



Universiteit
Leiden
The Netherlands

Towards an adaptable quality monitoring process for self-piercing riveting

Noller, V.; Walther, U; Meschut, G.; Bäck, T.H.W.

Citation

Walther, U., Meschut, G., & Bäck, T. H. W. (2021). Towards an adaptable quality monitoring process for self-piercing riveting. *Iop Conference Series : Materials Science And Engineering*, 1157. doi:10.1088/1757-899X/1157/1/012001

Version: Publisher's Version

License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)

Downloaded from: <https://hdl.handle.net/1887/3277244>

Note: To cite this publication please use the final published version (if applicable).

PAPER • OPEN ACCESS

Towards an adaptable quality monitoring process for self-piercing riveting

To cite this article: Vanessa Noller *et al* 2021 *IOP Conf. Ser.: Mater. Sci. Eng.* **1157** 012001

View the [article online](#) for updates and enhancements.

You may also like

- [Numerical simulation and experimental validation of self-piercing riveting \(SPR\) of 6xxx aluminium alloys for automotive applications](#)
F Hönsch, J Domitner, C Sommitsch et al.
- [Numerical study on the ideal interference of aircraft riveting joints](#)
S Sun
- [Research on automatic assembly technology for final assembly of helicopter fuselage](#)
Q D Xiao, L A Zhang, L Gan et al.



The Electrochemical Society
Advancing solid state & electrochemical science & technology

242nd ECS Meeting

Oct 9 – 13, 2022 • Atlanta, GA, US

Abstract submission deadline: **April 8, 2022**

Connect. Engage. Champion. Empower. Accelerate.

MOVE SCIENCE FORWARD



Submit your abstract



Towards an adaptable quality monitoring process for self-piercing riveting

Vanessa Noller^{1*}, Ulrich Walther¹, Gerson Meschut², Thomas Bäck³

¹ Mercedes-Benz AG Sindelfingen, Germany

² University of Paderborn, Germany

³ Leiden University, Netherlands

E-mail: vanessa.noller@ieee.org

Abstract. In the automotive industry self-piercing riveting (SPR) is a standard joining technology, especially for vehicles with a high mix of materials. The applied quality control system and the underlying quality decisions have hardly changed in the recent years. The commonly used combination of process monitoring and rework strategies leads to a large number of false positives and a high amount of manual work. The collected process data is not used comprehensively and misses the potential to improve the SPR process. This article introduces a quality monitoring method for SPR to show a proof of concept by using machine learning to predict faulty joining points. Based on numerous technical and statistical features extracted from the force-displacement curve of SPR, the trained model categorizes the observations in two rivet head height (RHH) classes. An evolutionary algorithm is used for the feature selection and a random forest model for classification. The resulting accuracy scores up to 84.4% and shows the potential of the developed random forest model. The potential application of this approach in the context of serial body-shop production improves the prediction of joining quality and the process availability significantly. This enables the adaptable rework of non-critical joining points depending on the classified RHH.

1. Introduction

The general requirements for production are rising related to quality, universality, efficiency, and individualization to endure in the international market. Moreover, the demand for environment-friendly technologies rises and the discussion about climate-neutral production technology is getting increasingly important. A big factor to save energy is to reduce the weight of a vehicle. Thereby, the body-shop of a vehicle has great potential by representing 40% of the car's weight. At the same time, light-weight materials are getting introduced more and more. This leads to an increase of mixed joints and a rising demand for mechanical joining techniques [1]. In body-shop production, the quality monitoring system is essential to ensure production efficiency and competitiveness. The commonly applied quality monitoring method uses a process monitoring and rework concept to verify the production quality and has hardly changed in recent years. The settings of the used process monitoring method do not observe the complete tolerance limits of the used components. They are varying depending on the general conditions in production for technologically identical joining tasks. This leads to a large number of records incorrectly classified as not OK, called *false positives* and a high amount of manual rework. The collected semi-tubular self-piercing riveting (SSPR) process data is not used comprehensively,



misses the potential to improve the error detection and avoid unnecessary rework. The growing digitalization offers the opportunity to improve and optimize the production efficiency with data-driven digitalization technologies to keep pace with the rising cost pressure [1]. The integration of intelligence into a lower machine level provides an opportunity to optimize the manufacturing processes, monitor the machine condition, and deal with unexpected events [2]. The SSPR joints characteristic is represented by the force-displacement curve as a time-series dataset, the analysis of which is challenging. Therefore, a methodology is required to extract the necessary information and decide whether the joint is qualitatively acceptable or not.

The goal of this work is the development of a quality monitoring method for SSPR to show the feasibility by using machine learning to predict faulty joining points. We propose a method to use the process data more efficiently and complete the existing data with the prediction of two rivet head height (RHH) classes. Based on the recorded process data, the RHH is predicted and the joining point quality is determined. Firstly, the process data is extracted, preprocessed and the features are extracted. The second step is the integration of a sensor into production to record data and generate data labels. A machine learning model is trained with the resulting dataset by using an evolutionary algorithm for feature selection. The classification between two RHH classes is realized with a random forest classifier. For a proof of concept, we apply the classification performance evaluation of the generated model as a quality monitoring process.

The outline of this paper is as followed: Section 2 provides background information on SSPR and the machine learning algorithm used. In Section 3 the related work for machine learning in joining technology is summarized. Section 4 provides the modeling approach, while Section 5 discusses the results. In Section 6 the industrial application is discussed and Section 7 concludes the paper and outlines future work.

2. Background

2.1. Semi-Tubular Self-Piercing Riveting

The processes of joining by forming are standardized in DIN 8593-5 [3], in which riveting is one of the cold joining methods. Self-piercing riveting (SPR) is used to join two or more sheets of materials by locking them together to form a mechanical joint [4]. SPR with semi-tubular rivet (SSPR) is widely used in the automotive industry. The SSPR process is described generally by the following four steps (see Figure 1). Firstly, joining parts are fixed by the blank holder and the bottom die. Secondly, the rivet is moved by the riveting punch to pierce slightly the top layer of punch-side material and the punched scrap is taken up into the cavity of the semi-tubular rivet. Thirdly, the riveting punch keeps moving the rivet to pierce the die-side material, until the lower part of the rivet is split open and the mechanical interlock is formed. Finally, the punch and blank holder return to leave the material for the next process. [4]

To ensure the joint quality and the reliability of SSPR joints, cross-section analysis, and process monitoring are the main methods used [4]. Cross-section analysis measures the joint criteria that reflect the tightness and the strength of the joint. With the in-process analysis, process monitoring offers the possibility of mapping the quality of the joint during the ongoing production process in a reproducible quality process. The process quality is quantified by comparing the force-displacement curve with a reference [4]. The first investigation on the SSPR quality monitoring process was carried out by Budde et al. [5], and the force-displacement curve (see Figure 1) was proposed to monitor joint quality.

2.2. Data Analysis Method

In the CRISP-DM standard, all steps of data analysis from data acquisition to the application of methods are summarized [6]. These include data understanding, data preparation, modeling, deployment and evaluation. The analysis of large amounts of data is summarized under the term big data analysis and can be understood to include data mining, machine learning, and

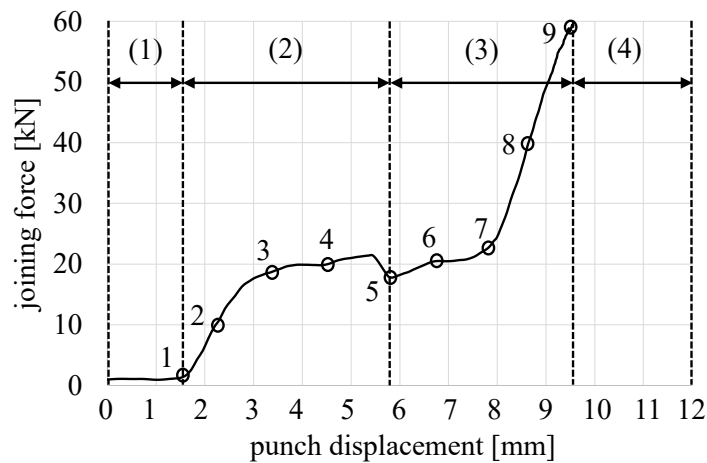


Figure 1. Force-displacement curve and process-steps (1) positioning, (2) applying of the rivet, (3) the shearing, spreading and foaming and (4) return of SSPR-process [4]. The markers (1, 2, ..., 9) represent prominent points in the joining process.

predictive analytics [7]. The analysis of data is a crucial step to generate insights from large amounts of data [8]. In the scope of this work, the field of machine learning is focused on. Machine learning is a method of data analysis that forms a subfield of artificial intelligence. It is roughly divided into supervised, semi-supervised and unsupervised learning, depending on the availability of input and output data [7]. If in supervised learning, the output is a class label (e.g. OK vs. not OK), the task is called classification. The following supervised learning methods are used for classification as part of this work.

Evolutionary algorithms are probabilistic optimization algorithms, modeling organic evolution through an iterative process of the variation operators 'mutation' and 'recombination', applied to the 'population' as a multiset of candidate solutions. They are evaluated by using the given objective function $f : D \rightarrow \mathbb{R}$ defining the optimization task ($f \rightarrow \min$ for example). Candidate solutions are selected for the next iteration or discarded from the population, based on their objective function value [9]. Depending on the domain D , one can define a simplifying, high-level difference between evolutionary strategies (ES), typically used when $D = \mathbb{R}^n$, and genetic algorithms (GA), typically used when $D = \{0, 1\}^n$ and in case of nominal variables, i.e., variables defined over a discrete, non-numerical domain without an order defined over the elements [10].

The *random forest* algorithm is an ensemble technique and combines different classifiers using aggregation techniques. It creates a compilation of several individual decision trees, whose results are averaged [8]. The decision-making basis in form of one or more rules can be derived from the resulting trees, making the decision technically understandable.

3. Related Work

The application of machine learning in joining technology is used in various areas. Here, we provide an overview of the currently existing applications. The work of Tan [11] shows the assessment of the load-bearing capacity of clinch connections using neural networks. The potential of machine learning in adhesive technology is examined in the work of Sierak [12]. Röske [13] investigates alternative sensor technologies and statistical analysis methods for improving error identification in screwing processes. Börsch et al. [14] shows a method to predict the welding spot diameter by monitoring process parameters. Karathanasopoulos et al. [15] show the usage of a neural network and finite element methods for predicting joining

characteristics. In particular, the application of data science approaches to improve the quality monitoring process shows great potential in joining technology. Kim et al. [1] present a literature survey of quality measures and prediction methods of self-piercing riveting. The implemented main approaches for SPR include a simulation-based improvement of quality monitoring as demonstrated in the work of Khrebtov [16] and Falk et al. [17]. In the scope of other joining techniques like spot welding, the examples of Pereira et al. [18], Bock et al. [19] and Li et al. [20] propose a data-driven model approach. Pereira et al. [18] propose a method for data-driven quality diagnostic and prognostics. Bock et al. [19] present in their work the first development for predicting the ultimate tensile force of friction riveting with a machine learning model. Li et al. [20] present an online weld-quality inspection method for ultrasonic composite welding based on artificial intelligence.

4. Modelling Approach

As presented in Section 2 and 3, the usage of data-driven modeling approaches in the context of joining techniques shows great potential for improving quality monitoring. It appears that there are hardly any approaches in the context of mechanical joining and especially SPR, which chases a data-driven model in the context of serial production [1]. Although a lot of data is recorded in serial body-shop production, the data quality typically is poorly suited for the use of a machine learning algorithm. The actual joining quality criteria are not detected and matched to the data records with a data label. The comprehensive extension of the process data with the automatically measured RHH offers an optimization possibility that enables additional categorization of the recorded datasets. The real quality status can be assigned for a large number of datasets and a data-driven model approach is made possible. The validation of the developed model is conducted with real-world process data.

4.1. Dataset and Preprocessing

The recorded raw data originates from a test facility of the Mercedes-Benz serial body-shop production in Sindelfingen. The first step based on CRISP-DM [6] process is the data preprocessing and the dataset analysis. The force-displacement curves (see Figure 1) are stored in a database and the appropriate parametrization is performed. The laser triangulation sensor, which was particularly integrated into the production for this use case, expands the process data by the contactless measured RHH. The measurement takes place after each joining process of serial production and the values are automatically stored in a cloud-based database. Figure 2 (a) shows the measurement setup, while (b) shows the distribution of the measured values, divided into three RHH areas. The result is a dataset consisting of more than 25,000 force-time series with two assigned RHH classes. The first part of the dataset was applied for the training process of the classification model, while the second part of the dataset was used for the validation of the developed model.

4.2. Feature Extraction

The following step of the CRISP-DM process includes the feature extraction from the force-time series. The marked points 2, 5, 7 and 9 shown in Figure 1 represent the characteristics of the curve and are used to extract 10 defined features. The extracted features are expanded using statistical characteristics. For this purpose, the force-time series is divided into 25 parts and 8 statistical features (maximum, minimum, mean, skew, kurtosis, standard deviation, average and variance) are calculated for each section. The Bravais-Pearson correlation analysis identified strongly correlating features and reduced the number of statistical features to 4 (average, standard deviation, skew, kurtosis) (see Table 1). The implementation is based on the python libraries SCIPY.STATS and NUMPY. The resulting extracted features are standardized

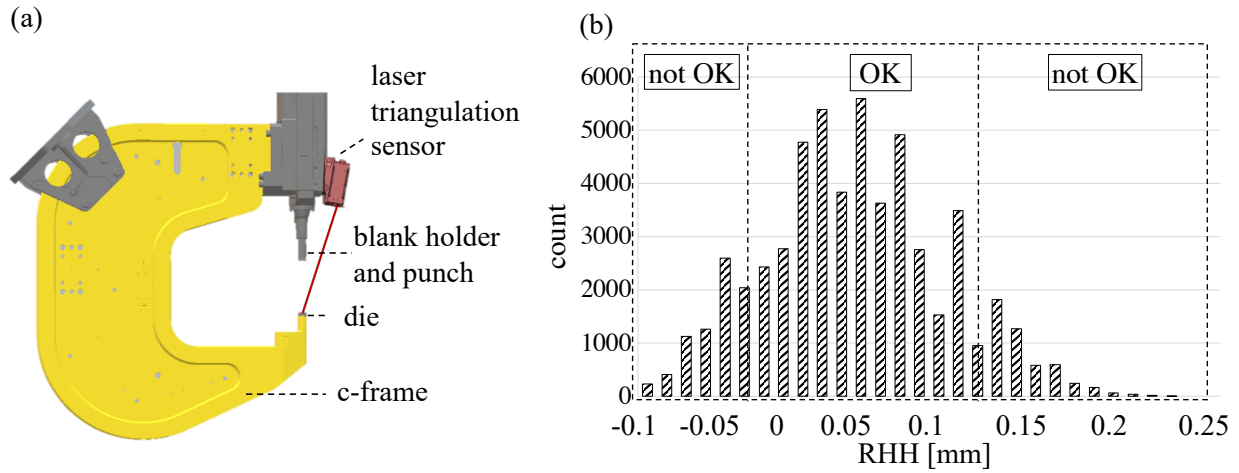


Figure 2. (a) Measurement setup of the RHH sensor element and (b) histogram of the measured RHH. Notice that the not OK class is split into too low and too high values.

by the `MINMAXSCALER` function of Sklearn. Incorrect data records that arise due to manual intervention or the tool check are removed.

Table 1. Definition of applied characteristics

Feature	Python-Library	Mathematical Description
Mean value	SCIPY.STATS [21]	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
Skew	SCIPY.STATS [21]	$\gamma = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^3$
Kurtosis	SCIPY.STATS [21]	$\omega = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^4$
Standard deviation	NUMPY [22]	$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$

4.3. Feature Selection and Classification

A genetic algorithm was selected for the implementation of the feature selection, as this is a suitable algorithm for optimizing parameters [23] and enables an increase in learning performance by reducing the data dimensionality [10]. The genetic algorithm selects the best combination of features by optimizing the classification quality. In our implementation, each individual consists of a set of selected features, represented by the feature names (i.e., nominal variables are used). The number of features (i.e., the dimensionality of the individuals) is varied between 5 and 25 (in steps of 5) across the different experiments that were performed, aiming at identifying the optimal number of features. The first step of the genetic algorithm then is the random generation of the initial population. The feature combinations are realized based on an array while the data is represented as a `DATAFRAME`. In the following steps, the feature combination is iteratively mutated (with four features per individual being randomly replaced by other features from the set of available features) and by applying recombination four times to the

parent population to generate a total of eight new solution candidates. For each combination of features, a classification model is trained and evaluated through cross-validation. The work presented in Section 3 shows very good results through the use of the decision tree and the random forest algorithm in joining technology, while they support the traceability of the decision. Due to its random sampling, the random forest algorithm reduces overfitting to the training data, and hence, shows an overall better performance [8]. Therefore, our fitness function is implemented by a random forest model for classification. The assessment of the best feature-model combination is determined by the F1-score. The genetic algorithm is optimized by varying the dimensionality of individuals (between 5 and 25) and the number of iterations n , with all other parameters being constant (and the parent population size $\mu = 10$). Iterative optimization was used as a technique for evaluation, while the F1-score of the corresponding model is used as the fitness score. The number of mutations (four) and number of individuals generated by recombination (eight) were defined by preliminary tests. For selection, the 10 best individuals of the union of parents ($\mu = 10$) and offspring ($\lambda = 12$) are deterministically selected (i.e., a $(10 + 12)$ selection is used). The classification model is implemented by a random forest model using the `SKLEARN.ENSEMBLE.RANDOMFORESTCLASSIFIER` in python. The model parameters `n_estimators`, `max_depth`, `max_features` and `criterion` are optimized by a grid search, as soon as the maximum number of iterations is reached.

5. Results and Discussion

The aim of developing the classification model is the quality prediction of the SSPR process to show the feasibility and technical possibilities through a proof of concept. As a first step, the classification model is optimized with a grid search. The resulting parameter configuration is applied to the classification model and the fitness function of the genetic algorithm is computed. In a second step, the feature selection is added and carried out until the maximum number of iterations is reached. The grid search of the parameters of the random forest model is carried out a second time with the resulting optimized feature combination. For academic validation, this developed classification model was evaluated using the created validation dataset. Figure 3 shows the resulting model quality for 10 independently performed model optimizations depending on the varied model parameters. The results of the respective parameter configurations are shown in form of a box plot diagram. The two diagrams at the left present the parameter variation of the genetic algorithm while the diagram to the right shows the most significant variation of the model parameters of the random forest classifier. The application of the genetic algorithm, as well as the grid search, shows a significant effect and improves the classification accuracy from 77% to 84% depending on the number of iterations. The variation of the dimension of each individual indicates a constant spreading and mean, while the variation with 15 features per feature combination shows slightly worse results. The maximum depth of the respective decision trees (`max_depth`) is varied between the values 10 and 50. The resulting box plot diagram shows the constant behavior in the scattering and the average classification quality. A maximum depth of 10 shows a different reduced value. The parameters `criterion` and `n_estimators` do not lead to any significant change or scattering behavior. When selecting the criterion through the grid search, 90% of the time the entropy is selected. The variation of the `max_features` parameter indicates a constant spread and a constant average performance for the auto, log2, and sqrt values. A reduction in the spread can be implemented by using any default value (none). The parameter configuration selected by the majority of grid search are `criterion: entropy`, `max_depth: 30`, `max_features: none`, and `n_estimators: 200`. It achieves a good classification score of 84.4%. In the production environment, many unquantifiable influences affect the measurement result. This includes, among other things, vibrations from neighboring technologies, and soiled or oiled material surfaces. The overall accuracy could be improved by leveraging more and better sensors with higher accuracy.

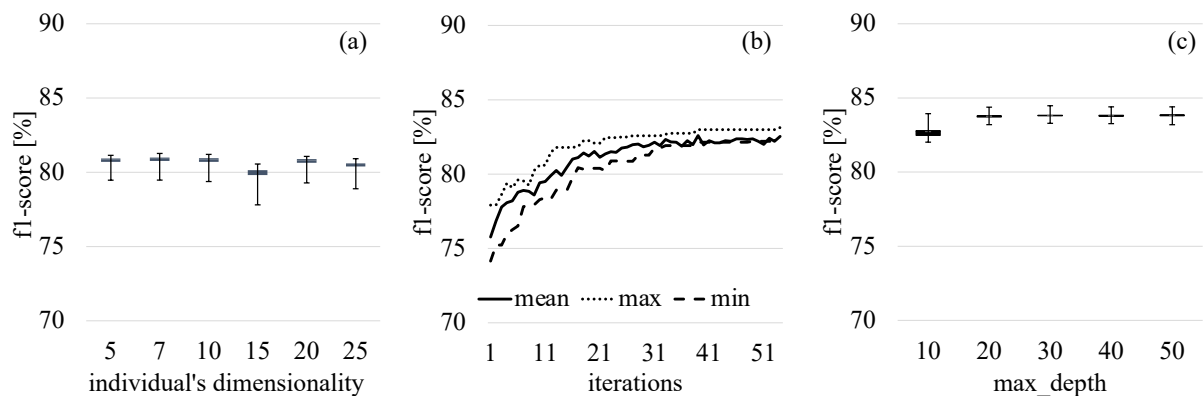


Figure 3. Influence of the variation of the model parameters (a) individuals dimension, (b) iterations and (c) max_depth for the 2 class RHH classification model.

6. Industrial Application

The developed classification model can be applied in a variety of production areas to improve the quality decision process. The distinction of the quality statement OK and not OK can be used in particular to develop a new quality control process in serial body-shop production. The recorded data can be utilized to generate additional knowledge about the joining process. The resulting information from process monitoring and accomplished rework is fed back and added to the existing dataset to improve the joining process. The validation of the developed model takes place as a prototype and demonstrates the possibilities. There are two resulting application scenarios, which enable an adaptive quality monitoring method to increase the process availability. In the first scenario, the automatic differentiation between errors and pseudo-errors is represented. The process monitoring is checked automatically and false-positive errors are detected. An increase in process understanding follows and the process availability is improved by the reduction of the pseudo-errors. The second scenario represents the further adaptive expansion of the SPR process quality control. For this purpose, reworking for non-safety-relevant joining points can be carried out depending on the classified RHH only at the necessary place. This means, that the RHH is classified after the quality assessment of the process monitoring. If the classified value is within the defined quality limits, reworking can be dispensed for non-safety-relevant joining points. Moreover, the dataset is marked for further analysis to enable the focus in data analysis on the relevant datasets. The downtimes in serial production are significantly reduced and the process availability increased. For a comprehensive application in serial production, it should be noted that the current classification accuracy of the model must be further improved in the future. The classification accuracy can be improved by increasing the data quality, especially of the data label. The comprehensive recording of process data over at least one year is necessary to include all seasonal fluctuations. In addition, the robustness of the model can be increased by clustering different technically similar joining tasks. In future, it is planned to implement the approach in the real-world production process.

7. Conclusion and Future Work

As part of this work, a data-driven machine learning model to classify the RHH was developed. A sensor element integrated into the serial body-shop production facilitates the generation of data labels. With the resulting dataset, a machine learning model consisting of a genetic algorithm and a random forest classifier is developed and used to predict the RHH. The promising result with a classification accuracy of 84.4% shows the feasibility of this proof-of-concept

implementation. In serial production, this can be used in two potential applications: (1) to improve the error detection and (2) to increase the production availability. The implementation of an automatic distinction between errors and pseudo-errors with the feedback of the reworking statement is a possibility to improve the process monitoring system. With the feedback of the reworking statement, an automatic distinction between errors and pseudo-errors is made and false-positive errors are recognized. An adaptive improvement of the process monitoring system of SSPR can be established with a machine learning model, which is trained on in serial production collected process data. Therefore, the manual rework can be reduced by focusing only on the necessary components. This is followed by a reduction of downtime and an increase in process availability. Overall, the developed mechanism already shows promising results and its feasibility. For future applications, it will be necessary to improve the classification model and in particular improve the data quality.

References

- [1] Kim K Y, Sim J, Jannat N E, Ahmed F and Ameri S 2019 *Procedia Manufacturing* **38** 1143–1150 ISSN 2351-9789
- [2] Susana Ferreira, Egoitz Konde, Santiago Fernández, Agustín Prado 2016 *Third European Conference of the Prognostics and Health Management Society*
- [3] Deutsches Institut für Normung eV 2003 *Fertigungsverfahren Fügen Teil 5*
- [4] Li D, Chrysanthou A, Patel I and Williams G 2017 *The International Journal of Advanced Manufacturing Technology* **92** 1777–1824 ISSN 1433-3015
- [5] L Budde, W Lappe, F Liebrecht 1992 *Bleche Rohre Profile* 310–314
- [6] Shearer C 2000 *Journal of data warehousing* **5** 13–22
- [7] Brühl V 2019 Big data, data mining, machine learning und predictive analytics
- [8] Rebala G, Ravi A and Churiwala S 2019 *An Introduction to Machine Learning* (Springer International Publishing) ISBN 3030157296
- [9] Bäck T and Schwefel H P 1993 *Evolutionary Computation* **Nr. 1** 1–23
- [10] Xue B, Zhang M, Browne W N and Yao X 2016 *IEEE Transactions on Evolutionary Computation* **20** 606–626
- [11] TAN Y 2004 *Beurteilung des Tragverhaltens von Clinchverbindungen mittels des Einsatzes künstlicher neuronaler Netze* Dissertation Universität Paderborn
- [12] Sierak P 2020 *Qualifizierung intelligenter Datenanalysemethoden bei vollautomatisierten Klebtechnik anwendungen* Dissertation Universität Paderborn
- [13] Röske J 2018 *Untersuchung alternativer Sensoren und Datenanalysemethoden für Schraubprozesse* Dissertation Universität Paderborn
- [14] Boersch I, Füßel U, Gresch C, Großmann C and Hoffmann B 2018 *The International Journal of Advanced Manufacturing Technology* 1085–1099 ISSN 1433-3015
- [15] Karathanasopoulos N, Pandya K S and Mohr D 2021 *Journal of Advanced Joining Processes* **3** 100040
- [16] Khrebtov P 2011 *Neuartiges Verfahren zur Online-Prozessüberwachung und -Fehlerklassifizierung beim Durchsetzfügeverbinden von Blechen* Dissertation Technischen Universität Clausthal
- [17] Falk T and Jäckel M 2019 *AIP Conference Proceedings* **2113** 050011
- [18] Pereira S, Baptista M and Henriques E M P 2018 *2018 IEEE Aerospace Conference* pp 1–15
- [19] Bock F E, Blaga L A and Klusemann B 2020 *Procedia Manufacturing* **47** 615–622
- [20] Li Y, Yu B, Wang B, Lee T H and Banu M 2020 *Materials and Design* **194** 108912 ISSN 0264-1275
- [21] Scipystats 2021 Statistical functions <https://docs.scipy.org/doc/scipy/reference/stats.html>
- [22] numpyorg 2021 Numpy - the fundamental package for scientific computing with python numpy.org
- [23] Mitchell T M 1997 *Machine Learning* (McGraw-Hill, Inc) ISBN 0070428077