Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 133 (2025) 138–143

www.elsevier.com/locate/procedia

20th CIRP Conference on Modeling of Machining Operations

Explainable AI for tool condition monitoring using Explainable Boosting Machine

Lorenzo Colantonio[*,a], Lucas Equeter[a], Pierre Dehombreux[a], François Ducobu[a]

[a]*Machine Design and Production Engineering Unit, University of Mons, Place du parc, 20, 7000, Mons, Belgium*

* Corresponding author. *E-mail address:* lorenzo.colantonio@umons.ac.be

Abstract

Machining is one of the most critical sectors in the manufacturing industry, but the quality of the parts produced is highly dependent on the condition of the cutting tools. Poor management of tool replacement can lead to increased production costs and reduced product quality. Various methods exist to monitor tool wear and optimize their replacement. However, most of these methods rely on ”black-box” AI models, which significantly limit their practical use. In this article, a “glass-box” method called Explainable Boosting Machine (EBM) is used to monitor the degradation of cutting tools for the turning operation. This method aims to be as accurate as ”black-box” models while being fully interpretable. A comparison between the EBM method and an Artificial Neural Network (ANN) approach is presented to evaluate the performance differences between these two models. The results indicate that although the EBM model’s performance is slightly lower than the ANN’s, it remains adequate for monitoring tool wear, with an average R^2 score only 2% lower. The global and local explainability of the model is also presented. The global analysis demonstrates that the model uses coherent features for estimating tool wear, proving that it has successfully understood the wear phenomena being monitored. The local explainability highlights the contribution of each input to the tool wear estimation. These two explainability analyses show results that are consistent with the physical phenomenon of wear, for example, the model identifies that an increase in cutting force implies an increase in tool wear. This provides explainability for its use, improving trust in the monitoring method.

Keywords: Machining; Turning; Monitoring; Explainable Artificial Intelligence; Boosting Machines

1. Introduction

Machining is an essential process in manufacturing, where the condition of the tools directly influences production quality. Tool wear affects the quality of parts, manufacturing costs, and the environmental impact of the process [1]. Poor management of cutting tools and their degradation can significantly impact production. Tool failures account for almost 20% of machining failures, and the cost of tools represents 3 to 12% of total production costs [2]. Therefore, many researchers are trying to monitor tool degradation to optimize their management.

In recent years, numerous monitoring methods have appeared in the literature, mainly based on artificial intelligence (AI) techniques [3]. AI models are commonly used because many signals directly related to tool wear can be collected during machining. These signals such as, the force, vibration,etc and their correlation with the flank wear are extensively reviewed in [4]. Availability of these signals makes AI models particularly well-suited and effective for determining the tool’s condition from these signals, whether through classification or regression.

Although these methods are effective, their real-world application remains limited. Indeed most AI methods are still ”black-boxes”, making it impossible to explain their performance. Therefore, the manufacturing industry is not keen to trust models blindly as it can have a serious impact on their production and operating costs. Unfortunately, the trade-off between explainability and performance remains a persistent issue in most AI methods [5]. Fortunately, interpretability methods have been developed recently. Interpretability can be divided into global and local interpretability. A global explanation describes how the data is used to perform the model’s task, offering insights into general trends and data interactions. Local explanation clarifies why a model made a specific decision, making each individual prediction explainable. Interpretability’s method can be either specific to a model or applicable to various models [6].

10.1016/j.procir.2025.02.025

In recent years, significant research efforts have been made to develop so-called “glass-box” AI methods. These methods aim to be explainable while being as accurate as ”black-box” approaches. An example of this is the Explainable Boosting Machine (EBM), which is a type of Generalized Additive Model that uses cyclic gradient boosting to enhance interpretability and accuracy [7]. This model can provide a global and a local explanation of its results and is applicable in several applications where interpretability is critical for decision-making.

In the context of cutting tool wear monitoring, some works on interpretability exists. In 2022, Guo et al.[8] used a deep learning method with attention mechanisms to provide explainability for model predictions based on cutting signals. They observed that due to the randomness of each training session, the interpretations varied. By averaging the results, they were still able to identify certain characteristic explanations [8]. However, this fluctuation in interpretability remains a limitation. In 2024, Kumar et al. [9] used deep learning models for image segmentation on tool images, incorporating a GRAD-CAM analysis to explain the model’s areas of interest. They combined this approach with human validation to guide the model in identifying the correct wear zones. They demonstrated that their approach was more effective than without interpretability analysis and human-AI collaboration [9]. However, this method is only applicable to models that process images of cutting tools and requires a person capable of performing this operation. In general, in the field of cutting tool monitoring, either explicability is not considered, or the methods used are not sufficiently transparent to be used industrially. In addition, an analysis of a "glass-box" method such as EBM has not been applied to cutting tool monitoring yet.

Therefore, the objective and novelty of this article is to utilize an EBM method and demonstrate its value in monitoring cutting tools degradation. To evaluate the model’s performance, it is tested on a turning database. The model’s hyperparameters are fine-tuned using a grid-search algorithm. Then, the monitoring performances are compared to a ”black-box” approach, specifically an Artificial Neural Network (ANN) that was previously developed using the same dataset [10]. The model’s explainability is then demonstrated, providing both global and local explanations of the results to showcase the capabilities of these models.

2. Methodology

The aim of this article is to monitor tool wear in real-time by employing an optimized EBM model that utilizes cutting forces acquired during machining. The monitored indicator is tool wear, characterized by the size of the flank wear, denoted as VB, and defined according to the ISO 3685 standard [11]. In addition to the monitoring of the flank wear, a global and local explanation analysis is also realized (Fig. 1). To achieve this objective, several steps are required:

- **Identify the database and the inputs of the model**. The model is evaluated using the turning database described in sec. 3, with all data provided as inputs. Explainable Boosting Machines (EBMs) can automatically determine which inputs are most correlated with their output. Additionally, EBMs can identify and include interactions between data points that, when combined, have a stronger correlation with the model’s output.
- **Train and test the model**. The database includes data on the lifespan of 30 tools. To test the model’s ability to predict tool wear, it was trained on the degradation data of 24 tools (80% of the database) and tested on the remaining 6 tools (20% of the database). The testing tools were selected for their variations in cutting conditions during their lifespan. This ensures that the model encounters different cutting speeds during testing, demonstrating its ability to generalize to variable conditions over the tools’ lifespans.
- **Optimize the model, to find its optimal performance**. Several hyperparameters of the model can be adjusted to achieve better results. To find the optimal hyperparameters, a grid-search algorithm was employed, testing all combinations within a specified range. The optimization process is realised on the training dataset with a K-Fold cross validation with 5 folds. The adjusted hyperparameters and their values are listed in Table 1. An explanation and recommendations for the model parameters can be found in [12].
- **Analyse and compare the results**. The EBM results are compared with those of an ANN model also trained on the same database and developed in [10]. The performances are compared, and a discussion on the interpretability provided by the EBM is presented.

Table 1. Parameters and their associated tested range and optimal value obtained with the grid-search algorithm.

Parameters	Max leaves	Learning rate	Smoothing rounds	Inner bags	Outer bags
Tested Range	[2, 3, 4]	[0.1, 0.01, 0.001]	[150, 200, 300]	[0, 20, 30]	[0, 14, 20]
Best value	3	0.01	200	20	14

These steps are described in the remainder of this article.

3. Database

The database is obtained from experimental cylindrical turning tests (Figure 2) and presented in [10]. The tests are conducted according to the ISO 3685 standard for tool life testing [11]. They consist of machining C45 steel bars on a Weiler E35 CNC lathe with the cutting conditions presented in Table 2, without cooling or lubrication. Cutting conditions are chosen

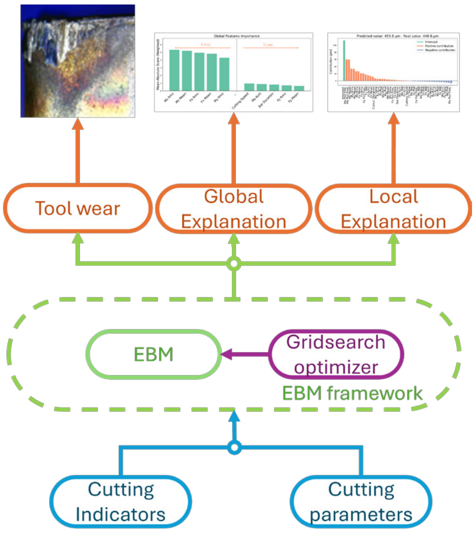

Fig. 1. EBM model framework. The model takes the cutting indicator and cutting parameters as input and output the value of the flank wear, a global explanation and a local explanation. The global explainability graph in this figure corresponds to Fig. 4, while the local explainability corresponds to Fig. 6.

to favour wear and reduce experimentation time, while respecting the cutting speeds recommended by SECO TOOLS. The C45 workpieces underwent a homogenizing heat treatment to achieve constant Vickers hardness throughout the section, ranging from 179 HV_{30} to 203 HV_{30}. The cutting equipment is selected from the SECO TOOLS catalog, with the tool reference CNMG120404-M3 TP40 and the tool holder reference DCLN L/R 2020K12-M. The tool holder sets the cutting edge angle κ at 95° and the rake angle γ at -6°. The tool's TP40 coating is the lowest grade to favor wear appearance and limit experimental time and material waste. The workpiece is machined from a diameter of 48 mm to 28 mm in 2 minutes and 30 seconds. After machining a bar, the turning operation is stopped, and the tool wear is inspected according to the ISO 3685 standard [11] using a Bysameyee EU-1000X 3 microscope. A tool is considered worn if its flank wear (VB) exceeds 300 μm. In total, 190 inspections were performed on the 30 tools, averaging 6 to 7 inspections per tool. The average lifespan of a cutting tool is 18 minutes and 30 seconds.

Table 2. Cutting conditions for the different test. One test correspond to one tool

Test N°	Material	Cutting Speed [m/min]	Feed [mm/rev]	Depth of cut [mm]
1 to 10	C45	260	0.2	1
10 to 15	C45	250	0.2	1
16	C45	240	0.2	1
17 to 20	C45	265	0.2	1
21 to 30	C45	Variable during machining: 240 to 260	0.2	1

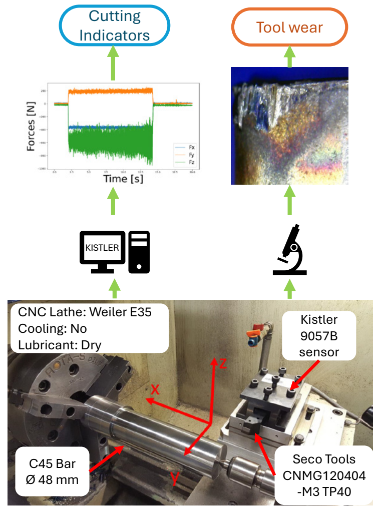

Fig. 2. Experimental setup and acquisition of the signal used in this article and [10]

During the tests, a dynanometer Kistler type 9057B measures the cutting forces (F) and torques (M) in all spatial directions (x, y and z) at a sampling rate of 10 kHz as shown in Fig. 2. In the reference frame, F_x corresponds to the feed force, F_y corresponds to the radial force and F_z corresponds to the cutting force. From the raw sensor signals, several indicators are calculated. The details of these calculations were described in previous article [13]. The statistical indicators (with their abbreviations) include: Root Mean Square value (RMS), standard deviation (std), skewness (skew), kurtosis (kurt), and mean value (mean). Frequency analysis of the signal also provides indicators. The selected ones are the dominant frequency of the power spectrum (PSD frequ.) and the amplitude of this dominant peak (PSD) [13]. In addition to these cutting indicators, some cutting parameters are also provided: the cumulative machining time (Cumul. Duration), the machining time of a bar (Bar Duration), and the cutting speed (Cutting Speed).

EBMs do not require pre-selection of features as they can dynamically determine the importance of their inputs. Therefore, all calculated indicators as well as the cutting conditions are directly provided as inputs to the EBM model. The data are not normalized because the model can handle non-normalized input data. This simplifies the explanations offered by the model, as it retains the physical meaning of the data.

4. Results

4.1. Comparison of performance

Fig. 3 compares the performance of the EBM approach developed in this article with the ANN approach developed in a previous article [10] across the different testing trajectories. Both models are trained on the same training dataset. This figure shows the target trajectory (measured during the tests) and the monitoring of the two methods. Table 3 summarizes the Mean Squared Error (MSE) and the coefficient of determination (R^2) metrics across the same testing trajectories between the EBM and ANN approaches. From the figure and table, it is observed that the performances are often similar, as illustrated by trajectories 27 and 28. However, trajectories 23 and 24 are less accurately tracked by the EBM model, especially during the mid-to-end life of the tool, where monitoring is crucial due to the higher likelihood of triggering a tool replacement. In contrast, trajectories 26 and 29 are both better tracked by the EBM approach, but it should be noted that on these trajectories, the ANN has the highest deviation compared to the other trajectories.

On average, according to the dataset and testing trajectories described in this article, the EBM approach is less effective than the ANN approach. However, despite being less effective, its performance can still be satisfactory, as the figures indicate that the general trend of degradation is accurately followed. Additionally, the table shows that the average metric values are similar, with an average R^2 difference of only 2%.

Although the EBM approach is less accurate than the ANN approach, the difference remains relatively small and is highly dependent on the trajectory. One advantage of the EBM approach is that it does not require feature selection to perform well. All database data is input into the model, which automatically ranks the most correlated features with tool wear. In contrast, ANN requires feature selection, making it more time-consuming. This is beneficial in applications where relevant data is difficult to identify and the interaction between data points is not obvious, such as machining with multiple sensors.

4.2. Global explanation

Fig. 4 illustrates the average contribution of the input features to monitor tool's degradation. The figure shows the first and last 5 features to illustrate the range of their average contribution to tool degradation. This hierarchy of importance is established by the EBM model during training, ensuring transparency about which features are most useful for predictions. Among the most significant values are: the torques around the Z axis (M_z) and the feed force (F_x RMS). In machining, it is well-established that cutting forces and torques are strongly correlated with the tool's condition, making these observations consistent with the physical phenomenon of tool wear. For instance, an increase in feed force (F_x) as a function of tool wear is a well-documented observation in the literature [14] and is automatically detected by the EBM model. Due to the database having only minor variations in cutting speed, this parameter contributes less to the estimation of the tool's condition and is among the least significant indicators. A Spearman's correlation analysis on this dataset also yields similar observations [13]. The model can also identify interactions between features. However, since the database contains data from only one sensor, interactions between signals from the same sensor do not provide any benefit. This is confirmed by the results, which show that the model did not detect any interactions within the database.

Fig. 5 shows the global importance of a single feature. Here, the global contribution of the feed force (F_X RMS) is selected to be shown. This feature is chosen due to its strong correlation with the tool's condition and its straightforward physical interpretation. The score (y-axis) is expressed in μm. This score represents the contribution of F_X RMS to the tool's wear estimation for different values (x-axis). The figure illustrates that depending on the value of the feed force, its contribution to the wear estimation changes. Higher cutting force values correspond to a greater impact on wear estimation, aligning with the physical phenomenon where a worn tool requires higher cutting force [14]. For example, when the F_X RMS value is 375 N, its contribution to the tool's condition estimation is approximately 15 μm. The model uses this value, along with all other feature values, to estimate the state of the tool. This figure therefore provides an overall summary of the contribution of F_X RMS for different input values. It should be noted that the trend observed in Fig.5 is generally seen in other features as well. Details about local explanations and wear estimation are described in section 4.3. Around the value (referred to as the score), there is a confidence interval indicating the variance of this variable. The density graph below shows the data distribution for that particular features. The height of the bar indicates the density of this value in the database (DB). This visualization is important as a model can perform very differently with large and small samples [12]. It is therefore important to know the density of the data as larger samples tend to provide more statistically significant interpretations.

4.3. Local explanation

In addition to global interpretation, local explanation is achievable with EBMs. Instead of looking at trends, the focus is on examining how a particular prediction is made. Fig. 6 represents the local explanation of the last wear prediction of trajectory 26 (Fig. 3). On this figure, the y-axis represents the feature's contribution expressed in μm and the x-axis corresponds to the feature name. In this example, the actual measured wear is 288.4 μm, and the model's prediction is 287.4 μm. To obtain this wear estimation, the model first provides the 'intercept' value (light green bar in Fig. 6), which is a constant determined during the training phase. Each input feature then adds a positive or negative contribution to this value. A positive contribution indicates that the tool is more worn compared to the average value (intercept), while a negative contribution indicates less wear. The sum of all contributions represents the wear estimation of the model expressed in μm. In this example, with a worn tool, features are expected to indicate high wear to the model, which is observed here. Indeed, most features have a positive contribution to the tool wear estimation. There are still some features that negatively influence the estimation, but their number and contribution are less significant compared to the others

5. Conclusion

This article compare the performance of an explainable "glass-box" EBM model with a "black-box" ANN model to

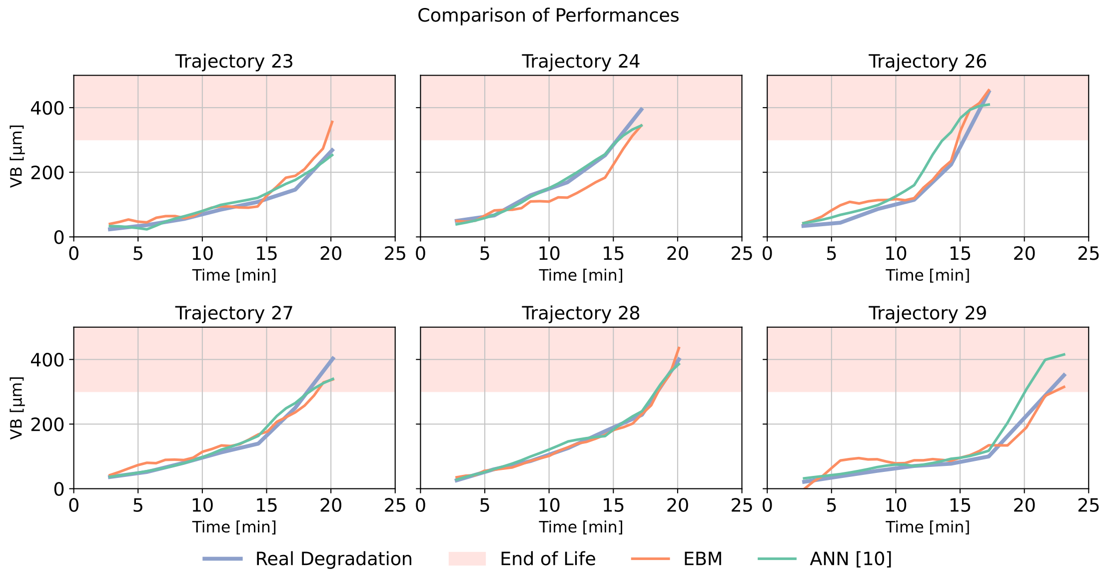

Fig. 3. Comparison of performances of the EBM model and the ANN model [10] trained and tested on the same dataset

Table 3. Comparison of performance metrics from the trajectories presented in Fig. 3

	Traj. 23		Traj. 24		Traj. 26		Traj. 27		Traj. 28		Traj. 29		Average	
	MSE	R^2	MSE	R^2	MSE	R^2	MSE	R^2	MSE	R^2	MSE	R^2	MSE	R^2
EBM	768	0.82	1591	0.84	911	0.93	569	0.94	96	0.99	763	0.88	783.00	0.90
ANN	190	0.95	194	0.98	2412	0.83	380	0.96	102	0.99	1082	0.83	726.67	0.92

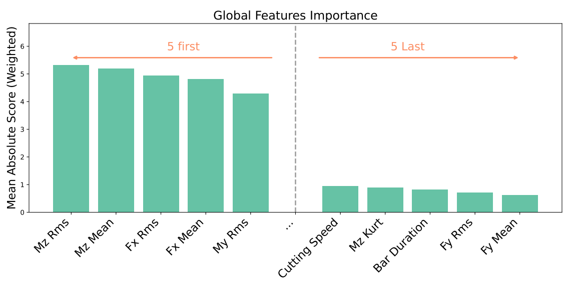

Fig. 4. Global explanation showing the 5 top and 5 last average contribution of all the input of the model

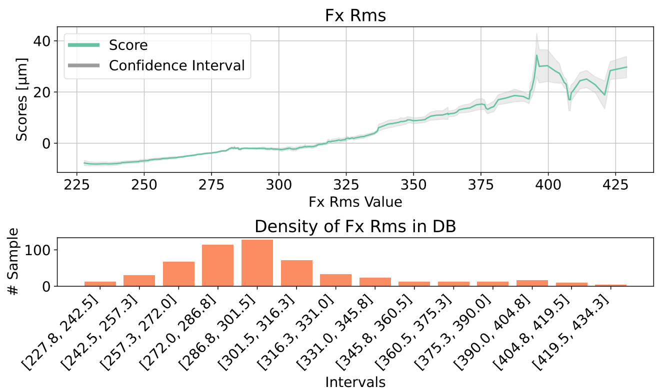

Fig. 5. Global explanation of the selected features F_x RMS and its contribution to estimation of the tool wear with respect to the input value. The graph below shows density of each interval inside the database

monitor the degradation of cutting tools from the cutting force signal. The comparison of performance shows that:

- On average, the EBM is less effective than the ANN approach by around 2% on the R^2 value. However, the monitoring results are still satisfactory to follow the tool wear.
- Feature selection is not necessary for the EBM model as it automatically select the most useful features. Furthermore, the model can work with non-normalized data which helps to keep the physical meaning of its explanations.

In addition to the performance, the EBM model provide, global and local explanation:

- The global explanation shows that F_x and M_z are most correlated with tool wear, aligning with the fact that increased tool wear requires greater cutting effort. This explanation confirm the model's understanding of tool degradation.
- As expected, the model does not detect any interaction between the features, as there is only one cutting force sensor in the database.
- The local explanation details how a specific prediction is made, showing the contributions of different features.

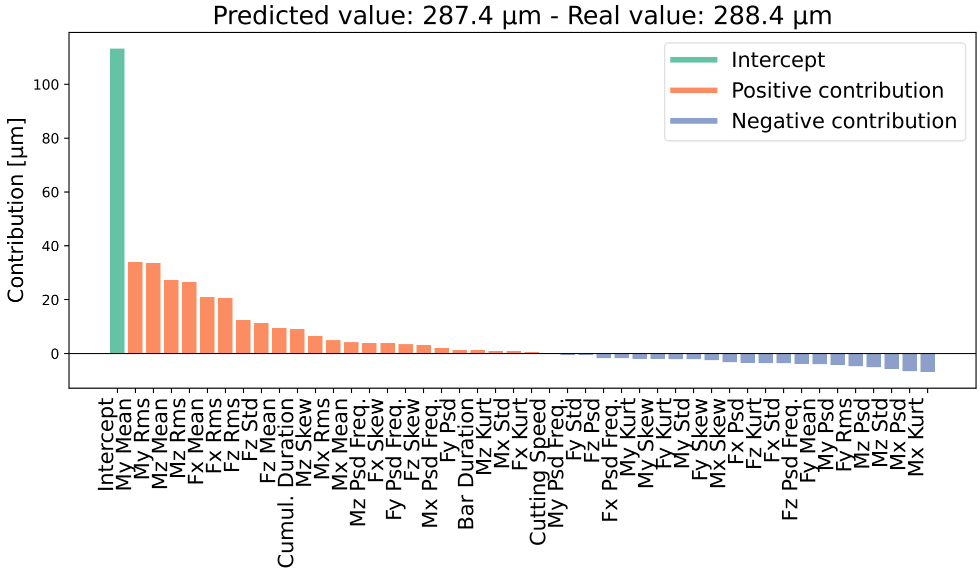

Fig. 6. Local explanation of one prediction of the EBM model showing the contribution of each input features to estimate the tool wear for the penultimate wear measurement of trajectory 29 (Fig.3)

This demonstrates the model's ability to detect higher or lower wear than average.

With this approach, an industrial user can ensure that the model is capable not only of accurately tracking tool degradation but also of correctly identifying the main features for estimating tool wear. This allows for trust in the models for making tool replacement decisions, thereby optimizing the economic and environmental management of cutting tools. Future work will take advantage of the flexibility of the EBM model to modify undesirable effects and apply the model to a wider range of cutting conditions and signals.

References

[1] C. Tian, G. Zhou, J. Zhang, and C. Zhang, 'Optimization of cutting parameters considering tool wear conditions in low-carbon manufacturing environment', J. Clean. Prod., vol. 226, pp. 706–719, Jul. 2019, doi: 10.1016/j.jclepro.2019.04.113.

[2] Baig RU, Javed S, Khaisar M, Shakoor M and Raja P. Development of an ANN model for prediction of tool wear in turning EN9 and EN24 steel alloy. Advances in Mechanical Engineering. 2021;13(6). doi: 10.1177/16878140211026720.

[3] L. Colantonio, L. Equeter, P. Dehombreux, and F. Ducobu, 'A Systematic Literature Review of Cutting Tool Wear Monitoring in Turning by Using Artificial Intelligence Techniques', Machines, vol. 9, no. 12, p. 351, Dec. 2021, doi: 10.3390/machines9120351

[4] A. Siddhpura and R. Paurobally, 'A review of flank wear prediction methods for tool condition monitoring in a turning process', Int. J. Adv. Manuf. Technol., vol. 65, no. 1–4, pp. 371–393, 2013, doi: 10.1007/s00170-012-4177-1.

[5] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, 'Explainable AI: A Review of Machine Learning Interpretability Methods', Entropy, vol. 23, no. 1, p. 18, Dec. 2020, doi: 10.3390/e23010018.

[6] Z. Chen, F. Xiao, F. Guo, and J. Yan, 'Interpretable machine learning for building energy management: A state-of-the-art review', Adv. Appl. Energy, vol. 9, p. 100123, Feb. 2023, doi: 10.1016/j.adapen.2023.100123.

[7] Y. Lou, R. Caruana, J. Gehrke, and G. Hooker, 'Accurate intelligible models with pairwise interactions', in Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, Chicago Illinois USA: ACM, Aug. 2013, pp. 623–631. doi: 10.1145/2487575.2487579.

[8] H. Guo, Y. Zhang, and K. Zhu, 'Interpretable deep learning approach for tool wear monitoring in high-speed milling', Computers in Industry, Vol. 138, 2022,P. 103638,doi: 10.1016/j.compind.2022.103638.

[9] A. S. Kumar, A. Agarwal, V. G. Jansari, K. A. Desai, C. Chattopadhyay, and L. Mears, 'HG-XAI: human-guided tool wear identification approach through augmentation of explainable artificial intelligence with machine vision', J. Intell. Manuf., Aug. 2024, doi: 10.1007/s10845-024-02476-2.

[10] L. Colantonio, L. Equeter, P. Dehombreux, and F. Ducobu, Comparison of artificial intelligence techniques for cutting tool condition monitoring, Materials Research Proceedings, Vol. 41, pp 1962-1971, 2024 doi: 10.21741/9781644903131-217.

[11] ISO 3685—Tool Life Testing with Single-Point Turning Tools. 1993. Available online: https://www.iso.org/fr/standard/9151.html (accessed on 15 november 2024).

[12] H. Nori, S. Jenkins, P. Koch, and R. Caruana, 'InterpretML: A Unified Framework for Machine Learning Interpretability', Sep. 19, 2019, arXiv: arXiv:1909.09223. Accessed: Nov. 08, 2024. [Online]. Available: http://arxiv.org/abs/1909.09223

[13] Colantonio L, Equeter L, Dehombreux P, Ducobu F. Confidence Interval Estimation for Cutting Tool Wear Prediction in Turning Using Bootstrap-Based Artificial Neural Networks. Sensors. 2024; 24(11):3432. https://doi.org/10.3390/s24113432

[14] Chuangwen X, Jianming D, Yuzhen C, Huaiyuan L, Zhicheng S, Jing X. The relationships between cutting parameters, tool wear, cutting force and vibration. Advances in Mechanical Engineering. 2018;10(1). doi:10.1177/1687814017750434