Development of an ML model for the classification of surface quality in a milling process

Max Simon Teubl
FH Joanneum University of Applied Sciences, Werk-VI-Straße 46, 8605 Kapfenberg, Austria; Email: max.teubl@edu.fh-joanneum.at

Vitaliy Mezhuyev*
FH Joanneum University of Applied Sciences, Werk-VI-Straße 46, 8605 Kapfenberg, Austria; Email: vitaliy.mezhuyev@fh-joanneum.at

Martin Tschandl
FH Joanneum University of Applied Sciences, Werk-VI-Straße 46, 8605 Kapfenberg, Austria; Email: martin.tschandl@fh-joanneum.at

ABSTRACT

Digitalisation is increasingly finding its way into the production process of manufacturing companies. The paper deals with the question of how manufacturing data of a milling process can be analysed using machine learning (ML) methods to classify surface defects at an early stage of production. The paper develops an ML model that classifies image data based on the surface roughness of produced parts. For this purpose, sample parts were produced on a milling machine in the learning and research factory for Industry 4.0, the Smart Production Lab at FH Joanneum, Austria. This resulted in a data set of about 38,500 images. The developed based on a convolutional neural network model achieved an accuracy of 82% in predicting surface quality. The model divides produced sample parts into quality classes with a surface roughness between Ra 0.3 and Ra 2.3. Being used in industrial processes, the developed ML model enables reliable prediction of surface quality without manual measurement and evaluation.

CCS CONCEPTS

• **Computing methodologies**; • **Machine learning**; • **Machine learning approaches**; • **Neural networks**; • **Applied computing**; • **Enterprise computing**;

KEYWORDS

machine learning, image classification, surface defects, milling process

ACM Reference Format:
Max Simon Teubl, Vitaliy Mezhuyev, and Martin Tschandl. 2023. Development of an ML model for the classification of surface quality in a milling process. In *2023 9th International Conference on Computer Technology Applications (ICCTA 2023), May 10–12, 2023, Vienna, Austria.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3605423.3605449

1 INTRODUCTION

In recent years, companies in industrialised countries have been increasingly facing the challenges of global competition. This international economic pressure is driving manufacturers to constantly make their production even more efficient and flexible or to develop new, digitised business models in the first place. Industry 4.0 or Smart Production can be helpful on all counts. The trend toward more efficient and, as a result, more automated production, requires, among other things, approaches for continuous and reliable monitoring of the machining processes to combine quality control with the actual production of parts. Such a combination of the manufacturing processes leads to more effective controllability of production, thus making it possible to achieve an essential efficiency goal for manufacturing companies, namely, to produce goods with the required quality at lower costs [1]. Thus, an important step is the combination of quality control with actual production, which has the goal of zero-defect production [2]. This is because quality defects have a particularly drastic effect on costs if they are only discovered at the end of the manufacturing process. At this point of production either expensive reworking or even more expensive scrap are the results [3].

Preventive detection of quality defects is a complex process that requires a variety of different methods [4]. Digitalization and the associated technologies enable companies to collect a large amount of data in their manufacturing processes [5]. This requires the integration of appropriate sensors and the implementation of analysis methods to detect trends and make valid statements for the process [6]. Modern machine learning (ML) approaches can represent such solutions [7, 8].

Used correctly, generated data can lead to quality improvements and/or cost reductions, among other benefits. Nevertheless, many companies are overwhelmed by the amount of data and associated analysis and can lose sight of the real goal. Often, the process data is just collected without being able to present an actual purpose for it [9]. An additional challenge is frequently posed by the actual tasks, which require expertise on the shop floor as well as know-how from IT experts and data scientists [10].

This paper describes data analysis of a digitized production process using ML methods for early quality assurance based on a concrete use case. To this end, the following research question should be answered: how can manufacturing data from a milling process be analysed using ML methods to classify surface quality defects at an early stage of production?

To answer this question, a comprehensive literature review was first conducted. The fundamentals of smart manufacturing, milling, ML and data processing were described [11]. In the use case, a milling machine from the Smart Production Lab of FH JOANNEUM University of Applied Sciences was used to produce sample parts. A smart factory or Industry 4.0 factory is a learning and research facility that uses advanced technologies such as the Internet of

*Corresponding author.

ICCTA 2023, May 10–12, 2023, Vienna, Austria

ACM ISBN 978-1-4503-9957-9/23/05.
https://doi.org/10.1145/3605423.3605449

Things (IoT), robotics, added manufacturing, vertical process integration, artificial intelligence, and automation – all this to improve production efficiency, quality, and flexibility in a manufacturing environment [12, 13]. Manufacturing numerous parts on a compact desktop CNC machine for training and education purposes generated enough data for the development and validation of the predictive model. To train the ML model, the machined parts were tested with a surface measuring device. Using a convolutional neural network (CNN), an attempt is then made to make a prediction statement about the surface quality of the manufactured components.

The paper is structured as follows. Following the introduction, the second chapter provides the methodology of the research. This part briefly describes the necessary steps from data acquisition to the final evaluation. Next, in the chapter "Use case: surface quality classification in a milling process" an ML model is developed and evaluated. In the final part of the paper, a conclusion is given. Recommendations for further work are also briefly addressed.

2 METHODOLOGY

The quality of the surface of produced parts is one of the most important specifications in a manufacturing process. A 100% inspection of produced parts is often not possible in practice. Thus, a possible approach is to develop an ML model making a reliable classification by identifying poor-quality parts at an early stage of the production process. The methodology follows the standard steps of an ML model development: data collection, data preparation, model training, model evaluation and optimization. In our approach, for quantitative measurability of the result, the parts were classified according to their surface roughness.

Data collection. After producing sample parts, their roughness was measured with the help of a surface-measuring device. To generate the database for ML training, the manufactured samples were photographed, and the images were saved in a structured way.

Data preparation. The generated data were pre-processed. In this step, the images were checked for quality and prepared for feeding into the model. The generated data were divided into training, testing and validation sets. 80% of the data was used for training purposes, 10% for testing and 10% for validation.

Model training. The training of the ML model took place using correspondingly the training data set. For the model training and its following evaluation, the KNIME tool [14] from the Swiss software company KNIME AG was used. In KNIME, a user can create an ML model by connecting nodes in an intuitive graphical interface, thus designing workflows for different applications.

Model evaluation and optimization. Finally, the developed ML model was assessed with the testing data set. Based on the evaluation metrics, the model was optimised in an improvement loop to increase its reliability.

3 USE CASE: SURFACE QUALITY CLASSIFICATION IN A MILLING PROCESS

3.1 Sample Production

To create the samples as resource-efficiently and effectively as possible, a part was chosen that could be used in another project at the Smart Production Lab. This made it possible to manufacture the pieces with as little waste as possible. The part comes from the assembly of a table clock and represents its base. The starting point to manufacture this base is an aluminium flat material with the dimensions 60mm x 20mm x 1000mm. The exact material designation is AlMgSi 0.5 / EN AW 6060 / DIN1770 / EN 755-5.

To be able to produce the samples on the milling machine, they had to be cut to size on an industrial band saw from Behringer in the first step. Here, the raw material was clamped on the machine bed and cut into 70mm long pieces. In the next step, the pieces were cleaned from cutting burrs so that there would be no dislocation due to contamination when the parts were clamped in the milling machine. In the last step, the cut parts were placed in the milling machine and fixed by means of the pneumatic clamping device.

To generate enough data for training the ML model, several workpieces were produced. During the production of the samples, the parameters of the milling machine were changed. In the various programmes created, the spindle speed (rpm), the feed rate (mm/min) and the depth of cut were adjusted to produce the respective quality of the surface. Each produced sample was labelled on the back with the respective machine parameters.

Table 1: Surface groups

Group	Group average surface [Ra]	Group average surface [Rz]
1	0.30	1.82
2	0.54	3.12
3	0.92	5.05
4	1.64	7.28
5	2.16	10.42
6	2.28	9.60

3.2 Roughness measurement

The MarSurf PS 10 C2 mobile surface roughness measuring device was used. Due to its robust and compact design, the device is particularly suitable for measuring under workshop conditions. The intuitive and simple operation of the device makes it particularly user-friendly. The measuring device offers a 4.3-inch touchpad, which can be used to control functions [15].

The unit must be calibrated before it can be put into operation. This is done with the help of the supplied measuring standard (Rz 9.520), which is located on the back of the unit. The determined correction value was 4.10%. After setting up and calibrating the measuring device, the measurements were carried out and processed. The values of the respective test series were then combined into six groups by an average roughness value (Table 1). This mean was then used to assign the image data (photographs of the parts) to the appropriate classes of the surface.

3.3 Model development

Five workflows were developed in KNIME: for image preprocessing, image processing, defining network architecture, model training, and model validation.

Image pre-processing. The workflow was designed to structure the images in the form of a data set for the application of the ML algorithm (specifically, a convolutional neural network, CNN).

Image processing. This workflow was created to pull the generated images from the database and prepare them for the CNN training.

Network architecture. With this workflow, the architecture and parameters of the respective CNN layers were defined:

- **Keras Input Layer.** Integration of the image data with a defined format of 32x32x1 into the network.
- **Keras Convolution 2D Layer.** A common size of 3x3 has been set for the size of the kernel. The ReLU function is used as the activation function. The number of filters has been set to 64.
- **Keras Max Pooling 2D Layer**. For the pooling layer, a size of 2x2 has been set with a stride of two steps. As the name suggests, this layer applies frequently used max pooling approach.
- **Keras Flatten Layer**. To prepare the feature map of the pooling layer for feeding into the network, it had to be transformed into a one-dimensional vector. This layer was used for this purpose.
- **Keras Dense Layer**. With the two dense layers, each unit of the layer input was connected with each output unit of this layer. For the first layer, the ReLU function was used. The second layer used a Softmax activation function.

Model Training. To train the model, three KNIME nodes were used:

- **Keras Network Learner**. This node was fed with the previously partitioned data. As a back-end for deep learning, the TensorFlow framework was used. The number of iterations over the input training data was set to 80 epochs. The number of training data rows that are used for a single gradient update during training was set to a batch size of 128. To improve the performance of the network, the training data was mixed according to each epoch.
- **Keras Network Executor**. With this node, the network runs on the selected backend. The input is connected to the learning node. In addition, 10% of the available data flows into the network as test data.
- **Keras Network Writer**. Finally, this node stores the trained network.

Model validation. This workflow was designed to evaluate the performance of the developed CNN:

- **Extract Prediction.** This node analyses the output of the network, which is available in a table form. The best results of the network are exported in a collected column. The result shows the actual class value (class label) as well as the predicted value in the column (Predicted Class).
- **Image Viewer.** This node is used to analyse the individual images in a table.
- **Rule Engine.** With the help of this node, the measured class and the predicted class are prepared for further analysis with the scorer node.
- **Scorer.** This node offers the possibility to carry out corresponding evaluations via its interactive view.

Table 2: Model evaluation

Overall accuracy	Overall error	Correctly classified	Incorrectly classified
77%	23%	2943	897

3.4 Model Evaluation

With the help of the KNIME, it was possible to calculate fundamental parameters for assessing the quality of the developed ML model. Table 2 shows the overall evaluation of the first developed model. Out of the 3840 inputs, 2943 were correctly classified, giving an overall accuracy of 77%. 897 images were incorrectly classified, resulting in an error of 23%.

Table 3 compares the actual surface groups (classes) with the predicted ones. To clarify the results, the table is coloured as a heat map. High values are marked yellow, whereas low values are coloured dark grey. The results of group 1 (those surfaces have an average Ra value of 0.3) have a minimal precision: of 629 images from class 1, 155 were assigned to class 2 and 148 to class 3 and only 541 were correctly classified. Surfaces from group 4 (Ra value is 1.64) have the maximum prediction accuracy. The table shows that 519 images were correctly classified (94%).

With the results from Table 3, an assessment of True Positive, False Positive, True Negative, and False Negative values and F-measure is made in Table 4. It follows that group 4 produced the results having the best precision.

In addition, ROC curves[1] were created for the individual classes. Figure 1 shows an example for class 4.

Subsequently, the AUC values (giving the area under the ROC curve) of all groups were calculated and compared. The higher the AUC, the better the model's performance at distinguishing between the classes. The AUC score of 1 means the classifier perfectly distinguishes between all class points. In our case, all ROC curves have AUC values of more than 0.935, giving an excellent prediction performance. It could also be seen here that groups four and five delivered the best results (Table 5).

Furthermore, the training and validation set were compared with the help of the learning monitor. This made it possible to determine whether there was overfitting or underfitting, and a statement could be made about how well the model can generalise. In the first step, the accuracy of two data sets was compared.

The error rate was assessed as follows:

The error rate of the training set $1 - Accuracy = 1 - 0.89 = 0.11$
The error rate of the validation set $1 - Accuracy = 1 - 0.78 = 0.22$

The results show that the error rate for the validation set is greater. Thus, a slight overfitting was detected[2]. In addition, the

[1]An ROC curve (receiver operating characteristic curve) is a graph showing the performance of a classification model. Classifiers that give points closer to the top-left corner indicate better performance. As a baseline, a random classifier is expected to give points lying along the diagonal (where the false positive rate equals the true positive rate).

[2]Overfitting occurs when the developed ML model has a high variance, i.e., the model performs well on the training data but does not perform accurately on the validation set.

Table 3: Confusion matrix

Group		Predicted 1	Predicted 2	Predicted 3	Predicted 4	Predicted 5	Predicted 6	Sensitivity
Actual	1	541	26	15	12	16	19	86%
	2	155	442	12	9	24	13	67%
	3	148	25	365	8	21	98	55%
	4	75	12	9	519	9	12	82%
	5	17	5	3	1	582	11	94%
	6	30	0	24	6	82	494	78%
Precision		56%	87%	85%	94%	79%	76%	

Table 4: Detailed statistics of the model

Group	TP	FP	TN	FN	Recall/Sensitivity	Precision	Specificity	F-measure
1	541	425	2786	88	86%	56%	87%	68%
2	442	68	3117	213	67%	87%	98%	75%
3	365	63	3112	300	55%	85%	98%	67%
4	519	36	3168	117	82%	94%	99%	87%
5	582	152	3069	37	94%	79%	95%	86%
6	494	153	3051	142	78%	76%	95%	77%

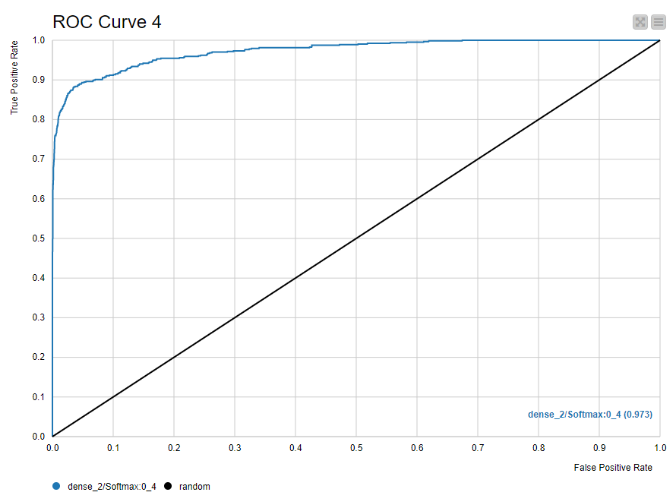

Figure 1: ROC Curve of Class 4

loss of the two data sets was compared[3]. The validation loss is greater than the training loss, as seen in Figure 3. A reason for this is that the model was trained for a long period. So, to prevent overfitting, training should be halted when the loss is low and stable (early stopping).

It could be seen that no more improvement was achieved after a training time in which the model had run through around 12,000 batches (for the test and validation sets). This confirms the results of the accuracy graph, as a plateau also occurs here at around 12,000 batches.

To improve the results of the model and its performance in the classification of the surfaces, an optimisation of the CNN architecture was carried out. For this purpose, the depth of the network was increased, two additional layers were added, as well as further fine-tuning was applied. The achieved results were compared with the first model. In the final optimised version of the model, an overall accuracy of 82% was achieved. Specifically, this means that 3157 surfaces were correctly classified and 683 were assigned incorrectly. This gives an improvement of 5%. The analysis of the confusion

[3] The training loss is a metric to assess how a ML model fits the training data, while the validation loss assesses the error of the model on the test/validation set.

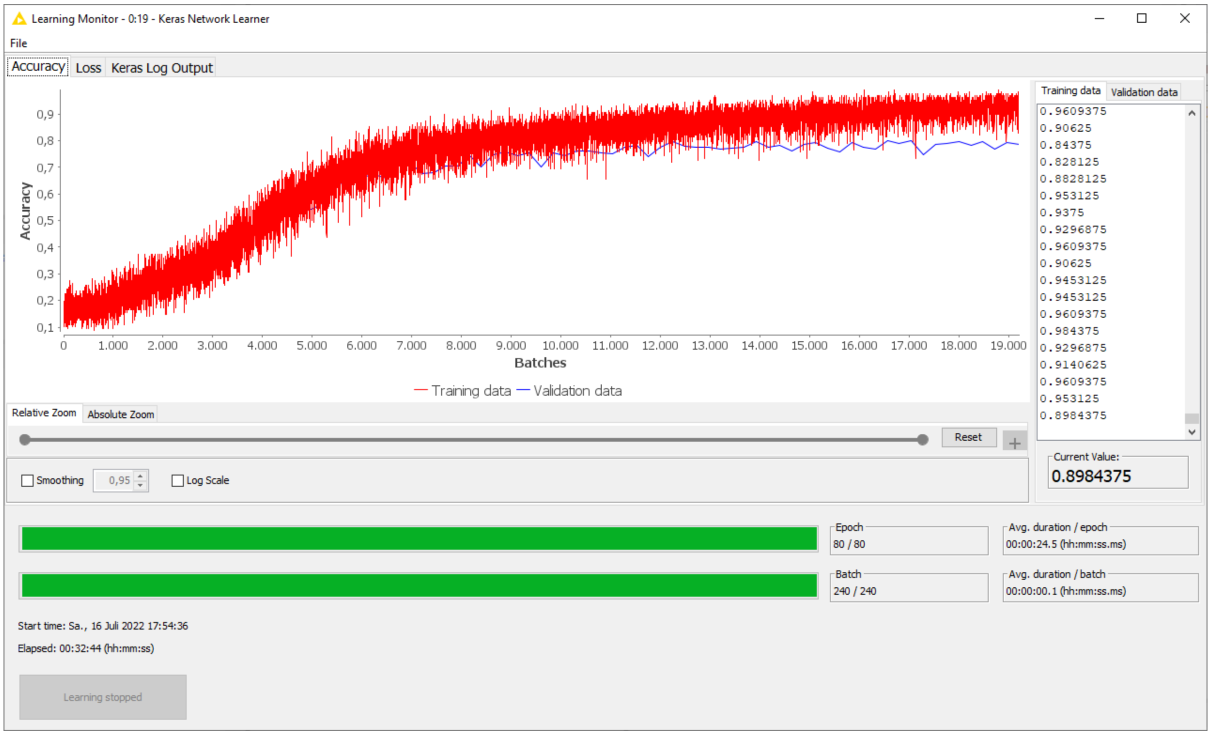

Figure 2: Learning monitor: the Accuracy

Figure 3: Learning monitor: the Loss

Table 5: AUC values comparison

Group	AUC
1	0,940
2	0,970
3	0,935
4	0,973
5	0,983
6	0,942

matrix also showed that there were fewer outliers in the prediction of the classes. The final optimized model achieved better scores across all classes.

4 CONCLUSION

The paper develops an ML model for the classification of the surface quality of samples, manufactured in the milling process. The model classifies the samples with a surface roughness value of Ra ~0.3 to ~2.3. The input for the model is an image, which can be taken on the shopfloor with a not-expensive smartphone camera. Thus, a machine operator could quickly check the produced parts for surface quality. In the future development, a downstream automatic inspection process would also be conceivable. The model will be improved, to have the accuracy, required at the level of industrial application.

Our future research will consider a broader data set with larger spans between the individual surface values. The plan is to use classes corresponding to Ra values in the ranges from 0.4, 0.8, 1.6, 3.2,.. to Ra 25. This would improve the generality and applicability of the ML model. Since only one specific milling method was used in this work, additional milling methods will be applied. The use of different materials would also be of interest, as the diverse physical properties influence the surface quality.

REFERENCES

[1] Rai, Rahul; Tiwari, Manoj Kumar; Ivanov, Dmitry; Dolgui, Alexandre: Machine learning in manufacturing and industry 4.0 applications, in: International Journal of Production Research, Volume 59, Issue 16, 2021, p. 4773-4778.

[2] Lu, Chen: Study on prediction of surface quality in the machining process. Journal of Materials Processing Technology, Volume 205, Issue 1-3, 2008, p. 439-450.

[3] Benteler: Mit Machine Learning zur Predictive Quality. My Factory, Volume 62, Issue 11, 2021, p. 1-42.

[4] M. Binder, V. Mezhuyev and M. Tschandl, Predictive Maintenance for Railway Domain: a Systematic Literature Review, in IEEE Engineering Management Review, doi: 10.1109/EMR.2023.3262282. https://ieeexplore.ieee.org/document/10082880

[5] Vitaliy Mezhuyev, Martin Tschandl, Matthias Mayr. Converting Manufacturing Companies into Data-Driven Enterprises: an Evaluation of the Transformation Model. Proceedings of 7th International Conference on Computer Technology Applications (ICCTA 2021), July 13-15, 2021, Austria. Pp. 80-85. https://doi.org/10.1145/3477911.3477924

[6] Raphael Hartner, Vitaliy Mezhuyev, Martin Tschandl, Christian Bischof. Data-Driven Digital Shop Floor Management: a Practical Frame-work for Implementation. ACM Proceedings of the International conference ICSCA 2020, February 18–21, 2020, Langkawi, Malaysia. pp. 41-45. https://doi.org/10.1145/3384544.3384611

[7] Raphael Hartner, Joachim Komar, Vitaliy Mezhuyev. An approach for increasing the throughput of CNN-based quality inspections systems in constrained environments. 11th International Conference on Software and Computer Applications (ICSCA 2022), February 24-26, 2022, Melaka, Malaysia. Pp. 179-184. https://doi.org/10.1145/3524304.3524330

[8] Powell, Daryl; Eleftheriadis, Ragnhild; Myklebust, Odd: Digitally Enhanced Quality Management for Zero Defect Manufacturing. Procedia CIRP, Volume 104, 2021, p. 1351-1354.

[9] Wuest, Thorsten; Weimer, Daniel; Irgens, Christopher; Thoben, Klaus-Dieter: Machine learning in manufacturing: advantages, challenges, and applications. Production & Manufacturing Research, Volume 4, Issue 1, 2016, p. 23-45.

[10] Lumma, Dirk; Häussler, Ute: Wer führt das Rennen am Shopfloor? Markt & Technik, Issue 49, 2021, p. 16-18.

[11] Max Simon Teubl. Development of a machine learning model for the prediction of surface quality in a milling process. Master Thesis. FH Joanneum University of Applied Sciences, Kapfenberg, Austria. 2022.

[12] Tschandl, Martin; Mayer, Barbara; Sorko, Sabrina Romina: An interdisciplinary digital learning and research factory: The Smart Produc-tion Lab, in: Procedia Manufacturing, Volume 45, 2020, Elsevier, 2020, S. 491-496.

[13] Dominic Welsh, Vitaliy Mezhuyev, Wolfram Irsa. Interdisciplinary Terminology Framework for Teaching and Research in Learning Facto-ries, Procedia Manufacturing, Volume 45, 2020, Pp. 301-306, ISSN 2351-9789, https://doi.org/10.1016/j.promfg.2020.04.021.

[14] Reifer, Abie: Big-Data-Analyse mit der KNIME Analytics Platform und zugehörigen Erweiterungen, https://www.computerweekly.com/de/ratgeber/Big-Data-Analyse-mit-der-KNIME-Analytics-Platform-und-zugehoerigen-Erweiterungen, [02.03.2023].

[15] MAHR GMBH: MarSurf PS 10 C2 - MarSurf Mobiles Rauheitsmessgerät, https://metrology.mahr.com/de/produkte/artikel/6910235-mobiles-rauheitsmessgeraet-marsurf-ps-10-c2, [02.03.2023].