IOP Conference Series: Materials Science and Engineering

PAPER • OPEN ACCESS

Machine Learning for Tool Wear Classification in Milling Based on Force and Current Sensors

To cite this article: M Schwenzer *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **520** 012009

IOP Conf. Series: Materials Science and Engineering **520** (2019) 012009 doi:10.1088/1757-899X/520/1/012009

Machine Learning for Tool Wear Classification in Milling Based on Force and Current Sensors

M Schwenzer[1,*], K Miura[1] and T Bergs[1]

[1] Laboratory for Machine Tools and Production Engineering (WZL) of the RWTH Aachen University, Campus-Boulevard 30, 52074 Aachen, Germany

*M.Schwenzer@wzl.rwth-aachen.de; +49-241-80-28021

Abstract. This paper presents how to classify the wear state of an end-mill based on force and current signals of the linear axes. The data is divided in binary classes based on the maximum flank wear. A support vector machine and a random forest are trained on orthogonal cutting experiments, but the validation is performed on arbitrary tool paths. To achieve this unique level of generality the signals are transformed into the rotation tool coordinate system. The features are extracted over five cutter revolutions. Support vector machines outperform random forests achieving 99,8% and 97% accuracy in the two classes on the test data and 98% and 61% accuracy on the validation data.

1. Introduction

Milling is a very flexible and economic production process, in particular suited for processing metals. Tool wear is as unavoidable as it is undesirable and influences manufacturing costs, workpiece quality and process safety (tool breakage). Nowadays, milling tools are replaced either after a fixed service life or based on the experience of the machine operator. Seldom, monitoring systems are used that trigger alarms if a fixed threshold is exceeded. This is only applicable for highly repetitive manufacturing tasks and the optimal choice of the threshold is a time-consuming task.

Only recently, data-driven models have turned into a hot topic in engineering with the rise of machine learning techniques, which have vastly spread throughout the academic community. They can learn abstract thresholds automatically and; thereby, enable for an automatic tool wear monitoring. However, most proposed classification models lack of generality.

This work trains models on data from orthogonal cutting but validates them on data from a complex tool path in one plane. To our known extend, this is the first work that dares to apply its model to structurally different data.

Though many works consider acoustic emission (AE) signals to classify tool wear [1–4], Kaya et al. [2] found that they only make a small contribution to an accurate classification of tool wear. In contrast to that, acceleration signals show a high correlation to all wear classes, whereas force is of particular interest to identify sharp tools. Kaya et al [2] use a genetic algorithm to optimize the selection of the features, which serve as input to a support vector machine. The selection could be made between 18 statistical values in time-domain of eight signals (force, acceleration, both in all Cartesian axes; torque, and AE). Analyzing the choice of the global optimization algorithm, they derived which signals have a high correlation to tool wear.

Nevertheless, the dominating input to models that classify tool wear is force, since its variation correlates to the wear of the tool [5]. Most approaches with machine learning techniques rely on force signals [3,4,6,7]. However, if the tool life duration is considered as input feature, it dominates the approximation [1]. This changes drastically if more than one tool is considered. Considering the tool life makes the approach prone to uncertainties in the production of the tool, the process, and the workpiece material.

The recent focus of learning tool wear lay on classification. Few studies approximate a tool wear curve by using a SVM for regression (SVR) [1,6]. However, a classification of the tool wear state was achieved afterwards via a customized threshold [6]. The current research focuses on support vector machines (SVMs) [1–3,8] and decision trees mostly random forest (RF) [3,4,9] as algorithms to classify the tool wear state.

For tool wear classification, popular features (signal describing characteristics) are the maximum, the mean, the median, and the standard deviation [2–4] of a signal. The choice of the features is key to the performance of every machine learning algorithm. Kaya et al. [2] extract less common features, such as the skewness or the kurtosis, and use a genetic algorithm to select the most appropriate features. On this background, the approach of Martínez-Arellano et al. [7] is exceptional since it does not extract features explicitly but uses deep learning on the raw signals. They transform the time-series signal to an image and apply a convolutional neural network (cNN) to classify the tool wear state. The approach considers the time-dependency of the force signal, making it unique for every cutting operation. This results in a lack of transferability to a new operation. However, the low level of generality is the major drawback of all works in this field. All studies in literature perform straight cuts for both, training and testing. None of the models was validated on structurally different data.

The described works present an active field with a strong focus on feasibility. No consolidation has yet taken place with regard to a method or certain features. Furthermore, most works lack of generality since training and testing data were taken from the same experiments. This work tackles the latter deficit and gathering the most promising features and algorithms from literature.

2. Approach

This work uses two of the most popular algorithms for classification from the field of machine learning [10]: Support vector machines (SVMs) and random forest (RF), which are basically decision graphs. The algorithms are particularly suited for problems with biased numbers of examples, i.e. many examples of one class (e.g. "good tool") and only a few examples for a different class (e.g. "dull tool").

2.1. Support vector machine

Originally designed for classification [11,12], SVMs have turned into a powerful tool to approximate arbitrary functions with a comparably small number of model parameters. This makes them less prone to the problem of "overfitting", a phenomenon that all supervised learning methods suffer from. It is best described by memorizing the training data revealing a lack of generalization. Models with a large number of parameters likely learn a training set without error but exhibit poor generalization, which makes the SVM of particular interest [11] in contrast to e.g. artificial neural networks (NNs) [13]. Today, it is among the most popular algorithms in machine learning because of its robustness and generalization.

SVMs determine the optimal hyperplane

$$\boldsymbol{w} \times \boldsymbol{x}_i + b = 0 \quad \text{with} \quad \boldsymbol{w} \in \mathbb{R}^n, b \in \mathbb{R} \tag{1}$$

that separates two groups of n data points $\boldsymbol{x}_i$ (were × denotes the dot product in the feature space). Assuming a set of data points with corresponding labels of

$$\{(\boldsymbol{x}_i, y_i) \mid i = 1, \dots, n; \boldsymbol{x}_i \in \mathbb{R}^n; y_i \in \{-1, +1\}\}, \tag{2}$$

an optimal hyperplane has to satisfy the condition

$$y_i = \text{sgn}(\boldsymbol{w} \times \boldsymbol{x}_i + b). \tag{3}$$

The optimal hyperplane maximizes the minimum distance (often called "margin") of all data points to the hyperplane. This corresponds to minimizing the Euclidian norm of the normal vector w of the

IOP Conf. Series: Materials Science and Engineering **520** (2019) 012009 doi:10.1088/1757-899X/520/1/012009

hyperplane. Those points, which lie on the margin – and thereby define the hyperplane – are the support vectors.

Boser et al. [11] point out that though a maximum margin optimization minimizes the probability of misclassification, it is prone to “atypical” data near the decision boundary. They argue that those data points have a very high influence on the decision boundary (through their high Lagrangian multiplier). This is counteracted by introducing so-called “slack variables”; variables that penalizes the extend of misclassification (i.e. the distance to the decision boundary) in the optimization function. In this way, a certain misclassification is allowed if it contributes to the overall accuracy and; thereby, allows for a more generic decision boundary.

2.2. Random forest

The RF consists out of multiple decision trees, of which everyone is build on a randomly selected subset of the feature space [14]. The RF obtains a decision by following the majority vote of its decision trees.

The algorithm does not use any pruning techniques on the decision trees. [14] argues that by pruning back decision trees they lose their accuracy on the training data, which gives little confidence of performing good on unseen data. The hope for generalization is bought at the cost of a lower accuracy on the training data. Instead, overfitting of a single classifier can be compensated by using multiple classifiers. This work builds a RF with 400 binary decision trees.

2.3. Experimental set-up

The experiments were conducted on a Mazak i600 VARIAXIS machining center with a Siemens SINUMERIK 840D sl controller. The tool was a two-fluted solid carbide end mill (diameter D = 10 mm, helix angle β = 45 °) and the material a high-alloyed steel (X5CrNi-18-10). In order to provoke tool wear, the process parameters were chosen somewhat more aggressive than suggested by the tool manufacturer: cutting speed v_c = 140 m/min, feed per tooth f_z = 0.1 mm. This results in a tool life travel path of 88 m, which consists mainly from straight cuts (550 times) and two simple impeller geometries with ten blades each, Figure 1. The idealized impeller geometry was chosen in order to represent a typical freeform surface geometry for milling and to emphasize the industrial applicability of this approach. It evaluates the transferability to a 2D cutting process.

The first impeller was manufactured after 350 cuts with a still sharp tool and the second impeller at the end with a dull tool. The tool state was checked at regular intervals on a microscope (Keyence VHX 5000). Those measures classify the tool whether it is sharp or dull. In this way the last 50 straight cuts and the last ten blade cuts were classified as “dull”.

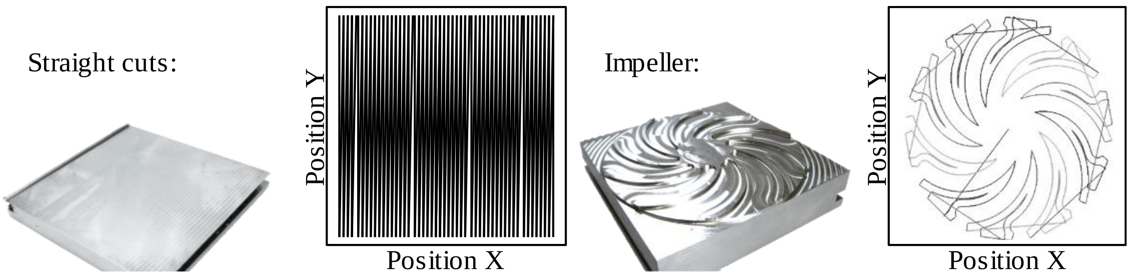

Figure 1. Overview of the two workpiece geometries: straight cuts and impeller.

The three-phase alternating current and voltage signals were measured at fs = 10,000 Hz with external sensors for the linear axes. To calculate the corresponding DC-components, the angular position of the motor was observed applying the method of [15]. In addition, the force was measured with a piezo-electric dynamometer (Kistler 9255) at fs = 10,000 Hz. Internal signals from the numeric controller (NC) of the machine were acquired via the compile cycle “axis data stream” (ADAS). This

IOP Conf. Series: Materials Science and Engineering **520** (2019) 012009 doi:10.1088/1757-899X/520/1/012009

enables us to record the position, velocity, acceleration, and current information of every axis. The compile cycle works at fs = 400 Hz.

2.4. Preprocessing

2.4.1. Transformation To generalize from orthogonal cutting data to arbitrary conditions, the signals are transformed to the coordinate systems of the cutting edges, i.e. being split into their tangential, radial, and passive components. For this it is necessary to know the feed direction angle α and the angle of every cutting edge, Figure 2. Since the positions of the cutting edges do not vary to each other – unless a severe failure of the cutting edge occurs – its position angle composes of the rotation angle $\varphi(t)$ and the starting position angle $\Phi_{0,z}$.

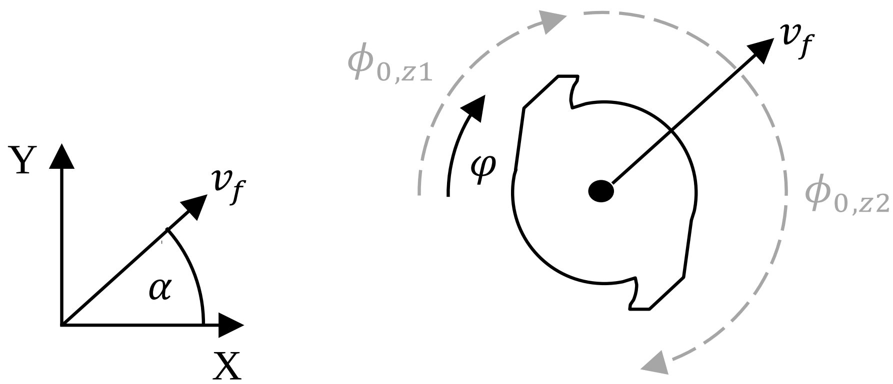

Figure 2. Schematic overview of the geometric values for the transformation from machine to tool coordinate system.

In general, both variables are unknown. While the starting position of the cutting edges must be identified through precise and laborious measurements, the angular position of the spindle is a vital information for the cascaded controllers of the spindle itself. However, modern machining centers work on encrypted BUS systems denying access to this information. Only external encoders with an open interface can provide the exact information. A practical approach to overcome both issues is to reconstruct the angular position of the cutting edges. We performed a cross correlation every ten cutter revolutions, restricting the allowed lag to $360/N_z$ of a revolution, where N_z is the number of teeth. This ensures to maintain the correct mapping between signal fraction and cutting edge.

Since the sampling rate is not a multiple of the rotational speed, the rotation angle $\varphi(t)$ varies and with it the uncut chip thickness h. This ensures that we can generalize from constant width of cut ae to other width of cuts via the uncut chip thickness.

2.4.2. Features Force is among the most popular signals to use to classify tool wear [1–4,6,7]. Furthermore, acceleration shows a high correlation to tool wear [1,2,4,6]. Cho et al. [6] showed that an additional electrical information (power) improves the classification accuracy by 20 %. García-Nieto et al. [1] demonstrate the correlation of the current to the tool wear state. Though they focused on the spindle signals, we chose to use the axis current as additional signals of all linear drives. The NC provides additionally the signals of the spindle.

The most common mechanistic force models in cutting technology – such as the exponential Kienzle model [16] or the linear Altintas model [17] – rely on the uncut chip thickness h and the uncut chip width b. This information is of particular importance to generate a generic classification since it scales the cutting force and all dependent signals. The uncut chip thickness h and the uncut chip width can be calculated through a simulation of the engagement conditions.

Table 1 provides the available signals of the experimental set-up.

Table 1. Overview of the available signals and calculated additional data.

External signals ($f_s = 10\ kHz$)	Force (F_t, F_r, F_p), current ($I_{q,t}$, $I_{q,r}$, $I_{q,p}$), uncut chip thickness h, uncut chip width b
Internal signals ($f_s = 400\ Hz$)	Torque (T_r, T_t, T_p, T_S), current ($I_{q,t}$, $I_{q,r}$, $I_{q,p}$, $I_{q,S}$), acceleration (a_t, a_r, a_p), uncut chip thickness h, uncut chip thickness b

The mean, the maximum, and the standard deviation were extracted over five cutter revolutions building the features space. Most of the literature uses those statistical values from time domain as features (see [2–4]). The accuracy of the two machine learning algorithms is tested on three different feature sets consisting out of the same data points:

- features extracted from internal signals,
- features extracted from external signals, and
- merging the previous data sets.

2.4.3. Data division The major drawback of all classification models proposed in literature is its lack of generality. Trained and tested on the same cutting geometry (often simply straight cuts), the model cannot be transferred to other workpiece geometries. To prove the ability to generalize of our approach, we train the model on straight cuts but validate it on the data from the idealized impeller, Figure 1, representing structurally different data. We use 70 % of the data from the straight cuts to train the classification model; the remaining 30 % is used for testing. Table 2 provides the absolute number of data points.

Table 2. Overview of the different data sets.

	Training	**Testing**	**Validation**
Total number of data points	65553 (Straight cuts)	28094 (Straight cuts)	6023 (Impeller)
Number of "sharp" data points – class A	59667	25572	3012
Number of "dull" data points – class B	5886	2522	3011

3. Results and discussion

3.1. Results

To evaluate the classification accuracy of the two machine learning algorithms on the different feature sets, Table 3 and Table 4 provide the confusion matrices of the test data and the validation data respectively. The row of a confusion matrix represents the estimated classes (indicated by the accent "^") and the columns represent the actual classes (class A: "sharp", class B: "dull"). A high value on the diagonal axis indicates a good performance. For the sake of convenience, we colored the cells of the matrix in traffic light colors. The secondary diagonal contains the number of wrong assignments.

Table 3. Results of the test data (straight cuts) for different feature sets (only signals from internal or external sensors and joint performance) and both algorithms: support vector machine (SVM) and random forest (RF). Class A: “sharp tool”, class B: “dull tool”.

Algorithm \ Feature set		Intern *A*	Intern *B*	Extern *A*	Extern *B*	Intern + Extern *A*	Intern + Extern *B*
SVM	$\hat{A}$	100 %	100 %	99,8 %	2,7 %	99,9 %	1,3 %
	$\hat{B}$	0 %	0 %	0,2 %	97,3 %	0,07 %	98,7 %
RF	$\hat{A}$	98 %	21 %	99,9 %	2,5 %	99,9 %	1,2 %
	$\hat{B}$	2 %	79 %	0,1 %	97,5 %	0,03 %	98,8 %

Table 4. Results of the validation data (idealized impeller geometry) for different feature sets (only signals from internal or external sensors and joint performance) and both algorithms: support vector machine (SVM) and random forest (RF). Class A: “sharp tool”, class B: “dull tool”.

Algorithm \ Feature set		Intern *A*	Intern *B*	Extern *A*	Extern *B*	Intern + Extern *A*	Intern + Extern *B*
SVM	$\hat{A}$	100 %	100 %	98 %	39 %	100 %	42 %
	$\hat{B}$	0 %	0 %	2 %	61 %	0 %	58 %
RF	$\hat{A}$	98,2 %	99 %	87 %	69 %	99 %	99 %
	$\hat{B}$	1,8 %	1 %	13 %	31 %	1 %	1 %

3.1.1 Feature sets Regarding the feature sets, the internal signals alone are not suitable for tool wear classification. The random forest is able to achieve a solid classification accuracy on the test data, which has the same structure as the data on which the algorithm is trained. However, it is unable to generalize from it (only 1 % accuracy on class B, “dull”, Table 4). Using the features extracted from external signals results in an outstanding accuracy on the test data and even a solid accuracy on the validation data. This is in good accordance with the literature – which only tests on data that has the same structure as the training data. The accuracy and the sensitivity of the external sensors (dynamometer, current, and voltage sensors) and the sampling frequency is much higher than the internal signals provided by the machine controller. This drastically affects the accuracy of the classification algorithms.

At the first glance, it is surprising that merging the features of the internal and the external signals does not lead the same or a higher classification accuracy as the features of the external signals alone. A possible reason for this is that the features extracted from the internal signals do not contain suitable information (bad accuracy if they are used alone) and increase the dimension of the feature space to

$19 \cdot 3 = 57$ dimensions. The problem of a reduced significance of the individual feature in a high-dimensional feature space is known as “curse of dimensionality”. The feature space of the external signals has $6 \cdot 3 = 18$ dimensions; the feature space of the internal signals $13 \cdot 3 = 39$ dimensions.

3.1.2. Classification algorithms Both, the SVM and the RF, achieve a very high accuracy on the test data. Therefore, for evaluation purposes, we focus on the validation data, Table 4. The way in which randomization is introduced in multiple decision trees – by using a random subset of the feature space – makes RF particularly prone to the curse of dimensionality. If the majority of the features does not contribute to the classification objective, many trees are grown that have a low classification accuracy. Since RF relies on the majority vote of these trees, this severely affects the classification accuracy of the forest. Using more classes eases this problem as the wrong classifications may distribute among the more classes and a lower number of correct trees is required for a majority vote. Solving an optimization problem, the SVM is less prone to the curse of dimensionality.

3.2. Limitations and discussion

Examining the points that are classified incorrectly more closer, Figure 3, reveals that both algorithms manage to classify sharp tools (class A) correctly: average accuracy 98.1 % (SVM) and 86.1 % (RF). Figure 3 depicts the accuracy rate in groups of 120 samples of the validation data. Note that the first 3012 samples represent a sharp tool, whereas the remaining 3011 samples represent a dull tool.

Both algorithms experience a severe drop in accuracy when the tool wear class changes to “dull” (class B). However, as the tool wear increases the SVM is able to identify the dull tool correctly: average accuracy of the last six blades 99.2 % (SVM) and 18.9 % (RF).

The arbitrary definition of a “sharp” and a “dull” tool may cause the problem of an ambiguous separation plane. Tool wear evolves continuously, which makes it difficult to define the exact point from where a tool is considered as “dull”. The sensor signals reflect this problem as they change gradually. Only if the tool suffers from breakouts, one may obtain a distinct separation plane.

The presented system itself suffers from the inaccuracy introduced by the missing rotational position of the spindle (which was reconstructed through cross correlation) and of the auxiliary drives (which was estimated through a Kalman filter). To exploit the full potential of this approach, the encoder signal of all motors must be known. In future research, the performance over multiple tools must be examined; as well as the sensitivity to more classes.

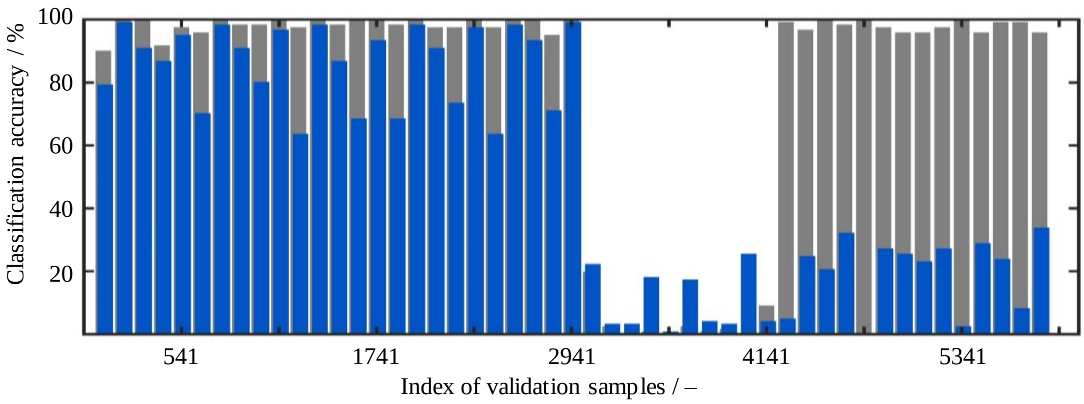

Figure 3. Evolution of the classification accuracy in the validation data. Blue bars: SVM, gray bars: RF

3.3. Notes to a practical application

Within the rising digitalization and activities such as “Industry 4.0” or an “Internet of Production”, the production process itself becomes transparent. Machine tool manufacturers open their interfaces and include more sensors into their products. This work stresses the significance of signal quality but is based on standard measurement equipment. Machine learning in particular provides a method to automatically analyze big data and robustly monitor e.g. tool wear. Considering the engagement conditions provided by some CAM programs and transforming all signals to the coordinate system of the tool enables to learn a model that is independent of the tool path and thus is even suited for single-batch-production. All preprocessing and training can be conducted onsite on an edge computer – the resulting machine learning model may even be transferred to other machine tools.

4. Conclusion

This paper presents the first approach on tool wear classification that is validated on structural different data. This previously unreached level of generality is achieved by transforming the signals from the fixed machine coordinate system to the rotating coordinate system of the cutting edges. It is the first to use classical process characteristics such as the uncut chip thickness h and the uncut chip width b. Features are extracted from signal segments of five cutter revolutions, contributing to the objective of creating a model as generic as possible. Furthermore, this results in more data points than any study in literature.

Two popular classification algorithms were examined: SVM and RF. Both achieve a very high accuracy on the test data (over 97 % for both classes). But SVM is able to obtain a more generic model from the straight cutting tests: 98 % for class A ("sharp") and 61 % for class B ("dull") on the blade geometry.

Examining features from internal sensors (torque and current at fs = 400 Hz) and external sensors (force and current at fs = 10 kHz) indicates that

- the higher quality and sampling rate of external sensors lead to a much more accurate classification; and that
- adding the features extracted from the internal sensor worsens the performance ("curse of dimensionality").

With the raise of the digitalization at shop floor level more data and of higher quality will become available. This work outlines an automatic analysis and lies the basis to profit from data.

5. Acknowledgments

The authors would like to thank the German Research Foundation DFG for the kind support within the Cluster of Excellence "Internet of Production" (Project ID: 390621612).

References

[1] García-Nieto P J, García-Gonzalo E, Vilán Vilán J A and Segade Robleda A 2016 A new predictive model based on the PSO-optimized support vector machine approach for predicting the milling tool wear from milling runs experimental data *The International Journal of Advanced Manufacturing Technology* **86** 769–80
[2] Kaya B, Oysu C, Ertunc H M and Ocak H 2012 A support vector machine-based online tool condition monitoring for milling using sensor fusion and a genetic algorithm *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture* **226** 1808–18
[3] Wu D, Jennings C, Terpenny J, Gao R X and Kumara S 2017 A Comparative Study on Machine Learning Algorithms for Smart Manufacturing: Tool Wear Prediction Using Random Forests *Journal of Manufacturing Science and Engineering* **139** 071018
[4] Wu D, Jennings C, Terpenny J and Kumara S 2016 Cloud-based machine learning for predictive analytics: Tool wear prediction in milling *2016 IEEE International Conference on Big Data (Big Data)* 2016 IEEE International Conference on Big Data (Big Data) (Washington DC,USA: IEEE) pp 2062–9
[5] Dimla D E 2000 Sensor signals for tool-wear monitoring in metal cutting operations—a review of methods *International Journal of Machine Tools and Manufacture* **40** 1073–1098

[6] Cho S, Asfour S, Onar A and Kaundinya N 2005 Tool breakage detection using support vector machine learning in a milling process *International Journal of Machine Tools and Manufacture* **45** 241–9

[7] Martínez-Arellano G, Terrazas G, Benardos P and Ratchev S 2018 In-process Tool Wear Prediction System Based on Machine Learning Techniques and Force Analysis 8th CIRP Conference on High Performance Cutting (Budapest, Hungary: Elsevier)

[8] Nieto P J G, García–Gonzalo E, Vilán J A V and Robleda A S 2015 Modeling the milling tool wear by using an evolutionary SVM–based model from milling runs experimental data International Conference on Computational Methods in Sciences and Engineering 2015 (ICCMSE 2015) (Athens, Greece) p 190021

[9] Wu D, Jennings C, Terpenny J, Gao R and Kumara S 2017 Data-Driven Prognostics Using Random Forests: Prediction of Tool Wear *Volume 3: Manufacturing Equipment and Systems* ASME 2017 12th International Manufacturing Science and Engineering Conference collocated with the JSME/ASME 2017 6th International Conference on Materials and Processing (Los Angeles, California, USA: ASME) p V003T04A048

[10] Wu X, Kumar V, Ross Quinlan J, Ghosh J, Yang Q, Motoda H, McLachlan G J, Ng A, Liu B, Yu P S, Zhou Z-H, Steinbach M, Hand D J and Steinberg D 2008 Top 10 algorithms in data mining *Knowledge and Information Systems* **14** 1–37

[11] Boser B E, Guyon I and Vapnik V 1992 A training algorithm for optimal margin classifiers *Proceedings of the 5th annual ACM workshop on computational learning theory* (ACM Press) pp 144–152

[12] Vapnik V, Golowich S E and Smola A J 1996 Support Vector Method for Function Approximation, Regression Estimation and Signal Processing *NIPS*

[13] Smola A J and Schölkopf B 2003 A Tutorial on Support Vector Regression

[14] Tin Kam Ho 1995 Random decision forests *Proceedings of 3rd International Conference on Document Analysis and Recognition* 3rd International Conference on Document Analysis and Recognition vol 1 (Montreal, Que., Canada: IEEE Comput. Soc. Press) pp 278–82

[15] Miura K, Döbbeler B and Klocke F 2018 Cutting Power Estimation via External Voltage and Current Sensors on Feed-Drive Axis for the Straight Milling Process *Procedia CIRP* CIRPe 2018- 6th CIRP Global Web Conferene

[16] Kienzle O 1952 Die Bestimmung von Kräften und Leistungen an spanenden Werkzeugmaschinen *VDI-Z* VDI-Z 81. VDI-Hauptversammlung der Fachgruppe Betriebstechnik vol 96 (VDI-Verlag) pp 299–305

[17] Altintas Y and Lee P 1996 A General Mechanics and Dynamics Model for Helical End Mills *CIRP Annals - Manufacturing Technology* **45** 59–64