Journal of Machine Engineering, 2024, Vol. 24, No. 3, 17–31
ISSN 1895-7595 (Print) ISSN 2391-8071 (Online)

Received*: 27 June 2024* / Accepted: *27 August 2024* / Published online: *25 September*

RUL estimation, ball-screw drives,
cloud environment, model uncertainty

Alexander BOTT[1*],
Bolin LIU[2],
Alexander PUCHTA[1],
Juergen FLEISCHER[1]

MACHINE LEARNING-DRIVEN RUL PREDICTION AND UNCERTAINTY QUANTIFICATION FOR BALL SCREW DRIVES IN A CLOUD-READY MAINTENANCE FRAMEWORK

In today's rapidly evolving industrial landscape, efficient predictive maintenance solutions are essential for minimizing downtime and enhancing productivity. This research introduces an adaptive cloud-based model pipeline for predicting the Remaining Useful Life (RUL) of machine components, specifically ball screws. The pipeline integrates local pre-processing, edge computing, and cloud-based adaptive model training, ensuring data privacy and reducing data transmission volumes. The system classifies wear states using various machine learning models and predicts RUL through regression analysis, incorporating uncertainty quantification for robust maintenance scheduling. The experimental setup includes accelerated degradation of ball screws, with data collected via a three-dimensional accelerometer. Feature extraction and data augmentation techniques are employed to enhance prediction accuracy. Random Forest and Gradient Boosting models demonstrate superior performance, with Random Forest selected for its robustness and uncertainty quantification capabilities. Empirical results indicate high prediction accuracy, with Random Forest achieving up to 91% accuracy in Phase 2. This cloud-ready predictive maintenance framework leverages scalable cloud infrastructure for efficient data processing and real-time updates, offering a practical solution for industrial applications. The proposed approach significantly advances the adoption of digital business models within the manufacturing industry, providing a reliable and efficient tool for predictive maintenance.

1. INTRODUCTION

Since introducing the concept of Industry 4.0 in 2011, many industrial applications have been created in various fields [1–5]. Multiple technologies from areas such as the Internet of Things, Big Data, AI (especially machine learning), and finally, cloud computing are proposed and effectively combined to create modern cyber-physical systems (i.e. a highly connected, intelligent, and flexible industrial production facility) [1, 3, 6–8]. This generally enables more adaptable and customized production with a remarkable increase in efficiency

[1] Institute of Production Science, Karlsruhe Institute of Technology, Germany
[2] Institute for Operations Research (IOR), Karlsruhe Institute of Technology, Germany
[*] E-mail: alexander.bott@kit.edu
https://doi.org/10.36897/jme/192681

[9–12]. In practice, however, implementation has many obstacles, especially in the production sector of small and medium-sized enter-prises (SMEs). In addition to a lack of financial resources, management expertise, and planning strategy, the most critical problem triggers are the lack of skilled workers and infrastructure [13–16].

One important example in this context is the condition monitoring or failure prediction of ball screw drives, which are often used in high-precision production machines to convert rotary movements into precise linear positioning. The lifetime of ball screws can deviate considerably from the standard lifetime [17–19]. For example, an over-supply or undersupply of lubricant can lead to an undesirable increase in frictional torque, which can severely affect the life of the ball screw and thus lead to premature failure of the ball screw and unexpected machine downtime, affecting the overall efficiency of the machine [20]. Against this background, it is essential to classify the general state of wear of the components as part of intelligent condition monitoring to solve the problem and, based on this, to create a forecast of the remaining service life.

An intelligent yet safe and trustworthy data infrastructure is essential to implement data-driven models (e.g. machine learning methods). Gaia-X offers one approach, launched in 2019 and aims to create a secure and transparent environment for data ex-change [21, 22]. The data-centric ecosystem brings together research institutions, technology providers and companies to generate new business ideas in a privacy-compliant and standardised way [23, 24]. The standardized data exchange facilitates and supports implementing Industry 4.0 applications such as predictive maintenance and real-time monitoring [25]. In this work, different machine learning models for the remaining useful life (RUL) prediction of ball screw drives are investigated, and a prediction pipeline is pro-posed, which takes the need for a cloud-ready approach into consideration and also includes the uncertainty in the prediction.

2. STATE OF THE ART

Recent studies have extensively explored various methods to monitor and analyse the health condition and wear of ball screws. Schopp (2009) observed increased vibration amplitudes and noise levels under artificial loads, akin to worn-out ball screws in industrial settings. Although the natural frequency decrease indicates preload loss, the high variance due to pitting and bearing ball orientation makes it a suboptimal health indicator. Ex-tending this research, Verl et al. (2009) conducted a run-to-failure experiment with a single axis ball screw under an artificial load. They used time and frequency domain features correlated with wear, such as reversal positioning error and vibration energy, to form a health condition indicator. However, they did not pursue RUL estimation. Furthering Verl et al.'s work, Walther (2011) included motor currents and position signals, finding distinct vibration energy patterns for different wear mechanisms. Nevertheless, the generalizability of these results was limited due to the lack of validation on new specimens and long-term trials. Maier (2015) incorporated feed drive motor torque monitoring and Hilbert-Huang transform features, noting that motor torque varies significantly with position for decom-missioned ball screws. Möhring and Bertram (2012) approached the problem differently by using direct measurement with strain gauges and temperature monitoring to investigate preload. While

long-term testing could detect preload loss, wear detection was hindered by sensor sensitivity to thermal changes. Building on these findings, J. Wen et al. (2018) used vibration signals to study ball screw health, extracting various features. The RMS of vibration signals served as a partial health indicator, showing an increasing trend through different wear stages. They created a health condition model using a multiple classifier system and local class accuracy technique, though its generalizability remains untested. Denkena et al. (2021) conducted experiments on a single-axis ball screw test bench with three preload values, using vibration and acceleration data collected via computer numerical control and an accelerometer. Principal component analysis on time and frequency domain features identified two key components for a preload classification model. Complementing these studies, Schlagenhauf et al. (2022) also performed destructive wear tests on ball screw spindles, using high-resolution pitting damage images for wear progression quantification. Classification methods enabled wear quantification based on visual pitting area characterization.

All the before-mentioned studies use a local processing approach with many inefficiencies against better scalable cloud computing approaches. Because of this, we look into some cloud-based approaches and industry scenarios for condition monitoring in manufacturing. Cloud-based prediction technologies are service-orientated and enable multiple companies to offer and manage prediction services over the Internet [26]. For example, Villalonga et al. [27] built a global monitoring system for a network of CNC machines us-ing local and global control. They evaluated the effectiveness of the architecture using bearing failure benchmarks. Using big data analytics or statistical methods in machine learning is an increasingly important trend in cloud-based condition monitoring. For instance, in [28] Arévalo et al. proposed a cloud-based architecture for condition monitoring using a fusion of classification methods based on the Dempster-Shafer Evidence Theory (DSET) [29]. The method's functionality has been successfully evaluated for small-scale bulk goods plants. In [30], a cloud-based framework for intelligent online diagnostic services was developed specifically for machining difficult-to-cut materials. In addition to a knowledge based detection algorithm for identifying the occurrence of CTF (Critical Tool Failure), the tool life was diagnosed in [30] using a pattern recognition method based on artificial neural networks. In [31], a new method for monitoring wear on CNC milling machines was developed by segmenting and classifying individual machining cycles based on path length, spindle speed, and cycle duration.

Besides the dominating focus on local processing approaches in the condition monitoring of ball screws, the functionally relevant prediction uncertainty aspect has received little attention in the literature. Still, the ability to explain data-driven algorithms and transparency by estimating and visualizing the uncertainty associated with the model's predictions is essential in machine learning to assess whether the model is wrong or doesn't know enough to solve the task [32]. Generally, methods for quantifying (epistemic and aleatory) uncertainty can be divided into Bayesian and frequentist approaches [32, 33]. While Bayesian methods define a priori a hypothesis space of plausible models, frequentist approaches consider how well the distribution over the observations implied by each hypothesis is consistent with the data [32]. Many methods for quantifying uncertainty have been developed in the context of machine learning. For example, formal statistical inference techniques for predictions generated by supervised learning ensembles have been developed

in [34] to quantify uncertainty in random forests through confidence intervals and hypothesis testing.

Despite the above methods, to the best of our knowledge, there is no comparable approach in the field of predictive maintenance of ball screw drives that combines cloud technology with machine learning and uncertainty quantification, which, in addition to high accuracy, can lead to efficient data storage, processing and updating in real-time.

3. PROPOSED PREDICTION PIPELINE FOR ADAPTIVE CLOUD-BASED MODEL USAGE

The proposed pipeline consists of three main components to address different issues in evaluating sensor signals and analysing them to predict the RUL of machine components. In most cases, sensor data for machine components has a relatively high sample rate with hundreds or thousands of measurement values per second. This leads to the need for a local pre-processing step, which can be carried out via edge computing. It enables simple anonymisation, which plays an important role in data protection issues for specific or highly relevant production processes. Another benefit of local preprocessing is the reduced volume of data that needs to be sent from the machine (on-premise) to the cloud solution, where evaluation models are often run as shown in state-of-the-art methods. As shown in Fig. 1, a message broker is the interface between the cloud-run data storage and the machine learning models for specific predictions regarding the monitored component or machine.

The main issue with most prediction tasks in the context of RUL predictions is the shortage of sufficient data points covering different wear states and operating conditions, as seen in section 4. The main reason is that rather complex, time-consuming, and expensive experiments are needed to track the degradation process for different components. Additionally, due to production tolerances and assembling factors, the intrinsic variety of mechanically produced parts increases the variety of possible sensor signals. This leads to the necessity of increasing the number of monitored components to increase and generalize the prediction quality. This issue can be compensated via adaptive model training like in [35], fed through a continuous stream of production data to cyclic train new models on a steadily increasing training base or meta-learning approaches as in [36]. This enables continuous improvement, especially in the generalization of the models. The different resulting model instances can be combined and connected in an ecosystem like Gaia-X to address various business models. These business models can accommodate different pricing or data exchange scenarios, such as “pay with data,” which benefits both the prediction service provider and the prediction consumer. The provider gains increased training data, enhancing model quality, while the consumer receives higher quality predictions for more efficient machine or component use. Additionally, using an ecosystem with specialized participants and services for tasks like storage or computing can increase the efficiency of the individual modules.

To enable such a scenario, the first step in implementing an example ecosystem to showcase the descriptive functionalities is conceptualizing and implementing the modules, as shown in Figure 1. In this work, we first focus on the data science perspective to investigate a suitable preprocessing and modelling approach. Furthermore, we evaluate different feature

based approaches in terms of their ability to predict RUL to consider the whole ecosystem in our further research. We used a laboratory set to acquire experimental data, as described below. Future work will focus on the capability of different approaches to continuously incorporate new training data and increase performance.

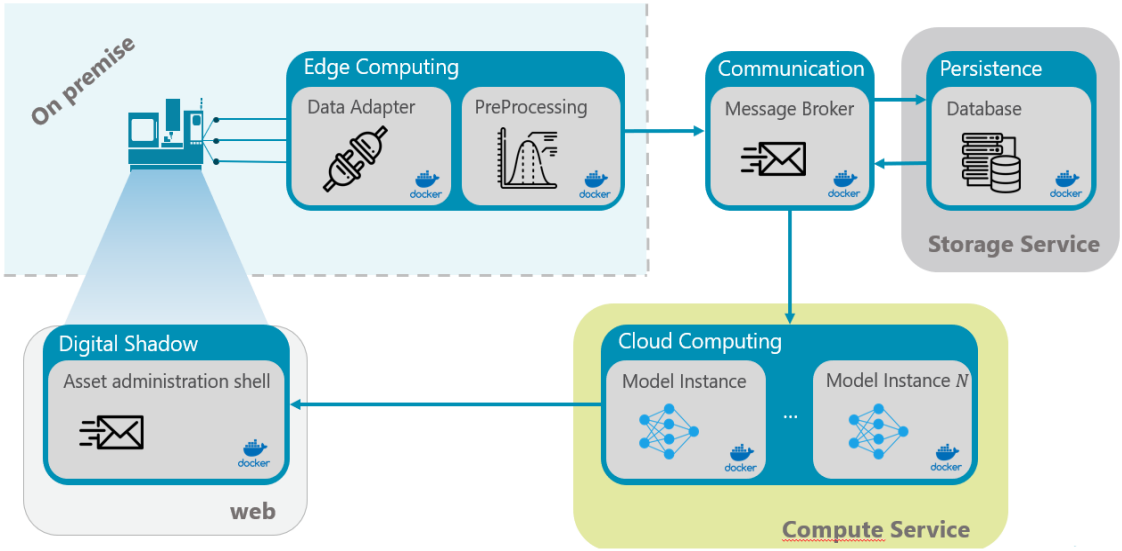

Fig. 1. Proposed prediction Pipeline for adaptive cloud-based model usage

4. EXPERIMENTAL SETUP AND DATA ACQUISITION

The experimental setup consists of one ball screw, one motor, one hydraulic cylinder to apply load and a pair of linear guide rails. The ball screw is mounted on bearings that are connected to the test bench and is driven by the motor. During the accelerated degradation processes, a force of 30 kN is applied through mechanical screws axially. A three-dimensional accelerometer is mounted on the screw nut. The x-direction is defined as the screw nut direction of motion, which is the axial direction of the spindle. The y-direction is the opposite direction of gravity. The z-direction is defined by the right-hand rule and the two directions mentioned above. Due to the durability of ball screws, the degradation process of the ball screw is accelerated. Driven by the motor, the screw nut cyclically moves back and forth with a velocity of 300 rpm between the two ends of the ball screw, which spans 600 mm. This degradation acceleration movement is performed continuously with intermittent interruptions for measurements triggered manually, and the experiment is terminated when the experiment operator deems the ball screw to be sufficiently degraded based on visual inspection of the metal debris.

The manually triggered data collection of vibration signals in the *x*-, *y*-, and *z*-direction is performed daily, with multiple measurements throughout the day. During measurement sessions, the hydraulic cylinder applies an axial load of 2 kN. Each time a measurement is made, the ball nut is moved from one end of the spindle to the other, i.e. end-to-end, with no turnaround, with a velocity profile shape as accelerated degradation. While the target velocity

for accelerated degradation is only 300 revolutions per minute, three measurement samples with individual end-to-end velocities (Speed 1 = 300 rpm / Speed 2 = 450 rpm / Speed 3 = 600 rpm) are collected during each measurement session. The time intervals between measurements are approximately 1 hour. The number of measurements per day is regular. Data is collected on two ball screws of the same model and specification, which all undergo the same accelerated degradation resulting in two data sets (dataset 1 and dataset 2). This paper uses the measurements with a target velocity of 450 rpm (Speed 2) for the following RUL modelling in the degradation process. Considering the possible anomalies from the measurement sensors, an outlier filtering is performed on the raw vibrational signal data as described in [37]. The first quantile (Q1, 25th percentile) and third quantile (Q3, 75th percentile) of the data are used for filtering. The interquartile range (IQR), calculated as the difference between Q1 and Q3, measures statistical dispersion and defines the range where most data points lie. Outliers are values significantly higher or lower than the majority of data points. The outlier boundaries are set as described in (1), which is a generous range that permits a wide range of data points to be considered non-outliers. Data points detected outside the outlier boundaries are not considered for the following feature generation.

$$Q1 - 6 * IQR \cup Q3 + 6 * IQR \tag{1}$$

5. FEATURE-BASED WEAR-STATE PREDICTION

Feature extraction is performed on the filtered experimental datasets. The evaluation and selection of a feature selection method are performed in tandem with the evaluation and selection of data augmentation methods to enhance the existing data. The performances of all combinations of the feature selection and data augmentation method candidates for phase-wise regression (wear state-specific regression model) are compared via the achievable accuracy and used as the basis for method and feature selection. Therefore, features are extracted from both experimental datasets and augmented datasets. In this pa-per, 28 common features for RUL-prediction tasks were chosen (see Table 1).

The precondition for selecting features is determining a feature selection method. The methods used for filter-type feature selection are Principal Component Analysis (PCA), which performs and evaluates a dimensionality reduction of correlated data, and the Laplacian score, which evaluates the ability to maintain the local structure [38]. The features considered are limited to only features from the sensor signal of the *x*-axis or features of the sensor signals from all three *x*-, *y*-, and *z*-directions. Table 2 shows the top 5 selected features for the considered signal cases. Rank 1 represents the Lapcian score ranking of a feature for speed 1, and correspondingly, Rank 2 represents the ranking of the feature for speed 2.

The subsequent data augmentation can enable coverage of the unexplored input space, avoid overfitting, and improve model generalization [39]. Since there is no standard for data augmentation in this field, the data augmentation settings used in the work of Um et al. [39] are used as the basis for setting the hyperparameters of the different data augmentation methods in this paper and the following methods from [40, 41] are applied:

- **Jittering**: slightly decreasing or increasing the oscillation data randomly to simulate real data fluctuations.
- **Interpolation**: pattern mixing generates extended data points by estimating new data points around known points.

Table 1. Common features for RUL-prediction tasks

Feature	Mathematical Description
Root Mean Square	$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_{iu}^2}$
Variance	$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_{iu} - \mu)^2$
Standard Deviation	$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_{iu} - \mu)^2}$
Power	$P = \frac{1}{N}\sum_{i=1}^{N} x_{iu}^2$
Peak	$peak = \max(\lvert x_{iu}\rvert)$
Peak to Peak	$P2P = \max(\lvert x_{iu}\rvert) - \min(\lvert x_{iu}\rvert)$
Peak Factor	$PF = \frac{\max(x_{iu})}{RMS}$
Impulse Factor	$IF = \frac{Peak}{\lvert\mu\rvert}$
Crest Factor	$CF = \frac{Peak}{RMS}$
Form Factor	$FF = \frac{RMS}{\mu}$
Margin Factor	$MF = \frac{Peak}{\sqrt{\frac{1}{N}\sum_{i=1}^{N} x_{iu}{}^2}}$
Energy	$E = \sum_{i=1}^{N} x_{iu}^2$
Envelope energy	$EE = \sum_{i=1}^{N} \lvert hilbert(x_{iu})\rvert^2$
Zero Crossing	$ZC = \sum_{i=2}^{N} \frac{\lvert sign(x_{iu}) - sign(x_{(i-1)u})\rvert}{2}$
Binned Entropy	$BE = -\sum_{i=1}^{20} p(x_{iu}) \log_2(p(x_{iu}))$
Skew	$Skew = \frac{\sum_{i=1}^{N}(x_{iu} - \mu)^3}{(N-1)\sigma^3}$
Quantile	$Q = \frac{100 \times \sum_{i=1}^{N} x_{iu} \le percentile(x_{iu}, 0.9)}{len(x_{iu})}$
Kurtosis	$Kurtosis = \frac{1}{N}\sum_{i=1}^{N}(\frac{x_{iu} - \mu}{\sigma})^4$
Pulse Indicator	$PI = \frac{peak}{\mu}$
Frequency Mean	$f, A = welch(x_{iu})$ $FMean = \frac{\sum_{i=1}^{N} f_i \times A_i}{\sum_{i=1}^{N} A_i}$
Frequency Main	$FMain = \frac{index\ of\ \max\ value\ in\ x_u}{len(x_u)}$
Frequency Variance	$FV = \frac{\sum_{i=1}^{N} A_i (f_i - FMean)^2}{\sum_{i=1}^{N} A_i}$
Fourier Fast Transformation Coefficient 1	$FFTCoe1 = FFT(x_u)[0]$
Fourier Fast Transformation Coefficient 2	$FFTCoe2 = FFT(x_u)[1]$
Fourier Fast Transformation Coefficient 3	$FFTCoe3 = FFT(x_u)[2]$
Fourier Fast Transformation Coefficient 4	$FFTCoe4 = FFT(x_u)[3]$
Spectral Density	$SD = \bar{A}$
Total Wavelet Energy	$C = wavedec(x_u, wavelet, level) \quad TWE = \sum_{i=1}^{M}\sum_{i=1}^{N} C_{level,i}^2$

Table 2. Top 5 selected features for speed 1-2 for Laplacian Score *x* and Laplacian Score *xyz*

	Laplacian Score *x*	Rank 1	Rank 2	Rank Sum	Laplacian Score *xyz*	Rank 1	Rank 2	Rank Sum
Speed 1	RMS_*x*	0	0	0	RMS_*x*	0	0	0
	Quantile_*x*	2	1	3	Total Energy_*y*	1	3	4
	Skew_*x*	3	2	5	Spectral Density_*y*	2	4	6
	Kurtosis_*x*	4	3	7	FFT Coe4_*y*	3	5	8
	Pulse Indicator_*x*	5	4	9	FFT COE3_*y*	4	6	10
Speed 2	Kurtosis_*x*	1	3	4	Energy Coe4_*y*	1	9	10
	Frequency Main_*x*	2	5	7	FFT Coe2_*y*	8	8	16
	Quantile_*x*	7	1	8	FFT Coe3_*y*	12	7	19
	Frequency Mean_*x*	3	6	9	Quantile_*y*	7	16	23
	FFT Coe2_*x*	4	7	11	FFT Coe4_*y*	18	6	24

Using the abovementioned augmentation methods, ten augmented datasets are generated from dataset 1 with each augmentation method and used as training data for a preliminary regression analysis. The synthetically generated data's feature set is the input for the regression model. Meanwhile, the target variable is the RUL value in the form of the percentage value of the full lifetime. For which the regression performance, see Fig. 2, is calculated via R2-score difference to mean prediction. Here, negative percentage values mean a poor ability to capture any data pattern for regression. They indicate a worse performance than predicting the mean of target variables for all observations. The best results can be achieved for the feature set by Laplacian Score *xyz* based on speed 1 data. This can be derived for all three speeds. For the following feature-based wear state prediction, the features from speed 1 for *xyz*-direction are extracted for the 20 augmented datasets used to train the classifier, identifying the given wear state, e.g., phase 1 or phase 2. The original dataset 2 is used for testing the classification algorithm. Dataset 1 is used as validation data in the following investigation.

In the next step, the classifiers listed in Table 4 are examined to classify the different wear phases. The performance is evaluated and compared using confusion matrices and the corresponding accuracy results. A corresponding function is also used to run through all combinations of the hyperparameter range under consideration. The unnaturally high accuracy of > 99% result is mostly due to the relatively small variance of the augmented data from the test set. This is a natural limitation when working with a limited data set, as in this paper.

Table 4 shows an excerpt of the comparison of the classification accuracy results for different combinations of classification methods, hyperparameters, and data split. The default values are presents for the standard functions from the respective libraries. Based on the validation results, the model with the highest accuracy is selected as the final phase wise regression classification model, in this case, the random forest. This model identifies via the classification in the first stage the wear state of the component. So, if a component is still in its steady wear phase or in an exponential wear phase at the later part of its lifetime.

Feature Selection Method	Regression Method	Augmentation Selection Method				
		Jittering		Interpolation		
		Phase 1	Phase 2	Phase 1	Phase 2	
Laplacian Score x	SV-R	-20%	12%	0%	19%	Standard Attempt
	RF-R	12%	-10%	15%	-14%	
	GB-R	5%	-12%	5%	-11%	
	NN-R	-40%	-50%	-62%	-60%	
Laplacian Score xyz	SV-R	-16%	-26%	-21%	1%	
	RF-R	-19%	-27%	-20%	-11%	
	GB-R	-18%	-18%	-17%	-3%	
	NN-R	-23%	-29%	-38%	-3%	
PCA x	SV-R	-10%	-10%	-20%	-5%	
	RF-R	-50%	24%	-56%	22%	
	GB-R	-1%	23%	-1%	30%	
	NN-R	-19%	15%	-12%	18%	
Laplacian Score x Speed 1 feature set	SV-R	-40%	-4%	-14%	-5%	Additional Attempt
	RF-R	-23%	52%	-17%	37%	
	GB-R	-16%	25%	-11%	27%	
	NN-R	-43%	33%	-45%	25%	
Laplacian Score xyz Speed 1 feature set	**SV-R**	**20%**	**10%**	**39%**	**16%**	
	RF-R	**52%**	**91%**	**59%**	**93%**	
	GB-R	**44%**	**88%**	**50%**	**85%**	
	NN-R	**29%**	**79%**	**47%**	**81%**	

Fig. 2. Preliminary regression accuracy

Table 3. Classification accuracy for Wear State Detection

Classification Method	Validation Data		Test Data
	with default Hyperparameters	with optimized Hyperparameters	with the optimized Hyperparameters
Decision Tree	99 %	100 %	92 %
Support Vector Machine	100 %	99 %	94 %
K-nearest Neighbor	100 %	100 %	78 %
Random Forest	**100 %**	**100 %**	**95 %**
Neural Network	NA	100 %	93 %

6. FEATURE-BASED RUL-PREDICTION

The classification results determine the current phase of the ball screw operation. The data from the feature space is fed into the regression model corresponding to the individual phase. For each model, hyperparameter optimization and, thus, a comparison between the standard model and the model optimized with hyperparameters are now carried out. The model with the better performance is selected based on the prediction's R2 value compared to

the true value. The R2 value measures the proportion of the variance of the target value, in this case, the remaining service life, in a regression model. The regression accuracies for the four investigated methods with different combinations of hyperparameters for the phases are shown in Table 1. Furthermore, it can be observed that the accuracies for phase 2 are significantly higher than in phase 1. The characteristics of the two phases likely cause this. phase 1 is relatively stationary, with a stable and roughly constant degradation rate. Degradation accelerates in phase 2. Therefore, the various features considered in the regression models may not vary sufficiently for the model to make distinctive RUL predictions based on those features in phase 1.

Table 4. Regression method accuracies comparison for Speed 2

	Regression Method	Validation Data			
		Phase 1		Phase 2	
		Default Hyperparameters	Optimized Hyperparameters	Default Hyperparameters	Optimized Hyperparameters
Speed 2	Support Vector Machine	25%	20%	4%	8%
	Random Forest	**57%**	**54%**	**91%**	**91%**
	Gradient Boosting	**47%**	**53%**	**78%**	**93%**
	Neural Network	NA	28%	NA	54%

The two regression methods, Random Forest and Gradient Boosting with the corresponding hyper-parameters, are selected for their similar good predictive performance on the validation (dataset 1) in Table 4 A clear correlation can be confirmed for the random forest RUL model using visualizations of the predicted and true RUL values, as seen in Fig, 3 (blue circles delta to the dotted center line in red). In phase 1, the points are more spread out but still show a general clustering trend around the line of ideal predictions. This suggests that the model's predictions are reasonably accurate over the entire range of RUL values in phase 1. In phase 2, the points are more tightly grouped around the optimal diagonal, indicating a high accuracy of the model's predictions. Similar observations can be made for the gradient boost model. For speed 2, although the values from the random forest generally follow the ideal line for both phase 1 and phase 2. Compared to the phase 1 of the random forest for speed 1, the predictions for phase 1 of speed 2 deviate more strongly, possibly due to the higher speed being closer to the critical speed. Similar observations can again be made for the gradient-boosting regressor. The random forest for the RUL prediction was chosen because it inherently enables uncertainty quantification.

Table 5. Validation accuracies for selected regression

	Regression Method	Test Data	
		Phase 1	Phase 2
Speed 2	**Random Forest with default HP (P1&P2)**	**68%**	**80%**
	Gradient Boosting with optimized HP (P1&P2)	64%	83%

7. UNCERTAINTY INVESTIGATION FOR MODEL PREDICTION

Random forests are enabled for uncertainty determination due to their ensemble learning approach, which constructs multiple decision trees and aggregates the individual trees' results. The individual trees are based on a random data subset and consider a random selection of features at each split. The randomness leads to variability and uncertainty amongst all the trees. Therefore, the aggregate result of all the trees reflects that uncertainty. The uncertainty of the random forest regression models is calculated after the approach from Mentch and Hooker [42]. Figure 3 visualizes the predictions vs. the relative remaining useful life values between zero and one and includes the 64%, 95%, and 99% confidence intervals around the prediction. Redcolored predictions are outside the possible relative RUL range and are, therefore, not plausible. Ideal predictions are represented through a red dotted line representing the optimal solution.

The uncertainty quantification is shown in Fig. 3. For the 1st phase, Fig. 3 (top), the range of uncertainty is superior towards the end of the phase, beyond the true RUL value of 0.2. This narrower and superior uncertainty range continues to phase 2, shown in Fig. 3 (bottom). This can be attributed to the more distinctive features of the 2nd phase, the exponentially increasing wear phase. This could suggest that this portion of predictions beyond the true RUL value of 0.2 should be classified into phase 2 instead. The rapid in-crease of RMS observed in the data at around 20% RUL is also a possible explanation for the sudden decrease in the uncertainty range starting from 20% RUL for both models. Considering the scale of uncertainties and behaviour of feature trends, the classification of this portion into phase 2 is further supported. The mean values, represented by the blue circles in Figure 3, deviate from the ideal line. However, the confidence intervals show that both phases' most frequent predicted value generally occurs near or on the ideal line.

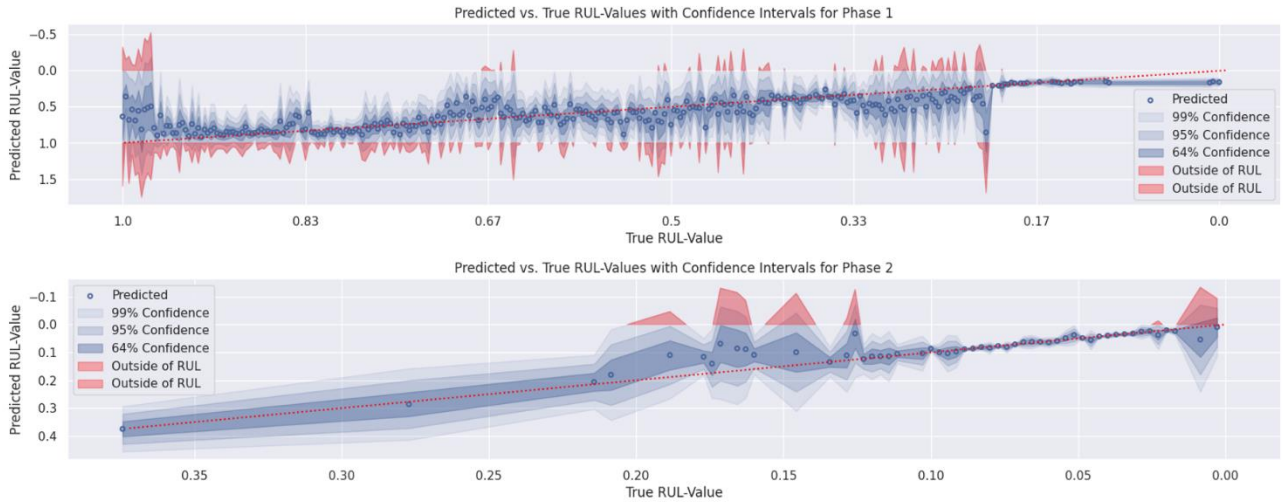

Fig. 3. Predicted vs true values for regression model with three confidence intervals for the two phases

In phase 1, reasonable confidence in the predictions can be observed. Here, the first predictions at the beginning of the lifetime must be critically evaluated. The resulting predictions, which lead to the assumption of a nearly complete degraded component, most likely

result from run-in effects. To compensate for such effects, a larger study regarding the investigated components and datasets is needed to enable differentiation between run-in characteristics and classical wear characteristics.

The predictions in phase 2 are usually very precise and highly confident. Predictions with a remaining useful lifetime of less than 15% are especially accurate and made with strong confidence.

8. SUMMARY AND CONCLUSION

The proposed prediction pipeline for adaptive cloud-based model usage is designed to evaluate sensor signals and predict machine components' remaining useful life (RUL). The pipeline comprises three main components: local preprocessing via edge computing, data transmission through a message broker, and adaptive model training in the cloud. Local preprocessing reduces data volume and enhances privacy, while adaptive model training improves prediction quality by continuously incorporating new production data. The experimental setup involves a ball screw test bench with accelerated degradation, where data is collected using a three-dimensional accelerometer. Various preprocessing steps, such as outlier filtering, are applied to the collected data to ensure quality. Feature extraction and data augmentation techniques are employed to enhance the dataset. Principal Component Analysis (PCA) and the Laplacian Score are used for feature selection. The extracted features are then used for phase-wise wear state prediction, with different classifiers tested to identify the optimal model. Multiple machine learning models were investigated for wear state prediction, including Decision Trees, Support Vector Machines (SVM), K-nearest Neighbors (KNN), Random Forests, and Neural Networks. These models were evaluated based on their classification accuracy using default and optimized hyperparameters. Ran-dom Forests and SVMs showed the highest accuracy, with Random Forests being selected as the final model due to their robustness and high accuracy in both default and optimized settings. Regression models such as Random Forest, Gradient Boosting, Support Vector Machines (SVM), and Neural Networks were evaluated for RUL prediction. These models were optimized using hyperparameters to improve accuracy, particularly for phase 2, where degradation accelerates. The Random Forest and Gradient Boosting models demonstrated the best performance, with the Random Forest selected for its superior accuracy and ability to quantify uncertainty. The regression models' accuracies were compared using R2 values, with Random Forest and Gradient Boosting showing strong predictive performance across different phases. Uncertainty quantification is conducted using the Random Forest model, providing prediction confidence intervals. The results show high accuracy and confidence in phase 2 predictions, while phase 1 predictions require further investigation due to run-in effects.

The proposed adaptive cloud-based model usage pipeline effectively addresses the challenges of predicting the remaining useful life of machine components. The pipeline ensures high-quality predictions while maintaining data privacy by incorporating local preprocessing, efficient data transmission, and adaptive model training. The experimental setup and data augmentation methods enhance the dataset, enabling more accurate and reliable wear state and RUL predictions. Investigating different machine learning models for

wear state and RUL prediction highlighted the robustness and accuracy of Random Forest and Gradient Boosting models. Random Forest, in particular, was selected for its ability to handle feature variability and provide uncertainty quantification, making it highly suitable for the RUL prediction task. The models' performance was validated through rigorous testing and comparison, demonstrating their effectiveness in capturing the degradation patterns of machine components. Overall, this research provides a robust frame-work for future implementations of adaptive cloud-based prediction models, offering significant benefits in predictive maintenance and machine component management. Future work will expand the dataset and refine the models to improve prediction accuracy and generalization across different operating conditions and component types and address the implementation of the different functionalities in an artificial Gaia-X ecosystem.

ACKNOWLEDGEMENTS

This research was supported by the Innovation Campus Future Mobility of the Ministry of Science, Research and the Arts of the State of Baden-Württemberg as part of the research project SdManuX - Gaia-X4ICM.

REFERENCES

[1] ZHENG T., ARDOLINO M., BACCHETTI A., Perona M., 2021, *The Applications of Industry 4.0 Technologies in Manufacturing Context: A Systematic Literature Review*, International Journal of Production Research, 59/6, 1922–1954.

[2] LU Y., 2017, *Industry 4.0: A Survey on Technologies, Applications and Open Research Issues*, Journal of industrial information integration, 6, 1–10.

[3] PIVOTO D.G., DE ALMEIDA L.F., DA ROSA RIGHI R., RODRIGUES J.J., LUGLI A.B., ALBERTI A.M., 2021, *Cyber-Physical Systems Architectures for Industrial Internet of Things Applications in Industry 4.0: A Literature Review*, Journal of manufacturing systems, 58, 176–192.

[4] MALIK P.K., SHARMA R., SINGH R., GEHLOT A., SATAPATHY S.C., ALNUMAY W.S., Nayak J. 2021, *Industrial Internet of Things and Its Applications in Industry 4.0: State of The Art*, Computer Communications, 166, 125–139.

[5] VEILE J.W., KIEL D., MÜLLER J.M., VOIGT K.I., 2020, *Lessons Learned from Industry 4.0 Implementation in the German Manufacturing Industry*, Journal of Manufacturing Technology Management, 31/5, 977–997.

[6] WANG L., WANG G., 2016, *Big Data in Cyber-Physical Systems, Digital Manufacturing and Industry 4.0*, International Journal of Engineering and Manufacturing (IJEM), 6/4, 1–8.

[7] BUENO A., GODINHO FILHO M., FRANK A.G., 2020, *Smart Production Planning and Control in the Industry 4.0 Context: A Systematic Literature Review*, Computers & Industrial Engineering, 149, 106774.

[8] WOLLSCHLAEGER M., SAUTER T., JASPERNEITE J., 2017, *The Future of Industrial Communication: Automation Networks in the Era of the Internet of Things and Industry 4.0*, IEEE industrial electronics magazine, 11/1, 17–27.

[9] JAVAID M., HALEEM A., SINGH R.P., SUMAN R., 2022, *Enabling Flexible Manufacturing System (FMS) Through the Applications of Industry 4.0 Technologies*, Internet of Things and Cyber-Physical Systems, 2, 49–62.

[10] TORN I.A.R., VANEKER T.H., 2019, *Mass Personalization with Industry 4.0 by SMEs: A Concept for Collaborative Networks*, Procedia manufacturing, 28, 135–141.

[11] XU L.D., XU E.L., LI L., 2018, *Industry 4.0: State of the Art And Future Trends*, International journal of production research, 56/8, 2941–2962.

[12] CHEN T., SAMPATH V., MAY M.C., et. al., 2023, *Machine Learning in Manufacturing Towards Industry 4.0: From 'for now' to 'Four-Know'*, Applied Sciences, 13/3, 1903.
[13] KHLIL A., SHI Z., UMAR A., MA B., 2023, *A New Industry 4.0 Approach for Development of Manufacturing Firms Based On DFSS*, Processes, 11/7, 2176.
[14] STENTOFT J., JENSEN K.W., PHILIPSEN K., HAUG A., 2019, *Drivers and Barriers for Industry 4.0 Readiness and Practice: A SME Perspective with Empirical Evidence.*
[15] MASOOD T., SONNTAG P., 2020, *Industry 4.0: Adoption Challenges and Benefits for SMEs*, Computers in Industry, 121, 103261.
[16] YU F., SCHWEISFURTH T., 2020, *Industry 4.0 Technology Implementation in SMES–A Survey in the Danish-German Border Region*, International Journal of Innovation Studies, 4/3, 76–84.
[17] SCHLAGENHAUF T., BURGHARDT N., 2021, *Intelligent Vision-Based Wear Forecasting on Surfaces of Machine Tool Elements*, SN Applied Sciences, 3, 1–13.
[18] MÜNZING T., 2017, *Auslegung von Kugelgewindetrieben bei oszillierenden Bewegungen und dynamischer Belastung*, Stuttgart: Institut für Konstruktionstechnik und Technisches Design.
[19] HILLENBRAND J., FLEISCHER J., 2021, *Unsupervised Detection of State Changes During Operation of Machine Elements*, Journal of Machine Engineering, 21/2, 36–46.
[20] FLEISCHER J., LEBERLE U., MAIER J., SPOHRER A., 2014, *Resource-Efficient Ball Screw by Adaptive Lubrication*, Procedia CIRP, 15, 50–55.
[21] BMWI B.M.B.F., 2019, *Das Projekt GAIA-X: Eine vernetzte Dateninfrastruktur als Wiege eines vitalen*, Europäischen Ökosystems.
[22] NIEBEL C., REIBERG A., KRAEMER P., 2022, *Gaia-X for SMEs*, 6. November 2022.
[23] KRAEMER P., NIEBEL C., REIBERG A., 2023, *Gaia-X and Business Models*, 25. February 2023.
[24] HOFFMANN F., WEBER M., WEIGOLD M., METTERNICH J., 2022, *Developing Gaia-X Business Models for Production*, Conference on Production Systems and Logistics: CPSL 2022, 583–594, Hannover, publish-Ing.
[25] NIEBEL C., SMOLEŃ, A., 2023, *AI And Gaia-X*, 30. November 2023.
[26] GAO R., WANG L., TETI R., DORNFELD D., KUMARA S., MORI M., HELU M., 2015, *Cloud-Enabled Prognosis for Manufacturing*, CIRP annals, 64/2, 749–772.
[27] VILLALONGA A., BERUVIDES G., CASTAÑO F., HABER R., 2018, *Industrial Cyber-Physical System for Condition-Based Monitoring in Manufacturing Processes*, 2018 IEEE Industrial Cyber-Physical Systems (ICPS), 637–642.
[28] ARÉVALO F., DIPRASETYA M.R., SCHWUNG A., 2018, *A Cloud-Based Architecture for Condition Monitoring Based on Machine Learning*, 2018 IEEE 16th International Conference on Industrial Informatics (INDIN), 163–168).
[29] ZHAO K., LI L., CHEN Z., SUN R., YUAN G., LI J., 2022, *A Survey: Optimization and Applications of Evidence Fusion Algorithm Based on Dempster–Shafer Theory*, Applied Soft Computing, 124, 109075.
[30] CAGGIANO A., 2018, *Cloud-Based Manufacturing Process Monitoring for Smart Diagnosis Services*, International Journal of Computer Integrated Manufacturing, 31/7, 612–623.
[31] STRÖBEL R., BOTT A., WORTMANN A., FLEISCHER J., 2023, *Monitoring of Tool and Component Wear for Self-Adaptive Digital Twins: A Multi-Stage Approach Through Anomaly Detection and Wear Cycle Analysis*, Machines, 11/11, 1032.
[32] BHATT U., ANTORÁN J., ZHANG Y., LIAO Q.V., SATTIGERI P., FOGLIATO R., XIANG A., 2021, *Uncertainty as a Form of Transparency: Measuring, Communicating, and Using Uncertainty*, Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, 401–413.
[33] HÜLLERMEIER E., WAEGEMAN W., 2021, *Aleatoric and Epistemic Uncertainty in Machine Learning: An Introduction to Concepts and Methods*, Machine Learning, 110, 457–506.
[34] MENTCH L., HOOKER G., 2016, *Quantifying Uncertainty in Random Forests Via Confidence Intervals and Hypothesis Tests*, The Journal of Machine Learning Research, 17/1, 841–881.
[35] MIRA J., MORENO I., BARDISBANIAN H., GORROÑOGOITIA J., 2023, *Machine Learning (ML) as a Service (Mlaas): Enhancing IoT with Intelligence, Adaptive Online Deep and Reinforcement Learning, Model Sharing, and Zero-Knowledge Model Verification*, Shaping the Future of IoT with Edge Intelligence, 63.

[36] MO Y., LI L., HUANG B., LI X., 2023, *Few-Shot RUL Estimation Based on Model-Agnostic Meta-Learning*. Journal of Intelligent Manufacturing, 34/5, 2359–2372.

[37] ILYUSHIN B.B., 2024, *On Applicability of IQR Method for Filtering of Experimental Data*, Journal of Engineering Thermophysics, 33/1, 1–8.

[38] MÜLLER A.C., GUIDO S., 2017, *Introduction to Machine Learning with Python: a Guide for Data Scientists* (First edition), O'Reilly Media Inc.

[39] UM T.T., PFISTER F.M.J., PICHLER D., ENDO S., LANG M., HIRCHE S., FIETZEK U., KULIĆ D., 2017, *Data Augmentation of Wearable Sensor Data for Parkinson's Disease Monitoring Using Convolutional Neural Networ*ks, 216–220, arXiv:1706.00527.

[40] IWANA B.K., UCHIDA S., 2021, *An Empirical Survey of Data Augmentation for Time Series Classification with Neural Networks*, PloS One, 16/7, e0254841.

[41] OH C., HAN S., JEONG J., 2020, *Time-Series Data Augmentation Based on Interpolation*, Procedia Computer Science, 175, 64–71.

[42] MENTCH L., HOOKER G., 2016, *Quantifying Uncertainty in Random Forests Via Confidence Intervals and Hypothesis Tests*, Journal of Machine Learning Research, 17/26, 1–41.