

Automatic tool wear inspection by cascading sensor and image data

Robbert Verbeke¹, Lars De Pauw², Fabian Fingerhut¹, Tom Jacobs³, Toon Goedemé², and Elena Tsviporkova¹

¹ EluciDATA Lab Sirris, Brussels, Belgium robbert.verbeke@sirris.be

² EAVISE KU Leuven, Sint-Katelijne Waver, Belgium lars.depauw@kuleuven.be

³ Precision Manufacturing Sirris, Brussels, Belgium

Abstract. This paper explores the possibility to apply data-driven methods to improve the replacement strategy of worn cutting tools used in milling industry. While the data was generated in a controlled environment, the conditions under which the data were generated varied to make them more realistic. Both indirect (sensor data) and direct (images) data was captured and individually modelled. We propose a cascading approach, combining both modalities in a sequential way, and show that this methodology leads to very accurate tool replacement strategy while keeping production efficiency high.

1 Introduction

In many manufacturing settings, the usage of cutting tools used during, e.g., milling or drilling, constitutes a non-negligible economic cost. They are typically replaced after manual inspection of an operator who makes a decision on the tool status. This monitoring is time consuming and subjective. Alternatively, the tool is replaced automatically after a fixed time of operating. Since a faulty tool can lead to lower quality of the produced part, or even severe damage when the tool breaks, this automatic replacement time is set very conservatively. This leads to an inefficient use of the tools and a waste of precious materials.

Continuous collection of diverse measurements of the tool status allows to use a data-driven approach to assess whether the tool should be replaced or not and thus can lead to an improved replacement strategy that is time and resource efficient. However, such continuous monitoring comes with a cost, not only in terms of investment related to the installation of the measurement infrastructure, but also in terms of potentially reduced production efficiency caused by obtrusive measurements.

In this paper, a two-phase approach is presented, using both sensor and image data, to accurately monitor the tool status. The two modalities work in a complementary way and we show that combining the two in a cascading fashion leads to a better performance than any of the individual models. In addition, the two-phase paradigm allows optimisation of the measurement workflow in such way that the production process is minimally affected by the executed measurements. The data was produced in a controlled real-life environment using varying machine settings.

2 Related Work

In this section, we give a non-exhaustive overview of some related works in the field of tool wear monitoring.

Kim et al. [1] and Ayman et al. [2] both give an exhaustive overview of, among others, the research done on the topic tool monitoring and tool wear prediction.

Benkdedjouh et al. [3] use Support Vector Machines (SVMs), taking cutting forces, vibration, and acoustic emission as input, to predict the tool wear and remaining useful life. ANNs using only image data as input have been used to predict the tool wear as well, achieving good results as shown in [4]. However, other authors (see e.g. [5]) found that Random Forests, using the same input features, can outperform Artificial Neural Networks (ANNs) and SVMs for tool wear prediction. More recently, deep learning was shown to be well suited for tool wear prediction as well [6].

In many application contexts, the exact value of the wear is less important than the tool condition (e.g., highly worn or not). Also for this purpose, the above methods can be applied, either on image data [7,8] or on sensor measurements [9,10,11,12] as input.

A main drawback in many of the above studies is that the data was generated under simplified controlled environments, typically using one fixed value for the machine settings like the cutting speed and feed rate. Transfer learning has been used to generalize better to different operating conditions [13], however the step to an integration of the approach in an industrial setting remains large. In real-life settings the machine settings may vary continuously, data will be noisy, and reliable labels will be very rare.

Due to the latter reason, unsupervised [14,15,16] and semi-supervised [17] anomaly detection methods have been applied for tool wear monitoring, instead. Despite the mentioned above challenges, promising results for unsupervised tool wear monitoring during real life manufacturing operations have been achieved [18].

In this paper, the data used is generated in a controlled setting, but the machine settings are varied, mimicking real industrial situations, to study their impact. Both sensor and image data are captured and used as input to separate models.

3 Materials and Methods

3.1 Use Case and Data

Use Case Scenario The use case considered in this paper is on the automatic monitoring of tool wear and recommending the optimal replacement time. This is important in manufacturing settings where tools are typically replaced in a less-than-optimal way, leading an inefficient use of materials. Rather than trying to predict the exact tool wear through a regression model, the aim is to come up with an accurate classification task that predicts whether the tool is

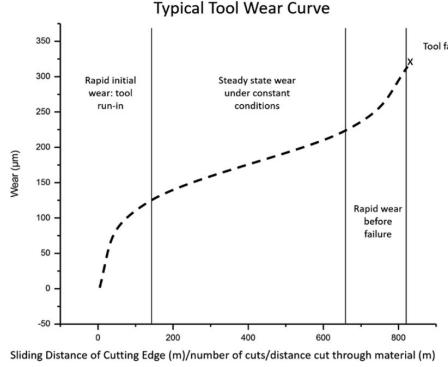


Fig. 1: The typical evolution of tool wear. Figure taken from Alhadef et al. 2019 [19].

operating under *safe* or *unsafe* conditions. Unsafe conditions are characterized by a rapid increase in tool wear, compared to a more steady increase during safe conditions. This is illustrated in Figure 1. In practice, this is treated as equivalent to predicting *low* or *high* tool wear based on some threshold due to convenience. However, these are not necessarily the same and the differences are discussed later on.

Data generation We collected 11 different data sets, using a different tool for each set and with different cutting speed V_c , feed rate f_z , spindle speed n , radial depth a_e , and axial depth of the cut a_p . Each set consists of a certain number of usages, all done with the same tool. The aim was to produce 100 usages per set. However, when the tool wear evolved too slowly (or too quickly), more (or less) usages were performed. These are all specified in Table 1. All tests were done on the same type of material, namely CK45 (steel). More information about the data generation process, as well as a download link for the full dataset used in this paper, can be found in [20].

As can be seen, sets 4 to 11 have the same V_c and n , while the other machine settings may vary. Sets 2 and 3 on the other hand, have different values for V_c and n , and the values for these parameters for set 1 are even unknown. This is the reason why in the (sensor) analysis, the first 3 sets will be analysed separately from the last 8.

Sensor Data 5 different sensor values are captured:

1. Acoustic emission (Vallen VS30-Sic-V2: 25-80 kHz sensor),
2. Vibration (multiple PCB 333B40 accelerometers),
3. Force in x, y, and z direction (Kistler 9257B and 9255C dynamometers).

All sensor data are gathered using a DAQ system from National Instruments in a synchronized way, with a sampling frequency of 1626 Hz. An overview of the set-up is shown in Figure 2a.

Set	V_c [m/min]	n [rev/min]	f_z [mm/rev]	V_f [mm/min]	a_e [mm]	a_p [mm]	n_{usages}
1	Unknown	Unknown	Unknown	Unknown	1.0	1.0	91
2	120	2547	0.08	203	1.0	1.0	82
3	150	3184	0.05	159	1.0	1.0	92
4	174	3705	0.05	185	1.0	1.0	75
5	174	3705	0.04	148	1.0	1.0	56
6	174	3705	0.04	148	1.0	1.0	139
7	174	3705	0.045	170	1.0	1.0	98
8	174	3705	0.048	178	1.0	1.0	97
9	174	3705	0.048	178	1.0	0.5	98
10	174	3705	0.05	185	1.0	0.5	98
11	174	3705	0.043	159	1.0	1.0	97

Table 1: Machine settings and number of usages for each set.

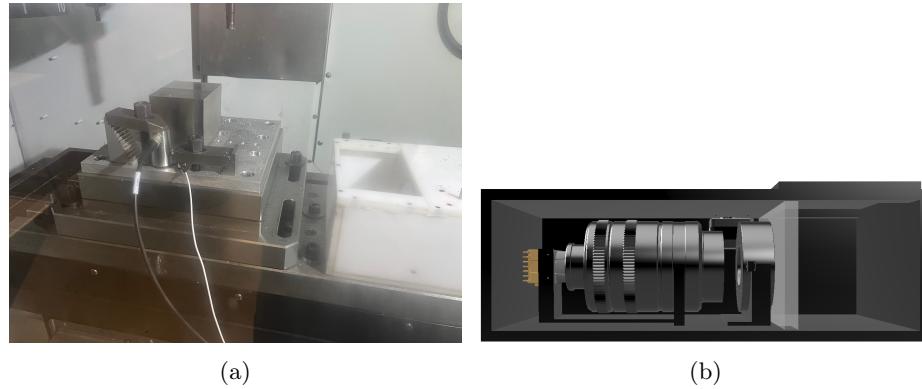


Fig. 2: (a) Sensor data capturing set-up. The acoustic sensor is connected with the black cable, the accelormeter with the white cable, and the plate with the holes is the dynamometer. (b) Cross-section of Camera box with used vision setup.

Image Data Image data acquisition on a scale of only few micrometers is prone to errors caused by small movements of the camera w.r.t. to the tool. This is addressed by creating a fully enclosed box which is bolted to the machine table of the milling machine. This box can be seen on the right side of Figure 2a. A firm connection to the table ensures the focal distance is respected on all images and generates stable images throughout the dataset. Figure 2b provides a visual of the used vision setup. Left to right are camera, lens, ring light and enclosed part for tool to be presented to the camera. Technical details for these parts are discussed next. Most left part is 20.5 MP Alvium 1800 U-2050 rgb camera. This is configured using the manufacturers framework Vimba to apply an exposure time of 1500ns and a gain of 15. These settings are adjusted together with the aperture on our macro lens to provide well lit images with a large enough depth of field to capture the worn edge on the tool. The used lens is a CA-LMHE0510 produced by Keyence. This lens is manually adjusted to a field of view of 17.8 x 14.3 resulting in a pixel resolution of 3.2 micron per pixel. A LED ring light (CA-DRW13M by Keyence) is positioned in front of this lens as close to the insert as possible to provide multi-angle lighting to the point of interest. A plexi screen is added in front of the light to protect the electronical systems from metal chips and cooling fluids coming from the milling process.

The results from this setup are shown in Figure 3. The left and right images on the top represent the first and second set of data respectively. Underneath them are displayed their cropped counterparts that are later used for training.

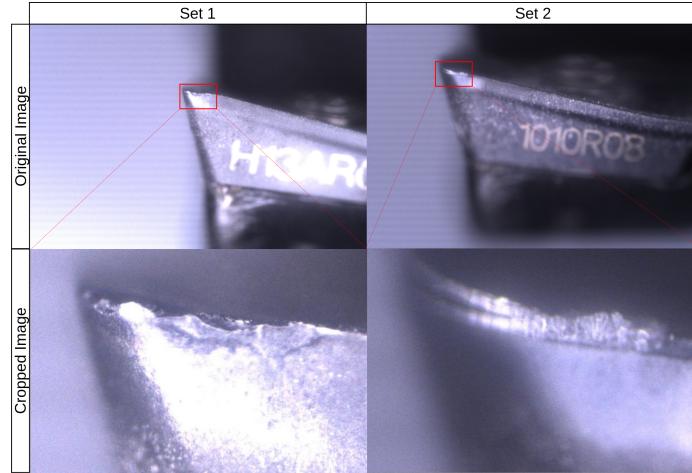


Fig. 3: Example of raw image captured with setup.

Tool wear data The values for the tool wear are measured by a domain expert from the captured images. Two types of tool wear are identified: flank wear and

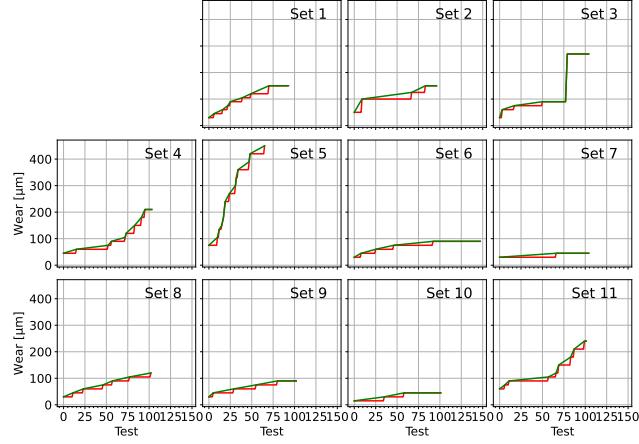


Fig. 4: Evolution of the annotated flank wear (red) and the corresponding interpolation (green) for all sets.

adhesive wear. The flank wear increases monotonically while the adhesive wear can also decrease as the particles sticking to the tool get removed.

The ground truth wear values are obtained by human assessment of the tool pictures taken. The resolution of these measurements is $15 \mu\text{m}$ which results in plateaus in the tool wear evolution. To make the labels more reliable, we linearly interpolate the *flank* tool wear during the plateaus from the first usage of the plateau to the moment the measured wear increases. No interpolation is done after the last measured increase.

We assume that adhesive wear will have very little effect on the sensor measurements, since this adhesion to the tool is typically very short lived. Because of this, wherever adhesive wear was measured, the labels were replaced by an estimate of the flank wear by linearly interpolating using the closest measured flank wear usages.

Looking at Figure 4, it seems that the rapid increase in wear occurs around $120 \mu\text{m}$ for sets 4 and 11. Set 5 seemingly was always in the unsafe zone, apart from the first few usages, reaching very quickly much higher wear values than the other sets. The tool wear in sets 1 and 2 seems to still increase at a safe, gradual rate, even after reaching $120 \mu\text{m}$. Set 3 has a sudden increase in wear from 90 to $270 \mu\text{m}$, but the wear rate gets stable after that. The other sets show a stable evolution and do not exceed $120 \mu\text{m}$.

Given all this, it seems that the safe zone depends on many different factors, such as the cutting speed and the current amount of wear. Furthermore, given that the sudden increase in wear in set 3 did not lead to any big increase in wear afterwards, it seems that the wear geometry plays a big role as well.

Despite this, for the sake of uniformity, we chose to use a threshold of $120 \mu\text{m}$ to label the usages as in the safe or unsafe zone. This seems valid at least for sets 4 to 11, but the above considerations will have to be taken into account while analyzing the results.

3.2 Tool wear inspection via a cascade of sensor and image models

This study presents a multi-modal data analysis framework for tool-wear monitoring. Namely, two different models for tool wear monitoring are used, originating from different modalities, which are deployed in a complementary way. It aims to shed light on the potential and limitations of the different data modalities, e.g. different types of sensor data, that are usually employed to monitor the wear rate of the cutting tools. The outcome of such analyses can help choosing the suitable data modality or set of data modalities for a given machine setting and cutting tool.

Cascade model To combine the two methodologies, a 2-phase cascade approach is proposed:

1. The sensor-based model decides whether the tool is in the safe or unsafe regime,
2. If the model output is unsafe, the image modality is switched on,
3. The image-based model reevaluates the safe or unsafe condition.

There are multiple advantages to this approach:

- It becomes a multi-modal approach, making use of all the available data
- The image capturing modality, which is time-consuming, is only switched on when it is deemed necessary.

Sensor data preprocessing Before any data exploration or modelling is done on the sensor data, some preprocessing steps are performed.

As a first step, the *active* part of the signal, corresponding to the times where the tool is in contact with the work piece, is identified. For this, we propose a very simple change point detection method. First, the standard deviation σ is computed over a rolling window of 200 timestamps. Next, all the timestamps where σ is above 1.5 times the median of σ are identified. The first and last of these timesteps are set to respectively be the start and end of the active part. This value of 1.5 was chosen based on a visual inspections that verified the accurate identification of start and end. The active part is typically about 30 seconds long.

Sometimes, drift can be present in the sensor measurements. This is purely an artifact and not a physical property. To account for this, the mean of the signals is calculated over a rolling window of 1000 timestamps and subsequently subtracted from the signal.

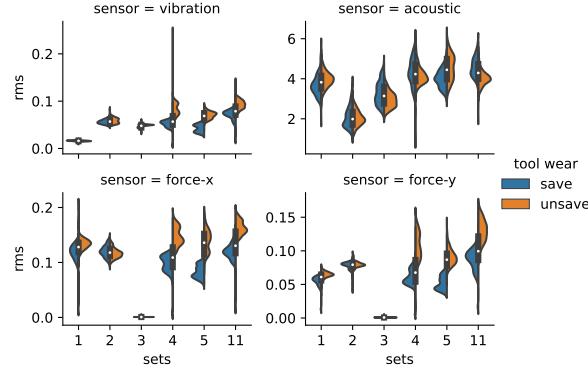


Fig. 5: Exploratory analysis of sensor data segments with respect to tool wear. Most statistical parameters are indicative of low and high tool wear.

As a final step, the active part, which has variable length, is split into overlapping segments of fixed length N . This has the advantage that the input samples to the model have a much shorter and fixed length. This also assumes that the signatures related to wear are present over this shorter time span. Another advantage is that the number of input samples available gets greatly increased. After experimenting with varying window sizes, $N = 2000$ was chosen empirically since basic characteristics, like the maximum amplitude, are more stable when looking at consecutive segments compared to using shorter segments. Given the sampling frequency, this corresponds to segments of 1.2 seconds. The stride was chosen to be 500 timestamps, resulting in overlapping segments, which further increases the number of input samples.

The tool wear labels of the segments are set to the value of the corresponding usage.

The above preprocessing results in 7400 to 15600 segments, depending on the set. The variation can mostly be explained by the different number of usages per set, but small differences in the length of the active parts also have an impact.

Sensor data exploration and profiling After an initial examination of the data, it was found that the force in the z-dimension often was a constant value. This feature was thus removed from any further analysis.

Subsequently, it was checked whether the extracted segments can be indicative of tool wear. Fig. 5 illustrates the root mean square (rms) value that was extracted from the identified segments, where the rms value for a given measurement vector \mathbf{x} is calculated as $rms = \sqrt{1/n \times \sum_{i=1}^n x_i}$ with $\mathbf{x} = [x_1, \dots, x_n]^T$. A binary value for safe and unsafe tool wear was chosen based on the previously defined tool wear threshold of $120 \mu m$ which was shown to be the point at which tool wear rapidly increases. We excluded two outlying instances, where the rms value for the vibration exceeded 2.

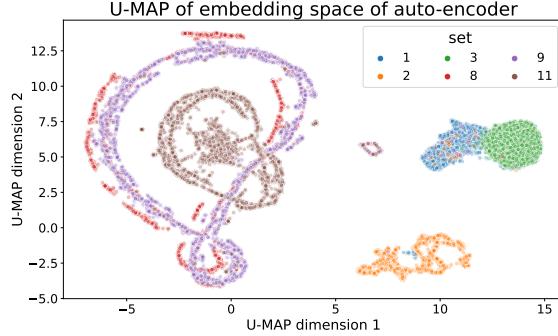


Fig. 6: The machine settings have a clear effect embedding of the auto-encoder. Set 1, 2 and 3 differ from the other sets with respect to V_c and n and are therefore embedded differently compared to sets 8, 9 and 11.

It can be observed that (1) within each setting, the low and high tool-wear segments differ in their rms distribution and that (2) the setting itself has an influence on the distribution. For instance, the rms value of the y-force differs substantially between safe and unsafe tool wear segments in set 11. The rms value for unsafe tool wear examples of set 1 is however typically lower than the rms values for safe tool-wear examples of set 11. Similar observations can be made with other statistical parameters (e.g. standard deviation, absolute maximum). We conclude, that with respect to the rms value (but also with respect to other statistical parameters that were not shown in Fig. 4) the segments are separable per set, but not separable over all sets. This hold with the exception of the acoustic emission which displays little differences between save and unsave tool-wear with regards to the rms value.

The effect of different tool settings cannot only be observed in the statistical metrics, but in data-driven models that take the given segments as input as well. In order to prove this point, an auto-encoder is trained, based on the *Deep clustering with Convolutional Autoencoders* (DCEC) network [21]. Fig. 6 displays the U-MAP [22] representation of the 10 dimensional embedding space of the auto-encoder trained on sets 4, 5, 6 and 10 and taking set 7 as validation set. It becomes apparent that the segments of sets 1, 2 and 3 are embedded differently from the other sets, following their different machine settings that they were run under (see Table 1).

Sensor data classification Using the sensor data, a model is trained to identify whether the tool is in the unsafe zone, i.e., where the wear is increasing much more rapidly.

Our classification model is a deep neural network architecture, built using 1D convolutional neural networks (CNNs) and taking the time series segments

as multi-channel input. The 1D CNNs were used to learn temporal patterns of the time-series sensor data which are essential for efficient tool-wear monitoring.

Given the meaningless values of the z-force measurements and the low separability based on tool wear values of the acoustic emission, only the vibration, x-force, and y-force were used as an input to the model. Each input sample will thus have the shape 2000×3 .

There are 5 convolutional blocks (consisting of a 1D CNN layer, a max-pooling layer, and a batch-normalization layer) using 8, 16, 32, 64, and 128 number of filters. All CNN layers have a kernel size of 5. After the last convolutional block, the output is flattened and followed by 5 fully connected (FC) layers (of size 64, 32, 16, 8, and 2). All CNN and FC layers use a ReLU activation function, except for the last one, which uses SoftMax. The model was trained over 200 epochs, with an initial learning rate of 10^{-5} . This was reduced when a plateau was reached for the validation accuracy, with a minimum learning rate of 10^{-11} .

Image Data Analysis Image data analysis is performed using a 2D CNN classifier trained to predict the safety of the tool. This safety is defined the same way as for the sensor classifier with a threshold at $120\mu\text{m}$. Resulting in a label 0 for safe, 1 for unsafe for the binary classification.

We selected the Resnet [23] architecture out of many different possible classification networks due to proven performance on classification tasks. On this backbone, the classifier has been modified from a 1000 class classifier to a 1 class binary classifier. To interpret the binary classification output value, a threshold needs be set between 0 and 1 for safe or unsafe classification, this threshold is chosen to be 0.5 by default. Pretrained weights trained on Imagenet have been loaded into the model before training.

Training was done on the same training sets used for training sensor classifier to prevent data leakage between the two approaches. Parameters for training are left to default except for the learning learning rate which was set to 3×10^{-3} and batch size set to 32. The Resnet family consists of multiple layer depths for the network. Resnet101 was empirically selected as best performing for this task.

Training runs were stopped by an early stopping algorithm to prevent overfitting by ending the training when validation loss stopped decreasing. To further optimize training, the learning rate was reduced by 1/2 after a plateau of 10 epochs was reached.

The imbalancedness of the dataset poses difficulties where the network would predict the same class for all samples presented. The dataset was balanced by random sampling from both classes (safe/unsafe) which results in a 50/50 distribution between the two classes. The test dataset was left imbalanced to evaluate on realistic data.

Training strategy To train the model using sensor data, great care has to be taken to prevent data leakage. The sensor data coming from each set can be seen as one continuous time series, interrupted when a picture is taken between each

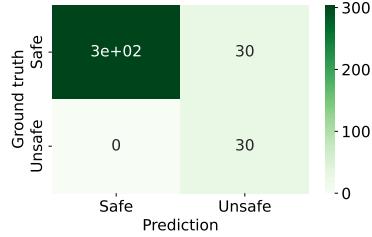


Fig. 7: Confusion matrix for the sensor-based model.

usage. As such, different usages in the same set should not be randomly divided between training and test set. It is indeed to be expected that subsequent usages with the same tool will result in very similar time series. In other words, these samples will be highly correlated. Similarly, as the wear increases gradually, the images can not be considered as independent samples either.

To avoid any data leakage, the split is done on the set level. All usages from sets 4, 5, 6, and 10 were used for training the models. These sets were chosen because they encompass the whole wear range of all sets. The decision of including four sets in the training is based on the fact that this is half of the number of sets with similar machine setting.

The remaining sets are used to evaluate the generalizability of the model. The first 3 sets were not included on purpose in the training set, to inspect how the results differ as the machine settings vary.

4 Results and Discussion

4.1 Sensor-based model

This section reports the results of the sensor data classification model, as described in section 3.2. The classification score is computed for all the segments of the sets, which are not included in the training set. However, we are interested mostly in the prediction on the usage level, not for individual segments. Therefore, the prediction scores are averaged for the segments of the same usage. The results are depicted in Figure 7. As can be observed, the majority (90 %) of the safe usages are labeled correctly, while *all* unsafe usages are identified as such correctly. In other words, there are no false negatives, which are much more catastrophic than false positives. The performance metrics are reported in the first column of Table 2.

4.2 Image-based model

The vision-based classifier was trained on the same data as the sensor-based was trained on. The results are discussed here, shown in Figure 8 and summarized

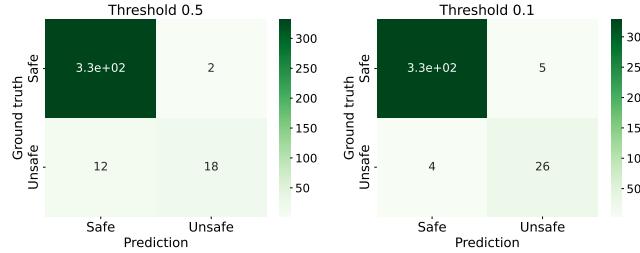


Fig. 8: Confusion matrix for the computer vision model with a decision threshold for the output of the model of 0.5 (left) and 0.1 (right).

in Table 2. When using the default threshold of 0.5 to decide which class the model outputs, the model reaches a higher precision than the sensor-based model (90% vs. 50%) but much lower recall (60% vs. 100%). By lowering this decision threshold, the recall can be increased, but with a lower precision as trade-off. When using a decision threshold of 0.1, the precision slightly decreases, but all other metrics increase. Several thresholds were explored and the value was chosen based on the best F1-score.

4.3 Cascade approach

When combining two modalities, using the best threshold for the vision-based model, in a cascade approach where the sensor network provides a first selection between safe and unsafe, the total accuracy is boosted further to 99%, with high recall, precision, and F1-score as well, as seen in the final column of Table 2.



Fig. 9: Confusion matrix for the 2-step classification model

An important question is where the misclassified usages are occurring. This is shown in Figure 10. As it can be seen, set 11 is the only set that gets into the unstable regime. For all other sets, the usages are correctly labeled as safe. For set 11, the incorrectly labeled usages are close to where the tool transitions

	Sensor model	Image model (0.5)	Image model (0.1)	Cascading model
Accuracy	92%	96%	98%	99%
Precision	50%	90%	84%	86%
Recall	100%	60%	87%	87%
F1-score	67%	72%	85%	91%

Table 2: Performance metrics of the sensor-based model, the image-based model, using both a decision threshold of 0.5 and 0.1, and the cascading model.

from safe to unsafe, which is naturally the most noisy region for classification. In other words, none of the misclassifications are catastrophic with respect to wear.

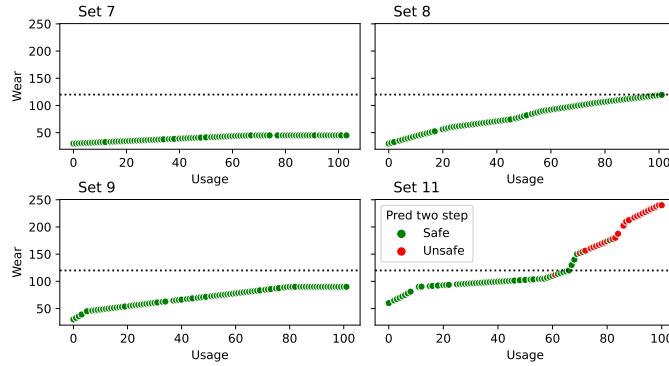


Fig. 10: Interpolated tool wear evolution of the four unseen sets with the prediction of the cascading model for each usage.

4.4 Generalization to different cutting speeds

Figure 11 shows the predictions of the sensor model for sets 1, 2, and 3. These sets used a different cutting speed, so it is not clear how well the model would generalize to these sets. As can be observed, the model predicts safe conditions for all usages of sets 2 and 3, and only just starts to predict unsafe conditions for the final usages of set 1. This is despite the fact that all sets contain usages well past the threshold of $120 \mu m$. However, as discussed in Section 3.1, it is not clear from the tool wear evolution that these sets have entered an unsafe regime. In this sense, the prediction of the model are accurate.

This raises the question what determines the transition from safe to unsafe conditions. From the current available sets, it seems that the threshold for this transition is dependent on the cutting speed, but too few sets where the transition happens are available to make this more concrete.

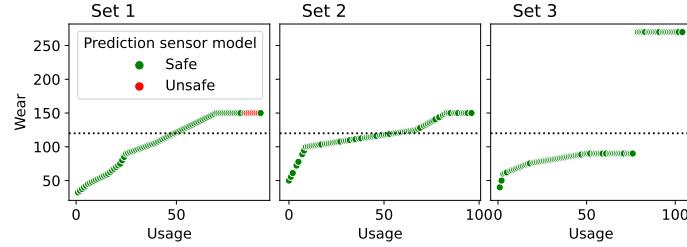


Fig. 11: Evolution of the interpolated tool wear and the prediction of the sensor model for Sets 1-3, which have different cutting speeds.

Unfortunately, with the current data it can not be evaluated whether the model would accurately pick up the transition also for sets with different cutting speeds.

The performance of the image model on Sets 1-3 is also evaluated. This is shown in Figure 12. The accuracy is 69% and the F1-score 49%. Varying the decision threshold did not significantly improve the results for these sets.

The model clearly performs much better than the sensor-based model on the same sets, which can be explained by the fact that the resulting images are not (directly) influenced by the cutting speed, unlike the sensor measurements. On the other hand, the performance is worse on these sets than of the same model on the sets previously discussed in 4.2. This implies that the resulting wear looks differently when the cutting speed is changed, making it harder for the image model to determine whether the wear is above the threshold of $120 \mu m$.

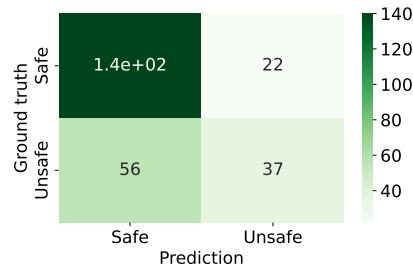


Fig. 12: Confusion matrix for the predictions of the image model on Set 1-3.

Note that, given the bad performance of the sensor model here, using the 2-phase, cascading approach would lead to bad performance as well.

5 Conclusion

In this paper, a measuring set-up and methodology was outlined to improve the replacement strategy for cutting tools. The approach involves predicting when the tool wear starts increasing more rapidly, leading to unstable behaviour. The data is generated in several sets, using different tools and machine settings. Both sensor and image data are combined to achieve a model that can very accurately identify when this rapid increase in tool wear starts, as long as the cutting speed is kept constant. For the sets where this is not the case, the results seem much worse. However, this is likely due to a mislabeling of safe or unsafe conditions, since this was purely based on a threshold on the wear. Unfortunately, only a few of the sets have different cutting speeds, so this could not be clearly validated.

Our two-step cascade approach shows to be a promising alternative to the typical regression-based approach for tool wear monitoring. It has a very good performance on the untrained data, while only employing the time-intensive image capturing module if the sensor-based model deems it necessary.

Acknowledgment

This research is funded by VLAIO (Flanders Innovation & Entrepreneurship) through the ATWI project and the Flemish Government through the AI Research Program.

References

1. D.-H. Kim, T. J. Kim, X. Wang, M. Kim, Y.-J. Quan, J. W. Oh, S.-H. Min, H. Kim, B. Bhandari, I. Yang *et al.*, “Smart machining process using machine learning: A review and perspective on machining industry,” *International Journal of Precision Engineering and Manufacturing-Green Technology*, vol. 5, no. 4, pp. 555–568, 2018.
2. A. Mohamed, M. Hassan, R. M’Saoubi, and H. Attia, “Tool condition monitoring for high-performance machining systems - a review,” *Sensors*, vol. 22, no. 6, 2022. [Online]. Available: <https://www.mdpi.com/1424-8220/22/6/2206>
3. T. Benkedjouh, K. Medjaher, N. Zerhouni, and S. Rechak, “Health assessment and life prediction of cutting tools based on support vector regression,” *Journal of intelligent manufacturing*, vol. 26, no. 2, pp. 213–223, 2015.
4. D. D’Addona, S. Ura, and D. Matarazzo, “Tool-wear prediction and pattern-recognition using artificial neural network and dna-based computing,” *Journal of Intelligent Manufacturing*, vol. 28, 08 2017.
5. D. Wu, C. Jennings, J. Terpenny, R. X. Gao, and S. Kumara, “A Comparative Study on Machine Learning Algorithms for Smart Manufacturing: Tool Wear Prediction Using Random Forests,” *Journal of Manufacturing Science and Engineering*, vol. 139, no. 7, 04 2017, 071018.
6. J. Ma, D. Luo, X. Liao, Z. Zhang, Y. Huang, and J. Lu, “Tool wear mechanism and prediction in milling tc18 titanium alloy using deep learning,” *Measurement*, vol. 173, p. 108554, 2021.

7. M. García-Ordás, E. Alegre, V. González-Castro, and R. Alaiz, “A computer vision approach to analyze and classify tool wear level in milling processes using shape descriptors and machine learning techniques,” *International Journal of Advanced Manufacturing Technology*, vol. 90, 05 2017.
8. H. Mamledesai, M. A. Soriano, and R. Ahmad, “A qualitative tool condition monitoring framework using convolution neural network and transfer learning,” *Applied Sciences*, vol. 10, no. 20, p. 7298, 2020.
9. P. Krishnakumar, K. Rameshkumar, and K. Ramachandran, “Tool wear condition prediction using vibration signals in high speed machining (hsm) of titanium (ti-6al-4 v) alloy,” *Procedia Computer Science*, vol. 50, pp. 270–275, 2015.
10. S. Painuli, M. Elangovan, and V. Sugumaran, “Tool condition monitoring using k-star algorithm,” *Expert Systems with Applications*, vol. 41, p. 2638–2643, 05 2014.
11. K. Guo, B. Yang, H. Wang, J. Sun, L. Lu *et al.*, “Singularity analysis of cutting force and vibration for tool condition monitoring in milling,” *IEEE Access*, vol. 7, pp. 134 113–134 124, 2019.
12. C. Shi, G. Panoutsos, B. Luo, H. Liu, B. Li, and X. Lin, “Using multiple-feature-spaces-based deep learning for tool condition monitoring in ultraprecision manufacturing,” *IEEE Transactions on Industrial Electronics*, vol. 66, no. 5, pp. 3794–3803, 2019.
13. R. Xie and D. Wu, “Optimal transport-based transfer learning for smart manufacturing: Tool wear prediction using out-of-domain data,” *Manufacturing Letters*, vol. 29, pp. 104–107, 2021.
14. B. Denkena, M.-A. Dittrich, H. Noske, and M. Witt, “Statistical approaches for semi-supervised anomaly detection in machining,” *Production Engineering*, vol. 14, no. 3, pp. 385–393, 2020.
15. T. Watanabe, I. Kono, and H. Onozuka, “Anomaly detection methods in turning based on motor data analysis,” *Procedia Manufacturing*, vol. 48, pp. 882–893, 2020.
16. J. Dou, C. Xu, S. Jiao, B. Li, J. Zhang, and X. Xu, “An unsupervised online monitoring method for tool wear using a sparse auto-encoder,” *The International Journal of Advanced Manufacturing Technology*, vol. 106, no. 5, pp. 2493–2507, 2020.
17. Y. Liang, S. Wang, W. Li, and X. Lu, “Data-driven anomaly diagnosis for machining processes,” *Engineering*, vol. 5, no. 4, pp. 646–652, 2019.
18. F. Fingerhut, C. Harsha, A. Eghbalian, T. Jacobs, M. Tabassian, R. Verbeke, and E. Tsiporkova, “Data-driven usage profiling and anomaly detection in support of sustainable machining processes,” 11 2022, pp. 127–136.
19. L. Alhadeff, M. Marshall, and T. Slatter, “The influence of tool coating on the length of the normal operating region (steady-state wear) for micro end mills,” *Precision Engineering*, vol. 60, 07 2019.
20. L. De Pauw, T. Jacobs, and T. Goedemé “Matwi: a multimodal automatic tool wear inspection dataset and baseline algorithms.” 2023, unpublished Manuscript.
21. X. Guo, X. Liu, E. Zhu, and J. Yin, “Deep clustering with convolutional autoencoders,” in *Neural Information Processing*, D. Liu, S. Xie, Y. Li, D. Zhao, and E.-S. M. El-Alfy, Eds. Cham: Springer International Publishing, 2017, pp. 373–382.
22. L. McInnes, J. Healy, N. Saul, and L. Großberger, “Umap: Uniform manifold approximation and projection,” *Journal of Open Source Software*, vol. 3, no. 29, p. 861, 2018. [Online]. Available: <https://doi.org/10.21105/joss.00861>
23. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.