



Application of Random Forest Algorithm for the Quality Determination of Manufactured Surfaces

Marcin Hinz^(✉), Jannis Pietruschka, and Stefan Bracke

Bergische Universität Wuppertal, Wuppertal, Germany
m.hinz@uni-wuppertal.de

Abstract. The optical perception of high precision, fine grinded surfaces is an important quality feature for these products. Its manufacturing process is rather complex and depends on a variety of process parameters (e.g. feed rate, cutting speed) which have a direct impact on the surface topography.

To improve the conventional methods of condition monitoring, a new image processing analysis approach is needed to get a faster and more cost-effective analysis of produced surfaces. For this reason, different optical techniques based on image analysis have been developed over the past years. Fine grinded surface images have been generated under constant boundary conditions in a test rig built up in a lab.

Within this study the image of each grinded surface is analyzed regarding its measured arithmetic average roughness value (Ra) by the use of random forest algorithms.

Keywords: Machine learning · Random forest · Computer vision · Condition monitoring

1 Introduction

The surface topography as well as the optical perception are important features for evaluating the quality of fine grinded knives. Parameters as the surface roughness, gloss or coloring are used for the quantification of these features. The measuring is implemented by the use of traditional methods, which are manual, time-consuming and cost-intensive. On top of that, the application of these methods for the condition monitoring of the ongoing process is rather limited. Therefore, a new, faster and more cost-effective approach is needed to improve the classical measurement methods. A conceivable approach could be based on image analysis.

Over the past years, different contactless image analysis based approaches have been developed to simplify the traditional roughness measurement methods. Some studies propose picture pre-processing and feature extraction in combination with machine learning algorithms. Neural Networks (NN) based methods are presented and tested on specific surfaces, as well (cf. [1–3]).

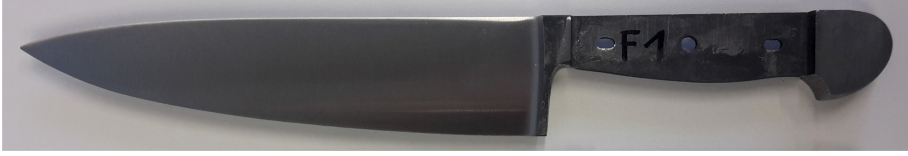


Fig. 1. An example image of a chef knife analyzed within this research activities

The overall goal of the presented research activities is the development of a condition monitoring tool which can be implemented in the ongoing grinding process of the knives. It should be used to ensure the knives quality and to reduce rejects by an immediate detection of deviations of the target values and the possibility to adapt the production process accordingly. For this reason, a data set based on cutlery samples has been generated and analyzed. During previous research activities, the extraction of features of the data set is presented. The features are used to train various machine learning algorithms with and without a combination of logged process parameters to evaluate the surface roughness (cf. [4, 5]). Another roughness prediction possibilities by the use of neural networks (NN) are presented in [6] and [7]. In this paper, the focus is set on the evaluation of the surface roughness (Ra) based on NNs. Therefore, a set of 851 chef's knives is used. An exemplary knife is shown in Fig. 1.

2 Data Generation

Within the presented research activities the analyzed data set embraces photographs of the surfaces of 851 8'' chef's knives (cf. Fig. 1). The surface images are taken within an experimental test rig which provides constant boundary conditions (cf. Sect. 2.2). Reference measurements of the surface roughness as well as gloss and coloring are taken of all the knives (cf. Sect. 2.1).

2.1 Reference Measurements

To describe the surface topography of the knife blades the surface roughness (Ra, Rq, Rz, Rt), gloss (GU) and coloring (CIELAB) are measured with traditional measurement methods. The roughness measurements are provided with the device PCE-RT11 over a sampling length of 6 mm. The arithmetic average roughness (Ra) can be calculated with Eq. (1). Here, $z(x)$ is the surface profile and l describes the length of the measured section (cf. [4, 8, 9]).

$$Ra = \frac{1}{l} \int_0^l |z(x)| dx \quad (1)$$

Since the Ra value is commonly used to define rejection limits in the regarded process, this parameter is used as the target value in this study. The measured Ra value of the data set varies between 0.07 μm and 0.47 μm .

2.2 Experimental Setup

In order to take comparable images of different knives, a test rig that provides consistent lighting conditions was designed and build up. The test rig is based on a design with white inside-walls, to keep ambient light outside and diffuse the light inside to prevent reflections on the knife's surface. The knives are mounted on a 3D printed fixation to ensure a proper positioning of each knife (Fig. 2).

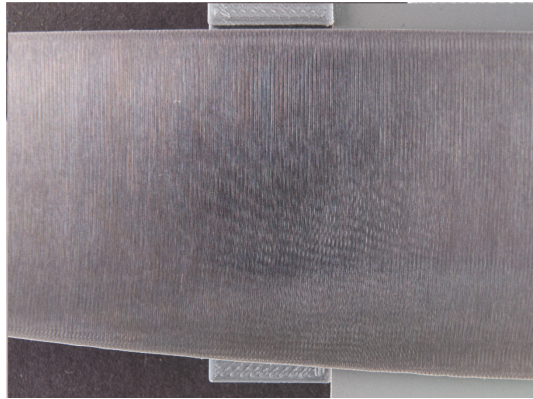


Fig. 2. An exemplary image of the photographed knife section

Two LED Spotlights, pointed at the walls on the top- and bottom ends of the knife, provide constant, diffuse light conditions inside the box. This light arrangement accentuates the characteristic marks left by the grinding process that appear in a 90° angle to the knife's length. The photos are taken by an Olympus E-520 DSLR, equipped with an Olympus Zuiko Digital 14–42 mm f 1:3.5–5.6 lens. The lens is placed normal to the knife's surface at a distance of 8 cm. Finally, to prevent reflections, a black screen at the height of 15 cm covers the camera body and the white ceiling, leaving only the lens visible from below. For a comprehensive description of the test rig cf. [4, 5].

3 Computer Vision Based Feature Extraction

Computer Vision (CV) describes the ability of perception of optical data by a computer. Since the investigation of optical data is extremely complex, picture pre-processing can be used to reduce the amount of information which will be studied to the most important characteristics for the available analysis task. The selection and application of appropriate pre-processing methods increase the quality and accuracy of the research. Therefore, within this study CV is used for analyzing the pictures of the knife surfaces and for the extraction of relevant features out of these images [1, 2, 10, 15].

As a first step the part of the knives, where the reference measurements had been taken, has to be identified on the pictures. In this way, scattering of the surface topography and inconsistent lightning conditions on the edge of the focused range don't affect the

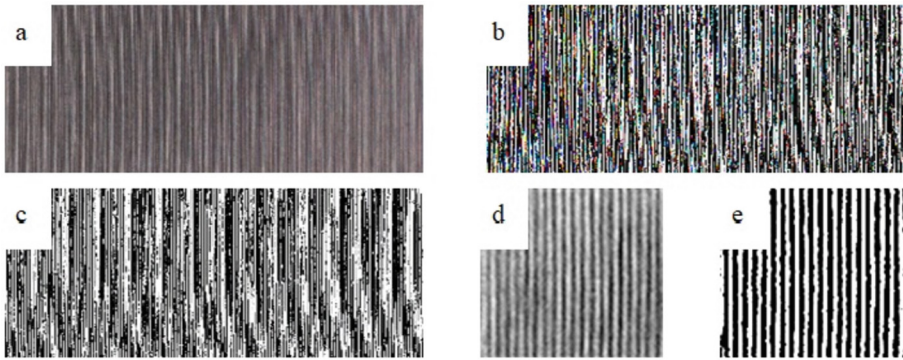


Fig. 3. (a) cropped image (b) sobel operator (c) contrast change and sobel operator (d) low pass mask and median blur (e) low pass mask, median blur and sobel operator

results of the further analysis. The images are cropped to a size of $1250 \text{ px} \times 550 \text{ px}$, which corresponds to an area of $2,5 \text{ cm} \times 2,0 \text{ cm}$.

The knife surfaces have a grooved structure, as it can be seen in Fig. 3(a). On all of the images the creases are rising vertically and parallel, but they differ in terms of the creases' width, depth and quantity along the considered picture section. As a result, information to determine these parameters will be detected within this research of the analysis of the surface roughness based on CV. Because the roughness is measured orthogonal to the creases, it makes sense to extract the features in the same way. The grooves are not consistent along their length. For this reason, ten uniformly distributed lines are drawn over the image height and the features are detected along [4, 8].

In order to get the best results, appropriate picture pre-processing filters and methods are utilized to reduce and adjust the kind of information which is inspected. The choices of the type of filters and their specification were made on base of researches and trials on the data base. For one part of the pre-processing the contrast of the images is changed. Besides that, low and high frequencies are filtered with approved filters. For the purpose of sharp separations between the individual grooves, the Sobel filter for the x-dimension with a kernel size of five is used. It is a well-known high pass edge detecting filter by which the kernel gets convolved with the image. As low pass filter, a filtering mask is used which is applied to the frequency domain of the image. On top of that, the median blur filter with a kernel size of three is selected to eliminate disruptive pixels by the comparison with its neighbors. Figure 3 shows the original and the pre-processed pictures [1–4, 11].

Subsequently, features are extracted. The creases' width and quantity are extracted from:

- the original image,
- the picture with the Sobel filter,
- the photo after the change of the contrast and the Sobel filter,
- the result after the application of the low pass mask and the median blur filter,
- the image with the use of the frequency mask, the median blur filter and the Sobel filter.

In addition to the absolute values, the averages and standard deviation of the creases' width of each line as well as their minimum, maximum and range are calculated. Between the results of each pre-processing combination obvious differences can be detected. Because the pictures are two-dimensional the grooves' depth cannot be gathered directly. Therefore, the lightness of each pixel is extracted by the use of the L*-value of the CIELAB color space. Since its course over the image width is comparable to a roughness profile, the course is used to determine parameters with the formulas of the roughness values (Ra, Rq, Rz, Rt). These parameters are calculated over the whole image width, over the width of the sampling length, and without the consideration of the edges of the images since these areas tend to show changing lightning conditions. That provides 42 features in total which are used within the following analysis:

- the number of creases,
- its average width and the associated standard deviation,
- the minimum and maximum width as well as the range of the crease's width from five differently pre-processed images (all above amount to 30 in total),
- roughness value formula calculated parameters (12 in total). cf. [5]

Exemplarily and according to [4], the average outputs of one line for two images can be found in the following Table 1.

Table 1. Results of the feature extraction on the basis of two exemplary picture for one line each

Filter	Image A		Image B	
	Average width of creases – line 1/pixels	Quantity of creases – line 1	Average width of creases – line 1/pixels	Quantity of creases – line 1
Original	4.34	287	3.70	335
Sobel operator	3.06	406	3.48	357
Contrast change	3.09	402	3.40	365
+ Sobel operator	14.77	31	14.32	37
DFT + median blur	13.82	39	12.71	42
DFT + median blur + Sobel operator				
R _a	0.11 μm		0.16 μm	
R _z	0.88 μm		1.56 μm	

It is obvious that all pre-processing stages result in different values. It is possible to detect less than 300 creases along one line from Image A out of the original picture. If the grooves are counted by human eye due to the verification purpose, the sum will be more than 400. That correlates with the result of the Sobel operator. Whether the contrast change is needed or not, depends on the individual photo. That is the reason

why both results are used. The comparison of the groove width outlines the undetected edges of the original picture, as they are measured about one and a half time as wide in the original picture as in the filtered one. Apparently, the use of the discrete Fourier transformation causes the blur of several grooves into one. The identification works even better after edge detecting processes. By comparing the results of image A and B, in consideration of the references of their measured surface roughness, differences can be seen. On the one hand, it seems that there are more grooves on surfaces with a lower Ra value after the use of a high pass filter. On the other hand, the low pass filter causes less creases for lower surface roughness.

In general terms, regarding the results of this feature extraction, it can be stated that there are analogies between the measured roughness values and the extracted ones.

4 Random Forest Theory

In the previous subsection, data base and data generation were described. The machine learning model used for this study is the random forest which was first published in [12]. This Algorithm is a widely used forecasting tool and can be assigned to the learning technique of supervised learning. The application of the random forest relates to the classification and regression of data.

Random forests are a combination of tree predictors, also known as decision trees, such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest [12]. These types of algorithms generate classifiers or regressors in form of a decision trees by synthesizing a model based on a tree structure. The C4.5, one of the most known decision trees, first published in [13], is one of the most known and well renowned decision tree algorithms. Generally, it uses a divide-and-conquer method to grow an initial tree based on recursively partitioned data sets with the splitting criterion of entropy.

Entropy, or information entropy, is a representation of how much information is encoded by given data. Measured by the highest information gain, the data is split into partitioned data sets. At each node of a decision tree, the entropy belonging to a class membership is given by the formula:

$$info(S) = - \sum_{j=1}^k \frac{freq(C_j, S)}{|S|} \times \log_2 \left(\frac{freq(C_j, S)}{|S|} \right) bits \quad (2)$$

where $freq(C_j, S)$ is the number of cases in S that belong to the class C_j . By applying formula 2 to a set of training classes, $info(TT)$ measures the average amount of information needed to identify the class in the case of TT as follows:

$$info_x(TT) = \sum_{i=1}^n \frac{|TT_i|}{|TT|} \times info(TT_i) \quad (3)$$

And

$$gain(X) = info(TT) - info_x(TT) \quad (4)$$

with the outcomes of a test X . A disadvantage of decision trees is that they are prone to overfitting. As a result of overfitting the model becomes too catered towards the training data set and performs poorly on testing data. This will lead to low general predictive accuracy [14]. Whereas combining decision trees in a random forest circumvents this problem. To create the forests, the algorithm uses various methods of ensemble learning. The Bagging is a method for generating multiple versions of a predictor and using these to get an aggregated predictor. The aggregation averages over the versions when predicting a numerical outcome and does a plurality vote when predicting a class. The multiple versions are formed by making bootstrap replicates of the learning set and using these as new learning sets. This method adds some of embedded test data set to the model, which improves the accuracy on unknown data [12]. Besides bootstrapping, random forest using randomly selected subsets of features for creating slightly different tree structures. The simplest random forest with random features is formed by selecting randomly, at each node, a small group of input variables to split on. Consequently, the following properties are the result of the random forest improvements:

- It has been considered as very good algorithm in accuracy, also with noisy data
- It is much more efficient on huge data sets without overfitting
- Works fine without a lot of hyper parameter tuning
- Very good generalization ability
- Can process numerical data as well as strings

Another property of the algorithm should be mentioned separately. In the combination of trees, it is possible to see the importance of the individual features calculated by the algorithm. The Importance of the different features are measured by the mean decrease of impurity. The meaning of impurity in this context is the information content in relation to the forecast of the data. Therefore, the computed entropies of the features of all trees are summarized and averaged. With the insight into the importance of the features, it is possible to identify particularly relevant features in the data set. This method is used by for the reduction of input parameters presented in the following section.

5 Feature Reduction

Based on the mentioned impurities, the feature/input parameter reduction can be performed. For this purpose, the data set is split into two subsets consisting of production parameters (PP) and based on CV extracted features (EF). For the sake of simplicity, the parameters are numbered and don't need to be described in detail.

Once performing the feature weighting based on the impurities, as it can be observed in Fig. 4, we take a look into the subset based only on EF and without PP.

It can be observed that the feature 1 is the most important one with the highest mean decrease in impurity by around 8%. A slightly different behavior can be observed by applying the same method in a regression problem. In this case (not shown in a figure) the most important features are 1, 6 and 3 according to the mean decrease in impurity. The analysis of both, classification and regression problem in term of feature importance gives a diffuse projection of features. It can be stated that the first 6 features have slightly

higher number than the remaining ones, though none of them (neither a single one nor a group) can be defined as dominant.

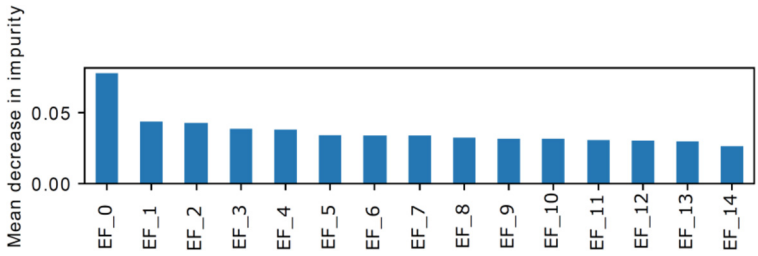


Fig. 4. Feature importance of classification with RF without PP

Once adding the PP into the data set, a different behavior can be observed (as shown in Fig. 5).

It is obvious in this case that the PP dominate the importance graph. The first extracted feature occurs at the amount of around 4% by a highest amount of the first production parameter (the temperature of cooling fluid) by around 13%. This parameter dominates even much more the importance in term of regression analysis with an amount of over 50%.

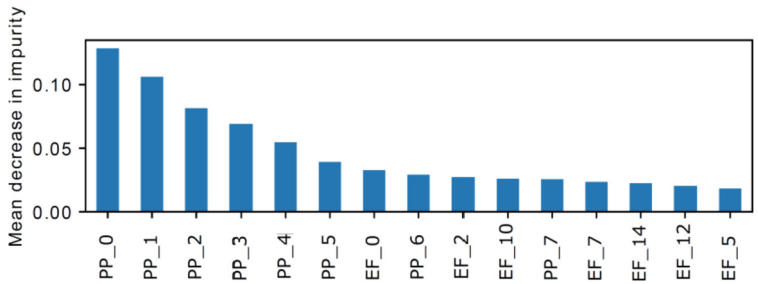


Fig. 5. Feature importance of classification with RF with PP

6 Analysis and Prognosis of the Data

In the following sections, a selection of numerical results of the machine learning analysis is discussed. Basically, all analysis processes were divided into many groups according to the certain problem. Therefore, we distinguish between different data sets (with and without PP) as well as between classification and regression problems, where the roughness values were classified by the specification limits into three groups.

All the forests were trained with 80% of the data. The remaining subset was divided equally into two groups for the validation and test purposes. All the forests were performed with the amount of 1000 trees.

6.1 Prognosis Without Process Parameters

In Fig. 6, an exemplary confusion matrix obtained by the classification analysis of data without PP is shown.

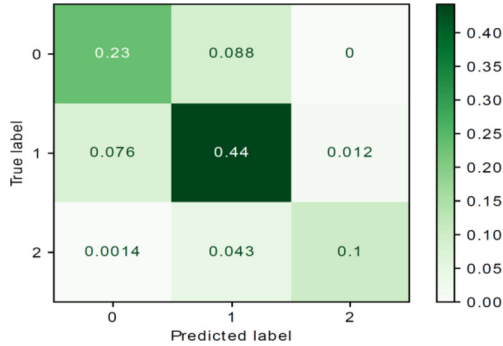


Fig. 6. Confusion matrix, classification problem, data set without PP

The overall efficiency of all perfumed analyses amounts to around 75%–77% which corresponds also to other tested algorithms that are not discussed within this paper (especially Support Vector Machines and Neural Networks). An important property of the analysis with RF is the fact that only very few data points are classified wrong with the distance of two classes (no wrong classified class 2 and very few in class 0). The most misclassification classes occurs especially at the borders of the specs which can let the conclusion of an inappropriate data amount.

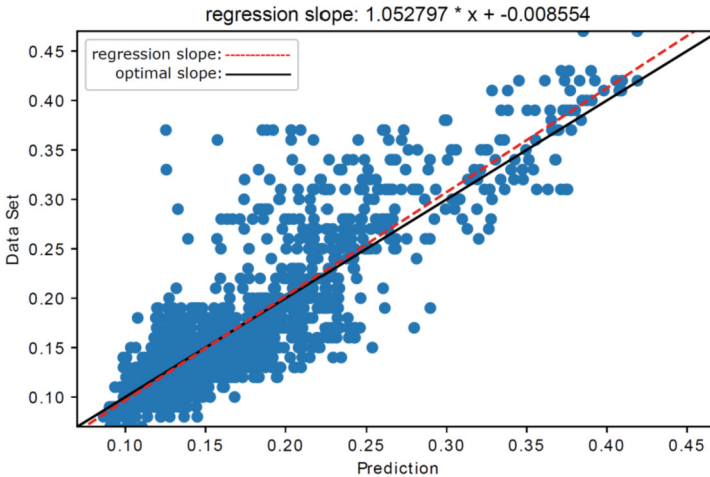


Fig. 7. Regression analysis of the data without PP – test phase

In Fig. 7 the results of the regression analysis in form of the predicted data as well as both models (predicted and real) in form of the two different lines are shown.

It shall be mentioned that the plot shows the test data which is separated only for the prediction purposes.

Once more it can be stated that the RF provide good results in terms of the model accuracy (mainly regarding the interpretation of both regression models but also based on quantified results which are not shown within this study).

6.2 Prognosis with Process Parameters

The achieved results by the analysis of the data without the PP can be significantly increased by the addition of PP (cf. Fig. 8).

It can be observed that the overall efficiency increases to 93% with the additional fact that no data is misclassified with the distance of two classes.

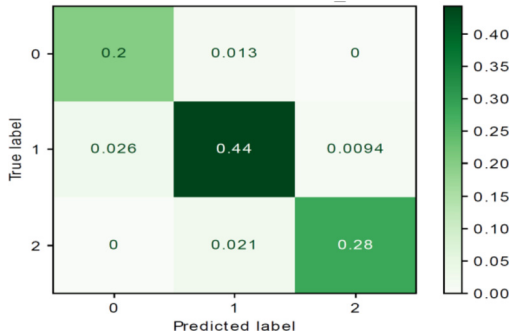


Fig. 8. Confusion matrix, classification problem, data set with PP

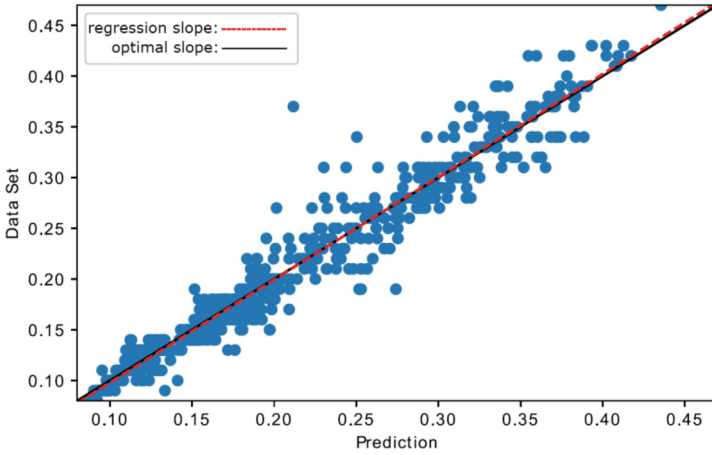


Fig. 9. Regression analysis of the data with PP – test phase

Also, the analysis of regression results leads to the conclusion that the addition of the PP increases significantly the modelling accuracy (as shown in Fig. 9).

Here, the scattering of the prediction decreases by a simultaneous increase of the model efficiency compared to the theoretical/optimal one.

7 Summary and Outlook

In this paper, the results of RF based analysis of fine grinded surfaces were shown. The training of the algorithms is based on feature extraction gathered by the means of CV. As summary, following declarations can be stated:

- RF provide comparable results to other machine learning algorithms which leads to the statement that the achieved efficiency of 75%–77% is mainly restricted by the feature extraction or even more particularly to the amount of unextracted information from the photos
- The inclusion of PP increases the efficiency of the algorithm to around 93%
- RF is a very good and well working algorithm for the feature reduction

In the further studies, more comprehensive feature extraction techniques shall be analyzed and performed. Here, a great number of potential features can be extracted and, with the help of FR, evaluated in terms of importance. A further possibility of the feature extraction is the application of LSTM [16] networks in terms of encoder – decoder algorithm. These networks with a recurrent part in the neurons itself give extract the features automatically, though not necessary with a higher technical understanding of the product.

Furthermore, an optimized camera system with an optimized lens within the test rig is planned to be tested. For an alternative measuring equipment in terms of the target variables, a confocal 3D measuring device shall be used.

References

1. Koblar, V., Pecar, M., Gantar, K., Tusar, T., Filipic, B.: Determining surface roughness of semifinished products using computer vision and machine learning. In: Proceedings of the 18th International Multiconference Information Society, Volume A, pp. 51–54 (2015)
2. Suen, V., et al.: Noncontact surface roughness estimation using 2D complex wavelet enhanced resnet for intelligent evaluation of milled metal surface quality. *Appl. Sci.* **8**, 381 (2018)
3. Rifai, A.P., Aoyama, H., Tho, N.H., Dawal, S.Z.M., Masruroh, N.A.: Evaluation of turned and milled surfaces roughness using convolutional neural network. *Measurement* **161**, 107860 (2020)
4. Hinz, M., Radetzky, M., Guenther, L.H., Fiur, P., Bracke, S.: Machine learning driven image analysis of fine grinded knife blade surface topographies. *Procedia Manuf.* **39**, 1817–1826 (2019)
5. Hinz, M., Guenther, L.H., Bracke, S.: Application of computer vision in the analysis and prediction of fine grinded surfaces. In: Baraldi, P., DiMaio, F., Zio, E. (eds.) Proceedings of ESREL 2020 PSAM 15. Research Publishing Services, Singapore (2020)

6. Pal, S.K., Chakraborty, D.: Surface roughness prediction in turning using artificial neural network. *Neural Comput. Appl.* **14**(4), 319–324 (2005)
7. Vasanth, X.A., Paul, P.S., Varadarajan, A.S.: A neural network model to predict surface roughness during turning of hardened ss410 steel. *Int. J. Syst. Assur. Eng. Manag.* **11**(3), 704–715 (2020)
8. DIN: DIN EN ISO 4287: 2010-07, geometrical product specifications (GPS) – surface texture: Profile method – terms, definitions and surface texture parameters (2010)
9. Bracke, S., Radetzky, M., Born, P.: Multivariate analyses of aperiodic surface topologies within high precision grinding processes. In: *CIRP Conference on Intelligent Computation in Manufacturing Engineering (ICME)*. Gulf of Naples, Italy (2018)
10. Szeliski, R.: *Computer Vision: Algorithms and Applications*. Texts in Computer Science, Springer, London (2011). <https://doi.org/10.1007/978-1-84882-935-0>
11. Kekre, H., Gharge, S.: Image segmentation using extended edge operator for mammographic images. *Int. J. Comput. Sci. Eng.* **2**, 1086–1091 (2010)
12. Breiman, L.: Random forests. *Mach. Learn.* **45**, 5–32 (2001)
13. Quinlan, J.R.: *C4.5: Programs for Machine Learning*. The Morgan Kaufmann Series in Machine Learning, Kaufmann, San Mateo (1993)
14. Awad, M., Khanna, R.: *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, The Expert’s Voice in Machine Learning. Apress Open, New York (2015)
15. Solem, J.: *Programming computer vision with Python: [tools and algorithms for analyzing images]*. Aufl. 1. ed. Beijing [u.a.]: O’Reilly (2012)
16. Yang, R., et al.: CNN-LSTM deep learning architecture for computer vision-based modal frequency detection. *Mech. Syst. Signal Process.* **144**, 106885 (2020)