Machine learning with domain knowledge for predictive quality monitoring in resistance spot welding

Baifan Zhou[1,2,3] · Tim Pychynski[1] · Markus Reischl[2] · Evgeny Kharlamov[3,4] · Ralf Mikut[2]

Abstract

Digitalisation trends of Industry 4.0 and Internet of Things led to an unprecedented growth of manufacturing data. This opens new horizons for data-driven methods, such as Machine Learning (ML), in monitoring of manufacturing processes. In this work, we propose ML pipelines for quality monitoring in Resistance Spot Welding. Previous approaches mostly focused on estimating quality of welding based on data collected from laboratory or experimental settings. Then, they mostly treated welding operations as independent events while welding is a continuous process with a systematic dynamics and production cycles caused by maintenance. Besides, model interpretation based on engineering know-how, which is an important and common practice in manufacturing industry, has mostly been ignored. In this work, we address these three issues by developing a novel feature-engineering based ML approach. Our method was developed on top of real production data. It allows to analyse sequences of welding instances collected from running manufacturing lines. By capturing dependencies across sequences of welding instances, our method allows to predict quality of upcoming welding operations before they happen. Furthermore, in our work we strive to combine the view of engineering and data science by discussing characteristics of welding data that have been little discussed in the literature, by designing sophisticated feature engineering strategies with support of domain knowledge, and by interpreting the results of ML analysis intensively to provide insights for engineering. We developed 12 ML pipelines in two dimensions: settings of feature engineering and ML methods, where we considered 4 feature settings and 3 ML methods (linear regression, multi-layer perception and support vector regression). We extensively evaluated our ML pipelines on data from two running industrial production lines of 27 welding machines with promising results.

Introduction

Technological advances in trends of Industry 4.0 (Kagermann 2015) and Internet of Things (ITU 2012), including technologies in sensoring, communication, information processing, and actuation, have opened horizons of new opportunities and challenges to change the paradigm of many industrial processes, such as manufacturing, oil and gas, chemical and process industries. Kagermann describes Industry 4.0 in the way that smart machines, storage systems and production facilities will be incorporated into aggregate solutions that are often referred to as Cyber-Physical Systems (Kagermann 2013; NSF 2010). Furthermore, such systems are then integrated in smart factories.
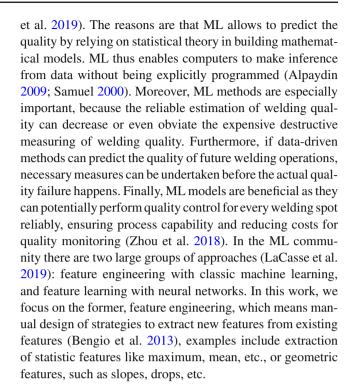
This trend towards smart factories unlocked an unprecedented volumes of data. Indeed, modern manufacturing machines and production lines are equipped with sensors that constantly collect and send data. The machines have control units that monitor and process the data, coordinate machines and manufacturing environment and send messages, notifications, requests. Such data generated during manufacturing (Chand and Davis 2010; Wuest et al. 2016) has led to a large growth of interest in data analysis for a wide range of industrial applications (Mikhaylov et al. 2019a, b; Zhou et al. 2017, 2019).

✉ Baifan Zhou
baifanz@ifi.uio.no

[1] Bosch Corporate Research, 71272 Renningen, Germany

[2] Karlsruhe Institute of Technology, 76344 Eggenstein-Leopoldshafen, Germany

[3] SIRIUS Centre, University of Oslo, 0316 Oslo, Norway

[4] Bosch Center for Artificial Intelligence, 71272 Renningen, Germany

Resistance spot welding. In this work, we investigate such data analysis for a particular industrial scenario of *Resistance Spot Welding (RSW)* and its applications at Bosch, a large manufacturing company that is one of the world leaders in automated welding in the automotive industry.

We illustrate RSW with Fig. 4, in which the two electrode caps of the welding gun press two or three worksheets between the electrodes with force. A high electric current then flows from one electrode, through the worksheets, to the other electrode, generating a substantial amount of heat as a result of electric resistance. The material in a small area between the worksheets will melt, and form a welding nugget connecting the worksheets, known as the welding spot. The quality of welding operations is typically quantified by quality indicators like spot diameters, as prescribed in international and German standards (ISO 2004; DVS 2016). To obtain the spot diameters or tensile shear strength precisely, the common practice is to tear the welded chassis apart and measure these two quality indicators DVS (2016), which is extremely expensive. Nevertheless, monitoring of welding quality has a great importance. Indeed, consider a scenario in the car factory, where cars are continuously produced in several production lines. In each production line, a series of chassis is produced, with up to six thousands of spots welded on each chassis. If a quality failure happens on one spot, the whole production line needs to be stopped, which means the loss of several cars, production down-time, and cost to bring the production line back to running. Thinking about the number of cars produced everyday, it reveals the huge economic benefit behind improving quality monitoring of RSW. Furthermore, if the technology developed for improving RSW can be generalised over other manufacturing processes or other industries, the industrial impact behind the research endeavour to improve RSW will be tremendous.

Machine learning for resistance spot welding. Bosch RSW solutions are fully automated and produce large volumes of heterogeneous data. Therefore, we focus on data analyses in this work, in particular, on Machine Learning (ML), for quality monitoring [1] for RSW. Note that ML approaches have proven their great potential for quality monitoring and thus they have received an increasing attention in industry (Zhao

et al. 2019). The reasons are that ML allows to predict the quality by relying on statistical theory in building mathematical models. ML thus enables computers to make inference from data without being explicitly programmed (Alpaydin 2009; Samuel 2000). Moreover, ML methods are especially important, because the reliable estimation of welding quality can decrease or even obviate the expensive destructive measuring of welding quality. Furthermore, if data-driven methods can predict the quality of future welding operations, necessary measures can be undertaken before the actual quality failure happens. Finally, ML models are beneficial as they can potentially perform quality control for every welding spot reliably, ensuring process capability and reducing costs for quality monitoring (Zhou et al. 2018). In the ML community there are two large groups of approaches (LaCasse et al. 2019): feature engineering with classic machine learning, and feature learning with neural networks. In this work, we focus on the former, feature engineering, which means manual design of strategies to extract new features from existing features (Bengio et al. 2013), examples include extraction of statistic features like maximum, mean, etc., or geometric features, such as slopes, drops, etc.

Limitation of previous works. Many previous studies have adopted the approach of data-driven models for quality monitoring in RSW. Many have studied the problem, to classify (Martín et al. 2007; Sun et al. 2017), estimate (El Ouafi et al. 2010; Wan et al. 2016), or optimise (Panchakshari and Kadam 2013) welding quality of each individual operations. In most studies, data were collected from laboratory or experimental settings (Summerville et al. 2017; Boersch et al. 2016). The typical data amount with labels (quality data) was less than 500 Zhang et al. (2017), very few of them above 3000 (Kim and Ahmed 2018). A number of different features were used for analysis, such as process curves like electrode displacement (Li et al. 2012), or scalar process parameters like coating (Yu 2015). Various methods of feature processing and ML modelling were explored, including classic ML methods like Random Forests (Sumesh et al. 2015), or Neural Networks (Afshari et al. 2014).

However, previous works treated the welding operations as independent events, ignoring the fact that welding is a continuous process with systematic dynamics, e.g. wearing of welding tools, production cycles caused by maintenance. The reason of this is because the characteristics of welding data were insufficiently recognised. Very few studies were conducted on real production data. It is questionable whether models developed from laboratory data are applicable in real production, as the welding conditions (such as cooling time and wear) are usually different. Furthermore, although some feature processing methods seemed to consider some domain knowledge of welding, the integration of domain knowledge in data analysis was limited, and the interpretation of data

---

[1] Note that quality monitoring that we investigate here falls into a wider well-known category of *condition monitoring* that is typically divided into two categories *machine health monitoring* (Zhao et al. 2019) that aims at maintaining the healthy state of an equipment, and quality monitoring that aims at ensuring the product quality within acceptable limits. The common quality monitoring practice is to estimate or assess the product quality after a manufacturing operation happens, so as to provide results for future manufacturing operations. While predictive quality monitoring, is to predict the quality of a manufacturing operation before the actual operation happens, so that necessary measures can be undertaken even before the quality deficiency happens (Zhou et al. 2020a).

Fig. 1 Welding workflow with quality prediction. If the predicted quality (Q-Value in this paper) of the next welding operation is good, the welding process continues. Otherwise necessary measures are to undertake, e.g. changing electrode caps, dressing electrodes ("Resistance spot welding" section)

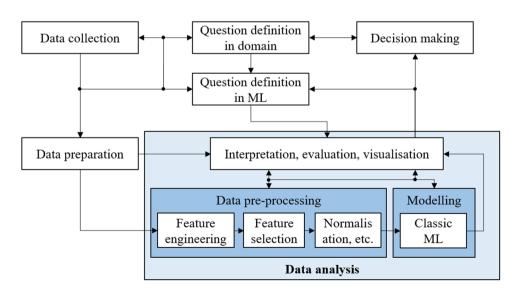analysis also provided limited insights from perspective of engineering know-how.

Our approach. In this work, we develop ML approaches to *predict* the welding quality before the actual welding process happens, considering the characteristics of welding data, especially the temporal dependency. The envisioned welding process is to undertake necessary measures before possible quality failures (Fig. 1). Furthermore, this work strives to integrate domain knowledge in machine learning modelling, combining views from data science and welding engineering know-how. We focus on feature engineering because it is more transparent than feature learning / deep learning, since transparency is very desired in a industrial setting. Four settings of engineered features are designed for machine learning modelling to explore and test whether and to what degree feature engineering can increase the prediction power. Three ML methods, Linear regression (LR), multi-layer perceptron with one hidden layer (MLP), and support vector regression (SVR) are studied as representative classic machine learning methods (LaCasse et al. 2019). The combination of the feature processing settings and ML methods gives 12 ML pipelines. The developed ML pipelines are extensively evaluated with data collected from two run-

ning *industrial production lines* of 27 welding machines at Bosch's plants or development partners. Results from two representative datasets collected from two welding machines are demonstrated in this paper. In total, results of 24 ML models are demonstrated. The algorithms in this paper are implemented with the MATLAB toolbox SciXMiner (Mikut et al. 2017; Bartschat et al. 2019) and Python (Rossum 1995).

The machine learning in this work (Fig. 2) is slightly adapted from the pipeline of Fayyad et al. (1996) and Mikut et al. (2006). The complete pipeline of ML for data analysis includes data collection, data preparation, data pre-processing, modelling, and interpretation. Data preparation refers to the activities that transform the industrial data from different sources and formats, e.g. csv, SQL data base, txt, xlsx, to a uniform format so that the data can be processed in a uniform way. Data pre-processing is basically the process of feature extraction (in this work feature engineering specifically), that is to change the representation of data (Bengio et al. 2013), so that it can be suitable for the subsequent machine learning modelling.

Our contributions. We now summarise the main contributions of our paper as follows:

- We conducted an in depth study and revealed characteristics of welding data collected from running production with domain knowledge. These characteristics are the natural results of welding production and reveal the temporal dependencies of welding operations, which to the best of our knowledge have never been discussed in depths in the literature (except for minor mentioning). The discussion and visualisation of the data are important for understanding the data from both perspectives of engineering and data science.
- We demonstrated that sophisticated feature engineering with support of domain knowledge can greatly improve

Fig. 2 Basic machine learning pipeline in this work (Fayyad et al. 1996; Mikut et al. 2006; LaCasse et al. 2019). Question definition is illustrated in Fig. 1. The collected data is described in "Data and description" section and prepared to the formats in Data in two format. The data pre-processing and modelling are studied in "methodology" section. Interpretation and visualisation are presented in "Experiment results" section and "discussion" section

the performance of ML methods for RSW. To this end we designed 4 settings with 4 levels of feature engineering. This in particular illustrates how domain knowledge can play an important role in ML-based data analysis.

- We developed and compared novel ML-based methods for predicting welding quality. The adopted ML algorithms include Linear Regression (LR), Multi-layer Perception (MLP) and Support Vector Regression (SVR). The combination of three ML algorithms and four settings of feature engineering gives twelve ML models for each dataset. We showed that, on the one hand, simplistic methods such as LR can have comparable performance to MLP, given that efficient feature engineering strategies are adopted; while on the other hand, LR has a very desired feature in manufacturing industry, which is transparency, compared to less transparent methods like MLP or Deep Learning.
- We conducted an extensive evaluation of the twelve ML models with real industrial data collected from two welding stations in running production lines, resulting in twenty four models in total. This in particular provides a guidance to a wider community on how to develop of ML-based quality monitoring systems in production.
- We interpreted the ML results and feature selection with engineering knowledge and provided insights that enable engineers to understand the data and the process from a data science perspective. This demonstrates the advantage of transparency of domain knowledge supported feature engineering, compared to approach of feature learning. Thus, ML can be a natural method in the toolbox of engineers for common engineering practice.

This work has been conducted as a part of the PhD study of the first author (Zhou 2021). The material presented in this paper significantly extends our previously presented conference paper (Zhou et al. 2020a) as follows. First, we provide an extensive systematic review of related work of the past 20 years. Second, we discuss characteristics of welding data that were only shortly mentioned in the conference paper, especially hierarchical temporal structures. Third, we incorporate domain knowledge more deeply involved in data handling, data splitting, and feature engineering for time series. Fourth, we design four feature engineering strategies: two of them are new, and the other two were only briefly discussed in the conference paper while here we present them in much more details. Fifth, here we present three types of classic ML pipelines (LR, MLP, SVR) and only one of them (that is, LR) was in the conference paper. Sixth, here we evaluate our 12 ML pipelines with a dataset that is much more complex and reflect more real complicated dynamics in production than the one used in the conference paper. Seventh, we introduce a new section that is entirely devoted to an extensive interpreta-

tion of modelling results to provide insights for engineering – this was not presented the conference paper.

Organisation of the paper. This paper is organised as follows. "Related work" section gives a detailed review on related work and reveals their limitation. The "Data and Problem Statement" section describes the Resistance Spot Welding process and the structures of the data collected from the production lines. "Methods" section introduces the strategies for handling the special data structures and for feature engineering. Section "Experiment settings" section describes the experiment settings for evaluating the feature engineering approaches. "Results and discussion" section presents and discusses the evaluation results. "Interpreting ML results for engineering insights" further interprets the features, visualisation and results to extract engineering insights. "Conclusion and outlook" section concludes the work and previews the future research directions.

Related work

Many previous studies have adopted the approach of data-driven models for quality monitoring (summarised in Table 1). It can be seen that regression (R) and classification (C) has been the focus in the past 20 years. The interest of machine learning for RSW has been growing in the recent years. This work will discuss and summarise them from four perspectives: question definition, data collection, feature processing, and machine learning modelling.

Question definition. There exist two aspects of question definition. The first one is which quality indicator is used as the target feature. Most previous works have studied estimation of the spot diameter (Boersch et al. 2016; Kim and Ahmed 2018), as this is the suggestion by standards. Many works studied estimation of tensile shear strength (Cho and Rhee 2000; Martín et al. 2009; Zhang et al. 2017; Sun et al. 2017), and other less common quality indicators like load (Tseng 2006), gaps (Hamedi et al. 2007), penetration (El Ouafi et al. 2010), pitting potential (Martín et al. 2010). All of these quality indicators are physical quantities that can be measured.

The second aspect is to study the question from the perspective of classification, regression, or optimisation. Many works (Lee et al. 2003; Cho and Rhee 2004; El-Banna et al. 2008; Yu 2015) treated the problem as classification, that is to predict the category of quality: good, bad, and sometimes the concrete failure types, which is the final important goal for quality monitoring. Some works (Lee et al. 2001; Martín et al. 2009; Afshari et al. 2014; Summerville et al. 2017) defined the question as regression, that is to assess a numerical value of the quality. This may seem to be unnecessary, but could be beneficial in many senses. The exact quality values are of great interest for process experts to gain better

Table 1 An overview of related works of machine learning in RSW. All studies are carried out with a **laboratory** data source on RSW except for (Kim and Ahmed 2018). #Data: number of data tuples. C: classification, R: regression, O: optimisation, TSS: tensile shear strength, D: diameter, h: height, E_{pitt}: pitting potential, SF: single features, FE: feature engineering TSFE: time series features engineered, DKFE: domain knowledge supported feature engineering, Re: resistance, F: force, t: time, I: current, s: displacement, U: voltage, MLP: multi-layer perceptron, NN: neural networks, LR: linear regression, SOM: self-organising maps, BN: Bayesin network, LVQ: learning vector quantisation, LDA/QDA: linear/quadratic discriminant analysis, kNN: k-nearest neighbours, GRNN: general regression neural networks, GA: genetic algorithms, LogisticR: logistric regression, ANOVA: analysis of variance, DT: decision trees, RF: random forests, SVR/SVM: support vector regression/machine, PolyR: polynomial regression PSO: particle swarm optimisation, KELM: kernel extreme learning machines, CART: classification and regression tree, GLM: generalised linear model, SAE: sparse auto-encoder, CNN: convolutional neural networks

	Question	#Data	Feature Processing	ML Methods
Cho and Rhee (2000)	R: TSS	50	TS: Re; TSFE: DKFE (slope, max, std, mxpo)	MLP
Li et al. (2000)	R: D	170	SF: F, t, I; TS: F, Re, s; TSFE: subsampling; FE: PCA, rms	MLP
Lee et al. (2001)	R: TSS	80	TS: Re, s; TSFE: subsampling, DKFE (slope, mxpo, range)	Fuzzy NN
Cho and Rhee (2002)	R: D, TSS	60	TS: Re; TSFE: DKFE (max, slope, min, std), stats	LR, MLP
Lee et al. (2003)	C: D	390	Image from SAM; ImageFE: dilation, quantisation	MLP
Cho and Rhee (2004)	C: TSS	10	TS: Re; TSFE: TS to bipolarised image	HopfieldNN
Junno et al. (2004)	C: D	192	TS: I, F, U; TSFE: means of segmented TS	SOM
Laurinen et al. (2004)	C: D	192	TS: I, F, U; TSFE: histogram, discretisation of quartiles	BN
Park and Cho (2004)	C: TSS, h	78	TS: s; TSFE: subsampling, slopes	LVQ, MLP
Podržaj et al. (2004)	C: spatter	30	TS: U, I, Re, s, F; TSFE: subsampling, std	LVQ
Haapalainen et al. (2005)	C: TSS	3879	TS: I, U; TSFE: transitions, segmentation mean; FE: PCA	LDA, QDA, kNN, LVQ
Tseng (2006)	R, O: load	25	SF: I, F, t, thickness	GRNN, GA
Hamedi et al. (2007)	R, O: gaps	39	SF: I, t, thickness	MLP, GA
Koskimaki et al. (2007)	C: TSS	3879	TS: I, U; TSFE: segmentation mean	kNN, SOM
Martín et al. (2007)	C: D	438	TS: ultrasonic oscillogram; TSFE: DKFE (heights & distances of echoes)	MLP
El-Banna et al. (2008)	C: D	1270	TS: R; TSFE: segmentation, DKFE (max, min, mean, std, range, rms, slope); FS: power of the test	LVQ
Haapalainen et al. (2008)	C: TSS	3879	TS: I, U: TSFE: transitions, segmentation mean; FE: PCA; FS: SFS, SBS, SFFS, SBFS, nBest	kNN
Martín et al. (2009)	R: TSS	242	SF: t, I, F	LR, MLP
El Ouafi et al. (2010)	R: h, penetration, D	54	SF: Thickness, I, F, t; TS: Re; TSFE: subsampling	MLP
Martín et al. (2010)	R: E_{pitt}	242	SF: t, I, F; FE: polynomial features	MLP
Li et al. (2012)	R: D	145	TS: s; TSFE: DKFE (geometric features)	MLP
Panchakshari and Kadam (2013)	R, O: D, TSS, load	25	SF: I, welding time, holding time, squeezing time; FS: ANOVA	LR, GA

Table 1 continued

	Question	#Data	Feature Processing	ML Methods
Afshari et al. (2014)	R: D	54	SF: F, t, I	MLP
Yu (2015)	C: TSS, spatter	473	SF: coating, F; TS: Power; TSFE: DKFE (mxpo, max, drop)	LogisticR
Pereda et al. (2015)	C: D	330	SF: t, I, material, treatment	ANOVA, DT, NN, RF, SVR, LogisticR, QDA
Zhang et al. (2015)	C: TSS, D	200	TS: s; TSFE: segmentation, mean, radar chart, geometric features	DT
Boersch et al. (2016)	R: D	3241	TS: I, U, T; TSFE: derivative, segmentation, scale-space filtering, geometric features, stats	LR, DT, RF, SVM, kNN
Pashazadeh et al. (2016)	R, O: D, h	48	SF: t, I, Pressure	PolyR, MLP, GA
Wan et al. (2016)	C, R: D, load	60	SF: F, I, t, TS: Re; TSFE: PCA	MLP
Summerville et al. (2017)	R: D	126	TS: Re; TSFE: PCA	LR
Summerville et al. (2017)	R: TSS	170	TS: Re; TSFE: PCA	PolyR
Sun et al. (2017)	C: TSS	67	SF: I, t, Pressure	PSO, KELM
Zhang et al. (2017)	C: TSS	120	TS: s; TSFE: DKFE stats, Chernoff images, binary features	HopfieldNN
Hou et al. (2017)	C: good/bad	88	X-ray images	SAE
Kim and Ahmed (2018)	R: D	3344	SF: material, thickness, t, coating, I	DT, CART
Gavidel et al. (2019)	R: D	39	SF: material, thickness, various surface conditions, F, I, t	CART, RF, MLP, kNN, SVM, GLM, Bootstrapping
Amiri et al. (2020)	R: TSS, fatigue life	60	SF: statistic features extracted from ultrasonic images	MLP
Dai et al. (2021)	C: good/bad	400	pictures of outlook of spots assisted visual inspection	CNN

insights in the influence of input factors on the quality indicator. After predicting the numerical values of spot diameters, a classification can still be made according to the tolerance bands, which usually vary for different welding conditions and user specifications. Other works (Tseng 2006; Hamedi et al. 2007; Pashazadeh et al. 2016) studied the optimisation problem, that is to optimise the influential factors so that the quality can improve.

Data collection. It is an important issue to discuss from two aspects. The first aspect is the data amount. The amount of data labelled with quality features is extremely limited (Zhou et al. 2018) due to the costly data collection process as discussed before. Therefore, most of the previous works used a relatively small amount of data. The number of welding spots ranges from 10 (Cho and Rhee 2004) to less than 4000 (Haapalainen et al. 2008). The typical data amount in literature is less than 500, e.g. in (Tseng 2006; Sun et al. 2017; Zhang et al. 2017; Summerville et al. 2017; Zhang et al. 2015; Martín et al. 2010; Laurinen et al. 2004).

The second aspect is the data source. There could be three major data sources: (1) simulation data (Zhou et al. 2018), where the most amount of data labelled with quality features could be produced with less cost, but the data conditions may deviate from production because it is very difficult to perfectly reproduce the real production conditions in simulations; (2) laboratory data (Summerville et al. 2017), where the welding condition could be more similar to real production but less labelled data can be produced; (3) and production data, which is the final target application of welding quality monitoring but there exist the most restrictions on amount of labelled data, cost for data collection and number of sensors. Almost all of the previous studies have collected data from laboratories or experimental settings (Lee et al. 2001, 2003; Koskimaki et al. 2007; Martín et al. 2010; Wan et al. 2016; Sumesh et al. 2015; Yu 2015), except for Kim and Ahmed (2018), which used about 3400 welded spots accumulated production data.

Feature Processing. In the literature, two types of features are commonly used. The first type is single features, which are recorded as (aggregated) constants for a welding operation. Examples include welding time (t) (Li et al. 2000), sheet thickness (thickness) (El Ouafi et al. 2010), sheet coating (coating) (Yu 2015), electrode force (F) (Tseng 2006), welding current (I) (Hamedi et al. 2007), pressure (Pashazadeh et al. 2016). The second type is sensor measurements, or process curves, which are physical quantities measured along time, and are therefore referred to as time series (TS). Examples include dynamic resistance (Re) (Cho and Rhee 2000), electrode force (F) (Junno et al. 2004), welding current (I) (Haapalainen et al. 2005), electrode displacement (s) (Park and Cho 2004), welding voltage (U) (Haapalainen et al. 2008), ultrasonic images (Martín et al. 2007; Amiri

et al. 2020), power. Besides the two common types, Lee et al. (2003) used images collected from Scanning Acoustic Microscopy (SAM).

Some authors have attempted to exploit domain knowledge for designing feature engineering strategies to extract features from process curves. The works (Cho and Rhee 2000, 2002) used statistic or geometric features of process curves, like slope, maximum, standard deviation, position of maximum, range, root-mean-squares of measurements (El-Banna et al. 2008). Some more specific features based on domain knowledge, such as heights and distance between echoes from ultrasonic oscillograms were studied (Martín et al. 2007). Geometric features (Li et al. 2012) were introduced later, e.g. "drop" in process curves (Yu 2015), which is the value decrease from a peak to the following valley. In recent years, richer features were extracted, e.g. derivative, filtering (Boersch et al. 2016), and Chernoff images drawn with statistic features from process curves (Zhang et al. 2017).

Feature Selection (FS), such as analysis of variance (ANOVA) Panchakshari and Kadam (2013), step-wise regression (Panchakshari and Kadam 2013), etc. was performed. The work (Haapalainen et al. 2008) discussed and compared various feature selection methods, e.g. Sequential Forward Selection (SFS), Sequential Backward Selection (SBS), Sequential Forward Floating selection (SFFS), Sequential Backward Floating Selection (SBFS) and N-Best Features Selection (nBest).

ML Modelling. Most of the methods used for machine learning modelling can be classified as classical machine learning methods (LaCasse et al. 2019), like Linear Regression (LR) (Cho and Rhee 2002; Martín et al. 2009; Panchakshari and Kadam 2013), Polynomial Regression (PolyR) (Pashazadeh et al. 2016; Summerville et al. 2017), or Generalised Linear Models (GLM) Gavidel et al. (2019), k-Nearest Neighbours (kNN) (Haapalainen et al. 2005; Koskimaki et al. 2007; Boersch et al. 2016), Decision Trees (DT) (Zhang et al. 2015; Kim and Ahmed 2018), Random Forests (RF) (Pereda et al. 2015; Boersch et al. 2016), Support Vector Machines (SVM), etc. Statistic methods like Linear or Quadratic Discriminate Analysis (LDA and QDA) are also used for classification. Bayesian Networks (BN), Genetic Algorithms (GA) (Tseng 2006; Panchakshari and Kadam 2013), and Particle Swarm Intelligence (PSO) Sun et al. (2017) are often used for optimisation. The Artificial Neutral Networks (ANN) used in previous studies, include Fuzzy Neural Networks (FuzzyNN) (Lee et al. 2001), Learning Vector Quantisation (LVQ) (Junno et al. 2004), Self-Organising Maps (SOM) (Junno et al. 2004), General Regression Neural Networks (GRNN) (Tseng 2006), Hopfield Neural Networks (HopfieldNN) (Zhang et al. 2017), Kernel Extreme Learning Machines (KELM) (Sun et al.

2017) and Multilayer-Perceptrons (MLP). Since these networks either have fewer than two hidden layers, or do not demonstrate the characteristic of hierarchical feature learning (Bengio et al. 2013), they can still be classified in the category of classic machine learning.

There exist also several works that use deep learning (DL) for quality monitoring in RSW. The work (Hou et al. 2017) applied sparse auto-encoder network for detecting welding defects from X-ray images. Convolutional neural networks (CNN) are proposed (Dai et al. 2021) to classify spot quality assist visual inspection for classifying spots into good/bad based on pictures of outlook of welded car body parts. Long short-term memory neural networks (Zhou et al. 2020a) are applied for learning abstract representation from the temporal data for predicting quality. Many more works exist that apply DL for quality monitoring or optimisation in other welding processes, such as laser welding (Mikhaylov et al. 2019a; Shevchik et al. 2020; Zhang et al. 2019a), arc welding (Nomura et al. 2021; Zhang et al. 2019b), resistance wire welding (Guo et al. 2017) etc. It can been seen there exist much more works about classic ML than DL for quality monitoring of RSW. In our industrial environment classic methods are also preferred than DL methods. One reason for that could be that the amount of collected data is large enough to demonstrate the advantage of DL methods, as deep learning usually requires very large amount of data LaCasse et al. (2019). DL becomes more useful when the analysed data are "heavier", e.g. X-ray images, pictures, and laser welding images. However, quality judged from X-ray images and pictures for RSW are less reliable (indirect quality indicators) and less precise (only good/bad instead of numeric values), compared to diameters and other numerical quality indicators. Another reason could be, the relationship between input features and target features is not very complex, or not very non-linear, so that deep learning methods cannot demonstrate their benefits (Bengio et al. 2013).

Summary. There exist the following aspects that the previous work has not covered sufficiently:

- The temporal dependency of welding data is limitedly discussed. Each welding operation has been deemed as an independent event in almost all literature, that is to say, the quality estimation of each welding operation is carried out independently from other operations. This is logically and physically true, but there could be some systematic trends between the welding operations performed continuously, e.g. the wearing effect of welding electrodes should also increase continuously.
- Due to the insufficient recognition of interdependency between welding operations, the literature has been focusing on estimating the welding quality with the data of each welding operation, but the prediction of future

welding quality has been largely ignored. If the functionality of estimation of welding quality is realised in a welding system, this system can make an quality estimation for each welding operation only *after* the operation, during which the quality failure may already happened.

- Most of the previous work has used a limited amount of data collected from experimental settings. It is questionable whether models developed from laboratory data are applicable in real production, as the welding conditions (such as cooling time and wear) are usually different. More studies should be conducted on a larger amount of data (e.g. more than 4000 welding spots) and data collected from real industrial production.
- Domain knowledge is insufficiently considered in ML analysis. This can be seen from the fact that most previous studies treated each welding operation independently. In addition, the design of feature engineering strategies can rely more on domain knowledge. Furthermore, feature selection has been performed, but was not used for understanding the importance of input features on quality features. The interpretation of ML analysis results and features should rely more on domain knowledge and generate more insights for a better understanding of the process and data.

Data and problem statement

This section gives a brief introduction to the resistance spot welding process ("Resistance spot welding"section), discusses the characteristics of welding data collected from running industrial production lines ("Hierarchical temporal structures in the data" section), their formats ("data into formats" section), and defines the problem of predictive quality monitoring ("Problem statement" section).

Resistance spot welding

Resistance Spot Welding (RSW) is widely applied in automotive industry for e.g. car chassis production. During the welding process, the two electrode caps, equipped at the end of the two welding electrodes (Fig. 4), press two or three worksheets (two layers of car chassis parts in automotive industry) with force. Then an electric current flows from one electrode, through the worksheets, to the other electrode, generating a large amount of heat due to electric resistance. The materials in a small area between the two worksheets, called the welding spot, will melt, and form a weld nugget connecting the worksheets. The electrode caps directly touching the worksheets wear quickly due to high thermal-electric-mechanical loads and oxidation, and need to be maintained regularly. After a fixed number of spots are welded, a very thin layer of the cap will be removed to restore the surface
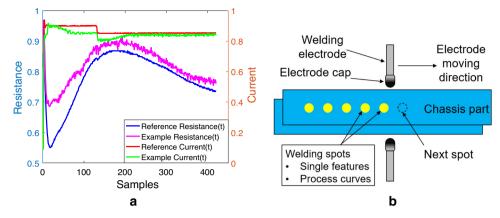
Fig. 3 (a) Examples of process curves. The adaptive control will try to force the process curves to follow the reference curves defined by the welding program. Left y-axis: Resistance (anonymised), right y-axis: Current (anonymised), x-axis: Samples (b) Consecutive welding operations constitute the welding operation level. For each operation, a data tuple consisting of single features on the welding operation level and process curves on the welding time level (Sect. 3.3) are recorded by the adaptive control system that tries to force the actual curves follow the reference curves
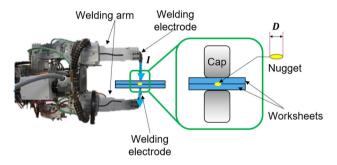
Fig. 4 A schematic illustration of a welding gun in Resistance Spot Welding (RSW) Zhou et al. (2018). The electrode caps, equipped at the end of two welding electrodes, press the worksheets (e.g. chassis parts in automotive industry) with force. A high electric current (the blue arrows) flows through, generating heat and forming a welding nugget, whose diameter D is an important quality indicator

condition of the electrode cap. This is called *Dressing*. After a certain number of dressings are performed, the electrode cap becomes too short and needs to be changed altogether. This is called *Cap Change*.

The welding process is controlled by the welding control system, which is an *Adaptive Control System* and also provides the function to store the collected data. It will try to force the electric current flowing through the electrodes and worksheets to follow a pre-designed profile. This profile is called reference current curve (Fig. 3a) in *Adaptive Control*, which is determined in the process development stage. Apart from the reference current curve, other reference curves, e.g. voltage, are also determined in the process development stage. A complete set of all reference curves are stored in a *welding program*. The welding program also prescribes other information, e.g. the welding position on the car part, the thickness, material, surface coating condition of the worksheets and the glue between the worksheets.

The diameter of the welding nugget is typically used as the quality indicator of a single welding act according to international standards (ISO 2004) and German standards DVS (2016), but it is difficult to measure precisely. Destructive methods, i.e. tearing the welded chassis apart, are expensive and can only have the quality partially controlled. Non-destructive methods, including ultrasonic wave and X-ray, are also costly, time-consuming and yield imprecise results (El Ouafi et al. 2010; Zhang et al. 2004; Cho and Rhee 2000). In industry, substitute quality indicators are often developed to describe the welding quality. For example, the *Q-Value* developed by *Bosch Rexroth*, will be studied in this work. Q-Value is an aggregated value calculated using process curves, by extracting their statistic, geometric and other features that incorporate domain knowledge. The exact way of calculation of Q-Value cannot be disclosed here since it is a know-how of Bosch Rexroth. The optimal Q-Value is one; values below one normally indicate quality deficiency, and values above one indicate energy inefficiency.

Hierarchical temporal structures in the data

Previous studies (Boersch et al. 2016; Zhang et al. 2004; Cho and Rhee 2000) have treated each welding operation independently. If we closely examine the data, e.g. Fig. 5a showing the Q-Value along a number of welding operations for an example welding machine, we can see clearly that the data has strong periodicity, which indicates that the data very likely have temporal dependencies. This section elaborates the multi-levels of temporal structures in data as an intrinsic result of the structure of production processes.

The first time level is the *welding time level* during a single welding operation, which are usually several hundreds of samples (Fig. 3a). For each welding operation, data of process
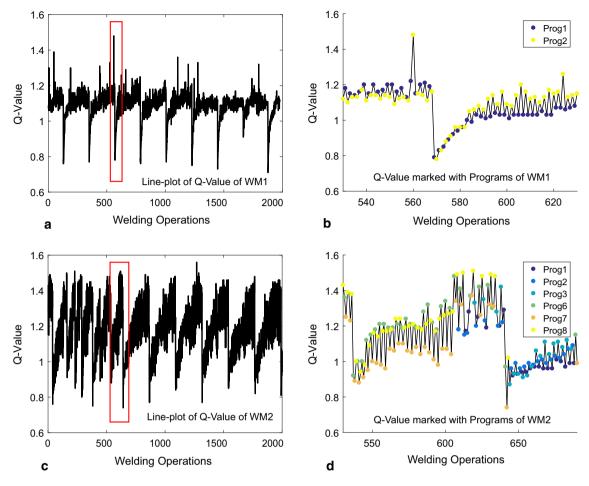
Fig. 5 (a) Q-Value along a number of welding operations for Welding Machine 1. The red rectangle indicates the area for a closer look in (b). Meaning of Q-Values, 1: optimal, <1: quality deficiency; >1: energy inefficiency. The Q-Values raise gradually due to wearing effects, and drop abruptly when there is a maintenance. (b) Welding operations with different welding programs are performed for spots of different welding positions on the car part, thus often possessing different dynamics, e.g. the means of Q-Values are different. (c) Q-Value along a number

of welding operations for Welding Machine 2 (partially shown). The irregular trends of Q-Value indicate complex production conditions. The red rectangle indicates the area for a closer look in (d). (d) Welding operations performed with different welding programs. Note that there exists a change of production arrangement. Before the 618th welding operation, only three welding programs were performed (Prog6, 7 and 8). After that three MORE welding programs were performed (Prog1, 2 and 3) due to a change of production plan
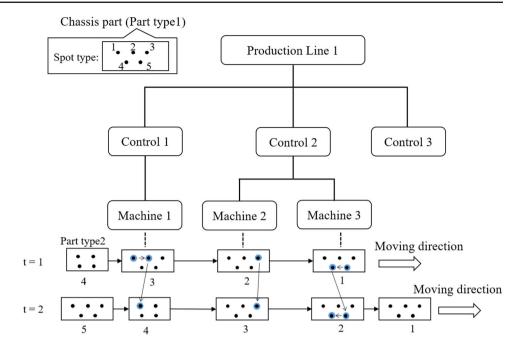
curves and single features are recorded (Fig. 3b). The consecutive welding operations constitute the second time level, the *welding operation level*. A data sample on this level contains the complete information of a welding operation. The welding operations are controlled by the adaptive control system and are operated with respective welding programs. A closer look at a small area reveals that the welding programs of the operations are arranged with a specific order (Fig. 5b and d), defined by the order of car part types manufactured in the production lines.

As the welding process goes on, the electrode wears. The wearing effect is quantified using the single feature WearCount (Fig. 7). When the WearCount reaches a fixed threshold pre-designed by the process experts, *dressing* is performed. A complete dressing-welding-dressing proce-

dure forms a *dress cycle*. Since the wearing effect repeats in each dress cycle, the consecutive dress cycles form the *dress cycle level*. A data sample on this level contains the complete information of all welding operations in a dress cycle. According to the domain expert, the strong periodicity of the Q-Value is caused by the wearing effect. The Q-Values in Fig. 5 begin with small values at each start of dress cycle, rises as the electrode cap wears, and ideally become stable at the end of dress cycles (Fig. 5a), but can also demonstrate complex trends under other conditions (Fig. 5c). After certain dress cycles, the electrode cap is changed. The behaviour of Q-Value is then influenced by the new electrode. All operations welded by one electrode cap constitute an *electrode cycle*. The consecutive electrode cycles comprise the *electrode cycle level*.

Fig. 6 Sequential structure of a simplified RSW production line, where multiple types of car chassis parts, each with a certain amount of welding spots with specified types, go through a sequence of welding machines. Each spot type is welded with one pre-defined welding program by one machine in a fixed order
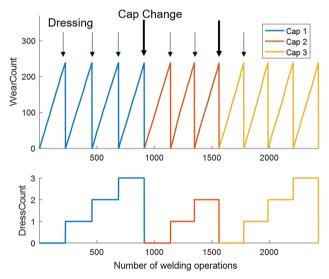

Fig. 7 Example of temporal structures in the RSW data quantified by WearCount, DressCount, and CapCount. WearCount is a counter on the welding operation level. It increases by one after one welding operation, and is reset to zero after a dressing is performed, or the cap is changed. DressCount is a counter on dress cycle level. It increases by one after one dressing operation, and is reset to zero after the cap is changed

Until here, the time levels of a single welding machine are explained. If we step further to see the multiple machines organised in production lines, we see two typical structures of production lines. Figure 6 illustrates a simplified production line for RSW with a sequential structure. This organisation of welding machines constitute *the machine level*. Depending on the structures of production lines, the collected data points need to be organised with different temporal structures.

There exist other latent time-levels, including the level of car chassis parts, production batches of parts and machines, suppliers of the materials, components and machines. Since this information is normally not available in the welding data, these time-levels will *NOT* be addressed in this work.

Data in two formats

In the data collected from two production lines with a total of 27 welding machines, we have selected two representative datasets for this paper to demonstrate the results, including Welding Machine 1 (WM1) with 2 welding programs and 1998 welding operations, and Welding Machine 2 (WM2) [2] with 6 welding programs and 5839 welding operations. The data come with inhomogeneous formats, including welding protocols generated in real-time, feed-back curves databases, reference curve databases, and meta settings.

After collection and preparation, the inhomogeneous data are prepared in two formats for each single welding operation (Fig. 3b). We use the term *Data Tuple (DT)* (Mikut et al. 2006) to indicate a single data instance that contains all features fully describing the instance. In this work, data tuples are on the welding operation level, and is comprised of the following two types of features.

– *Time series (TS)*, including eight effective features.

  – Four *Reference Curves* prescribed by the welding programs of the adaptive control system, e.g.

---

[2] Note this is a different welding machine with much more complex data than the WM2 in Zhou et al. (2020a).
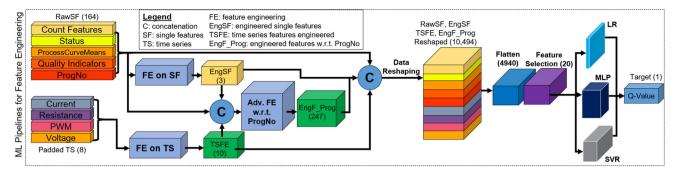
Fig. 8 Hierarchical feature extraction. RawSF and padded TS go through different feature engineering modules. The resulting features (EngSF and TSFE) are combined with RawSF and go through further advanced FE module with respect to ProgNo. All raw features (RawSF and Padded TS) and engineered features (EngSF, TSFE, EngF_Prog) are combined again and g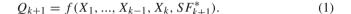o through data reshaping to accentuate short- time dependency. After that, features are flattened to be made suitable for classic ML methods. Feature selection reduces these features to a small amount (20). Three ML methods are studied for modelling. LR: linear regression, MLP: multi-layer perceptron, and SVR: support vector machine

reference current (I_{ref}), reference voltage (U_{ref}), reference resistance (R_{ref}), reference pulse width modulation (PWM_{ref}). The *reference curves* for a specific welding program usually remain identical unless they are manually changed.

– Four *Actual Process Curves*, which are actual measured process feedback curves, e.g. electric current (I), voltage (U), resistance (R), pulse width modulation (PWM).

– *Single features (SF)*, containing 164 effective features.

– Single features that describe the information of the temporal structures in data, including the aforementioned *WearCount*, *DressCount*, *CapCount*, and *Program Numbers (ProgNo)*, which are ordinal numbers of the welding programs.

– Other raw single features, examples including *Status*, describing the operating or control status of the welding operation, *ProcessCurveMeans*, which are the average values of the process curves and their welding stages calculated by the welding software system, and *QualityIndicators*, which are categorical or numerical values describing the quality of the welding operations, e.g. *HasSpatter* (boolean), *ProcessStabilityFactor (PSF)*, *Q-Value*.

Problem statement

The representation of the question defined in the "Introduction", Figure 1 is to *find a function* between the available information and the Q-Value of the next welding operation Q_{k+1}, shown in Eq. 1, where $X_1, ..., X_{k-1}, X_k$ include data tuples (single features and time series) of previous welding operations (from time step 1 to k) and known features of the next welding operation (SF^*_{k+1}, e.g. welding program)

$$Q_{k+1} = f(X_1, ..., X_{k-1}, X_k, SF^*_{k+1}). \tag{1}$$

In the following text, the subscripts $1, ..., k - 1, k, k + 1$ will be replaced by $1, ..., pre2, pre1, next1$ for a better understanding. Thus, Eq. 1 becomes Eq. 2.

$$Q_{next1} = f(X_1, ..., X_{pre2}, X_{pre1}, SF^*_{next1}) \tag{2}$$

Methods

This section elaborates on the machine learning methods studied for quality monitoring. This section first introduces the strategies to handle the temporal structures in the data and meaningfully combine features on different time levels ("Feature engineering" section), then describes feature selection ("Feature selection" section) and machine learning modelling ("Machine learningmodelling" section).

Feature engineering

Since the data exist in at least two levels: welding level and welding operation level, the feature engineering are also performed on different time levels, resulting in a ML pipeline of hierarchical feature extraction. We first take a glace over all levels, and dive in each module to understand the details.

Hierarchical feature extraction. To address the issue of temporal structures, feature extraction is performed on different time levels (Fig. 8). Features are first extracted from the padded time series on the welding time level. These extracted features (TSFE) can be seen as vectors containing compressed information from the time series. After feature extraction, the TSFE become features on the welding operation level, since they do not change along the welding time, but change for different welding operations. The TSFE
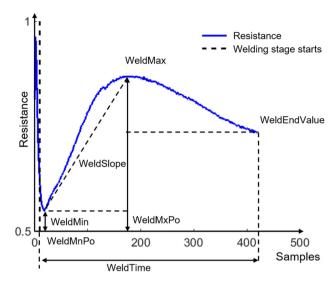
Fig. 9 Resistance curve as example to explain time series feature engineered (TSFE, partially shown). Resistance curve is the most complicated process curve and often deemed as the most informative time series feature (Cho and Rhee 2000; El-Banna et al. 2008; El Ouafi et al. 2010; Lee et al. 2001)

can be combined with the single features. The consecutive combined features again form time series on the welding operation level. The combined features can go through further feature extraction modules. This hierarchical feature extraction can continue on further time levels, depending on the desired granularity of time levels. In this work, feature extraction is performed on the first two time levels. After that, the extracted features will be reshaped on the welding operation level and then be used for ML modelling.

FE on the welding time level. After considering strategies proposed in the literature (Junno et al. 2004; Park and Cho 2004; Zhang et al. 2015; Lee et al. 2001; El-Banna et al. 2008; Yu 2015) and engineering knowledge, this work extract statistic and geometrical features to aggregate the information of time series (on the welding level) to time series features engineered (TSFE) on the welding operation level.

First, all eight time series are synchronised to the same time point of welding-start. The pre-welding-stage where nothing actually happens is chopped. Then these features are padded to the maximum length according to physical meaning. For example, current is padded with zero because after welding current becomes zero, while resistance is padded with the last value because resistance does not disappear after welding. Then, these features are extracted (Figure 9): length (WeldTime) (Panchakshari and Kadam 2013), maximum (WeldMax) (Cho and Rhee 2000), minimum (WeldMin) (El-Banna et al. 2008), maximum position (WeldMxPo) (Lee et al. 2001), minimum position (WeldMnPo), slope (WeldSlope) (Cho and Rhee 2002) ($slope = (max - min)/(mxpo - mnpo)$), mean (WeldMean) (El-Banna et al. 2008), median

(WeldMedian), standard deviation (WeldStd) (El-Banna et al. 2008), and end value (WeldEndValue). We can see these extracted statistic features can characterise the time series. Other time series, e.g. current curves and pulse width modulation curves, are much simpler than the resistance curve and can also be described by these features.

FE on the welding operation level. Strategies for feature engineering on single features are designed based on the meaning of features in domain knowledge, changing the representations of the raw features to facilitate machine learning modelling. Denoted as *Engineered Single Features (EngSF)*, 3 new features are generated, listed below.

- *WearDiff* is calculated as the difference between WearCount of two consecutive data tuples, characterising the degree of change of wearing effect. The value is normally ONE if the data is continuous; if some data tuples are missing, the value will be other numbers that correctly describe the wearing effect; and the value will be a large negative value after each fresh dressing.
- *NewDress* will be ONE after each dressing, and ZERO for other welding operations.
- *NewCap* will be ONE after each Cap Change, and ZERO for other welding operations.

Note that before the next welding operation happens, the WearCount, DressCount, and CapCount of the next operation are already known, since they are artificially designed incremental features. The EngSF based on them are therefore also known. These features corresponding to the next welding operation can therefore be used for predicting the Q-Value of the next welding spot.

According to the welding expert and observing from Fig. 5b, Q-Values with different welding programs have different behaviours, but this work does not consider Program Number (ProgNo) as a good feature for machine learning modelling. The same value of ProgNo would have different meanings for different welding machines in case of using raw values of ProgNo for modelling, and the number of features may change in case of *One-Hot Encoding*. Therefore, this work creates another type of features to incorporate the information of ProgNo implicitly, avoiding using the feature ProgNo (Fig. 11). Firstly, all single features that form time series on the welding operation level are decomposed to sub-time series, each only belonging to one ProgNo. Secondly, the aforementioned EngSF are extracted separately from each sub-time series. We give a name to this group of features: *Engineered Single Features considering ProgNo, (EngSF_Prog)*. Concretely, *WearDiff_Prog* is calculated as the difference between consecutive WearCounts that belong to the same ProgNo. *NewDress_Prog*, *NewDress_Prog*, and *NewCap_Prog* are created similarly.

Moreover, the following features are also created to implicitly incorporate the information of welding program.

- *RawSF_Prog* indicates the features generated by decomposing the *raw single features* of the data points belonging to the same ProgNo.
- *TSFE_Prog* indicates the features generated by decomposing the *time series features engineered* of the data points belonging to the same ProgNo.

EngSF_Prog, *RawSF_Prog*, and *TSFE_Prog* are grouped under the name *Engineered Features considering ProgNo (EngF_Prog)*.

Before feeding extracted features on the desired time level into machine learning algorithms, these features need to be reshaped to form small time snippets of a certain look-back length l.
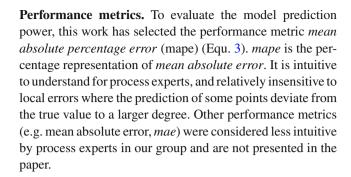
Feature selection

The number of features grow enormous because of feature engineering. From each of the 8 time series, 10 features are extracted, resulting in 80 features. From single features describing temporal structures, 3 new features are generated. There are 164 raw single features. After reshaping l previous operations, then flatting to table, the number of engineered features can grow to more than 2000. For example, suppose $l = 10$, the number of engineered features: $(164 + 3 + 80) \times 10 = 2470$. When adding the EngF_Prog, the number of features even doubles: $(164 + 3 + 80) \times 2 = 494$, $494 \times 10 = 4940$ (Fig. 8).

This work suggests using *step-wise forward feature selection* (Mikut et al. 2006) to keep the number of features for modelling small, for two purposes: 1) to retain the prediction power of ML models, especially in case where number of data points is less than number of features; 2) to attain the transparency of the ML models. After considering the time cost of feature selection, for linear regression models, this work applies a wrapper method to select the features. For MLP and SVR, a pre-selection with linear regression is performed.

Machine learning modelling

Three ML methods. Three types of classic machine learning methods are tested in this work. Linear Regression (LR) is selected as the representative algorithm of classic ML methods and extensively studied. Two non-linear methods, Multi-Layer Perceptrons with one hidden layer (MLP) and Support Vector Regression (SVR) are studied to see if non-linear methods can improve the performance further.

Performance metrics. To evaluate the model prediction power, this work has selected the performance metric *mean absolute percentage error* (mape) (Equ. 3). *mape* is the percentage representation of *mean absolute error*. It is intuitive to understand for process experts, and relatively insensitive to local errors where the prediction of some points deviate from the true value to a larger degree. Other performance metrics (e.g. mean absolute error, *mae*) were considered less intuitive by process experts in our group and are not presented in the paper.

$$mape = \frac{1}{N} \sum_{n=1}^{N} \left(|\frac{y_n - \hat{y}_n}{y_n}| \right) \times 100\%. \qquad (3)$$

Experiment settings

We now explain the data splitting strategies the benchmark feature engineering settings and finally the modelling.

Data splitting according to the temporal structures

Data splitting also needs to take the temporal structures of data in consideration. The splitting point should be chosen at complete units of some time levels. According to process experts, the deployment scenario will only be to test the developed machine learning methods on complete dress cycles in the future. It is therefore more meaningful to split the data also in this way. In this work, the data is split to training, validation, and test in a 0.8 : 0.1 : 0.1 ratio, rounded to complete dress cycles, illustrated in Fig. 10 for both welding machines using the Q-Value. It is also important to note that the validation data and test data should contain at least one complete dress cycle to ensure they cover the wearing effect through a full dress cycle. Details see Table 2.

Benchmarks

Three benchmarks are designed with process experts using intuition and domain knowledge, to provide baselines for evaluation of the performance of the ML models.

- *Benchmark 1,* is a simple estimation to predict the next Q-Value as equal to the previous Q-Value: $\hat{Q}_{next1} = Q_{pre1}$.

Table 2 Details of datasets

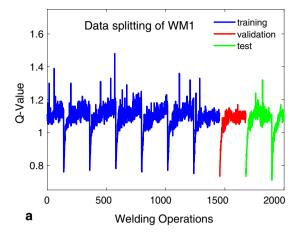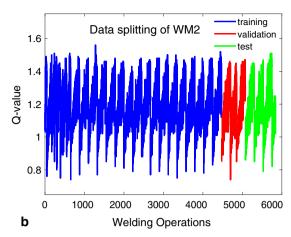| Welding Machine	#Training data	#Validation data	#Test data
Welding Machine 1	1456	219	323
Welding Machine 2	4474	602	763

Fig. 10 Example of data splitting rounded to complete dress cycles. Note the data of Welding Machine 2 is much more complicated than that of Welding Machine 1. Especially at the beginning cycles, compli-

cated dressing operations were performed. This may be caused by the change of production arrangement (Figure 5)
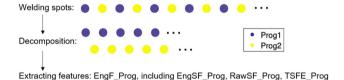

Fig. 11 Example for generating EngF_Prog on data of Welding Machine 1. Each dot indicates a welding spot and its data. Purple dots belong to Welding Program 1 and yellow dots belong to Welding Program 2. All single features that form time series on the welding operation level are decomposed to sub-time series, each only belonging to one ProgNo. The Engineered Features considering ProgNo (EngF_Prog) are extracted separately from each sub-time series

Table 3 Performance of benchmarks on *test set* evaluated using mape

Benchmarks		WM 1	WM 2
Benchmark 1	$\hat{Q}_{next1} = Q_{pre1}$	3.19%	10.41%
Benchmark 2	Average over WearCount	2.70%	5.13%
Benchmark 3	$\hat{Q}_{next1} = Q_{pre1\_Prog}$	2.48%	5.03%

four settings of features to study whether and to which degree feature engineering on the two types of features can increase model prediction power.

– *Benchmark 2, Average over WearCount* assumes the behaviour of Q-Value of the same welding program across all dress cycles should be nominally identical. The Q-Values with $WearCount = i$ in any dress cycle are therefore calculated as the average value of all Q-Values whose $WearCount = i$ and $ProgNo = P$, across all dress cycles in training data:

$$\hat{Q}_{next1|WearCount=i} =$$
$$mean(\{Q_{WearCount=i, ProgNo=P}, |Q \in TrainingSet\})$$

– *Benchmark 3* is a slight adaptation based on the Benchmark 1, but taking more domain knowledge into consideration. Thus: $\hat{Q}_{next1} = Q_{pre1\_Prog}$.

We have calculated benchmarks on two welding machines and report this in Table 3.

Four settings of feature engineering

Feature engineering can be performed on two types of features, time series features and single features (also time series in the time level of welding operations). We have designed

– *Setting 0*, no feature engineering. Only the *raw single features* will be used in machine learning modelling. Notice that the *ProcessCurveMeans* in the *raw single features* already provide some information of the time series. A total of 1640 (164×10) features are generated before feature selection.
– *Setting 1*, only performing feature engineering on time series. The resulting time series features engineered (TSFE) will be combined with raw single features (RawSF) in machine learning modelling. The TSFE serve as a supplement to the *ProcessCurveMeans*. A total of 2440 (($164 + 8 \times 10) \times 10$) features are generated before feature selection.
– *Setting 2*, performing feature engineering on time series and single features. The time series features engineered (TSFE), raw single features (RawSF) and engineered single features (EngSF) will be combined and used in machine learning modelling. A total of 2470 (($164 + 3 + 8 \times 10) \times 10$) features are generated before feature selection.

– *Setting 3*, performing a further step of feature engineering on time series and single features. The time series features engineered (TSFE), raw single features (RawSF), engineered single features (EngSF), and EngF_Prog (RawSF_Prog, TSFE_Prog and EngSF_Prog) will be combined and used in machine learning modelling. A total of 4940 ($(164 + 3 + 8 \times 10) \times 2 \times 10$) features are generated before feature selection.

ML model training and hyper-parameter selection

We trained the ML models on the training set and selected the hyper-parameters based on the performance evaluated on the validation set.

Two hyper-parameters are to be selected for linear regression, the *number of selected features* (ω) and *look-back length* (l). These two hyper-parameters are similar in the sense, that as they increase, the data amount provided to the ML model will increase and therefore the model performance evaluated on the training set should always increase. In industrial application, it is desired to find suitable hyper-parameters that provide relatively good performance, avoid overfitting, and ideally make the model insensitive to the hyper-parameters.

We performed a limited grid search to find the hyper-parameters that is to fix the first hyper-parameter and vary the second one; then select the second hyper-parameter in an area where the model performance is good and insensitive to the hyper-parameter; then fix the second hyper-parameter, vary the first one, and select the first one.

To make a fair comparison, we want to make the data amount delivered to the model the same, so that the performance difference is indeed caused by the quality of features, not the data amount. We therefore unify the hyper-parameters in the four models trained on the four feature settings. After a series of experiments we selected a *number of selected features* of 20 and a *look-back length* of 10 for the four feature settings. Figure 12a and b illustrate the model performance evaluated on validation set of Welding Machine 1. The models become insensitive to the hyper-parameters after more than approximately 20 features are selected and when the length of look-back window is longer than about 10. A further reason that 20 is selected is that we want to limit the *number of selected features* to retain model transparency, which is very desirable from the view of process experts. The performance of Setting 0 and Setting 1 is more sensitive to hyper-parameters than that of Setting 2 and Setting 3.

The same feature sets determined by LR are tested on MLP and SVR. The reasons are of two-fold: (1) the features selected by LR should already cover more than necessary information for a successful prediction of the next Q-Value. This hypothesis is confirmed by extra experiments of feature selection; (2) feature selection with MLP and SVR takes more time than LR, which makes it less desirable for the quick-adaptation of the methods to new datasets and industry application scenarios.

For MLP and SVR, other hyper-parameters are selected using limited grid search. MLP has two hyper-parameters, number of neurons in the hidden layer and the activation function. SVR has two or three hyper-parameters, kernel type, regularisation factor, and degree in case of polynomial kernel types (Table 4).

Results and discussion

This section presents the experiment results and discuss the performance of ML analysis. Four feature settings and three ML methods (LR, MLP, SVR) give 12 models for each dataset and 24 models in total.

After the hyper-parameters are determined, we trained the models again with the selected hyper-parameters on the combined set of training data and validation data, and tested the models on the respective test sets. The results of linear regression (LR) models trained with these four feature settings on dataset of Welding Machine 1 and Welding Machine 2 are presented in Table 5. The performance of the four settings are compared to Setting 0 and the best benchmark (Benchmark 3). Percentage improvements are calculated.

Results and discussion of linear regression

Firstly, from Table 5 one can conclude the model performance increases as the degree of feature engineering increases. We observe the effect of the features derived from domain knowledge that they help to improve the model performance by comparing the performance of Setting 0 to 3 on WM1 and WM2. The improvement becomes significant when we compare Setting 3 to Setting 0 (28.09% for WM1 and 29.31% for WM2), where Setting 3 has the most advanced features derived from domain knowledge. The domain knowledge derived features also reveal more insights from the engineering perspective, and help to make the model insensitive to hyper-parameters. These are discussed extensively in the "Interpreting ML results for engineering insights" section.

The performance of Setting 0 and Setting 1 on dataset of Welding Machine 1 shows deterioration compared to Benchmark 3. Observing from Figure 5a we can see the Q-Value behaviour of Welding Machine 1 is relatively stable, Benchmark 3 therefore works very well. The performance of Setting 0 and Setting 1 on dataset of Welding Machine 2 shows some improvement, because the Q-Value behaviour of Welding Machine 2 (Figure 5d) is much more complicated.

The performance improvement of Setting 1 compared to Setting 0 is insignificant, but consistent (Fig. 12), which
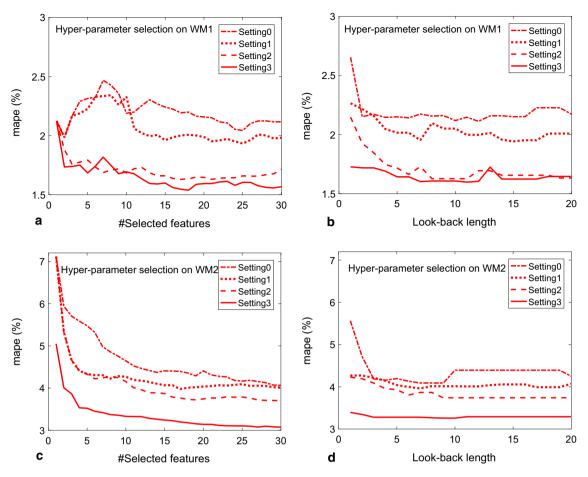
Fig. 12 (a) Performance of models evaluated on the validation set of Welding Machine 1. The models are trained on dataset of Welding Machine 1 with different *number of selected features* (ω) but a *fixed look-back length* (l) of 10. The performance changes less than 0.1% after approximately 20 features are selected. (b) Performance of models evaluated on the validation set of Welding Machine 1. The models are trained on dataset of Welding Machine 1 with different *look-back length* but a *fixed number of selected features* of 20. The performance changes less than 0.1% when the length of look-back window is about 10. To note in Setting 3, where the EngF_Prog is used, the effective look-back time step is *length of look-back window × number of welding program* (l × #Prog). (c) Performance of models evaluated on the validation set of Welding Machine 2. (d) Performance of models evaluated on the validation set of Welding Machine 2

Table 4 Hyper-parameter selection of ML methods. ω: #selected features, l: look-back length, λ: regularisation factor

Methods	Hyper-parameters	Selection method
LR	ω, l, features	Limited grid search
MLP	ω, l, features	Kept the same as LR
	#Neurons, activation function	Limited grid search
SVR	ω, l, features	Kept the same as LR
	Kernel type, λ, degree	Limited grid search

implies the features in the RawSF that contain time series information (the ProcessCurveMeans) already provide valuable information. A significant improvement begins with Setting 2, which indicates the feature engineering strategies on the welding operation level are meaningful. The performance difference between Setting 2 and Setting 3 analysed on dataset of Welding Machine 1 is rather small, but that on

dataset of Welding Machine 2 is evident. In Figure 12, we can see this difference is not a random effect but systematic.

A further inspection on the two datasets reveals that the welding programs of Welding Machine 1 are always arranged in a fixed interlaced order (Figure 5b), i.e. the ProgNo always repeat the same pattern: Prog1, Prog2, Prog1, Prog2, ... but the welding programs of Welding Machine 2 are not arranged in a fixed order. A change of production arrangement hap-

Table 5 Performance of the Linear Regression (LR) models evaluated on *test sets* of representative welding machines. Percentage improvements are calculated with respect to Setting 0 (the 4th column), and to the best benchmark (the 5th column). Note that the performance of the models trained with Setting 0 and 1 shows deterioration with respect to Benchmark 3. This phenomenon will be discussed in Sect. 6. mape: mean absolute percentage error. "Imprv. w.r.t.": "improvement with respect to"

Feature settings or benchmarks		Welding Machine 1 (WM1)			Welding Machine 2 (WM2)		
		mape (LR)	Imprv. w.r.t. Setting 0	Imprv. w.r.t. Benchmark3	mape (LR)	Imprv. w.r.t. Setting 0	Imprv. w.r.t. Benchmark3
Benchmark 3	$\hat{Q}_{next1} = Q_{pre1\_Prog}$	2.48%	7.12%	–	5.03%	−4.57%	–
Setting 0	RawSF	2.67%	–	−7.66%	4.81%	–	4.37%
Setting 1	RawSF, TSEF	2.55%	4.49%	−2.82%	4.55%	5.41%	9.54%
Setting 2	RawSF, TSEF, EngSF	2.10%	21.35%	15.32%	4.19%	12.89%	16.70%
Setting 3	RawSF, TSEF, EngSF, EngF_Prog	1.92%	28.09%	22.58%	3.40%	29.31%	32.41%

pened in the 618th welding operation (Fig. 5d), before which the production arrangement was comprised of three welding programs. After the 618th welding operation, three extra welding programs were added to the production arrangement, which is quite usual in manufacturing industry since the production arrangement could change at any time in agile manufacturing. This explains why Setting 3, in which the welding program information is specially handled, has a significant improvement on dataset of WM2, but less significant improvement on dataset of WM1.

Figure 13 illustrates the visualisation of the target values and estimated values. It can be seen that the Setting 0 and Setting 1 learned a general rising trend of the behaviour of Q-Values. However, the dynamics of behaviour of Q-Values are more complex than a simple rising trend. We can observe the Q-Value trend first rises then declines slightly, and remains stable at the end. These dynamics are better learned by more complicated feature settings.

Besides, although most of the Q-Values are predicted with small errors, there exist quite a few outliers that have an apparent different behaviour than the "normal" Q-Values. These outliers seem to be random, and cannot be explained with the trained models.

The complex dynamics of Q-Value behaviours implicate that the welding quality is not solely influenced by linear wearing effects. According to process experts, other influential factors include the statistic and stochastic variance caused by dressing, the chassis to be welded, etc.

Results and discussion of MLP and SVR

The results of MLP and SVR models are shown in Table 6. We have performed experiments of feature selection with these methods. The results show that the selected features are largely overlapping with those selected by LR. Thus, the MLP and SVR models are trained with the features determined by LR models and their hyper-parameters are selected using limited grid search.

Table 6 demonstrates that performance can indeed be improved by non-linear models. This means there exist non-linearity and interaction between the selected features, which cannot be described using the LR models.

The performance of MLP models is usually better compared to LR models (thus also better than the benchmarks). The improvement becomes quite significant for the complicated dataset (Welding Machine 2) with the highest degree of feature engineering (Setting 3). Conspicuous is the Setting 1 model performance is worse than that of Setting 0 for both welding machines. This indicates the time series features engineered may cause overfitting in some cases.

A closer look at results of LR and MLP models on the test set of Welding Machine 2 for two example welding programs with the Setting 3 is illustrated in Fig. 14. This reveals that
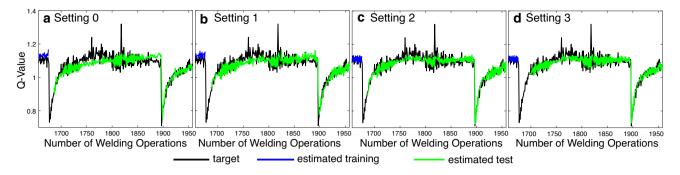
Fig. 13 Prediction results zoomed in on the *test set* area of Welding Machine 1, with the model of linear regression with the four settings of feature engineering, 20 features selected, and a look-back length of 10. The test set takes 10% of the data, and the training data area is not shown

although the different welding programs have very different behaviours for Q-values, both LR and MLP models are able to capture the different dynamics. From the figure the performance of the two models are not easy to differentiate. This indicates the importance of numerical performance metrics in Tables 5 and 6.

The performance of SVR models are on par with the LR models (and better than the benchmarks). Also conspicuous is that the performance of SVR models does not always improve as the degree of feature engineering increases. A further investigation of SVR models reveals that the performance of them fluctuates irregularly quite a few (up to 1.2% mape) as the regularisation factor changes, which means SVR models often fall into local optimums. Some SVR models perform well on training and validation sets but perform badly on test sets, which indicates a tendency of overfitting.

Generalisability over other welding machines

We have been testing our methods on a number of other welding machines (until the submission four extra welding machines) and other quality indicators (e.g. Process Stability Factor). The results have shown comparably promising results and similar diagrams as Figures 12, 13 and 14. Our evaluation has confirmed our hypothesis of generalisability of the proposed approaches over other welding machines and quality indicators.

Interpreting ML results for engineering insights

ML analysis in engineering should not only deliver results, but also provide insights that are helpful for engineering practice. Therefore, feature engineering has the advantage that the engineered features can efficiently represent the necessary information of the data and reveal influential factors and other insights. This section interprets benchmarks, selected

features, and visualisation of hyper-parameter selection to gain insights of ML analysis for engineering practice.

Interpretation of the benchmarks

Referring to Table 3, we can see that through simple consideration it is already possible to build very effective predictors for the next Q-Value. According to process experts, Benchmark 1, $\hat{Q}_{next1} = Q_{pre1}$ works to some degree because the behaviour of Q-Value should be normally stable and influenced by the wearing effect, while the wearing effect progresses gradually. The Q-Value therefore cannot change abruptly. Benchmark 2, Average over WearCount works well, because *dressing* should restore the change of surface condition caused by the wearing effect, and therefore also restores the behaviour of Q-Value. In another word, the behaviour of Q-Value should be nominally identical in different dress-cycles. Benchmark 3, $\hat{Q}_{next1} = Q_{pre1\_Prog}$ has the best performance because the welding operations of the same welding program provide more valuable information for prediction of future welding quality of that welding program than the other welding programs.

Extensive interpretation of selected features

Welding Machine 1.

Table 7 lists the 5 most important features with order of descending importance for Setting 0, Setting 1, Setting 2 and Setting 3 on dataset of Welding Machine 1, respectively.

Q-Value of the previous second welding spot is selected as the most important feature in three settings. Since the ProgNo always repeat the same pattern (see the "Results and discussion of linear regression" section): the Prog1, Prog2, Prog1, Prog2, ... The feature RawSF_Q_pre2 is therefore equal to EngF_Prog_Q_pre1. EngF_Prog_Q_pre1 is namely the Q-Value of the previous spot welded with the same welding program as the next welding spot (This feature is identical to Benchmark 3). This means the quality of welding usually does not have abrupt changes.

Table 6 Performance and hyper-parameters of the Multi-Layer Perceptrons (MLP) and Support Vector Regression (SVR) models evaluated on *test sets* of representative welding machines. Radial-Basis-Function has always been selected as the kernel types of SVR, and is therefore not listed in the table. mape: mean absolute percentage error, Act. function: activation function, λ: regularisation factor

Feature settings	Welding Machine 1 (WM1)					Welding Machine 2 (WM2)				
	mape (MLP)	Act. function	#Neurons	mape (SVR)	λ	mape (MLP)	Act. function	#Neurons	mape (SVR)	λ
Setting 0	2.27%	tanh	5	2.25%	8.17	3.87%	tanh	50	4.02%	19.02
Setting 1	2.42%	tanh	25	1.94%	10.19	4.07%	tanh	47	4.49%	21.29
Setting 2	2.03%	tanh	50	2.08%	6.15	3.57%	logistic	60	4.32%	14.79
Setting 3	1.92%	tanh	53	1.87%	8.17	2.97%	tanh	44	3.34%	16.00

Moreover, the features RawSF_WearCount_next1, EngSF_NewDress_next1, EngF_Prog_WearDiff_next1 are selected, which means the wearing effect has strong influence on the welding quality, and therefore quality prediction should use features characterising the wearing effect.

The features RawSF_I2_Mean_pre1, TSFE_R_WeldStd_pre1, TSFE_I_WeldMin_pre3, RawSF_I_Mean_pre3 are selected, which means the time series features extracted from the welding stage indeed contain some information for predicting the next spot quality. Note these features are of previous first or third spots, which are not welded with the same ProgNo as the next spot. This indicates that the information provided by historical spots that are not welded with the same ProgNo as the next spot are also important. The reason may be that the wearing effect taking place during these temporally adjacent welding operations have influence on the next welding quality.

No time series features extracted from the initial stage are selected as the most important features. This indicates the initial stage may be less important for the welding quality, which is also reasonable considering the time of the initial stage is short and the current is relatively too small to exerting effect.

In Setting 0 and 1, the selection of features of relative early operations, like RawSF_Q_pre9, TSFE_R_WeldStd_pre8, is questionable, because their influence should not be greater than temporally more adjacent features. In Setting 2 und 3 we can see these questionable features are no longer selected.
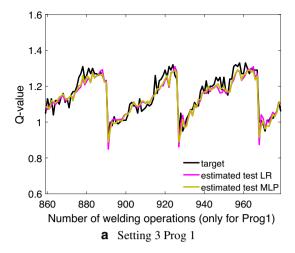
Welding Machine 2.

As for Welding Machine 2 (Table 8), the selected features are different.

The most obvious difference is that RawSF_Q_pre2 is no longer selected as the most important feature. As mentioned in see the "Results and discussion of linear regression" section, a change of arrangement of welding programs happened in the 618th welding operation. For the same reason, no EngSF is selected in the most important features, since EngSFs do not incorporate information of welding programs. Although the feature RawSF_WearCount_next1 can also describe the dependency of Q-Value on wearing effect to some degree, as it is selected through Setting 0 to 2, the performance of models trained on Setting 3 demonstrates a significant improvement (Table 5). This protrudes the advantage of the feature EngF_Prog_Q_pre1 in Setting 3.

Many TSFEs extracted from reference curves are selected as more important than those from actual process curves. Reference curves are prescribed by the welding programs, and are therefore always identical for a specific program. This implicates that the next spot quality is also dependent on the welding programs performed on the previous spots, rather than the corresponding actual process curves. This
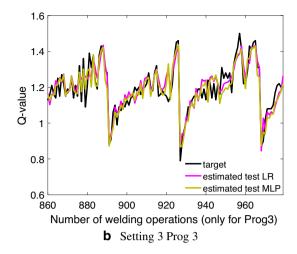
Fig. 14 Examples of prediction results on *test set* of Welding Machine 2 illustrated for two welding programs, with the model of MLP with the Setting 3 of feature engineering, 20 features selected, and a look-back length of 10

phenomenon is not evident for Welding Machine 1, since the welding program arrangement of which is fixed.

Similar to the case of Welding Machine 1, feature engineering avoids selection of questionable features such as RawSF_Power_Mean_pre10, which is far away in terms of temporal influence and should not be included, if considered from the view of engineering know-how.

Interpretation of hyper-parameter selection

The results of selection of hyper-parameters are illustrated in Fig. 12. The trend that the performance of the models increases as the look-back length increases (Figure 12b) imply that the hypothesis holds, that there exist temporal dependencies between welding spots. That is to say, the Q-Value of the next welding spot is indeed dependent on the previous welding operations.

We observe that when #Selected features $\omega > 15$, look-back length $l > 10$, the model performance does not change significantly for WM1 (Fig. 12a and b). The same phenomenon is revealed for Welding Machine 2 (Fig. 12c and d) that the model performance becomes insensitive in the areas of the selected hyper-parameters. The same hyper-parameters are therefore selected ($\omega = 20, l = 10$) for datasets of Welding Machine 2.

The performance of Setting 0 and Setting 1 is more sensitive to hyper-parameters than that of Setting 2 and Setting 3. This indicates again that the features derived from domain knowledge also make the model more insensitive the hyper-parameters.

Conclusion and outlook

Conclusion. This work firstly reveals characteristics of welding data that are little discussed in the literature, especially the hierarchical temporal structures in the production data that are important for quality prediction. Then machine learning (ML) approaches to deal with the hierarchical temporal structures and feature engineering with deep consideration of engineering knowledge are introduced. After that, the ML approaches are evaluated on two industrial production datasets to test the generalisability. A great advantage of our solution that is very desired in industry is that the ML approaches are insensitive to the hyper-parameters, number of features and lookback length. Our results demonstrate that the prediction power of even the most simplistic modelling method, linear regression, can be substantially enhanced through cunning design of engineered features. Furthermore, the transparency of feature engineering allows the interpretation of ML results to gain engineering insights. On the contrary, a blind training of ML models would select questionable and less robust features. The extensive interpretation of ML results enable a better understanding of the meaning of features, temporal dependency, crossing the board of two disciplines: ML and engineering and making them more deeply intertwined.

Outlook. We have been doing or plan to do investigation in the following directions which we believe are valuable from both industrial and academic points of view:

– Testing the proposed approach on more datasets to further verify the generalisability.
– Exploring other feature extraction strategies, especially feature learning (deep learning), to compare the (dis)advantages to feature engineering.

Table 7 The most important 5 features selected from feature settings in the analysis of dataset of Welding Machine 1, listed in ranking of descending importance. Note the score is evaluated on a multivariate basis, i.e. the importance of ω-th feature is the combined score of the 1st to ω-th feature (Sect. 4.2). Correlation coefficient between the model estimation and target value is chosen as the feature score for an intuitive comparison. The prefixes RawSF, TSFE, EngSF, EngF_Prog indicate the feature source, the suffixes indicate the time stamp of the features, and the stems indicate the physical meanings, e.g. Q for Q-Value, R for resistance, I for current, I2_Mean for ProcessCurveMeans of current, I_WeldMin for the minimum extracted from the *welding stage*

#	Setting 0	Score	Setting 1	Score	Setting 2	Score	Setting 3	Score
1	RawSF_Q_pre2	0.72	RawSF_Q_pre2	0.72	RawSF_Q_pre2	0.72	EngF_Prog_Q_pre1	0.73
2	RawSF_I2_Mean_pre1	0.78	RawSF_I2_Mean_pre1	0.78	EngSF_NewDress_next1	0.80	EngF_Prog_WearDiff_next1	0.86
3	RawSF_WearCount_next1	0.79	RawSF_WearCount_next1	0.79	RawSF_I2_Mean_pre1	0.84	EngF_Prog_I2_Mean_pre1	0.87
4	RawSF_Q_pre9	0.81	TSFE_I_WeldMin_pre3	0.81	TSFE_R_WeldStd_pre1	0.86	TSFE_R_WeldStd_pre1	0.88
5	RawSF_Q_pre3	0.82	TSFE_R_WeldStd_pre8	0.82	RawSF_I_Mean_pre3	0.87	EngF_Prog_Q_pre2	0.88

Table 8 The most important 5 features selected from feature settings in the analysis of dataset of Welding Machine 2, listed in ranking of descending importance. Feature scores are similar to Table 7. The prefixes RawSF, TSFE, EngSF, EngF_Prog indicate the feature source, the suffixes indicate the time stamp of the features, and the stems indicate the physical meanings, e.g. Q for Q-Value, I2_Mean for ProcessCurveMeans of current, R for resistance, RefU_WeldEndValue for the end value extracted from the *welding stage* of the reference curve of voltage, I_WeldMin for the minimum extracted from the *welding stage*, RefPWM for the reference curve of the *Pulse Width Modulation*

#	Setting 0	Score	Setting 1	Score	Setting 2	Score	Setting 3	Score
1	RawSF_WearCount_next1	0.75	RawSF_WearCount_next1	0.75	RawSF_WearCount_next1	0.75	EngF_Prog_Q_pre1	0.79
2	RawSF_Time_pre2	0.81	TSFE_RefU_WeldEndValue_pre2	0.85	TSFE_RefU_WeldEndValue_pre2	0.85	EngF_Prog_WearDiff_next1	0.91
3	RawSF_I2_Mean_pre2	0.85	TSFE_RefR_WeldMax_pre1	0.88	TSFE_RefR_WeldMax_pre1	0.88	TSFE_RefR_WeldSlope_pre2	0.92
4	RawSF_R_Mean_pre1	0.86	TSFE_I_WeldMin_pre3	0.90	TSFE_I_WeldMin_pre3	0.90	RawSF_WearCount_next1	0.93
5	RawSF_Power_Mean_pre10	0.87	TSFE_R_WeldMax_pre1	0.90	TSFE_R_WeldMax_pre1	0.90	TSFE_RefPWM_WeldMean_pre4	0.93

– Investigating other ML methods (Zou et al. 2021; Feng et al. 2020), e.g. artificial neural networks, especially recurrent neural networks, which are suitable for processing data with temporal structures.

– Predicting the Q-Value of the next welding spot as a probabilistic distribution. Before the next welding actually happens, the next Q-Value can in fact not be deterministically predicted. The prediction of the next Q-Value in this work is actually a prediction of the mean value. A better way is probabilistic forecasting.

– Using the prediction results as a basis for process optimisation. After the next spot quality is predicted, there exist several possible measures to undertake, e.g. flexible dressing, adaptation of the reference curves, or switching to non-adaptive control mode.

– Modelling of domain knowledge in ontologies (Svetashova et al. 2020a, b; Zhou et al. 2020b, 2021a), knowledge graphs (Zhou et al. 2021b, c) and rule-base systems (Kharlamov et al. 2017a, b, 2019; Horrocks et al. 2016), etc. Domain knowledge will be more deeply integrated in data integration (Jiménez-Ruiz 2015; Pinkel et al. 2015, 2018; Kalayci et al. 2020; Kharlamov et al. 2016), data query, reasoning (Thinh et al. 2018; Ringsquandl et al. 2018) and automatic construction of ML pipelines.

References

Afshari, D., Sedighi, M., Karimi, M. R., & Barsoum, Z. (2014). Prediction of the nugget size in resistance spot welding with a combination of a finite-element analysis and an artificial neural network. *Materiali in Tehnologije, 48*(1), 33–38.

Alpaydin, E. (2009). *Introduction to Machine Learning*. Massachusetts: MIT Press.

Amiri, N., Farrahi, G., Kashyzadeh, K. R., & Chizari, M. (2020). Applications of ultrasonic testing and machine learning methods to predict the static and fatigue behavior of spot-welded joints. *Journal of Manufacturing Processes, 52*, 26–34.

Bartschat, A., Reischl, M., & Mikut, R. (2019). Data mining tools. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9*(4), e1309.

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: a review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*(8), 1798–1828.

Boersch, I., Füssel, U., Gresch, C., Großmann, C., & Hoffmann, B. (2016). Data Mining in Resistance Spot Welding. In: The International Journal of Advanced Manufacturing Technology (p. 1–15).

Chand, S., & Davis, J. (2010). What is smart manufacturing. *Time Magazine Wrapper, 7*, 28–33.

Cho, Y., & Rhee, S. (2000). New technology for measuring dynamic resistance and estimating strength in resistance spot welding. *Measurement Science and Technology, 11*(8), 1173.

Cho, Y., & Rhee, S. (2002). Primary circuit dynamic resistance monitoring and its application to quality estimation during resistance spot welding. *Welding Journal, 81*(6), 104–111.

Cho, Y., & Rhee, S. (2004). Quality estimation of resistance spot welding by using pattern recognition with neural networks. *IEEE Transactions on Instrumentation and Measurement, 53*(2), 330–334.

Dai, W., Li, D., Tang, D., Jiang, Q., Wang, D., Wang, H., & Peng, Y. (2021). Deep learning assisted vision inspection of resistance spot welds. *Journal of Manufacturing Processes, 62*, 262–274.

DVS. (2016). *Widerstandspunktschweißen von Stählen bis 3 mm Einzeldicke - Konstruktion und Berechnung*. Standard, Deutscher Verband für Schweißen und Verwandte Verfahren e: V., Düsseldorf, DE.

El-Banna, M., Filev, D., & Chinnam, R. B. (2008). Online qualitative nugget classification by using a linear vector quantization neural network for resistance spot welding. *The International Journal of Advanced Manufacturing Technology, 36*(3–4), 237–248.

El Ouafi, A., Bélanger, R., & Méthot, J. F. (2010). An on-line ANN-based approach for quality estimation in resistance spot welding. *Advanced Materials Research, 112*, 141–148.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine, 17*(3), 37–37.

Feng, W., Zhang, J., Dong, Y., Han, Y., Luan, H., Xu, Q., Yang, Q., Kharlamov, E., Tang, J. (2020). *Graph random neural networks for semi-supervised learning on Graphs*. NeurIPS

Gavidel, S. Z., Lu, S., & Rickli, J. L. (2019). Performance analysis and comparison of machine learning algorithms for predicting nugget width of resistance spot welding joints. *The International Journal of Advanced Manufacturing Technology, 105*(9), 3779–3796.

Guo, Z., Ye, S., Wang, Y., & Lin, C. (2017). Resistance Welding Spot Defect Detection With Convolutional Neural Networks. In: *International Conference on Computer Vision Systems*, (p. 169–174). Springer.

Haapalainen, E., Laurinen, P., Junno, H., Tuovinen, L., & Röning, J. (2005). Methods for Classifying Spot Welding Processes: A Comparative Study of Performance. In: *International Conference on Industrial. Engineering and Other Applications of Applied Intelligent Systems* (p. 412–421). Berlin/Heidelberg, Bari, Italy: Springer.

Haapalainen, E., Laurinen, P., Junno, H., Tuovinen, L., & Röning, J. (2008). Feature selection for identification of spot welding processes. In Informatics in Control Automation and Robotics, (p. 69–79). Springer, Berlin/Heidelberg.

Hamedi, M., Shariatpanahi, M., & Mansourzadeh, A. (2007). Optimizing spot welding parameters in a sheet metal assembly by neural networks and genetic algorithm. *Proceedings of the Institution of*

Mechanical Engineers, Part B: Journal of Engineering Manufacture, 221(7), 1175–1184.

Hou, W., Wei, Y., Guo, J., & Jin, Y., et al (2017). Automatic Detection of Welding Defects Using Deep Neural Network. In: *Journal of physics: Conference series*, vol. 933, (p. 012006). IOP Publishing.

Horrocks, I., Giese, M., Kharlamov, E., Waaler, A. (2016). Using semantic technology to Tame the data variety challenge. *IEEE Internet Computing, 20*(6), 62–66 (2016)

ISO. (2004). Resistance welding-procedures for determining the weldability lobe for resistance spot. *International materials reviews., 49*(2), 45–75.

ITU (2012). Recommendation ITU – T Y.2060: Overview of the Internet of Things. Tech. rep. In: *International Telecommunication Union*.

Jiménez-Ruiz, E., Kharlamov, E., Zheleznyakov, D., Horrocks, I., Pinkel, C., Skjaeveland, M. G., Thorstensen, E., Mora, J. (2015). BootOX: Practical mapping of RDBs to OWL 2. *ISWC (2)*, 2015, 113–132

Junno, H., Laurinen, P., Tuovinen, L., & Röning, J (2004). Studying the Quality of Resistance Spot Welding Joints Using Self-organising Maps. In: *Fourth International ICSC Symposium on Engineering of Intelligent Systems (EIS)*, (pp. 705–711).

Kagermann, H. (2013). Recommendations for implementing the strategic initiative INDUSTRIE 4.0. Tech. rep., acatech–National Academy of Science and Engineering, Frankfurt, DE.

Kagermann, H. (2015). Change through digitization – value creation in the age of industry 4.0. In Management of Permanent Change, (p. 23–45). Springer.

Kalayci, E. G., Grangel-González, I., Lösch, F., Xiao, G., ul Mehdi, A., Kharlamov, E., Calvanese, D. (2020). Semantic integration of Bosch manufacturing data using virtual knowledge graphs. *ISWC (2), 2020*, 464–481.

Kharlamov, E., Brandt, S., Jiménez-Ruiz, E., Kotidis, Y., Lamparter, S., Mailis, T., Neuenstadt, C., Özçep, Ö. L., Pinkel, C., Svingos, C., Zheleznyakov, D., Horrocks, I., Ioannidis, Y. E., Möller, R. (2016). Ontology-based integration of streaming and static relational data with optique. *SIGMOD Conference, 2016*, 2109-2112.

Kharlamov, E., Hovland, D., Skjaeveland, M. G., Bilidas, D., Jiménez-Ruiz, E. Xiao, G., Soylu, A., Lanti, D., Rezk, M., Zheleznyakov, D., Giese, M., Lie, H., Ioannidis, Y. E., Kotidis, Y., Koubarakis, M., & Waaler, A. (2017a). Ontology based data access in Statoil. *Journal of Web Semantics, 44*, 3–36.

Kharlamov, E., Mailis, T., Mehdi, G., Neuenstadt, C., Özçep, Ö. L., Roshchin, M., Solomakhina, N., Soylu, A., Svingos, C., Brandt, S., Giese, M., Ioannidis, Y. E., Lamparter, S., Möller, R., Kotidis, Y., Waaler, A. (2017b). Semantic access to streaming and static data at Siemens. *Journal of Web Semantics, 44*, 54–74.

Kharlamov, E., Kotidis, Y., Mailis, T. , Neuenstadt, C., Nikolaou, C., Özçep, Ö. L., Svingos, C., Zheleznyakov, D., Ioannidis, Y. E., Lamparter, S., Möller, R., & Waaler, A. (2019). An ontology-mediated analytics-aware approach to support monitoring and diagnostics of static and streaming data. *Journal of Web Semantics, 56*, 30–55.

Kim, K. Y., & Ahmed, F. (2018). Semantic weldability prediction with RSW quality dataset and knowledge construction. *Advanced Engineering Informatics, 38*, 41–53.

Koskimaki, H. J., Laurinen, P., Haapalainen, E., Tuovinen, L., & Roning, J. (2007). Application of the extended *k* nn method to resistance spot welding process identification and the benefits of process information. *IEEE Transactions on Industrial Electronics, 54*(5), 2823–2830.

LaCasse, P. M., Otieno, W., & Maturana, F. P. (2019). A survey of feature set reduction approaches for predictive analytics models in the connected manufacturing enterprise. *Applied Sciences, 9*(5), 843.

Laurinen, P., Junno, H., Tuovinen, L., & Röning, J. (2004). Studying the quality of resistance spot welding joints using Bayesian networks.

In: *Proceedings of the Artificial Intelligence and Applications*, (p. 705–711). Toulouse, France.

Lee, H. T., Wang, M., Maev, R., & Maeva, E. (2003). A study on using scanning acoustic microscopy and neural network techniques to evaluate the quality of resistance spot welding. *International Journal of Advanced Manufacturing Technology, 22*(9–10), 727–732.

Lee, S. R., Choo, Y. J., Lee, T. Y., Kim, M. H., & Choi, S. K. (2001). A quality assurance technique for resistance spot welding using a Neuro-Fuzzy algorithm. *Journal of Manufacturing Systems, 20*(5), 320–328.

Li, W., Hu, S. J., & Ni, J. (2000). On-line quality estimation in resistance spot welding. *Journal of Manufacturing Science and Engineering, 122*(3), 511–512.

Li, Y., Zhao, W., Xue, H., & Ding, J. (2012). Defect recognition of resistance spot welding based on artificial neural network. *Advances in Intelligent and Soft Computing, 115*(2), 423–430.

Martín, Ó., López, M., & Martin, F. (2007). Artificial neural networks for quality control by ultrasonic testing in resistance spot welding. *Journal of Materials Processing Technology, 183*(2–3), 226–233.

Martín, O., Tiedra, P. D., López, M., San-Juan, M., García, C., Martín, F., & Blanco, Y. (2009). Quality prediction of resistance spot welding joints of 304 austenitic stainless steel. *Materials and Design, 30*(1), 68–77.

Martín, T., De Tiedra, P., & López, M. (2010). Artificial neural networks for pitting potential prediction of resistance spot welding joints of AISI 304 Austenitic stainless steel. *Corrosion Science, 52*(7), 2397–2402.

Mikhaylov, D., Zhou, B., Kiedrowski, T., Mikut, R., & Lasagni, A. F. (2019a). High accuracy beam splitting using SLM combined with ML algorithms. *Optics and Lasers in Engineering, 121*, 227–235.

Mikhaylov, D., Zhou, B., Kiedrowski, T., Mikut, R., & Lasagni, A.F. (2019b). ML Aided Phase Retrieval Algorithm for Beam Splitting With an LCoS-SLM. In Laser Resonators, Microresonators, and Beam Control XXI, Vol. 10904, p. 109041M. In: *International Society for Optics and Photonics*, San Francisco

Mikut, R., Bartschat, A., Doneit, W., Ordiano, J.Á.G., Schott, B., Stegmaier, J., Waczowicz, S., & Reischl, M. (2017). The MATLAB Toolbox SciXMiner: User's Manual and Programmer's Guide. arXiv:1704.03298.

Mikut, R., Reischl, M., Burmeister, O., & Loose, T. (2006). Data mining in medical time series. *Biomedizinische Technik, 51*(5/6), 288–293.

Nomura, K., Fukushima, K., Matsumura, T., & Asai, S. (2021). Burn-through prediction and weld depth estimation by deep learning model monitoring the molten pool in gas metal arc welding with gap fluctuation. *Journal of Manufacturing Processes, 61*, 590–600.

NSF. (2010). *NSF 11–516: Cyber-Physical Systems (CPS)*. National Science Foundation, Virginia, USA: Tech. rep.

Panchakshari, A. S., & Kadam, M. S. (2013). Optimization of the process parameters in resistance spot welding using genetic algorithm. *International Journal of Multidisciplinary Science and Engineering, 4*(3), 1438–1442.

Park, Y. J., & Cho, H. (2004). Quality evaluation by classification of electrode force patterns in the resistance spot welding process using neural networks. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 218*(11), 1513–1524.

Pashazadeh, H., Gheisari, Y., & Hamedi, M. (2016). Statistical modeling and optimization of resistance spot welding process parameters using neural networks and multi-objective genetic algorithm. *Journal of Intelligent Manufacturing, 27*(3), 549–559.

Pereda, M., Santos, J., Martín, Ó., & Galán, J. (2015). Direct quality prediction in resistance spot welding process: sensitivity, specificity and predictive accuracy comparative analysis. *Science and Technology of Welding and Joining, 20*(8), 679–685.

Pinkel, C., Binnig, C., Jiménez-Ruiz, E., May, W., Ritze, D., Skjaeveland, M. G., Solimando, A., Kharlamov, E. (2015). RODI: A

benchmark for automatic mapping generation in relational-to-ontology data integration. *ESWC, 2015*, 21–37.

Pinkel, C., Binnig, C., Jiménez-Ruiz, E., Kharlamov, E., May, W., Nikolov, A., Sasa Bastinos, A., Skjaeveland, M. G., Solimando, A., Taheriyan, M., Heupel, C., Horrocks, I. (2018). RODI: Benchmarking relational-to-ontology mapping generation quality. *Semantic Web, 9*(1), 25–52.

Podržaj, P., Polajnar, I., Diaci, J., & Kariž, Z. (2004). Expulsion detection system for resistance spot welding based on a neural network. *Measurement Science and Technology, 15*(3), 592.

Ringsquandl, M., Kharlamov, E., Stepanova, D., Hildebrandt, M., Lamparter, S., Lepratti, R., Horrocks, I., Kröger, P. (2018). Event-enhanced learning for KG completion. *ESWC, 2018*, 541–559.

Rossum, G. (1995). Python reference manual. Tech. rep., CWI (Centre for Mathematics and Computer Science), Amsterdam, The Netherlands, The Netherlands.

Samuel, A. L. (2000). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development, 44*(12), 206–226.

Shevchik, S., Le-Quang, T., Meylan, B., Farahani, F. V., Olbinado, M. P., Rack, A., Masinelli, G., Leinenbach, C., & Wasmer, K. (2020). Supervised deep learning for real-time quality monitoring of laser welding with X-ray radiographic guidance. *Scientific reports, 10*(1), 1–12.

Sumesh, A., Rameshkumar, K., Mohandas, K., & Babu, R. S. (2015). Use of machine learning algorithms for weld quality monitoring using acoustic signature. *Procedia Computer Science, 50*, 316–322.

Summerville, C., Adams, D., Compston, P., & Doolan, M. (2017). Nugget diameter in resistance spot welding: a comparison between a dynamic resistance based approach and ultrasound C-Scan. *Procedia Engineering, 183*, 257–263.

Summerville, C. D. E., Adams, D., Compston, P., & Doolan, M. (2017). Process monitoring of resistance spot welding using the dynamic resistance signature. *Welding Journal., 11*, 403–412.

Sun, H., Yang, J., & Wang, L. (2017). Resistance spot welding quality identification with particle swarm optimization and a kernel extreme learning machine model. *The International Journal of Advanced Manufacturing Technology, 91*(5–8), 1879–1887.

Svetashova, Y., Zhou, B., Pychynski, T., Schmidt, S., Sure-Vetter, Y., Mikut, R., & Kharlamov, E. (2020a). Ontology-enhanced machine learning: A Bosch use case of welding quality monitoring. In: *ISWC*, 2020, pp. 531–550.

Svetashova, Y., Zhou, B. Schmid, S., Pychynski, T., & Kharlamov, E. (2020b). SemML: Reusable ML models for condition monitoring in discrete manufacturing. In: *ISWC (Demos/Industry)*, Vol. 2721, 2020, pp. 213–218.

Thinh Ho, V., Stepanova, D., Gad-Elrab, M. H., Kharlamov, E., Weikum, G. (2018). Rule learning from knowledge graphs guided by embedding models. *ISWC (1), 2018*, 72–90.

Tseng, H. Y. (2006). Welding parameters optimization for economic design using neural approximation and genetic algorithm. *International Journal of Advanced Manufacturing Technology, 27*(9–10), 897–901.

Wan, X., Wang, Y., & Zhao, D. (2016). Quality monitoring based on dynamic resistance and principal component analysis in small scale resistance spot welding process. *The International Journal of Advanced Manufacturing Technology, 86*(9–12), 3443–3451.

Wuest, T., Weimer, D., Irgens, C., & Thoben, K. D. (2016). Machine learning in manufacturing: advantages, challenges, and applications. *Production & Manufacturing Research, 4*(1), 23–45.

Yu, J. (2015). Quality estimation of resistance spot weld based on logistic regression analysis of welding power signal. *International Journal of Precision Engineering and Manufacturing, 16*(13), 2655–2663.

Zhang, H., Hou, Y., Zhang, J., Qi, X., & Wang, F. (2015). A new method for nondestructive quality evaluation of the resistance spot welding based on the radar chart method and the decision tree classifier. *International Journal of Advanced Manufacturing Technology, 78*(5–8), 841–851.

Zhang, H., Hou, Y., Zhao, J., Wang, L., Xi, T., & Li, Y. (2017). Automatic welding quality classification for the spot welding based on the hopfield associative memory neural network and chernoff face description of the electrode displacement signal features. *Mechanical Systems and Signal Processing, 85*, 1035–1043.

Zhang, Y., Chen, G., & Lin, Z. (2004). Study on Weld Quality Control of Resistance Spot Welding Using a Neuro-Fuzzy Algorithm. In: *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, (p. 544–550). Springer.

Zhang, Y., You, D., Gao, X., Zhang, N., & Gao, P. P. (2019). Welding defects detection based on deep learning with multiple optical sensors during disk laser welding of thick plates. *Journal of Manufacturing Systems, 51*, 87–94.

Zhang, Z., Wen, G., & Chen, S. (2019). Weld image deep learning-based on-line defects detection using convolutional neural networks for Al alloy in robotic arc welding. *Journal of Manufacturing Processes, 45*, 208–216.

Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019). Deep learning and its applications to machine health monitoring. *Mechanical Systems and Signal Processing, 115*, 213–237.

Zhou, B. (2021). Machine learning methods for product quality monitoring in electric resistance welding. Ph.D. thesis.

Zhou, B., Chioua, M., & Schlake, J.C. (2017). Practical methods for detecting and removing transient changes in univariate oscillatory time series. *IFAC-PapersOnLine, 50*(1), 7987–7992.

Zhou, B., Chioua, M., Bauer, M., Schlake, J. C., & Thornhill, N. F. (2019). Improving root cause analysis by detecting and removing transient changes in oscillatory time series with application to a 1, 3-Butadiene process. *Industrial and Engineering Chemistry Research, 58*(26), 11234–11250.

Zhou, B., Pychynski, T., Reischl, M., & Mikut, R. (2018). Comparison of machine learning approaches for time-series-based quality monitoring of resistance spot welding (RSW). *Archives of Data Science. Series A (Online First), 5*(1), 1–27.

Zhou, B., Svetashova, Y., Byeon, S., Pychynski, T., Mikut, R., & Kharlamov, E. (2020a). Predicting quality of automated welding with machine learning and semantics: a Bosch case study. In CIKM.

Zhou, B., Svetashova, Y., Pychynski, T., & Kharlamov, E. (2020b). Semantic ML for manufacturing monitoring at Bosch. In: ISWC (Demos/Industry), Vol. 2721, p. 398.

Zhou, B., Svetashova, Y., Gusmao, A., Soylu, A., Cheng, G., Mikut, R., Waaler, A. and & Kharlamov, E. (2021a). SemML: Facilitating development of ML models for condition monitoring with semantics. *Journal of Web Semantics, 71*, 100664.

Zhou, D., Zhou, B., Chen, J., Cheng, G., Kostylev, E. V., Kharlamov, E. (2021b). Towards ontology reshaping for KG generation with user-in-the-loop: Applied to Bosch Welding. In: *IJCKG*.

Zhou, B., Zhou, D., Chen, J., Svetashova, Y., Cheng, G., Kharlamov, E. (2021c). Scaling usability of ML analytics with knowledge graphs: Exemplified with a Bosch welding case. In: *IJCKG*.

Zou, X., Zheng, Q., Dong, Y., Guan, X., Kharlamov, E., Lu, J. Tang, J. (2021) TDGIA: Effective injection attacks on graph neural networks. *KDD, 2021*, 2461–2471