Discover Applied Sciences

Research

In production system for tool wear prediction using multi-sensor time series and machine learning models

Jonathan Dreyer[1,2] · Stefano Carrino[1] · Hatem Ghorbel[1] · Paul Cotofrei[2]

Received: 16 October 2024 / Accepted: 6 May 2025
Published online: 07 July 2026

Abstract
In micro-machining milling, optimizing the process is a major key to reducing production costs. By monitoring tool wear and detecting when the quality of the machining deteriorates, the process could be more accurately adjusted. The present research explores four approaches that use different machine learning models to predict machining tool wear during the milling process. The study is based on a dataset created from a face milling operation on stainless steel round (AISI 303) material. The workpiece used for the machining process is conceived as a set of multiple stairs shape in brass and is milled using a 3 mm tungsten carbide tool. Three different types of sensors—acoustic emission, accelerometers and axis currents—are set-up to measure the wear of the tool. The article provides a description of the different sensors used, and the dataset collected. The performance of the tool wear prediction is evaluated using the F1-score and a customized weighted expected value. The extra-trees classifier yields the optimal F1-score result of 73% when evaluated across the five classes, indicating its superior performance in comparison to other classifiers. Based on this best model, the implementation feasibility and the potential performance of an in-production system predicting machining tool wear is evaluated.

Keywords Tool wear monitoring · In production · Milling machining · Multi-sensors time series · Machine learning

1 Highlights

- Multi-approach studies using different machine learning models to predict tool wear during the machining process.
- New "expected value" metric used in the machining context that adjusts for the effects of prediction errors.
- In-production implementation that demonstrated the feasibility and the processing time.

2 Introduction

This article examines the possibility of improving milling performance by monitoring and assessing tool wear without affecting the continuity of the machining process. A mixed method is employed to detect tool wear, which represents a state-of-the-art methodology. This method is compared with other algorithms not yet considered in the literature on the discipline. The utilization of sensors for in-process measurement has the potential to enhance both productivity and

✉ Jonathan Dreyer, jonathan.dreyer@he-arc.ch; Stefano Carrino, stefano.carrino@he-arc.ch; Hatem Ghorbel, hatem.ghorbel@he-arc.ch; Paul Cotofrei, paul.cotofrei@unine.ch | [1]Haute Ecole Arc Ingénierie, University of Applied Sciences and Arts Western Switzerland (HES-SO), Rue de la Serre, 7, 2610 Saint-Imier, Switzerland. [2]Information Management Institute, University of Neuchâtel, Neuchâtel, Switzerland.

manufacturing control [1, 2]. The custom production processes have the effect of reducing the demand for large volumes of production. The use of small-sized machine tools is an effective solution for micro manufacturing, as it enables the miniaturization of production systems while maintaining high accuracy [3].

The field of computer numerical control (CNC) machining covers a range of machining techniques, including turning, milling and grinding. More precisely, turning is a machining operation in which a cutting tool moves in a longitudinal direction while the workpiece completes a rotary motion. Milling entails the utilization of a rotary cutting tool for the material removal process, which involves the movement of the cutting tool into the workpiece. Finally, grinding is a machining process that employs a circular grinding wheel to remove material from the surface, thereby producing a smooth surface finish. The present study will focus on the milling process, specifically the use of micro-milling machining technology.

The micro-machining machines consume less energy due to the scaling down of their dimensions [4]. The reduction in size has the effect of reducing the moving masses and increasing the dynamic motion, as evidenced by Grimske et al. [3]. High-speed machining (HSM) is advantageous for micro machining and significantly reduces the necessity for cooling during milling, as a considerable amount of heat is naturally dissipated by chips. According to Jain [5], the HSM can increase productivity by accelerating milling and cutting speeds while reducing the need for multiple tools.

The prediction of tool wear in the context of micro milling using sensors represents a significant challenge. In order to tackle this issue, we put ahead a number of approaches based on machine learning classification models, which include convolutional neural networks (CNN). By mining vast datasets and identifying hidden patterns, these algorithms can improve tool wear detection. The aim of the study is to investigate various machine learning techniques capable to improve tool wear detection during the machining process without interruption, which will be accomplished by monitoring the process with external sensors. In order to evaluate the performance, various models used in the article are evaluated (based on a cross-validation procedure) using two evaluation metrics: the first is the F1-score, and the other is the expected value adapted to the context of machining. Furthermore, we propose an approach for the in-production prediction of the tool wear.

This work builds upon our previous research on tool wear prediction [6] and introduces a novel metric in the machining field, the expected value. This metric may be used for the purpose of comparing the performances of algorithms within the context of industrial machining. Furthermore, the state of the art is more comprehensively detailed, with the incorporation of two pivotal elements of machining: high-speed machining and tool wear profile. The current work uses the same dataset (actually available on Zenodo [7]) employed in our previous publication, but includes a more detailed description about the acquisition process. This paper outlines the feasibility and process of treatment during machining, as well as the treatment time. Finally, the potential for utilizing the methodology in a production setting has been explored.

The rest of the paper is structured as follows. The next section is focused on the state of the art of the machining (conventional & high-speed machining), tool wear (profile, sensing & prediction) and machine learning models. The milling process, the necessary materials (e.g., the sensors), the dataset description and the global methodology in the context of a machine learning approach (from dataset to model evaluation) are detailed in the Sect. 4. The following section is devoted to the analysis of the models' performances for each approach, whereas the Sect. 6 presents an approach for the in-production detection of milling process. The closing section contains a discussion about the achieved results, followed by the final conclusions.

3 State of the art

The objective of this study is to examine the potential correlation between tool wear and the data obtained from various sensors in the context of micro milling and high-speed milling. Our aim is to advance the detection of tool wear and identify the most effective sensors for tool wear detection.

In the field of machining, the wear of tools is dependent on a number of factors, such as physical constraints, milling materials, etc. [8]. Tool wear management is a crucial aspect of optimizing the milling process. Physical constraints, such as cutting forces, accelerations, and vibrations, can be quantified through the use of sensors, which provide insights into the condition of the tool [1].

3.1 High-speed machining

Recently, the field of high-speed machining has gained a new interest in the manufacturing industry due to its potential for enhancing production speed, lead time reduction, cost savings, and increasing part quality. HSM originated in the cutting speeds, which are significantly higher to those traditionally used in conventional machining, with a spindle rotation up to 100'000 rpm and even more [5]. This approach needs certain requirements, like special bearings on moving components, high feed rate capability, CNC system that anticipates the path to reduce the load on the tool, cutting tools-toolholders-spindle balanced for this high-speed rotation, coolant systems that provide enough flow pressures to dissipate tool heat produced, and chip control systems for high volume flows. The application of this machining is mainly used in aircraft industries, automotive, computers, medical and mold industry [9].

3.2 Tool wear profile

The tool wear degradation can be measured with the flank wear, and is generally a function of cutting time. The shape of the curve (illustrated in Fig. 1), can be decomposed into three consecutive stages designated as: break-in, steady-state, and failure [9, 10].

After a relatively short break-in period characterized by a fast wear, the tool will pass into the steady-state in which the wear slowly increases. The last phase of the tool life cycle, failure is characterized by a rapid deterioration of the tool at the end of which, the tool is no more usable [9].

3.3 Tool wear sensing and prediction

Adaptive control systems aim to estimate the remaining useful life (RUL) of a tool in order to enhance precise control and optimization of the machining process. CNC machine operators have the expertise to discern excessive tool wear by discerning the sounds and vibrations produced by the machine during operation. Therefore, the tool degradation can be indirectly observed without stopping the production by using information captured by external sensors. This non-destructive method permits the estimation of the current degree of wear of the tool. Furthermore, it enables the detection of instances where the tool has reached a point of excessive wear and requires replacement [11, 12].

The milling process and the related physical parameters can be monitored and recorded by a variety of sensors [13], each of which is capable of capturing a specific aspect of the process. These sensors are employed in the modeling of the behavior of one or multiple physical elements. For example, acoustic emission (AE) sensors are identified as a valuable source of information [14–18]. However, one limitation of AE sensors is the presence of noise in the signal, as well as the

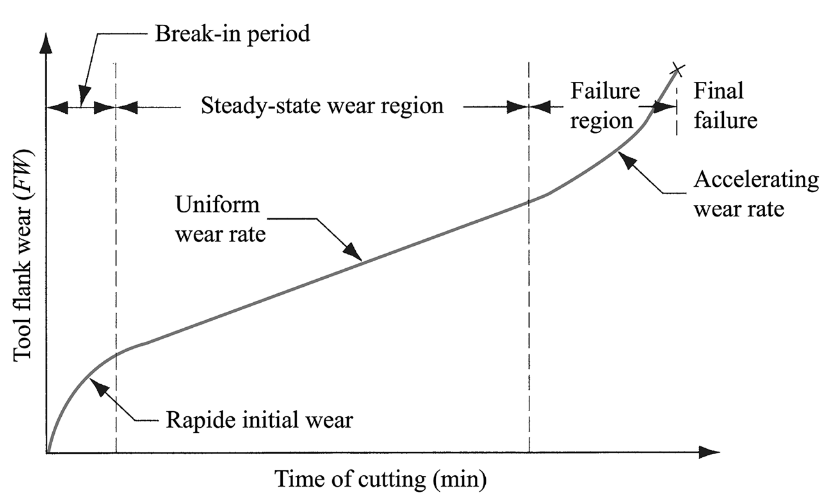

Fig. 1 Tool wear as a function of cutting time (Groover 2020). (Reprinted with permission, courtesy of Mikell P. Groover; This figure is not included under the article's CC BY license.)

potential influence of unrelated sounds and vibrations in the machining environment. Less sensitive to the environment noise than AE, the accelerometers placed on the machine axes can also provide significant signals which can be correlated with the tool wear process to estimate the wear [19, 20]. Accelerometers provide information on the vibration along the axes, which is a consequence of the interaction between the tool and the material. The physical effort generated on an axis can be correlated with the power consumed by that axis. This power consumption can be quantified by measuring the electric current drawn by each motor (axis and spindle). The correlation between consumption and tool wear has been demonstrated by a multitude of researchers [19, 21, 22]. All these available data sources can be employed individually or merged to create a multivariate dataset, as previously demonstrated by [12, 17, 19, 23, 24].

In order to retrieve pertinent details from data sources and reduce the data set size, several researchers have suggested the use of features in place of raw data. These researchers have shown that spectrograms can be used in place of raw acoustic emission time series [25–28]. The underlying concept is to process data in order to derive features that represent the frequency and amplitude over a specified time interval. Statistical features with a less computational cost can also be used instead of raw data [19].

3.4 Machine learning models

The literature review shows that tool wear detection can be addressed by a number of approaches, among which machine learning and artificial intelligence methods were widely studied. Krishnakumar et al. [17, 29] employed statistical features derived from vibration signals to train classification models, including decision tree, artificial neural network and support vector machine, with the objective of predicting the stage of tool wear. Cao et al. have demonstrated that features extracted from derived wavelet frames with a convolutional neural network (CNN) yield robust results [30].

The potential of deep learning for modelling tool wear states has been investigated by numerous researchers. In their work, Dou et al. [10] employed a sparse auto-encoder model, which was trained on vibration and force signals. Von Hahn et al. [31] proposed a disentangled-variational-autoencoder, combined with a temporal convolutional neural network. Liu et al. [32] developed a transformer-based neural network in addition to a long short-term memory network, utilizing temporal features extracted/derived from unprocessed signals.

4 Methodology

In the methodology section of this paper, we describe the context of this study, including the type of sensors, the collected dataset, and the milling process. We then enumerate the study's objectives, and detail the four approaches used to achieve these objectives.

4.1 Context

The project analyzes the *micro* milling machine,[1] micro^5, which has 5-axis of movement and less than 10 Kg of moving masses, with the ability to achieve high-speed machining capabilities at up to 60,000 revolutions per minute. The machine's kinematics consist of three linear axes (X, Y, Z) and two rotary axes (B, C), which correspond to the type 57 as defined by ISG-kernel.[2] The data utilized in this project are generated by sensors integrated into the machine.

A new dataset, comparable to that presented in Carrino et al. [26], was generated for the purposes of this study. The principal distinctions between the present study and its predecessor lie in the shape of the raw material (a cylindrical shape is employed in the present study, in contrast to a cubic one) and the composition of sensors (the number of AE sensors is reduced to one, and multiple accelerometers are utilized).

4.1.1 Sensors

The signals collected for this study are provided by three different types of sensors: electric current sensors (one for each axis), accelerometers (placed on two of the five axes) and an acoustic emission sensor. Figure 2 shows location of the accelerometers and the acoustic sensor on the machine.

[1] https://www.he-arc.ch/en/rd-projects/micro5/
[2] https://www.isg-stuttgart.de/en/

Fig. 2 Micro[5] machine with accelerometers & acoustic sensors locations. (Reprinted with permission, courtesy of Guillaume Perret; This figure is not included under the article's CC BY license.)

Fig. 3 C axis with accelerometers sensors [6]

The AE sensor is located within the milling machine, as illustrated in Fig. 2, in the region designated by III. It is positioned close to the raw material where the machining process takes place, but not directly glued to the raw material as in [26]. The Vallen VS45-H acoustic sensor is used for acquisition, which is carried out with an Advantech PCIe 1840/L. The sampling acquisition is limited to 200 kHz. The sensor is positioned as close as possible to the raw material on the B axis, which allows replacing the raw material without moving the AE sensor.

The spindle axis and the stator of the C axis are equipped with multiple accelerometer sensors. Two types of sensors *Brüel & Kjær* were used: a triaxial sensor (type 4520) and a uniaxial sensor (type 4507). The sensors that measure acceleration are positioned at I, II and III as shown in Fig. 2. The spindle has one 3-axis accelerometer sensor located at the bottom for the XYZ axis and two one-axis accelerometers located at the top for the XY axis. Three one-axis accelerometers are mounted on the stator of the C axis to measure the XYZ axis accelerations. The National Instruments NI-9234 modules are used to acquire data from accelerometer sensors, which are sampled at a rate of 2 kHz. The placement of the sensors is illustrated in Figs. 3 and 4.

The acoustic and accelerometer signals are synchronized with an external trigger that is recorded by the acquisition equipment. For these data sources, we have chosen acquisition frequencies using an equipment with high acquisition frequency and wide bandwidth to retain as much information as possible in the signal. The recorded signal is not digitally filtered in order to preserve as much of the embedded data as possible. When some of the values are missing, the ongoing milling operation is ignored, as explained in Sect. 4.1.4.

The axis motors are driven by a Triamec Servo Drive, operating at high-speed frequency, and the currents of each are stored during the milling. The acquisition sampling rate is limited to 1 kHz. The spindle is connected to Sieb & Meyer drive connected via digital and analog inputs/outputs. For this data source, a digital filter is used to reduce noise from the power supply and motor drive. Additionally, the motor acts as a filter due to its large inductive element.

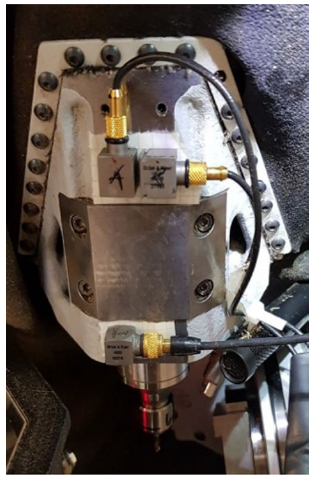

Fig. 4 Spindle axis with accelerometers sensors [6]

Table 1 shows some details (type, location and sampling rate) for each recorded signal. In this article, we will focus on acoustic emissions, accelerometers and axis/spindle currents.

4.1.2 Dataset

The signals collected by the sensors correspond to several milled parts produced under three different conditions (machining direction, number of stairs and lubrication type). Prior to machining the parts, two test pieces have been produced, but with some variations in the machining process. The number of stairs was adjusted to ensure sufficient tool wear. Defects in some machined parts have occurred due to errors in axis control, rendering them unusable. All the parts were all milled at the same spindle speed and with the same diameter tool. Further details on the milling parts description, summarized in Table 2, are described in the following section.

The collected signals are split into three datasets, one for each data source (acoustic emission, accelerometers and numerical control). The used CNC produces and stores a large amount of data relating to the control of the axes (e.g., temperature, axis position, axis following error), but this article focuses on the axis/spindle currents.

The dataset generated during several experiments on the micro[5] machine is freely available on Zenodo [7].

Table 1 Signals description

Signal type	Axes	Location (Fig. 2)	Sampling rate
Acoustic emission	Axis C	III	100 kHz
Accelerometer	Axis C - X	III	2 kHz
Accelerometer	Axis C - Y	III	2 kHz
Accelerometer	Axis C - Z	III	2 kHz
Accelerometer	Spindle (Top) - X	I	2 kHz
Accelerometer	Spindle (Top) - Y	I	2 kHz
Accelerometer	Spindle (Bottom) - X	II	2 kHz
Accelerometer	Spindle (Bottom) - Y	II	2 kHz
Accelerometer	Spindle (Bottom) - Z	II	2 kHz
Current	Axis X	-	1 kHz
Current	Axis Y	-	1 kHz
Current	Axis Z	-	1 kHz
Current	Axis B	-	1 kHz
Current	Axis C	-	1 kHz
Current	Spindle	-	1 kHz

Table 2 Milling parts description

Milling part ID	Milling direction	Stairs (front / rear)	Cooling	Remarks
1100.002	-	1 / 0	CO_2	Milling test
1100.003	-	1 / 0	CO_2	Milling test
1100.004	X	4 / 0	CO_2	
1100.005	X	4 / 0	Pressurised air	
1100.006	X	4 / 2	Pressurised air	
1100.007	X	5 / 2	Pressurised air	
1100.008	X	4 / 2	Pressurised air	Machining error
1100.009	Y	5 / 2	Pressurised air	
1100.010	Y	5 / 2	Pressurised air	
1100.011	X	5 / 2	Pressurised air	
1100.012	Y	3 / 2	Pressurised air	Machining error
1100.013	Y	3 / 2	Pressurised air	Machining error
1100.014	X-Y	5 / 2	Pressurised air	
1100.015	X-Y	5 / 2	Pressurised air	

4.1.3 Milling process

The machining process consists of milling the raw material in several stairs; to create each stair, the milling path is again divided into multiple linear passes (see Fig. 5). Because the piece is a cylinder, we can never have exactly the same milling width for two successive passes. The term **experience** will be used to refer to a single block of raw material that has been machined. This block starts with a new tool and ends when the tool breaks or becomes unusable.

For this study, a subset of the parts, described in Table 2 and using the same machining conditions, has been chosen. This subset consists of six experiences of the same machining path, namely 1100.007, 1100.009, 1100.010, 1100.011, 1100.014 and 1100.015. The cutting technique utilized for the manufacture of parts is conventional milling.

The machine is configured to follow a linear milling path along three orientations: X (experiences 1100.007 and 1100.011), Y (experiences 1100.009 and 1100.010), and X+Y (experiences 1100.014 and 1100.015). The experience parts are formed from a round bar and then sliced into cylinders. Given the cylindrical nature of the shape, the length of the linear passes outside the shape is shorter than that of passes located nearer to the center of the raw material, as shown in Fig. 6. For each pass, the tool is moving for the same distance, but because of the cylindrical shape, the length of actual machining inside (respectively outside) the material increases (respectively decreases) non-linearly among successive passes. During each experience, the direction of milling pass (e.g., right to left) is preserved.

The tool employed in the experiments has a nominal diameter of 3 mm and is manufactured from tungsten carbide (article number 3100d3.00, from Louis Belet manufacturer). At the beginning of each experiment, the tool is replaced with a new one. The spindle and tool rotation speed is set to 33,000 revolutions per minute (RPM) and is consistent across all experiments. The parts are milled using compressed air lubrication.

In order to automatically identify and delineate when the tool is milling the raw material, and not just moving without contact, a dynamic threshold is employed to assess the intensity of the acoustic signal. As a first step in dynamically fixing the threshold, the signal is smoothed with a moving average over 100 values. The difference between the maximum

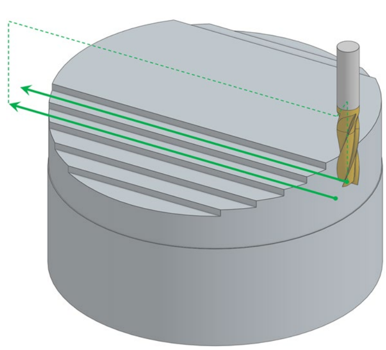

Fig. 5 Multiple linear passes [6]

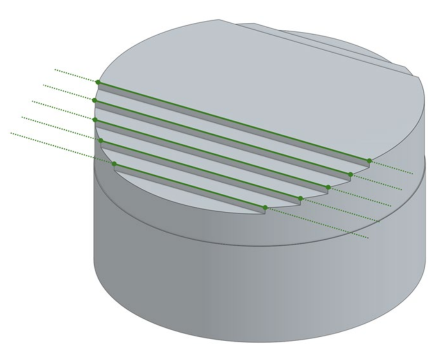

Fig. 6 Different lengths of linear passes [6]

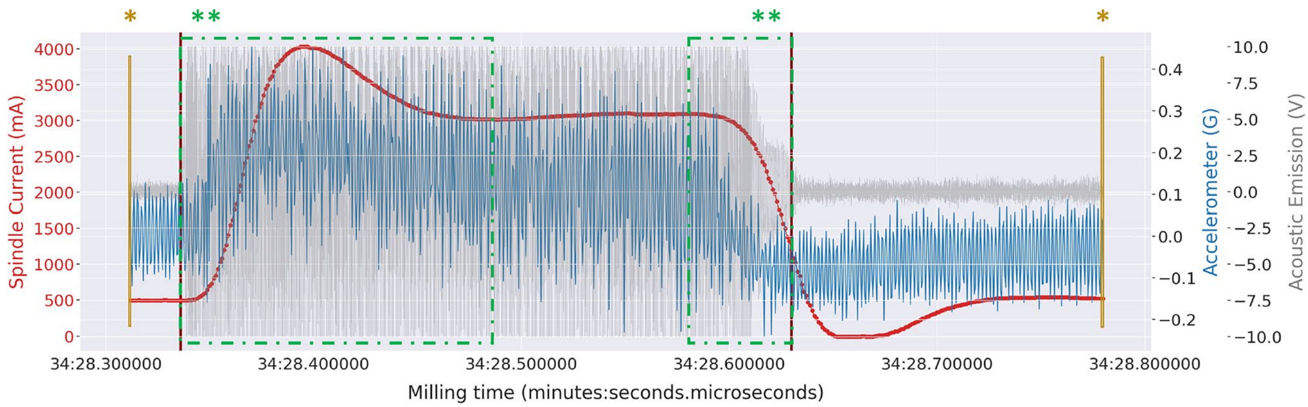

Fig. 7 Signals detected for a typical machining pass. * time window for maximal value detection (in yellow); ** transient phases during milling (in green). Three different signals are illustrated: the acoustic emission (gray), one of eight accelerometers (blue) and the spindle electric current (red) [6]

signal value in the part center of the pass and the maximum signal value at the beginning and end of the pass is quantified. The machining process design ensures that the tool does not touch the material at the beginning and end of each pass. The threshold for segmenting each linear pass is determined by using the midpoint of the difference. Specifically, the beginning and ending of a pass are delimited as the first 2000 values and the 2000 latest values, respectively. During these two time frames, the tool is not in contact with the material, and therefore no milling occurs. The time frame for 2000 values corresponding approximately to 1ms (refer to Fig. 7).

Figure 7 illustrates three distinct types of signals: the acoustic emission (shown in gray), one of the eight accelerometers (shown in blue), and the spindle electric current (shown in red). The green areas highlight the transient phases of a milling pass.

Each linear pass is characterized by the presence of a transient phase (when the tool impacts, enters the material and when the tool gradually moves away from the material) and a stationary phase (the tool is completely in the material) - as illustrated in Fig. 7. Thus each linear pass can be divided in three distinct phases: "entering" transient, stationary and "exiting" transient.

4.1.4 Objectives of the study

The focus of this article is the milling process, with a particular accent applied to the process during which the tool removes material. An experience is considered completed, if, at the end of the milling process, a tool is considered as no more usable (broken, burned out, etc.). The main objective is to develop a model that is capable of predicting the remaining useful life (RUL) of the tool. To apply classification models, it is necessary to first discretize the continuous variable RUL. Five states (or labels) will be used to perform a finer tool wear evaluation, denoted by numbers “0” to “4”. The number “0” corresponds to a new tool, while the number “4” corresponds to a non-usable tool. We have chosen that at the end of the milling operations, the tool is deemed unusable and labelled as “4”. The label “0” corresponds to the initial passes when the tool is new. In each experience, the tool label progresses from “0” at the beginning of the milling process to “4” at the end of the process. The distribution of the remaining labels is done using two different time-based methods. Each approach is explained in detail in the following section.

Table 3 presents the approaches or strategies considered and evaluated in this paper.

Table 3 Approaches evaluated in this article

Section	Title
4.2	Initial approach: dataset validation and baseline
4.3	Focus on machine learning algorithms using statistical features
4.4	Focus on stationary components of the signals
4.5	Mixing spectrogram approach and stationary components of the signals

As the dataset contains outliers and peculiar values due to the implicit process variability, it is shuffled and split into multiple folds to mitigate such effects. To assess the overall performance of the classification model, we employ a cross-validation methodology [33] with six folds, one for each experience. During the models training, we use four folds for training and two for testing. Therefore, the ratio between training and test sets is fixed at 66% for the training and 33% for testing. To mitigate the potential for performance variability across folds, all possible combinations of folds (with a ratio of 66–33%) are utilized, i.e. 15 training run.

The performance measures used to compare different classification models and approaches include two metrics returned by the cross-validation process: the mean of the F1-score and the mean of the expected value. The F1-score is a measure that balances the precision and recall of a classification model, providing an overall measure of its precision. The expected value is a metric that measures the performance of a model when the cost of incorrect predictions varies across classes.

Formally, after defining a weighted cost/benefice matrix of the same size as the confusion matrix, the two matrices are multiplied element-by-element, and the expected value is calculated as the sum of the elements of the product matrice, normalized by the number of instances in the test set. For our study, we decided to consider the expected value of the decision to replace/not replace the tool/part based on the prediction of a classification model with only two classes, "good" and "worn out", related to the tool quality. This model is obtained by binarizing the initial classification model with five classes/labels, by grouping the labels 0 to 2 in the class "good", and the labels 3 and 4 in the class "worn out". The binary confusion matrix of the new model can be obtained from the 5 x 5 initial confusion matrix as shown in Fig. 8.

The corresponding weighted cost/benefice matrix is displayed in Fig. 9. The matrix has a cost of -100 for the worst case, i.e., the false positive (FP) case, when the tool is worn out but the model predict a good tool. In this case both the tool and the part need to be scrapped with the worst impact on the milling process. The tool is replaced by default when the model predicts "worn out", but the matrix penalizes six time more the false negative (FN) case, where the replaced tool was in fact good, compared to the case true negative (TN) where the tool was really worn out.

Table 4 summarizes weighted values with description for each scenario. It should be noted that the weights represent relative costs, all being negative, excepting the weight for the best, true positive case, which is zero. Moreover, the expected value of a perfect model (with no prediction errors), which corresponds to the highest possible value for this performance metric, is -4 whereas the expected value for a baseline, random model is -40. In order to facilitate the interpretation of the results, the EV values are presented in the form of percentages. These percentages are inverted to reflect the impact on the production, with 0% representing the best result (highest EV possible value: -4) and -100% denoting the worst outcome (random model: -40). The processing time of classification is not considered because it is in the order of few milliseconds and does not impact the process.

It is worth noting that initially, in this multi-source/experience dataset, there are some missing data in different experiences. To avoid any impact on the comparison among data sources, missing values have been removed from the linear passes in all experiences. For instance, if data values are missing for an accelerometer, the entire pass is discarded, even if the signals corresponding to other data sources are not missing. This methodology is applied to all data sources, including acoustic emission and electric current signals. The removed passes represent approximately 5.5% of the entire dataset. This technique has been applied to all approaches listed in Table 3.

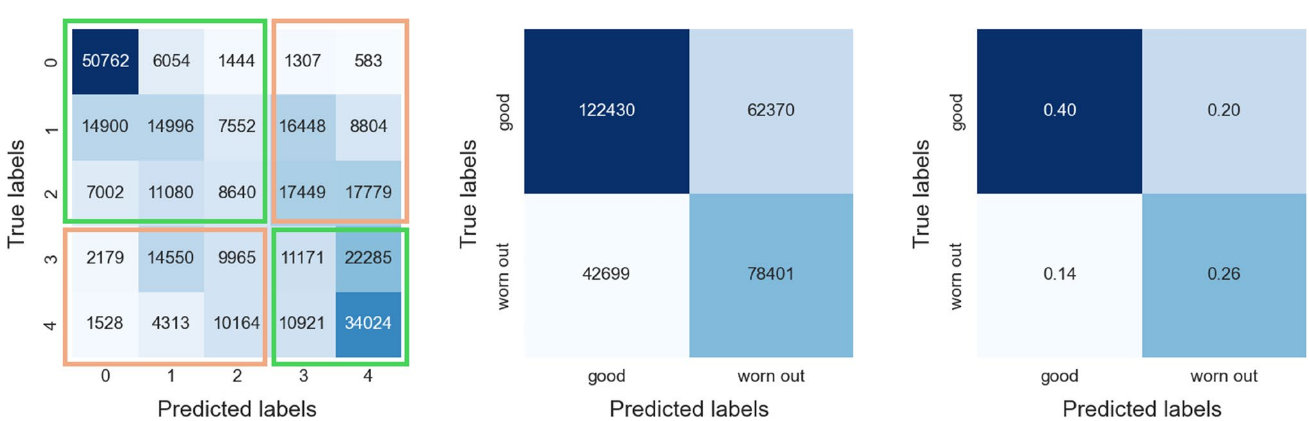

Fig. 8 Binarization for multiclass tool wear (results of approach 4.2)

Fig. 9 Weighted cost/benefice matrix

Table 4 Interpretation of the expected values weights

Confusion matrix metrics	Values	True & predicted labels	Production impact
TP	0	Unworn tool with a prediction "unworn"	Uninterrupted production with no loss.
FN	-60	Unworn tool with a prediction "worn"	Interrupted production (-10) with unneeded replacement tool (-50).
FP	-100	Worn tool with a prediction "unworn"	Uninterrupted production with defective work pieces.
TN	-10	Worn tool with a prediction "worn"	Interrupted production (-10) with needed replacement tool (0).

4.2 Initial approach: dataset validation and baseline

The objective of this initial approach is to validate the dataset and establish a baseline for use as an initial comparison point among each data sources. In this approach, the dataset is processed in order to extract spectrograms from the acoustic emission signal, which have a length of approximately 200 milliseconds (i.e., 39936 samples). In this approach, only acoustic emission data obtained during the milling process are utilized. The number of spectrograms extracted from each milling pass is contingent upon the length of the pass. One or two spectrograms are extracted from each pass, depending on the length of the pass. The spectrograms are converted into images, which are then employed as training data for a CNN. In consideration of the fixed dimensions of the image utilized by the CNN, specifically 144px by 144px, only those linear passes that are a multiple of the spectrogram size are employed. Linear passes of insufficient length are excluded from the spectrogram generation process. In order to achieve a stable learning process with cross-validation, the training of the CNN is repeated 10 times in this initial approach.

The labeling of spectrograms in five classes is based on a discrete uniform distribution. To provide further clarification, the sequence of all passes for each experience is organized into five equal-sized bins with the same classe. This proposed methodology will serve as the baseline for the investigation of acoustic emission datasets in the following sections. Moreover, an attempt was made to enhance the model performance by guiding the learning process to diagonalize the multiclass classification problem, based on the implicit class order ("0" > "1" >.. > "4"). This classification method is commonly referred to as *ordinal classification* or *ranking learning*. This concept is better aligned with the principal goal, which is to determine the remaining lifespan of the tool. In this instance, the consequence of an inaccurate prediction is more severe when the prediction is more divergent from the ground truth. For example, an incorrect classification of class 0 (indicating that the tool is new) as class 4 (indicating that the tool is totally burnt out) should be penalized more severely than an incorrect classification of class 0 as class 1 (indicating that the tool is new but has some sign of worn out). As Gaudette and Japkowicz (2009) [34] have previously explained, the metrics RMSE (root-mean-square error) or MSE (mean square error) are more appropriate for this type of classification task.

4.3 Focus on machine learning algorithms using statistical features

This approach, which is composed of two distinct sequential phases, is centered on the definition and utilization of statistical features for the investigation of the performance of machine learning algorithms utilizing multiple data sources.

The preliminary phase is devoted to an exploration of a large selection of machine learning algorithms and statistical features. The second phase is dedicated to the optimization of the most promising algorithms by fine-tuning the hyperparameters of models.

In the preliminary phase, a number of classification algorithms with their preset parameters are evaluated. This phase makes use of the *lazypredict* library [35] to assess a diverse range of algorithms based on scikit-learn models [36].

The features exploited by algorithms exert a considerable impact on the overall performance of the model. A previous study [26] utilizing the similar dataset demonstrated a dependable capacity to ascertain the extent of tool wear. Accordingly, the selected features are intentionally chosen with the objective of attaining a performance at least comparable to that previously achieved. In this previous study, the relationship between acoustic emission and tool wear has examined with a series of features, including the mean, standard deviation, minimum, maximum, first quartile, second quartile, and third quartile. All features are derived from the specified window of raw signals. The corpus of data is subdivided into three distinct datasets, namely acoustic emission, accelerometers and currents. The data extraction process creates three distinct data sources, each of which is composed of features extracted from the respective signal in the data source.

The labeling process applied in this approach is also based on five classes, but the method used to assign labels is distinct. In contrast to the previous approach, which distributed the sequence of linear passes equally across the five classes, this approach considers that only the actual machining time affects tool wear. This is based on the premise that when the tool is not in contact with the material, no wear occurs. In practice, the time interval between the beginning and the end of the experience is divided equally into five subintervals, with all passes occurring in a given subinterval receiving the same label. This method of labeling is more consistent with the tool wear profile (as presented in Sect. 4.2). This approach has a slight impact on the label balancing, but it allows for an accurate distribution between each class to be maintained.

Figure 10 illustrates the performance of various classification models utilizing electric current signals, with a 90% confidence interval for F1 scores and expected value. On the other hand, when the same algorithms are applied to other data sources, such as accelerometers or acoustic emission, they demonstrate a comparatively weaker performance.

In accordance with the F1-score metrics, Extra-Trees and Random Forest classifiers demonstrated the optimal performance and were selected for fine-tuning. The majority of the fine-tuning process was dedicated to the optimization of hyperparameters, using Optuna [37] (a hyperparameter optimization framework designed to automate the search for optimal hyperparameters for machine learning models). The tool employs efficient algorithms to efficiently explore the search space and identify optimal configurations, rendering it an interesting instrument for enhancing model performance. The identification of optimal hyperparameters is achieved through an iterative process, wherein the learning process is repeated 1000 times. The results obtained with the optimal hyperparameters are presented in Sect. 5.

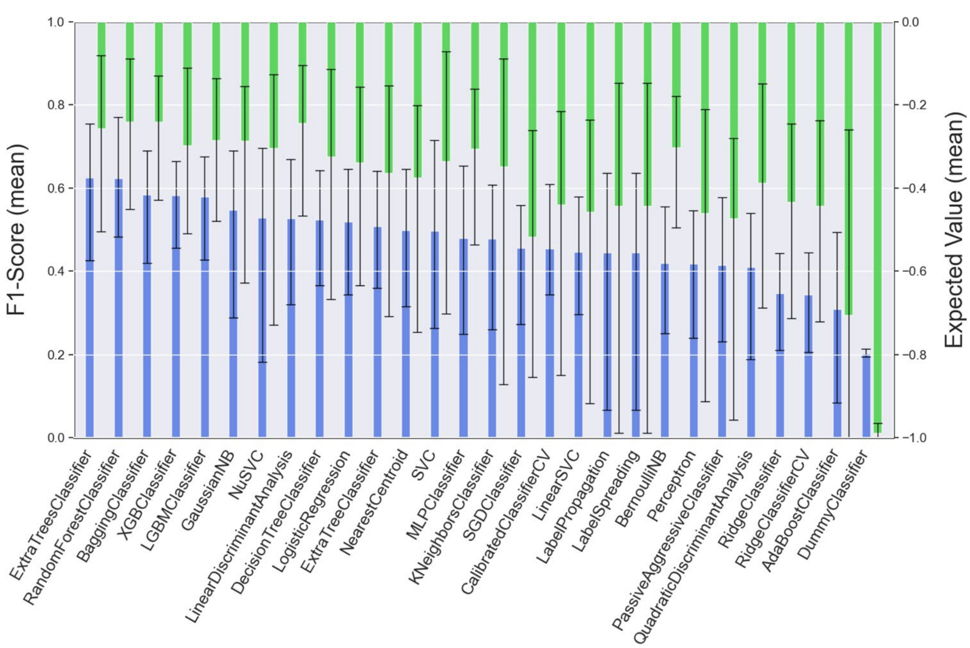

Fig. 10 Result of algorithms comparison using currents as data source

4.4 Focus on stationary components of the signals

In previous approaches, the analysis focused on all signals generated during machining phase, described as a sequence of three states ("entering" transient, stationary, and "exiting" transient, see Sect. 4.1.3). As illustrated in Fig. 7, the transient states affect the signals' shape by generating spikes and noise, as evidenced by the machining signal analysis.

In this approach, the transient states were excluded, and the focus was placed on the analysis using the features extracted from stationary components of the signals. The underlying rationale for this approach is that the transient regimes of each pass is too noisy and should be excluded. This decision was based on the observations that the spindle electric current and the shape of the transient regimes are entirely distinct when comparing transient and stationary states (Fig. 7).

In order to eliminate the transient signals, the timeframe of both transient states (entering and exiting the material) was determined through all acquired datasets. It was determined that the duration of the "entering" transient is less than 150 milliseconds, while the "exiting" transient is approximately 50 milliseconds. The transient states are removed in order to isolate and retain only the stationary component of each linear pass of the signal, which is used for feature extractions. This process has the potential to enhance both the quality of the extracted features and the classification performance. With regard to the electric current signals, it can be observed that the values of the spindle and non-moving axes remain constant throughout the milling interval. It is anticipated that the acoustic emission will yield a more consistent signal with stable frequencies. Furthermore, the accelerometer values indicate that the vibration should be more consistent as the force exerted remains at approximately the same level. As a concluding step, it was resolved that only stationary signals with a duration exceeding 100 milliseconds should be retained, in order to ensure the conservation of meaningful signals.

4.5 Mixing spectrogram approach and stationary components of the signals

The last approach examined in this article is the use of spectrograms and CNNs to improve the signal pre-processing step, as detailed in [26]. The objective is to retain only to the stationary components of the signals presented in Sect. 4.4, and to apply spectrogram extraction. The rationale underpinning this combined strategy is to eliminate the effects of transient phases on the acoustic emission signals.

The utilisation of solely the stationary components of the signal results in a spectrogram duration that is approximately half that of the original length. This approach ensures that the spectrogram representation does not contain the frequencies generated by the transient states of milling, which have the capacity to pollute and distort spectrographic representations.

The labeling method employed is identical to that used for stationary signals and introduced in Sect. 4.3.

5 Analysis of models' performances

The analysis section is divided into four parts, one for each approach. It should be noted that the same dataset is utilized for all approaches, though the data sources and pre-processing methods vary. For each approach, the performances are presented in detail, accompanied by a concise discussion. The scores are summarized in Fig. 12, located at the end of this paragraph.

Initial approach: dataset validation and baseline—The approach yielded an F1-score of 35% on five classes, rendering it unsuitable for practical applications. For the purpose of comparison with the other approaches, the second metric, the expected value, yielded a result of -67.8%. The potential cause of this substandard performance is the presence of erroneous acoustic data. The modification of the loss function to minimize the diagonal spread (*ordinal classification*) score in a reduction of the F1 score to 25%.

Focus on machine learning algorithms using statistical features—As shown in Fig. 10, the optimal performance observed in the comparison of algorithms was achieved by the *ExtraTreesClassifier*, with a F1-score of 62% (ex-aequo with the *RandomForestClassifier*). In the case of other data sources, it is important to note that only the highest algorithm score is presented. The merged approach achieves an F1-score of 58%. The F1-score for the AE data source is 43%, while the score for the accelerometers data source is 41%.

By optimizing all algorithm parameters described in scikit-learn with Optuna, the F1-score increased up to around 66% for both top mentioned algorithms. The classifiers have an expected value of -20.4% (*ExtraTreesClassifier*) and -21.2% (*RandomForestClassifier*).

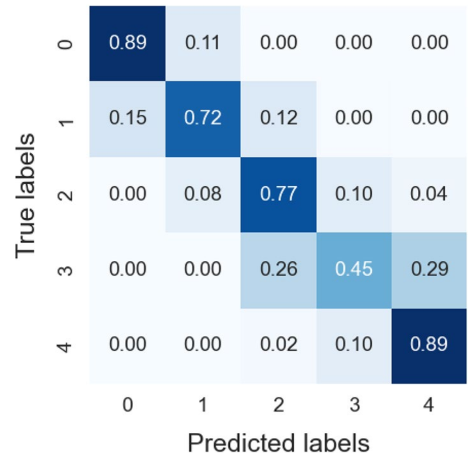

Fig. 11 Classification performance of *ExtraTreesClassifier* with Optuna optimization

			Baseline Score		***Optimized Score***	
Approach	***Data source***	***Model***	***F1-Score***	***Expected Value***	***F1-Score***	***Expected Value***
Initial approach	Acoustic	CNN	35.0%	-67.8%		
Approach based on statistical features	Currents	ExtraTreesClassifier*	62.5%	-25.6%	66.4%	-20.4%
	Merged	RandomForestClassifier*	58.2%	-27.1%		
	Acoustic	NuSVC*	43.3%	-58.1%		
	Accelerometers	LGBMClassifier*	41.4%	-29.6%		
Approach based on stationary components	Currents	ExtraTreesClassifier*	69.1%	-21.3%	73.5%	-18.2%
	Merged	ExtraTreesClassifier*	67.8%	-19.9%		
	Accelerometers	RandomForestClassifier*	49.0%	-44.5%		
	Acoustic	LogisticRegression*	46.9%	-32.9%		
Mixing approach	Acoustic	CNN	36.3%	-61.3%		

* Best models of 30 classifiers

Fig. 12 Summarize models performances

Focus on stationary components of the signals—This approach follows the same two-step methodology. For the classification of all algorithms, once again the *ExtraTreesClassifier* and the *RandomForestClassifier* reach first with a F1-score of ~69% with the electric current signals dataset. The merged data sources achieve an F1-score of 68%, while accelerometers achieve a lower F1-score of 49%. Finally, the F1-score of the AE data source is 47%.

The F1-score for the best algorithm optimized using Optuna has increased to 73%. The confusion matrix score is shown in Fig. 11. The expected value for the *ExtraTreesClassifier* is -18.2%, which is the best score achieved between among all the approaches.

Mixing spectrogram approach and stationary components of the signals—The approach achieves an F1-score of 36.3%, which can be compared to the baseline approach's score of 35%. There is a slight improvement in the initial approach, but it is not significant. The expected value for this approach reach is -61.3%.

The section's score indicate that the spectrogram approaches did not yield satisfactory performances, while the feature-based approaches showed promise. By optimizing the algorithms' hyperparameters with Optuna, the F1-score increased to 69.1% and 73.5%. The expected value can be improved from -21.3% to -18.2%.

6 Implementing an in-production predicting system

The objective of this section is to demonstrate the feasibility of an in-production prediction system that is capable of predicting the state of the tool wear after each pass, as opposed to once the machining is complete.

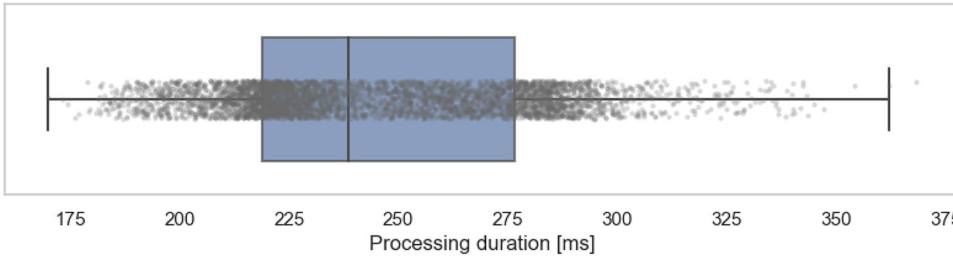

Fig. 13 In-production processing time

In order to demonstrate the feasibility of the proposed prediction system, the same dataset will be used, but the entire data processing pipeline will be executed on each pass. The system is composed of three stages: 1. The first stage involves the collection of recorded signals from each data source at the conclusion of the milling pass. 2. The second stage encompasses the processing of signals, as outlined in Sect. 4.1.3; this stage involves the detection of the milling part, the extraction of features, and the normalization of the signals. 3. The third stage involves the analysis of the signal and the prediction of the tool wear value.

The objective of the in-production predicting system is to achieve the fastest possible prediction of tool wear, composed by steps 1, 2 and 3. The time taken to record and store the data sources were not documented at the time of the experiments, and thus cannot be reproduced for analysis. It was decided that this particular period should be excluded from this system analysis, as the analysis showed that this operation represents only a small proportion of the pipeline processing time. For purposes of comparison, the storage of a single pass on the same hardware platform described in the following section takes approximately 40 milliseconds.[3] In this study, the loading time of the disk for each data source is considered for the step 1. The maximum time available for processing is determined by the interval between the end of the pass and the next pass. In the context of these experiences, the time available is at least 1.35 s. In the case of real part machining, it is pertinent to note that the available time could be shorter.

The in-production system uses the best, optimized model *ExtraTreesClassifier*, using stationary components (described in Sect. 4.4 and evaluated in Sect. 5). The processing time is only considered and measured when the system is able to predict the tool wear class. That is, when the stationary part of the milling pass exceeds 100 milliseconds. The inability of the system to predict when the pass is too brief is not overly constraining in this context, as such "brief" passes account for approximately 13.3% of the total milling time.

The processing times were obtained after implementing the system on a hardware platform using an Intel Broadwell generation processor released in 2016 (model E5-2697AV4), with a base frequency of 2.6GHz and a maximal frequency of 3.6 GHz. The memory utilized is DDR4-2400 with error correction code functionality.

In the methodology employed, acoustic signals are exploited to detect the ongoing milling process, whereas electric current signals are utilized to generate the features. These features are then used by the *ExtraTreesClassifier* model. The performances of the model, after simulating the application of the in-production system on all available datasets, are significantly better (F1-score of 89.4% and expected value of -5%) compared with the generalized performances estimated in Sect. 5 (73.5% for F1-score and -18.2% for expected value). This is due to the fact that the model used by the in-production system was trained and optimized on the entire dataset.

The variation in processing time, related to the volume of data to be processed, is illustrated in Fig. 13. The mean processing time is 245 milliseconds, with a maximum duration of 370 milliseconds. The relatively brief processing time enables the system to be deployed during the production process, allowing for the detection of tool wear.

7 Conclusion—discussion

This article presents different approaches to predicting tool wear during the milling process on a micro machine with high-speed machining capabilities optimized to improve the milling process and reduce manufacturing costs. The objective is to investigate different machine learning methods to improve tool wear detection/prediction by measuring three different data sources without interrupting the machining process.

[3] From memory into a solid-state drive (Samsung 980 PRO 1TB).

The paper describes the process of generating the dataset, using three data sources: acoustic emission, accelerometers, and axis currents. The dataset, freely available on Zenodo, was generated during several experiments on the micro[5] machine. The data sources are used separately or combined in the context of several machine learning approaches for tool wear monitoring.

The analysis of the performance of the models in these different approaches shows that the electrical machine currents are the most effective of the three data sources. It is worth noting that the approaches using only the acoustic emission data set did not yield reliable performances, contrary to the conclusions of other research. This could be caused by a recording problem or by a noisy environment for our experiments. The Extra-Trees and Random Forest algorithms using electrical current data sources produced the best performances and were ranked highest in the model comparison. By optimizing the hyperparameters of the Extra-Trees algorithm using the Optuna module, the model achieved an F1-score of 73%, while the expected value metric, which balances the impact of prediction errors with a manufacturing optimization perspective, achieved a score of -18% over two classes "good / worm out".

The capabilities of the in-production tool wear prediction have been demonstrated, with the ability to perform detection after each pass. The processing, which is composed by collecting, transient/stationary detection, feature extraction and prediction, takes about 245 milliseconds with a maximum of 370 milliseconds, thereby proving the possibility of performing the task during an ongoing milling operation.

As a future work, to explain the difficulty of achieving accurate results with the AE dataset, more milling tests would need to be performed to establish its relevance for use. In the studies conducted in this paper, no filtering step of the AE or accelerometer signals was performed, in order to preserve as much embedded information as possible. A specific study on signal filtering could be conducted to determine the impact on performance.

Acknowledgements This work constitutes an extended version of the conference paper [6]. The authors would like to acknowledge the contributions of the technical team, who were responsible for collecting the datasets, designing and implementing the machining plan, as well as the data analysis team, who pre-processed and explored the datasets. Gratitude is also addressed to Massimo De Santis, Edouard Goffinet, Jonathan Guerne and Célien Donzé for their contributions.

Funding The research presented in this paper was conducted with the support of the CHIST-ERA program (SOON project) and the University of Applied Sciences and Arts Western Switzerland (HES-SO).

Declarations

Ethics and consent to participate Not applicable

Consent to publish Not applicable

References

1. Dornfeld DA, Lee Y, Chang A. Monitoring of ultraprecision machining processes. Int J Adv Manuf Technol. 2003;21(8):571–8. https://doi.org/10.1007/s00170-002-1294-2.
2. Shokrani A, Dogan H, Burian D, Nwabueze TD, Kolar P, Liao Z, Sadek A, Teti R, Wang P, Pavel R, Schmitz T. Sensors for in-process and on-machine monitoring of machining operations. CIRP J Manuf Sci Technol. 2024;51:263–92. https://doi.org/10.1016/j.cirpj.2024.05.001.
3. Grimske S, Kong N, Röhlig B, Wulfsberg JP. Square foot manufacturing-a modular and mutable desktop machine tool system. Mech Based Des Struct Mach. 2014;42(3):386–97. https://doi.org/10.1080/15397734.2014.908728.
4. Wulfsberg JP, Redlich T, Kohrs P. Square foot manufacturing: a new production concept for micro manufacturing. Prod Eng Res Devel. 2010;4(1):75–83. https://doi.org/10.1007/s11740-009-0193-x.
5. Jain A, Bajpai V. Introduction to high-speed machining (HSM). In: High Speed Machining, pp. 1–25. Academic Press, Cambridge, MA 2020;. https://doi.org/10.1016/B978-0-12-815020-7.00001-1

6. Dreyer J, Carrino S, Ghorbel H, Cotofrei P. Tool Wear Monitoring Using Multi-sensor Time Series and Machine Learning. In: Moniz, N., Vale, Z., Cascalho, J., Silva, C., Sebastião, R. (eds.) Progress in Artificial Intelligence vol. 14116, pp. 497–510. Springer, Cham 2023;. https://doi.org/10.1007/978-3-031-49011-8_39 . Series Title: Lecture Notes in Computer Science
7. Dreyer J, Carrino S, Ghorbel H. Machine learning solution for machining quality prediction using acoustic emissions, accelerometers and current data. Zenodo. 2022. https://doi.org/10.5281/zenodo.6505073.
8. Dehen S, Segebade E, Gerstenmeyer M, Zanger F, Schulze V. Milling parameter and tool wear dependent surface quality in micro-milling of brass. Procedia CIRP. 2020;87:95–100. https://doi.org/10.1016/j.procir.2020.02.024.
9. Groover MP. Fundamentals of Modern Manufacturing: Materials, Processes, and Systems. 7th ed. Hoboken, NJ: John Wiley & Sons; 2020.
10. Dou J, Xu C, Jiao S, Li B, Zhang J, Xu X. An unsupervised online monitoring method for tool wear using a sparse auto-encoder. Int J Adv Manuf Technol. 2020;106(5–6):2493–507. https://doi.org/10.1007/s00170-019-04788-7.
11. Teti R, Jemielniak K, O'Donnell G, Dornfeld D. Advanced monitoring of machining operations. CIRP Ann. 2010;59(2):717–39. https://doi.org/10.1016/j.cirp.2010.05.010.
12. Mohamed A, Hassan M, M'Saoubi R, Attia H. Tool condition monitoring for high-performance machining systems-a review. Sensors. 2022;22(6):2206. https://doi.org/10.3390/s22062206.
13. Korkmaz ME, Gupta MK, Li Z, Krolczyk GM, Kuntoglu M, Binali R, Yasar N, Pimenov DY. Indirect monitoring of machining characteristics via advanced sensor systems: a critical review. Int J Adv Manuf Technol. 2022;120(11–12):7043–78. https://doi.org/10.1007/s00170-022-09286-x.
14. Diei EN, Dornfeld DA. A model of tool fracture generated acoustic emission during machining. J Eng Industr. 1987;109(3):227–33. https://doi.org/10.1115/1.3187123.
15. Pai PS, Rao PKR. Acoustic emission analysis for tool wear monitoring in face milling. Int J Prod Res. 2002;40(5):1081–93. https://doi.org/10.1080/00207540110107534.
16. Kishawy HA, Hegab H, Umer U, Mohany A. Application of acoustic emissions in machining processes: analysis and critical review. Int J Adv Manuf Technol. 2018;98(5–8):1391–407. https://doi.org/10.1007/s00170-018-2341-y.
17. Krishnakumar P, Rameshkumar K, Ramachandran KI. Acoustic emission-based tool condition classification in a precision high-speed machining of titanium alloy: a machine learning approach. Int J Comput Intell Appl. 2018;17(03):1850017. https://doi.org/10.1142/S1469026818500177.
18. Huang P-M, Lee C-H. Estimation of tool wear and surface roughness development using deep learning and sensors fusion. Sensors. 2021;21(16):5338. https://doi.org/10.3390/s21165338.
19. Harris K, Triantafyllopoulos K, Stillman E, McLeay T. A multivariate control chart for autocorrelated tool wear processes. Qual Reliab Eng Int. 2016;32(6):2093–106. https://doi.org/10.1002/qre.2032.
20. Yu J, Liang S, Tang D, Liu H. A weighted hidden Markov model approach for continuous-state tool wear monitoring and tool life prediction. Int J Adv Manuf Technol. 2017;91(1–4):201–11. https://doi.org/10.1007/s00170-016-9711-0.
21. Lauro CH, Brandão LC, Baldo D, Reis RA, Davim JP. Monitoring and processing signal applied in machining processes—a review. Measurement. 2014;58:73–86. https://doi.org/10.1016/j.measurement.2014.08.035.
22. Sayyad S, Kumar S, Bongale A, Kamat P, Patil S, Kotecha K. Data-driven remaining useful life estimation for milling process: sensors, algorithms, datasets, and future directions. IEEE Access. 2021;9:110255–86. https://doi.org/10.1109/ACCESS.2021.3101284.
23. Hanachi H, Yu W, Kim IY, Liu J, Mechefske CK. Hybrid data-driven physics-based model fusion framework for tool wear prediction. Int J Adv Manuf Technol. 2019;101(9–12):2861–72. https://doi.org/10.1007/s00170-018-3157-5.
24. Huang Q, Chunyan Qian Li C, Han Y, Zhang Y, Xie H. Tool Remaining Useful Life Prediction Method Based on Multi-Sensor Fusion under Variable Working Conditions. Machines 2022;10(10), 884
25. Ibarra-Zarate D, Alonso-Valerdi LM, Chuya-Sumba J, Velarde-Valdez S, Siller HR. Prediction of inconel 718 roughness with acoustic emission using convolutional neural network based regression. Int J Adv Manuf Technol. 2019;105(1–4):1609–21. https://doi.org/10.1007/s00170-019-04378-7.
26. Carrino S, Guerne J, Dreyer J, Ghorbel H, Schorderet A, Montavon R. Machining Quality Prediction Using Acoustic Sensors and Machine Learning. Proceedings 2020;63(1), 31 https://doi.org/10.3390/proceedings2020063031
27. Arslan M, Kamal K, Sheikh MF, Khan MA, Ratlamwala TAH, Hussain G, Alkahtani M. Tool health monitoring using airborne acoustic emission and convolutional neural networks: a deep learning approach. Appl Sci. 2021;11(6):2734. https://doi.org/10.3390/app11062734.
28. Munaro R, Attanasio A, Del Prete A. Tool wear monitoring with artificial intelligence methods: a review. J Manuf Mater Process. 2023;7(4):129. https://doi.org/10.3390/jmmp7040129.
29. Krishnakumar P, Rameshkumar K, Ramachandran KI. Tool wear condition prediction using vibration signals in high speed machining (HSM) of titanium (Ti-6Al-4V) alloy. Proc Comput Sci. 2015;50:270–5. https://doi.org/10.1016/j.procs.2015.04.049.
30. Cao X, Chen B, Yao B, Zhuang S. An intelligent milling tool wear monitoring methodology based on convolutional neural network with derived wavelet frames coefficient. Appl Sci. 2019;9(18):3912. https://doi.org/10.3390/app9183912.
31. Hahn TV, Mechefske CK. Self-supervised learning for tool wear monitoring with a disentangled-variational-autoencoder. Int J Hydromechatron. 2021;4(1):69. https://doi.org/10.1504/IJHM.2021.114174.
32. Liu H, Liu Z, Jia W, Lin X, Zhang S. A novel transformer-based neural network model for tool wear estimation. Meas Sci Technol. 2020;31(6):065106. https://doi.org/10.1088/1361-6501/ab7282.
33. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. Springer Series in Statistics. Springer, New York, NY 2009;. https://doi.org/10.1007/978-0-387-84858-7
34. Gaudette L, Japkowicz N. Evaluation Methods for Ordinal Classification. In: Gao, Y., Japkowicz, N. (eds.) Advances in Artificial Intelligence vol. 5549, pp. 207–210. Springer, Berlin, Heidelberg 2009;. https://doi.org/10.1007/978-3-642-01818-3_25 . Series Title: Lecture Notes in Computer Science

35. Pandala S R. Lazy Predict 2022;. https://github.com/shankarpandala/lazypredict
36. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: machine learning in python. J Mach Learn Res. 2011;12:2825–30.
37. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A Next-generation Hyperparameter Optimization Framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery And Data Mining 2019;. https://doi.org/10.1145/3292500.3330701