A Hybrid Model for Hot Rolling Pass Schedule Design Using Reinforcement Learning

Marco Kemmerling[1], **Nils Frahm**[1], **Christian Idzik**[2], **Anas Abdelrazeq**[1], **David Bailly**[2], and **Robert H. Schmitt**[1]
[1]Chair of Intelligence in Quality Sensing, RWTH Aachen University, Aachen, Germany
[2]Institute of Metal Forming, RWTH Aachen University, Aachen, Germany

Abstract

Hot rolling of metals requires the design of pass schedules to accommodate different products and arrive at the desired target properties. Pass schedule design is often still performed manually, although automated approaches are emerging in the literature. One such approach is the use of reinforcement learning, which requires extensive training on simulations. To allow for efficient development of suitable reinforcement learning agents, it is important to reduce the computational cost of each simulation step as much as possible. While some simulation models for hot rolling exist, they have not been developed explicitly for reinforcement learning purposes. Today, there is a lack of fast approximate models to allow for efficient (pre-)training of reinforcement learning agents. To address this gap, we propose a hybrid model consisting of a data-driven model augmented by physical approximations. Our hybrid model produces better predictions than either of its components and is efficient enough to allow for quick training of reinforcement learning agents. We demonstrate that reinforcement learning agents pre-trained on our hybrid model can be transferred to a more accurate simulation and achieve good results after a short additional training phase. The two-stage training approach significantly reduces the overall training while achieving similar agent performances. The reduction in training time enables faster iteration cycles in agent and environment design, thereby creating a foundation for the development of reinforcement learning approaches for a wider range of hot rolling scenarios in the future.

Keywords Hot Rolling · Reinforcement Learning · Hybrid Model · Pass Schedule

1 Introduction

In 2022, more than 1,855 million metric tons of crude steel were produced worldwide [1] and most of the produced steel is rolled at least once. Rolling is a well-established and highly automated metal-forming process that produces finished or semi-finished products such as plates, sheets, and coils for various industries. In rolling, the metal workpiece is passed through a gap between two or more rotating cylindrical rolls to reduce its thickness. In most applications, the metal is reduced to the desired thickness in several steps, called passes. Each pass is defined by among others, the applied thickness reduction, rolling velocity, and the start time of each pass. The process parameters of all passes in a process are defined in a so-called pass schedule. These pass schedules significantly influence the final material properties and they are often designed using very specific heuristics and process knowledge [22].

In recent times, research has focused on automatic pass schedule design using reinforcement learning [18, 7, 8], where a reinforcement learning agent is coupled with a fast rolling model (FRM) to create and evaluate pass schedules. While the computational cost of such FRMs is appropriate for applications of trained agents in practice, the training of reinforcement learning agents requires significant numbers of interactions with such FRMs, which leads to lengthy training times and slow iteration cycles in the design of rolling agents.

To foster more rapid experimentation in the development of rolling agents, we propose the use of a faster approximate model to simulate hot rolling processes. This approximate model can be used to pre-train reinforcement learning agents, such that the majority of training time is spent on the much faster approximate model with a short subsequent training phase of the more realistic FRM. The model we propose is a hybrid model, consisting of a data-driven core trained by supervised learning which is further augmented by a physical approximation describing selected parts of the hot rolling process.

We demonstrate the accuracy of our model with respect to the FRM and further show that it can be used for the pre-training of agents which can then be successfully transferred to the FRM, arriving at a much reduced total training time.

In the remainder of this work, we describe our approach and result in detail, starting with a brief background on hot rolling in Section 2 and a discussion of related work in Section 3. In Section 4, we describe our modeling approach and demonstrate the performance of the hybrid model, before describing experiments on training and transferring reinforcement learning agents in Section 5.

2 Hot Rolling

Among the many subcategories of rolling, this paper focuses on the hot rolling of long slabs or sheets with rectangular cross-sections. In hot rolling, the temperature of the rolled metal is above its recrystallization temperature. For steel, this is usually between 720 and 1260°C [15]. On the one hand, the high temperature allows to deform the metal more easily due to its lower strength and higher ductility. On the other hand, it allows to influence the microstructure and thus the final mechanical properties which are relevant for the final usage. However, the relationships between the selected process parameters, the material behaviour and the final microstructure as well as the mechanical properties are highly complex.

To design hot rolling schedules so that the final material properties are achieved, a model of the process is needed. Initial efforts were made early in the last century to describe the rolling process and the material behaviour in simplified analytical and semi-empirical equations. These equations were integrated into fast-rolling models (FRM) which are often used in the industry as well as in academia. In the hot rolling process, numerous FRMs have been developed to simulate and predict material properties during and after rolling based on the pass schedule and material parameters. More concretely, they predict the resulting forces, torques, temperature, and microstructure evolution, within seconds. One well-known rolling model called Slimmer (Sheffield Leicester Integrated Model) was developed by Caglayan and Seynon [3] to predict the microstructural evolution in rolling. Another example is the FRM presented by Schey in 2000 [19] where the goal of the FRM is to predict the rolling forces and torques for low carbon steels. Similar models from Bland and Ford in 1948 [2] and Roychoudhury and Lenard in 1984 [16] can be found in the literature.

In this paper, the classical analytical FRM called RoCaT (Rolling Calculation Tool) developed by Lohmar et al. [11] is used. It calculates the complete rolling process within a few seconds and allows the prediction of force, torque, crystallised fraction, and grain size, among others, including height resolution and the influence of shear.

However, when training reinforcement learning agents, a few seconds of pure calculation time for each interaction between agent and the FRM can lead to undesirably long training times, especially if typical deviations in material properties and process disturbances such as incorrect start temperature, longer pause times, etc. are also to be taken into account in the optimisation process.

3 Related Work

Currently, FRMs as described in the previous section are a typical choice to train reinforcement learning agents for hot rolling pass schedule design. While training of such agents using FRMs is possible, the incurred computational costs are significant and can lead to slow development cycles. For instance, Idzik et al. [7] report a training duration of roughly 24 hours when having direct access to a FRM and performing 20,000 training steps. In many cases, direct access to a FRM is not feasible due to intellectual property issues. In the absence of such direct access, training can be accomplished through a Software-as-a-Service (SaaS) architecture as proposed by Scheiderer et al. [18]. This allows for the use of the FRM while preserving the intellectual property of the domain experts, but incurs additional communication overhead. In our own experiments, the training duration increases to roughly one week when accessing a FRM model through a SaaS architecture. As multiple trainings are typically necessary to settle on a suitable reinforcement learning agent design, this impacts the development of reinforcement learning solutions significantly.

Instead of direct training on a FRM, a potential approach to speed up development time is to introduce a pre-training stage, in which a reinforcement learning agent is first trained on a faster, but less accurate data-driven model before being transferred onto the FRM for final adjustments. In recent years, data-driven models have increasingly replaced semi-empirical FRMs, as exemplified by Saravanakumar et al. [17] who use five inputs like coiling temperature and carbon equivalent for the final mechanical properties prediction. In a more advanced model developed by Xie et al. [21], a deep neural network is trained using 27 inputs to predict mechanical properties.

Data-driven models can additionally be augmented by mathematical or physical models to improve their accuracy and training speed. Such hybrid models can take a variety of different shapes, such as embedding physics into the input of data-driven models, using physics to inform the model selection and architecture, or embedding physics into loss functions [20].

In the context of hot rolling, hybrid models have previously been created to predict the roll force [13, 14, 6] by using a mixture of conventional features and features computed

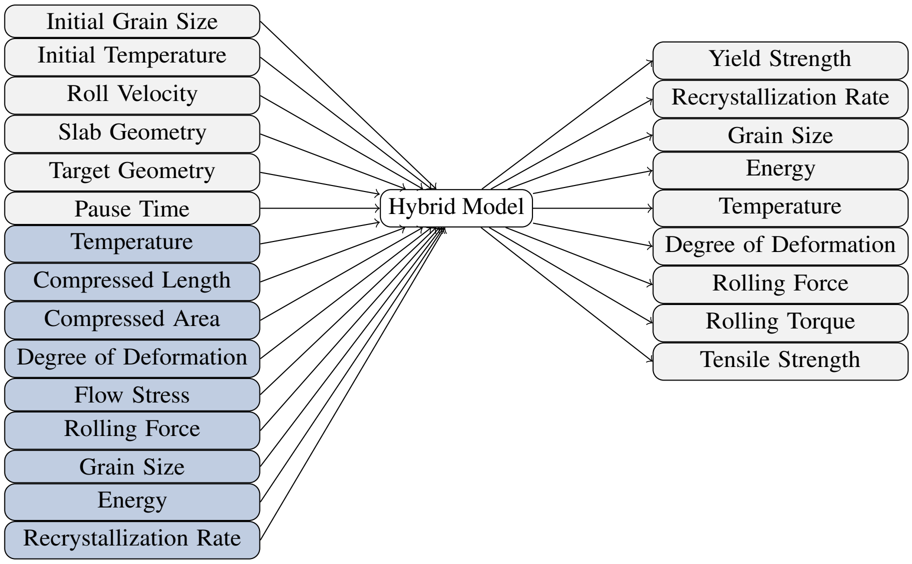

Figure 1: Overview of the proposed hybrid model. Inputs given in grey are regular features, while inputs given in blue are features that have been calculated by a physical approximation. For a more compact representation, several of the visualized individual inputs are actually groups of features.

by a mathematical model as input to a neural network or gradient boosting method. The resulting predictions are generally more accurate than either a purely data-driven or mathematical model. To the best of our knowledge, more comprehensive hybrid models that feature a set of target variables suitable for hot rolling pass schedule design have not been investigated before.

Although not considered in this work, explicit transfer learning techniques have been investigated before and may be helpful to facilitate a successful transfer of reinforcement learning agents from the pre-training stage to the FRM.

4 A Hybrid Hot Rolling Model

According to the classification by Wang et al. [20], our model is a physics-informed machine learning model, with physics embedded into the data-driven model as input in the form of simulation results. Our proposed model produces a set of outputs given a set of conventional features, as well as a set of features calculated by a physical approximation, as visualized in Figure 1. The overall model is actually a collection of individual models for each target variable, which may each be trained by different algorithms. Additionally, each pass is modelled by an individually trained model. As the pass number increases, the inputs for the corresponding model become more extensive, as all the inputs of previous passes are included as well.

The model is constructed along the procedure visualized in Figure 2, which starts with sampling the necessary training data from the RoCaT FRM through the SaaS architecture. We use a weighted random sampling strategy respecting the boundaries of RoCaT, which focuses on parameter values that are likely to be encountered during the design of pass schedules, while sampling more extreme values with reduced probability. An example

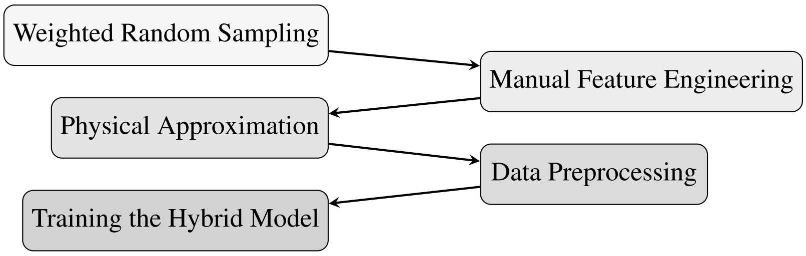

Figure 2: Overview of the construction procedure of the proposed hybrid model.

of the latter is a strong height reduction in a single pass, which leads to undesirable grain sizes. This effect only needs to be captured to the degree that a reinforcement learning agent trained on the model learns to avoid such scenarios and, consequently, very precise predictions of the effect are not required.

Different variables in the sampled data serve as features for the model (see Figure 1), with some being augmented by additional feature engineering to arrive at more suitable data-driven model inputs. Examples of this are (relative) height differences instead of absolute before and after height values, or the time between two passes instead of two separate before and after timestamps. Next to the conventional features, a separate set of features is then computed by a simple and efficient physical approximation of the process. This physical approximation part of the model consists of the equations used by Liebenberg [10, Section 4.1.2 & Appendix A], which are omitted here due to space constraints. These equations include procedures to estimate the temperature of the workpiece by considering its heat loss due to dissipation and contact with the roll, as well as formulae to calculate recrystallization rate, grain size, rolling force, and degree of deformation.

To prepare the data for the training, we standardize the features to have zero mean and unit variance. Furthermore, we perform features selection by using a tree-based approach to compute the impurity-based feature importances. The data is then split into training and test set to train a model for each target variable. For each target, three candidate models are trained using XGBoost [4], LightGBM [9], and CatBoost [12]. The final model for a target is then obtained by choosing the candidate model with the lowest average mean absolute error (MAE) on the test set.

The performance gain of the hybrid model in comparison to either of its two constituents is visualized in Figure 3 and displayed in Table 4. On average, the hybrid model yields far better predictions than the physical approximation alone, and achieves modestly better predictions than the data-driven model alone. The improvement over the data-driven model is especially pronounced in the prediction of the grain size, which is an important parameter due to it being one of the optimization targets in the pass schedule design.

While most predictions improve when incorporating the physical approximation into the data-driven model, the quality of some predictions actually decreases. This especially concerns the temperature predictions, which are almost always slightly worse in the hybrid model than in the purely data-driven model. This is not especially problematic, since the data-driven model can simply be used instead of the hybrid model for temperature predictions. Nevertheless, as Figure 5 shows, the quality of the temperature predictions of the hybrid model is not particularly low. Similarly, there is not much room for improvement for many of the prediction targets, with the exception of tensile and yield strength, recrystallization, and grain size.

Target Variable	Mean	MAE data-driven only	MAE hybrid	Performance Gain in %
Yield Strength	281	11.5	10.3	11
Recrystallization Rate	55.5	4.53	3.67	19
Grain Size	107	10.4	8.39	19
Energy	0.00281	0.000403	0.000318	21
Temperature	1061	2.75	3.20	-16
Degree of Deformation	0.107	0.00549	0.00521	5
Rolling Force	0.646	0.0257	0.0234	9
Rolling Torque	15.3	0.737	0.617	16
Tensile Strength	566	6.04	5.64	7

Table 4: All target variables with their mean values, the MAE of the predictions by the data-driven and hybrid machine learning approach, as well as the performance gain of the hybrid approach in %. The data has been averaged over all passes.

5 Pre-Training and Transfer of Reinforcement Learning Agents

The hybrid model described in the previous section is created to serve as a less computationally expensive replacement for the FRM. The intention is to use the hybrid model to quickly train reinforcement learning agents for pass schedule design and perform a subsequent transfer from the hybrid model to the FRM.

We use a similar agent and environment setup as described by Scheiderer et al. [18], where the environment offers an action space comprising the next workpiece height and the pause time after the next pass, as well as an observation space consisting of the current workpiece height, temperature, and grain size. The agent uses the soft actor-critic algorithm [5] and iteratively creates a pass schedule with up to 5 passes to transform a metal workpiece. To learn to design suitable pass schedules, the agent is rewarded after every t-th pass according to the reward function proposed by Scheiderer et al. [18]:

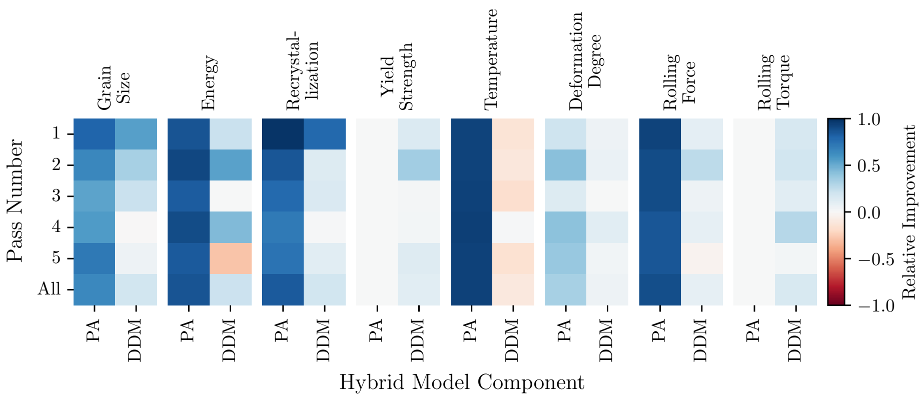

Figure 3: Relative improvement in MAE of hybrid model over each of its components: data-driven model (DDM) and physical approximation (PA). Yield strength and rolling torque are not computed by the physical approximation and the corresponding relative improvement of the hybrid model is hence denoted as 0.

$$R_t = 1 - \left(\frac{\alpha}{h_{max}} (\Delta h_t)^2 + \frac{\beta}{d_{max}} (\Delta d_t)^2 + \frac{\epsilon}{\epsilon_{max}} \sum_{k=1}^{t} e_k \right),$$

where h_t and d_t denote the distance to the goal height and grain size, respectively, while e_t corresponds to the used energy in the t-th pass. Each component of the equation is scaled by the associated maximal value h_{max}, d_{max} or e_{max}. The influence of each component can be controlled by adjusting α, β and ϵ. Note, that a reward of 1 is usually not possible, because the execution of a pass schedule requires energy.

As described in Section 3, current approaches to pass schedule design with reinforcement learning typically train agents directly on a FRM, which can be time-consuming and significantly slow down the development of new agent or environment designs. We replicate such an approach to serve as a baseline to which the pre-training and transfer approach is later compared. To facilitate this, we access the FRM RoCaT through a SaaS architecture and train an agent as described above for 2500 environment steps, which corresponds to a training time of 43.59 hours in our setup. As can be seen in in the learning curve displayed in Figure 6 (left), the agent shows good convergence behaviour in the achieved rewards, with some outliers due to exploration.

In contrast, training on the proposed hybrid model takes much less time. Again, we train an agent for 2500 environment steps, which corresponds to a training time of 0.17 hours on the hybrid model. As displayed in Figure 6 (middle), the agents converge to similar rewards on the hybrid model as on RoCaT. While the training time hence appears to be much reduced for similar results, the trained agent is expected to require some additional fine-tuning on RoCaT due to the approximate nature of the hybrid model.

To establish how much more efficient the overall approach consisting of pre-training and transfer phase is compared to sole training on RoCaT, the pre-trained agent is taken and trained on RoCaT for final adjustments. For this final phase, an additional 500 training steps are executed. As displayed in Figure 6 (right), the additional training

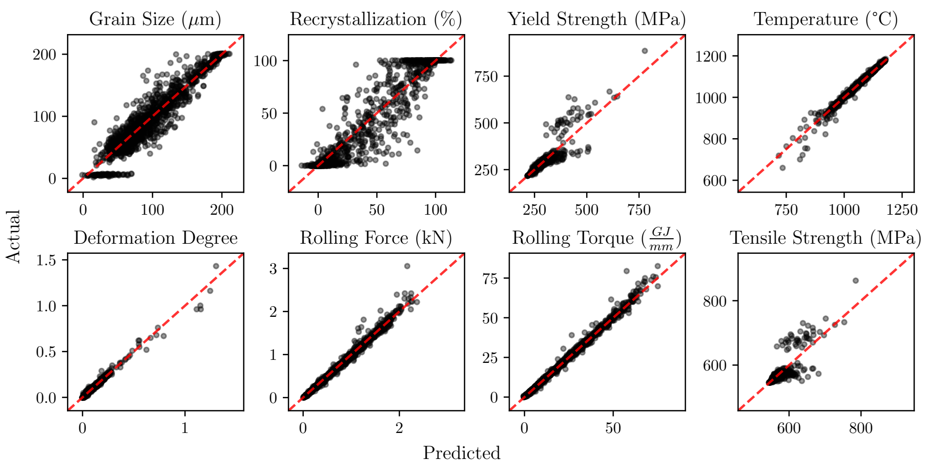

Figure 5: Scatter plots showing the discrepancies between predictions produced by our hybrid model and the actual values produced by the FRM for several different targets. Data points on the dashed red line indicate perfect predictions, while those above and below the line indicate under- and overestimation, respectively.

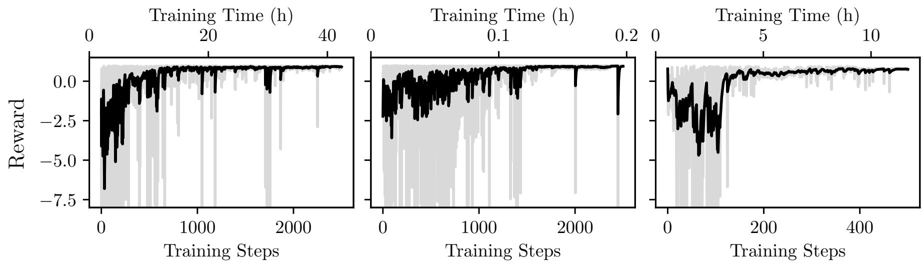

Figure 6: Learning curves of agents trained on RoCaT (left), on the hybrid model (middle), and transfer learning on RoCaT of an agent pre-trained on the hybrid model (right). The learning curves shown in black are exponentially weighted moving averages ($\alpha = 0.1$) with the original data being displayed in light grey. All curves are the average of three separately trained agents.

phase is needed, as the agent's rewards start significantly below those achieved on the hybrid model. The agent is nevertheless able to leverage its learned behaviour from the pre-training stage, which becomes apparent by the quickly rising rewards, which converge much more quickly compared to training from scratch. Overall, the combined training time of pre-training and transfer phase is 7.58 hours, while the training time of the baseline is 43.59 hours. It could be argued that the baseline training time is longer than necessary, as convergence behaviour can be observed starting at roughly half of the training duration. In any case, the two-stage training approach achieves a significant reduction in overall training time compared to the baseline.

6 Conclusion

We have proposed a hybrid model to approximate an FRM with the main purpose of using it to pre-train reinforcement learning agents for pass schedule design. Our model consists of a data-driven core, which is further augmented by a set of physical approximations of selected variables relevant for pass schedule design. The evaluation of the hybrid model shows that, on average, it outperforms each of its individual components.

While the predictions of our model are not as accurate as the FRM, its efficiency allows for quick training of reinforcement learning agents for pass schedule design. Although agents pre-trained on the hybrid model require some additional training on the FRM, the overall training time of this two-stage approach is significantly reduced while achieving similar results.

In the future, this reduction in training time will allow for more rapid development cycles of reinforcement learning agents. The agent used in our experiments only has access to a small, selected set of observations out of all the variables calculated by the FRM and our hybrid model. By incorporating additional observations, the performance of the agent can potentially be improved further in the future. Additionally, extending the observation and action spaces will particularly allow for the design of agents with more extensive generalization capabilities across hot rolling scenarios with varying initial geometries, materials, and material properties.

Acknowledgments

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2023 Internet of Production – 390621612.

References

[1] World crude steel production from 2012 to 2022. `https://www.statista.com/statistics/267264/world-crude-steel-production/`. Accessed: 2024-01-03.

[2] D. Bland and H. Ford. The calculation of roll force and torque in cold strip rolling with tensions. *Proceedings of the Institution of Mechanical Engineers*, 159(1):144–163, 1948.

[3] H. Caglayan and J. Beynon. Slimmer sheffield leicester integrated model for microstructural evolution in rolling. In *Int. Conf. Modelling of Metal Rolling Processes. The Inst. of Materials, Imperial College, London*, pages 274–282, 1993.

[4] T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[5] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor, 2018.

[6] R. Hwang, H. Jo, K. S. Kim, and H. J. Hwang. Hybrid model of mathematical and neural network formulations for rolling force and temperature prediction in hot rolling processes. *Ieee Access*, 8:153123–153133, 2020. Publisher: IEEE.

[7] C. Idzik, J. Gerlach, J. Lohmar, D. Bailly, and G. Hirt. Application of Reinforcement Learning for the Design and Optimization of Pass Schedules in Hot Rolling. In *Congress of the German Academic Association for Production Technology*, pages 71–80. Springer, 2022.

[8] C. Idzik, D. Hilger, N. Hosters, M. Kemmerling, P. Niemietz, L. Ortjohann, J. Sasse, A. Serafeim, J. Wang, D. Wolff, et al. Decision support for the optimization of continuous processes using digital shadows. In *Internet of Production: Fundamentals, Applications and Proceedings*, pages 1–22. Springer, 2023.

[9] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30, 2017.

[10] M. R. Liebenberg. Autonomous agents for the world wide lab: artificial intelligence in the manufacturing industry, 2021.

[11] J. Lohmar, S. Seuren, M. Bambach, and G. Hirt. Design and application of an advanced fast rolling model with through thickness resolution for heavy plate rolling. In *2nd International Conference on Ingot Casting Rolling Forging*, 2014.

[12] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin. Catboost: unbiased boosting with categorical features. *Advances in neural information processing systems*, 31, 2018.

[13] S. Rath, P. Sengupta, A. Singh, A. Marik, and P. Talukdar. Mathematical-artificial neural network hybrid model to predict roll force during hot rolling of steel. *International Journal of Computational Materials Science and Engineering*, 2(01):1350004, 2013. Publisher: World Scientific.

[14] S. Rath, P. Talukdar, and A. Singh. Methodology of developing mathematical-ANN hybrid model based mill setup model for hot strip rolling. *International Journal for Computational Methods in Engineering Science and Mechanics*, 22(5):358–365, 2021. Publisher: Taylor & Francis.

[15] S. Ray. *Principles and applications of metal rolling*. Cambridge University Press, 2016.

[16] R. Roychoudhury and J. Lenard. A mathematical model of cold rolling-experimental substantiation. In *Proc. 1st Int. Conf. Techn. Plast*, pages 1138–1145, 1984.

[17] P. Saravanakumar, V. Jothimani, L. Sureshbabu, S. Ayyappan, D. Noorullah, and P. Venkatakrishnan. Prediction of mechanical properties of low carbon steel in hot rolling process using artificial neural network model. *Procedia Engineering*, 38:3418–3425, 12 2012.

[18] C. Scheiderer, T. Thun, C. Idzik, A. F. Posada-Moreno, A. Krämer, J. Lohmar, G. Hirt, and T. Meisen. Simulation-as-a-service for reinforcement learning applications by example of heavy plate rolling processes. *Procedia Manufacturing*, 51:897–903, 2020.

[19] J. A. Schey. Introduction to manufacturing processes, 2000. *Mc Craw Hill*, 2000.

[20] J. Wang, Y. Li, R. X. Gao, and F. Zhang. Hybrid physics-based and data-driven models for smart manufacturing: Modelling, simulation, and explainability. *Journal of Manufacturing Systems*, 63:381–391, 2022. Publisher: Elsevier.

[21] Q. Xie, M. Suvarna, J. Li, X. Zhu, J. Cai, and X. Wang. Online prediction of mechanical properties of hot rolled steel plate using machine learning. *Materials Design*, 197:109201, 2021.

[22] A. Özgür, Y. Uygun, and M.-T. Hütt. A review of planning and scheduling methods for hot rolling mills in steel production. *Computers Industrial Engineering*, 151:106606, 2021.