Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 99 (2021) 502–507

www.elsevier.com/locate/procedia

14th CIRP Conference on Intelligent Computation in Manufacturing Engineering, CIRP ICME 2020

Models for emulating the human quality assessment of laser cut edges

Leonie Tatzel[a,b,*], Fernando Puente León[a]

[a]Karlsruhe Institute of Technology: Institute of Industrial Information Technology, Hertzstr. 16, 76187 Karlsruhe, Germany
[b]TRUMPF Werkzeugmaschinen GmbH + Co. KG, Johann-Maus-Str. 2, 71254 Ditzingen, Germany

* Corresponding author. Tel.: +49 7156 303-33775. *E-mail address:* Leonie.Tatzel@kit.edu

Abstract

With the increasing degree of automation in production, the ability to automatically assess and monitor the process quality is gaining in importance. In laser cutting, however, the quality of the cut edge is usually evaluated by humans. This paper aims to emulate their quality assessment based on objective criteria and to expand the understanding of which properties of the edge are relevant.
For this purpose, we carried out an expert survey: five experts rated the quality of 100 laser cut edges. Additionally, numerous objective features were measured for each edge. Based on this data, two models were developed that mimic the human quality assessment very well.

Keywords: Laser cutting; Edge quality; Expert survey; Machine learning

1. Introduction

Laser cutting is a complex process with numerous influencing factors. As some of them (e.g. the condition of the machine) are subject to change, the process quality needs to be checked regularly. A clearly defined concept of quality is the basis for every process optimization.

According to the ISO 9013 standard [1], two characteristic values describe the quality of a thermal cut edge: the mean height of the profile *Rz5* and the perpendicularity tolerance *u*. In the literature, quality is often associated with roughness: with *Rz* in [2] or with *Ra* in [3].

In practice the machine operator assesses the quality manually by looking at and touching the edge. This method is difficult to integrate into increasingly automated production processes. In addition, human quality assessments are always subjective and depend on the person and on the form of the day. It would be advantageous to automatically assess and monitor the quality based on objective, measurable criteria. To this end, we developed two data-based models: one for the surface of the cut edge and one for the burr. We aim to emulate the human perception and to answer the question whether *Rz5* and *u* sufficiently characterize the quality, or if additional features are needed.

There are numerous works that explore the human perception of quality, but most of them deal with media (e.g. the quality of videos [4]) or consumer products (e.g. sport shoes [5]) and not with manufacturing processes.

This paper is structured as follows. First, we explain how we generated the database of human quality ratings and objective features for 100 different cut edges. Then we present the model approach and the method to reduce the number of input variables. In section 4 we analyse the quality of the data and the performance of the models.

2. Generation of the database

Our database consists of 3 mm thick stainless steel edges that were cut with a state-of-the-art laser cutting machine (TruLaser fiber 5030). In the laser cutting machine, the beam is guided to the cutting head, then it passes through several focusing optics and exits the nozzle together with a coaxial gas stream. The focused beam melts or evaporates the underlying material; the gas jet helps to remove the melt from the kerf [6].

2212-8271 © 2021 The Authors. Published by Elsevier B.V.

10.1016/j.procir.2021.03.108

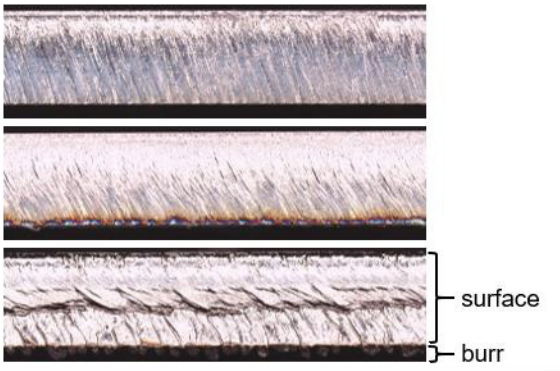

Fig. 1. Exemplary laser cut edges in a 3 mm thick stainless steel sheet metal.

In order to generate the entire bandwidth, from very good to very bad edges, four dominant process parameters were varied fully factorially. These are feed rate (13 to 29 m/min), gas pressure (9 to 21 bar), distance between the nozzle and the beam focus (−3.5 to −0.5 mm) and distance between the nozzle and the sheet metal (0.5 to 3 mm).

Less the miscuts (when the sheet cannot be separated due to wrong process parameters) this results in 834 cut edges. Three of them are shown in Fig. 1. In the following, we always distinguish between the surface of the edge and the burr.

2.1. Objective criteria

In order to obtain objective features, the cut edges were measured with an optical measuring system (3D profilometer VR-3200, Keyence Corporation). An RGB image and a 3D point cloud (height topography) were taken of each edge. The RGB image was used to determine the degree of burnt or oxidized area. From the 3D point cloud 1254 features were extracted for the cut surface and 20 for the burr. In the following, only those features will be discussed that proved to be particularly relevant. A complete list can be found in Appendix A.

According to the ISO 9013 standard [1] the quality of a thermal cut is described by the mean height of the roughness profile *Rz5* and by the perpendicularity tolerance *u*. *Rz5* indicates the absolute vertical distance between the highest profile peak and the deepest profile valley along the sampling length. *u* is the distance of two parallel straight lines between which the cut profile is within the theoretical angle (here: 90°). In this work, both, *Rz5* and *u* were determined on the basis of ISO 9013 [1] and ISO 4288 [7], except that instead of a stylus instrument the optical measurement system was used.

2.2. Expert survey

The rating of the cut edge quality cannot be done by laypersons but only by experienced experts. Due to their limited availability we selected 100 of the 834 samples for the survey. In order to make a representative selection of different edge qualities we used the *Rz5* values. In Fig. 2 the histogram of the whole database is shown.

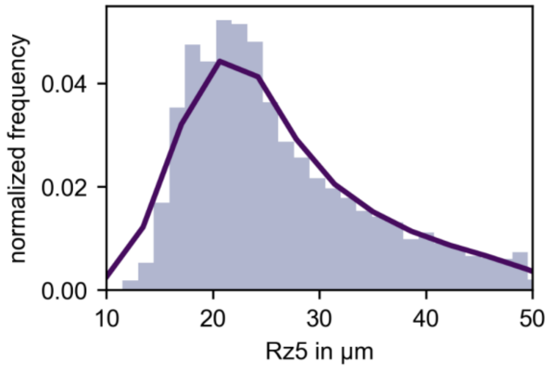

Fig. 2. *Rz5* distribution of the whole database containing 834 samples that were cut with different process parameter combinations; 18 % of the samples are not displayed in the graph, they have roughness values above 50 µm, scattered over a wide range.

The *Rz5* range was divided into 10 intervals, each containing 10 % of the samples, and from each interval 10 edges were selected randomly.

Five experts evaluated the selected edges. They could touch them, so they were able to evaluate the quality visually and haptically as they are used to. Two separate ratings should be given for each edge: one for the surface and one for the burr. They should each be assigned to one of six quality classes: 1 (perfect), 2 (good), 3 (satisfactory), 4 (inferior), 5 (bad) or 6 (very bad).

3. Models and methods

Based on this, two models were developed to emulate the human perception of the cut edge quality: the measured features served as input data and the human ratings were used as labels.

3.1. Preprocessing of the data

Since we aimed to train consensus models, the five individual ratings of each edge were combined, and the median was used as the label. To compensate the imbalanced class distribution (see Fig. 4 and Fig. 6), the majority class(es) were downsampled (downsampling in this context means using a subset of the class) and class 5 and 6, which are both representing unacceptable quality, were merged for the cut surface. In this way, the six-class problem was transformed into a five-class problem.

3.2. Ordinal classification with binary classification algorithm

There is an ordinal relationship between the classes. To take this into account and still be able to use a simple binary classifier, the approach described by Frank and Hall [8] was chosen: the five-class (resp. six-class) classification problem was transformed into four (resp. five) binary classifications: *target>1*, *target>2*, *target>3*, *target>4* (and *target>5*). After training the binary classifiers, their probabilities *Pr* were used to predict the actual class (*c*):

$$Pr(c = k) = Pr(target > k - 1) - Pr(target > k) \quad (1)$$

with $k = 1, 2, 3, 4, 5$ for the surface and $k = 1, 2, 3, 4, 5, 6$ for the burr. It is $Pr(target > 0) = 1$ and $Pr(target > 5) = 0$ resp. $Pr(target > 6) = 0$. Each of the binary classifiers was modelled with the logistic regression function:

$$\Pr(target > k) = \frac{e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n)}} \quad (2)$$

with n input features $X = (X_1, X_2, \ldots, X_n)$ and $n + 1$ coefficients $\beta_0, \beta_1, \beta_2, \ldots \beta_n$ [9].

3.3. Feature selection

As for the cut surface there are 1254 features but only 79 samples after preprocessing, a subset of the features had to be selected to reduce the model complexity and to avoid the curse of dimensionality [10]. To select the best subset, we used a hybrid approach of forward and backward stepwise selection. The basic idea of forward stepwise selection is to start with a model containing no features. Then the features are included, one at a time, until all of them are included or until the process is stopped prematurely. At each step, the feature that gives the largest additional improvement to the model performance is chosen. Backward stepwise selection works the other way around. Hybrid approaches combine both methods: features are added to the model one by one, but there is also the option to remove features that no longer improve the model performance [9].

3.4. Implementation details

To test the model cross-validation was applied: the data was split into four folds while preserving the percentage of samples for each class. Three folds were used to perform the feature selection and to train the model, the fourth fold was used for testing.

We implemented the binary classifiers with the Python library Scikit-learn [11] and the sequential forward floating feature selector with the Python library Mlxtend [12]. The number of selectable features was limited to 15 and 10 for the surface and the burr model, respectively. Unless otherwise specified, the default parameters were used.

3.5. Evaluation metrics

The confusion matrix [9] and the f1-score, which is the harmonic mean of precision and recall [13], are used to evaluate the performance of the models. These metrics only give information about whether a sample was classified correctly or incorrectly, but not about how far the predicted class was from the true one. So, we additionally plot the number of incorrectly classified samples over the distance between predicted and true class.

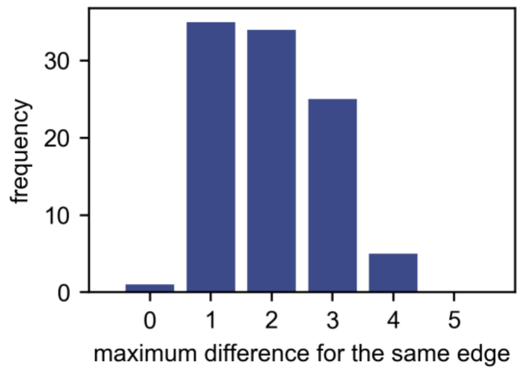

Fig. 3. Maximum difference between the expert ratings for the cut surface quality of the same edge.

4. Results and discussion

4.1. Expert ratings of the cut surface quality

There are five ratings (from the five experts) for each of the 100 edges. The experts were mostly but not always in agreement. In Fig. 3 the distribution of the maximum difference (difference of the best and worst grade) between the five ratings for the same edge is displayed. Only for one sample all five persons gave the same grade, for 69 edges the maximum difference is one to two classes, for 25 edges three and for 5 edges four classes. It is not possible to identify one expert who is, for example, significantly more critical than the others. The discrepancies seem to occur randomly. Differences of one to two classes are not serious, because in practice it is not necessary to differentiate between six quality levels, but differences of more than three classes are problematic. Here it becomes clear how much the individual perceptions of quality vary. It can be concluded that the process quality should generally not be monitored by a single person. Hereafter the median of the five ratings is used as the label.

The Pearson correlation coefficient between *Rz5* and the class label is 0.6 and between *u* and the class label it is 0.5. Thus, there is some relation, but the correlation is too low to assume that these two characteristics suffice to fully describe the quality. This is also indicated by the following result: the experts stated the main reason for their decision in their own words. In summary, there are four categories: homogeneity (34 % of the nominations), roughness (32 %), discoloration (18 %) and edge slope (14 %). Homogeneity was nominated more often than roughness (*Rz5*) and discoloration was more important than edge slope (*u*).

In Fig. 4 the distribution of the classes is shown. Most edges are of good or medium quality: there are 54 instances of class 1 and 2, 34 instances of class 3 and 4 and only 12 instances of class 5 and 6. To obtain a less imbalanced dataset, class 2 was downsampled by a factor of 2 and class 5 and 6 (both representing inacceptable quality) were merged. The resulting distribution is also displayed in Fig. 4.

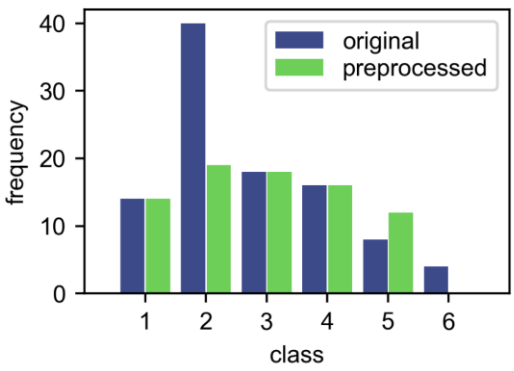

Fig. 4. Class distribution of cut surface quality before (original) and after downsampling of class 2 and combining class 5 and 6 (preprocessed).

4.2. Prediction of cut surface quality

4.2.1. Model performance

To emulate the (averaged) human quality rating the model (see section 3.2) was trained on three quarters of the data and tested on the remaining one. The confusion matrix of the model is shown in Table 1. In most cases prediction and label match or differ by only one class. The error is never greater than two classes.

Table 1. Confusion matrix of the model for the prediction of cut surface quality.

		predicted				
		1	2	3	4	5
true	1	12	1	1	0	0
	2	2	13	3	1	0
	3	1	2	12	3	0
	4	0	0	2	11	3
	5	0	0	1	1	10

The weighted f1-score of the model is 0.73. It indicates how many samples were classified correctly and how many were classified incorrectly. However, it does not give any information about the distance between predicted and true class. Therefore, in Fig. 5 we show how often and by how many classes the model and the individual experts deviate from the label.

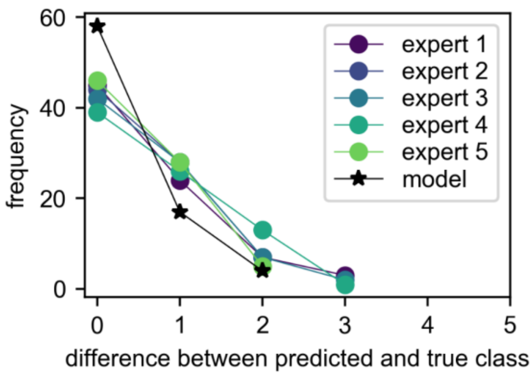

Fig. 5. Deviation of model and experts from the label for the cut surface.

This graph clearly demonstrates that the model represents the average quality better than any of the five individuals.

4.2.2. Selected features

The model is based on only 11 of the 1254 available surface features. Including more than 11 features hardly improves the performance. The selected features are listed with additional information in Table 2.

The first seven of the selected features are deduced from the line roughness (LR) profile, one from the surface roughness (SR) measurement and two from the waviness profile (W). The calculation of burnt_perc is not based on the 3D point cloud but on the RGB image.

Rt is particularly sensitive to outliers and might therefore be associated with individual defects e.g. splatters. *Rk*, *Rvk* and *Mr1* are derived from the Abbott curve, which is the cumulative probability density function of the profile height [14]. They could be related to the homogeneity of the edge. The same applies to *R*Δ*q*, which describes the mean slope of the profile. *RSm* is a measure of the width of the grooves and might be related to the roughness. The roughness *Rz* (*Rz5*) is also among the selected features.

Vvc is the only feature that contains information about the entire surface topography. All other features are deduced from measurement lines parallel to the direction of cutting. No feature was selected that is directly related to the perpendicularity of the edge. This coincides with the small number of nominations by the experts (see subsection 4.1). The slope of the edge is probably neither important nor clearly visible in a 3 mm thick sheet metal.

All selected features were deduced from measuring lines in the middle and lower area of the edge. This might indicate that the experts are particularly concerned about these areas.

Table 2. Selected features to predict the cut surface quality; LR: line roughness, SR: surface roughness, W: waviness; the line indicates the position of the measurement relative to the topside of the edge in mm, e.g. *Rz* on line 1.5 was measured in the middle of the 3 mm thick sheet metal.

name	description	measurement	line	ref.
Rt	total height of profile	LR	2.7	[15]
Rk	core height	LR	2.7	[16]
Rvk	reduced dale height	LR	2.7	[16]
RSm	mean width of profile elements	LR	1.2	[15]
Rz	maximum height of profile	LR	1.5	[15]
*R*Δ*q*	root mean square tilt angle	LR	1.5	[15]
Mr1	load length ratio that separates profile peak area and core	LR	1.8	[16]
Vvc	dale void volume	SR	-	[17]
Wsk	skewness	W	1.2	[15]
Wa	arithmetical mean height	W	0.9	[15]
burnt_perc	percentage of coloured pixels	RGB	-	-

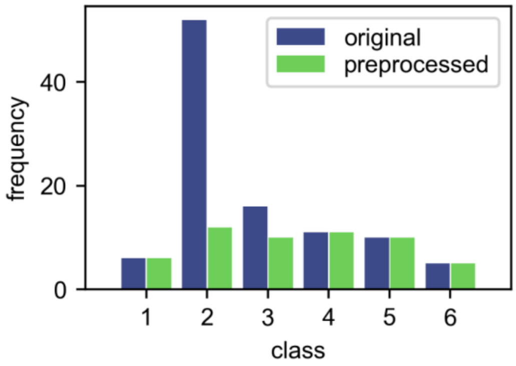

Fig. 6. Class distribution of burr quality before and after downsampling of class 2 and 3.

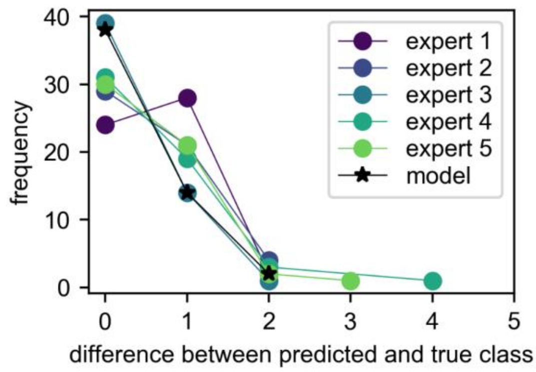

Fig. 7. Deviation of model and experts from the labels for the burr.

4.3. Expert ratings of the burr quality

The differences between the individual ratings of the burr quality for the same edge are smaller than those of the cut quality. For 94 % of the samples the maximum difference of the grades is two classes or less. Since the burr quality classes are even more unequally distributed than the surface classes, class 2 was downsampled by 75 %, class 3 by 30 %. The original distribution and the one after downsampling are displayed in Fig. 6.

4.4. Prediction of burr quality

4.4.1. Model performance

The confusion matrix of the burr quality model is shown in Table 3. Most of the samples are classified correctly, except for class 3. This could be caused by ambiguous data, because the experts might find it difficult to differentiate between the grades 2 and 3. The (weighted) f1-score is 0.69.

Table 3. Confusion matrix of the model for the prediction of burr quality.

		predicted					
		1	2	3	4	5	6
true	1	5	1	0	0	0	0
	2	3	9	0	0	0	0
	3	1	5	3	1	0	0
	4	0	1	0	7	3	0
	5	0	0	0	0	9	1
	6	0	0	0	0	0	5

The model reproduces the average of the expert ratings better than four of the five the individual persons. Only the ratings of expert 3 are closer to the label as it is displayed in Fig. 7.

4.4.2. Selected features

Unlike the model for the cut surface, which is based on 11 (out of 1254) features, only 4 (out of 20) features are needed here. Afterwards the model performance stagnates. The selected features are the median of the burr height, the median of the burr width and the mean and the standard deviation of the burr density. The so-called density indicates whether the burr consists of individual beads or is rather constantly high. Since the burr is a much simpler structure than the surface of the cut edge, it is plausible that fewer features contain all necessary information.

5. Conclusion

We developed two data-based models, one for the cut surface and one for the burr, that emulate the human quality assessment of laser cut edges very well. To deal with the small and imbalanced database, we greatly reduced the number of input features and used a simple algorithm that included the ordinal relationship of the classes.

Both models provide a more reliable statement than a single expert. The large deviations of the expert ratings for the same edge indicate that it is generally not recommended to let a single person monitor the process quality.

The results also show that, in order to fully describe the quality of a laser cut edge, it is necessary to consider more features than *Rz5* and *u*. This should also be taken into account when modelling and optimizing the laser cutting process.

The selected features give an impression of important quality-relevant properties, but the database on which the models were trained is very small. More experts from different industries should be interviewed to get a more reliable average. In addition, different processes (other materials, various sheet metal thicknesses) need to be included to find out how generalizable the models are.

Appendix A.

A.1. Line roughness (R) / waviness (W)

name	description
Ra, Wa	arithmetical mean height
Rz, Wz	maximum height of profile
Rp, Wp	maximum profile peak height
Rv, Wv	maximum profile valley depth
Rc, Wc	mean height of profile elements
Rt, Wt	total height of profile
Rq, Wq	root mean square deviation
Rsk, Wsk	skewness
Rku, Wku	kurtosis

name	description
RSm, WSm	mean width of profile elements
R∆q, W∆q	root mean square slope
Rmr(c), Wmr(c)	load length ratio
Rδc, Wδc	profile cut level distance
Rmr, Wmr	relative load length ratio
RzJIS	cross point average roughness
Rk	core height
Mr1	load length ratio that separates the profile peak area and core
Mr2	load length ratio that separates the profile peak area and core
Rpk	reduced peak height
Rvk	reduced dale height
HSC	high spot count
Pc/cm	peak count/cm
RPc, WPc	peak count
RLo, WLo	expansion length
Rlr, Wlr	expansion length ratio
R⊿a, W⊿a	arithmetic mean tilt angle
Rλa, Wλa	arithmetic mean wavelength
R⊿q, W⊿q	root mean square tilt angle
Rλq, Wλq	root mean square wavelength

A.2. Surface roughness

name	description
Sa	arithmetical mean height
Sz	maximum height of profile
Str	texture aspect ratio
Sq	root mean square height
Ssk	skewness
Sku	kurtosis
Sp	maximum profile peak height
Sv	maximum profile valley depth
Sal	autocorrelation length
Std	texture direction
Sdq	root mean square gradient
Sdr	developed interfacial area ratio
Spd	density of peaks
Spc	arithmetic mean peak curvature
Sk	core height
Spk	reduced peak height
Svk	reduced dale height
Smr1	upper material ratio
Smr2	lower material ratio
Sxp	peak extreme height

name	description
Vvv	dale void volume
Vvc	core void volume
Vmp	peak material volume
Vmc	core material volume

A.3. Burr and perpendicularity tolerance

name	description
burr height	height of the burr
burr width	width of the burr
burr density	burr continuous or single beads
u	perpendicularity tolerance

The features listed here were each measured at different positions on the edge. For more information see [15], [16], [17].

References

[1] DIN ISO 9013. Thermal cutting - Classification of thermal cuts - Geometrical product specification and quality tolerances; 2017.

[2] Tercan H, Khawli TA, Eppelt U, Büscher C, Meisen T, Jeschke S. Improving the laser cutting process design by machine learning techniques. Prod Engineer 2017; 11: 195–203.

[3] Sobih M, Crouse PL, Li L. Striation-free fibre laser cutting of mild steel sheets. Appl Phys 2007; 90: 171–4.

[4] Webster AA, Jones CT, Pinson MH, Voran SD, Wolf S. Objective video quality assessment system based on human perception. In: Proc SPIE. San Jose, USA; 1993.

[5] Tsiotsou R. The role of perceived product quality and overall satisfaction on purchase intentions. Int J Consum Stud 2006; 30: 207–17.

[6] Steen WM, Mazumder J. Laser Material Processing, 4th ed. London: Springer; 2010.

[7] DIN ISO 4288. Geometrical Product Specifications (GPS) Surface texture: Profile method Rules and procedures for the assessment of surface texture; 1997.

[8] Frank E, Hall M. A Simple Approach to Ordinal Classification. In: De Raedt L, Flach P, editors. Machine Learning: ECML 2001. Freiburg, Germany: Springer; 2001.

[9] James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning: With Applications in R. 1st ed. New York: Springer; 2017.

[10] Bellman R, Collection KMR. Adaptive Control Processes: A Guided Tour. Princeton, USA: Princeton University Press; 1961.

[11] Pedregosa F, Varoquaux G, Gramfort A et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res 2011; 12: 2825–30.

[12] Raschka S. MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. JOSS 2018; 3: 638.
[13] Kubat M. An Introduction to Machine Learning. 1st ed. Cham: Springer; 2015.

[14] Abbott EJ, Firestone A. Specifying Surface Quality—A Method Based on Accurate Measurement and Comparison. J Mech Eng 1933; 55: 569–72.

[15] DIN ISO 4287. Geometrical Product Specifications (GPS) – Surface texture: Profile method – Terms, definitions and surface texture parameters; 1996.

[16] DIN ISO 13565-1. Geometrical Product Specifications (GPS) – Surface texture: Profile method; Surfaces having stratified functional properties – Part 1; 1997.

[17] DIN ISO 25178-2. Geometrical product specifications (GPS) - Surface texture: Areal - Part 2; 2012.