Embedding Intelligent Features for Vibration-Based Machine Condition Monitoring

Christian Reich
Corporate Research
Robert Bosch GmbH
Renningen, Germany
christian.reich@de.bosch.com

Ahmad Mansour
Corporate Research
Robert Bosch GmbH
Renningen, Germany
ahmad.mansour@de.bosch.com

Kristof van Laerhoven
Ubiquitous Computing
University of Siegen
Siegen, Germany
kvl@eti.uni-siegen.de

***Abstract*—Today's demands regarding workpiece quality in cutting machine tool processing require automated monitoring of both machine condition and the cutting process. Currently, best-performing monitoring approaches rely on high-frequency acoustic emission (AE) sensor data and definition of advanced features, which involve complex computations. This approach is challenging for machine monitoring via embedded sensor systems with constrained computational power and energy budget.**

To cope with constrained energy, we rely on data recording with microelectromechanical system (MEMS) vibration sensors, which rely on lower-frequency sampling. To clarify whether these lower-frequency signals bear information for typical machine monitoring prediction tasks, we evaluate data for the most generic machine monitoring task of tool condition monitoring (TCM).

To cope with computational complexity of advanced features, we introduce two *intelligent preprocessing* algorithms. First, we split non-stationary signals of recurrent structure into similar segments. Then, we identify most discriminative spectral differences in the segmented signals that allow for best separation of classes for the given TCM task. Subsequent feature extraction only in most relevant signal segments and spectral regions enables high expressiveness even for simple features.

Extensive evaluation of the outlined approach on multiple data sets of different combinations of cutting machine tools, tool types and workpieces confirms its sensibility. Intelligent preprocessing enables reliable identification of stationary segments and most discriminative frequency bands. With subsequent extraction of simple but tailor-made features in these spectral-temporal regions of interest (RoIs), TCM typically framed as multi feature classification problem can be converted to a single feature threshold comparison problem with an average F1 score of 97.89%.

***Index Terms*—Tool condition monitoring, non-stationary signals, segmentation, mixture model, Internet of things (IoT)**

I. Introduction

To maximize efficiency in modern workshops' processing sequences, the processing time has to be reduced to a minimum under the constraint of optimal workpiece quality. The most important influencing factor for workpiece quality is condition of the cutting tool, which gradually becomes dull by processing of workpieces and has to be sharpened again.

Today, workpiece quality is typically monitored via manual inspection by the machine's operator. This is suboptimal, as manual inspection takes time in which workpieces could be processed and can be quite subjective. Automatic monitoring of process quality leads to a more stable quality of workpieces while freeing the operator from manual quality inspections.

Machine monitoring involves measuring physical properties of the processing sequence. Industrially established monitoring systems typically rely on high-frequency AE recordings (up to several MHz) [1] and evaluation of the AE root mean square (RMS) signal [2]. As the ultimate goal of our research is to develop an integrated multi-sensor system with built-in algorithms, we aim to monitor tool condition based on signals recorded with MEMS vibration sensors, which have a smaller sampling frequency and are thus expected to have lower energy demands.

Furthermore, we want to study the benefit of *intelligent preprocessing* techniques to address computational constraints of embedded sensor systems and complexity of given signals. Signals show recurrent structure due to reoccurring sequences of processing steps for all workpieces. Furthermore, signals behave non-stationary across these successive processing steps but stationary in each single processing step. As segments reoccur for all signals, identifying recurrent segments allows to extract features (i.e., computable properties of underlying signals) in parts of signals where they are most similar.

Currently, most monitoring applications for non-stationary machine tool signals rely on feature extraction for complete signals [3] and often involve costly computations e.g. of discrete wavelet transform (DWT) [4] or artificial neural networks (ANNs) [5] which are challenging for sensor systems with hard computational constraints. This computational burden might be reduced by intelligent preprocessing, i.e. finding relevant signal segments and most discriminative frequency bands automatically in a one-time bedding-in process of the cutting machine tool. Afterwards, results of detecting most discriminative RoIs both in time domain (TD) via segmentation and frequency domain (FD) via frequency band selection can be used to extract simple but meaningful, tailor-made features only in these RoIs. In this way, intelligent preprocessing involves a one-time additional computational burden but reduces the running computational cost and thus energy consumption introduced by feature extraction.

The benefit of intelligent preprocessing shall be evaluated for the TCM task and verified with data sets for common parameter variations. As the majority of data were recorded at centerless external cylindrical (CEC) grinding machines, we discuss the grinding process in the following section.

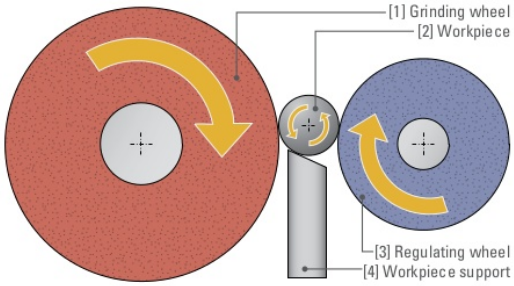

Fig. 1. Process of external cylindrical grinding [6].

II. Centerless External Cylindrical Grinding

The arrangement of CEC grinding machine parts is depicted in Fig. 1. The workpiece is situated between grinding wheel and regulating wheel on the workpiece support. The grinding wheel approaches the workpiece and starts the machining of the workpiece. Workpiece support and regulating wheel decelerate the workpiece. When approaching the workpiece, the grinding wheel feed (velocity of moving grinding wheel to workpiece) is reduced when assuming to be close to contact. This phase of reduced feed (caused by uncertainty about workpiece contact) is referred to as *air grinding*.

Due to processing of the workpieces, the grinding wheel's surface gradually becomes dull. To preserve high quality of the processed workpieces, the grinding wheel has to be sharpened (*dressing*) by removing the topmost grinding wheel layer by a dressing tool. Both dressing after a fixed number of processing cycles and dressing subject to visual inspections of workpieces are suboptimal regarding consumption of grinding wheels and subjectivity of the visual inspection, respectively. Thus, automatically finding the optimal point in time when dressing is needed is beneficial regarding processing efficiency.

III. Data

This study's data were recorded at different cutting machine tools using MEMS vibration sensors. The vibration sensors have a single degree of freedom and sample at a rate of 62.5 kHz. Sensors were mounted at different places of interest and connected via a gateway system. For TCM, the workpiece support proved to be a suitable measurement place.

Workpieces of different geometry and dimensions were processed with grinding wheels of different profile to verify generalizability of the suggested algorithms. Details are listed in Table I. While data sets DS1 to DS4 cover expected data variance by change of grinding wheels and workpieces, DS5 contains data of several month to incorporate unexpected variance not considered by parameter variations of former data.

All data records consist of an air grinding phase and a grinding phase. For some data sets, the grinding phase can be separated into different phases resembling the processing steps applied to the workpiece. Two DS2 sensor signals for a sharp and a dull grinding wheel and their corresponding short-time Fourier transform (STFT) spectrograms are shown in Fig. 2. Differences in sharpness can be observed best in amplitude of

TABLE I
Data sets and parameter variations

Data set	*Records*	*Machine type*	*Tool type*	*Workpiece*
DS1	320	CEC grinding 1	1	1
DS2	97	CEC grinding 1	2	2,3
DS3	1921	CEC grinding 2	3	4
DS4	280	CEC grinding 2	4	5
DS5	26770	CEC grinding 3	5,6,7	6,7
DS6	84	Turning	8	8

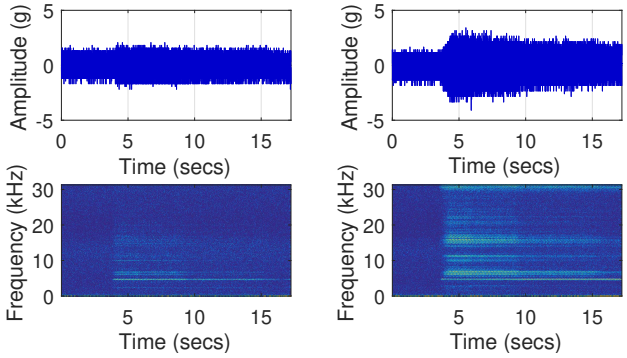

Fig. 2. Exemplary DS2 vibration sensor signals and corresponding STFT spectrograms for a sharp (left) and a dull (right) grinding wheel.

the TD signal's first contact phase and in FD by concentration of energy in different frequency bands. Thus, identifying both process phase segments and frequency bands with largest discriminative differences in spectral energy between signals for sharp and dull suggest to be important preprocessing steps to extract meaningful, tailor-made features.

IV. Related Work

A. Time Series Segmentation

Many algorithms for time series segmentation rely on piecewise linear approximation (PLA) [7]. These algorithms aim to reduce complexity of signals while preserving relevant information by representing segments of successive data points by linear approximations. However, for the given signals with deterministic segment structure (segments reflect process phases), PLA-based algorithms lead to overly complex segmentation results (more segments than process phases assigned). Furthermore, PLA-based algorithms do not lend naturally to model selection, i.e. selection of the model explaining the given data best without involvement of the machine operator. For the given signals with recurrent segments reflecting process phases, this model selection step is crucial.

Hidden Markov models (HMMs) are widely used models for time series classification and come with inherent segmentation of the signals [8]. HMMs are suited for a wide range of temporally structured prediction tasks by allowing to incorporate prior knowledge into the learning process of state transition probabilities a_{ij}. Learning a left-to-right (L2R) structure resembles the given signals by imposing constraints on state transitions to follow a strict temporal order. By

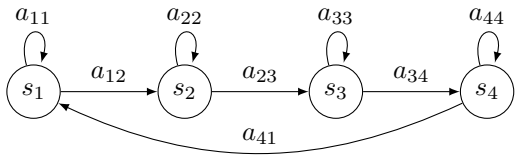

Fig. 3. 4-state-example of the proposed L2R postprocessing architecture. Here, states model the process phases we aim to detect. a_{ij} is the probability to transition from state s_i to state s_j.

allowing a transition a_{41} from the last state to the first we are able to mimic the recurrent structure of our data (Fig. 3).

Thus, HMMs lent themselves naturally for the given time series segmentation. While costly one-time training can be implemented on a connected gateway system with higher computational power, online prediction of streaming data on the sensor system with HMMs involves solving a complex Dynamic Programming problem (i.e., computing the Viterbi path). Gaussian mixture models (GMMs) are computationally less expensive and model class membership by probabilistic means similar to HMMs. However, they do not encompass temporal dependencies of successive data points in time series. Thus, after predicting most likely process phase cluster membership with the GMM, we postprocess these cluster estimates by an additional finite state machine (FSM) like depicted in Fig. 3 to mimic the HMM's L2R structure.

B. Features

Among others, typical TCM features involve measures based on autoregressive moving average (ARMA) models [4], Wavelet-based methods [9], and sparse dictionary learning [10]. A comprehensive overview of features is given in [4]. Lately, most of research is dominated by ANN techniques, e.g. self-organizing maps (SOMs) [11] and deep neural networks (DNNs) [5]. These advanced features are necessary to address non-stationary complexity of the signals. In this study, we aim to sustain expressiveness of these features while reducing computational burden by extracting simple but tailor-made features in automatically detected process phases and frequency bands.

V. Methods

A. Detection of Process Phases

The proposed phase detection algorithm is summarized in Algorithm 1. Based on extraction of features in successive signal segments of 1024 samples, we train a GMM to identify clusters in each TD signal. We chose features $MeanAbs_{TD} = \frac{1}{N}\sum_{i=1}^{N}|x_i|$ (x_i being *i*-th signal sample) and $Power_{FD} = \frac{1}{M}\sum_{i=1}^{M}U_i^2$ (U_i being *i*-th spectral magnitude).

The learning task for GMMs is to find the proportions π as well as mean μ and covariance structure $\mathbf{\Sigma}$ for each of a mixture of Gaussian distributions [12]. To find the optimal number of mixture clusters matching the (unknown) number of process steps, we compare Akaike information criterion (AIC) [13] and Bayesian information criterion (BIC) [14]. Both criteria consist of two terms: a goodness-of-fit term and a term penalizing model complexity. While AIC measures complexity only by the number of model parameters k, BIC incorporates the sample size N into the penalty term:

$$AIC = -2\,log\,L(\theta) + 2\,k, \tag{1}$$

$$BIC = -2\,log\,L(\theta) + k\,log(N), \tag{2}$$

where L represents the likelihood function and $\theta = (\pi, \mu, \mathbf{\Sigma})$ the set of learned model parameters. Both criteria identify the GMM explaining feature score clusters in the most sensible way, but do not allow to incorporate temporal information inherent to successive feature vectors. We make use of an L2R FSM (cmp. Fig. 3) to incorporate this information. In order to define the temporal order of FSM states s_i (representing the deterministic sequence of processing steps) we need to identify the corresponding GMM clusters. We do this for each cluster by computing the median value of all segments' temporal indices associated with this cluster. Thus, the order of clusters' median values reflects the order of processing steps. This allows to reduce the space of admissible state transitions and thus constrain transitions to follow a temporal L2R order.

B. Frequency Band Selection

The proposed frequency band selection algorithm is summarized in Algorithm 2. Discriminative differences in distribution of spectral energy for differently labeled signals (dull/sharp) can be found by computing the l_2-distance (Euclidean distance) of their respective fast Fourier transform (FFT) spectra.

Labels (sharp/dull) are allocated via an additional sensor at the dressing motor. For signals recorded directly before dressing, the grinding wheel is assumed to be dull, while for signals recorded directly after dressing, the grinding wheel is assumed to be sharp. This procedure yields pairs of signals labeled sharp and dull, for which single-sided spectra are computed via FFT. Computing the squared difference between each of these spectral pairs yields their l_2-distance. Smoothing and ensemble averaging of l_2-distance curves allows to compute reliable estimates of discriminative differences in spectral energy.

C. Tailor-made Features

Based on segmentation of signals into process phases (cmp. Algorithm 1) and determination of frequency bands reflecting the condition of the grinding wheel (cmp. Algorithm 2) we subsequently extract the feature $MeanAbs_{TD}$ across most relevant segments and frequency bands for each signal.

VI. Results

A. Detection of Process Phases

In Fig. 4, DS2 results for detection of number of process steps and corresponding GMM clusters are depicted for two features extracted in 1024 sample segments measuring both TD energy ($MeanAbs_{TD}$) and FD energy ($Power_{FD}$).

GMMs are trained for a number of cluster components between 1 and 10. AIC and BIC estimates are shown in the left part of Fig. 4. Assigning the optimal number of clusters in the elbow point of AIC and BIC plots like suggested in [15] results in an estimate of the best-fit model for a number of three

Algorithm 1: Process Phase Detection

Input: TD signal *currSig*, information criterion *IC*
Output: Process phase transitions in *currSig*

```
/* Identify clusters in currSig                    */
for each 1024 sample segment currSeg in currSig do
    featMat(:,1) = feat1(currSeg);
    featMat(:,2) = feat2(currSeg);
end
scaledFeat = scale(featMat);
for a number of mixture components nComp 1 to 10 do
    GMM(nComp) = trainGMM(featMat,nComp);
    valIC(nComp) = IC(GMM(nComp));
end
GMMSel = GMM(minAngle(valIC));
clustEst = max(posterior(GMMSel,scaledFeat));
/* Tie phase tags to clusters                      */
for all clusters nClust of GMMSel do
    medIndex(nClust) =
     median(indices(clustEst==nClust));
end
sequClust = sort(nClust,medIndex);
/* L2R FSM: Enforce temporal order of
   cluster estimates                               */
fsmMdl = defineL2R(nClust,sequClust);
fsmEst = fsmMdl(clustEst);
phaseTrans = transitions(fsmEst);
```

Algorithm 2: Frequency Band Selection

Input: TD signal segments detected by Algorithm 1, and corresponding labels (dull/sharp).
Output: Frequency bands capturing discriminative information of labeled TD signal segments *currSegDull* and *currSegSharp*.

```
for all pairs (currSegDull,currSegSharp) do
    specDull = singleSidedSpec(currSegDull);
    specSharp = singleSidedSpec(currSegSharp);
    spec += (specDull - specSharp)^2;
end
specMags = find(spec > mean(spec));
RoI = find(length(specMags) > length(spec)/100);
```

clusters. Unlike [15], we introduce finding the elbow point by searching for the smallest angle between connections of each index/value-tuple *(number of clusters, AIC/BIC score)* with its predecessor and successor tuple. Phase borders assigned with the best-fit GMM match the differences in TD envelope amplitude and FD energy distribution (cmp. left part of Fig. 5).

B. Frequency Band Selection

Spectral discriminative RoIs for wheel condition classes dull and sharp in DS2 are shown in the upper right subfigure of Fig. 5. The spectrum was normalized to its maximum magnitude (red dot). RoIs are marked with gray rectangles and

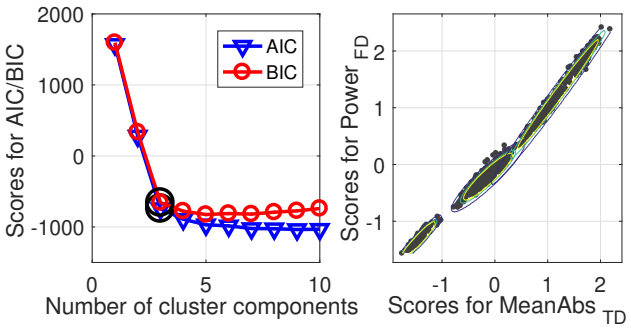

Fig. 4. GMM model selection results for DS2. Left: Best fit for AIC and BIC is found for three clusters (black circles). Right: Scores for features $MeanAbs_{TD}$ and $Power_{FD}$. Equi-probability estimates for mixture component membership found with the best-fit GMM are plotted as contour lines.

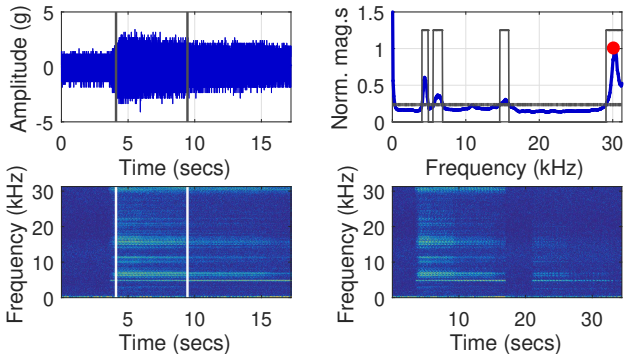

Fig. 5. Intelligent preprocessing results for DS2. Left: Signal of class dull, corresponding STFT spectrogram and phase borders (gray lines). Right: l_2-distance curve (blue line) for two representatives (bottom right) of classes dull (sec 1…17) and sharp (sec 17…34) and assigned RoIs (gray rectangles).

match differences in spectral distribution of the spectrograms in the lower right subfigure of Fig. 5. RoIs are defined relative to the spectrum's mean (dashed gray line), i.e. where the l_2-distance curve (blue line) is above the spectrum's mean for a period of at least 312.5 Hz (1 % of the spectrum's width).

C. Benefit of Intelligent Preprocessing for Feature Extraction

After computation of RoIs, TD signals were filtered in passbands defined by RoI borders. Afterwards, feature $MeanAbs_{TD}$ was computed like introduced in section V-A in signal segments detected by Algorithm 1.

The benefit of intelligent preprocessing for DS2 becomes apparent in Fig. 6. By extracting feature $MeanAbs_{TD}$ only in most discriminative frequency bands and process phase (right subfigure) a monotonic feature score trend between examples of classes dull (red) and sharp (blue) can be established. Also, labeled data points (red, blue) are separated clearly compared to when $MeanAbs_{TD}$ was extracted for the complete spectral range and across signals comprising all process phases.

For multi-contact DS4 signals, the benefit of process phase segmentation over simple contact detection is depicted in Fig. 7. Best results regarding feature score trend and separation of signals labeled dull and sharp are obtained when extracting feature scores only in the last process phase 4, where full

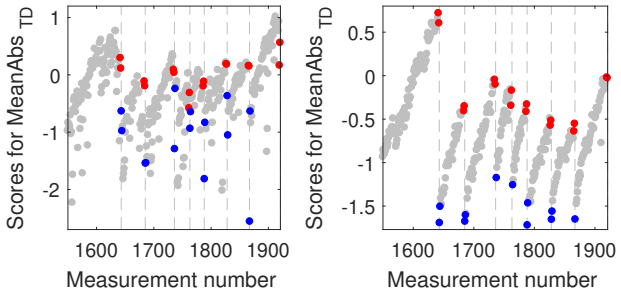

Fig. 6. Benefit of intelligent preprocessing for single-contact DS2 data. Left: Feature $MeanAbs_{TD}$ (gray) for complete signals and frequency range. Dressing times are plotted as vertical dashed lines. Right: $MeanAbs_{TD}$ extracted in most relevant process phase and discriminative frequency bands.

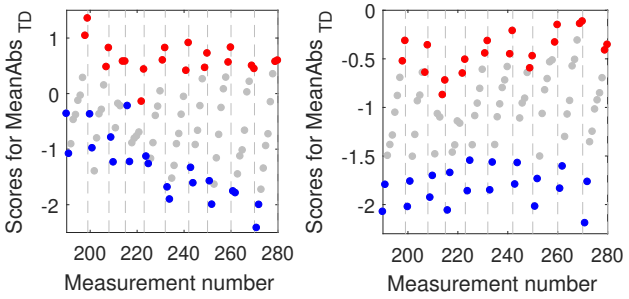

Fig. 7. Benefit of intelligent preprocessing for multi-contact DS4 data. Scores for feature $MeanAbs_{TD}$ are plotted for phase 2 (left) and phase 4 (right).

contact between grinding wheel and workpiece yields best transmission of energy from the former to the latter. Results for first contact phase 2 are plotted on the left for comparison.

Benefits for all data sets are summarized in table II. Two performance measures are listed. Monotonicity of the feature score trend is measured with root mean squared error (RMSE) between feature scores and the fit of a trend function to these data. As evaluations revealed restricted exponential growth of feature score trends (saturation effects by restricted dulling of grinding wheel), we fit to the function $y = a - (a - b)e^{-cx}$ with y being feature scores and x the measurement indices. Monotonicity of feature scores allows assessing tool condition based on simple threshold comparison. Performance of threshold classification is measured via F1 score. Results for tailor-made features based on intelligent preprocessing exceed results for non-preprocessed features for all but one data set.

TABLE II
PERFORMANCE MEASURES: RMSE, F1 SCORE

Data set	W/o intell. preprocessing		W/ intell. preprocessing	
	RMSE	***F1 score***	***RMSE***	***F1 score***
DS1	**0.0299**	**100%**	0.0392	**100%**
DS2	0.3195	93.68%	**0.0858**	**100%**
DS3	0.3288	88.95%	**0.1969**	**92.99%**
DS4	0.2200	74.30%	**0.1305**	**100%**
DS5	0.2803	94.14%	**0.2227**	**94.33%**
DS6	0.4484	75.00%	**0.3182**	**100%**
Average	0.2734	87.68%	**0.1798**	**97.89%**

VII. CONCLUSION

In this study, we validated the benefit of identifying spectral-temporal regions of interest (RoIs) in non-stationary signals for subsequent definition of simple but meaningful tool condition monitoring (TCM) features. To identify RoIs, we have contributed and validated two intelligent preprocessing techniques.

For detection of process phases (via GMMs), AIC and BIC proved suitable to identify best-fit GMMs. Incorporating information about sequentiality of data points by utilizing an L2R FSM increased robustness of process phase estimation.

Automatic selection of discriminative frequency bands (via ensemble-averaged l_2-distances of labeled FFT spectra) led to RoIs matching corresponding STFT spectrograms.

Subsequent extraction of feature $MeanAbs_{TD}$ only in discriminative process phases and frequency bands allowed to increase monotonicity of the feature score trend (from an average RMSE of 0.2734 to 0.1798) and improve separability of labeled signals with a threshold classifier (from an average F1 score of 87.68% to 97.89%).

REFERENCES

[1] J. F. G. Oliveira, E. J. Silva, C. Guo, and F. Hashimoto, "Industrial challenges in grinding," *CIRP Annals-Manufacturing Technology*, vol. 58, pp. 663–680, Aug. 2009.

[2] J. F. G. Oliveira and D. A. Dornfeld, "Application of AE contact sensing in reliable grinding monitoring," *CIRP Annals-Manufacturing Technology*, vol. 50, pp. 217–220, Jul. 2001.

[3] T. W. Liao, "Feature extraction and selection from acoustic emission signals with an application in grinding wheel condition monitoring," *Engineering Applications of Artificial Intelligence*, vol. 23, pp. 74–84, Feb. 2010.

[4] R. Teti, K. Jemielniak, G. O'Donnell, and D. A. Dornfeld, "Advanced monitoring of machining operations," *CIRP Annals-Manufacturing Technology*, vol. 59, pp. 717–739, Aug. 2010.

[5] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring: A survey," *Computing Research Repository (CoRR)*, vol. abs/1612.07640, pp. 1–14, Dec. 2016.

[6] K. C. S. Stoll and K. Otto, *Manual. Centerless External Cylindrical Grinding*. Leipzig, Germany: Schaudt Mikrosa GmbH, 2016.

[7] E. Keogh, S. Chu, D. Hart, and M. Pazzani, "Segmenting time series: A survey and novel approach," in *Data Mining in Time Series Databases*, Singapore, Singapore, Jun. 2004, pp. 1–21.

[8] C. M. Bishop, *Pattern Recognition and Machine Learning*. Berlin, Germany: Springer, 2006.

[9] J. Wang, P. Wang, and R. X. Gao, "Enhanced particle filter for tool wear prediction," *Journal of Manufacturing Systems*, vol. 36, pp. 35–45, Jul. 2015.

[10] S. M. del Campo and F. Sandin, "Towards zero-configuration condition monitoring based on dictionary learning," in *Proc. IEEE of European Signal Processing Conference (EUSIPCO'15)*, Nice, France, Aug. 2015, pp. 1306–1310.

[11] C.-L. Yen, M.-C. Lu, and J.-L. Chen, "Applying the self-organization feature map (som) algorithm to ae-based tool wear monitoring in micro-cutting," *Mechanical Systems and Signal Processing*, vol. 34, pp. 353–366, Jan. 2013.

[12] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, USA: The MIT Press, 2012.

[13] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, pp. 716–723, Dec. 1974.

[14] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, pp. 461–464, Mar. 1978.

[15] Q. Zhao, V. Hautamaki, and P. Fränti, "Knee point detection in BIC for detecting the number of clusters," in *Proc. IEEE International Conference on Advanced Concepts for Intelligent Vision Systems (ACIVS'08)*, Juan-les-Pins, France, Oct. 2008, pp. 664–673.