



Indirect cutting tool wear classification using deep learning and chip colour analysis

Luca Pagani¹ · Paolo Parenti² · Salvatore Cataldo² · Paul J. Scott¹ · Massimiliano Annoni²

Received: 6 May 2020 / Accepted: 7 September 2020 / Published online: 7 October 2020
© The Author(s) 2020

Abstract

In the growing Industry 4.0 market, there is strong need to implement automatic inspection methods to support manufacturing processes. Tool wear in turning is one of the biggest concerns that most expert operators are able to indirectly infer through the analysis of the removed chips. Automatising this operation would enable developing more efficient cutting processes that turns in easier process planning management toward the Zero Defect Manufacturing paradigm. This paper presents a deep learning approach, based on image processing applied to turning chips for indirectly identifying tool wear levels. The procedure extracts different indicators from the RGB and HSV image channels and instructs a neural network for classifying the chips, based on tool state conditions. Images were collected with a high-resolution digital camera during an experimental cutting campaign involving tool wear analysis with direct microscope imaging. The sensitivity analysis confirmed that the most sensible image channels are the hue value H that were used to teach the network, leading to performances in the range of 95 of proper classification. The feasibility of the deep learning approach for indirectly understanding the tool wear from the chip colour characterisation is confirmed. However, due to the big effects on chip colours of variables as the workpiece material and cutting process parameters, the applicability is limited to stable production flows. An industrial implementation can be foreseen by populating proper large databases and by implementing real-time chip segmentation analysis.

Keywords Image processing · Vision inspection · Deep learning · Tool wear · Chip analysis · Monitoring

1 Highlights

- A deep learning network is implemented to classify turning tool wear
- The procedure is based on chip colour analysis
- The method can be implemented online
- The approach supports automatic management of tool wear

2 Introduction

Minimising the number of defects in manufactured products is fundamental for production nowadays, toward the implementation of the Zero Defect Manufacturing (ZDM)

paradigm [1]. A quick and real-time monitoring of turning tool wear becomes therefore extremely important to keep the production under control, even during unmanned operations [2]. Since the olden times, one of the biggest issues in machining industry is tool wear. Keeping tools in good conditions is a fundamental practice to enable optimal production. The implementation of tool wear identification and monitoring strategies is therefore a key aspect. The most reliable, but less efficient, methods are based on the direct observation of the tools state by means of microscopes or laser instruments. The efficiency of the direct methods is limited by the fact that typically the cutting must be interrupted to measure the tools. Despite recently some online tool inspection systems for machining becoming available in the scientific literature [3], most of the industrial direct methods possess this strong limitation. The indirect methods, such as those based on the cutting force/torque or spindle current analysis, are more efficient and therefore very widespread in the industry [4–8]. Their accuracy, which is high only in very repetitive scenarios, and the system costs are their limiting factors. Other

✉ Paolo Parenti
paolo.parenti@polimi.it

indirect methods for tool wear identification are based on the inspections of the workpiece dimension (such as workpiece diameter) and surface finishing properties, such as roughness or aesthetic surface response [9]. Some of them exploit automatic imaging methodologies. Despite that it is common practice of the expert operators to look at the produced chip shape and colours for assessing the tool wear, not many attempts in literature have been found by the authors that investigate this concept. The study of chip colour itself as a proxy for assessing the tool wear level is something that have been tried in [10] and [11]. Only recently have some authors developed a tool for systematic chip colour analysis to produce an automatic wear assessment method of tools during milling operations [12]. Nowadays, the recent developments of the vision systems, the increasing computing performances of the GPU graphic chip sets and the capacity of the latest imaging processing algorithms, like deep learning, are opening the opportunities to exploit image processing on multiple industrial monitoring scenarios.

A review of the impact of convolutional neural networks in manufacturing can be found in [13] and [14]. As observed by the authors, although manufacturing was one of the last applications of deep neural networks, some companies (such as Nvidia Corporation, Jabil and eSmarts Systems) have already implemented artificial intelligence platforms to solve manufacturing needs. Manual tasks have been replaced by automatic tasks, allowing a better allocation of resources. Neural networks have been successfully implemented in defect detection, quality control, data fusion and other big data applications.

In [15], the authors proposed to use a convolutional neural network to extract features from a manufactured product, with the aim to automatically detect defects in the manufacturing process.

Neural networks have also been used since two/three decades ago for predicting the tool wear as reported in [16].

One of the most recent applications is from [17] which proposed a multisensor model, based on convolutional neural network, using raw machine signals to predict tool wear in machining.

To the authors' knowledge, nobody has tried to exploit the image analysis on the machining chips' colour to monitor the tool wear. In this paper, a turning tool wear indirect monitoring approach has been implemented based on the machine learning techniques applied on images taken on turning chips. The present analysis focuses on dry cutting processes. The paper starts with an explanation about the mechanisms involved in chip colour generation in Section 3. After an introduction of the proposed monitoring approach in Section 4, details about the implemented algorithms are given together with a description of the validation experiment and results in Section 5. The results are discussed in Section 6, while Section 7 concludes the paper.

3 Cutting process and chip colour generation mechanism

In machining processes, like turning, the material is removed by forming chip through the direct contact between a wear-resistant tool and the workpiece. Morphology and aspect of the chips produced by cutting shear deformation are affected by multiple process parameters [18]; see Fig. 1.

The chip morphology is one of the drivers in the tool design process and in the definition of optimal tool operative conditions. Chips can be continuous or discontinuous. The latter type is preferred in most of the industrial operations because it can be easily managed and removed from the cutting zone, avoiding re-machining and workpiece scratches but also damaging to the machine components. The introduction of the chip-breaker technology and the adoption of proper cutting parameters have to lead to (i) the

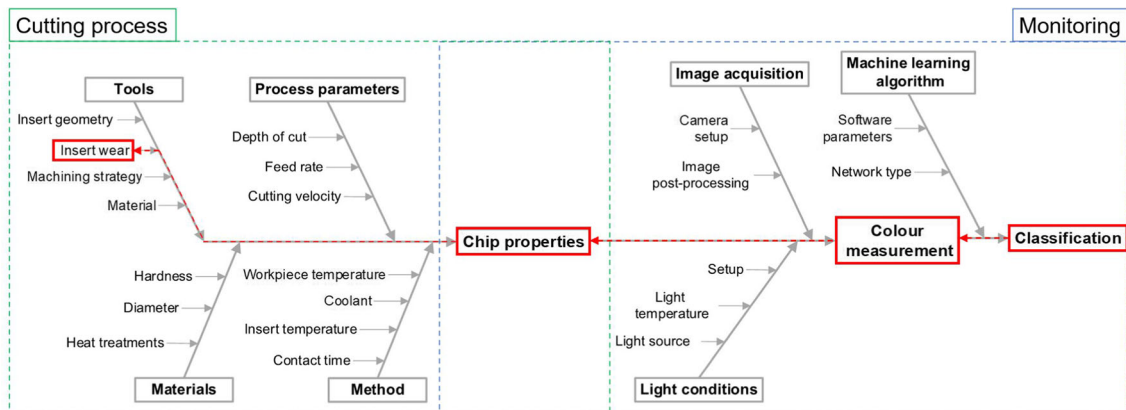


Fig. 1 Cause-and-effect diagram of chip properties and image based classification

generation of suitable chips, (ii) adequate tool life and (iii) high productivity. A good cutting setup is that where the large quantity of the generated heat is transferred into and removed by the chips, alleviating the part and the machine system from thermal stresses [19, 20]. This heat is related to the different alloy elements contained in the chip material and their oxidation leading to the generation of colour grading and colour striations on the chips. In consequence of that, any deviations from optimal cutting conditions, as those associated to the tool wear, are affecting chip shape and colour [21].

4 Material and methods

The study was based on five main phases: (i) performing turning tests to collect chips produced by tools at different wear conditions and with different cutting parameters, (ii) sampling images of the chips, (iii) colour extraction from the images, (iii) development of a machine learning tool to exploit the extracted signals and (iv) test the power fullness and robustness of the models.

4.1 Cutting experiments

4.1.1 Turning test and chip collection

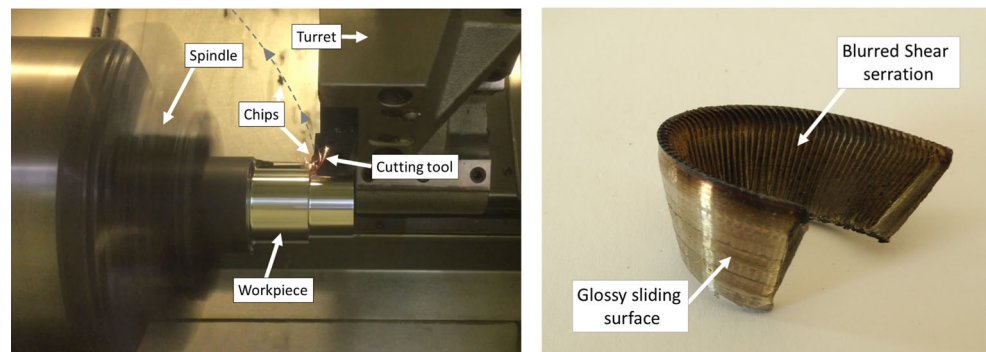
In order to stick to real industrial cases, a railway axle rough turning is considered as the manufacturing case. This scenario is one of the three selected cases in the ForZDM project, aimed at developing solutions for production and quality control of multi-stage manufacturing systems through the integration of multi-level system modelling, big data analysis, CPS (Cyber Physical Systems) and real-time data management. This kind of application, with large contact time on large components, is ideal for testing the developed approach since it gives enough cutting stability to prove the concept. Turning tests were therefore conducted

on an industrial lathe (Biglia B301) by using roughing CVD (TiCN+AL₂O₃+TiN) cutting inserts with 80° diamond shape (Sandvik CNMM 120612 - QR4325) (Fig. 2). A tempered and quenched low carbon steel (25CrMo4 - EN 10083-3), suitable for railway axles, was used for the specimens in the form of bars with initial diameter of 50 mm and length 114 mm. No tail-stock was used and a total length of 100 mm was cut from each bar (50 mm on one end and 50 mm on the other, after flipping the bar in the chuck), starting from the initial diameter 44 to 26 mm with 8 passes with depth of cut $ap = 1.5$ mm. In the wear tests, one insert was used until it got worn. In total, 18 bars were adopted by fully randomising the cutting experiments. This was done in order to confound any effects given by inhomogeneity of the bar material composition and hardness on the tool wear development. Preliminary tests were conducted by adopting cutting coolant whilst the final tests, used to teach and validate the monitoring approach, were performed in dry cutting conditions.

The other adopted process parameters were, as suggested by tool manufacturer, cutting speed $V_c = 180$ m/min and feedrate $f_z = 0.3$ m/rev, generating an overall Material Removal Rate (MRR) equal to 1350 mm³/s. The G-Code was designed in order to allow the operations to be interrupted for allowing chip collection. Chips were collected at each pass lasting around 5–8 s based on the actual bar diameter. In order to generate replicable cutting conditions, the initial temperature of the bars was kept constant by adopting an idle cooling time (in air) of 5 min between each cutting operation. The chips were collected in the middle of the workpiece in order to analyse stable cutting conditions thus avoiding transient colour deviations caused by the tool engagement phase and different chip geometry (longer chips) caused by the tool interaction with the workpiece square shoulder.

Preliminary tests on the presence of the coolant during cutting operations and on the adoption of different cutting parameters demonstrated that the chip colour in the analysed

Fig. 2 Cutting setup and machined chip



(a) Cutting setup

(b) Aspect of a machined chip: colour and morphology

case did not change significantly. The impact of the coolant on the chip colour was small, even if some minor effects were noticed about the position of the coolant supply tube.

4.1.2 Tool wear evaluation

Tool wear was assessed by interrupting the cutting tests after around a tool-workpiece contact time of 100–200 s (cutting length of about 200–450 mm depending on the cutting parameters). Tool inserts were removed and analysed at an optical microscope (Mitutoyo Quick Vison APEX 302 PRO3) where the flank wear land width (V_b) was extracted following the standard direct tool wear evaluation procedure ([22]). In the studied cases, the crater dimensions on the rake face, as the depth K_t , were less affected by the cutting contact time thus being a less sensible tool wear indicator. The first tool wear phase was the initial tool break-in, in which coating damages were observed. As can be seen in Fig. 3, this initial phase lasts for about 200 s. A steady-state of the wear then followed for about 800 s where a barely significant wear rate was noticed. After this phase, as expected, there was a sudden change of the wear rate visible not only in terms of V_b value but also in terms of rake face wear.

4.1.3 Image acquisition setup

The images of the chips were acquired on a dedicated overhanging stand (Fig. 4), equipped with a precise positioning stage and four light led bulbs (light temperature 6000 K). The lights were positioned at around 300 mm with inclinations of 30° and 45° (for the two sets of lamps) in order to remove as much as possible the shadows. Around 20–40 chips were disposed on a sheet of white

paper covering an area of around 25–35 cm². A compact digital camera (Nikon P7000 with a 1/1.7-in. CCD sensor with 10.1 megapixel) was adopted for taking the pictures. The exposure parameters were set in order to generate best possible images that preserve the chip colouring (focal distance 6 mm, exposure 1/50 s, numerical aperture 3.2, ISO 100).

For each batch of collected chips, three images were taken. In order to give some variability to the image acquisition, different chip arrangements were acquired: after each photo shoot, the chips were manually mixed to generate random overlapping and chip placement.

Examples of chips coming from tools with low and high wear are shown in Fig. 5. The chips appear coloured with different shades. They appear glossy, shining and polished at the outer surface because of the sliding on the tool rake face. On the contrary, the internal chip face appears corrugated, rougher and darker due to the shear mechanisms involved.

Seven (7) collections of pile chips were taken during the tool lifespan, one every two cutting passes (the collected chips then belong to four (4) diameters of the parts, i.e., 44 mm, 38 mm, 32 mm, 26 mm). Therefore, twenty-eight (28) pile chips were collected. For each of the chip pile, three (3) different images were taken bringing to a total of $7 \cdot 4 \cdot 3 = 84$ images that were collected. The image capturing was performed offline, i.e., at the end of the cutting tests. The acquisition sequence was totally randomised to confound possible effects of the human operators in displacing the chips under the camera.

In this paper, the chip colour solely was used as the wear quality index, basing on the feasibility given by the initial tests. The other visible characteristic of the chips, namely the chip shape, will be considered in the future; it was

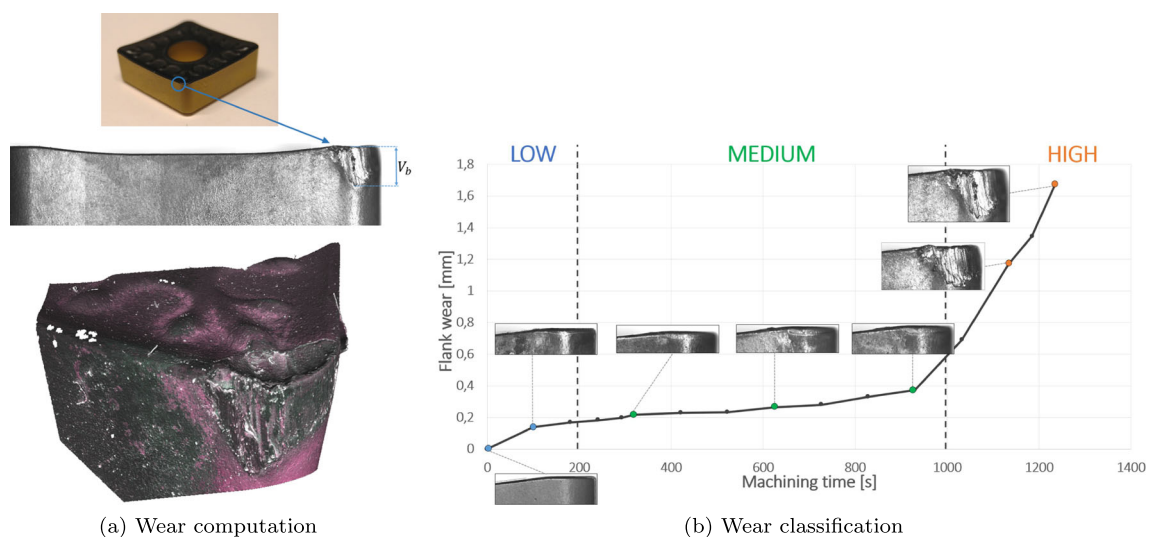


Fig. 3 Wear computation and classifications

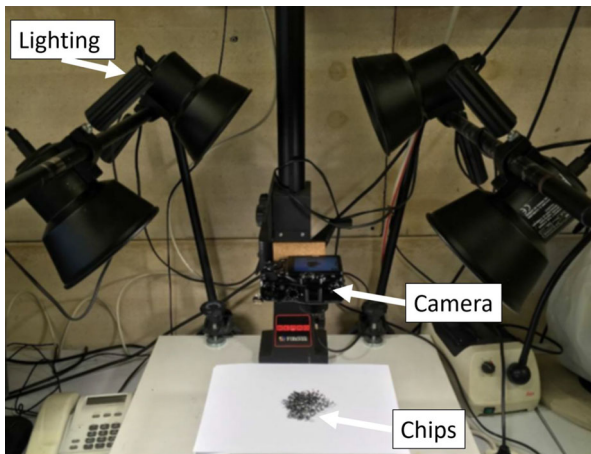


Fig. 4 Image acquisition setup

noticed from the initial testing that chip shape was strongly affected by a variation of cutting parameters and not only with tool wear thus increasing the monitoring complexity.

4.1.4 Image pre-processing

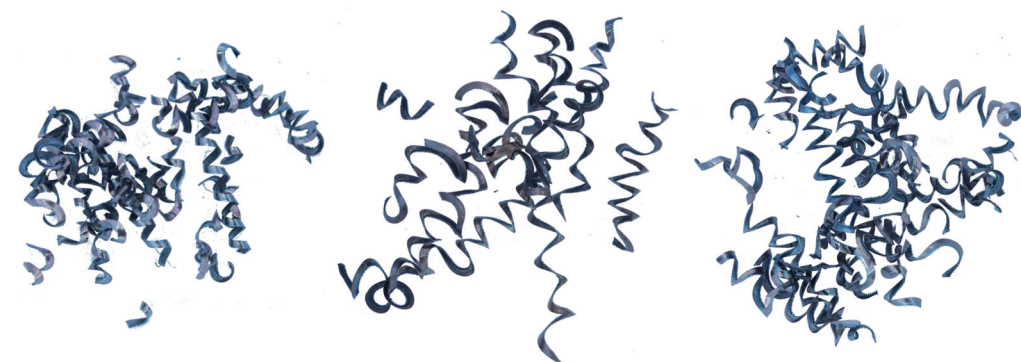
After acquiring the images, the background must be identified and deleted in order to let the algorithm focus on the solely colour features of interest. This operation

was manually carried out in GIMP [23] after performing a white balance operation. The chips were segmented in a semi-automatic fashion: the backgrounds were first selected using the fuzzy selection tool available in GIMP. The results were then manually refined (Fig. 6). For an automatic segmentation of the chips on the chip conveyor belt of the machine, other image processing methods can be applied. Future studies will be devoted to this aspect. Algorithms able to perform a pixel-based classification are called semantic segmentation [24]. After a good training of the neural network, those methods should allow the segmentation of an image in chips, lubricant and background. Semantic segmentation was not under the scope of this paper, that is to design a robust method to classify the tool wear based on the chip colour characteristics.

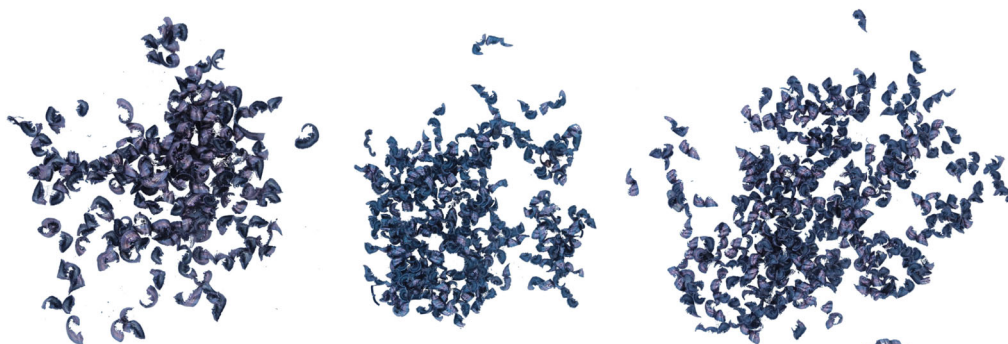
4.2 Technological requirements

Since, in industrial operations, the generated chips are continuously generated during the cutting and conveyed on a dedicated belt that brings them in a deposit tank, the algorithm should be:

- As simple as possible for supporting quick real-time analysis on low-cost computers;



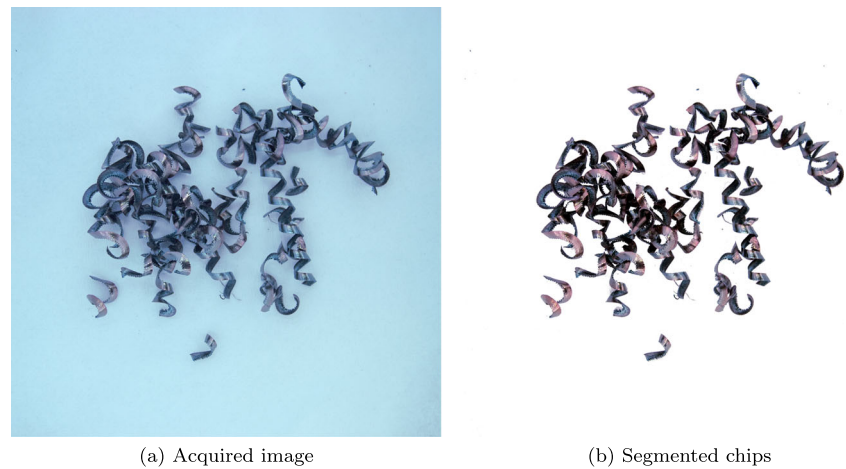
(a) Chips from low wear tool. The first on the left was acquired at the beginning of the machining, while the others after 180 s of machining



(b) Example of acquired images of chips produced with a tool with medium-high and high wear

Fig. 5 Example of acquired images of chips produced with a tool with low and high wear

Fig. 6 Acquired images and segmented chips



- Robust to the overlapped placement of the chips on the conveying belt (which is random)
- Robust to the chip shape variations (that are generated during tool life and cutting transients and that are related to cutting parameters)
- Robust to the presence of lubricating/coolant residues on the chips.

For instance, segmentation of each single chip must be avoided in order to keep the algorithm as simple as possible.

5 Machine learning algorithm

In this section, the algorithm used to analyse the data is presented. The flowchart of the methodology is shown in Fig. 7. Once the images were acquired, a data augmentation was applied to simulate new chips. The colour histogram of the augmented images was then smoothed using the kernel density estimation. Those curves were then used to estimate the parameters of the classification algorithm. The curves can be split into two sets, training and validation, to avoid the overfitting in the parameters' optimisation phase [25].

Two methods able to classify a curve, representing an acquired image, were investigated: (i) convolutional neural network [25] and (ii) supervised classification using functional data analysis [26].

The method used to generate more samples, necessary to perform the training of the algorithm, is first presented. It is followed by the algorithm used to extract the curve from the image histogram. Classification methods are then presented and the results for each channel of the HSV and RGB colour spaces are analysed.

5.1 Data augmentation

Data acquired by cameras may be affected by the ambient in which they are placed. During the training phase, in order to obtain a good model, the possible variability must be simulated to make the model robust. To generate more training data, each image was rotated and divided in four images of equal size that were then analysed. To simulate a camera with lower resolution, or not working in optimal conditions, a Gaussian blurring, with σ equal to 4 pixels, was applied to the acquired images [27]. In this study, it was observed that varying the value of σ does not affect the models' parameter estimation. Gaussian blurring was added in order to give more variability of the input data, leading to a better prediction ability. An image with and without the blurring is shown in Fig. 8. Rotations and splitting were then applied to the new set of generated images.

Each of the 84 acquired and blurred images was rotated by the arbitrary number of 10 different angles; each image was then divided to generate four images. For each image,

Fig. 7 Flowchart of the methodology

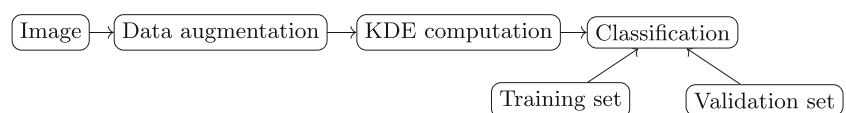
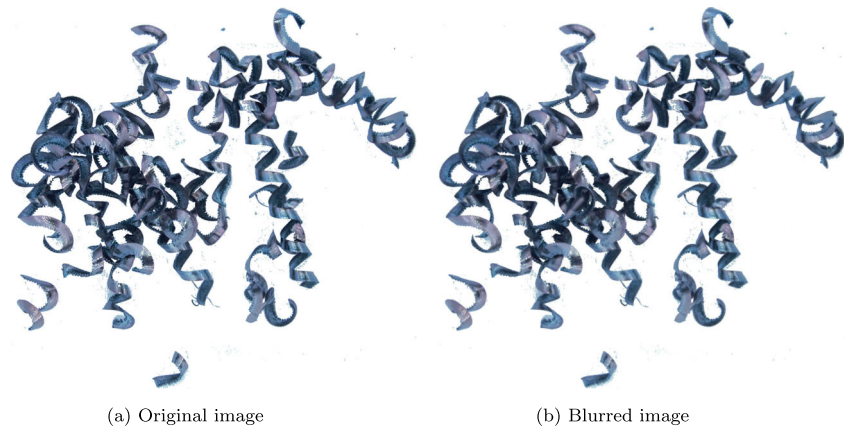


Fig. 8 Acquired images and image after blurring



$10 \cdot 4 = 40$ were generated, leading to a total number of $40 \cdot 84 \cdot 2 = 6720$ images.

The augmented images were used to estimate the models' parameters, while the acquired images will be used to check the performance of the investigated models.

5.2 Curve computation: from histogram to density function

Histograms are widely used in image analysis; since pixel values are usually stored by integer values ranging from 0 to 255, the histogram of an image is divided in 256 bins (one for each possible value). Since the colour may vary, due to the ambient factors, the histogram is not a robust approximation of the colour in the image. To avoid choosing the number of bins, the density function of each channel can be computed instead [28]. Both RGB and HSV colour scales were analysed [27]. Each channel was analysed separately. RGB represents the usual red, green and blue channels of an image, while HSV, hue, saturation and value, is a colour scale developed to be close to the human perception of colours.

A kernel density estimation (KDE) is a statistical method to estimate the density of a probability density function [29] of a random variable.

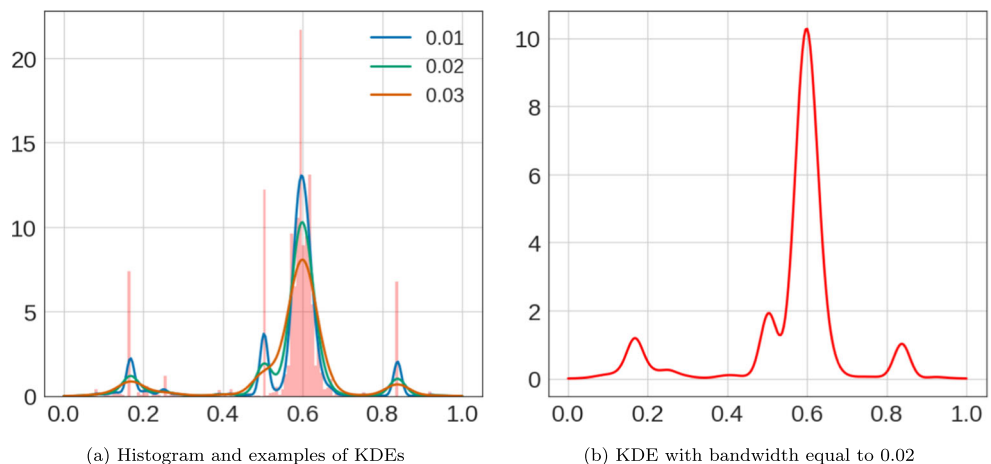
Let $\{x_i\}_{i=1}^n$ the sampled data. The density is computed as:

$$f(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i), \quad K_h(t) = \exp \left\{ -\frac{t^2}{2h} \right\}$$

where $K_h(t)$ is a kernel function. It depends on the parameter h , called bandwidth, that controls the region of influence of each point t , i.e. higher h will result in a larger number of points used in the smoothing process.

A comparison of the estimated density and the image histogram of the H channel is shown in Fig. 9. Using the histogram data may lead to non-robust results, while using a smoothing function helps to take into consideration also some ambient factors than can vary the colour of the acquired image. Figure 9 shows the density estimated using three different bandwidths. If this parameter is chosen too small, the estimated density function will follow the image histogram (blue curve), while if it is set to a high number it will smooth some details (orange curve). In this paper, the

Fig. 9 Histogram and KDE estimations



bandwidth was set to 0.02 (green curve and Fig. 9b) for all the analysed images.

Figure 10 reports the density plot for all the acquired images, for each channel in the HSV colour space. The H channel shows a shift of the highest peak as the wear increases. In the S channel, there is a distinction of the curves belonging to the medium class of the wear, but the curves of the low and high wear seem overlapping. The variability of the curves of the V channel appears high, and it appears difficult to see any visible distinctions. It was decided to analyse each channel independently to verify if it

is possible to predict the wear's class from an image of the chips.

5.3 Convolutional neural network

Neural networks [30] are an artificial intelligence tool designed based on the human brain and nervous system. In the frame of this study, the classification will be performed using convolutional networks. Convolutional networks, or convolutional neural networks, are a class of neural networks [25] in which there is at least a convolution

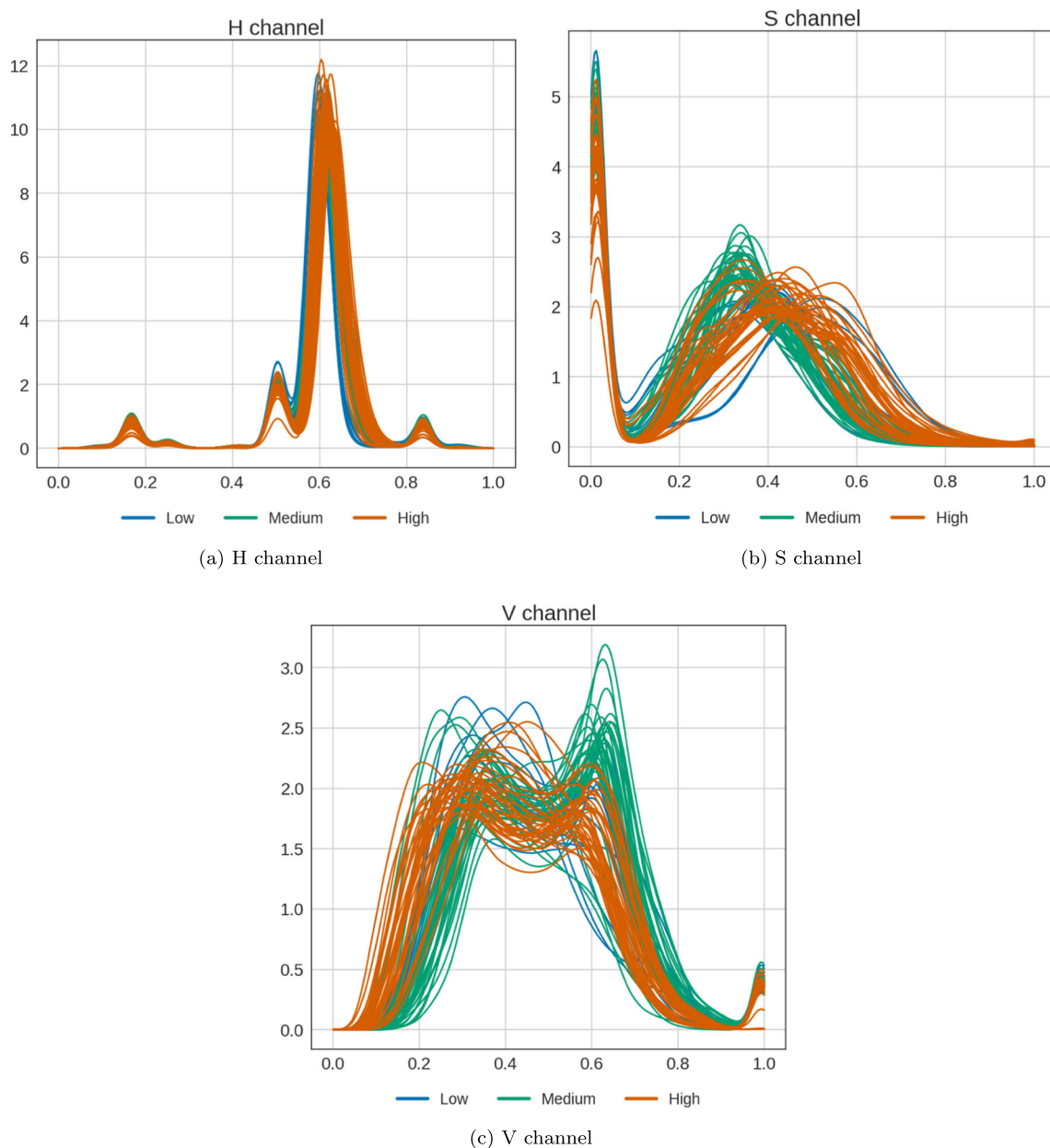


Fig. 10 Density estimation HSV color, for each level of wear (low, medium, and high)

layer. Convolutional layers allow the network to extract some features of interest from the curves estimated using the KDE at the previous step. Convolutional networks are usually characterised by some deep layers, allowing better classification of the results compared with the fully connected neural network. The discrete convolution operation, for a 1D signal, is defined as:

$$X(i) = \sum_{i=n}^N x(i - n) w(n) \tag{1}$$

where $x(i)$ is the i -th point of the original signal, $w(i)$ is the i -th weight of the kernel of size N and $X(i)$ the i -th point of the signal after applying the convolution operation. During the optimisation process, the weights $w(n)$ in Eq. (1) have to be selected in order to better classify the observed curve in the appropriate class. Usually N , the size of the kernel, is kept low to allow better optimisation. Each point of the output signal $X(i)$ is then a weighted average of its neighbourhood points.

Table 1 shows the designed network. It is composed of a total of three convolutional layers, each followed by a max pooling layer. A max pooling layer is used to down-sample the feature map. The feature vector is divided in sub-vectors of size n . For each of these, only the maximum value is stored on the feature vector of the following layer. Two dense layers are added to the end of the network to classify each curve in one of the wear classes. The output of the network is a vector of dimension three where each entry represents the probability of belonging to the corresponding class. The total number of parameters to estimate is equal to 3393. The network was implemented using TensorFlow [31].

Table 1 Neural network design and number of parameters to estimate

Layer (type)	Output Shape	Param #
conv_1_1 (Conv1D)	(None, 254, 30)	120
max_pooling_1_1 (MaxPooling1)	(None, 84, 30)	0
dropout_1 (Dropout)	(None, 84, 30)	0
conv_1_2 (Conv1D)	(None, 82, 20)	1820
max_pooling_1_2 (MaxPooling1)	(None, 27, 20)	0
drop_out_2 (Dropout)	(None, 27, 20)	0
conv_1_3 (Conv1D)	(None, 25, 10)	610
max_pooling_1_3 (MaxPooling1)	(None, 8, 10)	0
drop_out_3 (Dropout)	(None, 8, 10)	0
flatten_1 (Flatten)	(None, 80)	0
dense_1 (Dense)	(None, 10)	810
dense_2 (Dense)	(None, 3)	33
Total params: 3393		
Trainable params: 3393		
Non-trainable params: 0		

Table 2 reports the parameters used during the training phase of the network. The network is described in Table 1 and those parameters were used for all the analysed curves, showing good convergence results in the training phase. A total number of 6720 curves computed from the augmented images were split into two groups: the training and validation sets. Seventy percent were used to estimate the optimal network parameters, while the remaining 30% were used to check the goodness of the optimisation, to avoid under- and overfitting. The 84 acquired images will be used only to check the prediction accuracy. To allow better optimisation, the curves were standardised in order to have null mean and unitary variance.

Figure 11 shows the accuracy and loss functions versus the number of the epochs for the H channel. A good network can be chosen selecting the parameters corresponding to a point of the accuracy and/or the loss functions, after the curves have reached the steady state. The curves have a similar behaviour; the model after 350 epochs was used. It had a prediction accuracy of more than 95% on the validation set. The optimisation was run also with a lower number of training curves (6000), i.e., without using augmented curves coming from some of the acquired images.

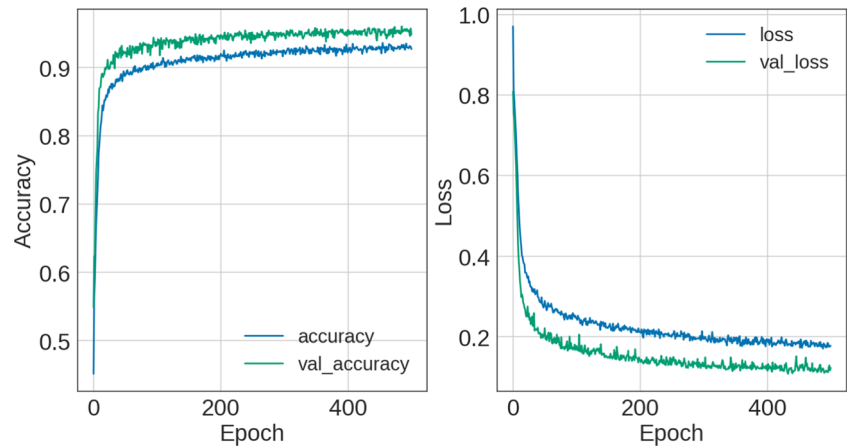
The acquired images, not used in the training and validation sets, were then used to check the goodness of the estimated model. The prediction probability is reported in Fig. 12; the dashed vertical lines separate the true classes, from left to right: low, medium and high. There was an error in the prediction in the transition between the low and medium wear classes. From the wear measurement in Fig. 3b, the trend is linear, so an error during the transition zone is admissible for a non-overfitting algorithm. Furthermore, if the wear is classified as medium, the tool can be used in the manufacturing process. On the contrary, the classification between the medium and the high classes is well marked, allowing a prompt tool replacement. The training time was around 1 h, while the prediction 0.01 s per image by using a computer with an Intel(R) Core(TM) i7-5930K CPU @ 3.50 GHz processor.

The difference between the features of a sample coming from a medium wear and one from low wear is shown in Fig. 13, while the difference between high and medium wear is shown in Fig. 14. Each colour represents a feature curve. The first layer enhances the difference caused by the

Table 2 Parameters used during the neural network training

Adam optimizer			Batch size	# epochs
LR	β_1	β_2		
0.001	0.9	0.99	32	500

Fig. 11 Loss and accuracy measures



shift of the higher peak on the H channel. Other differences are highlighted in the other layers.

5.4 Best channel identification

After showing the good results using the H channel as wear index, the neural network is applied to the S, V and RGB channels to identify the best model. All the models have good prediction results during the training phase: the accuracy on the validation set was higher than 95% and 90% for the S and V channels respectively. Looking at the prediction results of the S and V model, shown in Fig. 15, it is possible to draw the following conclusions:

- In the S channel, there is not a good separation between the medium and the high levels of wear; this may lead to wrong classification when the tool is at the end of its life;
- Due to the high variability of the curve, the prediction results using the V channel are poor; some of the chips

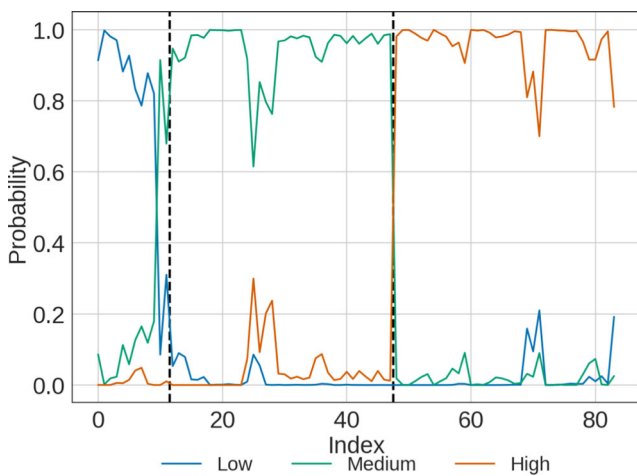


Fig. 12 Classification results using the H channel

produced by a tool with no wear were classified as coming from a tool with high wear.

Figure 16 shows the prediction results using the RGB channels. It is clear that it is not possible to achieve a good classification result using this colour space.

From these results, it is possible to conclude that it is possible to achieve a good classification using the neural network trained on the H channel. The results on the other channels show that the designed neural network is not capable of achieving good classification results because it is not possible to extract some distinctive feature for each wear class.

5.5 Classification using functional data analysis

Functional data analysis (FDA) [32] is a statistical tool aiming to analyse curves, surfaces, or high-dimensional manifolds. In FDA, the classical point-based supervised classification [33] can be generalised to curves [26]. The prediction is done using a distance measure between each tested curve and all the available classes. Given two curves, $x_i(t)$ and $x_j(t)$, the distance can be computed as:

$$d(x_i(t), x_j(t)) = \sqrt{\int_{\mathbb{R}} \phi_{\tau}((x_i(t) - x_j(t))^2)}$$

where $\phi_{\tau}(t)$ is a Gaussian kernel function depending on the bandwidth parameter τ :

$$\phi_{\tau}(t) = \frac{1}{2\pi} \exp\left\{-\frac{t^2}{\tau}\right\}$$

For an exhaustive definition of the classification of functional data, the reader is referred to [34]. As with the previous method, the augmented images are used to estimate the model parameters, while the acquired images were used to check the model prediction ability. `classiFunc` package, implemented in R [35], was used to perform the

(Medium - Low) wear feature maps

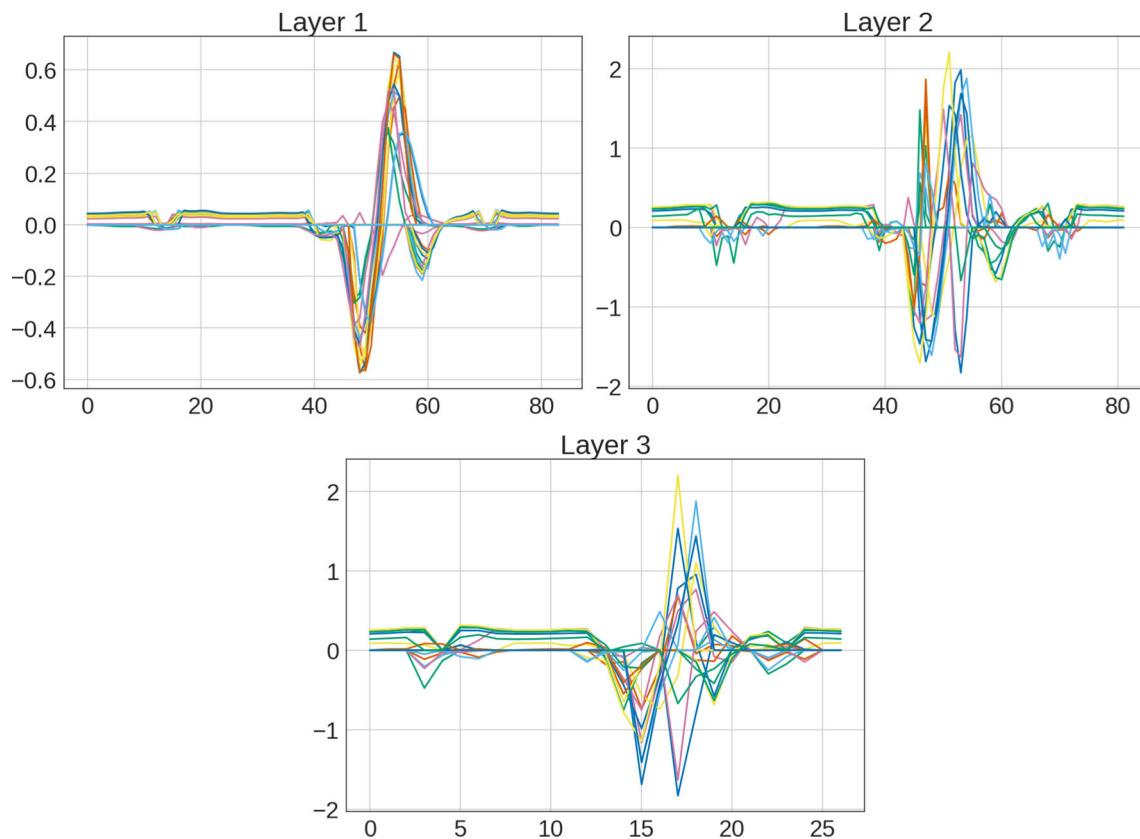


Fig. 13 Difference between the features of chips with medium and low wear using the H channel

functional classification [36]. L^2 metric and a Gaussian kernel, with a bandwidth equal to 1, were used. It was observed that, in this test case, the bandwidth did not influence the goodness of the prediction. The prediction accuracy during the training phase was equal to 88%, 66% and 54% for the H, S and V channels, respectively. The prediction results are shown in Fig. 17. The best accuracy can be achieved using the H channel, but it is low, equal to 56%. There is misclassification between the medium and the high wear, while no sample is classified as low wear. The models based on the S and V channel are obviously not able to classify the wear given a chip image.

5.6 Performance verification of the classification with respect to cutting parameter variations

In order to verify how much the chip colour, i.e. the feature used for the wear assessment, was affected by a change of the cutting conditions, a specific phase of the experimental testing was designed and carried out. In particular, five different new tools with five different cutting conditions were tested and the collected chips were analysed as in Table 3.

The obtained colour variations were low, and the algorithm was able to correctly classify the tools in the correct state, namely the new condition. The prediction probability is reported in Table 3. Since all the tools were new, they should be classified in the low wear class. This brings to conclude that, for the tested conditions, the colour feature is robust for the tool wear assessment since it varies much more with the wear than with all the tested cutting conditions.

6 Discussion of results

Based on the wide literature and infinite industrial experience out there, it can be stated that the chip colour characteristics reflect the actual tool wear state. In formal terms, the choice of the chip colour characteristics as the monitoring feature gives further reliability because the generated colour is directly coupled with the cutting temperatures that are strongly affected by tool wear onset. Based on this fact, the deep learning method is capable of processing the complex colour distributions on the chips and classifying the tool wear state. On the contrary,

(High - Medium) wear feature maps

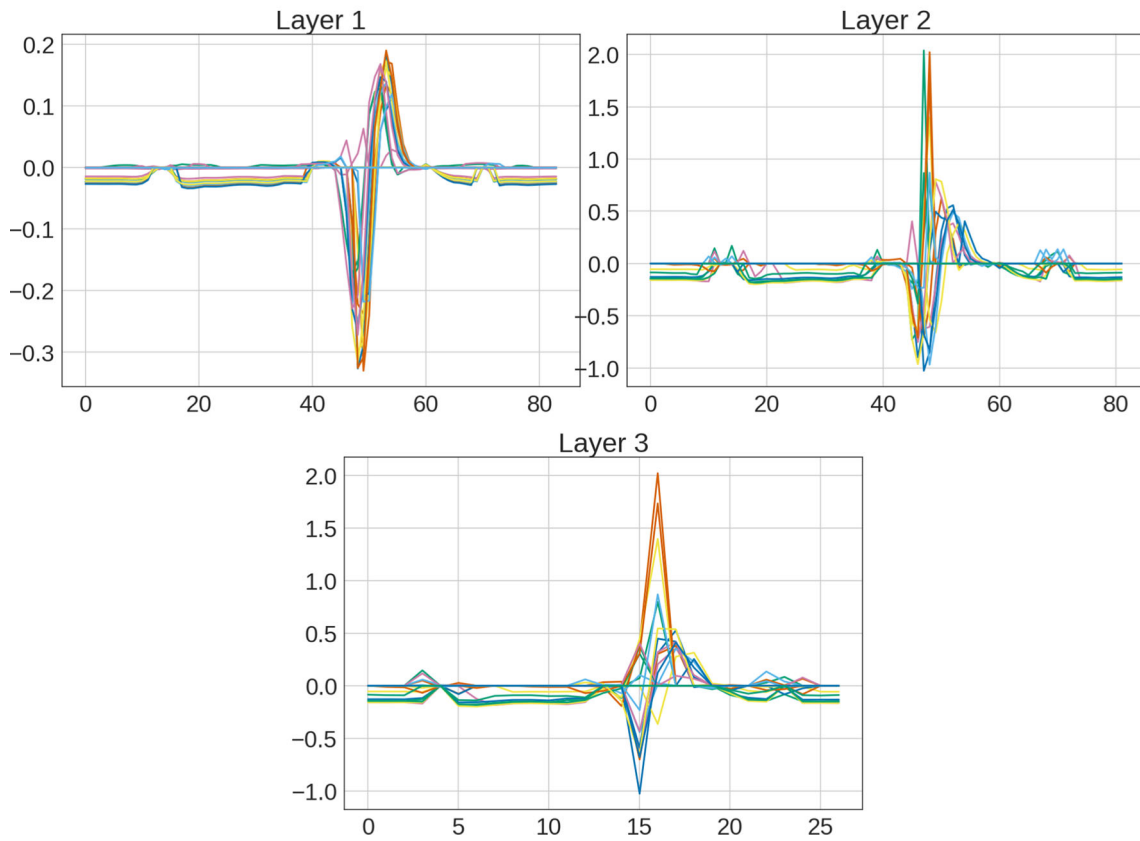


Fig. 14 Difference between the features of chips with low and medium wear using the H channel

other advanced statistical methodologies, such as the FDA, did not generate a proper classification, starting from the same input image data. This demonstrates the superior capacity of the deep learning network, which elevates it to be a suitable method to replace the eye and the

experience of a trained manufacturing operator. The method demonstrated adequate robustness to bring toward the development of automatic tool wear classification that does not impact the productivity of the machining systems. The obtained results confirm the suitability of the inspection

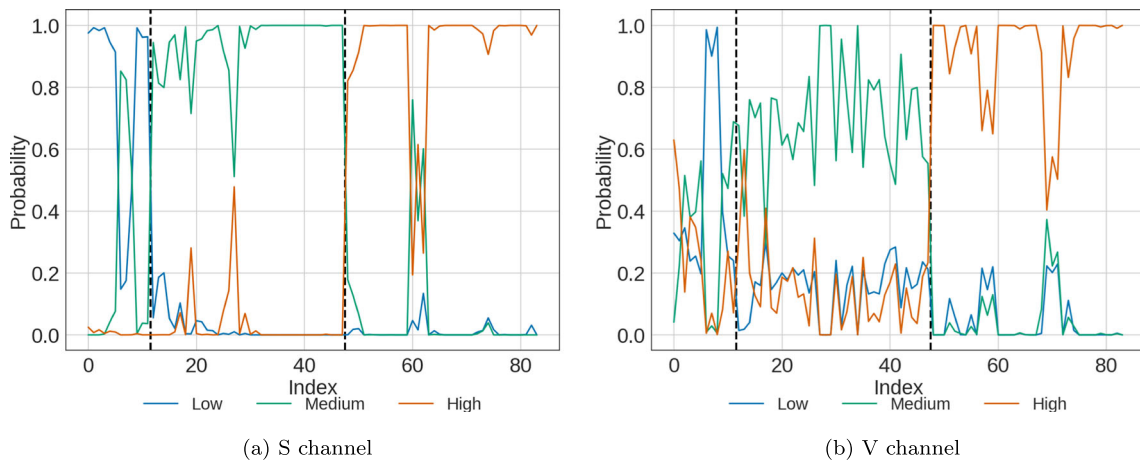


Fig. 15 Classification results using the S and V channels

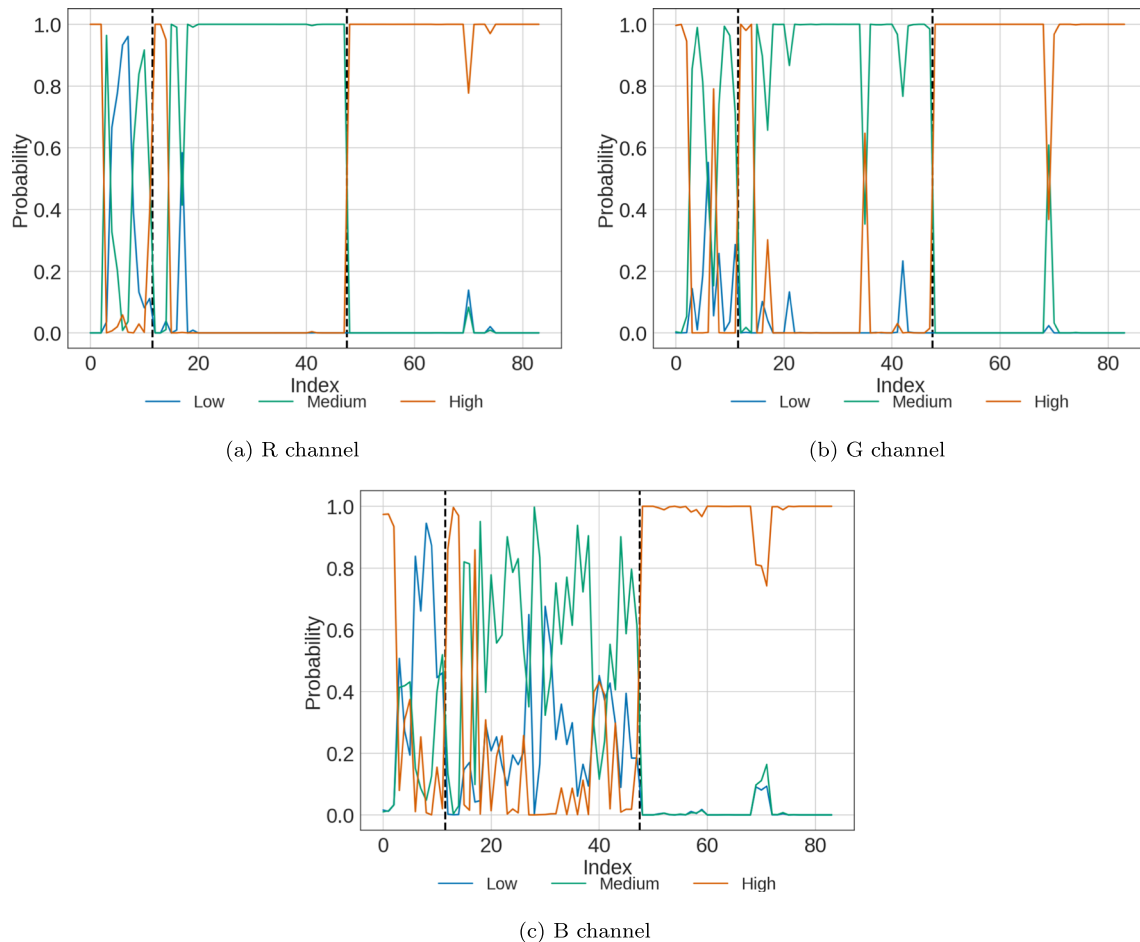


Fig. 16 Classification results using the RGB channels

method for the turning case of bars in roughing. The robustness showed by the method with respect to chip shape variations supports the belief that the method can address also more complex machining tasks such as milling operations. Good classification performances were showed with respect to workpiece diameters variations, which is a natural consequence of the material removal. Clearly, as in all of the neural network algorithms, much of the result strictly depends on the goodness of the learning phase; the representativeness of the chip images used to teach the network is fundamental. Every variation in the material coupling between tool and workpiece requires a new dedicated learning phase. The method is more reliable than offline standard wear predictions strategies (i.e. those based on prior empirical data collections such as Taylor law [22]) because it is based on the observation of a direct cutting process outcome, as the chips. It can be observed that offline monitoring methods work well when a small variability (caused by different sharpness levels or coating thickness/adhesion/durability) takes place within nominally

equal tool inserts. On the contrary, the proposed monitoring implementation is able to take this variability into account, keeping away from too premature or late prescriptions of tool changes. Another source of variability that makes this method stronger than offline prediction comes from the fact that in the studied scenario, i.e. the roughing turning of big forged shafts for the transportation sector, the forged workpieces typically shows quite irregular material allowance over the large part diameters; this fact generates intermittent tool contacts that make the wear unpredictable with offline methods, since the actual contact time of the tools is unknown. On the contrary, one of the limitations of the proposed method shared with the other prediction methods is related to the variations of the chip colour caused by workpiece property deviations such as chemical composition, microstructure and hardness, that sometimes are unavoidable in real application fields. In the presence of big changes, a new training of the network would be required. However, in advanced machining scenarios where only certified materials can be adopted, this will

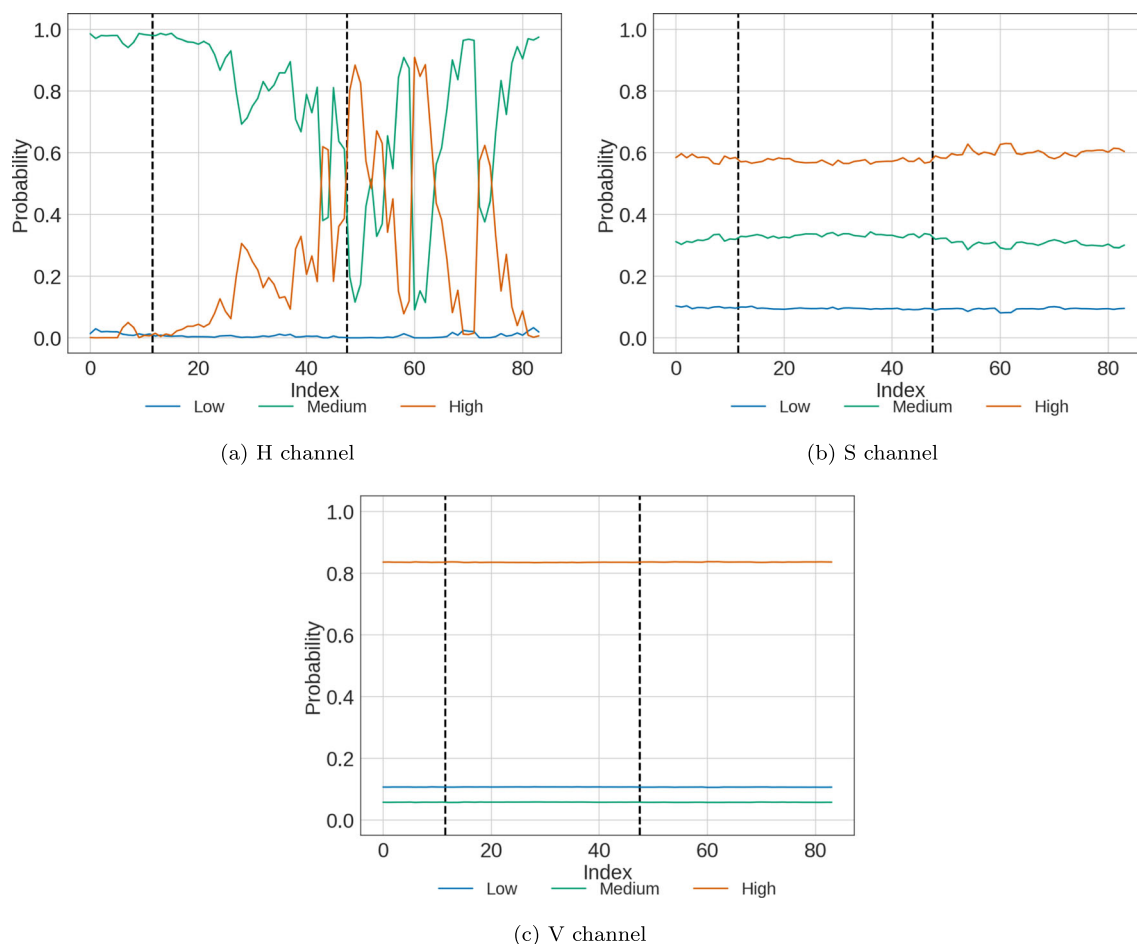


Fig. 17 Classification results of the functional clustering model using the HSV channels

constitute a minor issue, like in the addressed case. To verify the robustness of the computed solutions, both the training set and the parameters of the optimisation algorithm were changed, e.g. a training set with reduced number of images but the same network structure was applied. It was observed that those parameters have a small influence of the

found solution, i.e. the differences in the prediction were negligible. Other parameters that have to be set are the width of the kernel, in the KDE phase, and the number of points for each curve. To compute the width of the kernel, a good starting point is the automatic estimation [37]. The optimal width is optimised for the Gaussian distribution.

Table 3 Cutting experiments aimed at testing the influence on chip colour of cutting parameters together with obtained probability wear class prediction

	Sample #				
	2	3	4	5	
Cutting parameters					
V_c (m/min)	50	180	50	50	180
f_z (m/rev)	0.1	0.3	0.3	0.3	0.3
a_p (mm)	1.5	3.0	3.0	1.5	1.5
Predicted probability					
p_{low} (%)	99.72	99.25	99.93	99.22	99.76
p_{medium} (%)	0.28	0.75	0.07	0.72	0.24
p_{high} (%)	0.00	0.00	0.00	0.00	0.00

The values have to be changed to allow the smoothed curve to correctly follow the image histogram. The number of points of each curve, set to 256 in this paper, does not influence the prediction capability of the investigated models. This number should not be set to a small value, otherwise the machine learning algorithm would not be able to extract the useful features for a correct classification.

6.1 Industrial applicability

It should be noted that the vision system for automatic image capturing can be mounted on the conveyor belt of an industrial machine. Typically, the conveyor includes a washing step to cool down the chips which could be exploited to clean the chips from a part of the lubricant residues, thus preventing shadowing effect on the original chip colours. Industrial cameras can be adopted successfully as proven by some initial testing carried out aside from the experiments presented in this manuscript. Specific LED-light control boxes can be built to assure constant light conditions during the pic exposure. The conveyor speeds are typically low, about 0.25 m/s, and would even allow getting sharp images by adopting standard light intensity. Moreover, the typical homogeneous black background of the conveyors would help in the implementation of automatic background removal operations. The last remark is about the code implementation and transfer. The analysis presented in this paper was conducted by exploiting codes written by the authors which are easily implementable on industrial systems. To further foster the spread into the industrial field, the code can be re-written using the many multiple libraries that industrial coding sources offer.

7 Conclusions and future developments

An innovative deep learning method was proposed aimed at indirect tool wear classification using the colour of the produced chips in turning operations. This method can lead to automatic tool wear management that fosters the improvement of cutting system performances. It was shown that deep learning performs a good prediction accuracy on the available data, allowing a fast prediction on the acquired images, meaning that the method can be easily implemented on the production lines. Moreover, the proposed classification strategy does not depend on the chip shapes, making it suitable addressing a large number of manufacturing conditions. Since the approach is based on the image's colours, lighting conditions have an influence on the predicted wear class, i.e. the chips must be acquired under stable light conditions. Although the network was trained using blurred images to simulate a lower resolution camera, the performance using a commercial camera has to

be validated. Since the chip's colour signature depends on the work material, different models are needed to address multiple materials processed. The system works well in the tested dry cutting operations. Further developments are needed to extend the methodology to lubricated cutting operations due to the relevant influence that coolant (and coolant delivery system) can have on the cutting temperatures and chip colours. Future works will involve the development of automatic tools for chip segmentation in order to enable a fully automatic classification of images in chips, background and lubricant.

Acknowledgements Additional thanks goes to Eng. D.Gariboldi, M. Taccin (Sandvik) and P. Aquilino, S.Crespi (POLIMI).

LP and PJS acknowledge the EPSRC Fellowship in Manufacturing: Controlling geometrical variability of products in the digital and smart manufacturing era (ref: EP/R024162/1).

Funding Open access funding provided by Politecnico di Milano within the CRUI-CARE Agreement. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 723698. This paper reflects only the author's views and the Commission is not responsible for any use that may be made of the information contained therein.

Disclaimer This document reflects only the authors' views and the Commission is not responsible for any use that may be made of the information contained therein.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Eger F, Coupek D, Caputo D, Colledani M, Penalva M, Ortiz JA, Freiburger H, Kollegger G (2018) Zero defect manufacturing strategies for reduction of scrap and inspection effort in multi-stage production systems. In: *Procedia CIRP*, vol 67, pp 368–373. 11th CIRP conference on intelligent computation in manufacturing engineering, 19–21, July 2017, Gulf of Naples, Italy
2. Byrne G, Dornfeld D, Inasaki I, Ketteler G, König W, Teti R (1995) Tool condition monitoring TCM - the status of research and industrial application. *CIRP Annals* 44:541,567
3. García-Ordás MT, Alegre-Gutiérrez E, Alaiz-Rodríguez R, González-Castro V (2018) Tool wear monitoring using an online, automatic and low cost system based on local texture. *Mech Syst Signal Process*

4. Dimla DE, Lister PM (2000) On-line metal cutting tool condition monitoring. I: force and vibration analyses. *Int J Mach Tools Manuf*
5. Li X, A brief review: acoustic emission method for tool wear monitoring during turning. *Int J Mach Tools Manuf* (2002)
6. Axinte D, Gindy N (2004) Assessment of the effectiveness of a spindle power signal for tool condition monitoring in machining processes. *Int J Prod Res*
7. Seemuang N, McLeay T, Slatter T (2016) Using spindle noise to monitor tool wear in a turning process. *Int J Adv Manuf Technol* 86:2781–2790
8. Mir A, Luo X, Cheng K, Cox A (2018) Investigation of influence of tool rake angle in single point diamond turning of silicon. *Int J Adv Manuf Technol* 94:2343–2355
9. Lauro CH, Brandão LC, Baldo D, Reis RA, Davim JP (2014), Monitoring and processing signal applied in machining processes - A review
10. Cui X, Zhao B, Jiao F, Ming P (2017) Formation characteristics of the chip and damage equivalent stress of the cutting tool in high-speed intermittent cutting. *Int J Adv Manuf Technol*
11. Ortiz-de Zarate G, Sela A, Saez-de Buruaga M, Cuesta M, Madariaga A, Garay A, Arrazola PJ (2019) Methodology to establish a hybrid model for prediction of cutting forces and chip thickness in orthogonal cutting condition close to broaching. *Int J Adv Manuf Technol*
12. Shao-Hsien L, Zhi-Rong Chen. (2020) Study of using cutting chip color to the tool wear prediction. *Int J Adv Manuf Technol*
13. Wang J, Ma Y, Zhang L, Gao RX, Wu D (2018) Deep learning for smart manufacturing: methods and applications. *J Manuf Syst* 48:144–156
14. Kusiak A (2020) Convolutional and generative adversarial neural networks in manufacturing. *Int J Prod Res* 58:1594–1604
15. Weimer D, Scholz-Reiter B, Shpitalni M (2016) Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection. *CIRP Ann* 65:417–420
16. Sick B (2002), On-line and indirect tool wear monitoring in turning with artificial neural networks. A review of more than a decade of research
17. Huang Z, Zhu J, Lei J, Li X, Tian F (2019) Tool wear predicting based on multisensory raw signals fusion by reshaped time series convolutional neural network in manufacturing. *IEEE Access* 7:178640–178651
18. Belhadi S, Mabrouki T, Rigal J-F, Boulanouar L (2005) Experimental and numerical study of chip formation during straight turning of hardened AISI 4340 steel. *Proc Inst Mech Eng B* 219(7):515–524
19. da Silva LR, da Silva OS, dos Santos FV, Duarte FJ, Veloso GV (2019) Wear mechanisms of cutting tools in high-speed turning of ti6al4v alloy. *Int J Adv Manuf Technol* 103:37–48
20. Chuangwen X, Ting X, Huaiyuan L, Zhicheng S, Hongbing J, Mandong L (2017) Friction, wear, and cutting tests on 022cr17ni12mo2 stainless steel under minimum quantity lubrication conditions. *Int J Adv Manuf Technol* 90:677–689
21. Toh C (2005) Comparison of chip surface temperature between up and down milling orientations in high speed rough milling of hardened steel. *J Mater Process Technol* 167:110–118
22. *Metals Handbook ASM* (1989) *ASM Handbook volume 16: Machining*
23. The GIMP Development Team (2019) *Gimp*
24. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 3431–3440
25. Goodfellow I, Bengio Y, Courville A (2016) *Deep learning*. MIT Press, Cambridge. <http://www.deeplearningbook.org>
26. Fuchs K, Gertheiss J, Tutz G (2015) Nearest neighbor ensembles for functional data with interpretable feature selection. *Chemometr Intell Lab Syst* 146:186–197
27. Gonzalez RC, Woods RE (2006) *Digital image processing*, 3rd Edition. Prentice-Hall, Upper Saddle River
28. Peter HD (1985) Kernel estimation of a distribution function. *Commun. Stat. - Theory Methods* 14:605–620
29. Parzen E (1962) On estimation of a probability density function and mode. *Ann. Math. Statist.* 33:1065–1076
30. Gurney K (1997) *An introduction to neural networks*. Taylor and Francis, Bristol
31. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jia Y, Jozefowicz R, Kaiser L, Kudlur M, Levenberg J, Mané D, Monga R, Moore S, Murray D, Olah C, Schuster M, Shlens J, Steiner B, Sutskever I, Talwar K, Tucker P, Vanhoucke V, Vasudevan V, Viégas F, Vinyals O, Warden P, Wattenberg M, Wicke M, Yu Y, Zheng X (2015) TensorFlow: large-scale machine learning on heterogeneous systems. Softwareavailablefromtensorflow.org
32. Ramsden J, Allen D, Stephenson D, Alcock J, Peggs G, Fuller G, Goch G (2007) The design and manufacture of biomedical surfaces. *CIRP Ann* 56:687–711
33. Johnson RA, Wichern DW et al (2002) *Applied multivariate statistical analysis*, vol 5. Prentice Hall, Upper Saddle River
34. Ferraty F, Vieu P (2006) *Nonparametric functional data analysis: theory and practice*. Springer, Berlin
35. R Core Team (2018) *R: a language and environment for statistical computing*. R foundation for statistical computing, Vienna, Austria
36. Maierhofer T, Pfisterer F (2018) *ClassiFunc: classification of functional data*. R package version 0.1.1.
37. Bowman AW, Azzalini A (1997) *Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations*, vol 18. OUP, Oxford

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Luca Pagani¹ · Paolo Parenti² · Salvatore Cataldo² · Paul J. Scott¹ · Massimiliano Annoni²

¹ EPSRC Future Metrology Hub, University of Huddersfield, Huddersfield, Queensgate., HD13DH, UK

² Department of Mechanical Engineering, Politecnico di Milano, via La Masa 1, 20156, Milan, Italy