Key Engineering Materials *Submitted: 2025-12-18*
ISSN: 1662-9795, Vol. 1050, pp 9-20 *Revised: 2026-02-24*
doi:10.4028/p-GLpuV9 *Accepted: 2026-02-25*

Online: 2026-04-16

Prediction of Strip Width Deviation in Hot Strip Roughing Mills Based on Machine Learning Regression

Daniel Ewen[1,a *], Johannes Seitz [1,b], Sebastian Kallabis[2,c], Reinhold Franke[2,d], Joachim Denker[3,e], Lukas Lackmann[3,f] and Bernd Kuhlenkötter[1,g]

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

[2]thyssenkrupp steel europe AG, Duisburg, Germany

[3]Asinco GmbH, Duisburg, Germany

[a]ewen@lps.ruhr-uni-bochum.de, [b]seitz@lps.ruhr-uni-bochum.de (*corresponding author) [c]sebastian.kallabis@thyssenkrupp-steel.com, [d]reinhold.franke@thyssenkrupp-steel.com, [e]joachim.denker@th-koeln.de, [f]lukas.lackmann@asinco.de, [g]kuhlenkoetter@lps.ruhr-uni-bochum.de

Keywords: hot strip rolling, prediction, machine learning, predictive quality, width deviation

Abstract. Previous research shows that predicting width deviation inherits a central importance in hot rolling processes, so that the pass planning in the hot strip mill (HSM) can be optimized. These predictions can be enabled using machine learning, complementing analytical formulations of width spread. For reliable production, it is important for the plant operator to be able to control the geometry with high accuracy across the entire plant. Therefore, the width must be accurately known throughout the entire HSM. This paper aims on the prediction of width deviation in early product stages during the roughing mill processes, where the major deformation takes place, and thus also has the most significant influence on the width spread. Therefore, this paper takes industrial data into account which is also used for the roll passing planning. To achieve a prediction during rough rolling for the width after exiting the mill (future state of the strip width), various machine learning algorithms were implemented and tested. The prediction results are evaluated against an inline width measurement, where the XGB model performs best with a Root Mean Squared Error (RMSE) of 1.11 mm. Subsequently, feature importance analyses are used to examine which features are relevant for the prediction result and to elaborate which significance process- and geometry data has on the same strip.

Introduction

Hot strip rolling is an essential process in the steel industry. In the EU-27, hot-rolled steel strip accounts for the largest proportion of hot-rolled products, because it can be processed into a wide variety of other products in subsequent manufacturing steps [1]. At the same time, the HSM is highly relevant in terms of observing geometric tolerances. Studies have shown that defective HSM products, such as under-width and over-width, are problematic in further processing (e.g. cold rolling) and can result in rejection by critical customers [2, 3]. Consequently, controlling width accuracy is an essential challenge due to width spread during hot strip rolling [3]. In the context of process optimization in production systems, the use of machine learning (ML) is becoming increasingly important [4]. Many companies already store huge amounts of data that may contain hidden patterns which can be identified and used to optimize processes, such as hot strip rolling, through the use of ML amongst other data science approaches [5].

Conventional hot strip mills are usually based on two sections in which the product is processed differently. First, after leaving the furnace, the steel slab enters the roughing mill, which is responsible for the major forming work by passing the slab through a rolling stand in a reversing process until it results in a pre-strip. After leaving the roughing section, first geometric measurements are available to control the products' geometric accuracy. Subsequently, the operators can adjust the pre-strip by, for example, trimming the front and end. The final thickness is then achieved in the finishing mill, which usually consists of several rolls in direct succession. Before the strip is coiled, it passes through a cooling section and is then measured geometrically and in cross-section. The schematic structure of a HSM is shown in Figure 1.

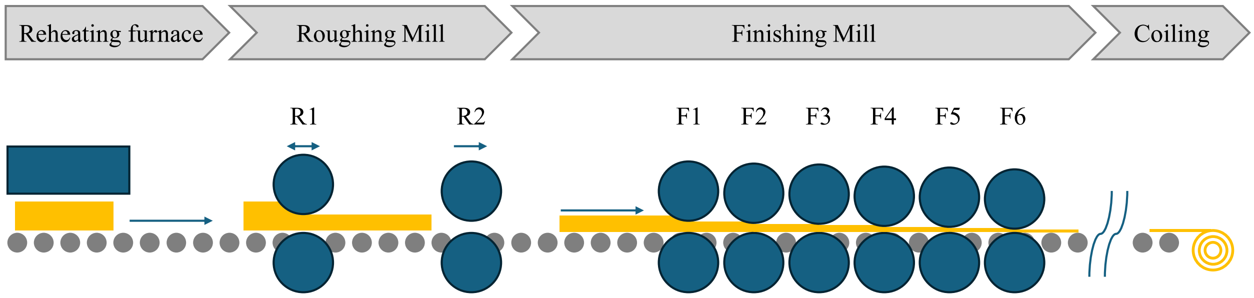

Fig. 1. Schematic structure of a HSM, containing the roughing mill, finishing mill and coiling, exemplified by the facility at thyssenkrupp Steel Europe in Duisburg

Within the scope of strip width control deterministic equations and mathematical models are available to calculate the width spread during hot rolling such as the *Shibahara Spread Model* (SSM) [6]. In addition to mathematical approaches, data-centric approaches are being used within the last decades for predictions due to increasing process data availability. An approach on width spread prediction in hot strip rolling was presented by *Hashemzadeh et al.* and is based on a combination of FEM analysis and Artificial Neural Networks (ANN) whereby they achieve high performance in the simulation environment [7]. When considering industrial data, both nominal (scalar) as well as time-dependent datasets are used for predictions. *Wu et al.* present a hybrid prediction model that incorporates historical and real-time (scalar) data into the prediction [8]. At the same time, they also highlight the absence of sufficient control equipment in HSM. Among the time-dependent approaches, *Latham et al.* developed a model for the categorization of width errors to make them accessible to operators in their decision-making [9]. Some authors, such as *Xin et al.* [10], who developed a time-dependent hybrid model approach, based on Convolutional Neural Networks (CNN) and Long-Short Term Memory (LSTM) models, criticize the accuracy of these purely mathematical models, which are currently often used in production planning. This statement is supported by *Gao et al.* [11], who mention the relevance of real-time measurement systems in the context of width prediction. In a heavy plate mill, the authors therefore investigate the prediction of width spread in finishing processes using Stochastic Configuration Networks (SCN) [11]. By considering two width gauge sensors, before and after the rolling pass, as well as process parameters like temperature, reduction ratio and rolling forces, *Dong et al.* demonstrate that the forming work plays a decisive role in width deviation [12]. They conclude that a high thickness reduction has a major influence on width deviation which essentially occurs at the roughing mill during the forming of the slab into pre-strip. Therefore, research is being also conducted in prediction of the width deviation during rough rolling. An example of this work is an hybrid approach by *Zhong et al.* [13, 14], in which the authors used ML to optimize the SSM for width prediction in a HSM roughing mill. Another approach on roughing mill exit width prediction is presented by *Ji et al.*, who optimized a linear regression model by using genetic algorithm for hierarchical clustering. Concurrently, they also emphasize the relevance of transparent prediction models for industrial applications and refer to some prediction models as black boxes [15].

This paper addresses the prediction of the width when leaving the roughing mill (as pre-strip) and uses industrial data as the basis for predictions. This investigation aims to develop a prediction model for the width spread at the end of the roughing mill and to identify relevant knowledge to achieve further improvements in the future. For this purpose, decision tree models and neural networks are created, optimized, and evaluated for prediction. Finally, the results are interpreted in an industrial context with regard to the predicted overall width compared to the measured width, and the relevance of various features is presented.

Methodology

The methodology for width deviation prediction for the exit of the roughing mill section is based on data preprocessing, followed by the construction of a data pipeline and the selection of appropriate ML models for prediction and comparison. The width predictions are based on industrial data and provided by a HSM from thyssenkrupp Steel Europe AG located in Duisburg, Germany. The dataset contains various features across the entire HSM (e.g. geometry measurement data, rolling forces, calculated process settings and target geometry data).

Data Preprocessing

The first step, therefore, was to reduce the data set to the important part for this research. Based on the feature names, the data set was first narrowed down manually. To do this, all process data that occurred after the roughing mill was removed by for example eliminating features that deal with the finishing mills (F1 – F{i}) or other parts of the HSM that are located behind the roughing section. This significantly reduced the number of features in the data set. The relevant plant section is illustrated in Figure 2, where everything to the left of the roughing section (reheating furnace) as well as the information from the roughing section itself is included and everything to the right (finishing mill) is excluded.

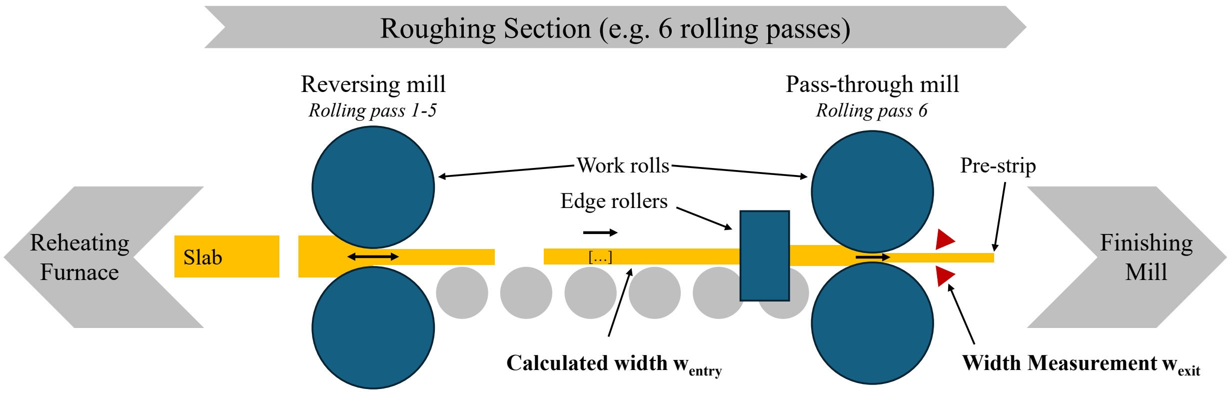

Fig. 2. Structure of the investigated HSM roughing section.

Prediction methodology

This work targets to predict the width spread on the final pre-strip, which can be calculated as the difference between the exit width of the pre-strip (shown in as red triangle in figure 2) and the entry geometry after the first five rolling passes. It is important to note, that the illustrated width measurement system in figure 2 is the first actual measurement of the width behind the furnace. The calculation for the width spread (target = y) is shown in equation (1).

$$y = w_{exit} - w_{entry} \tag{1}$$

Moreover, it is imperative for the plant operators to be aware of the selected target value at the earliest opportunity, thus enabling the assessment of product quality in advance and, if necessary, facilitating correction during the process. In order to predict the width spread, deterministic width calculations from the HSM according to the rolling passes 1-5 and process settings such as forces, roller settings, steel grades, etc. are considered in the prediction. The prediction methodology is therefore based on the calculated width after the fifth rolling pass and predicts the width after edge rollers and thickness reduction of the sixth pass-through rolling pass.

With regard to possible limitations in the accuracy of predictions, it must be said that w_{exit} and the deterministic width values such as w_{entry}, which represents the width after the fifth rolling pass (calculated width R5), are recorded with a maximum resolution of 1 millimeter. Both, the measured width w_{exit} and the calculated width w_{entry} (width after the 5th rolling pass) are average values over the entire strip length and are therefore independent of direction. Moreover, w_{entry} is based on a deterministic width formula, which is integrated into the plant control system by the plant manufacturer. This is a custom-developed formula, which is why it cannot be referenced. According to the plant manufacturer, this is based on variables such as temperature, thickness reduction, steel grade, forces, and roll gap settings. Hence, the performance of the width formula cannot be verified, for example, under different steel grades or through the combination of vertical edgers with thickness reduction and can therefore potentially generate biases in the prediction model. For this reason, the suitability of the calculated width as a basis for the prediction model will be revisited in the discussion. The complete data set is based on scalar data and comprises a total of 19,785 hot rolled strips and refers to the entire HSM (furnace, roughing mill, finishing mill, etc.). However, only some of these features can be used to avoid data leakage. Firstly, the data is reduced manually to only contain features from before the pass-through mill (R6). This means that all information recorded after the roughing mill, such as the target width of the roughing mill, information from the finishing mill, or even the final width of the finished coil, is removed from the data set. The data contained in the prediction model always refers only to the data of the strip currently being rolled. Table 1 shows a list of relevant features (X) available for predictions.

Table 1. Selection of features in the context of roughing mill

Features (X)	**Unit**
Rolling force (R1-R5)	kN
Temperature (R1-R5)	°C
Calculated width (R1-R5)	mm
Horizontal roller position (R1-R5)	mm
Vertical roller position (R1-R5)	mm
Steel grade	-
Number of reversing passes	-

For better understanding, the prediction methodology is illustrated in Figure 3, where the dashed green line marks up the feature space containing the complete data from the HSM roughing mill. This results in the input features X for further processing. Subsequently, the target (y) is defined, and the data set is divided into training, validation, and test data.

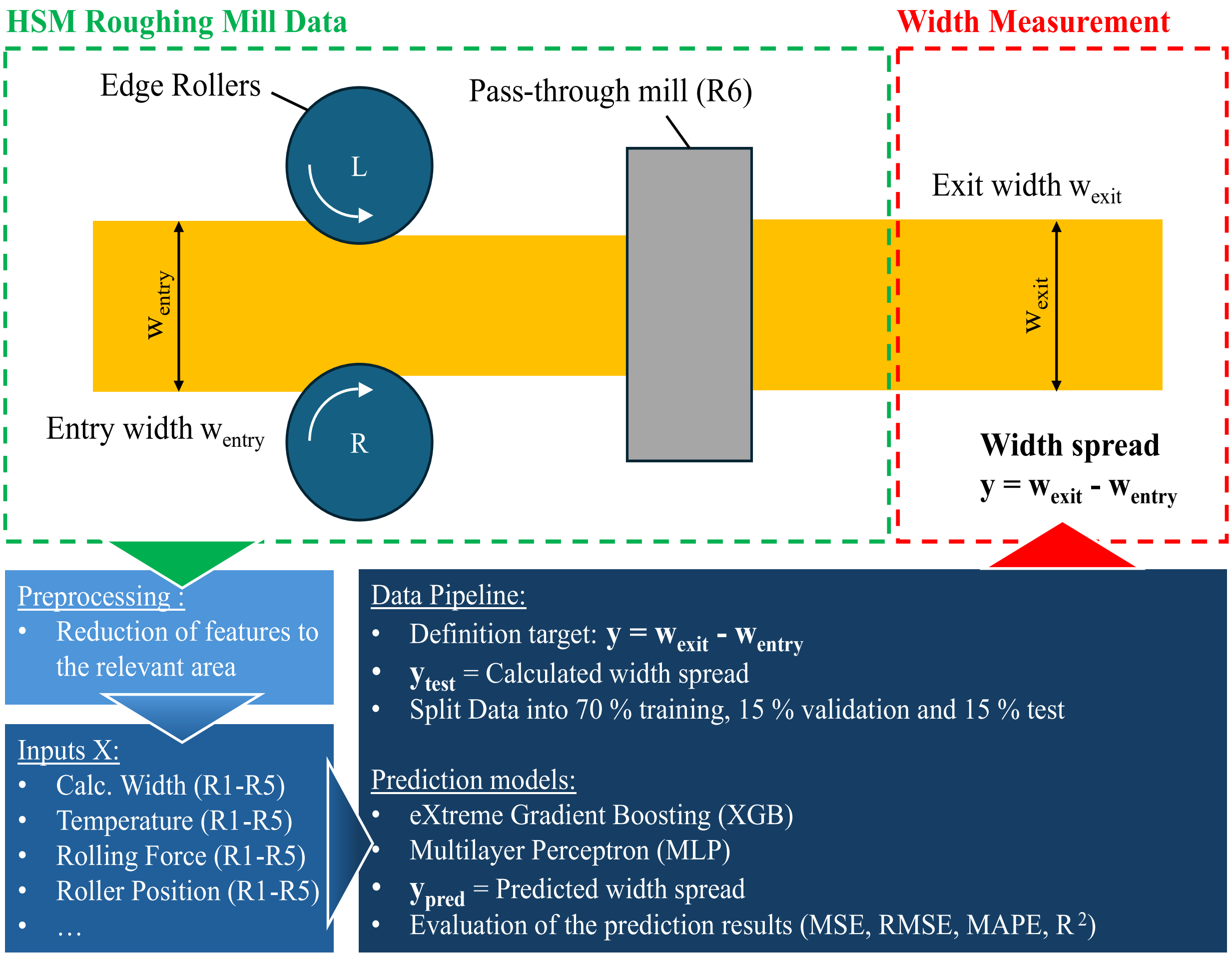

Fig. 3. Schematic illustration of the prediction methodology.

Initially, the data set was divided into 70 % training data, 15 % validation data, and 15 % test data and a minimum-maximum-scaler has been applied. Instead of presenting a broad survey of machine learning architectures for this task, this paper focuses on two distinct approaches to provide perspective, by following not only one approach. For this reason, a decision tree-based model and a neural network were used to compare the performance between the two models, as they differ in terms of their suitability for different approaches. It should also be noted that decision tree-based models, such as *Random Forest* (RF) or *eXtreme Gradient Boosting* (XGB), have the advantage of explainability. This means that influencing factors can be explored by calculating feature importance [16], for example. While these models may seem logical to the developer, they can also reveal hidden patterns in the data. Unlike decision tree models, who build a data-driven decision path and are therefore well suited for structured data sets, neural networks are highly suitable for discovering complex patterns in data sets, but they also require a large amount of data, and the predictions are not comprehensible to the developer [17]. In comparison to the decision-tree model, a *Multilayer Perceptron* (MLP) model and *Convolutional Neural Network* (CNN) was therefore also constructed as a neural network. Even though the models differ in their fundamental structure, the same method was used for optimizing the model parameters with 5-fold cross validation: *Bayesian optimization* (BO) [18]. Thereby, due to the fast convergence, 50 optimization trials were run to achieve the best parameter setting.

Evaluation

During training, it was observed that XGB outperformed RF and that the MLP outperformed CNN. Therefore, the evaluation will be presented based on XGB, a decision tree-based model, and on MLP, representing a deep learning network. Various error metrics are presented to evaluate the prediction results of the different models. These include Mean Squared Error (MSE), its root (RMSE), the coefficient of determination (R^2), and the Mean Absolute Error (MAE). While MSE, RMSE, and MAE are suitable for process-specific evaluation because they are expressed in absolute units (mm^2 for MSE, mm for RMSE and MAE), R^2 is more appropriate for comparisons across datasets, as it is dimensionless. The smaller the MSE, RMSE and MAE values and the closer the R^2 is to 1, the more accurate is the prediction model.

Results and Discussion

As previously explained, two different models were trained to predict the width spread for the final pre-strip, based on the results of the parameter optimization. The model configurations (settings) are shown in Table 2 (XGB) and Table 3 (MLP).

Table 2. Hyperparameter space and Optimization result for the XGB model

Parameter	Description	Feature Space	Setting
Estimators	Number of trees	{100, 1000, step size = 50}	800
Max depth	Max. tree depth	{3, 12, step size = 1}	5
Learning rate	Learning rate	{0.003, 3, log = true}	0.083
Samples per tree	Share of samples per tree	{0.5, 1.0}	0.668
Features per tree	Share of features per tree	{0.5, 1.0}	0.985
Gamma	Minimum loss per split	{0, 5}	0.566
Min. child weight	Sum of instance weights	{0, 10}	6.989
L1-regularization	L1-regularization	{$1*10^{-8}$, 10, log = true}	0.009
L2-regularization	L2-regularization	{$1*10^{-8}$, 10, log = true}	0.349

Table 3. Hyperparameter space and Optimization result for the MLP model

Parameter	Description	Feature Space	Setting
n_layers	Number of hidden layers	{1, 10}	4
Activation	Activation of neurons	{relu, tanh, elu}	elu
Dropout rate	Dropout rate	{0, 0.5}	0.175
Learning rate	Learning rate	{$1*10^{-4}$, $1*10^{-2}$, log = true}	$1.5*10^{-4}$
Batch Size	Batch Size	{16, 32, 64, 128}	128
Epochs	Training epochs	{50, 300}	140
n_Units_L0	Number of neurons layer 0	{32, 512, step = 32}	416
n_Units_L1	Number of neurons layer 1	{32, 512, step = 32}	384
n_Units_L2	Number of neurons layer 2	{32, 512, step = 32}	448
n_Units_L3	Number of neurons layer 3	{32, 512, step = 32}	384

The performances of the XGB and MLP on the test data is presented in Table 4, whose indicates, that the XGB model (decision tree) outperforms the prediction of the MLP model. The predictions are visualized in Figure 4 in a regression diagram.

Table 4. Evaluation of the prediction results.

Model	MSE [mm^2]	RMSE [mm]	MAE [mm]	R^2 [-]
XGB	1.2407	1.1139	0.8236	0.8204
MLP	1.3878	1.1780	0.8989	0.7991

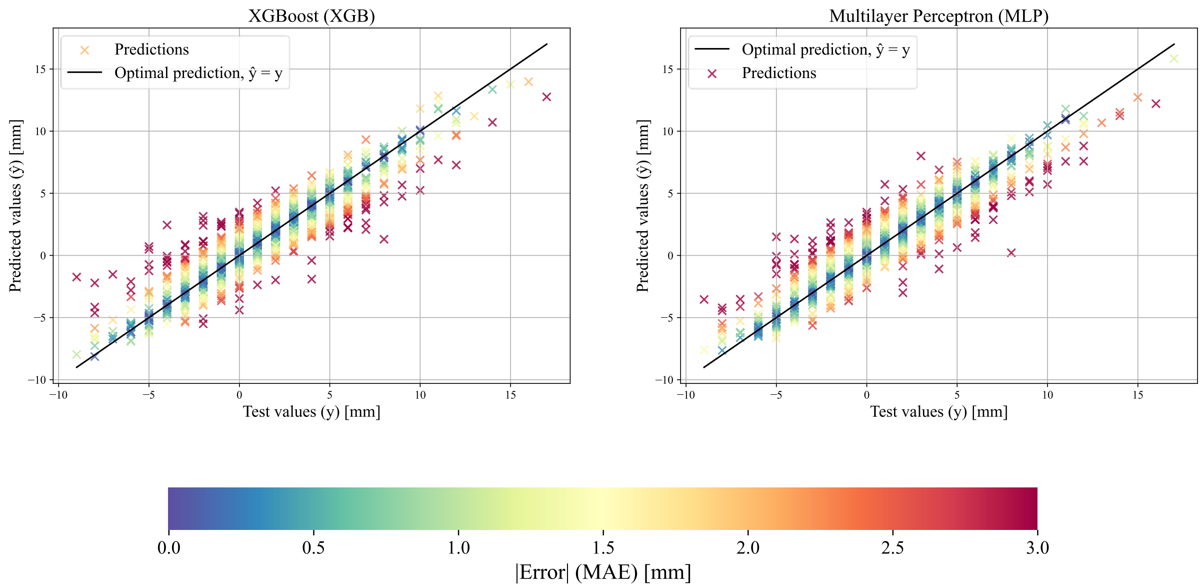

Fig. 4. Width Spread Prediction Results for the optimized models: XGB (left) and MLP (right).

Both models are capable of predicting the width spread, as can be seen in the regression diagram (Figure 4), where the prediction instances follow the bisector line. Deviations persists along the bisector line, but the overall trend is captured, concluding that a prediction sensitivity is given. The color scale shown is the MAE scale, which indicates the absolute prediction error in millimeters. The color matching is chosen to separate minor errors (blue) from major errors (red), with all errors greater than 3 millimeters being displayed in red. The results show that the XGB model provides the best prediction results with an RMSE of 1.1139 mm and a MAE of 0.8236 mm (R^2 = 0.8204). In order to better interpret the results in the context of the hot rolling mill, a comparison must be made with the total width. In this context, w_{ist} indicates the actual width measured by the width measurement system at the roughing mill exit and w_{pred} indicates the predicted width derived from equation (2). Statistical robustness is relevant for evaluation of prediction results in an industrial context. Therefore, Δw (3) is calculated as an evaluation metrics of the individual predictions and represents the deviation from w_{ist} to w_{pred}.

$$w_{pred} = w_{entry} + y_{pred} \tag{2}$$

$$\Delta w = w_{ist} - w_{pred} \tag{3}$$

Thus, it is possible to predict the final width at the roughing mill after the fifth rolling pass, which also represents an important input variable for the finishing mill. The prediction accuracy (standard deviation) is approximately 1.1 millimeters (XGB). In relation to DIN ISO 10051, which pertains to the specifications for hot-rolled flat products, the permissible range of width is delineated as a tolerance of between 0 millimeters and a maximum of 20 millimeters in over width [19]. The present research findings suggest that the determined prediction scale can be considered significantly helpful in decision-making. However, it should be noted that absolute deviations of 3 millimeters and more still require a careful review of the prediction result. This requires further investigations, such as segmentation of steel grades and/or input geometry clusters.

The statistical evaluation of the hyperparameter optimized models in Figure 5 shows that the standard deviation on the training data set is significantly better than the test data set. The training RMSE of the hyperparameter-optimized models is 0.69 mm (MLP) and 0.61 mm (XGB), while the test and validation data are both between 1.1 mm and 1.2 mm for XGB and MLP (see Figure 5). This results in the assumption that overfitting is in evidence and that the models lack effective generalization. To check the influence of hyperparameter optimization on overfitting, the models are simplified. As a result, the search spaces for hyperparameter optimization were considerably reduced, leading to a decrease in the model's depth and its complexity. It was determined that a simplified MLP model with one hidden layer and 96 neurons (compare Table 3) and a simplified XGB model with 50 estimators and a maximum depth of 8 (compare Table 2) performed slightly worse but in a similar range of results (Train-RMSE: approx. 0.7 [mm], Test-RMSE: approx. [1.3] mm). Summarized, hyperparameter optimization did not result in more overfitting of the XGB and MLP model than the simplified model and has slightly improved the prediction performance for both models. It can be derived that the high deviation in model performance between train and validation or test data is therefore not due to optimization-related overfitting, but rather to the heterogeneous data structure, which suggests that the width spread cannot be generalized for comparable roll gap settings (work roll position or forces) and strip input conditions (temperature, entry width, etc.). Moreover, the choice of the machine learning algorithm, such as MLP or XGB, and its underlying learning strategy can also influence data-dependent predictive performance. The root cause of this must be investigated by conducting more in-depth research into the data structure in future research work.

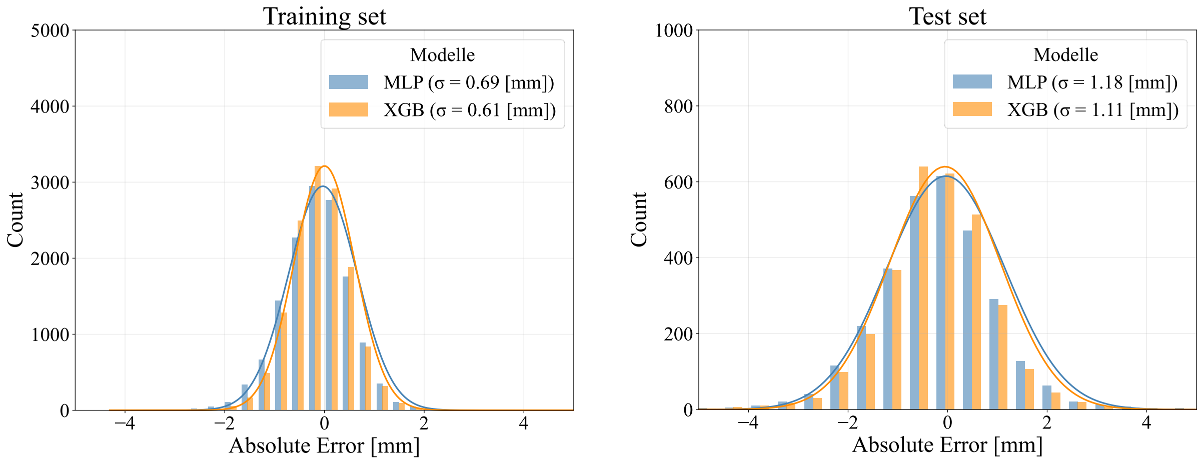

Fig. 5. Absolute error distribution of the optimized XGB (yellow) and MLP (blue) models.

Overall, it can be concluded that the hyperparameter optimized XGB model has the best prediction performance as the standard deviation on the training data set archives 0.61 mm, while on the test data set it yields 1.11 mm. In an industrial context, the research results presented show that the exit width of the roughing section can be predicted with a standard deviation of 1.11 mm.

As already introduced in the presentation of the various models, a decisive advantage of tree-based prediction models is that the importance of the features can be evaluated. For this reason, a feature importance analysis was performed to explain which features have a high influence on the prediction. The top 6 results of the feature importance analysis are illustrated in Figure 6, as these six features have the most significant influence. The illustration is sorted by the top 6 features of the best performing XGB model. It can be seen that a major influence on decision-making comes from the calculated entry width for the pass-through rolling mill (top 3), as well as the positioning of the edge rollers (top 2). As already explained in the prediction methodology, the calculated width R5 is a deterministic value based on roll gap setting, temperature, forces and steel grades. The high influence factor is therefore explainable and at the same time shows that the calculated width R5 (which may contain uncertainties, among other things with regard to the maximum resolution of 1 mm) does not cause an extremely large bias. Nevertheless, future work should also investigate the extent to which width measurements between the 1st and 5th roll passes, as compared to the current calculated width values, improve prediction performance. Further research is also needed into the resolution of the measurement system after the rough rolling section, as a resolution of 1 mm versus a resolution of 0.1 mm can reduce model performance due to rounding errors.

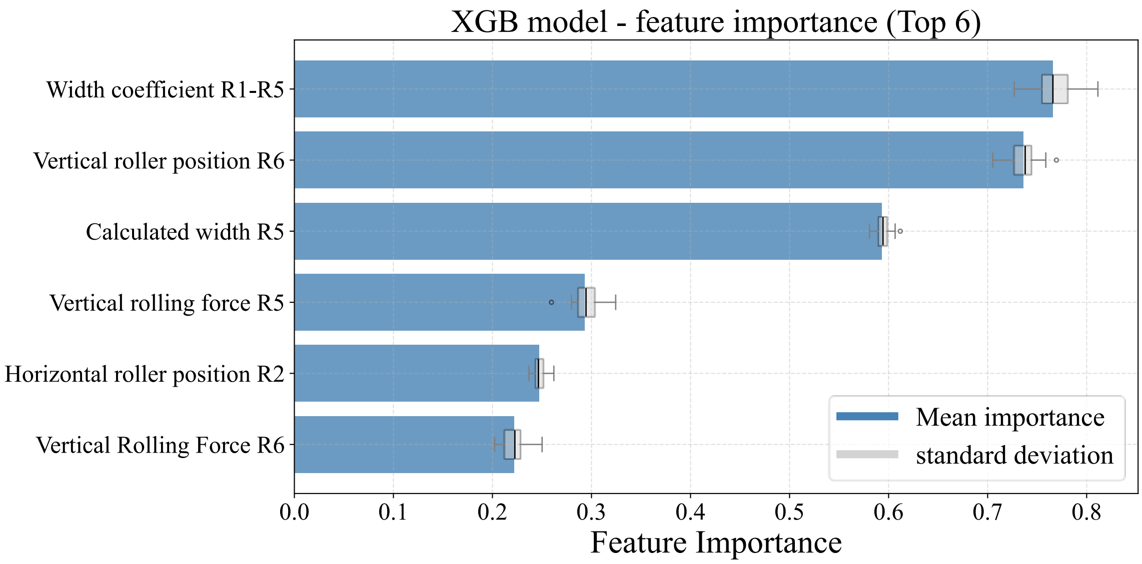

Fig.6. Feature Importance Analysis sorted by influence of the top 6 for XGB predictions

The greatest influence derives from a width coefficient of the reversing mill (top 1). This is an experience-based correction factor based on a matrix consisting of steel grades and thickness clusters. The correction factor "Width coefficient R1" is based on historical experience from plant operators and is adjusted if the difference between the target and actual width at the reversing stand becomes too large. This demonstrates the relevance of combining expert knowledge with data-driven approaches. It is evident that the rolling forces of the edge rollers prior to the pass-through mill (top 6), after the fifth pass on the reversing mill (top 4), and the vertical setting of the working rolls from the second reversing pass (top 5), contribute significantly to the prediction result. This leads to the assumption that information from process steps that have already been executed can be relevant for the prediction. In summary, it must be stated that with the width coefficient R1-R5 and the calculated width R5, two of the top three features are based on calculated models. As previously stated, the availability of real width measurement data as an input variable in the R6 rolling stand would be pertinent in order to entirely negate bias based on previous modelling.

A comparison of the presented solution with other research on width spread prediction in roughing mills remains challenging, due to limited availability of data, and only two studies with direct comparability were identified. For this reason, studies on width prediction on the finishing mill were also considered for benchmarking.
As already introduced in the beginning of this research paper, *Zhong et al.* [13, 14] provide a hybrid approach for modelling the width spread in the roughing mill section by using ML-methods to improve the SSM. In their first research work, *Zhong et al.* yielded a RMSE of 3.4777 mm with their proposed improved width spread model, compared to MLP with a RMSE of 5.6417 mm [13]. The authors significantly improved this model as Improved SSM (ISSM) in another research paper published around the same time, achieving an RMSE of 1.6440 mm (R^2 = 0.9655) against SSM with 6.1212 mm (R^2 = 0.5219), as well as MLP with a RMSE of 5.7743 mm (R^2 = 0.5745) [14] on the test data [14]. Another study that can be referred, as they use a similar procedure but on the finishing mill is the research by *Wu et al.* [8]. The authors attempt to predict the output geometry of the finishing line based on the input geometry in the first rolling stand, considering all process parameters. Thereby, the authors achieve a MSE of 2.6508 [mm^2], which results in a RMSE of 1.6303 mm as well as a MAE of 1.1095 mm (R^2 = 0.8091) [8].
A comparison of the presented results in this work with those in the literature suggests that a better prediction accuracy is achieved for this dataset. Considering the approach taken by [13, 14] with our approach, a main differences in feature selection can be identified. While in this paper the historical process data from the furnace to the last pass are involved, [13, 14] only consider the entry geometry and process setting of one rolling pass. This supports the earlier assumption that historical influences

have a significant impact on future prediction results. In order to emphasize this assumption, the data set was also shortened to the prediction of the last sample on a trial basis, which also reduced the prediction accuracy and resulted in slightly poorer performance than [14]. Finally, the comparison with existing studies suggests that including historical process information improves prediction accuracy and provides additional insight into the mechanisms influencing width spread.

Conclusion

This study addresses a current challenge in the field of hot strip rolling in industrial plants, investigating the width spread in the roughing mill. Since the roughing mill is responsible for the essential forming work in the HSM, it is of major importance in the context of width spread during rolling. For this reason, this work provides a hyperparameter optimized XGB and MLP to evaluate their capability for the exit strip prediction of a roughing mill. The results demonstrate that both models possess the capacity to recognize the prediction trend, with the XGB model exhibiting a marginally improved performance, yielding a RMSE of 1.11 mm ($R^2 \approx 0.82$) and the MLP model performing slightly worse with a RMSE of 1.18 mm ($R^2 \approx 0.80$). From a scientific perspective, this shows that predictions at the roughing mill are feasible, and the results are marginally improved with this dataset from comparable publications. Moreover, the feature importance analysis of the XGB model reveals that parameters from previous steps contribute to the performance of the model, which is also an important insight for industrial applications as the predictions are comprehensive and offer more insights. Also relevant in an industrial context is the conclusion that the prediction model is capable of predicting the measured exit width of the roughing mill within a standard deviation of 1.11 mm. Even though the XGB model already achieves good prediction results, limitations remain, as the defined target has a maximum resolution of 1 mm. Furthermore, due to the lack of in-line measurements the prediction method considers calculated width data for the first five rolling passes, the accuracy of which cannot be verified, as they are calculated by deterministic equations from the HSM pass schedule. Finally, it can be shown that previously recorded process and geometry data from the same strip not only have an influence on the prediction result but also improve it. Looking ahead, this motivates further investigation into considering feature correlation analysis as well as the development of high-resolution width measurement data for roughing mills to enhance further improvements.

Acknowledgements

The investigations were carried out as part of the *WBW-Smart* research project, funded as part of the EFRE/JTF program NRW 2021-2027. It is also co-financed by the European Union.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The industrial data set is confidential and therefore will not be shared.

References

[1] Wirtschaftsvereinigung Stahl, "Statistisches Jahrbuch der Stahlindustrie: 2024 | 2025," [Online]. Available: https://www.wvstahl.de/publikationen/statistisches-jahrbuch-der-stahlindustrie-2024-2025/

[2] H. Tiensuu, S. Tamminen, O. Haapala, and J. Röning, "Intelligent methods for root cause analysis behind the center line deviation of the steel strip," *Open Engineering*, vol. 10, no. 1, pp. 386–393, 2020, doi: 10.1515/eng-2020-0041.

[3] LLM Group, *Discussion on the production process and common quality problems of hot rolled strip steel.* [Online]. Available: https://www.lmmgroupcn.com/discussion-on-the-production-process-and-common-quality-problems-of-hot-rolled-strip-steel/ (accessed: Oct. 22 2025).

[4] D. Weichert, P. Link, A. Stoll, S. Rüping, S. Ihlenfeldt, and S. Wrobel, "A review of machine learning for the optimization of production processes," *International Journal of Advanced Manufacturing Technology*, vol. 104, 5-8, pp. 1889–1902, 2019, doi: 10.1007/s00170-019-03988-5.

[5] X. Shu and Y. Ye, "Knowledge Discovery: Methods from data mining and machine learning," *Social science research*, vol. 110, p. 102817, 2023, doi: 10.1016/j.ssresearch.2022.102817.

[6] T. Shibahara, Y. Misaka, T. Kono, M. Koriki, and H. Takemoto, "Edger Set-up Model at Roughing Train in Hot Strip Mill," *Tetsu-to-Hagane*, vol. 67, no. 15, pp. 2509–2515, 1981, doi: 10.2355/tetsutohagane1955.67.15_2509.

[7] A. Hashemzadeh, F. E. Bock, C. Hol, K. Schutte, A. Cometa, C. Soyarslan, B. Klusemann, and T. van den Boogaard, "An analytical predictor machine learning corrector scheme for modeling lateral flow in hot strip rolling," in *Materials Research Proceedings*, 2025, pp. 2002–2011, doi: 10.21741/9781644903599-215.

[8] W. Wu, W. Peng, Y. Liu, J. Liu, X. Li, D. Zhang, and J. Sun, "Predictive modeling of strip width based on incremental learning and adaptive-weight fusion during the hot rolling process," *Journal of Manufacturing Processes*, vol. 142, pp. 157–176, 2025, doi: 10.1016/j.jmapro.2025.03.091.

[9] S. Latham and C. Giannetti, "A Tool to Combine Expert Knowledge and Machine Learning for Defect Detection and Root Cause Analysis in a Hot Strip Mill," *SN Computer Science*, vol. 4, no. 5, 2023, doi: 10.1007/s42979-023-02104-5.

[10] Y. Xin, Z. Zhang, Z. Zhong, and Y. Li, "Lateral spread prediction based on hybrid CNN-LSTM model for hot strip finishing mill," *Materials Letters*, vol. 378, p. 137594, 2025, doi: 10.1016/j.matlet.2024.137594.

[11] H. Gao, Y. Qin, H. Yuan, X. Li, J. Cao, F. Luan, and D. Zhang, "Temporal online self-learning stochastic configuration networks: A study on strip deviation prediction," *Information Sciences*, vol. 689, p. 121446, 2025, doi: 10.1016/j.ins.2024.121446.

[12] Z. Dong, X. Li, F. Luan, C. Cui, J. Ding, and D. Zhang, "Rolling theory-guided prediction of hot-rolled plate width based on parameter transfer strategy," *ISA transactions*, vol. 146, pp. 352–365, 2024, doi: 10.1016/j.isatra.2024.01.013.

[13] Y. Zhong, J. Wang, J. Xu, and J. Rao, "Strip width spread prediction in rough rolling process based on mechanism modeling and optimization," *Journal of Iron and Steel Research International*, vol. 30, no. 12, pp. 2416–2424, 2023, doi: 10.1007/s42243-023-01085-2.

[14] Y. Zhong, J. Wang, J. Xu, J. Rao, and K. Dang, "Data-driven width spread prediction model improvement and parameters optimization in hot strip rolling process," *Applied Intelligence*, vol. 53, no. 21, pp. 25752–25770, 2023, doi: 10.1007/s10489-023-04818-8.

[15] Y. Ji, S. Liu, M. Zhou, Z. Zhao, X. Guo, and L. Qi, "A machine learning and genetic algorithm-based method for predicting width deviation of hot-rolled strip in steel production systems," *Information Sciences*, vol. 589, pp. 360–375, 2022, doi: 10.1016/j.ins.2021.12.063.

[16] G. Louppe, L. Wehenkel, A. Sutera, and P. Geurts, "Understanding variable importances in forests of randomized trees," *Advances in neural information processing systems*, vol. 26, 2013. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2013/file/e3796ae838835da0b6f6ea37bcf8bcb7-Paper.pdf

[17] P. Roßbach, *Neural Networks vs. Random Forests – Does it always have to be Deep Learning?* [Online]. Available: https://blog.frankfurt-school.de/de/neural-networks-vs-random-forests-does-it-always-have-to-be-deep-learning/ (accessed: Nov. 5 2025).

[18] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Anchorage AK USA, 2019, pp. 2623–2631, doi: 10.1145/3292500.3330701.

[19] *DIN EN 10051:2024-07, Continuously hot-rolled strip and plate/sheet cut from wide strip of non-alloy and alloy steels - Tolerances on dimensions and shape*, Berlin.