Improving quality prediction in radial-axial ring rolling using a semi-supervised approach and generative adversarial networks for synthetic data generation

Simon Fahle¹ · Thomas Glaser¹ · Andreas Kneißler² · Bernd Kuhlenkötter¹

Received: 1 June 2021 / Accepted: 24 August 2021 / Published online: 7 September 2021
© The Author(s) 2021

Abstract

As artificial intelligence and especially machine learning gained a lot of attention during the last few years, methods and models have been improving and are becoming easily applicable. This possibility was used to develop a quality prediction system using supervised machine learning methods in form of time series classification models to predict ovality in radial-axial ring rolling. Different preprocessing steps and model implementations have been used to improve quality prediction. A semi-supervised approach is used to improve the prediction and analyze, to what extent it can improve current research in machine learning for quality prediction. Moreover, first research steps are taken towards a synthetic data generation within the radial-axial ring rolling domain using generative adversarial networks.

Keywords Radial-axial ring rolling · Time series classification · GAN · Semi-supervised · Machine learning

Abbreviations

ACGAN	Auxiliary classifier generative adversarial network
BOSS	Bag of symbolic-Fourier-approximation-symbols
CGAN	Conditional generative adversarial network
COTE	Collective of transformation-based ensembles
DCGAN	Deep convolutional generative adversarial network
ECTS	Early classification on time series
GAN	Generative adversarial networks
HIVE-COTE	Hierarchical vote collective of transformation-based ensembles for time series classification
LSTM-FCN	Long short-term memory fully convolutional network
RARR	Radial-axial ring rolling
ROCKET	Random convolutional Kernels

SL	Supervised learning
SSL	Semi-supervised learning
Std	Standardization
TRTS	Train-real-test-synthetic
TSC	Time series classification
TSF	Time series forest
TSTR	Train-synthetic-test-real

Nomenclature

RWG	Ring-growth-rate
OD	Outer diameter
F_{rad}	Radial force
F_{ax}	Axial force

1 Introduction

The presented research at hand is building on earlier studies of the authors. These studies introduced a quality prediction approach in the domain of Radial-Axial Ring Rolling (RARR) with regard to form errors and especially ovality [1]. Moreover, this approach was enhanced by a domain specific preprocessing approach [2] and an evaluation on the best performing model for a Time Series Classification (TSC) task was performed [3]. Additionally, the authors shifted the problem definition from a TSC task into an Early Classification on Time Series (ECTS) task in their most recent work that is accepted for publication as well. This

✉ Simon Fahle
fahle@lps.rub.de

¹ Chair of Production Systems (LPS), Ruhr-Universität Bochum, Universitätsstr. 150, 44801 Bochum, Germany

² Fachbereich Maschinenbau, Umwelt- und Gebäudetechnik, Lehrgebiet Produktionssysteme, Westfälische Hochschule, Neidenburger Str. 43, 45897 Gelsenkirchen, Germany

ECTS approach now enables not only prediction but also prevention of form defects and can thus further improve the process of RARR. This research, especially the early classification approach, showed promising results, but the model performance still has potential to be optimized. For this refinement, two approaches are presented in this research, focusing on two different aspects:

- using unlabeled data in combination with labeled data
- generating synthetic data using generative models

These approaches try to address the main difficulty the authors had to face during their extensive work within the domain and are prominent throughout all machine learning fields: labeled data acquisition. As further detailed in section 3.1, the underlying data set consists of different rollings of an industrial rolling machine at *thyssenkrupp rothe erde Germany GmbH*. Different blank geometries as well as a wide variety of rolled height to wall-thickness ratios are present. For each ring there is an individual measurement using a line laser unit. Each ring's outer shape is measured and thus a target for each rolled ring regarding form errors is produced. Yet the laser unit is costly and requires constant maintenance hence it is not running all the time, but unlabeled data is produced automatically and is therefore acquired by the authors as well. This unlabeled data will be used within the semi-supervised approach to improve classification accuracies on the baseline TSC approach. Nonetheless, the authors establish a first attempt at synthetic data generation for different process data channels. Within this generation, the authors conducted an expert interview with process experts to further improve the generation task.

To recap, our contributions are:

- Enhancing ML in RAW using Semi-Supervised Learning (SSL)(self-training) from the real world production plant
- Using Generative Adversarial Networks (GANs) to increase the data sets even further with synthetic data sets

Both approaches are depicted in Fig. 1 in the middle part of the figure enhancing the baseline TSC approach.

2 Related work

The following section briefly discusses current trends and the state of the art applications regarding RARR, quality prediction, as well as the machine learning topics of time series classification, semi-supervised approaches and generative adversarial networks. The discussion starts with the following section about RARR, as it is the used domain of interest for the proposed approach within this research.

2.1 RARR

Radial-axial ring rolling is an important process for the production of seamlessly rolled rings. Even though the technology already exists for many decades, process improvements are still being researched. Improvements range from simulations, innovative combination of processes and materials as well as quality related issues. As for innovative combinations with other processes, current research by MICHL ET AL. investigates the possibility to use wire arc additive manufacturing to produce better pre forms. This approach shows promising results and has potential to increase process efficiency and lower process expenses [4]. Another innovative combination of traditional approaches is pursued by KUHLENKÖTTER ET AL. by using a combination of RARR with thermal spraying. Their intention is to compact coatings by rolling a sprayed ring. Their experimental results indicate that a final intact coating has yet to be rolled, still they managed to induce higher residual stress into the ring and reduce the porosity of the sprayed layer [5]. Moreover, GÜNTHER ET AL. focus on a combination of roll bonding and ring rolling to produce rings tailored to specific application. They support their innovative research using finite-element simulations [6]. In addition, simulations are used by LIANG ET AL. for an intelligent rolling simulation of titanium alloys.

Fig. 1 Schematic representation of the present research approach

This is done by taking the material temperature into account within a simulation model implemented in Abaqus/Explicit [7]. With regard to research in the quality prediction area, recent advances have been published by FAHLE ET AL. The authors investigated all necessary steps towards a data driven analysis using time series classification approaches. An initial study in 2019 was conducted to present the current state of data usage in RARR [1]. This was enhanced by a comparable study on domain specific preprocessing [2] and was followed up by a full time series classification. The taken approach was further improved by formulating an ECTS approach to not only predict, but also prevent the ovality form error in RARR.

2.2 Quality prediction

Due to massive improvements in artificial intelligence, quality prediction approaches increased in other manufacturing processes as well. For example, the utilization of unsupervised methods for a quality monitoring in metallic powder presses [8] or quality prediction regression models in rolling by KIRCH ET AL. [9]. Another in-process quality monitoring approach is proposed by BAUERDICK for machine tools [10]. Further, an approach by TANGJITSITCHAROEN ET AL. in ball-end milling achieved up to 92 % accuracy by using different machine data, e.g. the feed rate and dynamic cutting force ratio, as input features [11] or similarly the approach by ASILTÜRK ET AL. using neural networks for a surface roughness prediction [12]. Further, a similar approach was taken by LIU ET AL., using five different sensors and fifteen features in total to achieve up to 91 % accuracy with a support-vector machine in an welding application [13]. Lastly, a holistic approach for quality inspection using edge cloud computing is proposed by SCHMITT ET AL. They implement and validate their approach in a real industrial manufacturing use case for surface mounted technology and demonstrate that inspection volumes can be reduced as well as economic advantages can be achieved [14].

2.3 Time series classification

Time series classification is an approach where a set of feature-target tuples is used. The features represent different measurements taken over time and the targets represent discrete classes. Time series classification is different with regard to classical relational data classification and the differences are clearly outlined by LÖNING ET AL. In addition, they elaborate different tasks within the time series domain [15]. Due to the raising availability of time series data [16] a wide variety of algorithms have been proposed that are able to analyze time series data.

Within the Python programming language, many useful implementations of algorithms and even full libraries have been made available open-source, such as *sktime(-dl)* by LÖNING ET AL. [15] or *tsfresh* by TAVENARD ET AL. [17]. Both are partially used within the conducted research at hand.

Time series models can be categorized following BAGNAL ET AL. [18]:

- Whole series approaches compare series using different distance measures. Best performances were reached using the Dynamic Time Warping similarity measure [18].
- Interval based approaches use features that are time dependent and derived from intervals of each series. A promising representative of interval based approaches is the Time Series Forest (TSF) by DENG ET AL. [19].
- Dictionary based approaches try to discriminate between a whole series by using a representation of sub-series and their frequencies [20]. A famous representative of this category is the Bag of Symbolic-Fourier-Approximation-Symbols (*BOSS*) algorithm proposed by SCHÄFER ET AL. [21].
- Shapelet based models try to find unique and distinctive shapelets within time series. These shapelets (sub-sequences) are local, phase independent, and are used as a discriminative feature for another classifier [22]. Yet, according to FAWAZ ET AL. the training complexity of shapelet algorithms is high and thus they are not competitive for bigger data sets or real world applications [20].
- Combined transformations are ensembles of different classifiers that differ in their data representation. Example models are named Collective of Transformation-Based Ensembles (COTE) [23] or its improvement Hierarchical Vote Collective of Transformation-Based Ensembles for Time Series Classification (HIVE-COTE) [24].

The used models within this research are Random Convolutional Kernels (ROCKET) [25] and a long short-term memory fully convolutional network (LSTM-FCN) model [26]. Both will be used within this research as baseline classifiers as they proved to provide sufficient accuracies as well as good inference times in an earlier study on the data sets in the domain of RARR. The LSTM-FCN model reached about 88 % accuracy on the RARR data set, whereas the ROCKET model achieved 87 %. Next to these models a TSF model achieved a slightly better accuracy than the ROCKET model with 87.5 %, yet it does not include a native approach to use multivariate data and scaled very poorly with higher dimensional data [3]. As the intent of the underlying research is to use even more highdimensional data the TSF model is not considered

anymore. Another study by the author (accepted for publication in CPSL 2021 Proceedings) shows that the findings of [3] are also applicable in an extension of the TSC task towards an ECTS approach.

2.4 Semi-supervised learning and GANs in TSC

The main intention of using SSL is to improve the baseline supervised task with an utilization of expensively acquired labeled targets using easier to acquire unlabeled samples. The usefulness of this approach can be seen in the Computer Vision field and can be categorized into inductive and transductive approaches according to VAN ENGELEN ET AL. [27].

An early work on SSL for TSC was proposed by WEI ET KEOGH in 2006 using pseudo-labelling [28]. Another work by WANG ET AL. proposes the semi-supervised learning of shapelets. Their model learns shapelets from both labeled and unlabeled time series and is in contrast to kernel-based methods such as the aforementioned approach by WEI ET KEOGH [29]. The advances in neural networks and their success in other domains continues in SSL as well. In 2017, ZENG ET AL. proposed a semi-supervised convolution based approach for human activity recognition and were able to increase the mean F1-score for selected data sets by up to 17.6 percentage points from 48.7 % in the supervised case to 66.3 % in their proposed semi-supervised approach [30]. A multi-task network structure combining latent representations between the forecasting and classification task was trained by JAWED ET AL. in a semi-supervised way. They managed to outperform state of the art baselines on different data sets [31]. Another recent model called *TapNet* was proposed by ZHANG ET AL. in 2020. The architecture of the model consists of a combination of convolutional-layers combined with a recurrent LSTM-unit, concatenated before a fully-connected layer. This is the dimension permutation part and is followed by a time series encoding. This encoding is lastly followed by an attentional prototype learning section. In addition, a SSL approach is taken where unlabeled data is used to help finding the class prototypes of the labeled data set [32].

The state of the art, in the field of synthetic data generation using GANs, is still very limited due to its still very short history. Within physics GANs have already been successfully used for research in the field of black matter [33] and particle showers [34]. Furthermore, GANs have been used for the generation of so-called deepfakes. Deepfakes are intended to deceive either the human, the machine, or both. This has been done, for example, by synthetically altering cancer diagnostic images or the well-known deepfake videos of politicians. A comprehensive review regarding deepfakes can be taken from MIRSKY ET AL. [35]. GANs for generating time series is a less explored topic, but some successes have already been achieved and individual GAN

structures have been developed. One of these structures is *TimeGAN* by YOON ET AL. which combines the unsupervised learning advantages of GANs and the supervised advantages of autoregressive models. Through this combination, they achieved the state of the art on various time series datasets [36]. Furthermore, especially within medicine, attempts were made to produce time series by using GANs [37]. Another model called *SeqGAN* was proposed by YU ET AL. and is used to generate text sequences [38]. Due to the promising success of the GANs in the other areas, these are now being transformed into the area of RARR.

3 Baseline quality prediction approach

In order to increase productivity and to ensure a constant competitiveness of RARR, the following section describes the concrete implementation to reduce quality-related costs and rework.

3.1 Problem definition and data set

The machine learning problem is defined as a (semi-) supervised time series classification task. For a formal definition of TSC refer to [15] and for a domain specific RARR definition of the TSC task see [2]. All in all, the task is to predict whether a rolled sample lies within a defined threshold for ovality or not. The data set consists of 1256 samples of real world production data from *thyssenkrupp rothe erde Germany GmbH*. On top of the 1256 labeled data samples with a measurement, there are 2414 samples without an explicit measurement, but from the same machine. This leads to a supervised data set of 1256 samples and a semi-supervised data set of 3670 (1256 real-labeled and 2414 pseudo-labeled) data samples.

For both data sets, the used features are identical. In sum, more than 100 features (i.e. forces, torques, geometric and control values, currents of motors etc.) are available. Yet only 50 of them are used as input features for the label prediction. This subset of features was elaborated in earlier research and led to the best performances regarding the classification accuracy for the detection of ovality in RARR. As for preprocessing, all data samples were scaled to a fixed length, as required by many approaches in TSC. To ensure the fixed length, a domain specific rolling phase scaling approach [2] is used to make individual rollings more comparable. The phase scaling approach is performed using statistical values derived from the data set. The phase scaling approach is a RARR domain specific approach. The rolling process consists of four idealized ring rolling phases and the phase scaling approach makes use of this. An overall mean length of the rolling phases was investigated and for every rolling, all four phases were linearly scaled separately

so that there is a better comparison between all four phases. This results in four distinct phase lengths that all rollings are scaled to, producing equal length time series. All results are gathered using a five-fold stratified random shuffle split. Due to a non-disclosure agreement, the data set must not be made available to the public. The data set is not perfectly balanced, meaning that there is no perfect 50/50 split between oval and non-oval samples. This is due to the internally set and machine related threshold regarding ovality to constantly increase the ring quality. The split is roughly 54/46, which means that a naive classifiers accuracy should be about 54 % accuracy and has not yet learned to discriminate the underlying information correctly.

4 Experiment section

For an implementation into a fully-automated production line in a real industrial setting as a substitution for a costly measurement unit, prediction accuracy should be increased. This is researched using the approach described below, which is a semi-supervised approach and thus making use of the initial 1256 labeled data samples as well as the 2414 unlabeled data samples.

4.1 Semi-supervised approach

The used SSL approach at hand is a mix of a so-called self-training and pseudo-label approach as used by XIE ET AL. [39]. A classifier is trained on a labeled training data set and is then used to pseudo-label data that is not labeled. Both, the labeled train instances as well as the pseudo labeled instances are then combined to retrain a classifier and finally evaluate its performance on the holdout labeled test set. This is done using a five fold random stratified shuffle split with

five initializations each for both classifiers for a generalized performance evaluation.

The present implementation uses an ensemble consisting of two deep-learning models as well as a non deep-learning model to initially pseudo-label the unlabeled data. The ensemble was taken, because it represents a good mix of models with high single accuracy with moderate standard deviation as well as high accuracy with low standard deviation. One retrained classifier is a bigger version of the LSTM-FCN used in the pseudo label approach, since a recent study by XIE ET AL. proposed a noisy student model, that proposes to use noise in addition with an even bigger network as the student/retrained model [39]. Thus, the size and depth of the initial LSTM-FCN model was increased before it was used to be retrained on the combined SSL data set. Moreover, a ROCKET-model was used as a second baseline model. The used noise was separated into model and input noise. Model related noise was always applied by using dropout layers within the model architecture in the LSTM-FCN model, whereas the input noise was introduced using random Gaussian noise with a mean of zero and a standard deviation of 0.05. All trainings were performed using pre-processing with or without standardization. All these factors are indicated in Fig. 2 with the abbreviation of “SSL” for “semi-supervised learning” for the used SSL approach. “SL” for “supervised learning” as a baseline using no SSL to see whether it improves the classification performance. “Noise” if random gaussian noise was added to the data and “Std” if standardization was used. If “Noise” and “Std” are missing in the description, this indicates that they were not part of the approach in that specific run.

Fig. 2 Comparative results between different semi-supervised learning approaches and the baseline supervised learning approach using noise-addition as proposed in [39]

4.2 Semi-supervised evaluation

The results of the experiments are shown in Fig. 2 and illustrated by the approaches used. It can be seen that the ROCKET model per se gives better results than the LSTM-FCN model. The best result of both models was obtained by the SSL-Noise approach. In both cases, this improved the prediction accuracy on the test data of the two models per se. In general, standardization within the approach worked consistently worse for both models. Furthermore, it can be observed that exclusively the use of the SSL approach resulted in no (ROCKET) or only a very small improvement (LSTM-FCN), but the use of Noise led to an improvement in both cases. In conclusion, the increase of the prediction accuracy by the SSL-Noise approach is not very large, but it is present. The actual increase is from 88.76 % in the SL approach to 88.84 % in the SSL-Noise approach. From this, it is concluded that the general potential is present and needs to be explored through further research.

4.3 Synthetic data generation using GANs

Building on the success of the semi-supervised approach, an attempt was made to generate synthetic data, which in turn will be used to improve the baseline TSC task in the future as well. To generate synthetic data, a GAN approach is used and builds on recent successes and advances described in Sect. 2.4. Three typical architectures are depicted in Fig. 3. Within Fig. 3 x depicts real input rolling samples, G the generator, z the latent space, c the class label (e.g. ovality or no ovality) and D the discriminator. The actual implementation consists of a deep convolutional GAN (DCGAN) architecture, based on a traditional GAN (cf. Fig. 3a), to generate samples of semi-supervised, non-labeled data and is used to see whether

process experts can already be fooled by this approach. In a second implementation, the authors make use of Conditional GAN (CGAN; cf. Fig. 3b) as well as an Auxiliary Classifier Gan (ACGAN; cf. Fig. 3c). Both CGAN and ACGAN are both extensions of the GAN architecture and represent conditioning with respect to the class label. In the case of the CGAN architecture, both a point in latent-space and the class label are passed to the generator. The discriminator is then passed the generated data sample and the associated class. In the case of ACGAN, the generator is still passed a point in latent space as well as the class label, but the discriminator must give both the statement of real and generated samples, as well as a prediction of the class label [40].

The evaluation by process experts in Sect. 4.3.1 is inspired by the popular Inception-Score, where a trained classifier-network (*Inception*) classifies the generated images [41], yet there is no pre-trained network for TSC in RARR. This is why within this research this Inception-Score is mixed with the Hype-Metric proposed by ZHOU ET AL. in 2019 to substitute the pre-trained model. The Hype-Metric stands for human eye perceptual evaluation of generative models and consists of two approaches, one with and one without time constraints [42]. As the task of classifying time series of machines is a difficult task, the approach at hand is chosen without a time constraint as not even process experts would be able to differentiate a real and fake time series in e.g. just 500 ms. Moreover, in contrast to the HYPE approach, the actual distribution of real and fake images was not revealed to the experts.

To give a realistic chance to the process expert, specific and highly process relevant features have been selected to be generated. These features were radial and axial force, ring-growth-rate as well as outer diameter. These features are typically looked at when evaluating different rollings

Fig. 3 Architectures of used GAN implementations; inspired by [40]

and are thus used for the expert evaluation process. To evaluate the produced time series using machine learning models, common metrics called Train-Synthetic-Test-Real (TSTR) as well as Train-Real-Test-Synthetic (TRTS) proposed by ESTEBAN ET AL. [37] were used. TSTR is the more important metric as it shows the usefulness of generated data for the underlying use case.

4.3.1 Process expert evaluation

As stated before, a DCGAN architecture was used to produce realistic RARR data. The approach started by generating a single feature at a time, thus using four individual DCGANS to provide univariate time series for each feature. This approach was also performed as a multivariate approach suffered from different problems that will be elaborated later. Figure 4 depicts an excerpt of the survey that was sent to a handful of process experts within the RARR section. The survey consisted of 20 samples that were randomly taken from either generated time series and from the real world data set. Within Fig. 4 sample one displays a real sample whereas sample two and three are generated samples using GANs. The process experts were asked to try to distinguish real from synthetic samples and, if possible, to provide feedback to the author. The feedback should elaborate on which explicit representation

they focused on, so that an improvement of the GAN architecture by additional domain knowledge can be made. It must be noted that the individual skills of these process experts can hardly be formalized and therefore differences in expertise between these experts may occur. The survey samples were evenly sampled with 10 real and 10 generated samples, yet the process experts did not know this. The results of the survey show that the DCGAN already managed to deceive four out of nine process experts resulting in accuracies ≤ 50 %. However, four other experts found explicit indicators that allowed for a differentiation between the existing synthetic and real samples, achieving 85 % accuracy or higher. One major indicator was the missing time dependence between ring-growth-rate and outer diameter which are directly linked due to the process nature. This lack of connection is a direct cause of the individual generation of those features and will be addressed in a future iteration using the domain-specific knowledge. An initial approach using direct multivariate generation failed, due to an adaption of the sometimes oscillatory nature of the radial and axial forces as well as ring-growth-rate into the outer diameter channel, which is unintended and physically implausible and thus directly indicative of a generated sample. The failure of the multivariate approach is also evident in the evaluation process using TSTR and TRTS below. Moreover, a direct reply from the process experts was a lack of information about

Fig. 4 Excerpt from the conducted survey within german process experts in the domain of RARR displaying the first three samples that were shown to the process experts

Table 1 Survey results by expert

Expert no.	Correct	False	Accuracy (%)
E01	18	2	90
E02	6	14	30
E03	7	13	35
E04	18	2	90
E05	10	10	50
E06	18	2	90
E07	17	3	85
E08	14	6	70
E09	10	10	50

other process channels such as the axial and radial feed rates, yet these additional channels were omitted due to the already difficult task to generate four out of all possible channels using GANs and will be investigated in the future. Other findings of the process experts were gathered and will be taken into consideration for future deployments and research regarding GANs for RARR (Table 1).

4.3.2 GAN model performance evaluation

In contrast to the first evaluation method using a human evaluation process, in the following section a model-based evaluation using the TSTR and TRTS approach is pursued. As stated before, for the underlying use case of a synthetic data generation to increase the accuracy and enhance the TSC-task to predict form errors in RARR, the TSTR metric is more viable than the TRTS, as the main goal is to improve the prediction on the real world data samples. A class conditional architecture for the GANs had to be used to enable this approach. This class conditional approach allows a specific generation of wanted

classes and is thus required for the supervised learning approach. For the evaluation process 3200 synthetic data samples were produced using the GANs. The classifiers used are the earlier mentioned ROCKET and LSTM-FCN models.

4.4 Univariate data generation approach

Looking at the univariate evaluation regarding TSTR in Fig. 5, it can be seen that the CGAN architecture managed to produce useful samples. CGAN achieves up to 77.4 % accuracy when trained and validated on generated data samples of the ring-growth channel and tested on real data and slightly less accuracies using the radial and axial forces. Yet, only the outer diameter channel was not generated in a useful manner. The accuracy did not overcome the underlying imbalance ratio of the real world samples of 54/46 resulting in a prediction that only matches the class distribution of the data. Moreover, the auxiliary classifier GAN (ACGAN) architecture failed to produce useful samples at all and even produces misleading samples regarding outer diameter and ring-growth-rate. For the TSTR univariate approach there is a clear preference regarding the used classifier towards using the deep-learning LSTM-FCN model architecture instead of the ROCKET model as the LSTM-FCN model outperforms ROCKET on all three relevant features that were generated in a useful manner (cf. CGAN_F_ax, CGAN_F_rad, CGAN_RWG).

Similar results can be seen considering the second metric, TRTS using the univariate approach, depicted in Fig. 6. As the classifiers are trained and validated on real world samples and tested on generated samples, the (im-)balance ratio is perfectly split between round and oval samples, thus accuracies of 50 % do not represent a learning process by the classifier. Especially the radial force produced by the

Fig. 5 Comparison of TSTR results regarding an univariate generation ordered by the used GAN architecture

Fig. 6 Comparison of TRTS results regarding an univariate generation ordered by the used GAN architecture

Table 2 Multivariate test set accuracy ordered by classifier and approach

Approach	Architecture	Classifier	Accuracy (%)
TSTR	ACGANMTS	ROCKET	0.477
TSTR	ACGANMTS	LSTM_FCN	0.425
TSTR	CGANMTS	ROCKET	0.461
TSTR	CGANMTS	LSTM_FCN	0.658
TRTS	ACGANMTS	ROCKET	0.499
TRTS	ACGANMTS	LSTM_FCN	0.505
TRTS	CGANMTS	ROCKET	0.523
TRTS	CGANMTS	LSTM_FCN	0.500

CGAN allows for accuracies up to 97.4 % for the LSTM-FCN model. The CGAN_F_ax feature was evenly useful for both classifiers to test on the synthetic data, yet CGAN_RWG was significantly better for the LSTM-FCN model. Nevertheless, the outer diameter channel was not usefully generated resulting in a classification by chance for both classification models.

4.5 Multivariate data generation approach

Comparing results for the multivariate approaches depicted in Table 2, the multivariate approach still needs further research and tuning. Only the TSTR score for the LSTM-FCN classifier using the CGAN generated multivariate time series results in at least slight improvements compared to the naive classification results of the imbalance ratio of the real world samples. The TRTS approach failed completely, as neither a classifier managed to learn from the data nor a GAN architecture was able to produce data sufficiently for a viable learning process. Comparing results regarding GAN

architectures, and especially throughout all comparisons depicted in Figs. 5 and 6 and Table 2, the ACGAN architecture failed to generate reliable and realistically looking RARR data samples. The CGAN architecture, on the other hand, performed better and at some points produced sufficient synthetic data samples of RARR time series samples for a first approach.

5 Conclusion

Within the present research work, on the one hand, an approach to improve time series classification by means of a semi-supervised machine learning approach was pursued, while simultaneously evaluating the usability of a future extension of this approach by synthetically generating data using GANs. The results of the SSL approach show improvements with respect to the baseline TSC approach and may help to further improve prediction accuracy. The approach of synthetic data generation using a CGAN architecture already led to partially useful results, which have been evaluated by both human process experts as well as by TSTR and TRST metrics. Within the present use case of RARR, it can be stated that for a generation of synthetic data, the CGAN architecture performed significantly better than the ACGAN architecture. The hoped-for stabilization of the training process, which was expected from the ACGAN architecture in contrast to the CGAN architecture, did not occur in the case of the RARR. At the current state, it is not known why the CGAN architecture provides the better results and this will be pursued in future work. The possibility of the CGAN architecture to output the class label directly is of great value for the productive integration into a system for quality prediction or even fault-prevention. This generated data could be additionally used in the semi-supervised approach

shown to further enhance the increase in accuracy provided by the SSL approach and potentially integrate its own form of noise.

In the course of further research in this area, the presented first approaches will be further improved and prepared for conductive use as the ability to produce high amounts of new and in the case of CGANs labeled data is of high interest to further improve the process efficiency by using bigger data sets. Moreover, it is an interesting field of research to explore the latent space for the domain of RARR further as it was done before with image data. A better understanding of the latent space could lead to a data generation for specific parameters such as material or rolled and preform geometry. Further research fields in the future could be the applicability and generalizability of the approaches to a wide variety of rolling mills and even different processes very similar to RARR, such as cold rolling.

Acknowledgements This research work was funded by the *German Research Foundation (DFG)* within the research project “Fehlervermeidung in Radial-Axial Ringwalzprozessen durch Online-Analyse der Zustandsdaten” (KU 1543/20-1, project number 404517758). The authors thank the DFG for promoting and facilitating the research. Moreover, the authors would like to thank *thyssenkrupp rothe erde Germany GmbH* for their support and the possibility of data acquisition at their production plant. Moreover, the authors would like to thank all process experts for their participation and their kind feedback.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Fahle S, Kuhlenkötter B (2020) A framework for data integration and analysis in radial-axial ring rolling (1st Conference on Production Systems and Logistics):127–136:10.15488/9654
- Fahle S, Kneißler A, Glaser T, Kuhlenkötter B (2021) Research on preprocessing methods for time series classification using machine learning models in the domain of radial-axial ring rolling. In: Behrens BA, Brosius A, Hintze W, Ihlenfeldt S, Wulfsberg JP (eds) *Production at the leading edge of technology Lecture Notes in Production Engineering*. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp 487–496. https://doi.org/10.1007/978-3-662-62138-7_49
- Fahle S, Glaser T, Kuhlenkötter B (2021) Investigation of machine learning models for a time series classification task in radial-axial ring rolling. In: Daehn G, Cao J, Kinsey B, Tekkaya E, Vivek A, Yoshida Y (eds) *Forming the future the minerals, metals & materials series*. Springer International Publishing, Cham, pp 589–600. https://doi.org/10.1007/978-3-030-75381-8_48
- Michl D, Sydow B, Bambach M (2020) Ring rolling of preforms made by wire-arc additive manufacturing. *Procedia Manuf* 47:342–348. <https://doi.org/10.1016/j.promfg.2020.04.275>
- Kuhlenkötter B, Glaser T, Fahle S, Husmann S, Abdulgader M, Tillmann W (2020) Investigation of compaction by ring rolling on thermal sprayed coatings. *Procedia Manuf* 50:192–198. <https://doi.org/10.1016/j.promfg.2020.08.036>
- Guenther S, Schwich G, Hirt G (2020) Investigation of bond formation behaviour in composite ring rolling. *J Mater Process Technol* 275:116364. <https://doi.org/10.1016/j.jmatprotec.2019.116364>
- Liang L, Guo L, Wang Y, Li X (2019) Towards an intelligent fe simulation for real-time temperature-controlled radial-axial ring rolling process. *J Manuf Process* 48:1–11. <https://doi.org/10.1016/j.jmapro.2019.09.032>
- Weiß I, Vogel-Heuser B, Trunzer E, Kruppa S (2020) Unsupervised product quality monitoring in hydraulic metal powder presses using a minimal sample of sensor and actuator data. *ACM Trans Int Technol*
- Kirchen I, Vogel-Heuser B, Hildenbrand P, Schulte R, Vogel M, Lechner M, Merklein M (2017) Data-driven model development for quality prediction in forming technology. pp. 775–780. <https://doi.org/10.1109/INDIN.2017.8104871>
- Bauerdick CJH (2020) *Methodische Entwicklung einer in-Prozess Qualitätsüberwachung bei der Drehbearbeitung basierend auf maschineninternen Daten [1. auflage]*. edn. Schriftenreihe des PTW: Innovation Fertigungstechnik. Shaker Verlag, Düren
- Tangjitsitharoen S, Thesniyom P, Ratanakuakangwan S (2017) Prediction of surface roughness in ball-end milling process by utilizing dynamic cutting force ratio. *J Intell Manuf* 28(1):13–21. <https://doi.org/10.1007/s10845-014-0958-8>
- Asiltürk İ, Çunkaş M (2011) Modeling and prediction of surface roughness in turning operations using artificial neural network and multiple regression method. *Expert Syst Appl* 38(5):5826–5832. <https://doi.org/10.1016/j.eswa.2010.11.041>
- Liu G, Gao X, You D, Zhang N (2019) Prediction of high power laser welding status based on pca and svm classification of multiple sensors. *J Intell Manuf* 30(2):821–832. <https://doi.org/10.1007/s10845-016-1286-y>
- Schmitt J, Bönig J, Borggräfe T, Beitingner G, Deuse J (2020) Predictive model-based quality inspection using machine learning and edge cloud computing. *Adv Eng Inf* 45:101101. <https://doi.org/10.1016/j.aei.2020.101101>
- Löning M, Bagnall A, Ganesh S, Kazakov V, Lines J, Király FJ (2019) sktime: A unified interface for machine learning with time series. <https://arxiv.org/pdf/1909.07872>
- Silva DF, Giusti R, Keogh E, Batista Gustavo E. A. P. A (2018) Speeding up similarity search under dynamic time warping by pruning unpromising alignments. *Data Min Knowl Disc* 32(4):988–1016. <https://doi.org/10.1007/s10618-018-0557-y>
- Romain T, Johann F, Gilles V, Felix D, Guillaume A, Chester H, Marie P, Roman Y, Marc R, Kushal K, Eli W (2020) Tslern, a machine learning toolkit for time series data. *J Mach Learn Res* 21(118):1–6. <http://jmlr.org/papers/v21/20-091.html>
- Bagnall A, Lines J, Bostrom A, Large J, Keogh E (2017) The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Min Knowl Disc* 31(3):606–660. <https://doi.org/10.1007/s10618-016-0483-9>
- Deng H, Runger G, Tuv E, Vladimir M (2013) A time series forest for classification and feature extraction. *Inf Sci* 239:142–153. <https://doi.org/10.1016/j.ins.2013.02.030>

20. Ismail Fawaz H, Lucas B, Forestier G, Pelletier C, Schmidt DF, Weber J, Webb GI, Idoumghar L, Muller PA, Petitjean F (2020) Inceptiontime: Finding alexnet for time series classification. *Data Min Knowl Disc* 34(6):1936–1962. <https://doi.org/10.1007/s10618-020-00710-y>
21. Schäfer P (2015) The boss is concerned with time series classification in the presence of noise. *Data Min Knowl Disc* 29(6):1505–1530. <https://doi.org/10.1007/s10618-014-0377-7>
22. Hills J, Lines J, Baranauskas E, Mapp J, Bagnall A (2014) Classification of time series by shapelet transformation. *Data Min Knowl Disc* 28(4):851–881. <https://doi.org/10.1007/s10618-013-0322-1>
23. Bagnall A, Lines J, Hills J, Bostrom A (2015) Time-series classification with cote: the collective of transformation-based ensembles. *IEEE Trans Knowl Data Eng* 27(9):2522–2535. <https://doi.org/10.1109/TKDE.2015.2416723>
24. Lines J, Taylor S, Bagnall A (2018) Time series classification with hive-cote. *ACM Trans Knowl Discov Data* 12(5):1–35. <https://doi.org/10.1145/3182382>
25. Dempster A, Petitjean F, Webb GI (2019) Rocket: exceptionally fast and accurate time series classification using random convolutional kernels. <http://arxiv.org/pdf/1910.13051v1>
26. Karim F, Majumdar S, Darabi H, Harford S (2019) Multivariate lstm-fcns for time series classification. *Neural networks: the official journal of the International Neural Network Society* 116:237–245. <https://doi.org/10.1016/j.neunet.2019.04.014>
27. van Engelen JE, Hoos HH (2020) A survey on semi-supervised learning. *Mach Learn* 109(2):373–440. <https://doi.org/10.1007/s10994-019-05855-6>
28. Wei L, Keogh E (2006) Semi-supervised time series classification. In: Ungar L, Craven M, Gunopulos D, Eliassi-Rad T (eds.) *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '06* p.748. ACM Press, New York. <https://doi.org/10.1145/1150402.1150498>
29. Wang H, Zhang Q, Wu J, Pan S, Chen Y (2019) Time series feature learning with labeled and unlabeled data. *Pattern Recogn* 89:55–66. <https://doi.org/10.1016/j.patcog.2018.12.026>
30. Zeng M, Yu T, Wang X, Nguyen LT, Mengshoel OJ, Lane I (2017) Semi-supervised convolutional neural networks for human activity recognition. In: 2017 IEEE International Conference pp 522–529. <https://doi.org/10.1109/BigData.2017.8257967>
31. Jawed S, Grabocka J, Schmidt-Thieme L (2020) Self-supervised learning for semi-supervised time series classification. In: Lauw HW, Wong RCW, Ntoulas A, Lim EP, Ng SK, Pan SJ (eds) *Advances in Knowledge Discovery and Data Mining, Lecture Notes in Computer Science*, vol 12084, pp 499–511. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-47426-3_39
32. Zhang X, Gao Y, Lin J, Lu CT (2020) Tapnet: multivariate time series classification with attentional prototypical network. *Proc AAAI Conf Artif Intell* 34(04):6845–6852. <https://doi.org/10.1609/aaai.v34i04.6165>
33. Mustafa M, Bard D, Bhimji W, Lukić Z, Al-Rfou R, Kratochvil JM (2019) Cosmogon: creating high-fidelity weak lensing convergence maps using generative adversarial networks. *Comput Astrophys Cosmol*. <https://doi.org/10.1186/s40668-019-0029-9>
34. Erdmann M, Glombitza J, Quast T (2019) Precise simulation of electromagnetic calorimeter showers using a wasserstein generative adversarial network. *Comput Softw Big Sci*. <https://doi.org/10.1007/s41781-018-0019-7>
35. Mirsky Y, Lee W (2021) The creation and detection of deepfakes. *ACM Comput Surv* 54(1):1–41. <https://doi.org/10.1145/3425780>
36. Jinsung Y, Daniel J, Mihaela van der S (2019) Time-series generative adversarial networks (33rd Conference on Neural Information Processing Systems (NeurIPS 2019))
37. Esteban C, Hyland SL, Rätsch G (2017) Real-valued (medical) time series generation with recurrent conditional gans. <http://arxiv.org/pdf/1706.02633v2>
38. Yu L, Zhang W, Wang J, Yu Y (2016) Seqgan: Sequence generative adversarial nets with policy gradient. <http://arxiv.org/pdf/1609.05473v6>
39. Xie Q, Luong MT, Hovy E, Le VQ (2019) Self-training with noisy student improves imagenet classification. <http://arxiv.org/pdf/1911.04252v4>
40. Mino A, Spanakis G (2018) Logan: generating logos with a generative adversarial neural network conditioned on color. In: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), pp 965–970, IEEE. <https://doi.org/10.1109/ICMLA.2018.00157>
41. Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A, Chen X (2016) Improved techniques for training gans. <http://arxiv.org/pdf/1606.03498v1>
42. Zhou S, Gordon ML, Krishna R, Narcomey A, Fei-Fei L, Bernstein MS (2019) Hype: a benchmark for human eye perceptual evaluation of generative models. <http://arxiv.org/pdf/1904.01121v4>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.