Proceedings of the 33rd European Safety and Reliability Conference (ESREL 2023)
Edited by Mário P. Brito, Terje Aven, Piero Baraldi, Marko Čepin and Enrico Zio

doi: 10.3850/978-981-18-8071-1_P408-cd

Uncertainty quantification of different data sources with regard to a LSTM analysis of grinded surfaces

Marcin Hinz
Chair of Artificial Intelligence in Mechanical Engineering, Department of Mechanical, Automotive and Aeronautical Engineering, Munich University of Applied Sciences, Germany. E-mail: marcin.hinz@hm.edu

Jannis Pietruschka
Chair of Reliability Engineering and Risk Analytics, Faculty of Mechanical and Safety Engineering, University of Wuppertal, Germany. E-mail: jannis.pietruschka-hk@uni-wuppertal.de

Stefan Bracke
Chair of Reliability Engineering and Risk Analytics, Faculty of Mechanical and Safety Engineering, University of Wuppertal, Germany. E-mail: bracke@uni-wuppertal.de

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.
In this study, fine grinded surface images have been generated under constant boundary conditions in a test rig built up in a lab. The gathered image material in combination with the classical measured surface topography values is used as the training data for machine learning analyses. The image of each grinded surface is analyzed regarding its measured arithmetic average roughness value (Ra) by the use of Recurrent Neural Networks (in this case LSTM). LSTMs are a type of machine learning algorithms which can particularly be applied for any kind of analysis based on time series. In this paper a possible optimization potential of the available databases is analyzed. For this purpose, two different sets of images with various resolutions were taken under the same conditions. Since the data plays an essential role for the training of machine learning models, the challenge in the application is often to find cost-efficient, fast and at the same time process-adaptable measurement methods that also have sufficient accuracy. Thus, the target values recorded with tactile measurement method are compared to a more precise confocal / optical measurement method. This results in two data sets with unequal distributions and different statistical variance.

Keywords: Machine Learning, Neural Networks, LSTM, Uncertainty quantification, Konfocal Measurements.

1. Introduction

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. Therefore, the durable quality of a product can be improved by an optimized configuration of the process parameters.
In this study, a variety of cutlery samples with different surface topographies are manufactured with a variety of process parameters of the high precision fine grinding process. Surface topographies are measured by the use of classical methods like roughness measuring device or confocal measuring device.
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. For the purpose of this study, fine grinded surface images have been generated under constant boundary conditions in a test rig built up in a lab. The gathered image material in combination with the classical measured surface topography values (tactile and confocal) is used as the training

and target data for machine learning analyses.
Within this study, the image of each grinded surface is analysed regarding its measured arithmetic average roughness value (Ra) by the use of Recurrent Neural Networks (in this case LSTM). LSTMs are a type of machine learning algorithms which can particularly be applied for any kind of analysis based on time series. The novelty of the proposed method is the treatment of the pictures as time series. In general, along many horizontal lines drawn along the original picture of the surface, the development of the lightness value of the line provides a time series analogical signal that can be treated by recurrent network.
The entire parameter study regarding the network topology and parameter settings was performed prior this study. In this paper, only the most performant settings are used as starting points for the further optimization and uncertainty quantification. The approach of optimizing the algorithm results and identifying a reliable and reproducible LSTM model, which operates well independent of the choice of the random sampled training data, is presented in detail. Finally, the performance of the models trained with the optical measured data is compared with the models from the tactile measured database.

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

2. Data generation

The presented research activities within the analyzed data set embrace photographs of the surfaces of 812 8” slicer knives (cf. figure 1) and 851 8” chef’s knives (cf. figure 2) The surface images are taken within an experimental test rig which provides constant boundary conditions (cf. section 2.2). Reference measurements of the surface roughness as well as gloss and coloring are taken of all the knives (cf. section 2.1).

Fig. 2. An example image of a chef knife analyzed within this research activities

2.1. *Reference measurements*

The term classical measurements is used in this work for this type of measurements that can be performed with common and available stand alone measurement devices without the need of an application of any further process steps. As a matter of principle, some devices, as the confocal system presented in this chapter, need third-party software to edit and provide the results but there is no need of additional programming or applying of mathematical algorithms beyond this software. There are two main devices used in this study:

- Roughness Tester
- MarSurf CM mobile

Roughness Tester is a typical roughness measuring device that works based on piezoelectric micro probe principle. Basically, a test head moves along a line of $6mm$ and measures the unevenness of surface height along one line, which means that one measurement provides one single value. According to the product data-sheet, the tester has the following specifications:

- Roughness parameters Ra, Rz, Rq, Rt
- Accuracy $\pm 15\%$
- Repeatability $< 12\%$
- Measuring range Ra $0.05...10\mu m$
- Scanning path in total $6mm$

The second device used for the measurement of roughness as target values for the further analysis is an optical 3D microscope that uses the confocal measuring principle. Basically, the device measures a given number of 2D edge slices in a specified height and joins all of them to a 3D shape. For a comprehensive explanation of this method refer to (Price, 2011).
The main advantages of such a system are the

Fig. 3. MarSurf CM mobile - confocal measuring device build up in the lab

high precision (much higher than in case of a piezoelectric roughness measuring device) as well as the art of the measurement. It is performed in a predefined area (and not along a single line) by so called stitching, which is moving of the lens along the area, measuring and merging the results. Therefore, the measured value is the average roughness of a surface defined as:

$$S_a = \frac{1}{A}\iint_A |Z(x,y)|dxdy \qquad (1)$$

The main advantage of a surface roughness is that additionally to the measured value itself we obtain also the scattering of the roughens within this area. Device used in this study is the MarSurf CM mobile produced by Mahr. The assembled device used in the lab with a 320L lens, as shown in fig. 3, has the following technical specifications:

- Maximum number of metering points within one measurement: $1200 \cdot 1200 = 1.44mil$
- Lens magnification: $50x$
- Lateral measurement range $x \cdot y(mm)^2$: 0.1024
- Extended lateral measurement range $x \cdot y(mm)^2$ with the use of stitching: 84.6
- Measurement uncertainty based on an exemplary measurement of $R_a = 0.079\mu m$: uncertainty $U = 0.01\mu m$, standard deviation $\sigma = 0.0022\mu m$

Standard deviation is calculated based on 25 measurements. For more technical specifications refer to the Mahr homepage.

The measurement area was placed $1cm$ below the mentioned $6mm$ line to avoid spacial uncertainties. In order to assure that all the measurements are performed at the same position, a 3D printed holder was designed and used for the confocal measurement as well as the test rig presented designed for the purpose of this study. The grey holder is placed below the knife blank of the MarSurf CM mobile picture presented in figure 3. (Hinz et al., 2019, 2020)

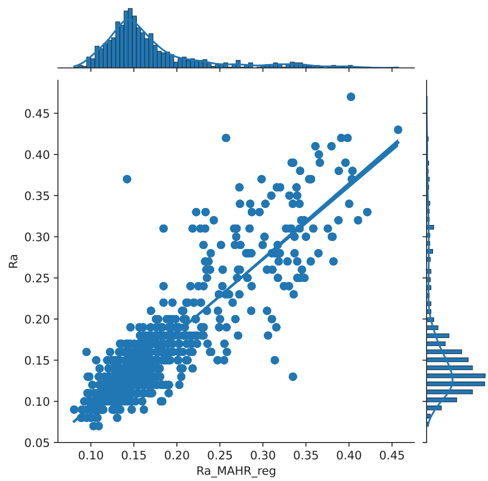

Fig. 4. Comparison of the distributed values out of both measurements: tactile and confocal

The Distribution of the values measured with both measuring devices is shown in Fig. 4. It can be observed that the values differ in many cases. This is mainly caused by the uncertainty of both measuring systems. It needs to be highlighted that both systems have different amount of uncertainty which leads to different target values. This will, obviously, have an impact on the overall results presented in section 4.

2.2. *Experimental setup*

In order to take comparable images of different knives, two similar test rigs that provide consistent lighting conditions was designed and build up for the purpose of this study. As a matter of principle it is one test rig that was adopted for two different experiments. 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.

Two LED Spotlights, pointed at the walls on the top- and bottom ends of the knife in the first experiment and at the knife itself in the second one (second camera needs much more light), 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 with:

- Olympus E-520 DSLR, equipped with an Olympus Zuiko Digital 14-42mm f 1:3.5-5.6 lens in the first experiment (pictures with lower resolution of $3648px$ x $2736px$ cropped to $1250px$ x $550px$)
- Canon EOS 77D camera with a Canon MP-E 65mm f/2.8 1-5x macro lens in the second experiment ((pictures with much higher resolution of $6000px$ x $4000px$ without any need of cropping due to a better lens).

The lens is placed normal to the knife's surface at a distance of 8 cm. For a comprehensive description of the test rig cf.

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. (Koblar et al., 2015; Suen et al., 2018; Szeliski, 2011)

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 results of the further analysis. The images are cropped to a size of 1250 px x 550 px, which corresponds to an area of 2,5 cm x 2,0 cm.

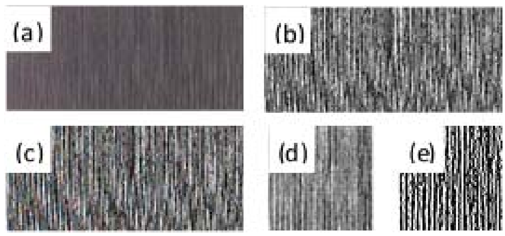

Fig. 5. (a) cropped image (b) sobel operator (c) contrast change and sobel operator (d) low pass mask and median blur (e) low pass mask, median blur and sobel operator

The knife surfaces have a grooved structure, as it can be seen in figure 5(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. (DIN, 2010; Hinz et al., 2019)

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 5 shows the original and the pre-processed pictures. (Hinz et al., 2019; Kekre and Gharge, 2010; Koblar et al., 2015; Solem, 2012; Suen et al., 2018)

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.

4. Model analysis

In this section a detailed analysis of the data sets is performed and discussed. The data sets used within this study is based on cropped images mentioned in chapter 3 and has the structure as follows: As already stated in (Hinz et al., 2022), the slicer dataset includes 8120 data points extracted from 812 knives, while the chef dataset contains about 8510 data points. From both types of knives, the entries along the surface can be analyzed on ten evenly set lines. Nevertheless, the three target classes based on target variables are not distributed equally. Fig. 6 shows the distributions of the classes for both types of knives. Here it can be clearly seen that for both types of knives a shift of the class distributions results from the confocal measurement. While the chef's knives have a more concise middle class, the proportions of the slicer class 0 as well as 1 decrease and the upper class expands. Since the set of examples plays a crucial role for the training of machine learning algorithms, all data series were used despite the unequal class distribution.

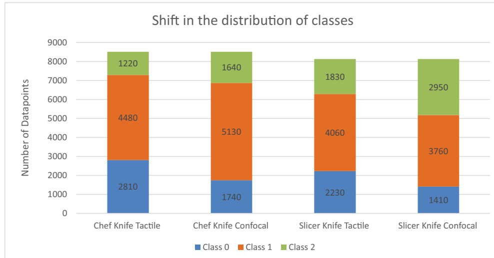

Fig. 6. Data Set Class Distribution

In addition to the varying measurement methods of the target value, there are also differences in the image resolution when comparing the two sets of measurement data analyzed here. Each labeled slicer knife time series consists of 1250 pixels with an assigned ouput value of tactile roughness measurement. On the other hand, the chef's knife data used here has an increased resolution of the images and thus each labeled time series consists of 6000 pixels. The assigned label then results from the confocal measurement. The input variables provide information about the surface properties by the lightness of the individual data point (the grey value lies between 0 and 255). In the output classes, the classification is based on the average roughness "Ra" in µm and classified according to the natural specification limits. Obviously, the middle class which represents the well produced knifes is the most frequent one. In any other case, the company would produce more waste than salable products and probably get into serious financial problems after a short period of time. The output variable is classified into three classes and subsequently binarized by the use of softmax function regarding the output neurons for numerical purposes.

In the previous study, two different time series sizes were tested. To keep the number of training data as high as possible, the pixel rows were not only read in as a whole, but also divided into five 250 px rows. (cf. figure 7) shows an example of these series. The red dashed lines mark the split points of the pixel rows, which also correspond to the original label of the whole pixel row of 1250. This results in a data set size of 40,600 examples.

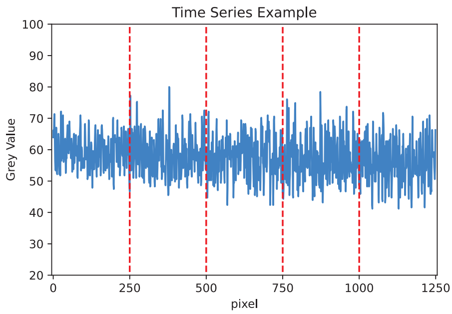

Fig. 7. Lower resolution Slicer Knife: Time series example with dividing lines

As seen in Fig. 8), the same approach was used for the data augmentation of the chief knife pixel rows. However, here the data was merely doubled by a single division, because a key discovery from the previous study was that longer pixel rows led to better accuracy. Since longer time series led to considerably longer calculation times and computational effort, the maximum possible length was shortened from 6000 pixels to 3000. This results in 17,020 examples that can be used for the machine learning training.

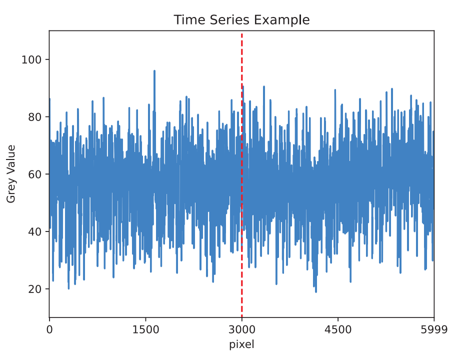

Fig. 8. Higher resolution Chef Knife: Time series example with one dividing lines

Based on the best performance regarding validation and test accuracy, a specific parameter combination was found for the slicer knife classification. Since the datasets have comparable properties, this was used as a reference point for the investigations of the chef knife classification. So the models were trained with the maximum possible pixel sequence length of 3000 px. The initial parameters are:

- neurons: 250
- learning rate of 1e-5
- batch size of 128
- epochs: 500

The final Slicer validation accuracy of this model is 77.46% with in test accuracy of 75.62%. With this combination of parameters, now the chef knife data were trained. Except for a large variance, the chef knife models performed comparably well without further adjustments. Based on the experiences with the slicer knives, obvious parameters such as neurons, batch size, and learning rate were adjusted, and a different optimizer (RMSprop) was tested using a best guess approach. An exemplary learning curve is shown in Fig. 9). It is clear to see here that the model becomes highly overfit after about 120 epochs and learns non-generalizing through the strong increase in validation loss.

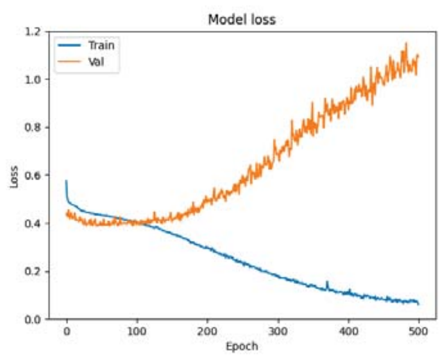

Fig. 9. Chef Knife dataset: Overfitting model with 500 epochs

To optimize the models that tended to overfit, the number of epochs was reduced to 120 and the learning rate was increased to 0.01. Finally, a suitable parameter setting was tested ten times to determine the stability of the model. Figure 10) shows a fully trained learning curve of these models, which makes it apparent that the learning process is stable and has fully converged, because no changes in the curve are visible.

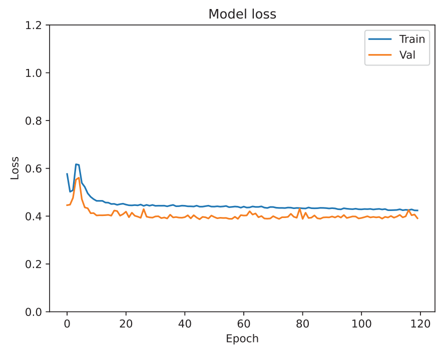

Fig. 10. Chef Knife dataset: Optimized learning curve

4.1. *Model comparison*

In direct comparison, the new chef knife models showed significant performance improvements compared to the slicer models. With a slight increase in the number of neurons to 300 in a single layer, an average validation accuracy of $84.45\% \pm \sigma = 3.68e-3$ was achieved. Thus, the results are almost 7% more accurate than those of the slicer knives. The test accuracy could also be improved by 7.5% to $83.11\% \pm \sigma = 2.1e-3$. The loss values also decreased by more than half: Validation Loss 0.399 $\pm\sigma = 7.31e-3$; Test Loss 0.434 $\pm\sigma = 3.42e-3$. The low standard deviations of all results thus demonstrate high stability of the models.

4.2. *Validation of results*

In general, it can be stated that the optimized models, which were validated with one product type, are suitable to be transferred to similar products. However, different measuring systems that provide varying amounts of uncertainty will cause the models to differ from each other in terms of the overall accuracy.

Observing the results leads to the conclusion that more precise measuring systems, such as the confocal system used in this study, result in much higher accuracy. This can be logically understood: a more accurate measuring system with higher-quality pictures leads to more accurate results. Taking into account the uncertainties arising from the manufacturing processes and product materials, an overall accuracy of approximately 85% can be considered satisfactory.

5. Summary and outlook

This research paper focuses on the importance of high precision fine grinded surfaces in cutlery products. The authors highlight the need for a new image processing analysis approach to improve conventional methods of condition monitoring for these surfaces. The paper presents a study in which fine grinded surface images are generated and analyzed using Recurrent Neural Networks (LSTMs) to determine the arithmetic average roughness value (Ra). The study includes a comprehensive discussion of the parameter study for identifying a reliable and reproducible LSTM model, which operates well independent of the choice of the random sampled training data. The authors compare the performance of the models trained with confocal measured data with the models from the tactile measured database. Overall, the aim of the study is to develop a condition monitoring tool that can be used to ensure the quality of knives and reduce rejects by detecting deviations of the target values and adapting the production process accordingly.

Acknowledgement

We would like to express our gratitude to Mr. **Mario Adam** and the entire **Mahr GmbH** company for their support in our research work. In particular, we would like to thank them for lending us their **MarSurf CM** confocal measuring system, which played a critical role in our experiments. The use of this machine allowed us to obtain valuable data that contributed significantly to our research findings. We appreciate their assistance and support throughout the duration of the project. Without their generous help, this work would not have been possible.

References

DIN (2010). Din en iso 4287: 2010-07, geometrical product specifications (gps) – surface texture: Profile method – terms, definitions and surface texture parameters iso 4287:1997 + cor 1:1998 + cor 2:2005 + amd 1:2009; german version en iso 4287:1998 + ac:2008 + a1:2009.

Hinz, M., L. H. Guenther, and S. Bracke (2020). Application of computer vision in the analysis and prediction of fine grinded surfaces. In P. Baraldi, F. Di Maio, and E. Zio (Eds.), *Proceedings of ESREL 2020 PSAM 15*. Singapore: Research Publishing Services.

Hinz, M., J. Pietruschka, and S. Bracke (2022). Lstm based condition monitoring of fine grinded surfaces. In *Proceedings of the 32nd European Safety and Reliability Conference (ESREL 2022)*, Singapore. Research Publishing Services.

Hinz, M., M. Radetzky, L. H. Guenther, P. Fiur, and S. Bracke (2019). Machine learning driven image analysis of fine grinded knife blade surface topographies. *Procedia Manufacturing*.

Kekre, H. and S. Gharge (2010). Image segmentation using extended edge operator for mammographic images. *International Journal on Computer Science and Engineering 2*.

Koblar, V., M. Pečar, K. Gantar, T. Tušar, and B. Filipič (2015). Determining surface roughness of semifinished products using computer vision and machine learning. In *Proceedings of the 18th International International Multiconference Information Society*, Volume A, pp. 51–54.

Price, R. L. (2011). *Basic Confocal Microscopy*. New York, NY: Springer New York.

Solem, J. E. (2012). *Programming computer vision with Python: Tools and algorithms for analyzing images* (1. ed. ed.). Beijing u.a.: O'Reilly.

Suen, V., B. Yao, B. Chen, Y. He, X. Cao, T. Zhou, and H. Liu (2018). Noncontact surface roughness estimation using 2d complex wavelet enhanced resnet for intelligent evaluation of milled metal surface quality. *Applied Sciences 8*, 381.

Szeliski, R. (2011). *Computer vision: Algorithms and applications*. Texts in computer science. London u.a.: Springer.