Key Engineering Materials
ISSN: 1662-9795, Vol. 1050, pp 81-91
doi:10.4028/p-0V5rM0

Submitted: 2026-01-18
Revised: 2026-02-18
Accepted: 2026-02-18
Online: 2026-04-16

Long Short-Term Memory to Predict the Quality Parameters in the Hot Rolling Process

Tamoghna Majumder[1,a*], Nilesh Thakare[1,b], Holger Brüggemann[1,c], Emad Scharifi[1,d] and Junhe Lian[1,e]

[1]Institute of Metal Forming (IBF), RWTH Aachen University, Germany

[a]tamoghna.majumder@ibf.rwth-aachen.de, [b]nilesh.thakare@ibf.rwth-aachen.de, [c]holger.brueggemann@ibf.rwth-aachen.de, [d]emad.scharifi@ibf.rwth-aachen.de, [e]junhe.lian@ibf.rwth-aachen.de

Keywords: Hot Rolling, Quality-relevant material parameters, Long Short-Term Memory.

Abstract: Accurate prediction of quality-relevant material parameters, such as thickness and grain size, ensures product quality in hot forming processes. This task becomes especially complex in hot rolling, where the sequential and time-dependent nature of the process results in pass schedules with varying numbers of passes and grain size evolution that depends on the deformation history. To address this complexity, this study aims to develop and train a deep learning model based on Long Short-Term Memory (LSTM) networks, which are well-suited for modelling sequential data. As input features for the model, process parameters such as rolling force and rolling temperature are used. Simulation data for both material and process parameters are acquired using Simulation as a Service (SaaS) through the Fast Rolling Model (FRM) called Rolling Calculation Tool (RoCaT), focusing on steel grade S355. RoCaT calculates rolling force, rolling temperatures, austenite grain size, strain, and strain rate by taking the geometry of the material, rolling speed, and pass schedule as input. The performance of the LSTM model is evaluated by analyzing loss curves over training epochs and comparing predicted values to reference data. The maximum relative percentage error for thickness between the LSTM predicted value and the thickness values entered as input to the RoCaT model is 18.81% and 16.56%, respectively, for pass schedules of 15 and 17 passes, with a starting thickness of 205mm. Both the maximum relative percentage error value for the pass schedules is obtained at the initial passes, which indicates that the model finds difficulty in performing the predictions accurately during the initial passes, as compared to the later passes. The percentage of relative error values for grain size is also more pronounced during the initial passes as compared to later passes for both pass schedules. The statistical validation is performed on the denormalized data using metrics such as Root Mean Squared Error (RMSE), Normalized Root Mean Square Error (nRMSE), Mean Absolute Error (MAE), and R-Squared (R^2), which demonstrate the model's ability to predict key material properties in the hot rolling process reliably. The RMSE, nRMSE, MAE, and R^2 values for thickness are obtained as 45.21mm, 0.11 (-), 18.75mm, and 0.81 (-), respectively. For grain size, the corresponding values are 25.82µm, 0.09(-), 13.43µm, and 0.91 (-).

Introduction

Hot rolling is a key thermo-mechanical process in metal forming, in which the thickness of a sheet material is reduced through successive deformation passes at elevated temperatures until the target gauge is attained. The sequence of these passes is known as the pass schedule. Designing a pass schedule requires appropriate consideration of both material and process parameters, as these influence the final product quality. Among them, the most critical quality-controlled parameters are the grain size and the thickness of the hot-rolled strip. The reduction in thickness and grain size of the hot-rolled strip influences mechanical properties, surface quality, and dimensional accuracy of the rolled product, which makes both quality parameters important.

Smart manufacturing uses the acquired data from sensors and performs accurate decision-making using digital technologies, resulting in an improvement in process efficiency, reliability, and product quality. The value of the manufacturing data lies in its ability to provide insights into process

behaviour, enabling the development of various data-driven models using advanced Information Technology (IT). Innovations such as the Industrial Internet of Things (IIoT), cloud computing, mobile internet, and Artificial Intelligence (AI) are driving automation and efficiency in modern manufacturing systems. Real-time data acquisition is achieved by deploying sensors throughout production environments. With AI-driven solutions, decision-making in smart factories increasingly shifts from humans to intelligent systems [1]. AI is a broader field that consists of machine learning and deep learning approaches. Machine learning involves algorithms that identify patterns within data to generate predictive or analytical outcomes. Deep learning is a subset of Machine Learning that uses neural networks with multiple layers, known as Deep Neural Networks (DNNs), to model complex relationships. Artificial Neural Networks (ANNs), composed of interconnected artificial neurons using nonlinear activation functions, form the foundation for DNNs. Among these, Convolutional Neural Networks (CNNs) preserve local spatial relationships and retain spatial information using convolutional layers [2].

On the other hand, Recurrent Neural Networks (RNNs) are neural networks that are applied to sequential data to identify the patterns present in the data. RNNs consider both previous and current inputs because of their recurrent connections that pass information from the current step to the next step [3]. A simple RNN can only retain information about short sequences because of its short-term memory. To retain information about long sequences, advanced RNNs are developed that include Long Short-Term Memory (LSTM), bi-directional LSTM, Gated Recurrent Unit (GRU), bidirectional GRU, Bayesian GRU, and others [4]. A unidirectional LSTM and GRU perform the sequence processing only in the forward direction, which captures the dependencies from earlier to later time steps, whereas a bidirectional LSTM and GRU perform the sequence processing both in forward and backward directions, which learns from past and future time steps simultaneously.

Several RNNs have been widely developed and used in the hot rolling process to perform predictions and anomaly detection, which are summarized below in Table 1.

Table 1. Summary of the RNN models used in the hot rolling process

Sources	Models	Purpose	Results
[5]	A Deep RNN model (2 LSTM models), Convolutional Neural Network (CNN), and the Deep Belief Network (DBN)	To predict the remaining useful life of a roller in a hot strip mill	RMSE:11.05, 18.97, and 16.23, respectively MAE: 9.85, 15.74, and 14.37, respectively
[6]	One GRU layer and two fully connected layers of 32 neurons	To predict the mean flow stress in industrial rolling	RMSE: 0.038 Flow stress value with a mean RMSE: 7.5 MPa
[7]	Inception-LSTM Neural Network	To predict the rolling force in the hot rolling	MAE, RMSE, and R^2 score: 3.969, 5.487, and 0.985, respectively
[8]	LSTM, Support Vector Machine (SVM), Random Forest, and ANN	Online prediction of the mechanical properties, such as Yield Strength (YS), Ultimate Tensile Strength (UTS), and Elongation (EL)	R^2 values: 0.96, 0.935, 0.940, and 0.942, respectively YS, UTS, and EL error values using LSTM: 19.68 MPa, 18.18 MPa, and 1.840%, respectively

The previously discussed state-of-the-art indicates a wide usage of RNNs in the hot rolling process, but predicting the chosen quality-controlled parameters, thickness, and grain size together using the proposed LSTM architecture has not been investigated. The motivation for using LSTM for modeling the hot-rolling process is its ability to learn from sequential data effectively compared to normal neural networks. For each pass, the model predicts the thickness and grain size by taking the rolling force and temperature as inputs. As thickness and grain size in the hot rolling process at any given pass are influenced by the deformation history from earlier passes, it is important to retain information across the sequence. This memory retention justifies the use of an LSTM approach. The trained LSTM model learns the functional relationship between the process parameters and the material parameters, which are treated as the inputs and outputs of the model, respectively. After accurate learning of the relationship, the model can predict both the quality parameters for any given values of rolling force and temperature. In the end, the values obtained by predictions using the trained and developed LSTM model are compared with thickness values entered as input to the RoCaT model and the grain size values calculated as output from the RoCaT model, and statistical validation is performed to ensure the reliability and robustness of the developed LSTM model.

Methodology

Data acquisition is one of the main steps before training and developing Machine Learning and Deep Learning models. In this research, data acquisition is performed using a Fast Rolling Model and a previously developed Simulation as a Service. FRMs are developed based on mechanical assumptions called the slab method, coupled with semi-empirical equations. One such FRM is the Rolling Calculation Tool (RoCaT). [9, 10].

RoCaT is used to calculate strain, strain rate, rolling forces, and austenite grain size for a complete pass schedule as shown in Figure 1. It consists of physics and semi-empirical-based process, temperature, and microstructure evolution of the material. The inputs to the RoCaT are pass schedule, rolling mill data like the rolling speed, material parameters like the flow stress, density, and heat transfer coefficient, and material calibrated parameters for semi-empirical models. A major assumption that is taken into account while developing RoCaT is that it considers static recrystallization.

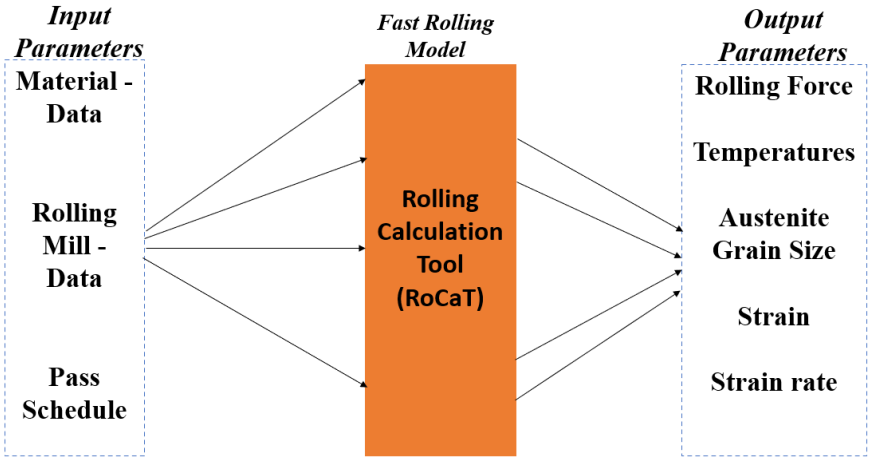

Fig. 1. Schematic structure of the Fast Rolling Model RoCaT [9]

RoCaT calculates the rolling force based on the equation developed by Sims and Wright [11], which depends on the contact length (l_d), the width of the rolling stock (b), the mean flow stress (σ_m), and the geometric factor (Q_p) that compensates for inaccuracies due to friction and shear, as discussed by Seuren et al. [12] and Lohmar et al. [9], shown in equation 6:

$$F = l_d \cdot b \cdot \sigma_m \cdot Q_p \quad (1)$$

RoCaT takes into consideration the static recrystallization, and the grain growth after static recrystallization is calculated using semi-empirical equations introduced by Beynon and Sellars [13]. The grain size after full static recrystallization (d_{srx}) calculation includes the accumulated equivalent strain (φ_{acc}), the average temperature (T_m), and the austenite grain size ($d_{\gamma,0}$) as shown in equation 7:

$$d_{srx} = d_{\gamma,0}^{0.4} \cdot \varphi_{acc}^{-0.5} \cdot \exp(-4500/8.31 \cdot T_m) \quad (2)$$

Data Acquisition & Preprocessing. In this work, RoCaT is used to calculate different pass schedules for hot rolling as training data. The pass schedule consists of a thickness reduction in each pass, inter-pass time, and the rolling velocity. While designing the pass schedule, the maximum allowed force and torque, energy consumption, and allowed tolerances of the ultimate tensile strength are considered [14]. The concept of Simulation as a Service is to remotely request simulations and help in achieving data for desired pass schedule. In this case, a SaaS [15] was developed for RoCaT, where a request is sent to RoCaT from the client-side, and RoCaT on the server-side performs the simulations for every pass, stores data in JSON format, and sends the result JSON back to the client. This kind of setup allows a process expert and a machine learning expert to collaborate efficiently.

The material and process parameters that are considered in the hot rolling process are rolling thickness (H), grain size (D_{GS}), rolling force (F), and rolling temperature. In rolling temperature, three rolling temperatures are considered: The temperature before the material comes in contact with the rollers (T_{Before}), the temperature just after the material leaves the roller (T_{After}), and the temperature of the material at the end of the transport table (T_{End}), respectively. The dataset consists of 3700 pass schedules of varying numbers of passes with starting thicknesses of 75, 85, 95, 105, 205, 305, and 405 mm. The target grain size that is chosen is below 30 micrometers. Each pass schedule is given a series_id number for identification, so there are 3700 series_ids. As the dataset consists of a varying number of passes, the concept of padding and masking is used. In the preprocessing step, the pass schedule with the maximum sequence length is taken, and the rest of the pass schedules are padded to that maximum length with zeros, so that all the sequences are of the same length, which makes it easier for batch training in LSTM. A masking layer is used in the model to ensure that the LSTM ignores padded zeros during training. The dataset consists of material and process parameters for the steel grade S355, which were previously determined [14] . While training the LSTM model, the data is split in the ratio of 70:30, where 70% of the data is used for training, and 30% is used for testing. Normalization is performed using MinMaxScaler, which is performed after data splitting. In order to prevent data leakage, MinMaxScaler is fit on training data, and using the same scaling parameters, the test data is also scaled before training is performed. In this study, a 3-fold cross-validation is used, in which the training dataset is split into three folds. One fold is used once as a validation set, while the remaining two folds are used for training. The final model is trained on the fold with the best validation performance.

Long Short-Term Memory (LSTM) Architecture

LSTM consists of three types of gates, the Input, Output, and forget gates. The purpose of the forget gate is to decide which information is required to keep or discard from the previous steps. The purpose of the input gate is to control the amount of new information in the current input, and the output gate is to control the output based on the current input. LSTM consists of a memory unit that helps to retain long sequences of memory. GRU consists of two gates, a reset gate and an update gate. The purpose of the Reset gate is to control the amount of past information that is going to combine with the current input. The purpose of the Update gate is to control the amount of new information that will replace the past information [16, 17]. The mathematical equations between the input and output of the LSTM [18] are defined by the following equations:

$$i_t = \sigma(W_{xi} \cdot x_t + W_{hi} \cdot h_{t-1} + W_{ci} \cdot c_{t-1} + b_i) \quad (3)$$

$$f_t = \sigma(W_{xf} \cdot x_t + W_{hf} \cdot h_{t-1} + W_{cf} \cdot c_{t-1} + b_f) \quad (4)$$

$$c_t = f_t \cdot c_{t-1} + i_t \tanh (W_{xc} \cdot x_t + W_{hc} \cdot h_{t-1} + b_c) \quad (5)$$

$$o_t = \sigma (W_{xo} \cdot x_t + W_{ho} \cdot h_{t-1} + W_{co} \cdot c_t + b_o) \quad (6)$$

$$h_t = o_t \cdot \tanh(c_t) \quad (7)$$

Equation 3 describes the update of the input gate, which takes into account the output (h_{t-1}) of the last time step, the input of the current time step (x_t), the memory cell value of the last time step (c_{t-1}), and a bias term (b_t). Equations 4 to 6 represent the output of the forget gate, input gate, and output gate, respectively. Equation 7 represents the output of the current time step. W_{xi}, W_{hi}, W_{ci}, W_{xf}, W_{hf}, W_{cf}, W_{xc}, W_{hc}, W_{xo}, W_{ho}, W_{co} are different weight matrices, and b_i, b_f, b_c, b_o are the bias terms used in equations 3 to 7.

In this section, an LSTM architecture is proposed for predicting the thickness and grain size simultaneously in the hot rolling process, as shown in Figure 2. The architecture of the LSTM model consists of one LSTM layer and uses four inputs and two outputs for the entire dataset. Input parameters of the LSTM model are considered as those that sensors can capture in real time. Output parameters influence the quality of hot-rolled sheets. The aim of developing the LSTM neural network is to capture the nonlinear functional relationship between the input parameters, which are rolling force and rolling temperatures, and the output parameters, which are thickness and grain size. By accurately learning the functional relationship, the developed LSTM model can predict both the output parameters closer to the RoCaT value. The learning of any Neural Network and model complexity is dependent on hyperparameters. Hyperparameters are those parameters that are defined before training machine learning models. Hyperparameters regulate model complexity with the number of layers and nodes, and control the neural network learning using batch sizes, learning rate, initialization conditions, and momentum decay parameters [19]. Hyperparameters that are used in this research for the training of the proposed LSTM architecture are units, dropout, and batch size. Optimization of the hyperparameters is performed using optuna [20]. 50 to 128 units, 0.1 to 0.4 dropout rates, and batch sizes of 8, 16, and 32 are used as the search space for optuna. 20 trials are conducted, and the optimal hyperparameter is obtained with 60 units, a dropout rate of 0.16, and a batch size of 16. Dropout is used to prevent overfitting of the neural networks. Overfitting occurs when a model performs well on the training dataset but fails to generalize to the test dataset [21]. The value of the units (nodes) and dropout used in the proposed LSTM are 60 and 0.16, respectively. The total number of epochs the LSTM model was trained on is 25, and the loss function used is Mean Squared Error (MSE).

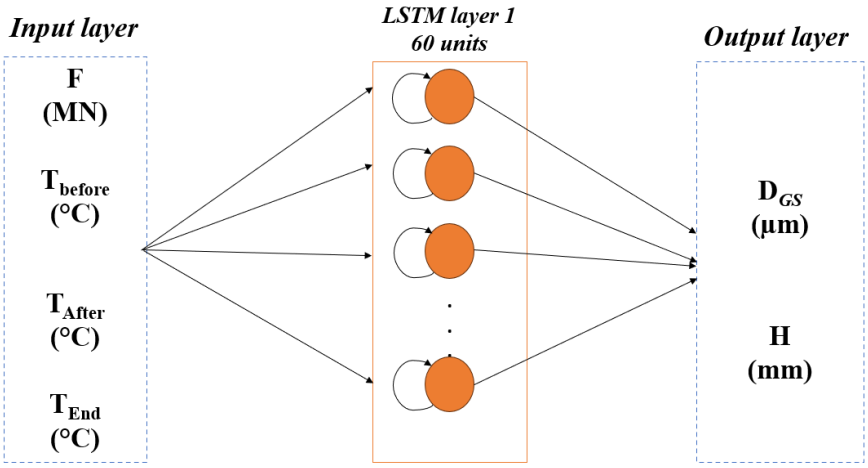

Fig. 2. Developed Long-Short Term Memory Architecture for predicting quality-controlled parameters of hot-rolled strip

Table 2 summarizes the architecture of the developed LSTM model. The proposed LSTM model uses a maximum sequence length of 38 passes present in the dataset. The input feature dimension is four, and the Dense output layer maps the LSTM outputs to two prediction targets. The batch size is "None," which indicates that a flexible batch dimension is used during the construction of the model. The model contains a total of 15,722 trainable parameters. Trainable parameters are the internal weights and biases of the LSTM and Dense layers that the model learns, and using backpropagation, it gets updated during training. These parameters influence the learning capacity of the model and indicate the ability of the developed LSTM to capture patterns in the data.

Table 2. Summary of the developed Long-Short Term Memory model

Layer (type)	Output Shape	Param #	Connected to
input_layer_3 (InputLayer)	(None, 38, 4)	0	-
not_equal_3 (NotEqual)	(None, 38, 4)	0	input_layer_3[0]…
masking_3 (Masking)	(None, 38, 4)	0	input_layer_3[0]…
any_3 (Any)	(None, 38)	0	not_equal_3[0][0]
lstm_3 (LSTM)	(None, 38, 60)	15,600	masking_3[0][0], any_3[0][0]
dropout_3 (Dropout)	(None, 38, 60)	0	lstm_3[0][0]
dense_3 (Dense)	(None, 38, 2)	122	dropout_3[0][0]

Results & Discussion

Loss curves are useful to analyze the training behavior of machine learning models as they can track changes in loss over each epoch and provide information regarding the performance and learning progress of the developed models. Figure 3 shows the pattern of training, validation, and test loss curves across epochs during the training of the LSTM model. The training and validation loss curve is obtained on a 70% training dataset that is split using a 3-fold cross-validation, and the best fold is shown, the test loss curve is obtained on a 30% test dataset. At the beginning of the training, MSE loss is high as the weights of the LSTM are randomly chosen, resulting in less accurate predictions. At the time of the training, adam optimizer adjusts LSTM weights to minimize the MSE loss. With the progress of the training, predictions by LSTM are closer to the RoCaT values, and the loss decreases. It can be concluded that as the training progresses, the model is learning the patterns present in the data better and improving, hence the loss curves show a downward trend. In training loss, a sharp drop is observed at the initial epochs because the model is able to quickly learn the pattern present in the data, then the curve decreases gradually to zero, and in the end stabilizes, which is an indicator of successful convergence of the model. Similarly, the validation and the test loss decrease gradually towards zero and eventually stabilize. The successful training of the LSTM model is indicated by the stabilizing nature of all three loss curves, and overfitting does not occur.

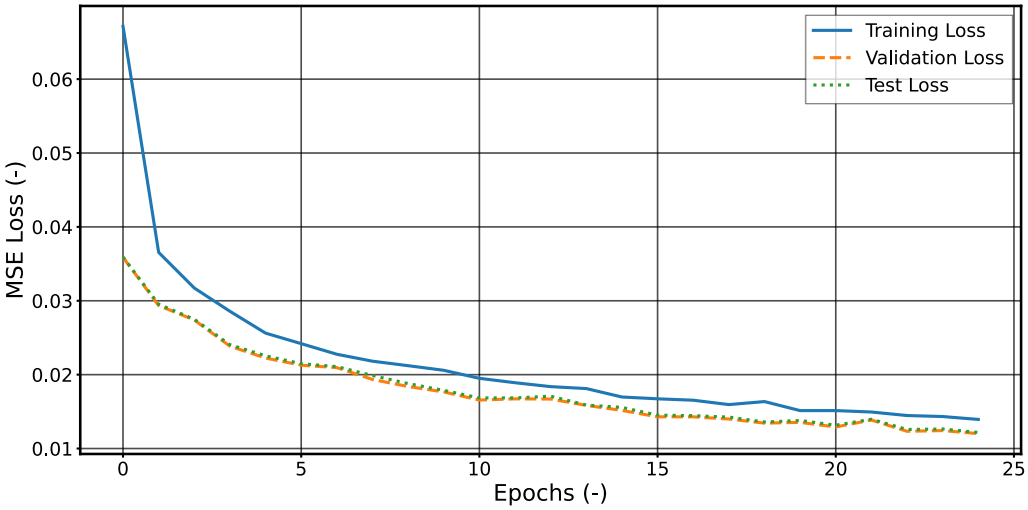

Fig. 3. Loss curves of the developed LSTM

Using MSE as the loss function, the LSTM model is trained on the 70% dataset containing 3700 pass schedules. After training, the developed LSTM model is used to predict thickness and grain size evolution along 500 different pass schedules with a starting thickness of 205 mm using the same rolling force and temperature as in the original dataset. Figure 4a and 4b show two examples for the starting thickness of 205 mm with 15 and 17 passes to compare the thickness and grain size values obtained from RoCaT and the predictions performed by the LSTM across the pass schedules. In both figures, it is observed that the developed LSTM model can capture the trend of both the quality parameters for the entire pass schedule. However, in Figure 4a, the maximum percentage of relative error for thickness between the thickness values entered as input to the RoCaT model and the predicted value by LSTM is obtained as 18.81%, and in Figure 4b, it is obtained as 16.56%. Table 3 shows the relative percentage error of the thickness between the predicted value by LSTM and the thickness values entered as input to the RoCaT model in each pass. The significance of the maximum relative percentage error in both pass schedules is that, as both errors are observed during predictions performed at the early passes, the LSTM finds difficulty during early passes in performing accurate predictions. It is also observed that during predictions of both thickness and grain size, the LSTM can accurately describe the thickness evolution compared to the grain size evolution. For both figures 4a and 4b, the grain size evolution shows significant deviations between the grain size value calculated as output from the RoCaT model and the predicted value by LSTM, also higher during the initial passes, as compared to later passes, showing that the LSTM also finds difficulty in retaining information during initial passes. The time required to perform the calculations of rolling force, temperatures, and austenite grain size using the RoCaT model for the corresponding thickness shown in Table 3 is 1.5 and 1.6 seconds, respectively, for both pass schedules. In comparison, the time required by the trained LSTM model is 0.14 and 0.56 seconds, respectively, to perform the predictions of both quality parameters. Time comparison is performed on a processor specification with 16gb RAM, 13th Gen Intel ® Core ™ i5- 1335U processor, and 128 MB Intel® UHD Graphics. Nevertheless, the predictions performed by LSTM for a varying number of passes overall justify the predictive capability of the developed model.

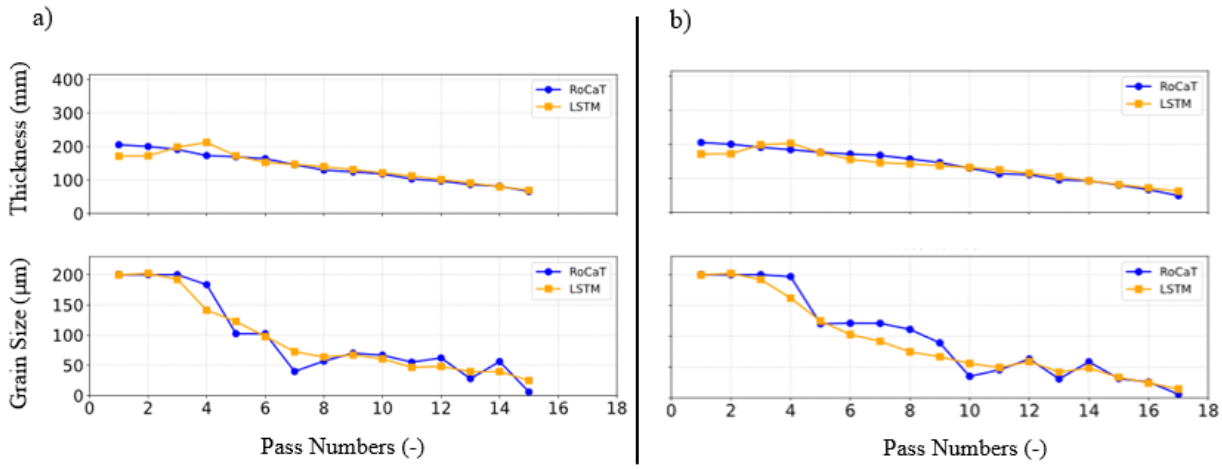

Fig. 4. Predictions performed by the developed LSTM with: a) Pass schedule with 15 passes, b) Pass schedule with 17 passes

Table 3. Thickness predictions and percentage error for a) pass schedule of 15 passes, b) pass schedule of 17 passes

RoCaT (Thickness in mm) (a)	Predicted with LSTM (Thickness in mm) (a)	Relative Percent Error of LSTM (-) (a)	RoCaT (Thickness in mm) (b)	Predicted with LSTM (Thickness in mm) (b)	Relative Percent Error of LSTM (-) (b)
205	171.04	16.56	205	171.04	16.56
200	171.98	14.00	200	171.98	14.00
190.8	198.06	3.80	190.9	198.18	3.81
172.7	205.2	18.81	183.9	202.78	10.26
169	171.71	1.60	176.2	175.33	0.49
163.4	152.40	6.72	170.8	155.49	8.96
145.7	146.39	0.47	167.5	146.04	12.80
129.4	139.44	7.76	156.7	142.13	9.29
124.2	130.84	5.35	146.1	137.14	6.13
117.7	121.14	2.92	129.8	131.78	1.53
102.9	111.94	8.78	113.1	124.62	10.18
96.8	101.19	4.53	110.6	114.77	3.77
86.2	90.60	5.11	95.3	104.68	9.84
81	79.73	1.55	92.7	92.83	0.14
65.6	69.09	5.33	80.1	81.67	1.96
			66.5	71.71	7.83
			55.1	62.14	12.79

The statistical validation is performed on denormalized data of the unseen test dataset to justify the reliability and robustness of the machine learning models. Performance metrics techniques such as Root Mean Squared Error, Normalized Root Mean Squared Error, Mean Absolute Error, and R-squared score are used to assess prediction accuracy. Figure 5 shows the values of performance metrics obtained on the 30% test datasets, which are split during data preprocessing for both the output parameters, thickness and grain size. Here, RMSE calculates the square root of the average of the squared differences between the actual and predicted values as shown in Equation (8) and Equation (9), with n describing the number of observations, Y_k the actual values, and $\acute{Y}_k$ the predicted

values. Normalized Root Mean Squared Error (nRMSE) is a dimensionless error metric that calculates the prediction error with respect to the range of actual data, as shown in Equation (10), where Y_{max} and Y_{min} are the maximum and minimum values of the actual data. MAE calculates the average sum of the absolute differences between the actual values and the predicted values shown in Equation (11). The R^2 is a dimensionless metric that calculates the variance in the dependent variable, which is predicted by the independent variables, as shown in Equation (12), with SSR describing the sum of squares of residuals and TSS the total sum of squares. The value ranges from 0 to 1, where 0 indicates the model is a poor fit and 1 indicates an ideal fit of the model to the data [22]

$$\mathrm{MSE} = 1/n \sum_{k=1}^{n} \left(Y_{\mathrm{k}} - \acute{Y}_{\mathrm{k}}\right)^2 \tag{8}$$

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}} \tag{9}$$

$$\mathrm{nRMSE} = \left(\frac{\mathrm{RMSE}}{Y_{\mathrm{max}} - Y_{\mathrm{min}}}\right) \tag{10}$$

$$\mathrm{MAE} = 1/n \sum_{k=1}^{n} \left(\left| Y_{\mathrm{k}} - \acute{Y}_{\mathrm{k}} \right| \right) \tag{11}$$

$$R^2 = 1 - \frac{SSR}{TSS} \tag{12}$$

The results of error metrics and the coefficient of determination on the denormalized data indicate that the model's performance is acceptable in predicting both thickness and grain size. For thickness, the RMSE, nRMSE, and MAE are obtained as 45.21mm, 0.11(-), and 18.75mm, respectively. The predictions performed by the LSTM model closely match the values of RoCaT, while the R^2 value obtained is 0.81(-), which indicates a strong variance despite some deviations. Similarly, for grain size, the RMSE, nRMSE, and MAE are 25.82µm, 0.09(-), and 13.43µm, respectively, and the higher R^2 value of 0.91(-) compared to the thickness indicates strong predictive capability of the developed LSTM model. In this study, RMSE, MAE, and nRMSE for thickness are more compared to grain size, both in absolute and relative measures. Grain size has larger absolute errors, but a low value of nRMSE indicates that the model predicts it more accurately relative to its typical range. The variance of the target variable has a strong influence on the coefficient of determination (R^2). The variance for the thickness is smaller, so even moderate prediction errors result in a larger drop in R^2, whereas the variance for the grain size is higher, so it is less affected by similar errors. Overall, these metrics show that the developed LSTM model verifies the reliability. However, the predictions for both the thickness and grain size could be further improved by using an attention layer, which is one of the core components of transformer architectures.

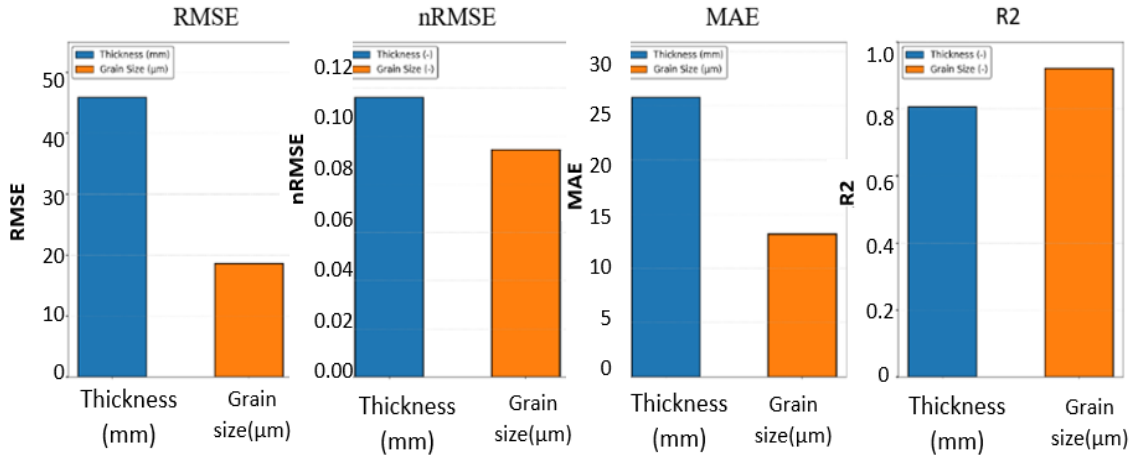

Fig. 5. RMSE, nRMSE, MAE, and R^2 score of the developed LSTM

Summary and Outlook

The effect of rolling force and rolling temperatures on the quality-controlled parameters, thickness, and grain size in the hot rolling process is investigated in the present work. Considering the aim to predict both the quality parameters accurately, a Long Short-Term Memory is developed. Real-time application of the developed LSTM architecture is to perform online process monitoring during the hot rolling process by taking rolling force, rolling temperature as input, and thickness as output, and all three parameters can be captured by sensors in real time. Conclusions that can be drawn are that the prediction performed by Long Short-Term Memory can capture the trend of both the quality parameter during the entire pass schedule. However, deviations between the RoCaT values and the predicted value by the developed LSTM are observed. Another conclusion is drawn based on statistical validation. The validation of the developed LSTM model is performed using Root Mean Squared Error, Normalized Mean Squared Error, Mean Absolute Error, and R-Squared value. It is observed that the value of Root Mean Squared Error and Mean Absolute Error for thickness is high compared to the grain size, whereas the R-Squared value for grain size is better compared to the thickness. The RMSE, MAE, nRMSE, and R^2 values suggest that the obtained results can be further improved. Both conclusions indicate that further investigation is required to make the predictions of both quality parameters more accurate. Further research will investigate the usage of other RNN-based models, such as the Gated Recurrent Unit (GRU) and the combination of LSTM and GRU with an attention layer, which is one of the core components of the Transformer-based architecture. This combination of RNNs with an Attention layer will be performed to further increase the accuracy of the RNN models, as well as reduce the relative error percentage for both the quality parameters.

Acknowledgement

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

References

[1] F. Tao, Q. Qi, A. Liu, and A. Kusiak, "Data-driven smart manufacturing," *Journal of Manufacturing Systems*, vol. 48, pp. 157–169, 2018, doi: 10.1016/j.jmsy.2018.01.006.

[2] S. J. Plathottam, A. Rzonca, R. Lakhnori, and C. O. Iloeje, "A review of artificial intelligence applications in manufacturing operations," *J Adv Manuf & Process*, vol. 5, no. 3, 2023, doi: 10.1002/amp2.10159.

[3] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," *Physica D: Nonlinear Phenomena*, vol. 404, p. 132306, 2020, doi: 10.1016/j.physd.2019.132306.

[4] T. Perumal, N. Mustapha, R. Mohamed, and F. M. Shiri, "A Comprehensive Overview and Comparative Analysis on Deep Learning Models," *JAI*, vol. 6, no. 1, pp. 301–360, 2024, doi: 10.32604/jai.2024.054314.

[5] R. Jiao, K. Peng, and J. Dong, "Remaining Useful Life Prediction for a Roller in a Hot Strip Mill Based on Deep Recurrent Neural Networks," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 7, pp. 1345–1354, 2021, doi: 10.1109/JAS.2021.1004051.

[6] A. G. Zinyagin, A. V. Muntin, V. S. Tynchenko, P. I. Zhikharev, N. R. Borisenko, and I. Malashin, "Recurrent Neural Network (RNN)-Based Approach to Predict Mean Flow Stress in Industrial Rolling," *Metals*, vol. 14, no. 12, p. 1329, 2024, doi: 10.3390/met14121329.

[7] G. Niu, M. Zhang, Y. Yang, and Z. Huang, “A Hot Rolling Full Process Rolling Force Prediction Method Based on Transfer Learning and Inception-LSTM Neural Network,” *ISIJ Int.*, vol. 65, no. 1, pp. 97–103, 2025, doi: 10.2355/isijinternational.ISIJINT-2023-446.

[8] Z. Yang *et al.,* “Online Prediction of Mechanical Properties of the Hot Rolled Steel Plate Using Time-series Deep Neural Network,” *ISIJ Int.*, vol. 63, no. 4, pp. 746–757, 2023, doi: 10.2355/isijinternational.ISIJINT-2022-383.

[9] J. Lohmar, S. Seuren, M. Bambach, G. Hirt, “Design and application of an advanced fast rolling model with through thickness resolution for heavy plate rolling,” 2014.

[10] S. Seuren, J. Willkomm, M. Buecker, M. Bambach, G. Hirt, “Sensitivity analysis of a force and microstructure model for plate rolling,” 2012.

[11] Sims, R. B., & Wright, H., “Roll force and torque in hot rolling mills,” 3(5), 261–269., 1963.

[12] Seuren, S., Bambach, M., Hirt, G., Heeg, R., & Philipp, M., “Geometric factors for fast calculation of roll force in plate rolling,” 2010.

[13] Beynon, J. H., & Sellars, C. M., “Modelling microstructure and its effects during multipass hot rolling,” 32(3), pp. 359–367, 1992.

[14] C. Idzik, A. Krämer, G. Hirt, and J. Lohmar, “Coupling of an analytical rolling model and reinforcement learning to design pass schedules: towards properties controlled hot rolling,” *J Intell Manuf*, vol. 35, no. 4, pp. 1469–1490, 2024, doi: 10.1007/s10845-023-02115-2.

[15] C. Scheiderer *et al.,* “Simulation-as-a-Service for Reinforcement Learning Applications by Example of Heavy Plate Rolling Processes,” *Procedia Manufacturing*, vol. 51, pp. 897–903, 2020, doi: 10.1016/j.promfg.2020.10.126.

[16] S. Nosouhian, F. Nosouhian, and A. Kazemi Khoshouei, *A Review of Recurrent Neural Network Architecture for Sequence Learning: Comparison between LSTM and GRU*, 2021.

[17] P. Tang, H. Wang, and S. Kwong, “Deep sequential fusion LSTM network for image description,” *Neurocomputing*, vol. 312, pp. 154–164, 2018, doi: 10.1016/j.neucom. 2018.05.086.

[18] Y. Gao and Glowacka Dorota, “Deep Gate Recurrent Neural Network,” pp. 350–365, 2016.

[19] Z. S. Kadhim, H. S. Abdullah, and K. I. Ghathwan, “Artificial Neural Network Hyperparameters Optimization: A Survey,” *Int. J. Onl. Eng.*, vol. 18, no. 15, pp. 59–87, 2022, doi: 10.3991/ijoe.v18i15.34399.

[20] J. Almomani, G. Nasserddine, O.A.Khatib, R.R. Kala, M.Nasserddine, "Time Series Forecasting Approach for Power Demand Prediction Based on Feature Engineering, Optuna-Based Hyperparameter Tuning, and Ensemble Learning*",2025 IEEE PES Conference on Innovative Smart Grid Technologies - Middle East (ISGT Middle East)*: IEEE, 2025, doi: 10.1109/ISGTMiddleEast65737.2025.11314432.

[21] Haider Khalaf Jabbar, “Methods to avoid over-fitting and under-fittingin supervised machine learning(comparative study),” pp. 163–172.

[22] Abhishek V Tatachar, “Comparative Assessment of Regression Models Based On Model Evaluation Metrics,” 2021.

The publication of this article was funded by the RWTH Aachen University 10.13039/501100007210