Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Simulation-Driven Machine Learning for Ring Rolling: A Proof of Concept on Transfer and Hybrid Learning Approaches

**Johannes Seitz**

seitz@lps.ruhr-uni-bochum.de

Ruhr University Bochum: Ruhr-Universitat Bochum https://orcid.org/0009-0007-9133-8267

**Luca Quagliato**
**Qinwen Wang**
**Mattia Perin**
**Helin Siedo**
**Tobias Moser**
**Alexander Brosius**
**Bernd Kuhlenkötter**

**Research Article**

**Keywords:** Ring rolling, Physics-informed machine learning, Transfer learning, Domain adaptation, Process modeling

**Posted Date:** November 10th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11399786/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

# Simulation-Driven Machine Learning for Ring Rolling: A Proof of Concept on Transfer and Hybrid Learning Approaches

Johannes Seitz [1*], Luca Quagliato [2†], Qinwen Wang [3], Mattia Perin [4], Helin Siedo [1], Tobias Moser [1], Alexander Brosius [3], and Bernd Kuhlenkötter [1]

[1] Chair of Production Systems, Ruhr-University Bochum, Bochum, Germany

[2] Department of Industrial Engineering, University of Trento, Trento, Italy

[3] Chair of Forming and Machining Technology, TUD Dresden University of Technology, Dresden, Germany

[4] Department of Management and Engineering, University of Padua, Vicenza, Italy

*Corresponding author's details: Email: seitz@lps.ruhr-uni-bochum.de; Tel.: +49 234 32 27627

†Co-corresponding author's details: Email: luca.quagliato@unitn.it; Tel.: +39 0461 283728

## Abstract

Finite element analysis (FEA) is widely used in forming technologies to enhance process modeling and generate process-level insight. However, discrepancies between numerical simulations and experimental or industrial reality remain, depending on both the specific process and the simulation setup. When machine learning (ML) models are trained exclusively on FEA-generated data, these discrepancies are implicitly embedded in the training dataset and may limit the model's predictive reliability on real process data. Motivated by the complementary strengths and limitations of heterogeneous data sources, this study investigates hybrid approaches that aim to maximize data utilization when experimental data is scarce and improve model performance rather than relying on a single data source. Accordingly, the paper examines the integration of FEA data, the most commonly used simulation data type, into machine-learning-based modeling of experimental ring rolling processes. In this regard, two major concepts are considered: hybrid training and transfer learning. This proof of concept discusses the methods on two different use cases: A general introduction and elaboration of an unsupervised representation learning task in the radial-axial ring rolling (RARR) process, as well as the supervised prediction of the out-of-roundness error in the tangential cold ring rolling (TCRR). Metrics are evaluated across tasks and methods, mechanisms analyzed, and applications presented. Ultimately, this study establishes a simulation-informed machine learning framework for ring rolling, showing that transfer learning and targeted data pre-processing can effectively bridge FEA-based knowledge and real process data to improve predictive modeling performance.

**Keywords:** Ring rolling; Physics-informed machine learning; Transfer learning; Domain adaptation, Process modeling

## 1. Introduction

The radial-axial ring rolling (RARR) of flat rings is a well-known process used to manufacture seamless rings for various industries, such as aerospace, aeronautics, heavy machinery, and construction [1-3]. Over the years, numerous scholars have focused on analytical, finite element analysis (FEA), and, more recently, machine learning (ML) approaches to model the RARR process. Predominantly, the analytical and FEA models focus on predicting the ring expansion during the process and especially the forces arising due to contact between the ring and the various dies involved [4]. In this regard, one of the first attempts to model the force and torque in the RARR process was made by Hawkyard et al. [5]. They employed the slip-line method to create reference charts that correlated the thickness and outer diameter of flat rings with a pressure factor that defined the resistance to plastic deformation. This was subsequently employed in a SLAB formulation for force and torque estimations. Later, Mamalis et al. [6] improved upon this solution by directly correlating it with the contact length between the ring, mandrel, and main roll. More recently, Quagliato et al. [7, 8] further expanded this approach by providing a more precise estimation of the contact arc and a numerical evaluation of the slip-line field during both radial and axial deformations. Additional analytical formulations for force estimation in the RARR process are primarily based on the upper bound method [9, 10] or the SLAB method [11].

However, considering the complexity of the interactions between the ring and the dies in the ring rolling mill, as well as the strain, strain rate, and temperature-dependent material behavior, since their introduction, finite element simulations have been developed and progressively perfected throughout the years to model the RARR process. In this regard, an interesting analysis is presented in Yea et al. [12] where a custom FEA program, named SHAPE-RR$^{TM}$, was employed to predict ring spread, forming load, and contact pressure with remarkable accuracy compared to experimental results. Further studies saw an extensive use of pre-processing operations aimed at decreasing the simulation-to-reality gap as well as the computational time. For instance, Kim et al. [13] employed a combination of a Spatial Mesh System (SMS) and Actual Mesh System (AMS) to reduce the computational cost by having a fixed and fine mesh only in the deformation gaps, achieving good accuracy and a reduction in the computational time. A similar dual-mesh approach by Kim et al. [14] investigated the optimal number of subdivisions of the ring into discrete slices to determine the best meshing strategy to maximize accuracy while minimizing the computational time. A comprehensive work on FEA simulation setting for the RARR process was proposed by Wang et al. [15] where a set of rules to control all the tools in the ring rolling mill was employed as input for an optimization algorithm leading to a FEA-driven production with a reduction of 28% of the rolling time. A convergence between analytical and FEA approaches was afterwards proposed by Berti et al. [16] where a fully analytical set of rules for the set-up and control of the RARR process was proposed, from the billet optimal size estimation up to the initial and final feeding speeds' definition for the mandrel and the upper axial roll, subsequently used as the preprocessing algorithm in [7, 8].

As summarized thus far, although the ring rolling process has attracted scholars' attention for several decades, it remains a key manufacturing approach across multiple industrial sectors [17, 18]. Besides, due to recent advancements in artificial intelligence (AI), machine learning (ML), and concepts like smart manufacturing [19, 20], researchers have begun shifting their focus towards ML-driven modeling approaches for the ring rolling process. One of the first attempts for rolling in general is attributed to Yang et al. [21] where a classic approach of employing FEA simulation results as the training dataset for an Artificial Neural Network (ANN) model, demonstrating significant potential. In the same direction, Rohani Raftar and Parvizi [22] employed a modeling approach based on the ANN architecture coupled with Genetic Algorithm (GA) and Particle Swarm Optimization

(PSO) to determine the best process parameters to optimize both load and torque during the RARR process grounding their work on the results of FEA simulations as training dataset. Although ANN models are strong and reliable, as demonstrated in Mirandola et al. [23], ensemble approaches, such as Gradient Boosting and Extreme Gradient Boosting [24], are also a promising alternative for both force prediction and minimization in the RARR process.

Other interesting works concerning ML and RARR processes are represented by the modeling and prediction of the rolls' fatigue [25] and quality prediction by semi-supervised approaches [26]. In addition, smart FEA simulation approaches [27] and soft sensor-based feedback to control the microstructure evolutions during the process [28] have also been proposed in the literature relevant to the modeling for ring rolling processes of flat rings. In the context of RARR, Seitz et al. [29] investigated transfer learning strategies for exploiting limited target-domain data, including parameter transfer between rolling mills, autoencoder-based learning from unlabeled process data, and synthetic data-based model training.

As summarized so far, despite the scientific community's interest in both RARR and ML, two key challenges remain. The former is related to the need to generate a robust training dataset, irrespective of the ML model used [30]. The latter, and perhaps most critical, pertains to the minimization of the simulation-to-reality gap, which can vary when the ML model is applied to cases different from those it was initially trained upon.

To this end, recent studies in manufacturing and computational mechanics increasingly address these limitations by moving from isolated data-driven or physics-based models toward hybrid and simulation-informed learning strategies [31]. In parallel, the integration of ML with FEA has gained increasing attention as a means to accelerate simulations, support inverse analysis, optimize process parameters, and exploit numerical results as structured training data, although the reliability of such models remains strongly dependent on the quality of the simulation database [32, 33]. In the broader field of metal forming, FEA-driven ML approaches have been used for surrogate modeling, process design, and benchmark dataset generation, including GPR-based blank design in deep drawing [34] and parametrized FE datasets for sheet metal forming [35], confirming the value of simulation data while highlighting the need for robust, reproducible, and application-specific databases. Moreover, as previously mentioned, FEA-generated data in conjunction with TL approaches have been shown to compensate for scarce experimental data and improve model generalization, implicitly addressing the simulation-to-reality gap, core issue of simulation modeling [36-38]. Recent work on data-scarce datasets for stamping and RARR processes further supports this direction, showing that expert- or proxy-physics-informed learning strategies can improve prediction robustness beyond the available experimental–FEA training space [39, 40].

Our previous research demonstrated the feasibility of transfer learning from FEA simulations to experiments for a supervised prediction task in RARR, and examined its performance under varying data availability [40]. Building on these findings, the present study extends the investigation to unsupervised process representation in order to adopt a more general approach and enable the explanation of domain-specific effects. The study focuses particularly on domain adaptation, latent representations and the effects of targeted pre-processing. The study also compares hybrid and transfer learning across different data configurations, extending the analysis to a supervised TCRR use case to assess the applicability of these approaches to different ring-rolling processes and learning tasks.

The study first considers RARR and evaluates transfer learning from FEA simulations to experimental time-series data with a targeted domain-adaptation analysis aimed at quantifying the domains discrepancies and how to account for them. This task is formulated as an unsupervised representation-learning problem based on autoencoders, where the model reconstructs multivariate process time series including ring geometry, rolling forces,

and other key variables rather than predicting an explicitly labelled quality target. This was chosen to address a general ML task, enabling broad analysis of underlying effects and building a reliable environment for generalizing findings.

The investigation is then extended to TCRR, where FEA-only, experimental-only, hybrid learning, and transfer learning strategies are compared in a supervised prediction task for out-of-roundness. In doing so, this research clarifies the reciprocal role of numerical and experimental data in training pipelines and demonstrates that FEA data are not a direct substitute for real process data, and yet they can significantly improve ML performance when integrated through transfer learning or suitable database consolidation strategies.

Overall, the results indicate that FEA data can substantially improve ML-based ring rolling models when integrated through transfer or hybrid learning and targeted pre-processing, with consistent benefits observed from unsupervised RARR process representation to supervised TCRR out-of-roundness prediction, while also confirming that simulation data must be adapted rather than used as a direct substitute for experimental data. Fig. 1 presents the overall framework of this research.

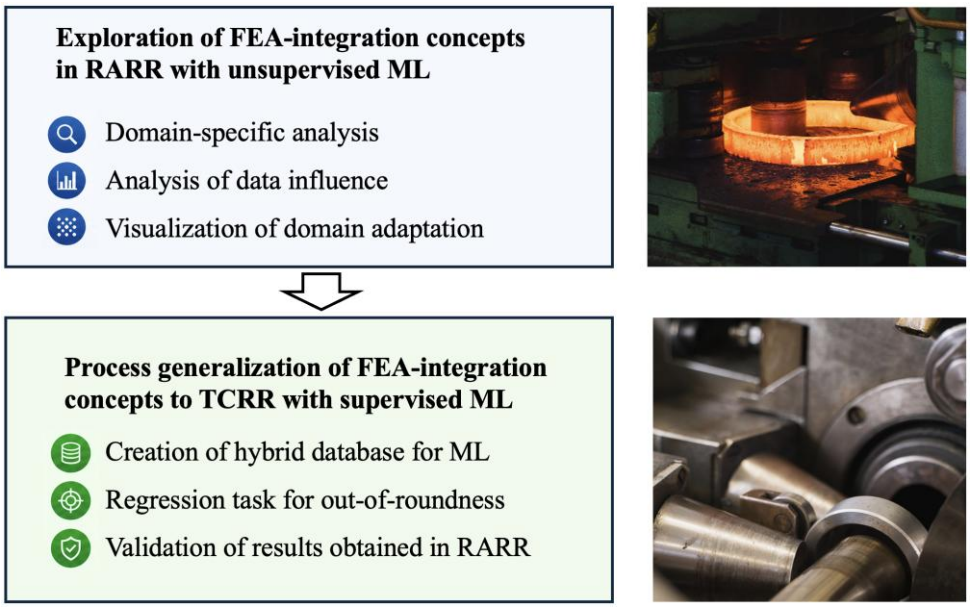



***Figure 1*** *– Schematic summary of the research implementation steps aimed at investigating the influence of FEA data in machine learning modeling relevant to the radial-axial and tangential cold ring rolling.*

## 2. Materials & Methods: Exploration of Transfer and Hybrid Learning in RARR

This chapter summarizes the technical aspects for the exploration of transfer and hybrid learning within RARR (first investigation, Fig. 1). To this end, this study develops an approach for accurately representing the radial and axial forming forces, ring diameter and thickness evolution, and process time with unsupervised learning-based autoencoders. The focus of this research is to analyze the effects of integrating FEA simulation data in machine learning modeling within a generalized task. The potential of learning from FEA simulations in a transfer and hybrid learning scheme is elaborated and broadly analyzed to study the effects. To this end, section 2.1 details the FEA simulations whereas section 2.2 focuses on the flat RARR experiments, section 2.3 provides an in-depth explanation of the implementation of machine learning. Finally, section 2.4 states the differences in integration of FEA data in modeling, specifically the transfer and hybrid learning. The research implementation is summarized in Fig. 2.

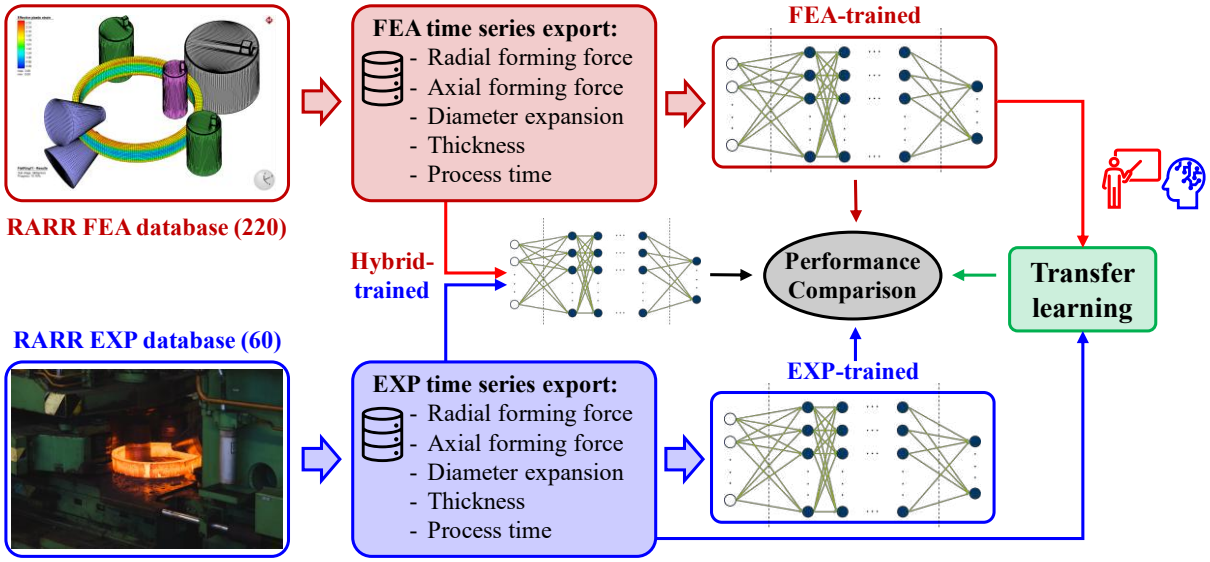



***Figure 2*** *– Schematic summary of the implemented research considering FEA- and EXP-trained models and the combined utilization in the transfer learning scheme.*

## *2.1 RARR FEA simulations*

To simulate the RARR process of flat rings, Fig. 3a, thermo-mechanical FEA simulations were implemented in the Simufact Forming 15, as shown in Fig. 3b, and are based on the results of a previous contribution of the authors [23], extensively validated against analytical solutions and experimental results [7, 8]. In the FEA simulations, a total of six dies were included and modeled as rigid bodies with heat transfer with both the environment and the ring. The tools' dimensions are reported in Table 1, together with the other simulation settings.

To expand the features' range, five ring materials commonly used in the RARR process were included in defining the FEA database. For each material employed in the setting of the FEA simulations, the strain ($\varepsilon$), strain rate ($\dot{\varepsilon}$), and temperature ($T$)-dependent Hensel-Spittel [41] flow stress model, as given by Eq. (1), was employed. The model constants for the flow stress of Eq. (1), defined in terms of $C_1$, $C_2$, $m_1$, $n_2$, $L_1$, $L_2$, $m_1$, and $m_2$, were derived from the MATILDA® (Material Information Link and Database Service) database available in Simufact Forming 15, and are reported in Table 2.

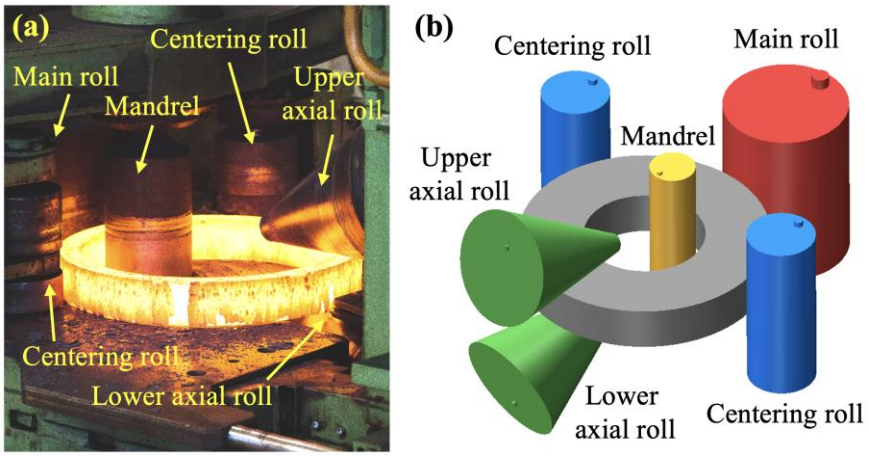



***Figure 3*** *– (a) RARR process experiment at the Chair of Production Systems of Ruhr-University Bochum and (b) FEA implementation by Simufact Forming 15 with details of the dies involved.*

***Table 1*** *– Tools' dimensions and common process settings for the RARR process FEA simulations [7, 8].*

| **Parameter** | **Value** |
|---|---|
| Main roll radius [mm] | 325 |
| Mandrel radius [mm] | 125 |
| Axial rolls' length [mm] | 596 |
| Axial rolls' vertex angle [°] | 35 |
| Centering rolls' radius [mm] | 175 |
| Tools' temperature [°C] | 150 |
| Environmental temperature [°C] | 50 |

$$\sigma_f = C_1 \cdot e^{C_2 \cdot T} \cdot \varepsilon^{(n_1 \cdot T + n_2)} \cdot e^{(L_1 \cdot T + L_2)} \cdot \dot{\varepsilon}^{(m_1 \cdot T + m_2)} \quad (1)$$

Concerning the ring dimensions, outer diameters from 429 up to 1993 mm were considered and randomly assigned to five materials, as summarized in Table 2, along with final height and thickness range to address different ring sections. The blank geometries were changed accordingly for each simulation to obtain a smooth and stable ring evolution during the total process [16], and the dimensions' ranges are listed in Table 2 as well.

In addition to that, considering the different typologies of materials, the initial temperature of the ring preforms assigned in each FEA simulation was set considering the range of validity of the Hensel-Spittel model of Eq. (1), as summarized in Table 2. As a final note, the process settings for the main roll, mandrel, and axial rolls were calculated according to [16] and are summarized in Table 2. Considering the huge volume of data, the detailed settings for each FEA simulation are not included here but are available on request to the corresponding author. In the FEA simulations considered in this research, friction was modeled according to Eq. (2) by setting the friction factor to m = 0.85 for the contact between the ring and the mandrel and the ring and the main roll whereas to m = 0.6 for the contact between the ring and both axial rolls and centering rolls [11, 42]. Considering different friction factors accounts for the fact that, due to the process settings, the friction developing between the main roll, mandrel, and the ring is higher than that between the ring and axial rolls or centering rolls [43].

$$\tau = m \cdot k = m \cdot \frac{\sigma_y}{\sqrt{3}} \quad (2)$$

FEA simulations were modeled with 3D arbitrarily distorted 8-nodes brick hexahedral elements (MARC® element type 7) having a mesh size of 16.0 mm (axial), 12.0 mm (radial), 16.0 mm (tangential), for a total number of elements depending on the initial size of the ring and no remeshing strategy during the process was employed. The MUMPS (MUltifrontal Massively Parallel sparse direct Solver) was employed for the solution of the RARR FEA simulations. In the MUMPS, the nodes' positions at each integration step are updated through a Lagrangian approach, suitable in case of large deformation, such as the RARR process. To control the contact between the ring and the tools, a 1/20 dimension to the element size criteria was considered, and, in case the relative velocity between two objects is less than 0.001 mm/s, a sticking condition is assumed. Though not directly implemented for the research presented in this contribution but gathered from a consolidated database [23], each of the 220 FEA simulations required a setting time of approximately 2 hours and a calculation time of 10 hours each. Summarizing, the FEA database employed in this research comprises a combination of different preform dimensions, final ring geometries, material types, temperature conditions, and processing parameters, all aimed at generalization of the modeling solution for the RARR process of flat rings.

***Table 2*** *– Tools' dimensions and process settings for the RARR process FEA simulations [23].*

| Category | Parameter | 42CrMo4 | AlMgSi1 | E355 | IN-718 | S355J2 |
|---|---|---|---|---|---|---|
| Hensel-Spittel flow stress model constants and validity range | C1 | 5290.5 | 378.5 | 2549.5 | 10501.1 | 2478.7 |
| | C2 | -3.70E-03 | -4.92E-03 | -3.4E-03 | -3.08E-03 | -2.98E-03 |
| | n1 | -3.34E-04 | -1.07E-04 | 1.9E-01 | -1.82E-04 | -4.24E-04 |
| | n2 | 2.06E-01 | -2.57E-02 | -4.9E-05 | 5.44E-01 | 3.66E-01 |
| | L1 | -8.27E-05 | 6.04E-05 | 1.1E-02 | -2.18E-05 | -7.51E-05 |
| | L2 | 2.89E-02 | -2.55E-02 | 1.2E-04 | 2.38E-02 | 3.16E-02 |
| | m1 | 3.01E-04 | 3.45E-04 | -4.3E-01 | -2.67E-06 | 2.61E-04 |
| | m2 | -1.56E-01 | -3.15E-02 | 3.0E-04 | 9.75E-02 | -1.37E-01 |
| | Validity range [°C] | 800~1200 | 250~500 | 800~1250 | 950~1200 | 800~1250 |
| Rings preform sizes and initial temperature ranges | Initial temperature range [°C] | 900~1200 | 300~450 | 950~1200 | 950~1150 | 950~1200 |
| | Outer diameter range [mm] | 518~990 | 549~897 | 429~593 | 491~898 | 400~440 |
| | Thickness range [mm] | 104~292 | 112~149 | 110~183 | 83~149 | 95~116 |
| | Height range [mm] | 112~386 | 112~219 | 190~325 | 112~219 | 400~431 |
| Final rings sizes | Outer diameter range [mm] | 649~1993 | 647~1427 | 521~920 | 593~1657 | 540~572 |
| | Thickness range [mm] | 62~233 | 62~100 | 81~120 | 59~100 | 70~90 |
| | Height range [mm] | 100~352 | 100~190 | 152~290 | 100~190 | 380~400 |
| Average Process settings | Main roll rotational speed [rad/s] | 2.9 | 3 | 2 | 3.2 | 1.7 |
| | Mandrel initial speed [mm/s] | 7.8 | 4.7 | 1.4 | 7 | 4.8 |
| | Mandrel final speed [mm/s] | 5.8 | 3.1 | 0.9 | 5.1 | 3.9 |
| | Upper axial roll initial translational speed [mm/s] | 9.7 | 6.5 | 3.8 | 6.2 | 0.5 |
| | Upper axial roll final translational speed [mm/s] | 7.1 | 4.2 | 2.5 | 4.1 | 0.3 |
| | Axial rolls' rotational speed [rad/s] | 9.3 | 9.6 | 5.6 | 10.5 | 4.6 |

## *2.2 RARR experiments*

The experimental dataset employed at different stages in this research was collected throughout various research projects at the Chair of Production Systems at the Ruhr-University Bochum. The ring rolling mill employed for the experiments is the ThyssenKrupp Wagner type 20/16-1000/160 with a maximal radial force of 200 kN and maximum axial force of 160 kN, whereas the ring dimensions under standard operating conditions are limited by 1 m outer diameter and a height of 160 mm. The ring rolling process is preceded by heating up of the preform, Fig. 4b, followed its setup in the RARR machine, Fig. 4c, until full process completion, Fig. 4d. The

experimental dataset consists of 60 steel rings. Every ring has multiple recorded time series, such as the time-dependent radial and axial forming force, the diameter and thickness of the ring, and the process time. Moreover, the database also includes time-independent scalars employed to identify the ring, namely the initial and final geometries, the material information, and the process parameters for each of the experimental cases. A summary of the key parameters is reported in Table 3.

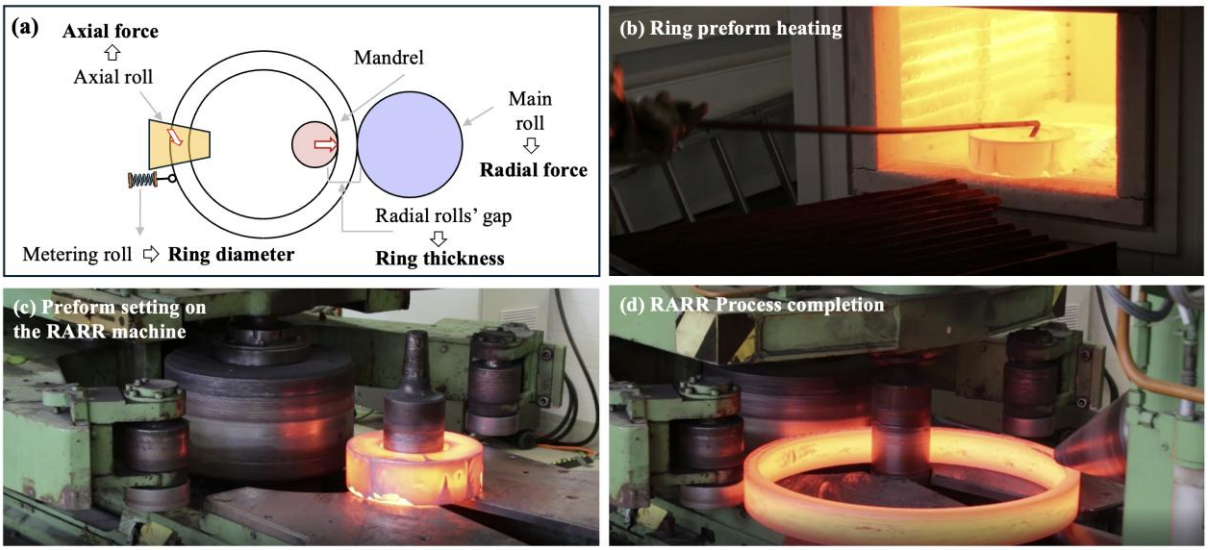



***Figure 4*** *– (a) Schematic representation of the signals' export from a ring rolling mill. (b) Ring preform heating, (c) set-up of the ThyssenKrupp Wagner type 20/16-1000/160 RARR machine, and (d) RARR process completion at the Chair of Production Systems at the Ruhr-University Bochum.*

***Table 3*** *– Summary of settings and ranges for the experimental dataset.*

| | | | |
|---|---|---|---|
| **Ring** | Temperature [°C] | Max | 1200°C |
| | | Min | not measured |
| | Diameter [mm] | Max | 999.55 |
| | | Min | 253.28 |
| | Thickness [mm] | Max | 57.32 |
| | | Min | 17.24 |
| | Height [mm] | Max | 73.09 |
| | | Min | 40.02 |
| **Tools** | Mandrel diameter [mm] | | 120, 130 |
| | Main roll diameter [mm] | | 450 |

## *2.3 Autoencoder modeling*

This section introduces the fundamentals of the ML architecture developed in this research, specifically the multi-layer autoencoders. The autoencoders were used to investigate the effects of FEA data integration in modeling, which is specified in Chapter 2.4. Autoencoders, first described by Rumelhart et al. [44], use neural networks to reconstruct input data and represent it meaningfully through unsupervised learning. Their primary function is to compress data to increase information density, followed by decompression to its original form. This process involves an encoder, a bottleneck (or latent space) where data is maximally compressed, and a decoder to restore the original data structure. In this research, three ML architectures are employed and following introduced: Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTMs) and Transformers.

In terms of reconstructing time series, only certain neural networks, such as RNNs, can process sequentially structured data sets [45] since they can map temporal relationships. This capability arises from recurrent connections within the network, which allow information from previous time steps to influence the current output. The backpropagation for updates in training is designed with backpropagation through time (BPTT), a variant of classic BP to count for time sequence inputs. However, classic RNN formulations can lose accuracy and modeling capability when considering long-term dependencies [46]. This issue, also known as the *vanishing gradient*, is addressed and reduced by the employment of LSTM architecture [48], which offers a way to capture long-term dependencies and reduce the effects of the vanishing gradient problem [49]. Thanks to their architecture, LSTM can selectively either retain or forget individual pieces of information from a time series and previous time steps. In LSTM, three components are crucial: the cell state $\bar{m}_t$, which acts as an internal memory and is updated after a new time step is received, the hidden state $\bar{d}_t$, which contains information from previous time steps $t-1$, and the gate functions, implemented in the form of activation functions $h$, mainly *sigmoid* functions. The gate determines whether the information is to be recorded in the cell state (input gate $\bar{\imath}_t$), whether is to be forgotten in the cell state (forget gate $\bar{f}_t$), or whether is to be forwarded to the next computing unit/time step (output gate $\bar{o}_t$) [49]. LSTMs network parameters are also optimized with BPTT.

Transformers, unlike RNNs and LSTMs, do not process sequences sequentially. Instead, they operate on all time steps in parallel and rely on positional encodings to inject information about token order into the model. These encodings enable the network to capture temporal or sequential relationships without recurrence, since the position is added for training. In addition, the self-attention mechanism allows the model to learn diverse relationships among elements in the sequence (here: timesteps in time-series data), attending simultaneously to different positions and to different feature subspaces [50].

This work employed autoencoder for the RARR process with five timesteps, meaning the networks aimed to learn relatively short-term relationships in the temporal sequences. The input shape for the training was defined as [*sequences (number of timesteps in the dataset), timesteps, features*]. The *features* part of the input shape vector contained the time series of the radial and axial force, the diameter and the thickness of the ring, and the computational or experimental time. Having fixed the timesteps to 5, the model received five timesteps in one vector and generates a reconstruction for these timesteps. The five timesteps have been chosen to analyze time-based inputs and to allow for possible deployment in a production environment, where predictions need to be made during operation. The term *sequences* refer to the number of instances passed to the model. Having 5 features and 5 timesteps, the input vector contained 25 data points. Within the autoencoder, the input space got compressed by 80%, to 5 latent space points. The autoencoders were trained with TensorFlow [47] and evaluated with the root mean squared error (RMSE), as well as the coefficient of determination ($R^2$), averaged across all five features. The

data were divided into training, validation, and test sets with proportions of 70%, 15%, and 15%, respectively. Hyperparameters were selected using Bayesian optimization with *Optuna*, iterating over 50 trials [51]. Moreover, before training, the features were normalized with *MinMax* scalers. Accuracies are presented for a 3-fold-cross validation, as well as the test data accuracies. To propose a summarized, yet informative, representation of the whole database, Fig. 5 shows the normalized mean experimental time series together with the FEA minimum and maximum forces correlated with the relevant minimum and maximum times, for the plotting of the FEA scope area. The radial forming force data are reported in Fig. 5a whereas those relevant to the axial forming force in Fig. 5b, respectively.

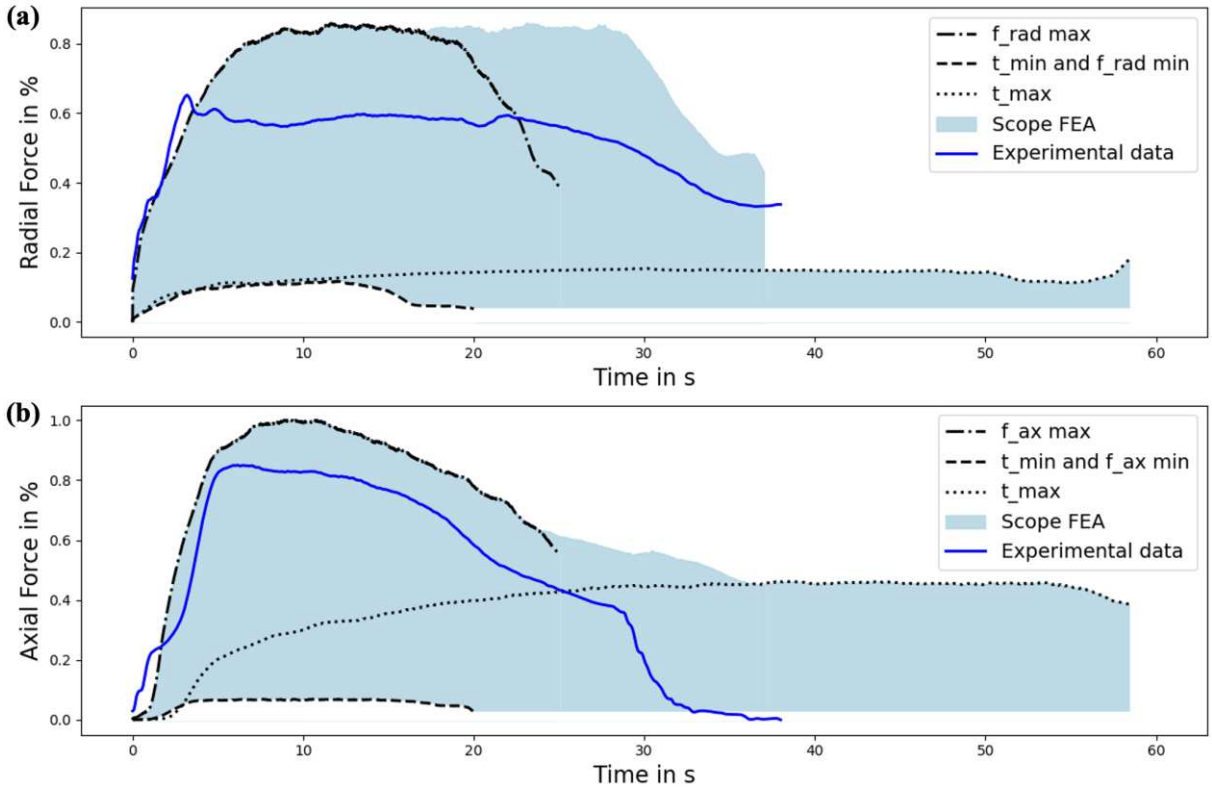



***Figure 5*** *– Comparison of the normalized experimental and simulation data of the (a) radial and (b) axial force. The blue line shows the mean course of the radial force, averaged by all experiments available.*

## *2.4 Training approaches to integrate FEA data*

To assess the role of FEA data in experimental-domain modeling, this study considers and compares how different learning strategies and dataset sources are incorporated into the training process, as follows:

1) **Experimental learning (EXP):** Learning using experimental data without FEA integration. This modeling approach will be considered the baseline and its models will be annotated with 'EXP'.
2) **FEA learning (FEA):** Learning with FEA data separately and test on experimental data. Target scaler of experimental data is used for experimental test data inference, to account for scaling differences.
3) **Hybrid learning (Hybrid):** Combining FEA data and experimental data in a combined training database. FEA time series are only added to the training data, validation and test data remain experimental.
4) **Transfer learning (TL):** Pre-Training of a FEA-model. Re-training of FEA model weights within experimental data. This transfer learning scheme is defined as *parameter transfer* [52] or known as *fine-tuning*. The learning rate is set to EXP training, while all layers are retrained. Therefore, none of the

model's parts are frozen. The base model (FEA) is selected based on the lowest RMSE across five runs, as tested on the EXP training and validation data. Test data is excluded from the selection process.

Fig. 6 summarizes the training approaches with respect to the utilized database and its training utilization. Moreover, the transfer learning approach is detailed for an LSTM unit-cell, showing the update of trainable parameters.

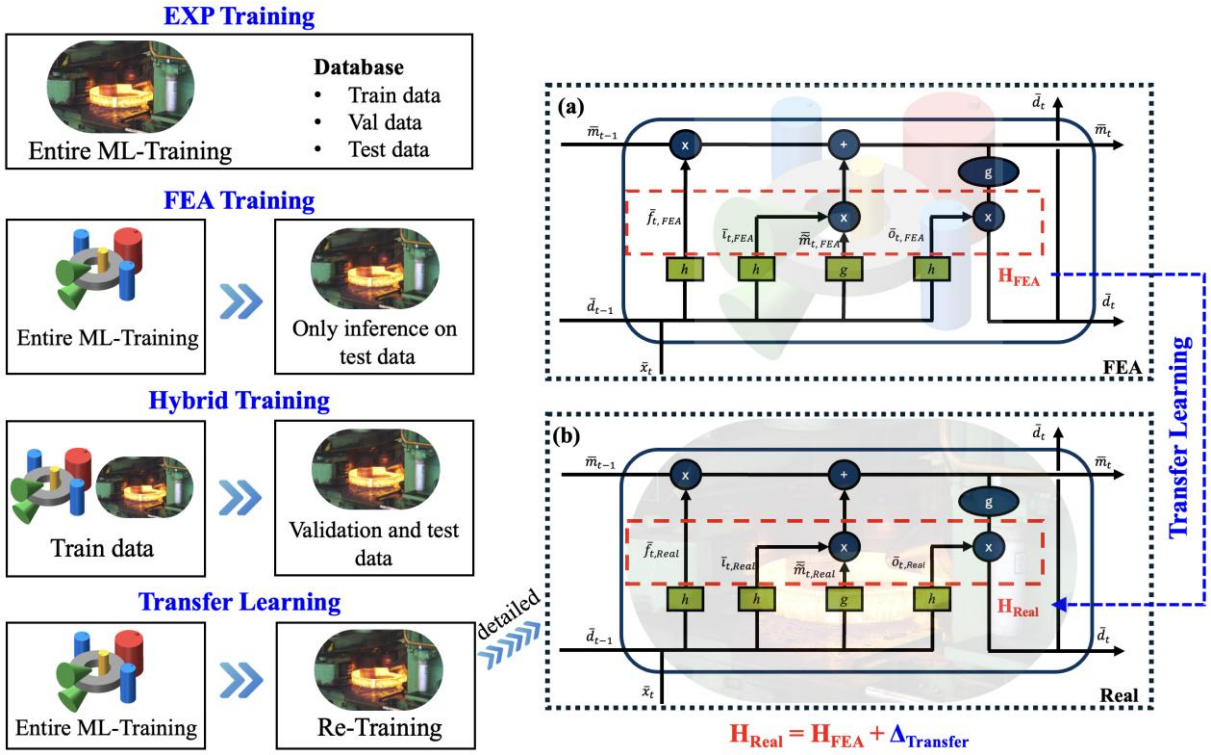



***Figure 6*** *– Visualization of training approaches used in this research. Transfer learning (TL) with parameter transfer is presented in detail for a LSTM unit cell.*

These training methods were used for all three architectures: LSTM, RNN and transformer. Besides the effects of the architecture chosen, this setup allowed especially the comparison between the training strategies and aiming to address the research question of this section: *How to learn from FEA simulations in RARR for experimental modeling in machine learning?* Firstly, this question was addressed by analyzing a setup where rare experimental data is available. To enable a consistent comparison of learning strategies and model architectures, a subset of five experiments was initially selected from the 60 available experimental cases (5 EXP). Subsequently, the full set of 60 experiments was considered to investigate the effect of increasing experimental data availability and to assess how database size influences the integration and contribution of FEA data within the ML framework.

## 3. Results: Exploration of Transfer and Hybrid Learning in RARR

### *3.1 Autoencoder training on experimental data*

The training of the autoencoder model is applied to the database composed of only experimental (EXP) results with five experiments (5 EXP) using LSTMs, RNNs and transformer models to simulate a scenario where few experimental data are available. The three different modeling techniques yield slightly different performance levels, as presented in Table 4. The presented RMSEs are dimensionless, due to the different scales of each feature. Moreover, the optimized hyperparameters (Bayesian optimization) for each model are presented in Appendix A.

As shown in Table 4, the models considered exhibit comparable performance, with RMSE differences within ± 0.003. Similar error levels are observed for both cross-validation and test datasets, indicating no significant overfitting. To facilitate interpretation of the scaled RMSE values, Fig. 7 provides a visual comparison between predicted and measured time series, while the corresponding RMSE values in the original units of each feature are reported for the transformer model.

***Table 4*** *– Performance of different models for five experiments (5 EXP). The accuracy is averaged and stated with its standard deviation. The best performing approach is bold highlighted.*

| | Architecture | Cross validation [-] $\pm \sigma$ | Test data [-] $\pm \sigma$ |
|---|---|---|---|
| RMSE | RNN | $0.038 \pm 0.008$ | $0.040 \pm 0.003$ |
| | **LSTM** | $\mathbf{0.036 \pm 0.006}$ | $\mathbf{0.038 \pm 0.005}$ |
| | Transformer | $0.037 \pm 0.011$ | $0.041 \pm 0.004$ |
| $R^2$ | RNN | $0.98 \pm 0.011$ | $0.94 \pm 0.014$ |
| | LSTM | $\mathbf{0.98 \pm 0.009}$ | $0.95 \pm 0.025$ |
| | Transformer | $0.98 \pm 0.019$ | $\mathbf{0.95 \pm 0.005}$ |

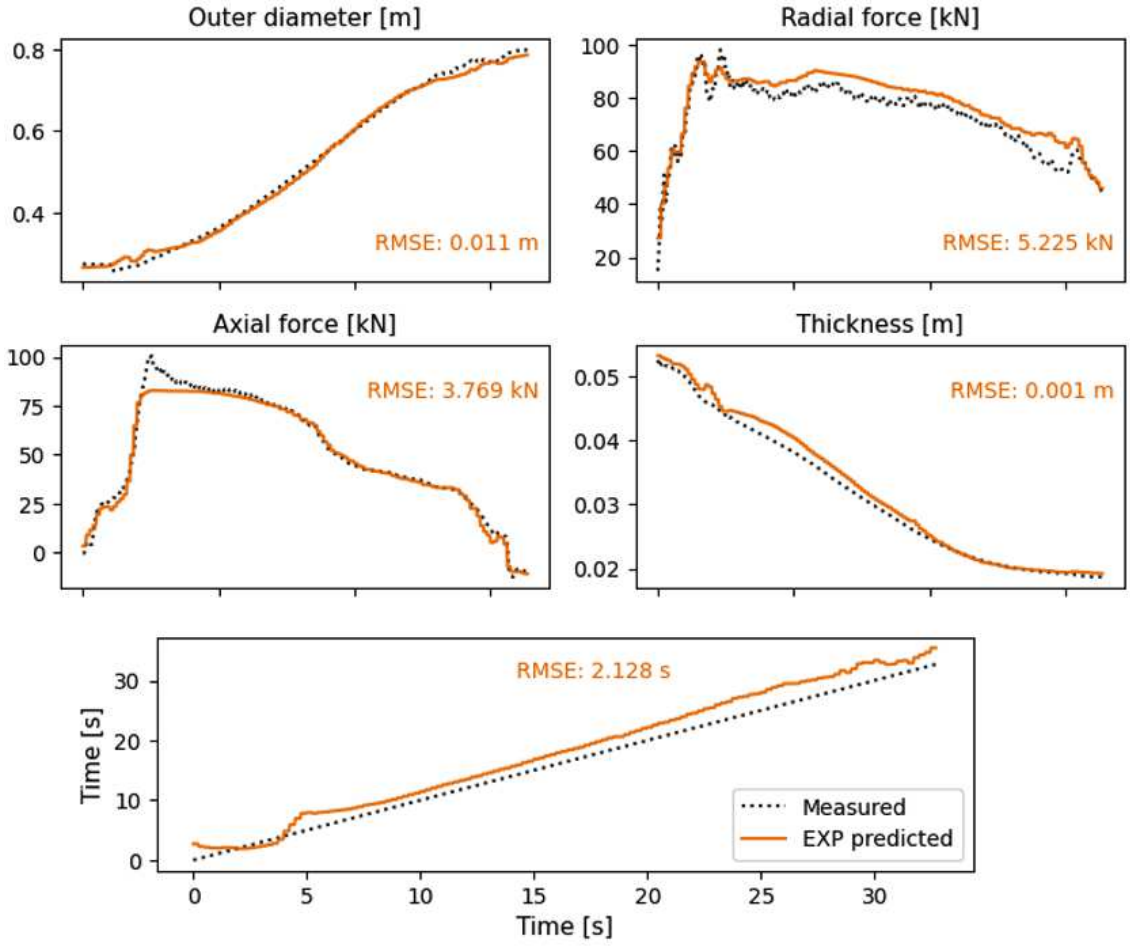



***Figure 7*** *– Predictions of the transformer baseline model compared to the measured time series data (test data). Moreover, the RMSE is assigned in its original units.*

The overall course of the targets is well predicted, although local errors and offset persist. For instance, the axial force does not meet the maximal force accurately and for the radial force the force is predicted to be too high. Moreover, the time prediction is for most timesteps predicted too great. Therefore, the integration of FEA data is

sought, to investigate its influence on predicting the time series. Especially the different data scales in its original metrics (refer to section 2.4) sets a challenge on transferring knowledge to the experimental cases. Thus, three different approaches are investigated on how to integrate the FEA data in the model.

### *3.2 Integration of FEA data for modeling using transfer and hybrid learning*

This section compares three different approaches to enhance training with FEA data: training only with FEA data (FEA), training within a hybrid database (Hybrid) and applying a transfer learning scheme (TL) (compare section 2.4, training concepts). Hyperparameters kept the same to the optimized, as presented in Chapter 3.1, to investigate the direct influence of FEA integration. The results for the test data are presented in Table 5, whilst the results for cross validation are stated in Appendix B.

***Table 5*** *– Performance of different models for five experiments (5 EXP) for the test data. The accuracy is averaged and stated with its standard deviation. The best performing approach is bold highlighted.*

| | Architecture | EXP | FEA | Hybrid | TL |
|---|---|---|---|---|---|
| RMSE | RNN | $0.040 \pm 0.003$ | $0.085 \pm 0.030$ | $0.050 \pm 0.004$ | $0.018 \pm 0.001$ |
| | **LSTM** | $\mathbf{0.038 \pm 0.005}$ | $\mathbf{0.058 \pm 0.021}$ | $\mathbf{0.033 \pm 0.003}$ | $\mathbf{0.015 \pm 0.001}$ |
| | Transformer | $0.041 \pm 0.004$ | $0.128 \pm 0.019$ | $0.061 \pm 0.014$ | $0.018 \pm 0.001$ |
| $R^2$ | RNN | $0.94 \pm 0.014$ | $0.80 \pm 0.121$ | $0.92 \pm 0.015$ | $0.991 \pm 0.001$ |
| | **LSTM** | $\mathbf{0.95 \pm 0.025}$ | $\mathbf{0.90 \pm 0.109}$ | $\mathbf{0.97 \pm 0.009}$ | $\mathbf{0.993 \pm 0.001}$ |
| | Transformer | $0.95 \pm 0.005$ | $0.64 \pm 0.120$ | $0.90 \pm 0.053$ | $0.991 \pm 0.001$ |

Table 5 reveals that FEA-trained models cannot outperform EXP models. With regards to performance, LSTM stands out as best performing on the experimental dataset (RMSE: 0.038) whilst the transformer and RNN models offer greater prediction errors. This could be caused by the differences in ring dimensions (compare Chapter 2.3) and will be further picked up in the discussion. By combining the databases for hybrid training, improvements can be seen in comparison to training with FEA only for all models. Still, transformers and RNN models do not outperform the EXP models, whilst LSTM leads to a slight improvement (RMSE: 0.033). By integrating the FEA data with a TL-scheme, the lowest RMSE can be found for all models. The RMSE reduces to 30-40% in comparison to the experimental and hybrid models. The best model performances are reached once again by LSTMs (0.015) followed by RNNs and the transformers (both 0.018). LSTM models outperform RNNs and transformers, particularly in FEA and hybrid modeling, but also for experimental and TL modeling. Therefore, LSTM seems to be the most suitable architecture for the presented ML task. Summarized for the LSTM model and its performances tested on the experimental dataset, it can be brought forward, that:

- FEA modeling inherits greatest prediction errors
- Integrating FEA into a hybrid training scheme can improve training and slightly outperforms the EXP model using the LSTM architecture (from 0.038 to 0.033)
- TL significantly reduces the RMSE compared to hybrid and EXP training (from 0.038 to 0.015). The reasoning behind the introduced effects is analyzed in more detail in Chapter 3.4.

To assess the improvements of TL in comparison to EXP modeling, Figure 8 captures the predictions in a regression diagram, where predicted values are plotted over real (measured) values. Therefore, an optimal prediction would follow a diagonal line (45°) in the scatter plot.

The TL model catches accurately the measured time series, as can be seen in Fig. 8, where for most of the timesteps the predicted values are located close to the diagonal line. One particular characteristic, the maximal axial force, is still not met with precision (Fig. 8, right). This is the only observable deviation with the TL model. However, the RMSE remains lower than the EXP one (reduction from 3.910 kN to 2.584 kN). In comparison to the EXP model, the predictions are much closer towards the measured values, especially for the radial force and time predictions. This is also quantified in the RMSE for each time series (compare Fig. 8), such as the reduction from 4.975 kN to 1.708 kN for the radial force with TL. The closer alignment of the TL model predictions in Fig. 8 indicates that transfer learning enables the autoencoder to learn a more informative representation of the underlying time series within the same train data. This suggests that the bottleneck latent space captures the relevant temporal characteristics more effectively, resulting in improved reconstruction performance. Since the architecture and hyperparameters remain unchanged, the performance gain can be attributed to the knowledge transferred during pre-training on the FEA dataset rather than to differences in model capacity. The implications of this finding for learning ring rolling characteristics are discussed in more detail in the Discussion section.

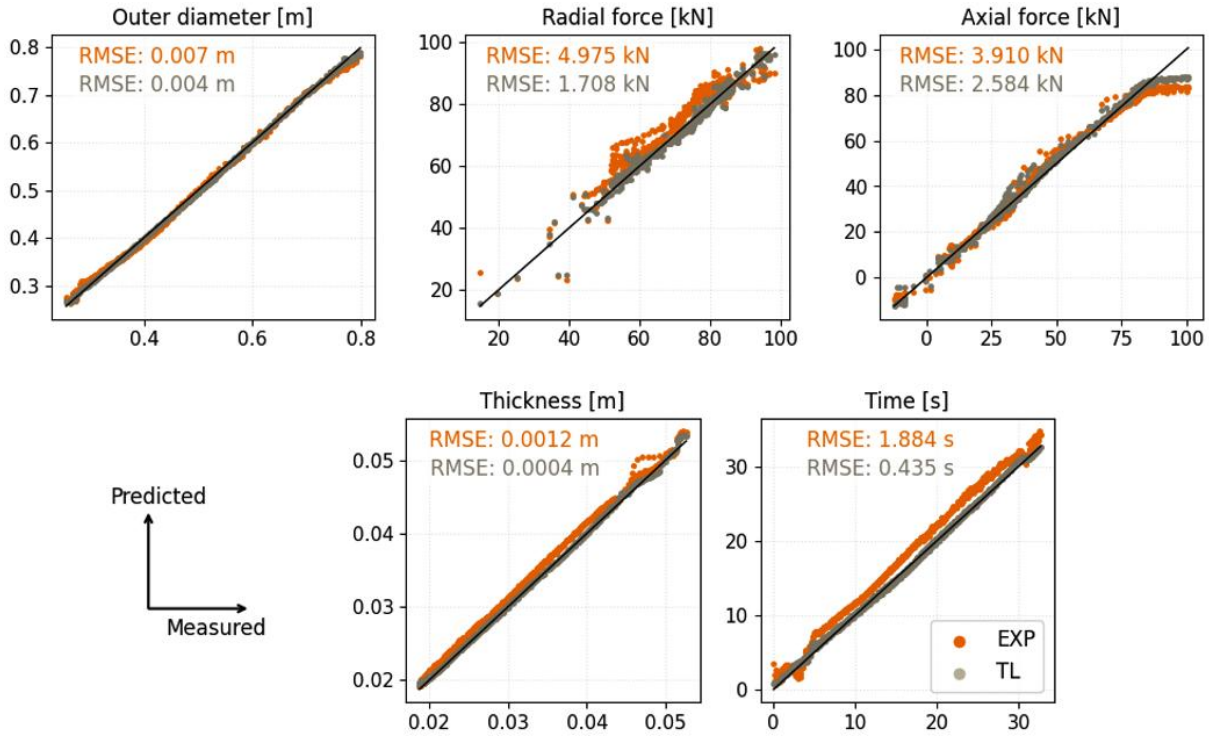



***Figure 8*** *– Predictions of the EXP model compared to the measured time series data and the TL model. Moreover, the RMSE is assigned in its original units for both ML (EXP and TL) approaches.*

As the results showed, TL offers a way to customize FEA models for experimental cases. Hybrid learning can be utilized as well, and depending on the architecture, be beneficial. For the present ML task, more precisely the model representation through autoencoders, TL shows the most accurate prediction results. The investigation was designed with respect to a few experimental cases (5 EXP) to address scenarios where little data is available, and to learn from the available knowledge. This has been demonstrated in this section. The next chapter analyses what happens when more experimental cases are available, and to what extent TL in these cases offers value.

## *3.3 Variational analysis – Influence of experimental cases on TL*

When increasing experimental data amount, hence having more cases to train on, the relations derived in the previous chapter may change. Therefore, starting with a few experiments, experimental cases are added to the train

database, and prediction accuracies are stored. Test data remains through all splits the same and equals 11 cases, unseen data in the training process. Starting from three cases for the training base (train and validation), cases are iteratively added until 49 cases are reached. This is the maximum amount of data available in this work (49 train + validation, 11 test). For this analysis, only TL in comparison to baseline training is displayed, since it showed best performances in the previous section. LSTM is chosen as the architecture for the same reason. Fig. 9 shows the trend of prediction accuracies (average and standard deviation) over the number of experimental cases in the train database.

As already presented in section 3.2 and visible in Fig. 8, TL outperforms EXP training when few cases are available. The RMSE is reduced here to approximately 50%. When extending data available for training and validation, this behavior remains, until 30 added experiments are reached. From this threshold on, EXP training led to comparable average prediction errors, while even inheriting fewer standard deviation. Therefore, above this threshold, it appears, that FEA integration is even reducing slightly reproducibility of ML modeling (standard deviation), which must be related to the domain gap (simulation-to-reality gap, Sim2Real). Although, with fewer experimental cases TL enables significantly better performance. The persistent domain gap after TL decreases improvement by adding experimental cases after a threshold of 30 experiments.

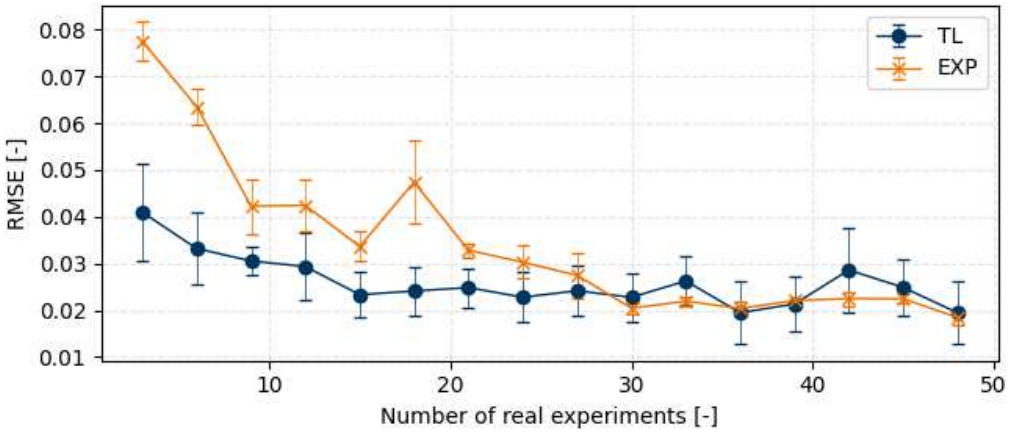



***Figure 9*** *– Variation of real experiments available for training. Presented is the RMSE of ML predictions in respect to the measured values for the EXP and the TL approach. Test data remains the same with 11 cases (approx. 15% of all cases) for all splits.*

A pre-processing step applied to the FEA data can potentially reduce the initial domain gap. Here, this is achieved by applying a *MinMax* scaler to each FEA time series individually, mapping the minimum and maximum of each feature to 0 and 1, thereby removing geometry-specific information. In summary, this pre-processing pipeline decouples the FEA data from its geometric characteristics. Returning to the variational study (Fig. 9), pre-processing visibly improves transfer learning (TL) performance. In particular, as more experimental samples are included, TL consistently outperforms EXP-only training across all configurations (see Fig. 10). This demonstrates that targeted data pre-processing can enhance TL, likely by reducing the initial domain gap between the FEA and experimental datasets.

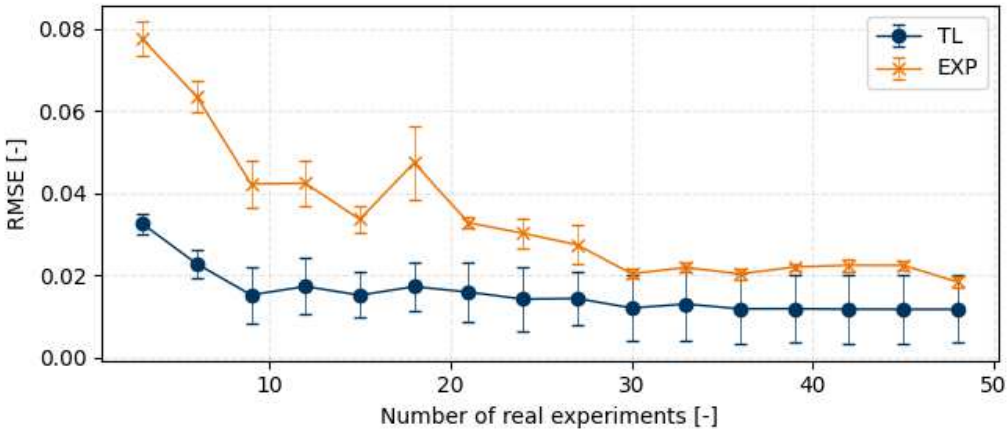



***Figure 10*** *– Variation of real experiments available for training, while pre-training on pre-processed FEA data. Presented is the RMSE of ML predictions in respect to the measured values for the EXP and the TL approach. Test data remains the same with 11 cases (approx. 15% of all cases) for all splits.*

## *3.4 Effects of TL*

After exploring the positive influence of TL on integrating FEA data, especially with LSTMs, this section aims to provide information on transfer characteristics. For this, the modeling is shifted away from representing local patterns with 5 timesteps, towards global patterns. These global patterns aim to represent the entire ring rolling time series. For this, all-time series (FEA and experimental) are resampled along the time axis to 500 timesteps. The sequence length of LSTM input is extended from 5 to 500, meaning that one entire time series is one input. To maintain compatibility, the same data is used as before in Chapter 3.2 with five experimental and 220 FEA time series. To visualize the effect of TL, following steps are implemented and calculated:

1. Training of a FEA-model;
2. TL to experimental data. Fine-tuning only of the decoder, encoder remains with weights trained of the FEA-training. TL employes the *domain shift*, so it shifts from FEA dataspace to experimental dataspace;
3. Re-apply TL-model to FEA data. This model learned the domain shift and therefore applies the domain shift back to FEA data. Thus, TL model is named in this context *domain shift model*.

The reason for only re-training the decoder (2), is to retain input-space of FEA-data, so it can be re-used on FEA data at a later point. Applying the domain shift model (3) to FEA data, its data distribution is theoretically changed to the target (experimental) distribution. These effects are analyzed further with two series: the mean FEA- and experimental curves. These curves are calculated as the means of the entire databases, representing an average ring rolling behavior. The domain shift model transforms the mean FEA curves to a different time-depending course. This is plotted in Fig. 11 for the radial force and the outer diameter. The time series follows more the trend of the experimental data. This can especially be observed for the force, where the initial mismatch has been greater and the transformed time series matches more closely the experimental data. For all process parameter time series, Table 6 states the $R^2$ of FEA to experimental average curves, as well as the domain shifted mean curves. The $R^2$ is significantly maximized for the forces, whilst the geometrical time series show less improvement in $R^2$.

For the outer diameter, the original FEA data already achieved an $R^2$ of 0.99. However, as visually presented in Fig. 11, the characteristic time-dependent change of non-linear ring growth in the main rolling phase between the $100^{th}$ and the $400^{th}$ timestep is caught better with the domain shifted model. This concludes that also for

geometrical expansion, the domain shift model changes temporal behavior closer to reality. The presented transformation of FEA data presents visually the effect of TL: adapting time dependent behavior and characteristics closer to the target (experimental) distribution.

***Table 6*** *– Coefficient of determination ($R^2$) from FEA, domain shifted FEA and experimental data.*

| Approach | Outer diameter | Thickness | Axial force | Radial force |
|---|---|---|---|---|
| EXP | 1 | 1 | 1 | 1 |
| FEA to EXP | 0.99 | 0.97 | 0.63 | 0.1 |
| FEA domain shifted to EXP | 0.99 | 0.98 | 0.90 | 0.97 |

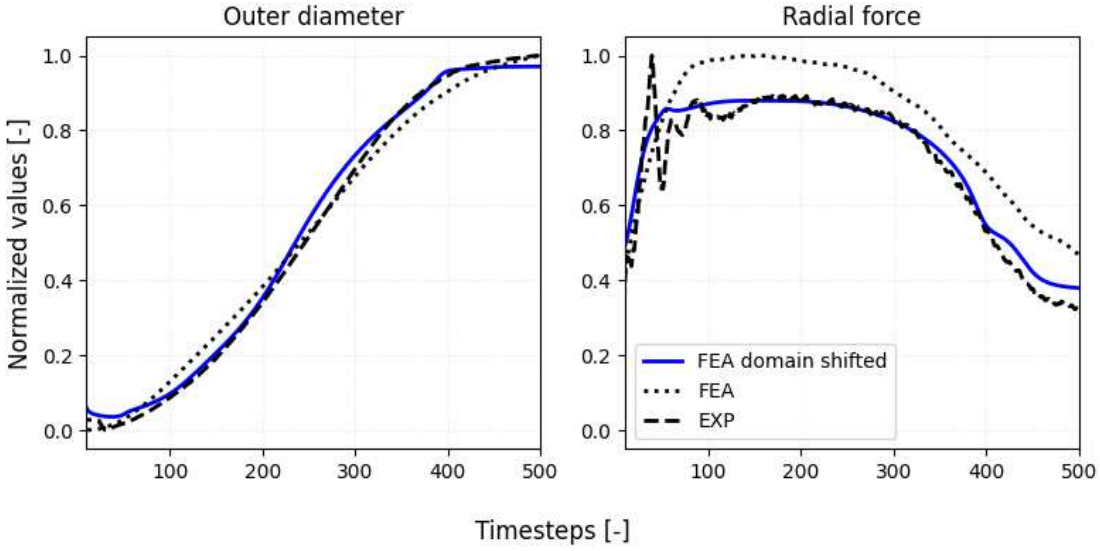



***Figure 11*** *– Domain shift of FEA data towards experimental data, represented by the average curves. Shift is achieved by inverse applying a TL autoencoder to its source (FEA).*

## *3.5 Summary*

The results reported in the various sub-sections of this chapter show that integrating FEA data in ML modeling can be beneficial. For the application presented with five experimental cases for system representation via autoencoders, training only on FEA data does not show a benefit in comparison to training on experimental data (LSTM test RMSE: EXP 0.038 ± 0.005 vs. FEA 0.058 ± 0.021). For this, hybrid learning and TL offer a suitable way for integrating data (LSTM test RMSE: Hybrid 0.033 ± 0.003 vs. TL 0.015 ± 0.001). Here, TL outperformed hybrid learning, but in general both integrations worked. Moreover, LSTM outperformed RNNs and transformers for all models. By analyzing the effect of TL, when experimental cases available are increased, it appears that after 30 experiments added, a threshold is reached. Here, TL shows no more improvement in comparison to EXP training. By pre-processing FEA data, this threshold has not been reached with all 60 experiments available. This pre-processing removes information that does not occur in the experimental database, which relates mainly in these cases to the geometrical spectra of pre- and final forms and the process time. For instance, final ring geometries in the FEA data can yield 2 m, whilst experimental cases stay around 1 m. As stated, this pre-processing improved TL performance, especially when the experimental database extends. To this end, the effect of TL can be visualized by training a domain shift model, that learns the simulation-to-reality gap from FEA to EXP, whilst keeping weights of FEA (encoder). Applying this model inversely to FEA data, it shifts its temporal relationships towards

the target distribution of EXP data (e.g., the $R^2$ increase in radial force is from 0.1 to 0.97.). This effect proves the capability of domain bridging with TL.

The presented results aim to provide an understanding of FEA integration in ML modeling for RARR. This is motivated by the data landscape occurring in ring rolling, where FEA data is widely used for process design and research, whilst experimental cases can be rare. Beyond TL exploration, the domain shift model derived in this work, can be used theoretically to post-process FEA data, to align it closer to reality. As in this research, only a limited number of experimental cases, such as five, are needed to perform this. This can also take account for aligning data between different ring rolling mills, geometrical spectra and so on. Moreover, in theory, the autoencoder modeling, that was used for exploring TL, could be applied in various ways in industry. For instance, anomaly detection can be performed, where the autoencoder outputs greater predictions errors, since the anomaly is out-of-distribution. It must be stated though that this is not the objective of the investigation presented and is therefore not further presented. An example of an anomaly detection for instabilities in ring expansion, performed with the trained autoencoders of this research, can be found in Appendix C.

To broaden the study of FEA integration in ring rolling modeling, the transferability of the concepts is examined within a different ring rolling process (**process generalization**) here instantiated by the smaller-scale TCRR with radial-only forming. Furthermore, the machine learning task shifts from a generalized perspective (**representation learning**) toward a supervised learning use case (**application**).

## 4. Process generalization and application in TCRR

The previous sections described the application of transfer learning and hybrid training in RARR, where the effectiveness of the proposed methodology was systematically validated. Building upon these findings, transfer and hybrid training were further applied to tangential cold ring rolling (TCRR) for the prediction of form errors. Compared to RARR, TCRR is primarily employed to manufacture seamless rings with smaller dimensions, typically ranging from a few centimeters to several decimeters in diameter. To reduce material consumption and energy waste, recent developments in data-centered manufacturing methods aim to predict product quality either prior to or at early stages of the manufacturing process. ML applications generally require a large amount of data, which can be obtained from experiments, FEA simulations, and analytical models. To reduce the number of costly experimental trials, simulation and analytical approaches are often used to augment the available datasets. However, based on preliminary investigations, a direct combination of data from different sources leads to limited prediction accuracy. In contrast, significantly improved performance can be achieved by introducing transfer learning, as discussed in the previous sections. One major form error in ring rolling is the out-of-roundness (or non-circularity) of the final ring shape, caused by critical bending moments [53]. Within the TCRR setup used in this study, this form error is the most dominant and is therefore selected as the supervised learning target. The application presented in the following successfully demonstrates the effectiveness of transfer and hybrid learning to predict form errors in TCRR using a hybrid database consisting of experimental and synthetic FEA data.

### *4.1 TCRR simulations and experiments*

First, the working principles of TCRR are briefly introduced. In this study, all experimental investigations were carried out on a retrofitted UPWS 31,5.2 from Profiroll Technologies GmbH, as shown in Fig. 12. The ring blanks were manufactured from the 100Cr6 steel (1.3505, AISI 52100), a material commonly used in the bearing industry due to its high hardness and excellent wear resistance.

During the forming process, the main roll is controlled to move towards the mandrel at a defined feed rate while rotating at a specified speed, which causes an expansion of the ring diameter and a reduction in the ring thickness. The mandrel is free to rotate with the ring while the guide rolls serve to stabilize the ring during the process. Only the final ring thickness is determined by the indentation geometry of the rolls. The forming process terminates when the main roll reaches the predefined target position.

The machine is equipped with different sensors to measure the process parameters, including main roll displacement, ring growth, rolling force and rolling torque. The displacement of the main roll is recorded with a linear displacement transducer, ring growth is measured using two metering rolls and a potentiometer, while both rolling force and torque are measured using strain gauges. All measurement signals were acquired via a data acquisition device (HBM QuantumX).

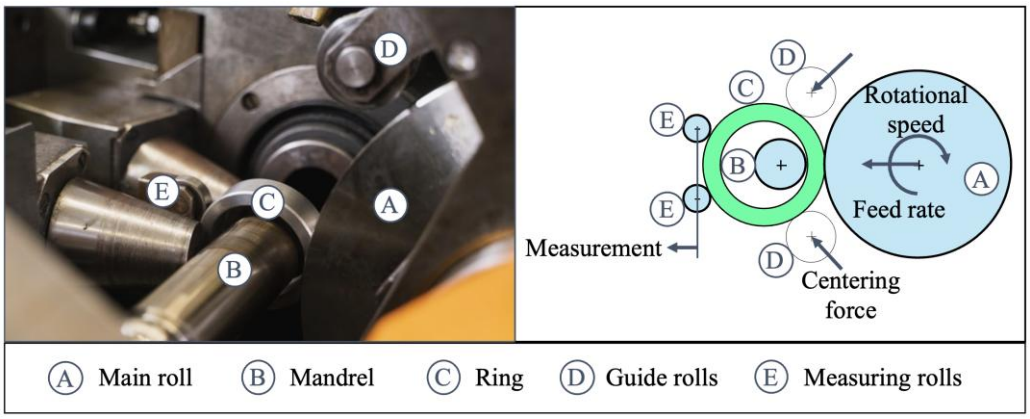



***Figure 12*** *– Photograph of the TCRR machine located at the Technical University of Dresden (left) and Schematic illustration of the process (right), adapted from [54].*

As previously mentioned, experimental data can be augmented by synthetic data. A 3D FEA model for TCRR was developed and calculated using the commercial software LS-DYNA version 14.0.0. The material properties were determined by performing compression tests on cylindrical samples at room temperature. The compression tests were conducted on cylindrical specimens (Ø 5.8 mm x 6.3 mm) following the guidelines of DIN 50106. Deformation of the compression samples was recorded using an ARAMIS 4M system for digital image correlation. The Voce hardening model was employed to represent and extrapolate the material behavior, expressed as $\sigma_p = 518.64 + 337.78 \cdot (1 - e^{-20.27 \cdot \varepsilon_p})$. Since the temperature during cold TRR does not exceed 100 °C, thermal effects were neglected.

In the FEA simulations, the ring was discretized into approximately 13k solid elements using brick solid element with 8 nodes and one integration point, ELFORM 1 in LS-DYNA, while a mass time scaling factor of 2 was used to further reduce computational effort. The contact conditions between the ring and the tools were modeled using the Coulomb model with a friction coefficient of 0.15, calibrated by the experimental results. The simulation was implemented for an average of approximately 330 core hours, utilizing 32 cores on an Intel® Xeon® Platinum 8470 processor within the high-performance computing system of the Technical University Dresden. The relevant process parameters were extracted from the simulation results. Fig. 13 compares the FEA results with experimental measurements for three different feed rates. The deviation between simulation and experiment is within approximately 15%, while the overall trends of the simulations are in good agreement with the experimental observations, indicating the reliability of the FEA model.

Within the framework of this research, a hybrid database was generated. The experimental design was structured into two phases: a screening phase and a data generation phase. The primary objective of the screening phase is to identify the factors that significantly influence the final product quality. Therefore, a full two-level five-factorial plan was performed for the first phase. Based on the data analysis from this phase, three key factors were selected for the subsequent data generation phase. In this second phase, a Central Composite Design (CCD) with three repetitions was employed to generate data suitable for training the machine learning model. The detailed experimental design is summarized in Table 7. For every parameter configuration, the corresponding FE simulation was carried out.

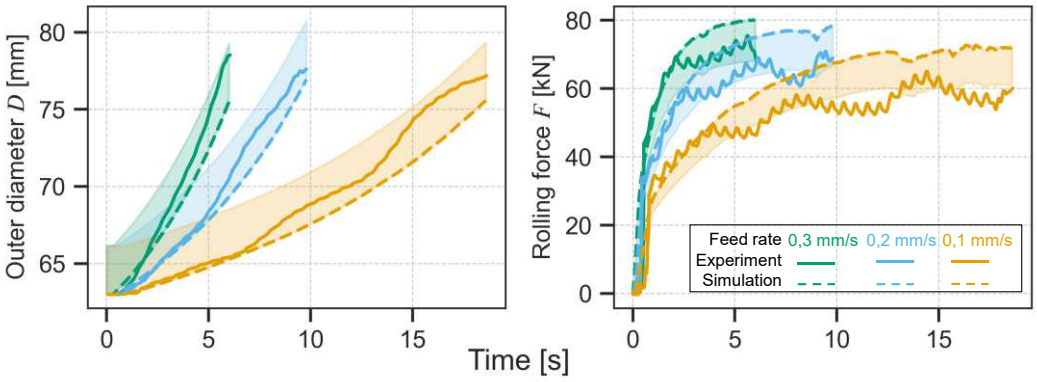



***Figure 13*** *– Validation of the 3D FE model: Comparison of FE and experimental results at three feed rates.*

***Table 7*** *– Data generation plan using design of experiments (DoE).*

| Factors | Screening phase | | Data generation based on CCD | | | | |
|---|---|---|---|---|---|---|---|
| | Level | | Level (coded) | | | | |
| | - | + | -2 | -1 | 0 | 1 | 2 |
| Ring outer diameter [mm] | 61 | 63 | 60 | 60.75 | 61.5 | 62.25 | 63 |
| Feed rate [mm/s] | 0.1 | 0.3 | 0.05 | 0.1 | 0.15 | 0.2 | 0.25 |
| Rotational speed of main roll [rpm] | 100 | 125 | 50 | 63 | 80 | 100 | 125 |
| Roll gap [mm] | 4.2 | 4.4 | - | - | - | - | - |
| Guide roll pressure [bar] | 7 | 15 | - | - | - | - | - |

## *4.2 Post process measurement of out-of-roundness*

The process configurations presented in Table 7 serve as features for ML modeling of the out-of-roundness. With respect to the target labels, the final ring geometry is measured post-process using an ATOS Q 255 MV170 optical measurement system. The measurement system exhibits a maximum uncertainty of 3 µm in the acceptance test in accordance with VDI/VDE 2634. In the FEA simulations, the final ring geometry can likewise be exported as a point cloud, as shown in Fig 14. Based on this representation, the form errors are calculated using the algorithm established by Wang and Seitz et al. [55].

After importing the ring as a point cloud, the ring is usually not aligned to the Cartesian coordinate system. Therefore, additional data preprocessing is required to ensure a consistent calculation of the form errors. For this purpose, an optimization-based alignment algorithm was developed, which aims to minimize the absolute distances in the vertical direction (z-axis), so that the lower ring surface is aligned with the xy-plane. Once the ring is aligned within the plane, the form error calculations can be performed. This calculation accounts for all

deviations from an ideal circular shape, defined as ($d_{\mathrm{k,ideal}} - d_{\mathrm{k,measured}}$). The proposed formula for out-of-roundness (non-circularity) $\varepsilon_r$ is given in Fig. 14, where n denotes the number of calculation points. Further information is available in Wang and Seitz et al. [55].

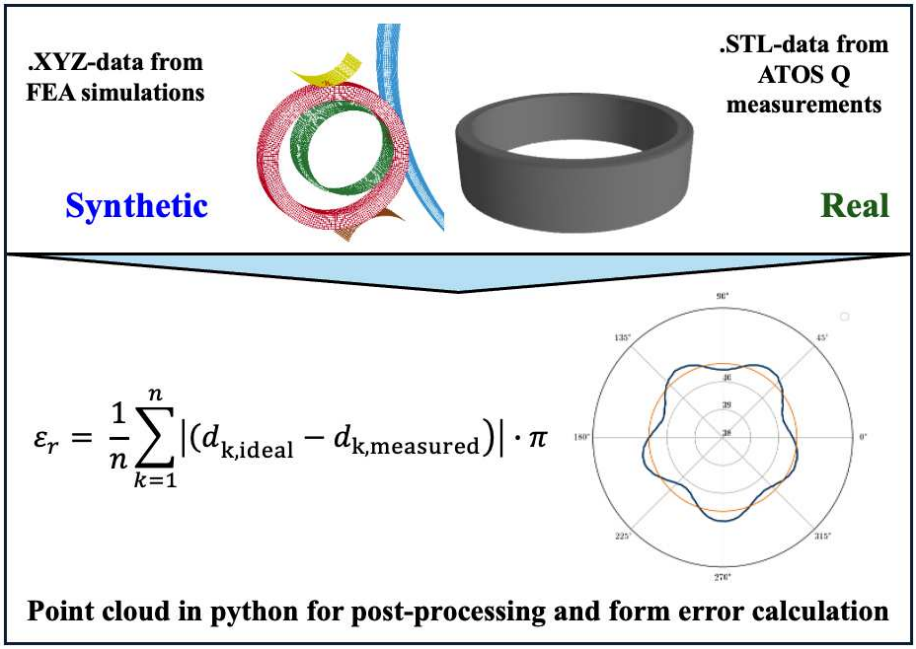



***Figure 14*** *– Conversion of geometry files from different formats to create a hybrid database. The simulation data is exported as .XYZ files, while the experimental data is provided in .STL format following optical measurement (adapted from [55]).*

### *4.3 TCRR database and machine learning modeling*

The experimental TCRR database contains 130 rings. Since repetitions were carried out for each unique parameter configuration, the out-of-roundness was averaged per configuration, resulting in 55 unique configurations. Consequently, the model was trained to predict the average out-of-roundness per unique scalar machine input configuration. Additionally, predicting the standard deviation per configuration allows for reproducibility assessment, which is elaborated further in the Discussion. The input features included the initial outer diameter, tool geometry, main roll rotational speed, feed rate, radial rolling force, guide roll pressure, and end diameter. The end diameter was included as a discretized geometry parameter (± 1 mm), guiding the model toward a desired final shape rather than relying on exact measurements. This is intentional: the model is designed to support process optimization and design, rather than to serve as a quality assurance tool based on precise measured values. Moreover, given the nature of process modeling for optimization, it is important for the model to detect significant out-of-roundness errors. Therefore, prediction performance is reported in terms of RMSE, as this metric penalizes large deviations more strongly than MAE. Moreover, the results remain interpretable in the original unit of millimeters (mm). $R^2$ and MSE were considered in this research but were not reported to provide a comprehensive presentation of the results. However, these metrics showed comparable trends to RMSE.

The 55 ring configurations of the **experimental dataset** were split into training (70%), validation (15%), and test data (15%), yielding 38 training, 8 validation, and 9 test samples. Most values fell within 0–0.4 mm, with occasional higher values reaching up to 0.6 mm. The **synthetic dataset**, obtained through FEA simulations (see Chapter 4.1), followed the same 70/15/15 split, yielding 52 training, 11 validation, and 12 test samples.

Compared to the experimental data, the out-of-roundness value range was approximately three times larger (single ring comparison, not averaged), reaching up to 6 mm. This simulation-to-reality gap must be accounted for when modeling experimental instances. Similar to the RARR case (Section 3), a *MinMax* scaling was applied to both features and targets, mapping all values to the range [0, 1] and ensuring comparability with the scaled experimental data. Combining both data sources into the **hybrid dataset** yielded 130 rings in total. The training set merged the experimental training samples with the full FEA dataset, while the validation and test sets were drawn exclusively from experimental data. Strict separation for validation and test data was motivated by the target domain of experimental/real data. This resulted in 113 training samples, eight validation samples, and nine test samples.

A Multilayer Perceptron (MLP), a feedforward neural network architecture, was employed to evaluate the integration of FEA data for predicting out-of-roundness from scalar input parameters. The MLP served to validate process generalization within a supervised learning framework incorporating FEA data. To maintain focus on this objective, no results from alternative ML architectures were presented in this section. Each MLP was trained five times, and the resulting performance metrics were averaged to mitigate the influence of model initialization and batch ordering. Furthermore, due to the limited dataset size, ten distinct training, validation, and test splits were generated and evaluated to reduce the effects of data distribution across the splits. Five random seeds were used to generate five different data splits. To maintain comparability between the validation and test sets, these subsets were exchanged to create an additional five splits, resulting in a total of ten different data splits. Performance metrics were reported based on the five model repetitions for each of the ten splits. The hyperparameters of the MLP were optimized using *Optuna* and 5-fold cross-validation [51]. The parameters included the number of hidden layers and neurons per layer, as well as the learning rate, $L_2$ regularization factor, dropout rate and initial output bias. The search space, as well as the optimized hyperparameters, are defined in Appendix D. Model training was executed using MSE loss and the Adam optimizer.

The training strategies used to incorporate FEA data in the RARR unsupervised learning case (Section 3) were analogously applied to the supervised TCRR case to evaluate potential improvements: experimental modeling only (EXP), FEA modeling using the experimental scaler for test data evaluation (FEA), hybrid training (Hybrid) and pre-training on FEA and fine-tuning on experimental data (TL). Model accuracies were compared to assess integration performance of the FEA data.

## *4.4 Results of generalization to TCRR*

The performance of the four training approaches for predicting the out-of-roundness, evaluated on the same left out experimental test and validation data set of ten dataset configurations, is presented in Table 8.

***Table 8*** *– Mean RMSE for the EXP, FEA, Hybrid and TL model. The specific dataset construction (see Chapter 4.3) results in the same RMSE for the validation and test data in the FEA training since no re-training is applied to new configurations.*

| Training approach | RMSE validation [mm] | RMSE test [mm] |
|---|---|---|
| **EXP** | 0.074 ± 0.033 | 0.095 ± 0.04 |
| **FEA** | 0.108 ± 0.03 | 0.108 ± 0.03 |
| **Hybrid** | 0.065 ± 0.030 | 0.089 ± 0.04 |
| **TL** | **0.056 ± 0.027** | **0.085 ± 0.03** |

As reported in Table 8, and similar to the RARR unsupervised case, modeling performance differs between the different training approaches. The TL approaches demonstrate superior performance. The RMSE of the test data decreases from 0.095 mm with the EXP approach to 0.085 mm with the TL approach, while the validation RMSE decreases from 0.074 mm to 0.056 mm. Although hybrid learning shows slightly worse performance than the TL approach, it still has a lower RMSE than the EXP approach. FEA has the greatest modeling errors, with an RMSE of 0.108 mm in the test data. Similar dynamics can also be seen in the validation performances, where the same order of performance is obtained: FEA, EXP, Hybrid, and then TL. Despite applying sensitive hyperparameter selection using Bayesian optimization with cross-validation, overfitting may have occurred. For example, the EXP validation data RMSE is 0.074 mm, which is 0.021 mm lower than the test data RMSE of 0.095 mm. This circumstance, as well as the observed differences in performance, will be discussed further in the Discussion section, where it will be embedded in the dynamics observed in the RARR case. Figure 15 shows an example of a regression diagram for one of the baseline predictions (left), as well as the TL predictions (right). Visually, the transfer-learned predictions align more closely with the measured values than the EXP-only predictions do.

If the experimental data amount for training is reduced to 50%, to address use cases with sparse data availability, the TL and Hybrid approach remained similar in prediction accuracy. The Hybrid approach even improves slightly. However, EXP modeling decreases in accuracy. Fig. 16 captures this in percentage improvement/worsening in comparison to Table 8. The EXP modeling shows the greatest changes, as indicated by 24.2% worsening of the RMSE metric. In comparison, the hybrid approach even improves the RMSE slightly by 3.9%, and the TL approach worsens by only 2.2%. Therefore, the performance of the Hybrid approach is now comparable to the TL approach. Thus, Fig. 16 shows that EXP modeling worsens noticeably when using less data, whereas TL and Hybrid learning are hardly affected by reducing the database. These findings will be cross compared with those of the unsupervised case in the RARR in the Discussion.

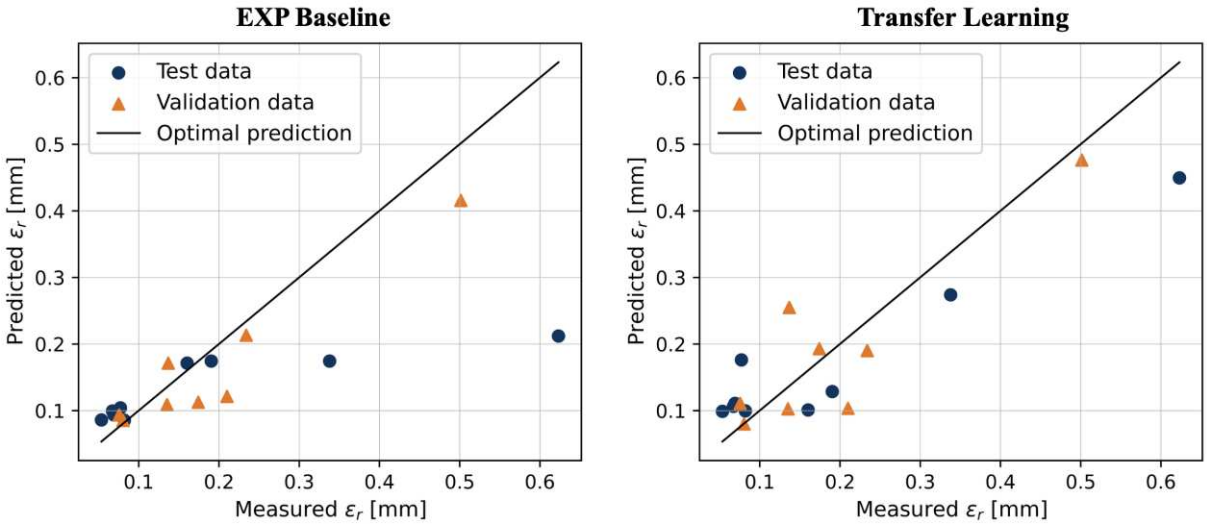
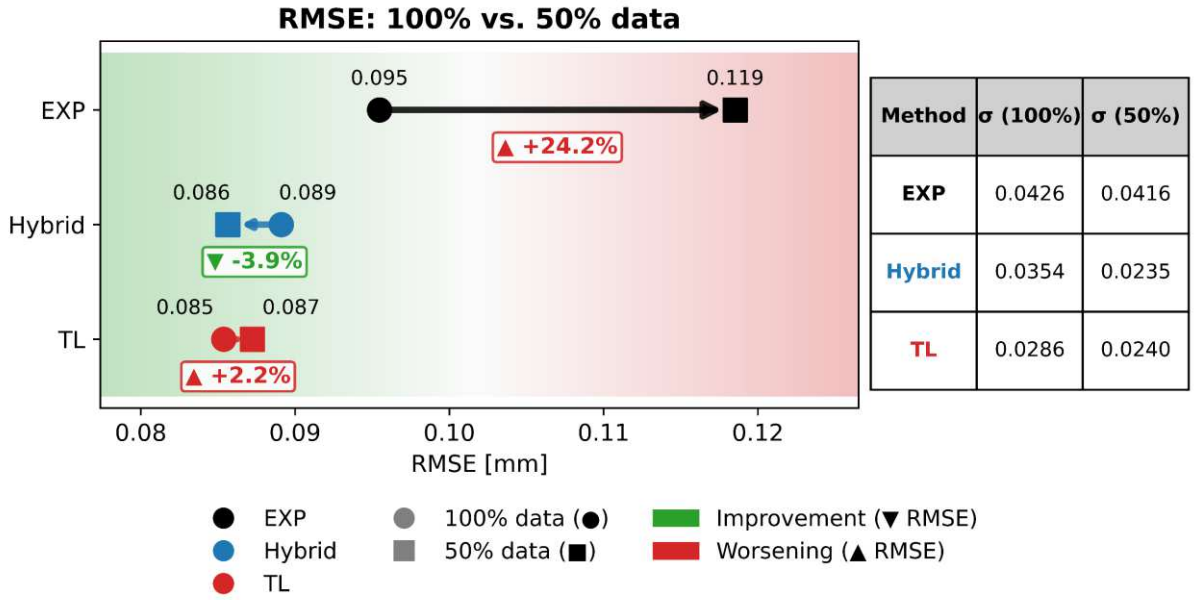



***Figure 15*** *– Instance regression diagram for one data split for the EXP-based, as well as the transfer learned prediction of the out-of-roundness using MLPs.*




| Method | σ (100%) | σ (50%) |
|---|---|---|
| EXP | 0.0426 | 0.0416 |
| Hybrid | 0.0354 | 0.0235 |
| TL | 0.0286 | 0.0240 |

***Figure 16*** *– Change of evaluation metric RMSE from 100% data availability to 50% data availability. The change is indicated by the arrows and annotated by the percentage relative change. The visualization shows on the one hand absolute comparison (x-axis) for the different approaches (EXP, Hybrid, TL – y-axis) and on the other hand relative comparison through the annotated arrows for the RMSE.*

## 5. Discussion and Results Cross-Comparison

The results presented in the previous two chapters show that FEA integration for ring rolling applications using ML benefits model performance. This has been observed in both the unsupervised RARR and the supervised TCRR cases. Therefore, this finding appears to be process- and paradigm-independent in the ring rolling domain. For both investigations, and considering the most performant algorithms, the order of model performance can be summarized as follows: FEA < EXP < Hybrid < TL (increasing performance).

In the RARR case, the latent representations learned from FEA data were successfully adapted through inverse inference toward the experimental distribution via TL, indicating that the network learned and minimized the simulation-to-reality gap (see Fig. 11). This explains the consistently superior performance observed in both RARR and TCRR. Thus, it satisfies the need for suitable integration in FEA modeling, which is most likely due to the domain gap being too large in the analyzed applications. Considering the RARR time series representation, it is especially striking that on the one hand quantitative metrics are transferred to improved reconstruction (compare Fig. 8, e.g. radial force), but also the time-dependent behavior and transients. This is particularly evident in the reconstruction of the outer diameter and thickness in Fig. 8. There, TL shifts time-dependent fluctuations closer to the measured time series, thereby improving the ability to learn temporal relationships. Therefore, the results suggest that TL transforms quantitative relationships (RARR and TCRR cases) and temporal relationships (RARR) toward improved predictions and representations by overcoming the domain gap and reducing the simulation-to-reality gap.

This assumption aligns with the findings regarding the effect of data preprocessing in the RARR case. Reducing the simulation-to-reality gap by unifying the FEA data and removing out-of-distribution information respecting the EXP dataspace improves model performance. Therefore, the network can focus more on process-specific learning and less on learning the simulation-to-reality gap. Using pre-processing enabled superior model

performance for all possible data configurations with TL, whereas no additional pre-processing encountered a threshold (30 experiments), where prediction performances equalized. Accordingly, additional pre-processing must be considered when seeking further model improvement. In comparison, for the TCRR case, only the target space is pre-processed using *MinMax*-scaling to align in the boundaries of the EXP target values, which is comparable to the baseline pre-processing in the RARR case that inherited the threshold. Further pre-processing of the features was not possible, since the machine input spaces are the same for EXP and FEA modeling. Although, using only baseline pre-processing for the TCRR supervised case, the threshold, similar to RARR, has not been exceeded. Improved performance was still achieved by utilizing 100% of the data for modeling using the hybrid and TL approach. However, performance in comparison to EXP modeling relatively improved more, when only considering 50% of the data available. Here, in contrast to the RARR case, the hybrid learning approach outperforms the TL approach, yielding comparable performance (RMSE). Thus, the hybrid approach may benefit from having less EXP data and relies more on FEA data. However, considering the number of data points – 500 per test data experiment in RARR and one per test data experiment in TCRR – it is difficult to make a direct comparison. These findings suggest that hybrid learning may be particularly beneficial when the amount of experimental data is insufficient for effective fine-tuning in TL. However, dedicated studies with controlled data regimes are required to validate this hypothesis. To summarize the data point of view, FEA integration is of particular interest when data is scarce, such as when a new product variant is introduced. However, it can also provide new information when a certain type of information is underrepresented in EXP data but covered in FEA data, which often occurs in industrial EXP databases where processes are standardized and performed only within certain ranges.

The representation learning models (autoencoders) of the RARR task can be used for anomaly detection (Appendix C) and *ab initio* FEA data pre-processing (Chapter 3.4). The proposed results, as well as the performance metrics and regression diagrams in Fig. 8, indicate that the trained models enable such applications. For example, see Appendix C for the anomaly detection. The improved reconstruction of quantitative metrics, especially the conversion of 4.975 kN to 1.708 kN, combined with improved time relationship detection, enables accurate tracking of changes in rolling regimes for downstream applications. The TCRR supervised use case is per definition easier to place in its industrial application, since the model was designed with surrogate model-based process optimization in mind. Here, optimization algorithms, such as genetic algorithms, could optimize the machine input to derive stable process configurations. On top of the prediction of the average out-of-roundness per configuration, the assessment of the standard deviation (repeatability) can be included in the optimization. The RMSE of 0.085 mm achieved with the test data, as well as the orientation in the regression diagram with sensitive prediction capability, demonstrate the feasibility of using the proposed model as a surrogate model. Furthermore, similar prediction accuracies have been achieved using 50% of the experimental data with TL and Hybrid training. Consequently, EXP effort is reduced to enable process optimization while maintaining similar outputs. This is particularly relevant when optimizing new product variants.

While the observed benefits of FEA integration were evident, the supervised TCRR case was limited by the small size of the experimental dataset, as evidenced by residual overfitting. However, the small datasets reflect the intended application scenario of data-scarce industrial environments. The overfitting occurred, although cross-validation and Bayesian optimization have been used for hyperparameter selection. Moreover, reducing hyperparameter to very little networks (e.g. 1 hidden layer), slight overfitting still remains. This may therefore be related to the small dataset utilized in this research. In future work, other ML algorithms, such as Random Forest,

should be elaborated, to investigate if these can reduce overfitting. To this end, extension to other ML architecture has not been performed in this research to provide only information about the different FEA integrations. In the unsupervised task in RARR, no overfitting has been found. The feasibility of industrial application of the derived models is presented and discussed within this contribution, the actual deployment is beyond the scope of the paper, which focuses strictly on FEA data integration mechanisms. Future work should address the deployment in an industrial setting and its linked model comparisons.

The presented approaches can be considered as a specific type of hybrid modeling that integrates physical knowledge into machine learning. In this case, physical knowledge is represented by the results of numerical finite element analysis (FEA) simulations and is integrated from a data and transfer learning (TL) perspective. In the future, we can seek the direct integration of physics into physics-informed machine learning models. In the unsupervised case, this could be achieved by incorporating governing equations into the loss function of known relationships. A preliminary study based on the results obtained in this paper demonstrated this approach for the supervised prediction of process parameters. This study indicated good agreement when little data was available [40]. Nevertheless, TL achieved better predictive performance when a sufficiently large FEA dataset was available, supporting its use under these conditions. Furthermore, in the supervised TCRR case, we could include analytical models that provide deterministic predictions of process stability. Combining these approaches with FEA integration could improve explainability and generalization capability. Additionally, future work may involve integrating more low-fidelity approaches, such as analytical or semi-analytical modeling. These approaches could serve as a potential supplement to those proposed in this research.

## 6. Conclusions

This study examines the integration of finite element analysis (FEA) data in machine learning (ML) modeling for ring rolling processes. To this end, four distinct training approaches are considered: FEA training only and inference on experimental data, experimental training only (EXP), hybrid training (Hybrid), and transfer learning with parameter transfer (TL). In general, the following order of performance has been observed in both use cases: **FEA < EXP < Hybrid < TL (increasing performance)**. Thus, simulation-informed models show superior performance for the presented and analyzed tasks. The work was designed and presented in two steps: (I) a general study using an unsupervised learning approach via autoencoders for process representation in RARR and (II) a supervised learning approach in TCRR for predicting out-of-roundness. This enables cross-comparison between different tasks and processes within ring rolling. The main findings can be summarized as follows:

(I) TL learns to minimize the simulation-to-reality gap, enabling predictions in different data regimes to those pre-trained on, such as for different ring geometries. This behavior can also be used to transform the FEA data space into the EXP data space by reapplying the learned shift to the FEA data.

(II) The positive effect of FEA data integration is especially significant when EXP data is scarce or not well represented in the database (i.e., out-of-distribution).

(III) Pre-processing can help to reduce the simulation-to-reality gap before training, leading to improved model performance.

(IV) TL demonstrated superior performance in most training configurations. The Hybrid approach reached only similar or slightly better metrics for one setup with little data in the TCRR case with 50% of the data.

The trained models can be used in industrial applications. The RARR autoencoder can be used for anomaly detection and FEA preprocessing. The TCRR supervised out-of-roundness model can be used as a surrogate model for process optimization and process mapping. However, future work needs to address these settings further since the investigations presented primarily focus on deriving effects and rules for FEA integration in ML training. This study demonstrated that FEA-derived knowledge enhances ML models for ring rolling, especially when data is scarce. Transfer learning appears to be the most robust integration strategy because it enables adaptation across the simulation-to-reality gap.

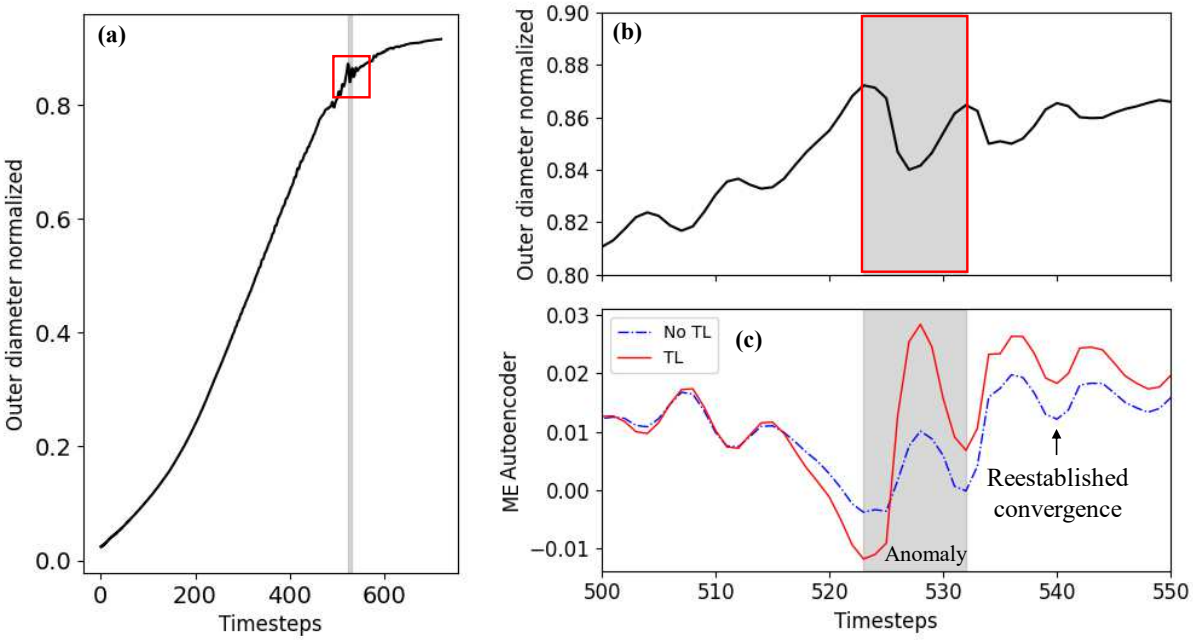
## Appendices

***Appendix A*** *– Optuna optimization results for the autoencoder. The maximal units were limited to 1200 per layer. The network is designed symmetrically (decoder vs. encoder). The number of attention heads equals to 2 for the transformer after optimization. Moreover, the batchsize is set to 64 and the patience of early stopping to 50.*

| **Architecture** | **Model depth** | **Dropout** | **Learning rate** |
|---|---|---|---|
| RNN | [638, 319,5,319,638] | 0 | 0.0001 |
| Transformer | [672, 336,5,336, 672] | 0 | 0.001 |
| LSTM | [1013, 507,5,507,1013] | 0 | 0.001 |

***Appendix B*** *– Performance of different models for five experiments (5 EXP) for cross validation accuracies. The accuracy is averaged and stated with its standard deviation. The best performing approach is bold highlighted.*

| | Architecture | EXP | FEA | Hybrid | TL |
|---|---|---|---|---|---|
| RMSE | RNN | $0.038 \pm 0.008$ | $0.049 \pm 0.003$ | $0.049 \pm 0.003$ | $0.020 \pm 0.004$ |
| | **LSTM** | $\mathbf{0.036 \pm 0.006}$ | $\mathbf{0.043 \pm 0.005}$ | $\mathbf{0.035 \pm 0.007}$ | $\mathbf{0.016 \pm 0.003}$ |
| | Transformer | $0.037 \pm 0.011$ | $0.146 \pm 0.010$ | $0.043 \pm 0.006$ | $0.022 \pm 0.005$ |
| $R^2$ | RNN | $0.98 \pm 0.011$ | $0.96 \pm 0.007$ | $0.97 \pm 0.005$ | $0.994 \pm 0.003$ |
| | **LSTM** | $\mathbf{0.98 \pm 0.009}$ | $\mathbf{0.98 \pm 0.008}$ | $\mathbf{0.98 \pm 0.010}$ | $\mathbf{0.996 \pm 0.002}$ |
| | Transformer | $0.98 \pm 0.019$ | $0.71 \pm 0.010$ | $0.97 \pm 0.010$ | $0.992 \pm 0.003$ |




***Appendix C*** *– (a) Ring diameter expansion of a test data sample containing an instability. This instability is considered as an anomaly and sought to be predicted by the autoencoders. (b) Zoom-in of the instability region with the relevant deviations in terms of normalized outer diameter of the ring. (c) Anomaly detection employing the ME variation as threshold for the EXP trained LSTM and TL models. The time steps range highlighted in gray are considered the anomaly manifestation interval. Modeling with all experiments (60) and constructing the LSTM model with 1 timestep to account for changes in small time windows.*

***Appendix D*** *– Optuna search space and optimal hyperparameter configuration for the baseline modeling.*

| | **Layers** | **Units** | **Learning Rate** | **Batch Size** | **Dropout** | **$L_2$ regularization** |
|---|---|---|---|---|---|---|
| Search space | [1, 2] | $[1, 12] \cap \mathbb{N}$ | [0.01, 0.001, 0.0001] | [8, 16] | [0, 0.05, 0.1, 0.2] | 0.001, 0.0001, 0.00001] |
| Optimum | 2 | 11 | 0.01 | 8 | 0.1 | 0.001 |

## References


[1] Allwood JM, Tekkaya AE, Stanistreet TF. The development of ring rolling technology - Part 1: Process design and production equipment. Steel Res Int. 2005;76(2-3):111-20. doi:10.1002/srin.200505981.

[2] Hua L, Deng J, Qian D. Recent development of ring rolling theory and technique. Int J Mater Prod Technol. 2017;54(1-3):65-87. doi:10.1504/IJMPT.2017.080566.

[3] Guo L, Yang H. Towards a steady forming condition for radial-axial ring rolling. Int J Mech Sci. 2011;53(4):286-99. doi:10.1016/j.ijmecsci.2011.01.010.

[4] Allwood JM, Tekkaya AE, Stanistreet TF. The development of ring rolling technology - Part 2: Investigation of process behaviour and production equipment. Steel Res Int. 2005;76(7):491-507. doi:10.1002/srin.200506044.

[5] Hawkyard JB, Johnson W, Kirkland J, Appleton E. Analyses for roll force and torque in ring rolling, with some supporting experiments. Int J Mech Sci. 1973;15(11):873-93.

[6] Mamalis AG, Johnson W, Hawkyard JB. Pressure distribution, roll force and torque in cold ring rolling. J Mech Eng Sci. 1976;18(4):196-209. doi:10.1243/JMES_JOUR_1976_018_033_02.

[7] Quagliato L, Berti GA, Kim D, Kim N. Contact geometry estimation and precise radial force prediction for the radial-axial ring rolling process. Int J Mater Form. 2018;11(6):789-805. doi:10.1007/s12289-017-1388-x.

[8] Quagliato L, Berti GA, Kim D, Kim N. Slip line model for forces estimation in the radial-axial ring rolling process. Int J Mech Sci. 2018;138-139:17-33. doi:10.1016/j.ijmecsci.2018.01.025.

[9] Parvizi A, Abrinia K. A two dimensional upper bound analysis of the ring rolling process with experimental and FEM verifications. Int J Mech Sci. 2014;79:176-81. doi:10.1016/j.ijmecsci.2013.12.012.

[10] Ranatunga V, Gunasekera JS, Vaze SP, De Souza U. Three-dimensional UBET simulation tool for seamless ring rolling of complex profiles. J Manuf Process. 2004;6(2):179-86.

[11] Parvizi A, Abrinia K, Salimi M. Slab analysis of ring rolling assuming constant shear friction. J Mater Eng Perform. 2011;20(9):1505-11. doi:10.1007/s11665-010-9824-9.

[12] Yea Y, Ko Y, Kim N, Lee J. Prediction of spread, pressure distribution and roll force in ring rolling process using rigid-plastic finite element method. J Mater Process Technol. 2003;140(1-3):478-86. doi:10.1016/S0924-0136(03)00721-0.

[13] Kim N, Kim H, Jin K. Optimal design to reduce the maximum load in ring rolling process. Int J Precis Eng Manuf. 2012;13(10):1821-8. doi:10.1007/s12541-012-0239-4.

[14] Kim B, Moon H, Kim E, Choi M, Joun M. A dual-mesh approach to ring-rolling simulations with emphasis on remeshing. J Manuf Process. 2013;15(4):635-43. doi:10.1016/j.jmapro.2013.05.002.

[15] Wang ZW, Fan JP, Hu DP, Tang CY, Tsui CP. Complete modeling and parameter optimization for virtual ring rolling. Int J Mech Sci. 2010;52(10):1325-33. doi:10.1016/j.ijmecsci.2010.06.008.

[16] Berti GA, Quagliato L, Monti M. Set-up of radial-axial ring-rolling process: Process worksheet and ring geometry expansion prediction. Int J Mech Sci. 2015;99:58-71. doi:10.1016/j.ijmecsci.2015.05.004.

[17] Hua L, Deng J, Qian D. Precision ring rolling technique and application in high-performance bearing manufacturing. MATEC Web Conf. 2015;21:03002. doi:10.1051/matecconf/20152103002.

[18] Qi H, Li Y. Research status and developing trends on the ring rolling process of profile ring parts. Procedia Eng. 2017;207:1260-5. doi:10.1016/j.proeng.2017.10.880.

[19] Phuyal S, Bista D, Bista R. Challenges, opportunities and future directions of smart manufacturing: A state of art review. Sustain Futures. 2020;2:100023. doi:10.1016/j.sftr.2020.100023.

[20] Cao J, Bambach M, Merklein M, Mozaffar M, Xue T. Artificial intelligence in metal forming. CIRP Ann Manuf Technol. 2024. doi:10.1016/j.cirp.2024.04.102.

[21] Yang YY, Linkens DA, Talamantes-Silva J, Howard IC. Roll force and torque prediction using neural network and finite element modeling. ISIJ Int. 2003;43(12):1957-66.

[22] Rohani Raftar HR, Parvizi A. Prediction and optimization of load and torque in ring rolling process through development of artificial neural network and evolutionary algorithms. J Comput Appl Mech. 2018;49(2):292-303. doi:10.22059/jcamech.2018.246800.215.

[23] Mirandola I, Berti GA, Caracciolo R, Lee S, Kim N, Quagliato L. Machine learning-based models for the estimation of the energy consumption in metal forming processes. Metals. 2021;11(5):833. doi:10.3390/met11050833.

[24] Lee S, Park J, Kim N, Lee T, Quagliato L. Extreme gradient boosting-inspired process optimization algorithm for manufacturing engineering applications. Mater Des. 2023;226:111625. doi:10.1016/j.matdes.2023.111625.

[25] Moser T, Seitz J, Alp E, Kuhlenkotter B. Identification of investigation procedures to predict work roll fatigue for developing machine learning applications - A systematic literature review. In: Proceedings of the 5th Conference on Production Systems and Logistics; 2023. p. 268-81.

[26] Fahle S, Glaser T, Kneissler A, Kuhlenkotter B. Improving quality prediction in radial-axial ring rolling using a semi-supervised approach and generative adversarial networks for synthetic data generation. Prod Eng. 2022;16(1):175-85. doi:10.1007/s11740-021-01075-x.

[27] Liang L, Guo L, Wang Y, Li X. Towards an intelligent FE simulation for real-time temperature-controlled radial-axial ring rolling process. J Manuf Process. 2019;48:1-11. doi:10.1016/j.jmapro.2019.09.032.

[28] Lafarge R, Hutter S, Tulke M, Halle T, Brosius A. Data based model predictive control for ring rolling. Prod Eng. 2021;15(6):821-31. doi:10.1007/s11740-021-01063-1.

[29] Seitz J, Moser T, Fahle S, Prinz C, Kuhlenkotter B. Transfer learning approaches in the domain of radial-axial ring rolling for machine learning applications. In: Proceedings of the 5th Conference on Production Systems and Logistics; 2023. p. 447-58.

[30] Quagliato L, Seitz J, Perin M. Machine learning modeling for material science and manufacturing: Overview and perspectives for the future. ELSP AI Mater. 2025. doi:10.55092/aimat20250005.

[31] Kasilingam S, Yang R, Singh SK, Farahani MA, Rai R, Wuest T. Physics-based and data-driven hybrid modeling in manufacturing: A review. Prod Manuf Res. 2024;12(1):2305358. doi:10.1080/21693277.2024.2305358.

[32] Nath D, Ankit, Neog DR, Gautam SS. Application of machine learning and deep learning in finite element analysis: A comprehensive review. Arch Comput Methods Eng. 2024;31(5):2945-84. doi:10.1007/s11831-024-10063-0.

[33] Li C, Yang S, Zheng H, Zhang Y, Wu L, Xue W, et al. Integration of machine learning with finite element analysis in materials science: A review. J Mater Sci. 2025;60:8285-307. doi:10.1007/s10853-025-10913-7.

[34] Lee S, Lim Y, Galdos L, Lee T, Quagliato L. Gaussian process regression-driven deep drawing blank design method. Int J Mech Sci. 2024;265:108898. doi:10.1016/j.ijmecsci.2023.108898.

[35] Heinzelmann P, Baum S, Riedmuller KR, Liewald M, Weyrich M. A comprehensive benchmark dataset for sheet metal forming: Advancing machine learning and surrogate modeling in process simulations. MATEC Web Conf. 2025;408:01090. doi:10.1051/matecconf/202540801090.

[36] Pareek KA, May D, Meszmer P, Abo Ras M, Wunderle B. Synthetic data generation using finite element method to pre-train an image segmentation model for defect detection using infrared thermography. J Intell Manuf. 2025;36(4):1879-905. doi:10.1007/s10845-024-02326-1.

[37] Fernandez F, Saravanan S, Omongos RL, Troncoso JF, Galvez-Aranda DE, Franco AA. Transfer learning assessment of small datasets relating manufacturing parameters with electrochemical energy cell component properties. npj Adv Manuf. 2025. doi:10.1038/s44334-025-00024-1.

[38] Xiao X, He Y, Su J, Hu K. Domain adaptation from simulation to reality: A GAN- and MK-MMD-based transfer learning approach for bearing fault diagnosis. Appl Sci. 2026;16(3):1407. doi:10.3390/app16031407.

[39] Modanloo V, Jang S, Lee T, Quagliato L. Gradient enhanced-expert informed neural network (GE-EINN) for forming depth prediction from a small-scale metal stamping dataset. J Manuf Process. 2025;140:224-40. doi:10.1016/j.jmapro.2025.02.052.

[40] Quagliato L, Seitz J, Perin M, Malejka D, Moser T, Zanella C, et al. Proxy-physics-informed and transfer learning networks for radial-axial ring rolling (RARR) under varying data availability. Key Eng Mater. 2026;1049:13-28. doi:10.4028/p-B4tPZr.

[41] Hensel A, Spittel T. Kraft- und Arbeitsbedarf bildsamer Formgebungsverfahren. 1st ed. Leipzig: Deutscher Verlag fur Grundstoffindustrie; 1978. (In German language).

[42] Sun ZC, Yang H, Ou XZ. Thermo-mechanical coupled analysis of hot ring rolling process. Trans Nonferrous Met Soc China. 2008;18(5):1216-22.

[43] Lee KH, Ko DC, Kim DH, Lee SB, Sung NM, Kim BM. Control method for centering rolls in radial-axial ring rolling process. Int J Precis Eng Manuf. 2014;15(3):535-44. doi:10.1007/s12541-014-0368-z.

[44] Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. In: Rumelhart DE, McClelland JL, editors. Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1, Foundations. Cambridge (MA): MIT Press; 1986. p. 318-62.

[45] Lipton ZC, Berkowitz J, Elkan C. A critical review of recurrent neural networks for sequence learning. arXiv [Preprint]. 2015. arXiv:1506.00019.

[46] Lillicrap TP, Santoro A. Backpropagation through time and the brain. Curr Opin Neurobiol. 2019;55:82-9. doi:10.1016/j.conb.2019.01.011.

[47] TensorFlow Developers. TensorFlow. Zenodo; 2026. doi:10.5281/zenodo.4724125

[48] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735-80. doi:10.1162/neco.1997.9.8.1735.

[49] How DNT, Loo CK, Sahari KSM. Behavior recognition for humanoid robots using long short-term memory. Int J Adv Robot Syst. 2016;13(6):1-14. doi:10.1177/1729881416663369.

[50] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Advances in Neural Information Processing Systems 30; 2017. p. 5998-6008.

[51] Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2019. p. 2623-31. doi:10.1145/3292500.3330701.

[52] Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng. 2010;22(10):1345-59. doi:10.1109/TKDE.2009.191.

[53] Xie D, Ouyang Q, Wang Y, He L, Xu W. Radial-axial feeding process design for large-scale flat ring by quantitative estimation of ring roundness error. Thin-Walled Struct. 2023;189:110887. doi:10.1016/j.tws.2023.110887.

[54] Seitz J, Wang Q, Quagliato L, Moser T, Lee T, Brosius A, Kuhlenkötter B. Synthetic process modeling for ring rolling via physics-guided GANs and analytical transfer learning. J Intell Manuf. 2026. doi:10.1007/s10845-026-02904-5.

[55] Wang Q, Seitz J, Brosius A, Kuhlenkötter B. Algorithm for a Comparable Calculation of Form Errors Based on a Hybrid Database in Cold Ring Rolling. In: Production at the Leading Edge of Technology; Drossel W-G, Ihlenfeldt S, Dix M, editors. 2025. p. 234–241. doi:10.1007/978-3-031-86893-1_26.

## Statements and Declarations

**Funding:** The results were mainly developed in the Project 499350001, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation).

**Conflict of interest:** The authors have no relevant financial or non-financial interests to disclose.

**Data availability statement:**

The datasets used for the TCRR modeling presented in this study are publicly available at the following persistent GitLab link:

https://gitlab.ruhr-uni-bochum.de/lps-projects/kodaml/tcrr-sim2real/-/tree/f70a3403082820673c1b77121a9b9c087f83e4b9/

The RARR datasets are not publicly available due to confidentiality but are available from the corresponding author on reasonable request.

**Author contributions:** All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Johannes Seitz, Luca Quagliato, Qinwen Wang and Mattia Perin. The first draft of the manuscript was written by Johannes Seitz, with contributions to the TCRR section by Qinwen Wang and Helin Siedo. Luca Quagliatio and Tobias Moser coordinated the internal revisions. Resources and supervision were provided by Alexander Brosius and Bernd Kuhlenkötter. All authors read and approved the final manuscript.