Unterberg, Martin; Becker, Marco; Niemietz, Philipp; Bergs, Thomas

Data-driven indirect punch wear monitoring in sheet-metal stamping processes

Journal of Intelligent Manufacturing

Journal of Intelligent Manufacturing (2024) 35:1721–1735
https://doi.org/10.1007/s10845-023-02129-w

Data-driven indirect punch wear monitoring in sheet-metal stamping processes

Martin Unterberg[1] · Marco Becker[1] · Philipp Niemietz[1] · Thomas Bergs[1,2]

Received: 20 January 2023 / Accepted: 7 April 2023 / Published online: 8 May 2023

Abstract
The wear state of the punch in sheet-metal stamping processes cannot be directly observed, necessitating the use of indirect methods to infer its condition. Past research approaches utilized a plethora of machine learning models to infer the punch wear state from suitable process signals, but have been limited by the lack of industrial-grade process setups and sample sizes as well as their insufficient interpretability. This work seeks to address these limitations by proposing the sheared surface of the scrap web as a proxy for the punch wear and modeling its quality from acoustic emission signals. The experimental work was carried out in an industrial-grade fine blanking process setting. Evaluation of the model performances suggests that the utilized regression models are capable of modeling the relationship between acoustic emission signal features and sheared surface quality of the scrap webs. Subsequent model inference suggests adhesive wear on the punch as a root cause for the sheared surface impairment of the scrap webs. This work represents the most extensive modeling effort on indirect punch wear monitoring in sheet-metal stamping both from a model prediction and model inference perspective known to the authors.

Keywords Sheet-metal stamping · Machine learning · Punch wear · Data mining

Introduction

Sheet-metal stamping processes are among the most frequently utilized manufacturing technologies due to their high resource efficiency (Klocke, 2014). Unexpectedly high tool wear can lead to unplanned machine down times, excess waste due to scrap production, costly finishing treatments required to compensate for product impairment, and high repair or replacement costs in case of excessively damaged critical tool components, such as the punch (Unterberg et al., 2021).

The punch wear state is not directly observable in industrial sheet-metal stamping processes. Therefore, the industrial approach towards controlling the punch wear currently consists of (i) *preventive* and (ii) *reactive maintenance*. The punch is either (i) replaced or redressed within fixed maintenance intervals or (ii) replaced or redressed when necessary, e.g. when impaired product quality is observed or an unplanned machine downtime occurred.

Based on the actual punch wear state, this approach has different implications. If the punch is replaced or maintained *too early* (non-critical wear state), both the planned machine downtime and the unnecessary maintenance of the punch lead to excess cost and wasted resources. If the punch is replaced or maintained *too late* (critical wear state), this does not only lead to unplanned machine downtimes, but also endangers other tool components through changes in the kinematic system induced by the punch wear.

Tool wear (and thus punch wear) does not progress linearly with production time (Behrens et al., 2016) and can be roughly separated into 3 phases: a *break-in phase*, a *steady-*

✉ Martin Unterberg
M.Unterberg@wzl-tf.rwth-aachen.de

Marco Becker
M.Becker@wzl-tf.rwth-aachen.de

Philipp Niemietz
P.Niemietz@wzl-tf.rwth-aachen.de

Thomas Bergs
T.Bergs@wzl-tf.rwth-aachen.de

[1] Laboratory for Machine Tools and Production Engineering (WZL), RWTH Aachen University, Campus-Boulevard 30, 52074 Aachen, Germany

[2] Fraunhofer Institute for Production Technology (IPT), Steinbachstraße 17, 52074 Aachen, Germany

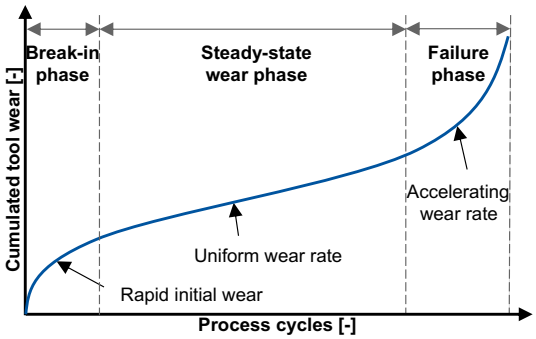

Fig. 1 Theoretical tool wear over the tool life

state wear phase and a *failure phase* as visualized in Fig. 1 (adapted from Groover (2010)).

Ideally, the punch wear state would be known at every stroke to infer optimal maintenance intervals. However, frequent direct measurements of the punch wear state (*direct tool wear monitoring*), e.g. by disassembling the tool and using a suitable wear evaluation metric, impair the economic viability of the process.

Indirect tool wear monitoring denotes leveraging process signals, such as force-displacement or acoustic emission (AE) signals, to infer the wear state of the tool from suitable signal characteristics. Although indirect tool wear monitoring is a frequently researched topic within the scientific community, no solution has prevailed so far in industrial sheet-metal stamping processes, highlighting the need for industrial applicable and validated approaches (Niemietz et al., 2022).

Unsupervised machine learning approaches do not require costly label generation in the form of punch wear evaluation to infer structural changes within a physical system. Prior works of the authors showed that representation learning approaches, such as UMAP and PCA embeddings (Unterberg et al., 2021) for AE signals and reconstruction error sequences of autoencoders (Niemietz et al., 2021) for both AE and force-displacement signals in fine blanking resemble the expected punch wear, lack, however, validation in the form of an approximated ground truth of the (either indirectly or directly) observed tool wear as a benchmarking tool for the developed approaches.

Niemietz et al. derived wear estimators from suitable representations of domain-specifically preprocessed force-displacement signals to model the punch wear in a fine blanking process. The validity of the estimators was examined with sparse wear stage information, motivating more extensive data collection. Moreover, the authors state that existing approaches to inferring punch wear information largely depend on artificially introduced wear states. This highlights the need for further research with industrial-grade process setups and emphasizes the importance of modeling and understanding transitions between wear stages (Niemietz et al., 2022). A similar representation learning approach to Niemietz et al. (2021) was followed by Asahi et al. The authors used the reconstruction error sequence of an autoencoder trained on collected "normal" (as in an unworn punch state) time series data from various sensor sources during a stamping process. The authors showed that the reconstruction error, calculated for each process cycle, can be interpreted as an anomaly score for the respective process cycle (Asahi et al., 2021). The authors used the reconstruction error as an input for classification models with 3 class labels, representing 3 consecutive wear stages. Kubik et al. used 5 stages of artificially introduced abrasive punch wear states as class labels in a stamping process. The authors used support vector machines to model the correlation between extracted features from force-displacement signals and the wear states. The signals were gathered for varying punch wear states, stroke frequencies, and sensor positions. Kubik et al. evaluated the model performances on the basis of their (i) classification accuracies and (ii) class separabilities. The authors emphasize the importance of leveraging domain knowledge during data acquisition, preprocessing, and transformation to optimize for model performance. Moreover, the authors highlight the need to evaluate different process signals, such as AE, as correlation variables for the punch wear (Kubik et al., 2022).

AE signals are based on transient elastic waves generated by the sudden release of energy, e.g., crack propagation, plastic deformation, or impacts (DIN EN 1330-9:2017-10, 2017). In contrast to force sensors, AE sensors are inexpensive and easy to mount and maintain, making them suitable for industrial usage. Compared to force-displacement signals, extracting relevant information from AE signals typically requires more effort due to their complex structure (Kollment et al., 2018). This highlights the need to obtain a high signal-to-noise ratio for AE signals with suitable preprocessing and transformation techniques, such as *frequency-domain analysis*. Frequency-domain analysis decomposes a time-domain signal into its frequency components. Past research showed that under sliding conditions, the behavior of these frequency components of AE signals changes with the wear state (Baccar & Söffker, 2015).

Shanbhag et al. used various frequency characteristics of AE signals from a semi-industrial sheet-metal stamping process setup to detect the onset and severity of galling, a phenomenon caused by adhesive wear. The authors state that the mean frequency feature, a scalar value representing the averaged frequency content of a signal weighted by the amplitudes, can be used to identify galling. Furthermore, the authors identified the frequency range 100–200 kHz as the frequency range that contained the most relevant information about the onset and severity of galling for their process setup (Shanbhag et al., 2020).

In summary, the following research deficits are identified and addressed within this work:

1. There is a lack of research for *industrial-grade sample sizes and process setups* on punch wear monitoring with AE signals.
2. For industrial-grade sheet-metal stamping processes, a *suitable proxy for isolating the punch wear* has yet to be identified in literature, necessitating disassembling and reassembling the tool to assess the punch wear state and thus affecting the process setup.
3. Past unsupervised approaches towards punch wear monitoring *lack a sufficient amount of labels* to evaluate the performances of said approaches.
4. The *number of class labels* utilized in supervised approaches to punch wear monitoring does not account for the transient nature of wear processes within sheet-metal stamping, neglecting transitions between wear stages and impeding inference from model decisions.
5. While recent machine learning approaches with highly flexible architectures towards punch wear monitoring are promising, they *lack interpretability*.

This work seeks to address these deficits as follows. The sheared surface roughness of the *scrap web*, which denotes the scrap part that results from the shearing operation, is proposed as a proxy for the punch wear in sheet-metal stamping processes and the evolution of this proxy over the lifespan of two punches is analyzed. Moreover, the correlation of this proxy with features of AE signals from an industrial-grade process setup and data set size is modeled with different approaches. Finally, model performances and model predictions are evaluated regarding the prediction quality and the feature importances for a comprehensive and interpretable feature space.

The developed methodology is laid out in the section "Methodology". The results of the analyses are presented and discussed in the section "Results and discussion". The section "Summary and outlook" summarizes this work and suggests open research questions.

Methodology

The following section describes the developed methodology. The section "Process setup" presents the process setup for the experiment. Following, the section "Sensoric setup" describes the sensory setup for the experiment. The section "Scrap web measurements" presents the measurement process of the sheared surface roughnesses of the scrab webs. The domain-specific data preprocessing approach (segmentation, filtering, and feature extraction) is documented in the section "Data preprocessing". To account for missing data, *multiple imputation* was used with different approaches, described in the section "Synthetic label generation". The developed analysis pipeline to (i) analyze the correlation between the sheared surface roughnesses of the scrab webs and the AE signal features with regards to the process phase and to (ii) infer feature importances from the resulting models is presented in the section "Data analysis pipeline".

Process setup

The experimental series was carried out on an industrial-grade fine blanking machine of type *Feintool XFT 2500 speed*. In total, 8 high-strength sheet-metal coils of grade 58CrV4 with a sheet-metal thickness of 5 mm were processed with a tool cavity of 2 and stroke frequency of 50 min^{-1}. The supplier reported a tensile strength of 568 MPa for the sheet-metal coils. A counter holder force of 300 kN and a press plate force of 500 kN were chosen according to the material specifications. No V-ring indenter was used. The punch steel was chosen as a powder metallurgical cold work steel of grade *K490* with a titanium aluminum carbon-nitride coating. The chosen lubricant was a chlorine-free lubricant of type *WISURA FMO 5020*. The experiment was carried out over the span of 3 consecutive days.

Sensoric setup

An acoustic emission sensor of type *Kistler 8152C0* with a frequency response of 50–400 kHz was mounted structurally close to the punch to avoid dampening effects, similar to the setup presented in Unterberg et al. (2021). A *Kistler 5125C* piezotron coupler was used to process the sensor signal. The AE signal was sampled with a sampling frequency of 1 MHz. Furthermore, a *Mikroepsilon optoNCDT* laser triangulation sensor was used to retrieve information about the process phase that corresponds to the AE signal in a given timeframe ex post. This distance signal was sampled with a sampling frequency of 10 kHz. Both distance and AE measurements were stopped whenever the process was stopped, e.g. due to coil changes, and were written to *.tdms* files with a shared timestamp. The sensory setup on tool site is shown in Fig. 2.

Scrap web measurements

In fine blanking, the shearing edges of the punch and the sheared surface of the scrap web represent a tribological system during the *shearing* (see Fig. 3a) and *stripping* phase (see Fig. 3b) with the lubricant as an intermediate and the punch wear as an initial size (Voigts, 2021). Followingly, the quality of the sheared surface of the scrap web is a result of the interaction of the punch with the surface during the shearing and stripping phase and is, from a physical standpoint, a suitable proxy for the punch wear. F_V denotes the V-ring

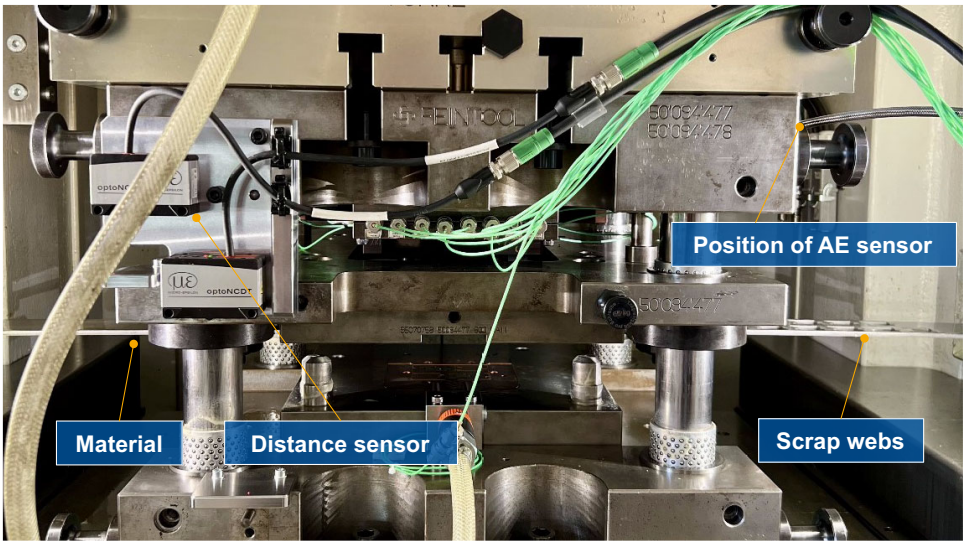

Fig. 2 Sensoric setup for the experimental stroke series

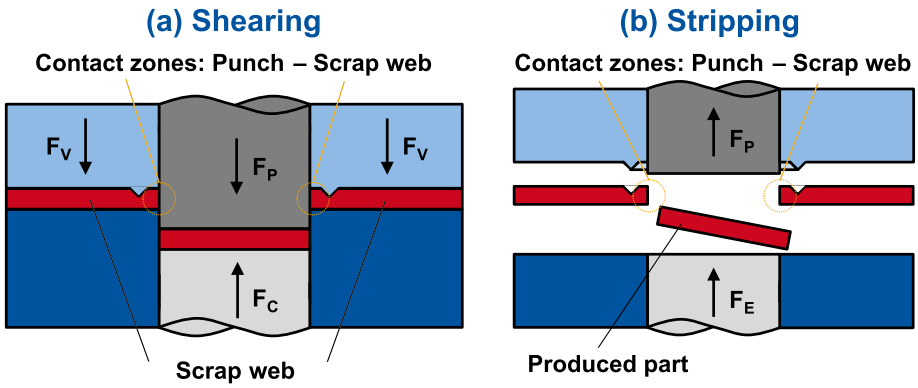

Fig. 3 Visualization of the **a** shearing and **b** stripping phase in fine blanking

indenter force, F_C the counter holder force, F_P and F_E the resulting punch and ejection forces in Fig. 3.

Scrap webs were sampled every 200th stroke, beginning with the first and ending with the last stroke of the experiment. In total, 203 scrap webs were sampled.

The most severely worn shearing edge surfaces of the punches were identified qualitatively ex post with a scanning electron microscope (SEM) at 40-fold magnification, see Fig. 4a. The identified surfaces showed severe delamination, abrasive and adhesive wear. Subsequently, the roughness of the corresponding surfaces of the scrap webs was tested at scale.

The roughness was tested with a stylus-based tactile profilometer of type *Jenoptik Waveline W920RC*. The selected surfaces are shown in Fig. 4b and the testing setup is shown in Fig. 4c. The length of the profile measurement section was chosen as 2.4 mm. To robustify the measurements towards extreme profile spikes and to account for inhomogeneous surfaces:

1. The arithmetic average Ra of the profile height deviations from the mean line was chosen to characterize the surfaces.
2. Per surface, 3 measurements were carried out with an offset to each other of 0.8 mm.
3. The mean of all 12 measurements (3 measurements per surface, 2 surfaces per punch, 2 punches) of a given scrap web was calculated and chosen as a representation of the scrap web roughness at the given stroke number.

Data preprocessing

To increase the signal-to-noise ratio and reduce computational load, the quasi-continuous AE signals were segmented in two subsequent segmentation steps and followingly, a feature extraction was performed, as visualized in Fig. 5.

In a first step (*rough segmentation*), strokes (i.e. one closing and the subsequent opening cycle of the tool) were detected using threshold values based on the known sheet-

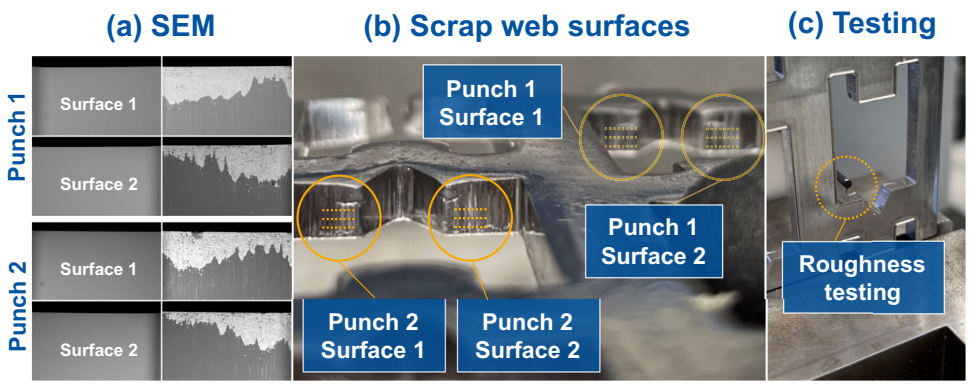

Fig. 4 **a** SEM of the punches before/after, **b** Selected sheared surfaces of the scrap webs, and **c** Roughness testing setup

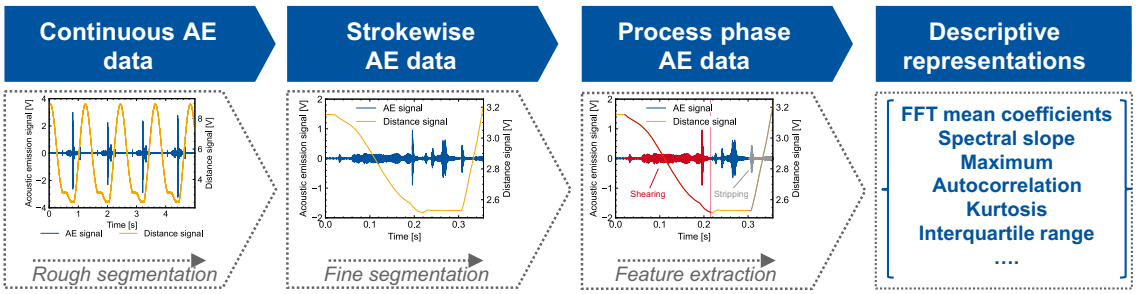

Fig. 5 Segmentation and feature extraction pipeline

metal thickness of 5 mm in the distance signals. The top dead center (TDC) was then identified as a local minimum in the distance signal, representing the end of the shearing phase. From the TDC, both the shearing and the stripping phases were identified and extracted by manually segmenting the first stroke of the experiment (determining the distance of the shearing and stripping phases relative to the TDC) and then using this distance with an automated Python script that (i) identifies the TDC of a stroke and (ii) extracts the shearing and stripping phase of a stroke (*fine segmentation*), see Fig. 6.

Since the AE and distance signals were sampled with different sampling frequencies, misalignment occurred within each sequence of process phases. Misalignment in this context means that for each stroke i, the shearing and stripping phases of all other strokes k with $k \neq i$ are shifted in time relative to stroke i. This shift is equal for both process phases and varies stroke-wise. An implementation of *Mueen's Algorithm for Similarity Search* (MASS) (Yeh et al., 2016) from the Python library *stumpy* was used to quantify the shifts within the sequence of stripping phases to a reference stroke, since the initial burst in the stripping segments is easily detectable, see Fig. 6. This information was then used to extract the aligned shearing and stripping segments from the rough segmentation.

To efficiently process the raw AE and distance data, the Python libraries *npTDMS*, *pandas* and *NumPy* were used.

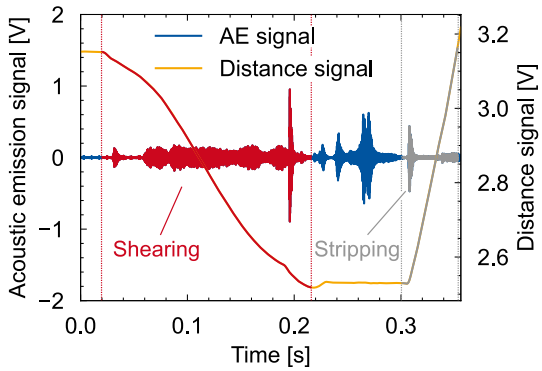

Fig. 6 Segmented stroke of the experimental series

To reduce noise in the AE signals, the resulting segments were band-pass filtered with a passband frequency range of 50–400 kHz after fine segmentation, which matches the frequency response of the AE sensor. The digital filtering was carried out with a *Butterworth* filter of order 6 using the Python library *SciPy*. For the feature extraction, the Python library *TSFEL* was used with all 390 available statistical, temporal, and spectral features (Barandas et al., 2020). To reduce the data load for subsequent data modeling, instead of a sliding window, the whole signal was used for the feature extraction. A further reduction of the feature space through a subsequent feature selection step was omitted to not introduce any further bias into the pipeline.

Synthetic label generation

The following assumptions were made prior to the experiment:

1. The roughness evolution of the sheared surfaces of the scrap webs has an *underlying physical causation* and is correlated with punch wear phenomena, thus, it is not completely stochastic. Subsequently, the missing data can be imputed with suitable models at the cost of introducing further uncertainty.
2. The mean of the *actual roughness measurements* for a scrap web is, even in the presence of measurement errors, *the most reliable proxy for this physical causation* at a certain stroke available in this work. Therefore, the actual roughness measurements should be present in all models.

To reduce the total amount of bias introduced into the subsequently trained models with the selection of an imputation model and to account for stochastic effects in the underlying physical process, various amounts of synthetic data were generated and different imputation approaches were carried out. This approach is visualized in Fig. 7. The regression models were trained with a combination of the 203 averaged sample roughnesses and for each model run randomly drawn $s \in \{0; 1000; 2000; 4000; 8000; 16{,}000; 32{,}000; 40{,}247\}$ synthetic labels. The synthetic labels were generated with interpolation methods with ascending model complexity:

1. *Data repetition* (assign roughness measured for scrap web 1 to grids 1–200, roughness for grid 200 to grids 201–400 etc.)
2. *Linear interpolation*
3. *Cubic spline interpolation* (smoothing factor chosen as 0, thus interpolating through each data point)
4. *Gaussian process regression*

The kernel for the Gaussian process regression (GPR) was chosen as a sum-kernel consisting of a *white noise kernel* and a *Matérn kernel* to introduce measurement noise and model the underlying physical process. A Gaussian process was fit to the roughness measurements with the *GaussianProcessRegressor* class from the *sklearn* library. The synthetic data was then sampled from the Gaussian process for each model run.

Data analysis pipeline

The data analysis pipeline for one model run is visualized in Fig. 8.

The task of modeling the roughness measurements and the synthetic roughness data from the AE features was treated as a regression problem. Prior to selecting the modeling approaches, a list of 3 criteria for the selection was compiled:

1. To infer the nature of the relationship (linear vs. nonlinear) between AE features and roughness data, both a *linear* and a *nonlinear modeling approach* should be chosen and compared.
2. Since no prior feature selection step was performed, both modeling approaches should utilize *regularization* to avoid overfitting.
3. To allow a high number of model runs for large data sets, both modeling approaches should be *highly scalable*.

Based to these criteria, two modeling approaches were chosen, namely *eXtreme Gradient Boosting* regression (XGBoost) and *Least absolute shrinkage and selection operator* regression (Lasso).

XGBoost is an end-to-end gradient tree boosting method with regularization. It uses both L_1 and L_2 regularization, controlling final model complexity, and avoiding overfitting even in the presence of a large number of input features (Chen & Guestrin, 2016). As a gradient tree boosting method, XGBoost is capable of learning nonlinear relationships between predictor variables and labels, and has shown on-par performance or even outperformed recently proposed deep learning models for classification and regression in tabular data sets (Shwartz-Ziv & Armon, 2022). XGBoost has the ability to utilize multiple CPU cores in parallel, speeding

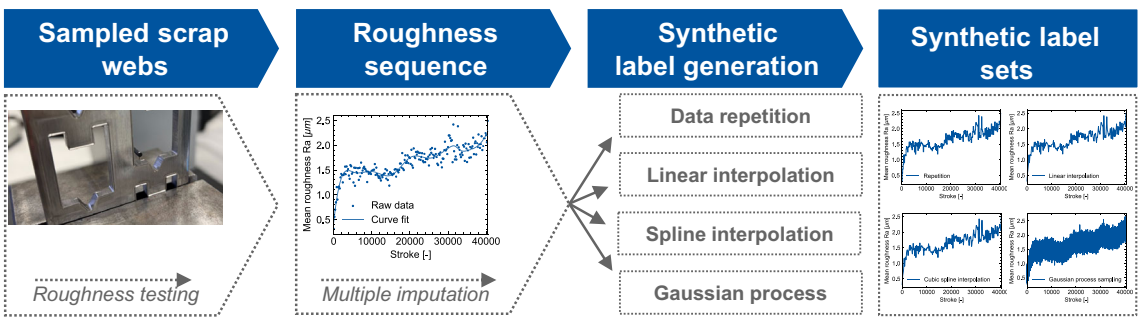

Fig. 7 Synthetic label generation pipeline

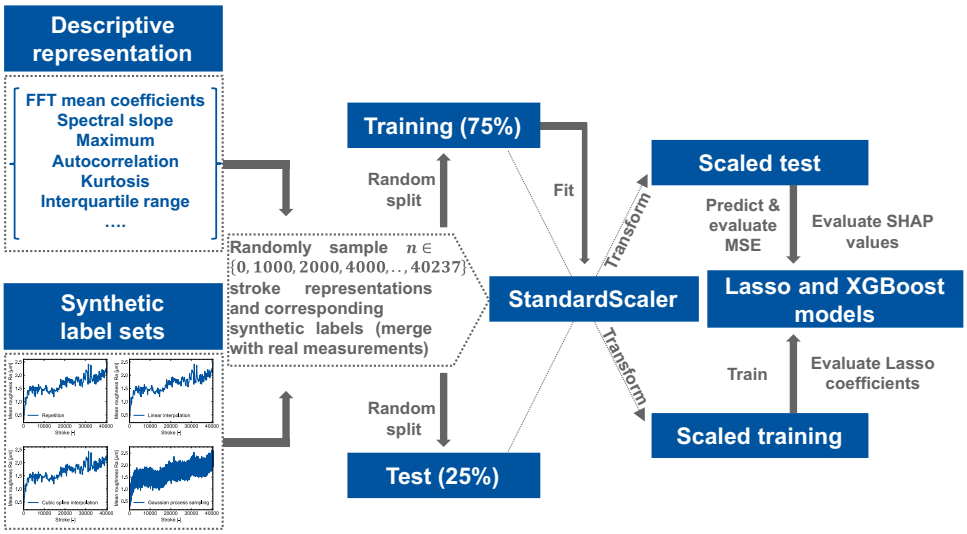

Fig. 8 Data analysis pipeline

up training and prediction substantially (Chen & Guestrin, 2016). Moreover, tree-based methods offer the advantage of efficient interpretation of model predictions in terms of feature importance, i.e. through the *TreeSHAP* algorithm. Compared to the model-agnostic *KernelSHAP*, TreeSHAP reduces computational complexity from exponential to polynomial time (Lundberg et al., 2020). Overall, XGBoost is highly scalable for both prediction and inference purposes and thus a suitable choice for this work.

Lasso is an even more restrictive modeling approach than linear regression, since in addition to modeling linear relationships between predictor variables and labels, Lasso produces coefficients that are exactly 0 through L_1 regularization. This allows inference of which features are ruled out as unimportant for the regression task and which are seen as important after the training process and effectively performs feature selection (Tibshirani, 1996). As a linear regression with L_1 regularization, Lasso is scalable to large data sets and a high number of model runs.

For every

- *data set size* (e.g. 1203 labels, consisting of 1000 synthetic labels and 203 averaged actual measurements),
- *process phase* (e.g. shearing) and
- *imputation method* (e.g. cubic spline),

100 random train/test splits (75%/25%) were drawn. Additionally, the synthetic labels from the GPR were sampled repeatedly for every model run from the Gaussian process.

Subsequently, the *StandardScaler* module from *sklearn* was used to standardize the model input for each model run. A StandardScaler object was fit to and then used to transform (standardize) the training set and subsequently transform (without another fitting procedure) the test set to prevent data leakage from the test to the training set.

For the XGBoost regressions, the Python library *xgboost*, developed and maintained by the authors of Chen and Guestrin (2016), was used. To reduce computation time, the default library parameters were used as documented in XGBoost (2023) without further hyperparameter optimization. For Lasso regressions, 5-fold cross-validation was carried out with 25 different regularization parameters λ between 0.02 and 0.5 for each model run with the *LassoCV* module from sklearn. This range has been determined empirically.

To evaluate model performances, the *mean squared error* (MSE) for every drawn test set has been evaluated.

For Lasso regressions, the feature importance was directly derived by identifying nonzero coefficients and inferring the important features from the trained model for each model run. For the resulting XGBoost models, the *SHapeley Additive exPlanations* (SHAP) method was used. SHAP is a method proposed by Lundberg and Lee towards model interpretability that is rooted in coalitional game theory. Features can be seen as players (in a game-theoretical sense) that coalesce with each other in a game (prediction task). Shapely values are a fair distribution of the game pay-out, thus Lundberg and Lee propose utilizing Shapely values as a measure of feature importance in machine learning tasks. Features with high Shapely values have a higher contribution to model predictions (thus, they receive a higher pay-out in a game-theoretical sense) than features with low Shapely values (Lundberg & Lee, 2017). The Python library *shap*, developed and maintained by the authors of Lundberg and Lee (2017), was used. The library contains an implementation of TreeSHAP for tree-based models.

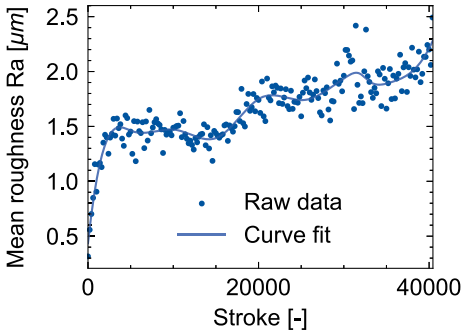

Fig. 9 Samplewise mean scrap web surface roughness

The following approach was taken for each model run $i \in \{1; 2; 3; \ldots; 100\}$ for a fixed process phase, data set size, imputation method, and model configuration with $M(MSE_{model})$ denoting the median MSE for all model runs within the configuration:

1. If $MSE_{XGB_i} < M(MSE_{XGB})$: Extract feature names with the highest 25 SHAP values of the XGBoost model, evaluated on the test set. *Every extracted feature name scores one point.*
2. If $MSE_{Lasso_i} < M(MSE_{Lasso})$: Extract all feature names with nonzero coefficients from Lasso model. *Every extracted feature name scores one point.*

The scores were calculated across all configurations and used to calculate the relative frequency of each feature that occurs in the most important features. The following section presents the results of the described approach.

Results and discussion

The section "Scrap web roughness development" presents the results of the scrap web surface roughness measurements and visualizes the generated synthetic label sets. The model performances are presented and discussed in the section "Model performances", followed by an evaluation of the feature importances in the section "Global Feature Importances".

Scrap web roughness development

Figure 9 shows the mean roughness values as outlined in the section "Scrap web measurements" for all sampled scrap webs.

Until the approximately 3400th stroke, the mean roughness of the examined scrap web surfaces increases rapidly, possibly resembling the break-in phase of the punch as visualized in Fig. 1. Until approximately stroke 15,400, the trend of the mean surface roughness appears to be slightly decreasing but rapidly increases from there on.

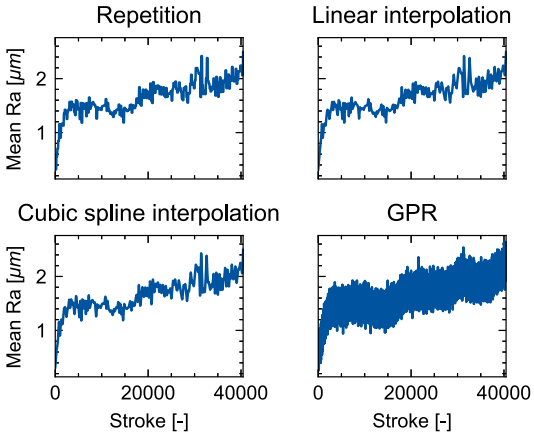

Fig. 10 Resulting synthetic label sets

Interestingly, the AE data shows a high-energy event within the shearing phase of stroke 15,401. The energy of this AE signal evaluates to 7,997.83 V^2, while the second highest energy during shearing evaluates to 457.55 V^2 at stroke 39,845 - with a standard deviation within all 40,450 strokes of 53.12 V^2, this energy difference represents a deviation of almost 142 standard deviations from the mean. It is likely, that this high-energy event is connected to the rapid increase of measured surface roughness of the scrap webs (e.g. small fractures on the shearing edges of the punches). However, this connection is not ascertainable with certainty ex post, and further analysis is out of the scope of this work.

The rapid increase in mean surface roughness continues to approximately stroke 21,000. After stroke 21,000 the trend shows an increase towards the end of the stroke series, with the last mean roughness representing the maximum of the data set.

Figure 10 shows the resulting sets of synthetic labels as outlined in the section "Synthetic label generation".

For the GPR, this represents a randomly drawn sample from the Gaussian process, clearly introducing a high amount of noise in comparison to the remaining interpolation approaches.

Model performances

Figure 11 visualizes the distribution of MSEs for the shearing segments and the respective examined configurations. Interestingly, the dispersion is exceptionally high for Lasso models trained and evaluated on small data sets for the shearing segments, with multiple training and test set configurations yielding an MSE of more than 10. Therefore, the distribution of MSEs was additionally visualized with a cut-off MSE of 0.05 in Fig. 12. The distribution of MSEs for the stripping segments is visualized in Fig. 13.

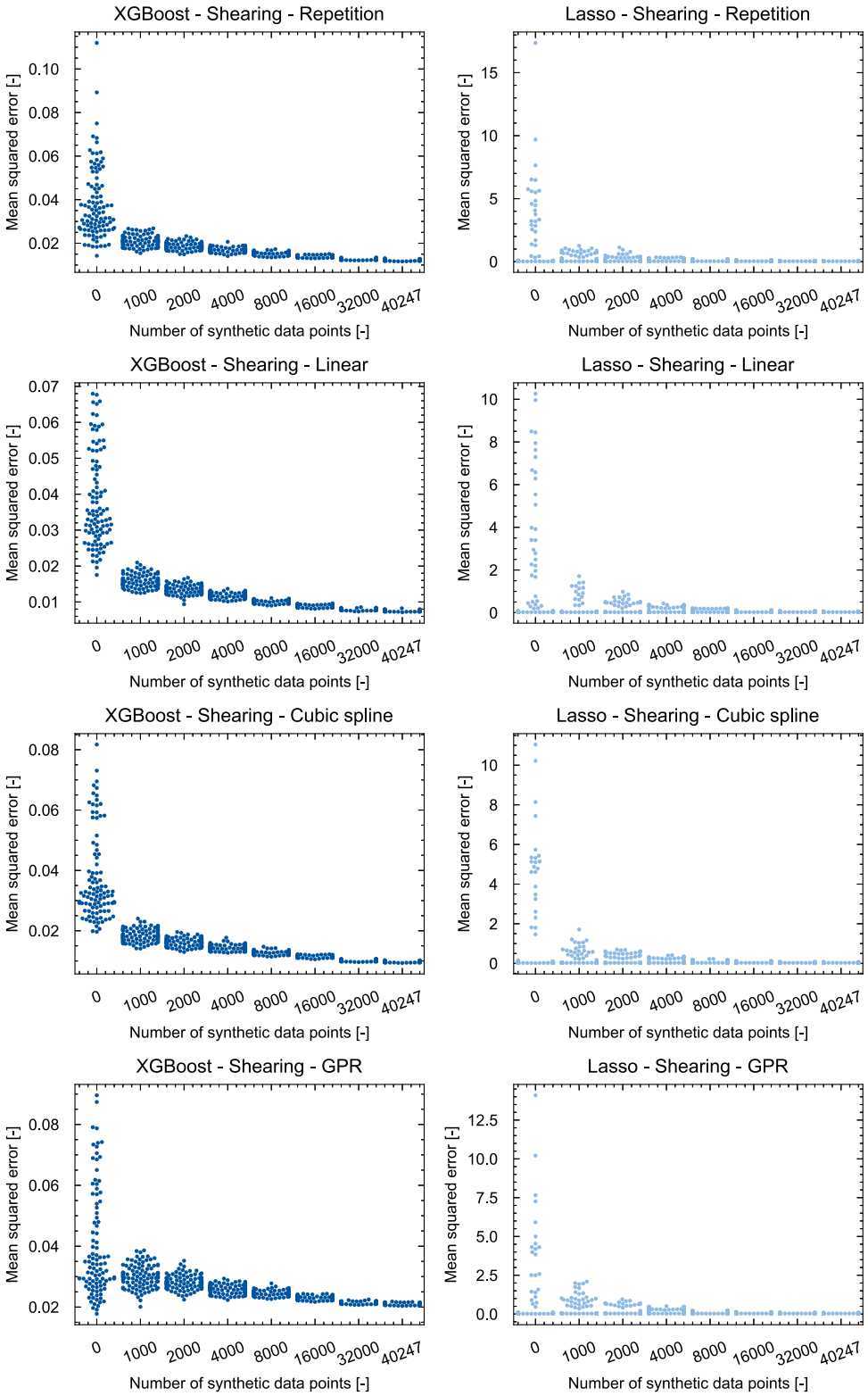

Fig. 11 Comparison MSEs shearing

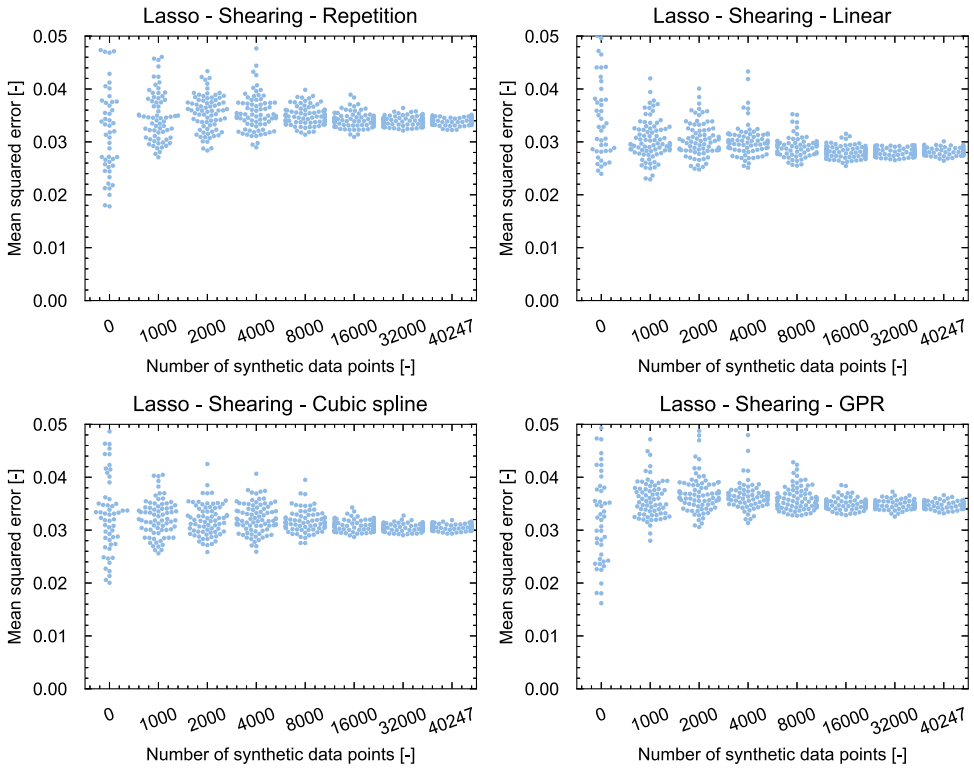

Fig. 12 Limited view Lasso MSEs shearing

The dispersion for Lasso models trained and evaluated on smaller stripping data sets is clearly lower than for the shearing segments, with an on-par performance of Lasso with XGBoost for training and test sets sampled from the 203 averaged roughness measurements (number of synthetic data points 0). In contrast, for the stripping segments, the linear models appear to be robust toward the selection of the training and data set even for small data sets.

Comparing both the absolute MSE values and the trend with increasing data set sizes, it becomes apparent that XGBoost models outperform Lasso models across all imputation methods. While the errors evaluated on the test sets of XGBoost models decrease with increasing data set (and thus both training set and test set) sizes, the same does not hold for Lasso models. This is further emphasized by comparing the statistical quantities mean, median and standard deviation (SD) for the MSEs across all model runs within a fixed model and imputation method configuration as presented in Table 1. These statistical quantities confirm the qualitative findings derived from Figs. 11 and 13: While both modeling approaches appear to be able to predict the test set labels well from the given features, XGBoost outperforms Lasso both in terms of mean and median of the MSE. Furthermore, comparing the mean and median of Lasso models for the shearing segments confirms the identified dispersion in Fig. 11 since, depending on the imputation method, the mean MSE is 6–8 times higher than the median. The same does neither hold true for the XGBoost models trained and evaluated on the shearing segments, nor for the XGBoost and Lasso models trained and evaluated on the stripping segments.

As documented in the section "Scrap web roughness development", stroke 15,401 showed exceptionally high signal energy during the shearing phase. Since this stroke was included in every data set either as a training or test sample (see section "Synthetic label generation") and its features showed, upon further investigation, substantially different values from all other strokes, its impact on the model performances was additionally examined by excluding the shearing and stripping phase of this stroke from the data sets and evaluating the resulting model performances.

For shearing, the resulting Lasso models yielded a mean MSE of 0.032 with a standard deviation (SD) of 0.006, which represents a substantially better performance than for the data sets that included the shearing segment of stroke 15,401 (mean MSE 0.229, SD 0.952). In contrast, the performance of the XGBoost models for the shearing segments did not change substantially (mean MSE 0.018, SD 0.009 without stroke 15,401 vs. mean MSE 0.019, SD 0.010 with stroke 15,401). Excluding the stripping segment of stroke 15,401 did not influence the performance of Lasso and XGBoost models, indicating that only the shearing segment of stroke 15,401 represents an extreme value. Further investigation revealed that the high MSEs of Lasso models (see Fig. 11) were caused by the presence of stroke 15,401 in the test sets.

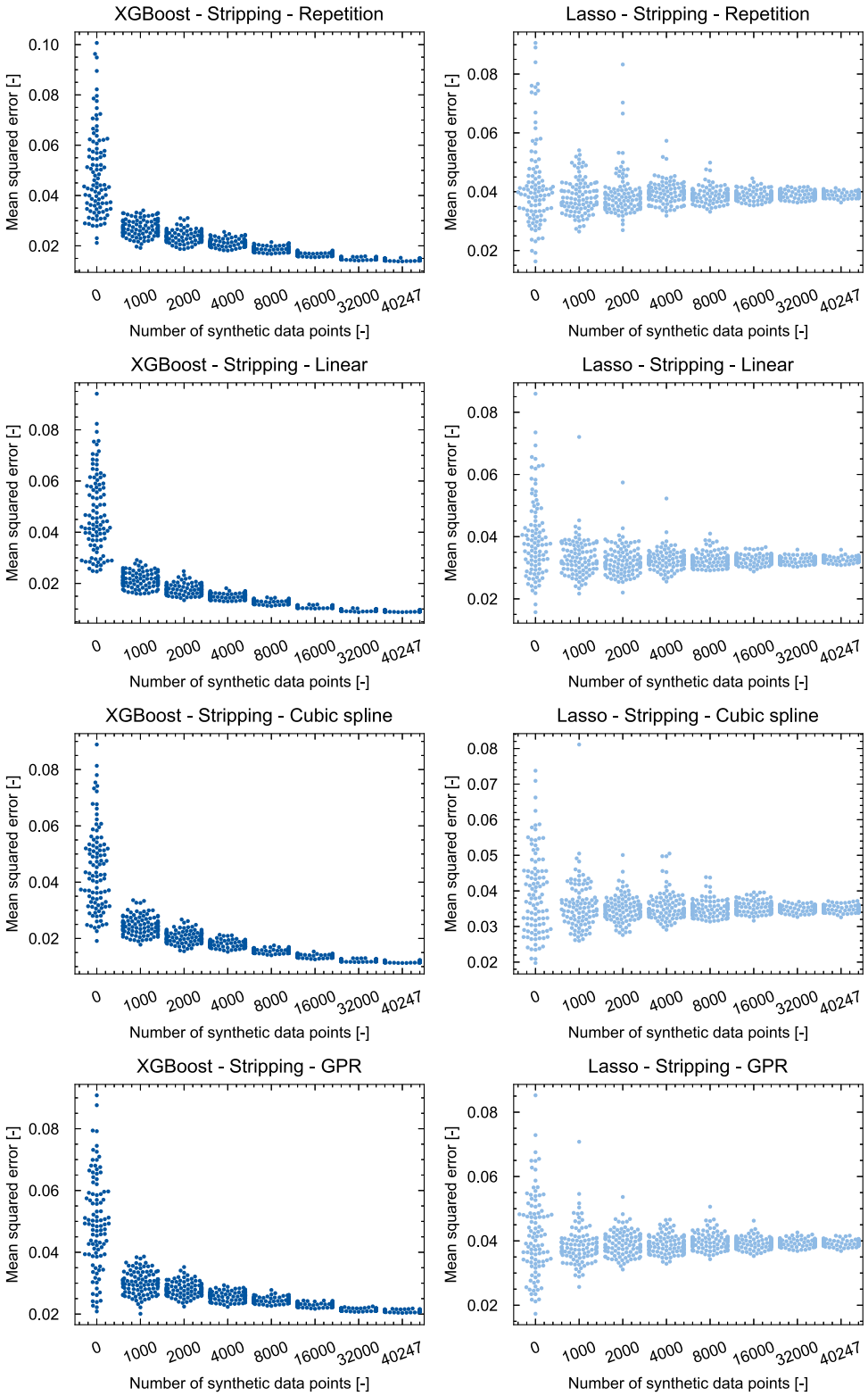

Fig. 13 Comparison MSEs stripping

Table 1 Mean, median and standard deviation (SD) of the MSEs across all model runs and data set sizes for the respective configurations

	Shearing					
	XGBoost			Lasso		
Interpolation	Mean	Median	SD	Mean	Median	SD
Repetition	0.019	0.016	0.010	0.230	0.035	1.005
Linear	0.014	0.011	0.011	0.240	0.030	0.992
Cubic spline	0.016	0.013	0.010	0.223	0.032	0.921
GPR	0.027	0.025	0.009	0.222	0.036	0.883
	Stripping					
	XGBoost			Lasso		
Interpolation	Mean	Median	SD	Mean	Median	SD
Repetition	0.023	0.020	0.012	0.039	0.039	0.007
Linear	0.018	0.014	0.013	0.034	0.033	0.006
Cubic spline	0.020	0.016	0.011	0.036	0.035	0.005
GPR	0.028	0.025	0.010	0.040	0.039	0.005

Other than the tree-based XGBoost, Lasso models failed to predict within the usual label value range when encountering the substantially different feature values of stroke 15,401 due to their linear structure, which explains the unusually high MSEs in test sets containing this stroke.

From a physical standpoint, while the AE signal measured during the stripping phase only contains information about the friction between the punch and the scrap web, the shearing segments contain, besides the friction between the punch and the scrap web, information about the plastic deformation of the material. While XGBoost models outperformed Lasso models, both modeling approaches were capable of modeling the relationship between the AE signal features and the scrap web surface roughnesses.

Global feature importances

Figure 14 presents the 25 features that show the highest relative frequencies (as described in the section "Data analysis pipeline", further denoted as *feature importances*) for the shearing segments across all data set sizes, interpolation methods, and models that resulted in a better-than-median performance (as in lower-than-median MSE) of a respective configuration. The analogous feature importances for the stripping segments are presented in Fig. 15. The feature importances for the shearing segments excluding stroke 15,401 are shown in Fig. 16. Since the feature importances for the stripping segments did not change after removing stroke 15,401, a further visualization is omitted.

The features labeled *FFT... kHz* represent the *FFT mean coefficient* features from TSFEL.

The identified feature importances suggest that across all model runs, similar features contribute to model decisions with a better-than-median performance, indicating a strong correlation of the contained information with the measured and synthetic roughness values of the examined scrap web surfaces.

A discussion of all the features represented in the presented feature importances is not in the scope of the scientific contribution of this work, and will therefore be omitted. However, evidently, the information contained in certain FFT mean coefficient features contributed frequently to the predictions of the better-than-median performing models, which

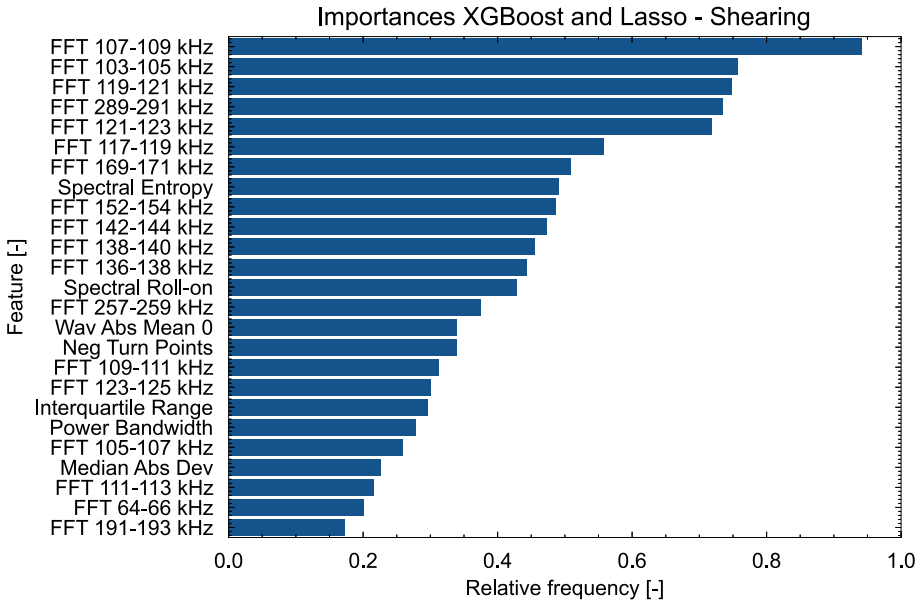

Fig. 14 Global feature importances shearing

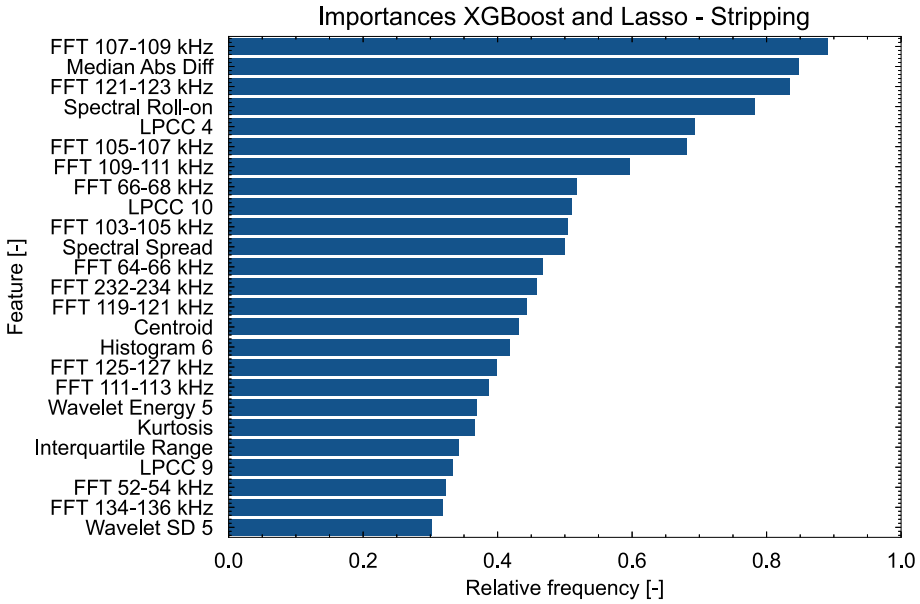

Fig. 15 Global feature importances stripping

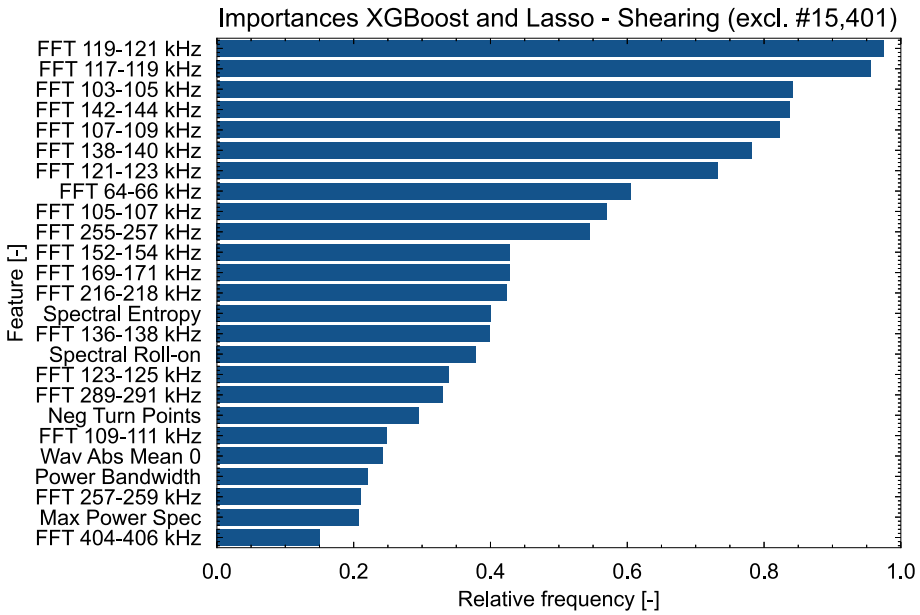

Fig. 16 Global feature importances shearing (stroke 15,401 removed)

interestingly holds true for both process phases. Therefore, the source code from the TSFEL library for the computation of these features is described here:

(i) Compute a spectrogram of the time series utilizing SciPy's *spectogram* function, yielding a matrix with rows corresponding to frequency ranges and columns corresponding to time ranges and

(ii) calculate the arithmetic mean with NumPy's *mean* function along the columns, i.e. the time ranges, for a given frequency range (matrix row). In total, the frequencies were divided into 256 ranges within this work, yielding 256 FFT mean coefficient features.

While both for shearing and stripping segments frequency ranges such as 64–66 kHz or 257–259 kHz appear, the findings suggest that for both shearing and stripping, the frequency range 103–144 kHz carries the most relevant information to enable model predictions (with a higher-than-median performance) of the scrap web roughnesses. For the data sets that were sampled from all strokes, the FFT mean coefficient for the frequency range 107–109 kHz represents the feature with the highest feature importance across all model runs. Removing stroke 15,401 from the data sets slightly changes the feature importances for the shearing segments. The FFT mean coefficient for the frequency range 107–109 kHz places 5th, whereas the FFT mean coefficients

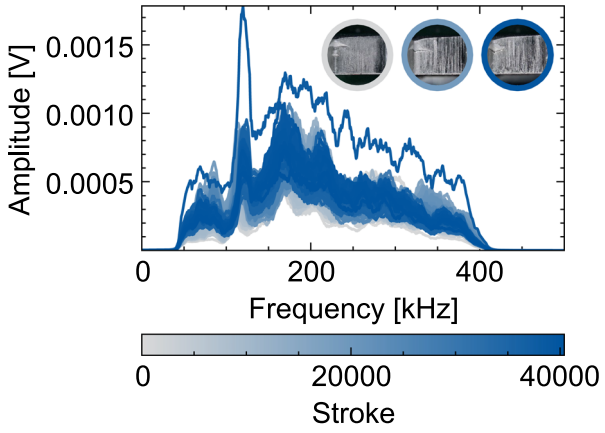

Fig. 17 Frequency content of every 100th stripping segment

for the frequency ranges 119–121 kHz and 117–119 kHz represent the features with the highest and second highest feature importances.

Figure 17 visualizes the frequency content of every 100th stripping segment of the stroke series using consecutive FFTs and 3 resulting scrap web surfaces from the beginning, middle, and end of the experiment, color-coded with their respective stroke number. The FFTs have been smoothed for better visibility with a rolling mean (window size of 600).

As laid out in the section "Introduction", Shanbhag et al. experimentally identified the frequency range 100–200 kHz as the frequency range that carries the most important information about the onset and severity of galling, caused by adhesive wear on the punch. Although more research is needed to understand the causes for the evolution of the scrap web roughness, the identified important frequency range 103–144 kHz points to adhesive wear on the punch and resulting galling as a root cause for the development of the sheared surface roughness of the scrap webs. This would imply significant adhesive wear on the punch during the presumed break-in phase, see Fig. 9.

While similarities in the feature importance between the shearing and stripping phases can be observed, substantial differences occur, e.g. for the *median absolute difference* (MAD), a measure of statistical dispersion of a signal. MAD has a feature importance for stripping segments, however, it does not appear in the top 25 for the shearing segments.

The importance of the frequency-domain features becomes apparent when evaluating the top 10 features. For stripping, 9 of the 10 most important features are frequency-domain features. For shearing, all 10 of the 10 most important features are frequency-domain features. These findings further emphasize the importance of frequency-domain representations for AE signals.

Summary and outlook

The objective of this work was to examine the development of the roughness of selected sheared surfaces of scrap webs (the scrap part resulting from the shearing operation) throughout the life of two fine blanking punches as a proxy for the punch wear and to analyze the correlation of this development with acoustic emission signal features for the shearing and stripping process phases.

The experimental work was carried out using an industrial-grade fine blanking plant with acoustic emission and laser triangulation signals collected for 40,450 strokes. The scrap webs were sampled every 200th stroke and the roughnesses of selected sheared surfaces of these scrap webs were tested. Different imputation methods were used to account for missing data between two subsequent scrap web samples. The acoustic emission signals were segmented into shearing and stripping phases and a comprehensive set of features was extracted for both process phases of each stroke.

In total, 12,800 XGBoost and Lasso regression models were trained on randomly sampled and differently sized training sets and evaluated regarding the mean squared errors of their test set predictions.

Both XGBoost and Lasso were able to model the scrap web surface roughnesses from the features of the acoustic emission signals. XGBoost outperformed Lasso for both process phases and across all data set sizes and imputation methods. Lasso models exhibited high mean squared errors for test sets containing the shearing phase of a stroke with extreme feature values compared to the other strokes (with a label in the usual range) due to their linear structure. Removing the stroke resulted in substantially improved Lasso model performances for the shearing segments, especially for smaller data set sizes.

Evaluation of feature importances suggests that information contained in the frequency range 103–144 kHz was most relevant for successful predictions (models with lower-than-median error) for both the shearing and stripping phases. This finding suggests that the development of the sheared surface roughness of the scrap webs was mainly caused by adhesive punch wear progression and manifested in the acoustic emission signal during both the shearing and stripping phases.

In conclusion, several open research questions and potential for future work arise:

1. For which process configurations (material, punch geometry, stroke frequency,...) and process signals are the *findings reproducible*, and for which not?
2. Which punch wear mechanisms *explain the development of the sheared surface roughness* of the scrap webs?
3. How do *extreme feature values influence model performances and model inference* in different process settings and how can this be handled robustly?

4. In what way do *feature importances and model performances change* with different feature spaces (i.e. time-frequency features)?
5. How does the *proposed approach perform compared to existing punch wear estimators*?

Funding Open Access funding enabled and organized by Projekt DEAL. The authors would like to thank the Federal Ministry for Economics and Climate Action for the kind support within the Project "SPAICER" (Project ID: 01MK20015A). Computational work was performed with computing resources granted by RWTH Aachen University under project RWTH0840.

Declarations

Conflict of interest The authors have no competing interests.

Ethical approval The authors declare that there are no conflicts with the ethical standards by Springer and the research conducted in this research paper.

Consent to participate A consent was obtained from all individuals included in the study.

Consent to publish The publisher has the consent to the authors to publish the given article.

References

Asahi, S., Karadogan, C., Tamura, S., Hayamizu, S., & Liewald, M. (2021). Process data based estimation of tool wear on punching machines using TCN-autoencoder from raw time-series information. *IOP Conference Series: Materials Science and Engineering, 1157*(1), 012078. https://doi.org/10.1088/1757-899X/1157/1/012078

Baccar, D., & Söffker, D. (2015). Wear detection by means of wavelet-based acoustic emission analysis. *Mechanical Systems and Signal Processing, 60–61*, 198–207. https://doi.org/10.1016/j.ymssp.2015.02.012

Barandas, M., et al. (2020). TSFEL: Time series feature extraction library. *SoftwareX, 11*, 100456. https://doi.org/10.1016/j.softx.2020.100456

Behrens, B.-A., Bouguecha, A., Vucetic, M., & Chugreev, A. (2016). Advanced wear simulation for bulk metal forming processes. *MATEC Web of Conferences, 80*, 04003. https://doi.org/10.1051/matecconf/20168004003

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785.

DIN EN 1330-9:2017-10. (2017). Non-destructive testing - terminology - part 9: testing (German version). https://doi.org/10.31030/2607064

Groover, M. (2010). *Fundamentals of modern manufacturing: Materials, processes, and systems* (4th ed.). Wiley.

Klocke, F. (2014). *Manufacturing processes 4: Forming*. Springer.

Kollment, W., O'Leary, P., Harker, M., Klünsner, T., & Eck, S. (2018). Force and acoustic emission measurements for condition monitoring of fine blanking tools. In *2018 IEEE international instrumentation and measurement technology conference* (pp. 1–6). IEEE. https://doi.org/10.1109/I2MTC.2018.8409569.

Kubik, C., Knauer, S., & Groche, P. (2022). Smart sheet metal forming: importance of data acquisition, preprocessing and transformation on the performance of a multiclass support vector machine for predicting wear states during blanking. *Journal of Intelligent Manufacturing, 33*(1), 259–282. https://doi.org/10.1007/s10845-021-01789-w

Lundberg, S., et al. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence, 2*(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

Lundberg, S., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Proceedings of the 31st international conference on neural information processing systems* (pp. 4768–4777). Curran Associates Inc.

Niemietz, P., Kornely, M., Trauth, D., & Bergs, T. (2022). Relating wear stages in sheet metal forming based on short-and long-term force signal variations. *Journal of Intelligent Manufacturing, 33*(7), 2143–2155. https://doi.org/10.1007/s10845-022-01979-0

Niemietz, P., Unterberg, M., Trauth, D., & Bergs, T. (2021). Autoencoder based wear assessment in sheet metal forming. *IOP Conference Series: Materials Science and Engineering, 1157*(1), 012082. https://doi.org/10.1088/1757-899X/1157/1/012082

Shanbhag, V., Rolfe, B., & Pereira, M. (2020). Investigation of galling wear using acoustic emission frequency characteristics. *Lubricants, 8*(3), 25. https://doi.org/10.3390/lubricants8030025

Shwartz-Ziv, R., & Armon, A. (2022). Tabular data: Deep learning is not all you need. *Information Fusion, 81*, 84–90. https://doi.org/10.1016/j.inffus.2021.11.011

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B, 58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Unterberg, M., et al. (2021). Wear monitoring in fine blanking processes using feature based analysis of acoustic emission signals. *Procedia CIRP, 104*, 164–169. https://doi.org/10.1016/j.procir.2021.11.028

Voigts, H. O. (2021). Feinschneiden mit Hartmetallstempeln: Fine blanking with cemented carbide punches. Apprimus.

XGBoost parameters: xgboost 1.7.2 documentation. Retrieved January 4, 2023, from https://xgboost.readthedocs.io/en/stable/parameter.html

Yeh, C.-C. M., et al. (2016). Matrix profile I: All pairs similarity joins for time series: A unifying view that includes motifs, discords and shapelets. In *2016 IEEE 16th international conference on data mining* (pp. 1317–1322). IEEE. https://doi.org/10.1109/ICDM.2016.0179.