euspen's 24th International Conference & Exhibition, Dublin, IE, June 2024
www.euspen.eu

Surface roughness monitoring and prediction based on audible sound signal with the comparison of statistical and automatic feature extraction methods in turning process

Yaoxuan Zhu[1], Amir Rashid[1], Tomas Österlind[1], Andreas Archenti[1]

[1]KTH Royal Institute of Technology, Department of Production Engineering, Manufacturing and Metrology Systems Division, 11428, Sweden

Yaoxuanz@kth.se

Abstract

In the turning process, the surface roughness of the machined part is considered a critical indicator of quality control. Provided the conventional offline quality measurement and control is time-consuming, with slow feedback and an intensive workforce, this paper presents an online monitoring and prediction system for the effective and precise prediction of surface roughness of the machined parts during the machining process. In this system, the audible sound signal captured through the microphone is employed to extract the features related to surface roughness prediction. However, owing to the nonlinear phenomena and complex mechanism causing surface quality in the whole process, the selection of statistical features of the sound signal in both the time and frequency domains varies from one case to another. This variation may lead to false prediction results as sufficient domain knowledge is required. Therefore, the versatile and knowledge-independent features extraction method is proposed, which exploits deep transfer learning to automatically extract sound signal features in the time-frequency domain through pre-trained convolution neural networks (pre-trained CNN). The performance of prediction models based on two feature extraction methods – statistical feature extraction and automatic feature extraction was further tested and validated in the case study. The results demonstrate that the performances of the prediction model built on the automatically extracted features outperformed that developed with the statistical feature method concerning the accuracy and generalization of the prediction model. In addition, this study also provides solid theoretical and experimental support for developing a more precise and robust online surface quality monitoring system.

Keywords: Data-driven monitoring, surface roughness prediction, transfer learning, audible sound, automated feature engineering

1. Introduction

During the machining process, the surface quality monitoring system has huge potential for online detecting and predicting surface roughness, which is regarded as a fundamental indicator of surface quality control of machined workpieces. To improve prediction performance of such system, various sensing techniques have been applied and investigated [1]. Due to easy access and low cost, the application of audible sound signals captured via a microphone has attracted more attention to the development of machining monitoring systems [2]. However, the main challenge is how to extract hidden information characteristics from the sound signal that correlates to surface roughness. An approach to tackle this is based on feature engineering – feature extraction. Although varied feature extraction methods have been independently discussed, the performance comparison of them under a unified dataset has, to the authors' knowledge, not been studied and published.

2. Methodology

To quantify the performance comparison of different feature extraction scenarios, a methodology is proposed and demonstrated as in the flowchart shown in **Figure *1***. Firstly, the raw sound signal in the time domain was recorded during the experiment (detailed description in **Section 3**) for each cutting test. Afterwards, it was divided into constant time interval length (10 s) blocks with corresponding surface roughness (R_a) measurements. The surface roughness was used as the prediction label. To enlarge the amount of dataset for model training and testing, each collected 10 s sound signal was further subdivided into segments with three different time lengths (1 s, 5 s, and 10 s), determining the total amount of dataset. Following this, each subdivided segment was separately analysed and converted into a frequency domain by power spectrum density (PSD) and time-frequency domain in the form of a generated 2D RGB image (256x256x3) - spectrogram by short-time Fourier transformation (STFT).

To acquire hidden information characteristics of sound signal correlated to the surface roughness, two feature extraction scenarios are proposed and compared. In the statistical feature-based scenario, features of the sound signal segment in both time and frequency domains were extracted from defined statistical features shown in **Table *1***. As a mature convolution neural network (CNN) architecture, VGG16 was employed to achieve automated feature extraction in this case, which was initially developed for object recognition by Oxford's Visual Geometry Group (VGG) and then widely applied to transfer learning tasks. In transfer learning-based scenarios, each generated spectrogram was fed into pre-trained VGG16 [3], in which its architecture was modified by removing the top two layers - the fully connected and classification layer and other layers were reserved and equipped with pretrained weights acquired in the training process of ImageNet dataset. Two feature groups of each spectrogram were respectively generated from nontrainable VGG16 with fixed original pre-trained weight in each layer and from trainable VGG16, in which all weights were fine-tuned during the training process.

Features extracted based on different scenarios associated with corresponding surface roughness were used as input data. They were split into training data (70% of all data), validating data (20% of all data) and testing data (10% of all data) for the establishment of two prediction models: support vector regression (SVR) and artificial neural networks (ANN). During the training process, Bayesian optimization [4] was used for automatic hyper-parameters tuning and mean squared error (MSE) was utilized as the loss function. Iterative runs resulted in a well-trained model with optimal hyper-parameters to obtain predicted surface roughness values.

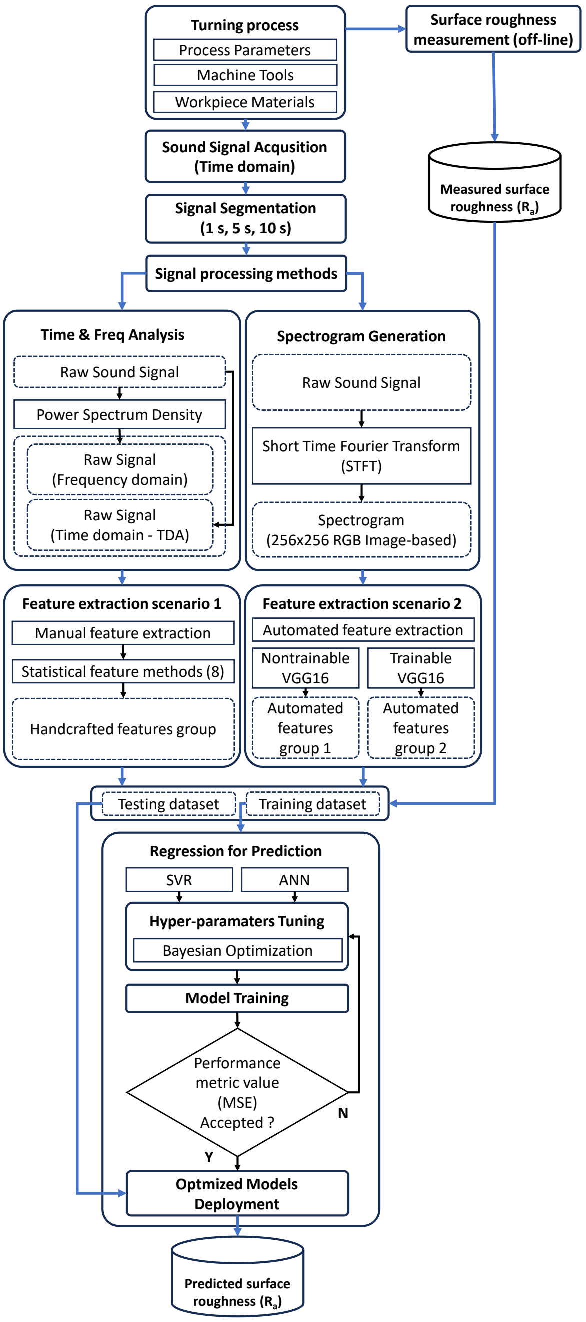

Figure 1. Flowchart of methodology.

Table 1. The type of statistical features (8 types in total) extracted from sound signal in the time domain (TDA) and frequency domain (PSD).

Statistical Feature	Formula ($\mu = \frac{\sum_{i=1}^{N=i} x_i}{N}$, $\sigma = \sqrt{\frac{\sum_{i=1}^{N=i}(x_i-\mu)}{N}}$)	Extracted Sound feature
Maximum	f1 = $max\|x_i\|$	$S1_{TDA}^i, S1_{PSD}^i$
Minimum	f2 = $min\|x_i\|$	$S2_{TDA}^i, S2_{PSD}^i$
Root means square	f3 = $\sqrt[2]{\sum_{i=1}^{N} \frac{x_i^2}{N}}$	$S3_{TDA}^i, S3_{PSD}^i$
Peak to peak	f4 = f1 − f2	$S4_{TDA}^i, S4_{PSD}^i$
Energy	f5 = $\sum_{i=1}^{N=i} \frac{\|x_i\|^2}{N}$	$S5_{TDA}^i, S5_{PSD}^i$
Skewness	f6 = $\frac{\sum_{i=1}^{N=i}(x_i-\mu)^3}{(N-1)\sigma^3}$	$S6_{TDA}^i, S6_{PSD}^i$
Kurtosis	f7 = $\frac{\sum_{i=1}^{N=i}(x_i-\mu)^4}{(N-1)\sigma^4}$	$S7_{TDA}^i, S7_{PSD}^i$
Entropy	f8 $= -\sum_{i=1}^{N=i} p(z_n)\log_2 p(z_n)$	$S8_{TDA}^i, S8_{PSD}^i$

3. Case study

The experiment was conducted to collect data containing surface roughness and sound signals in dry turning operation in a CNC lathe (SMT Swedturn 300) without significant background noise. The workpiece material is the hardened and tempered tool steel - Toolox33. The workpieces are cylindrical bars with a length of 550 mm and a diameter of Ø124 mm. The tooling system incorporates the insert (CNMG 12 04 08-PM 4425, Sandvik Coromant) with a nose radius (RE) of 0.8 mm and the tool holder (DCLNL 2525M, CoroTurn). During the machining process, two cutting parameters, cutting speed (320 and 280 m/min) and feed rate (0.4, 0.3 and 0.2 mm/rev), were set as variable factors to develop a Taguchi orthogonal experiment with six parameter combinations. The depth of cut was constant at 1 mm. Under each parameter combination, the test was replicated twice, in which the workpiece was machined from one run to another with the total cutting length 480 mm per run until flank wear of the cutting tool reached 0.3 mm as standard tool worn-out criteria, which was measured through the digital microscope (Dino-lite RK-10A). During each cutting run, the audible sound signal was captured through a microphone (Microtech GEFELL MKS 211) located at the turret, which was later processed by a data acquisition system (Siemens LMS SCADAS Mobile SCM01) with a 40 kHz sampling frequency. Subsequently, the captured signal was subdivided into each single segment with constant time interval (10 s) as each sampling area (seg.1, seg.2 …) where corresponding surface roughness – R_a (referred as arithmetic average value of surface roughness) were measured three times by profilometer (Mitutoyo SJ-210) at different angles (0°, 120°, 240°) around the cylinder bar then averaged as the input label to the prediction models, see **Figure 2**.

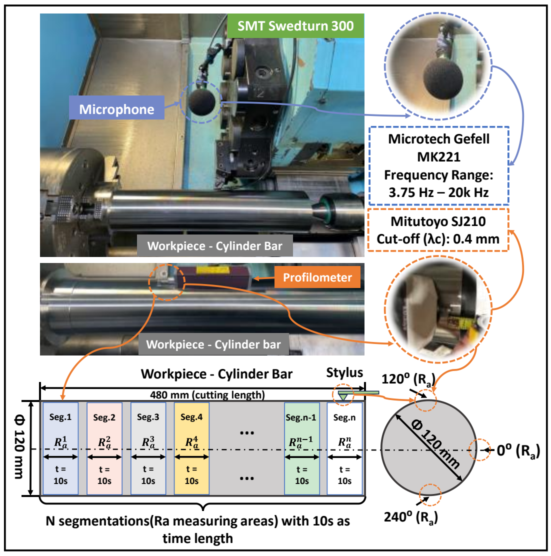

Figure 2. Experiment setup and surface roughness (R_a) measurement.

4. Numerical results analysis and discussion

The performance metrics were employed to estimate and compare the accuracy and reliability of two predictors: support vector regressor (SVR) and artificial neural networks (ANN) used for surface roughness prediction. The proposed metrics include means absolute error (MAE), mean square error (MSE), relative error (ER), average value and standard deviation of prediction accuracy, and coefficient of determination (R^2), which are described in Eqs. (1)-(6), respectively, where y_i denotes the actual value of measured or observed surface roughness collected in the experiment, $\hat{y}_i$ expresses the predicted surface roughness value as each single output of applied predictor, and n represents the total amount of testing data.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (1)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (2)$$

$$Relative\ error = \left|\frac{(y_i - \hat{y}_i)}{y_i}\right| \quad (3)$$

$$\overline{Accuracy} = \frac{1}{n}\sum_{i=1}^{n}\left(1 - \left|\frac{(y_i - \hat{y}_i)}{y_i}\right| \times 100\%\right) \quad (4)$$

$$STD(Accuracy) = \sqrt{\frac{\sum_{i=1}^{i=n}(Accuracy_i - \overline{Accuracy})^2}{n}} \quad (5)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (6)$$

Based on defined performance metrics, the surface roughness prediction performance of two predictors was quantitatively measured and compared on the testing dataset with different feature extraction scenarios from two aspects. Firstly, in each predictor, prediction performance was compared under three distinct feature extraction scenarios when sound signal segments were at the same time interval. Secondly, in each feature extraction scenario, the performance of each predictor was further analysed under sound signal segments with different time interval lengths.

As seen in **Figure 3**, when SVR was chosen as the prediction model, the application of input features extracted via trainable VGG16 from the sound signal segment with all three different time interval lengths (1 s; 5 s; 10 s) provided the superior prediction performance, as measured by MAE (0.10; 0.13; 0.24), MSE (0.03; 0.04; 0.12), average prediction accuracy (96.1%; 94.6%; 89.6%), and R^2 (0.97; 0.95; 0.86). These results were tightly followed by the SVR model trained with input features extracted from nontrainable VGG16, while the SVR model developed with statistical features rendered the worst prediction performance estimated and analysed with all defined performance metrics. Besides, the same conclusion is also reflected in **Figure 4**, which depicts the relative error distribution of predicted data points. With the lowest average value (3.9%; 5.4%; 10.4%) and narrowest range of relative error, the SVR model coupled with features extracted from trainable VGG16 presented the lowest prediction error regardless of time interval lengths (1 s; 5 s; 10 s) of sound signal segments applied for feature extraction. Moreover, as shown in **Figure 5** and **Figure 6**, similar behaviour occurred in the ANN model. Features extracted via trainable VGG16 achieved the best prediction performance exerting sound signal segments in all three varied time interval lengths, which can be verified with lowest MAE (0.09; 0.15; 0.21), MSE (0.02; 0.05; 0.08), average relative error (3.4%; 5.9%; 8.7%), highest prediction accuracy (96.5%; 93.6%; 90.8%) and R^2 (0.97; 0.61; 0.90). Additionally, the exception appeared in the performance comparison between features extracted from nontrainable VGG16 and statistical features, which was different from the SVR model. When sound signal segment with 1 s and 10 s time interval lengths, it was concluded that the ANN model combined with features from nontrainable VGG16 outperformed the ANN model trained with statistical features. Nonetheless, when sound signal segments with 5 s time interval length were employed, this conclusion was the opposite: that the ANN model developed with statistical features achieved better performance than features extracted from nontrainable VGG16.

Within each feature extraction scenario, the influence of sound signals with different time interval lengths on the performance of each prediction model was further compared. In both SVR and ANN, features obtained from sound signal segments with shorter time interval lengths were prone to offer better prediction performance. One exception appeared in the ANN model trained with features extracted from nontrainable VGG16, that sound signal segments with 5 s and 1 s time intervals provided almost the same prediction performance as illustrated in **Figure 4** & **Figure 6**.

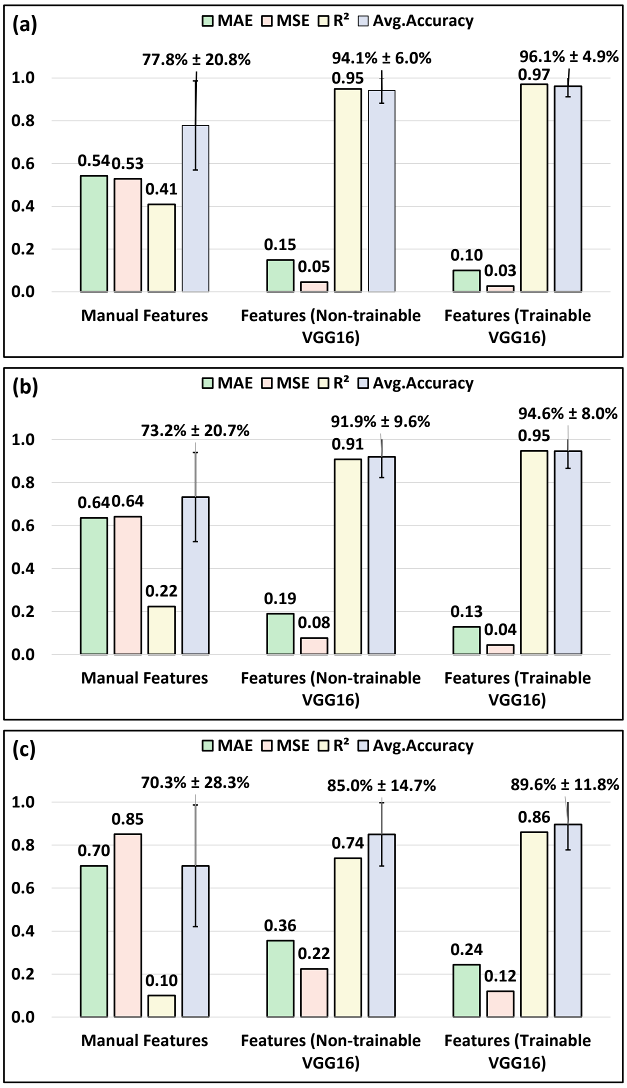

Figure 3. Performance metrics of prediction results based on multiple feature extraction methods from SVR as a predictor; (a): 1 s time length

of the sound signal as input data; (b) 5 s time length of the sound signal as input data; (c) 10 s time length of the sound signal as input data.

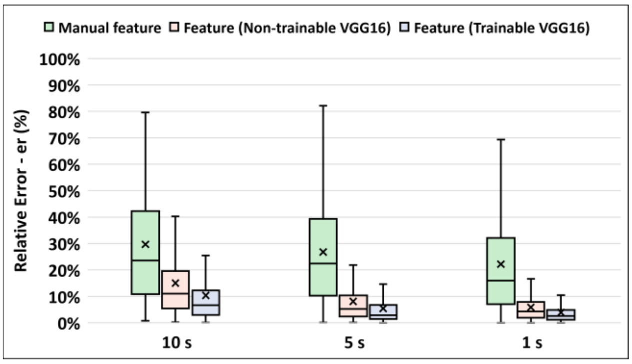

Figure 4. Relative error - er (%) of surface roughness prediction in SVR based on different feature extraction scenarios as input data with different time interval lengths of sound signal segments.

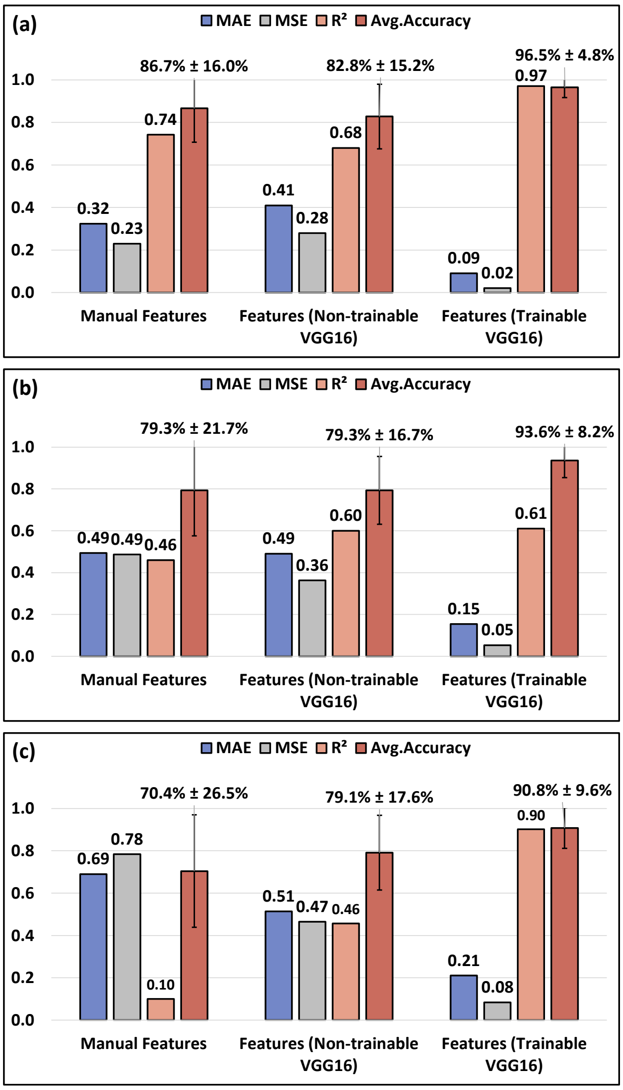

Figure 5. Performance metrics of prediction results based on multiple feature extraction methods from ANN as predictor; (a): 1 s time length of the sound signal as input data; (b) 5 s time length of the sound signal as input data; (c) 10 s time length of sound signal input data.

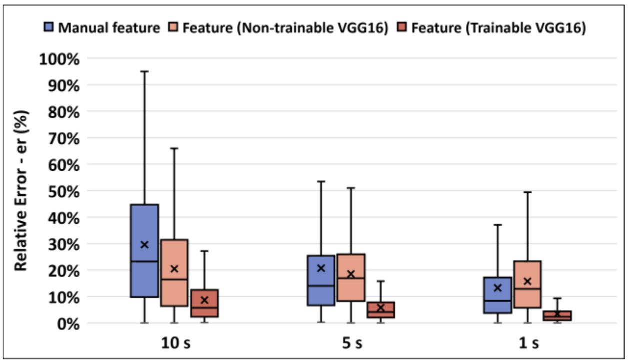

Figure 6. Relative error - er (%) of surface roughness prediction in ANN based on different feature extraction scenarios as input data with different time interval lengths of sound signal segments.

5. Conclusion

This paper proposes a novel approach to verify and compare the influence of varying feature extraction scenarios applied to the sound signal segment with different time interval lengths, including (1) statistical features from both the time and frequency domain of sound signal, (2) automated features directly extracted from sound signal spectrograms via nontrainable pre-trained VGG16 and (3) automated features extracted from sound signal spectrograms via trainable or fine-tuned pre-trained VGG16 on performances of surface roughness (R_a) prediction in two predictors – support vector regression and artificial neural networks. The overall results indicate that compared with statistical features, the automated feature enables the extraction of more valuable hidden information characteristics from the sound signals, representing a stronger correlation to the final prediction target – surface roughness. Based on its superior performance, automated feature engineering is conducive to the establishment of a surface quality monitoring system in terms of improved prediction accuracy, generalization, and versatility with a low requirement for domain expertise in the condition of a large dataset. Given that only one pre-trained CNN was applied in this case, future work will be focused on the exploration of other more advanced pre-trained CNNs, including but not limited to ResNet, Inception-ResNet and vision transformer.

References

[1] K. He, M. Gao, and Z. Zhao, "Soft Computing Techniques for Surface Roughness Prediction in Hard Turning: A Literature Review," *IEEE Access*, **vol. 7**, pp. 89556–89569, 2019, doi: 10.1109/ACCESS.2019.2926509.

[2] P. J. Papandrea, E. P. Frigieri, P. R. Maia, L. G. Oliveira, and A. P. Paiva, "Surface roughness diagnosis in hard turning using acoustic signals and support vector machine : A PCA-based approach," *Appl. Acoust.*, **vol. 159,** p. 107102, 2020, doi: 10.1016/j.apacoust.2019.107102.

[3] C. Chung *et al.*, "Published as a conference paper at ICLR 2015 VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION Karen," *Am. J. Heal. Pharm.*, **vol. 75**, no. 6, pp. 398–406, 2018.

[4] V. Nguyen, "Bayesian optimization for accelerating hyper-parameter tuning," *Proc. - IEEE 2nd Int. Conf. Artif. Intell. Knowl. Eng. AIKE 2019*, **pp. 302–305**, 2019, doi: 10.1109/AIKE.2019.00060.