J_{D_{KL}}(X_i^*) &= \max_{X_i \in F} J_{D_{KL}}(X_i) \\ &= \max_{X_{i+1} \in F} \max_{\substack{f \in X_{i+1}: \\ X_i = X_{i+1} \setminus f}} J_{D_{KL}}(X_{i+1} \setminus f) \end{aligned} \quad (18)$$

470 The task for DP here, is to rank the machining settings (features) in order of significance for anomaly detection using spindle power or cutting force measurements. For this purpose, we chose the Kullback–Leibler divergence between healthy and faulty system output signals.

3.4.2. Backward and Forward Feature Ranking as Approximate Dynamic Programming (ADP)

475 The dynamic programming as backward recursion has exponential computational complexity $O(2^N - 1)$. To reduce the complexity, “knock-out” strategies, including “backward” and “forward” feature selection methods, are used as the approximations of DP. These methods only consider $1 + \sum_{i=k}^{N-1} (i + 1)$ combinations for the backward and $\sum_{i=1}^k (N - i + 1)$ for the forward method.

Forward selection starts with a (usually empty) set of variables and adds variables to it, until the stopping criterion or the required number of the features k is met [56][57]. At stage i ($i = 1, \dots, k$), the forward strategy adds the best feature f_i^* which provides the largest D_{KL} to the feature set of the previous stage. The feature subset of i optimal features, X_i^* , and the the most effective feature, f_i^* , are shown in (19) and (20), respectively:

$$X_i^* = X_{i-1}^* \cup f_i^* \quad (19)$$

$$f_i^*(D_{KL}) = \arg \max_{f_i \in F \setminus X_{i-1}^*} J(X_{i-1}^* \cup f_i) \quad (20)$$

Similarly, the backward method starts with a (usually complete) set of variables and then excludes variables from that set, with the optimal features and the the most effective feature f_i^* as shown below:

$$X_i^* = X_{i+1}^* \setminus f_i^* \quad (21)$$

$$\begin{aligned} f_i^*(D_{KL}) &= \arg \max_{f \in X_{i+1}^*} (J(X_{i+1}^* \setminus f_i) - J(X_{i+1}^*)) \\ &= \arg \max_{f \in X_{i+1}^*} J(X_{i+1}^* \setminus f_i) \end{aligned} \quad (22)$$

The rank or importance of the features using the backward feature selection would be $\{f_1^*, f_2^*, f_3^*, \dots, f_k^*\}$. In total, there are $\frac{(2N-k+1)k}{2}$ subsets to evaluate for the forward method $1 + \frac{(N+k+1)(N-k)}{2}$ subsets for the backward. The backward and forward feature selection strategy can be used as a heuristic in rollout and Monte Carlo Tree Search-based approximate dynamic programming algorithms to provide approximately-optimal feature subsets [58].

3.5. Cumulative Sum (CUSUM)-based Anomaly Detection

This section presents an online sequential anomaly detector, based on the cumulative sum (CUSUM) test proposed by [59]. The CUSUM test is developed to detect small shifts in the power and force signal residuals generated by the machine models (hybrid in our case) in real time. Sequential change-point detection plays a crucial role in detecting whether a process is still working under normal operating conditions (usually termed in-control) or not (out-of-control). Using the hybrid system model, we calculate the difference between the predicted outputs and the actual observations. The key hypothesis here is that any deviations between the hybrid system model and the machine data are due to an anticipated fault. The CUSUM test is one of the control charts that monitors process variable means over time to quickly detect anomalies by calculating the cumulative sum of the sequence predictions and reduces the risk of false alarms [60][61].

Both positive and negative changes between the predicted outputs and the observations are calculated in our case. For a time sequence data $Z(t)$ ($t = 1, 2, 3, \dots, n$), the equations to calculate the high cumulative sum $C^+(t)$ and the low cumulative sum $C^-(t)$ at time t are shown below:

$$\bar{Z}(t) = \frac{1}{WL} \sum_{j=0}^{WL-1} Z(t-j) \quad (23)$$

$$C^+(t) = \max(0, C^+(t-1) + \bar{Z}(t) - \mu_0 - k) \quad (24)$$

$$C^-(t) = \max(0, C^-(t-1) - \bar{Z}(t) + \mu_0 - k) \quad (25)$$

where $\bar{Z}(t)$ is the t^{th} sample mean, WL is the window length, $Z(t)$ is the t^{th} sample measured with a target pre-defined distribution $Z \sim N(\mu_0, \sigma_0)$ under the nominal condition, k is the allowable slack and usually set as $k = \frac{\delta\sigma_0}{2}$, and the tunable parameter δ decides the amount of the shift in process mean that one seeks to detect. The Upper Control Limit (UCL) and the Lower Control Limit (LCL) are included to determine whether an anomaly has occurred or not. When either $C^+(t)$ or $C^-(t)$ exceeds the pre-defined threshold h , the process is said to be out-of-control.

3.6. Performance Evaluation

The performance of residual-based CUSUM detectors is assessed on the synthetic data set with different levels of tool wear in terms of classification accuracy, missed detection rate (FNR) and the false positive rate (FPR). In (26)-(28), TP is the number of true positives defined as positive instances classified correctly, FP is the number of false positives defined as negative instances misclassified as positive, TN is the number of negative instances correctly classified, and FN is the number of positive instances misclassified as negatives [62].

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (26)$$

$$\text{Missed Detection Rate: } FNR = \frac{FN}{TP + FN} \quad (27)$$

$$\text{False Alarm Rate: } FPR = \frac{FP}{FP + TN} \quad (28)$$

4. Results and Discussion

In this section, the residual generation and anomaly detection processes described earlier are applied to 1,000 (operating conditions) $\times$ 11 (tool wear levels) simulation data runs generated by the proposed hybrid model and 20 experimental milling processes run on a healthy tool. Width of cut a_e , depth of cut a_p , feed rate f , and spindle speed v_s are used as the attributes of regression models and spindle power and forces in the x and y directions are measured and used to evaluate the tool wear.

4.1. Hybrid Model generated data

A case study is presented to demonstrate the precision improvement from the hybrid approach of Section 3.3.2 by comparing its accuracy with that obtained by the physics-based model alone. The performance of the physics-based model on a subset of the experimental data (the testing set) is displayed in Fig. 6, including the prediction results and the residual between the actual and predicted values. It is evident that the physics-based model by itself does not fit the transient part of the experimental data well.

After incrementally updating the model, based on residuals from the experimental data and the physics-based model, the hybrid model (physics-based or data-driven model predictions + data-driven model predictions on the discrepancies) yields lower MSE and fits the experimental data much better as shown in Fig. 7. From the performance comparison, it is evident that the proposed hybrid model outperforms the data-driven model alone described in Section 3.2 and the physics-based model alone. The hybrid model achieves a smaller deviation between the predicted value and the observed value, reflected by the residuals and the values of mean square error (MSE). The mean value

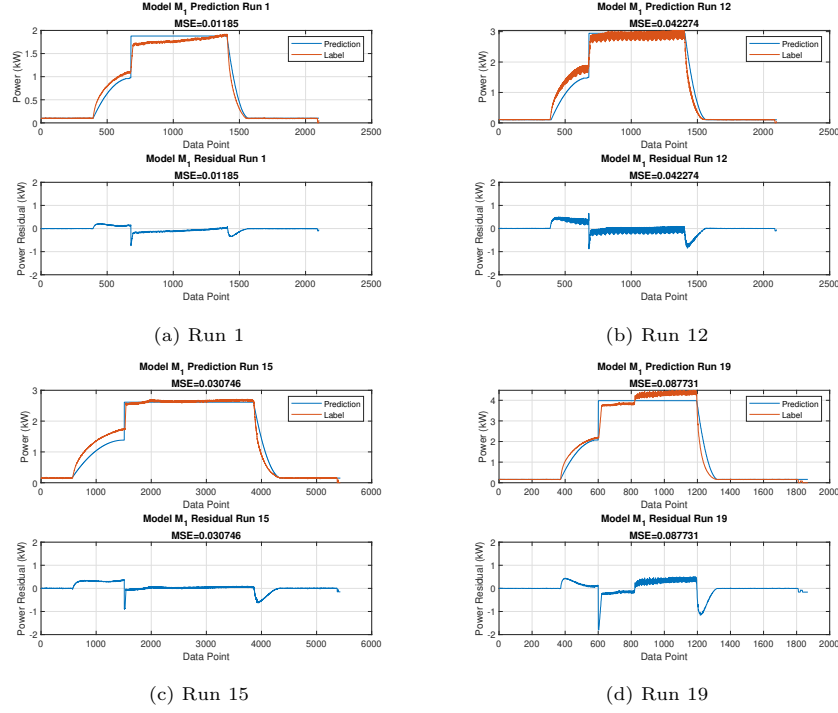


Figure 6: Physics-based model, M_1 , predictions and residuals for the experimental measurements of spindle power of the testing set.

of the MSE, averaged over 1000 simulation runs, decreases from 0.04315 for the physics-based model alone, and 0.0445 for the regression tree model alone, to 0.02405 and 0.0246 for the two hybrid models, respectively.

The performance of the hybrid Model 1 is slightly better than hybrid Model 2. The MSE is reduced by 2.2%, from 0.0246 (for hybrid model 2) to 0.02405 (for hybrid model 1). Based on the results of Table 5, we chose the hybrid model 1 for further implementation in the anomaly detection study.

4.2. Feature Ranking

The feature ranking step is crucial in order to find which features are the most significant and reliable for anomaly detection. In practice, measuring all features may be costly and thus feature selection is needed to select a subset of relevant features for use in regression models. If the input vector is large, training and validation steps could be computationally expensive. However, normally increasing the number of parameters in a regression model reduces the prediction error, e.g., MSE. This trade-off between model complexity and performance is also a problem for models that describe the relationship between the cutting parameters in precision machining and the response variables, such as power and forces. Here, four features; namely, the width of cut, a_e , depth

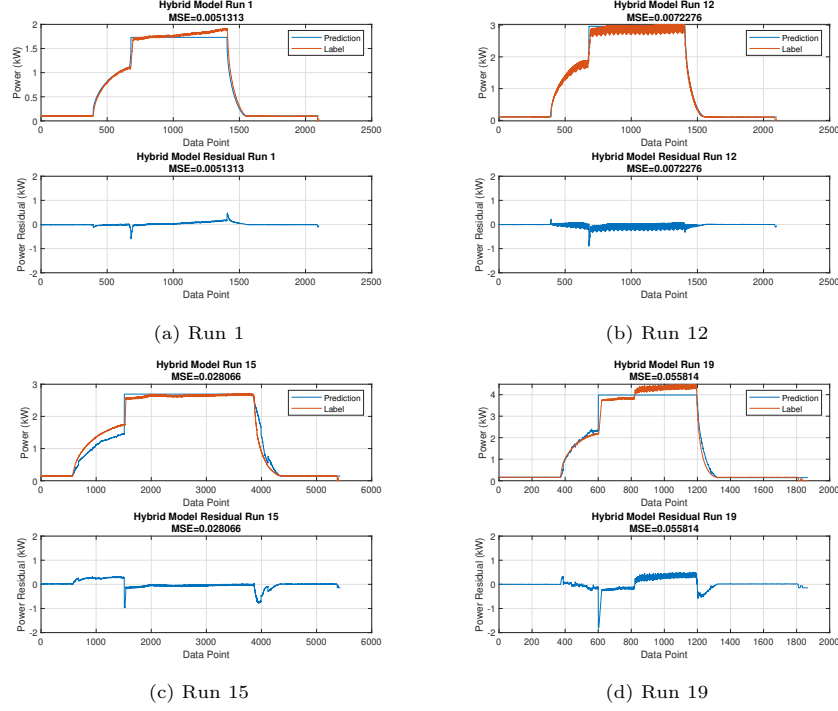


Figure 7: Hybrid model (physics-based + data-driven incremental) predictions and residuals for the experimental measurements of spindle power of the testing set.

of cut, a_p , feed rate, f , and spindle speed, v_s , were evaluated and ranked to find the optimal inputs to the regression models. Dynamic programming with backward and forward feature selection, developed in Section 3.4, were used to select the most salient features. Table 6 shows the results, wherein the “Remaining Features” lists the features selected, the “Deleted Features” lists the features deleted according to the algorithm and the “ D_{KL} using DT” column shows the Kullback–Leibler divergence between the residuals of the healthy data ($W = 0\text{mm}$) and faulty data ($W = 0.03\text{mm}$), using the decision tree model, and “ D_{KL} using NN” column shows the D_{KL} between the residuals of the healthy and faulty data using the neural network. The initial candidate set $\{a_p, a_e, f, v_s\}$ included all four features.

The number of selected features directly affects the D_{KL} , used as a measure of distinguishability of the power signal in response to a healthy and a degraded tool. The best ranked feature subsets of 1 to 4 features are $\{a_p\}$, $\{a_p, a_e\}$, $\{a_p, a_e, f\}$ and $\{a_p, a_e, f, v_s\}$ implemented using both dynamic programming and the “knock-out” strategy. It can be seen that more features give better discrimination in terms of D_{KL} .

Evidently, the model with ordered feature subset $\{a_p, a_e, f, v_s\}$ achieved the best accuracy. The dynamic programming chose all the features in this case,

Table 5: Comparison of the power prediction models in term of MSE

		Run 1	Run 12	Run 15	Run 19	Mean
Physics-based	MSE	0.0119	0.0423	0.0307	0.0877	0.0431
Data-driven	MSE	0.0124	0.0429	0.0320	0.0907	0.0445
Hybrid 1*	MSE	0.0051	0.0073	0.0280	0.0558	0.0240
	% change from Physics-based	-57.14%	-82.74%	-9.39%	-36.37%	-44.33%
	% change from Data-driven	-58.87%	-82.98%	-12.50%	-38.48%	-45.96%
Hybrid 2**	MSE	0.0051	0.0072	0.0306	0.0555	0.0246
	% change from Physics-based	-57.14%	-82.98%	24.10%	-36.72%	-43.06%
	% change from Data-driven	-58.87%	-83.22%	19.06%	-38.81%	-44.72%

*Model fusing physics-based model and data-driven model on residuals.

**Model fusing data-driven model and data-driven model on residuals.

which was expected since all four features are machine settings and directly affect the power consumption and forces. To verify this feature ranking, we performed the CUSUM test using the highest ranked feature sets with 1 to 4 features. The accuracy of detection, per (26) after tuning the CUSUM threshold for each case, is plotted as a function of tool wear severity in Fig. 8. In summary,

Table 6: Dynamic Programming based Feature Ranking using DT at $W = 0.03mm$

Feature Num.	Remaining Features	Deleted Features	D_{KL}	D_{KL} using NN
4	a_p, a_e, f, v_s	-	0.0744	0.0473
3	a_p, a_e, f	v_s	0.0569	0.0470
	a_p, a_e, v_s	f	0.0377	0.0454
	a_p, f, v_s	a_e	0.0065	0.0015
	a_e, f, v_s	a_p	0.0004	0.0005
2	a_p, a_e	f, v_s	0.0328	0.0028
	a_p, f	a_e, v_s	0.0067	0.0015
	a_p, v_s	a_e, f	0.0052	0.0012
	a_e, f	a_p, v_s	0.0004	0.0004
	a_e, v_s	a_p, f	0.0002	0.0002
	f, v_s	a_p, a_e	0.0002	0.0002
1	a_p	a_e, f, v_s	0.0017	0.00065
	a_e	a_p, f, v_s	0.0003	0.00025
	f	a_p, a_e, v_s	0.0002	0.00025
	v_s	a_p, a_e, f	0.0002	0.00024

the proposed methodology aims to balance accuracy with model complexity in an objective manner. Although it costs more in computation time, the model with four features provided both higher Kullback–Leibler divergence and higher accuracy across all levels of noise. Therefore, we concluded that the feature set

$\{a_p, a_e, f, v_s\}$ provides the best anomaly detection performance.

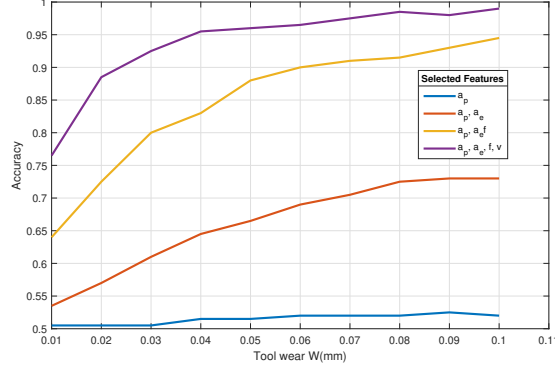
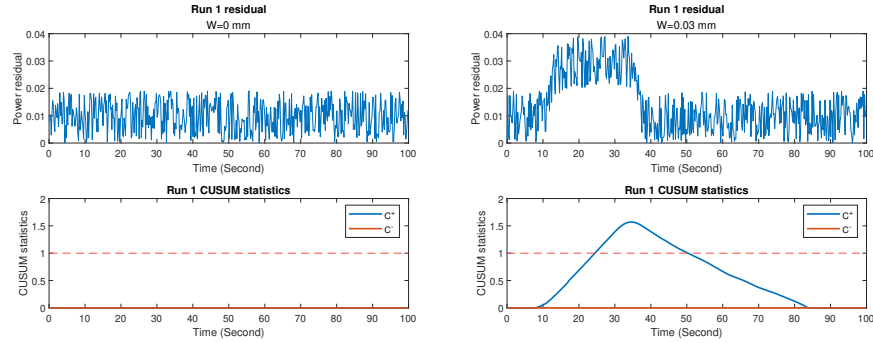


Figure 8: Detection accuracy for different numbers of features at various tool wear levels.

4.3. Anomaly Detection Model

As explained in Section 3.5, only healthy data are used for training, while
590 healthy and faulty ($W = 0.01, 0.02, \dots, 0.1$ mm) data are used for testing to
see if an anomaly has occurred. Residuals and CUSUM detection statistics for
mean changes of Run #1 of Table 4 are in Fig. 9. The high CUSUM value
(upper statistic) detects a positive anomaly (mean of residual increases) and
low CUSUM value (lower statistic) detects a negative anomaly. That is, if the
595 process mean shifts upward, the upper CUSUM test statistic will eventually
drift upwards, and vice versa if the process mean decreases.



(a) Residuals and CUSUM Results for Healthy Data $W = 0$ mm (b) Residuals and CUSUM Results for Faulty Data $W = 0.03$ mm

Figure 9: Run #1 hybrid model residuals and CUSUM statistics: Run #1 $a_p = 1.5875$ mm, $a_e = 31.75$ mm, $v_s = 1299$ RPM, $f = 647.7$ mm/min.

The spindle power was estimated over one milling pass using the hybrid model under a healthy condition ($W = 0$ mm) as described in Section 4.1.

A significantly worn tool affects the power consumption. ΔP , which are the deviations of power measurements from the predictions, were used as indicators for detection. We used a cumulative sum (CUSUM) control chart to detect a change in the mean of a moving window of residuals (a window size of 20 samples) that may indicate an out-of-control process. The CUSUM method was applied on each simulation run and it captures the first out-of-control situation when the thresholds are breached. If there is any point out-of-control in a simulation run, the tool is said to be worn in that simulation run. Thus, for each tool wear level, we had 2,000 runs, 1,000 under a healthy condition and 1,000 under the faulty condition of a specified severity level, to evaluate the anomaly detection performance. The results of this analysis are shown in Fig. 10. Here, the CUSUM δ is defined as two standard deviations from the training P residual mean and the threshold h was selected as 2.5.

In general, the example of Fig. 9 illustrates that combining the hybrid model with the CUSUM test detects anomalies with a high accuracy at a relatively early stage of tool wear. The detection accuracy tends to decrease and missed detection rate increases with larger noise levels or a smaller tool wear levels. When the tool wear (W) is equal or larger than 0.02 mm, CUSUM is able to detect more than 90% of faulty cases and almost all healthy cases are judged to be healthy, yielding greater than 90% accuracy at 0.1% noise level. However, due to the impact of noise and a wide range of operating conditions, runs with smaller tool wear, $W < 0.02$ mm, are hard to detect accurately.

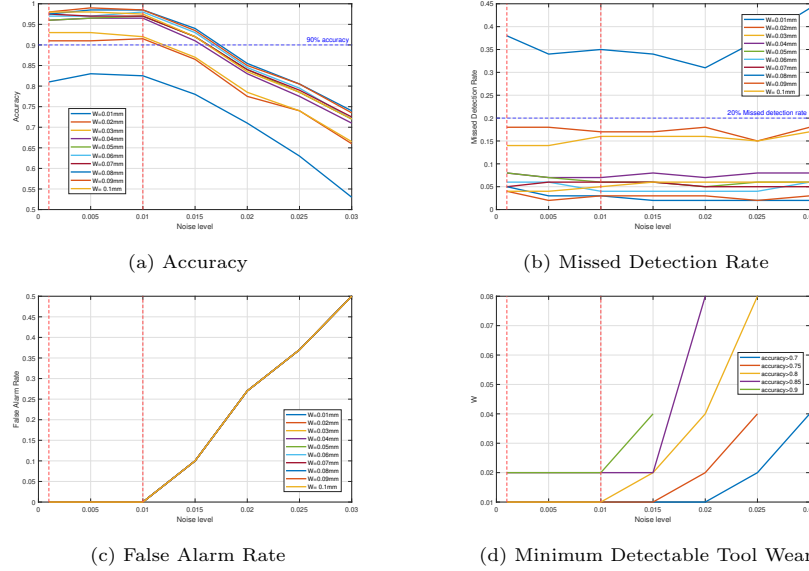


Figure 10: Sensitivity Analysis of CUSUM Anomaly Detector with the hybrid model on simulated data: $\delta = 2$, window size= 20, $h = 1$.

For the quantitative comparison of anomaly detection performance, a sensi-

tivity analysis of CUSUM test with the hybrid model, a physics-based model, DT-CUSUM and NN-CUSUM detectors using accuracy, missed detection rate, false alarm rate and the minimum detectable tool wear was performed and the results are shown in Table 8. Different regression methods were tried and the performance of the anomaly detectors with noise level of 0.01 (1%) and tool wear (W) of 0.03 mm are shown. The approach of combining a physics-based model with an incremental data-driven model to account for discrepancies under nominal conditions and a CUSUM detector are the highest accuracy at a relatively low computational cost among the methods. The approach can be implemented in real-time by monitoring the time series signals, spindle power and cutting forces; it takes 2.2 seconds for a 100 second run or about 4.4 msec per sample with a sampling interval of 200 msec, less than 2.5% of the sampling interval to compute.

Table 8: Comparing the Performance of Anomaly Detection Algorithms using P : Noise level = 0.01 (1%), $W = 0.03$ mm

Method	Accuracy	FNR	FPR	Min. Detectable Tool Wear (Accuracy > 0.8)	Cost over 2000 runs (Seconds)
Hybrid	0.92	0.16	0	0.01	3870
Physics-based	0.89	0.14	0.08	0.01	2956.17
Decision Tree	0.86	0.20	0.08	0.03	2219.72
Neural Network	0.8	0.3	0	0.03	4628.68

The detection performance of the anomaly detector at 10 different levels of tool wear with 1% noise level using $\delta = 2$ in the CUSUM test are compared in Fig. 11. The detection performance reflects different values of the threshold h of CUSUM, discussed in Section 3.5, governing the relationship between risk (false alarm rate) and detection delay. Evidently, for a given false alarm level (which can be controlled by adjusting the threshold h), lower tool wear level has higher detection delay, as it should.

Runs with tool wear less than 0.03 mm have higher risk of false alarms and take longer time to detect anomalies, which might represent a transition from a new to a worn tool state. When the tool wear becomes larger, the power and force signals deviate from the healthy runs significantly and the CUSUM can detect the shift within 25 seconds on average. The methods were also applied to the 20 experimental datasets and CUSUM decisions corresponding to a healthy state were obtained using the hybrid model.

In summary, the proposed approach integrates domain knowledge, synthetic data from a hybrid model, and an incremental data-driven model that accounts for discrepancies between nominal model predictions and the experimental data. The unexplained hidden information in real-world data is accounted for by the incremental data-driven model of the hybrid modeling approach. The generalization and robustness of the system model (physics-based or data-driven) are improved with the use of experimental data. On the other hand, the restriction of the lack of experimental data with tool wear at the initial stages of designing

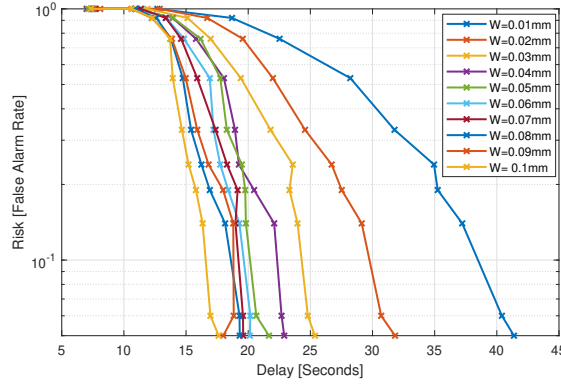


Figure 11: Operation curve - risk (false alarm rate) versus mean delay of detection at 1% noise level for 10 tool wear levels

anomaly detectors is overcome by the use of synthetic data from the hybrid model.

4.4. Experimental Validation

In order to verify the efficacy of the proposed hybrid tool wear detection method, a validation run-to-failure experiment was performed on a 3-axis Haas Mini Mill CNC machine with a maximum cutting power of 7.5 hp and a maximum spindle speed of 6000 RPM. A circular milling cutter with two inserts was used to cut a circular block of AISI 4340 steel. The milling machine recorded signals of machine power, spindle power, cutting force and 3-directional spindle vibrations in real time at sampling frequency of 16 Hz. The experimental database was compiled under various machine settings, that resulted in the 27 data sets of single-pass circular cuts shown in Table 9. Tool wear became significant (defined here as problematic for the stability of machining, i.e., leading to intense vibrations) in the last run (Run #27) with a value of $W = 0.1461$ mm, and was regarded as the failure, which the algorithms presented earlier needed to detect. This data collection system continuously recorded the power consumption at a 0.0625 s sampling interval.

Table 9: Settings of machining experiments from the experiments: $a_p = 2.54mm$, $a_e = 10.16mm$

No.	Cutting Speed v_s (RPM)	Feed rate f (mm/min)
1-16	2330	710
17-27	3184	970.5

A comparison of actual spindle power measurement and physics-based model estimation of the power consumption (for Run #18) is shown in Fig. 12. It was noted that the estimated power consumption of the physics-based model did not match the actual measured data during transients and it also underestimated the

power consumption. This could be due to dynamic effects in the progression of cutting. Also erroneous was potentially the model accuracy in representing the effect of spindle speed, which changed at Run #16, whereas the physics-based model was calibrated with experiments with a new tool only (Run #1). The mismatch was corrected by the hybrid model, which provided a better estimate of spindle power.

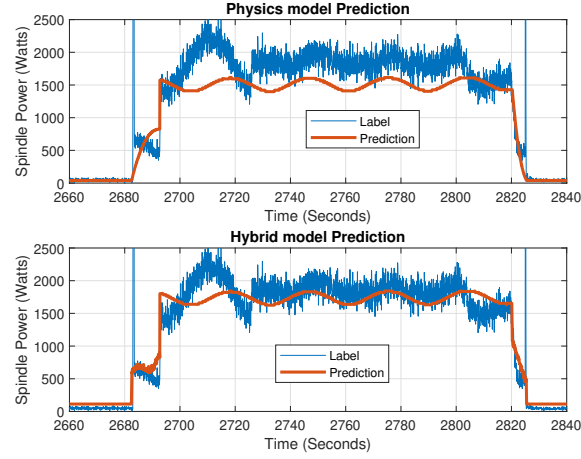
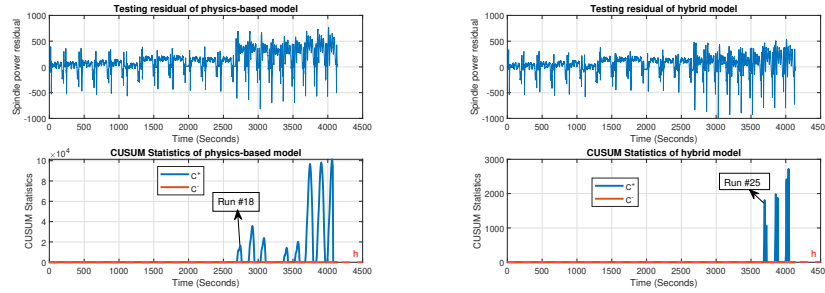


Figure 12: Example of the physics-based model prediction, M_1 , and hybrid model prediction for Run #18 of the tests reported in Table 8

The tool wear detection results obtained by the physics-based and the hybrid models considered in this study are shown in Fig. 13. The CUSUM statistics based on physics-based model mislabeled the change in operating conditions and gave a false alarm at Run #18. However, the proposed hybrid model was able to deal with the change in spindle speed and detected a 0.07 mm tool wear at Run #25, before the cutting tool actually failed at Run #27.



(a) Physics-based model residuals and CUSUM (b) Hybrid model residuals and CUSUM statistics

Figure 13: Tool wear detection for Haas data

5. Conclusions

We proposed an inferential anomaly detector for tool wear monitoring in precision machining using power or force measurements. The anomaly detector uses a model composed of a physics-based model and decision trees or neural networks for incremental error correction. The hybrid model was coupled with a Page’s CUSUM test of the residuals for incipient tool wear detection. Healthy data was used to train the regression models and the training residuals yield the mean and standard deviation values of their distribution under normal conditions, termed the baseline. The residuals from the test sets (healthy and faulty datasets) were used in a CUSUM detector. The framework enabled optimizing the sensors via an optimal feature ranking algorithm and to quantify the minimal detectable tool wear. The goal was to optimize the detector architecture via sensor optimization, minimize the cost of poor quality of parts and avoid machine downtime. With this approach, the prediction is updated using the measurements and eliminates the error in the physics-based model. Data from 20 experiments and 11,000 simulations (1000 simulation runs each for the healthy and ten tool wear conditions) were used to validate the anomaly detection approaches. The following conclusions can be drawn from the results presented earlier: 1) all four features, the width of cut, a_e , depth of cut, a_p , feed rate, f , and spindle speed, v_s , are needed to build an accurate model, which directly affect the power consumption and forces; 2) compared with the results of the standalone physics-based or data-driven model, the MSE in machine power prediction decreases from 0.04315 to 0.02405, and 0.0445 to 0.0246, respectively; 3) the hybrid approach was able to achieve an overall 92% accuracy in data with 1% noise; 4) the proposed approach is also validated on a Haas Mini Mill CNC machine dataset with progressive tool wear, and the approach demonstrated the ability to compensate for errors in physics-based model predictions. In the future, we plan to pursue a number of research avenues, including 1) introducing different types of tool failures and diagnosing the fault types; and 2) couple the anomaly detector with a prognostic algorithm to estimate the residual useful life of the tool.

Acknowledgement

This material is based upon work supported by the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy (EERE) under the Advanced Manufacturing Office Award Number DE-EE0007613. The work of Krishna R. Pattipati was also supported in part by the U.S. Office of Naval Research, in part by the U.S. Naval Research Laboratory under Grant N00014-18-1-1238, N00014-21-1-2187 and Grant N00173-16-1-G905, and in part by the Space Technology Research Institutes from National Aeronautics and Space Administration’s (NASA’s) Space Technology Research Grants Program under Grant 80NSSC19K1076.

Disclaimer: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government

nor any agency thereof, nor any of their employees, makes any warranty, express
or implied, or assumes any legal liability or responsibility for the accuracy,
735 completeness, or usefulness of any information, apparatus, product, or process
disclosed, or represents that its use would not infringe privately owned rights.
Reference herein to any specific commercial product, process, or service by trade
name, trademark, manufacturer, or otherwise does not necessarily constitute
or imply its endorsement, recommendation, or favoring by the United States
740 Government or any agency thereof. The views and opinions of authors expressed
herein do not necessarily state or reflect those of the United States Government
or any agency thereof.

References

- [1] B. Luo, H. Wang, H. Liu, B. Li, F. Peng, Early fault detection of machine tools based on deep learning and dynamic identification, *IEEE Transactions on Industrial Electronics* 66 (1) (2018) 509–518.
- [2] A. M. Khorasani, M. R. S. Yazdi, M. S. Safizadeh, Tool life prediction in face milling machining of 7075 Al by using artificial neural networks (ANN) and Taguchi design of experiment (DOE), *International Journal of Engineering and Technology* 3 (1) (2011) 30.
- [3] D. F. Hesser, B. Markert, Tool wear monitoring of a retrofitted CNC milling machine using artificial neural networks, *Manufacturing Letters* 19 (2019) 1–4.
- [4] M. Nouri, B. K. Fussell, B. L. Ziniti, E. Linder, Real-time tool wear monitoring in milling using a cutting condition independent method, *International Journal of Machine Tools and Manufacture* 89 (2015) 1–13.
- [5] R. Teti, K. Jemielniak, G. O'Donnell, D. Dornfeld, Advanced monitoring of machining operations, *CIRP Annals* 59 (2) (2010) 717–739.
- [6] M. Canizo, E. Onieva, A. Conde, S. Charramendieta, S. Trujillo, Real-time predictive maintenance for wind turbines using big data frameworks, in: *2017 IEEE International Conference on Prognostics and Health Management (ICPHM)*, IEEE, 2017, pp. 70–77.
- [7] C. Drouillet, J. Karandikar, C. Nath, A.-C. Journeaux, M. El Mansori, T. Kurfess, Tool life predictions in milling using spindle power with the neural network technique, *Journal of Manufacturing Processes* 22 (2016) 161–168.
- [8] G. Xu, J. Chen, H. Zhou, J. Yang, P. Hu, W. Dai, Multi-objective feedrate optimization method of end milling using the internal data of the CNC system, *The International Journal of Advanced Manufacturing Technology* 101 (1) (2019) 715–731.
- [9] R. Corne, C. Nath, M. El Mansori, T. Kurfess, Study of spindle power data with neural network for predicting real-time tool wear/breakage during Inconel drilling, *Journal of Manufacturing Systems* 43 (2017) 287–295.
- [10] P. Stavropoulos, A. Papacharalampopoulos, E. Vasiliadis, G. Chryssolouris, Tool wear predictability estimation in milling based on multi-sensorial data, *The International Journal of Advanced Manufacturing Technology* 82 (1–4) (2016) 509–521.
- [11] M. Kious, A. Ouahabi, M. Boudraa, R. Serra, A. Cheknane, Detection process approach of tool wear in high speed milling, *Measurement* 43 (10) (2010) 1439–1446.

- [12] T. Pfeifer, L. Wieggers, Reliable tool wear monitoring by optimized image and illumination control in machine vision, *Measurement* 28 (3) (2000) 209–218.
- 785 [13] T. Pfeifer, J. Elzer, Measuring drill wear with digital image processing, *Measurement* 8 (3) (1990) 132–136.
- [14] W. Wang, G. Hong, Y. Wong, Flank wear measurement by a threshold independent method with sub-pixel accuracy, *International Journal of Machine Tools and Manufacture* 46 (2) (2006) 199–207.
- 790 [15] K. Zhu, X. Yu, The monitoring of micro milling tool wear conditions by wear area estimation, *Mechanical Systems and Signal Processing* 93 (2017) 80–91.
- 795 [16] M. T. García-Ordás, E. Alegre, V. González-Castro, R. Alaiz-Rodríguez, A computer vision approach to analyze and classify tool wear level in milling processes using shape descriptors and machine learning techniques, *The International Journal of Advanced Manufacturing Technology* 90 (5-8) (2017) 1947–1961.
- [17] Z. Liu, Y. Guo, M. Sealy, Z. Liu, Energy consumption and process sustainability of hard milling with tool wear progression, *Journal of Materials Processing Technology* 229 (2016) 305–312.
- 800 [18] A. Bustillo, D. Y. Pimenov, M. Mia, W. Kapłonek, Machine-learning for automatic prediction of flatness deviation considering the wear of the face mill teeth, *Journal of Intelligent Manufacturing* 32 (3) (2021) 895–912.
- 805 [19] Y. Zhou, B. Sun, W. Sun, Z. Lei, Tool wear condition monitoring based on a two-layer angle kernel extreme learning machine using sound sensor for milling process, *Journal of Intelligent Manufacturing* (2020) 1–12.
- [20] N. Ghosh, Y. Ravi, A. Patra, S. Mukhopadhyay, S. Paul, A. Mohanty, A. Chattopadhyay, Estimation of tool wear during cnc milling using neural network-based sensor fusion, *Mechanical Systems and Signal Processing* 21 (1) (2007) 466–479.
- 810 [21] S. Han, N. Mannan, D. C. Stein, K. R. Pattipati, G. M. Bollas, Classification and regression models of audio and vibration signals for machine state monitoring in precision machining systems, *Journal of Manufacturing Systems* 61 (2021) 45–53.
- 815 [22] P. Palanisamy, I. Rajendran, S. Shanmugasundaram, Prediction of tool wear using regression and ann models in end-milling operation, *The International Journal of Advanced Manufacturing Technology* 37 (1-2) (2008) 29–41.

- [23] Z. Huang, J. Zhu, J. Lei, X. Li, F. Tian, Tool wear predicting based on multi-domain feature fusion by deep convolutional neural network in milling operations, *Journal of Intelligent Manufacturing* (2019) 1–14.
- [24] W. Cai, W. Zhang, X. Hu, Y. Liu, A hybrid information model based on long short-term memory network for tool condition monitoring, *Journal of Intelligent Manufacturing* 31 (6) (2020) 1497–1510.
- [25] M. S. Kan, A. C. Tan, J. Mathew, A review on prognostic techniques for non-stationary and non-linear rotating systems, *Mechanical Systems and Signal Processing* 62 (2015) 1–20.
- [26] G. Quintana, J. Ciurana, Chatter in machining processes: A review, *International Journal of Machine Tools and Manufacture* 51 (5) (2011) 363–376.
- [27] A. Bustillo, R. Reis, A. R. Machado, D. Y. Pimenov, Improving the accuracy of machine-learning models with data from machine test repetitions, *Journal of Intelligent Manufacturing* (2020) 1–19.
- [28] D. Y. Pimenov, A. Bustillo, T. Mikolajczyk, Artificial intelligence for automatic prediction of required surface roughness by monitoring wear on face mill teeth, *Journal of Intelligent Manufacturing* 29 (5) (2018) 1045–1061.
- [29] C. Liu, Y. Li, J. Li, J. Hua, A meta-invariant feature space method for accurate tool wear prediction under cross conditions, *IEEE Transactions on Industrial Informatics* 18 (2) (2021) 922–931.
- [30] M. Marei, S. El Zaatari, W. Li, Transfer learning enabled convolutional neural networks for estimating health state of cutting tools, *Robotics and Computer-Integrated Manufacturing* 71 (2021) 102145.
- [31] Y. Li, C. Liu, J. Hua, J. Gao, P. Maropoulos, A novel method for accurately monitoring and predicting tool wear under varying cutting conditions based on meta-learning, *CIRP Annals* 68 (1) (2019) 487–490.
- [32] H. Hanachi, W. Yu, I. Y. Kim, J. Liu, C. K. Mechefske, Hybrid data-driven physics-based model fusion framework for tool wear prediction, *The International Journal of Advanced Manufacturing Technology* 101 (9) (2019) 2861–2872.
- [33] J. Wang, Y. Li, R. Zhao, R. X. Gao, Physics guided neural network for machining tool wear prediction, *Journal of Manufacturing Systems* 57 (2020) 298–310.
- [34] U. Awasthi, N. Mannan, Z. Wang, K. Pattipati, G. M. Bollas, Physics-based and information-theoretic sensor and settings selection for tool wear detection in precision machining, In review (2021).
- [35] T. L. Schmitz, K. S. Smith, *Machining Dynamics: Frequency Response to Improved Productivity*, Springer US, 2008.
URL <https://books.google.com/books?id=Z5s0mpX2KxQC>

- [36] S. C. Lin, R. J. Lin, Tool wear monitoring in face milling using force signals, *Wear* 198 (1) (1996) 136–142. doi:[https://doi.org/10.1016/0043-1648\(96\)06944-X](https://doi.org/10.1016/0043-1648(96)06944-X).
- 860 [37] D. J. Waldorf, S. G. Kapoor, R. E. DeVor, Automatic recognition of tool wear on a face mill using a mechanistic modeling approach, *Wear* 157 (2) (1992) 305–323. doi:10.1016/0043-1648(92)90069-K.
- [38] T. Chang, F. Gan, A cumulative sum control chart for monitoring process variance, *Journal of Quality Technology* 27 (2) (1995) 109–119.
- 865 [39] P. Castagliola, A. Artiba, P. Castagliola, G. Celano, S. Fichera, A new cusum-s2 control chart for monitoring the process variance, *Journal of Quality in Maintenance Engineering* (2009).
- [40] W. L. Martinez, A. R. Martinez, Computational statistics handbook with MATLAB, Vol. 22, CRC press, 2015.
- 870 [41] T. Nguyen, T. H. Chan, D. P. Thambiratnam, Controlled monte carlo data generation for statistical damage identification employing mahalanobis squared distance, *Structural Health Monitoring* 13 (4) (2014) 461–472.
- [42] G. K. Tso, K. K. Yau, Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks, *Energy* 32 (9) (2007) 1761–1768.
- 875 [43] L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone, *Classification and Regression Trees*, Boca Raton, FL: Chapman & Hall, 1984.
- [44] J. R. Quinlan, Induction of decision trees, *Machine learning* 1 (1) (1986) 81–106.
- 880 [45] L. Breiman, Random forests, *Machine learning* 45 (1) (2001) 5–32.
- [46] O. Sagi, L. Rokach, Ensemble learning: A survey, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8 (4) (2018) e1249.
- [47] L. Torgo, Functional models for regression tree leaves, in: *ICML*, Vol. 97, Citeseer, 1997, pp. 385–393.
- 885 [48] W. S. Sarle, *Neural networks and statistical models* (1994).
- [49] D. Karayel, Prediction and control of surface roughness in cnc lathe using artificial neural network, *Journal of materials processing technology* 209 (7) (2009) 3125–3137.
- 890 [50] R. Battiti, Accelerated backpropagation learning: Two optimization methods, *Complex systems* 3 (4) (1989) 331–342.
- [51] M. D. Zeiler, Adadelta: an adaptive learning rate method, arXiv preprint arXiv:1212.5701 (2012).

- [52] C. M. Bishop, et al., Neural networks for pattern recognition, Oxford university press, 1995.
- 895 [53] N. Acir, Classification of ecg beats by using a fast least square support vector machines with a dynamic programming feature selection algorithm, Neural computing & applications 14 (4) (2005) 299–309.
- [54] M. T. Emmerich, A. H. Deutz, A tutorial on multiobjective optimization: fundamentals and evolutionary methods, Natural computing 17 (3) (2018) 585–609.
- 900 [55] Z. Zhen, X. Zeng, H. Wang, L. Han, A global evaluation criterion for feature selection in text categorization using kullback-leibler divergence, in: 2011 International Conference of Soft Computing and Pattern Recognition (SoCPaR), IEEE, 2011, pp. 440–445.
- 905 [56] S. Abe, Modified backward feature selection by cross validation., in: ESANN, Citeseer, 2005, pp. 163–168.
- [57] G. Borboudakis, I. Tsamardinos, Forward-backward selection with early dropping, The Journal of Machine Learning Research 20 (1) (2019) 276–314.
- 910 [58] D. P. Bertsekas, Reinforcement learning and optimal control, Athena Scientific, Belmont, MA, 2019.
- [59] E. S. Page, Continuous inspection schemes, Biometrika 41 (1/2) (1954) 100–115.
- 915 [60] J. Goh, S. Adepu, M. Tan, Z. S. Lee, Anomaly detection in cyber physical systems using recurrent neural networks, in: 2017 IEEE 18th International Symposium on High Assurance Systems Engineering (HASE), IEEE, 2017, pp. 140–145.
- [61] F. Kadri, F. Harrou, S. Chaabane, Y. Sun, C. Tahon, Seasonal arma-based spc charts for anomaly detection: Application to emergency department systems, Neurocomputing 173 (2016) 2102–2114.
- 920 [62] D. Chicco, G. Jurman, The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation, BMC genomics 21 (1) (2020) 6.