International Journal of Material Forming (2026) 19:7
https://doi.org/10.1007/s12289-025-01972-x

RESEARCH

Coupled process chain modelling to minimise the earing formation in industrially rolled aluminium strips for beverage can production using data-based methods

Nilesh Thakare[1] · Kai Karhausen[2] · Hans-Reimund Müller[3] · Emad Scharifi[1] · David Bailly[1]

Received: 16 July 2025 / Accepted: 6 December 2025

Abstract
The production of aluminium strip involves a long sequence of thermomechanical processing steps that significantly influence the material's mechanical properties and may induce anisotropy. This anisotropy can manifest as earing during deep drawing operations - such as those used in beverage can manufacturing - resulting in increased trimming scrap, process downtimes, and reduced economic viability. To assess formability and quantify earing, the cup drawing test is employed as a standard evaluation method. Understanding and minimizing earing formation requires comprehensive modelling of the entire process chain, which is traditionally performed manually by domain experts - a time-consuming, error-prone, and costly effort. This study presents a novel, scalable, and flexible approach to model a process chain by integrating production data with process models on the Microsoft Azure Databricks platform. The proposed method is validated on an industrial aluminium strip production line, demonstrating its capability to automate data processing, extract actionable insights, and support process optimisation. The approach successfully identifies an optimum processing route that minimises the earing integral, as determined by a dedicated evaluation function.

Keywords Industry 4.0 · Digital twin · Machine learning · Hot rolling · Deep drawing · Earing formation · Forming technology

Introduction

Aluminium is used in a wide range of applications, ranging from transportation to packaging industries, due to its properties like lightweight, high corrosion and fatigue resistance, high electrical and thermal conductivity, as well as recyclability [1–5]. According to the International Aluminium Institute (IAI), around 72,758 thousand metric tonnes of primary aluminium was produced in 2024, which is around 33% more compared to 2014, thereby showing the increasing demand and usage of aluminium around the world. In parallel, there is a growing emphasis on increasing the use of secondary aluminium through scrap recycling, driven by both economic and environmental considerations [6, 7]. However, aluminium recycling is also leading to the accumulation of impurities, like Si, Mg, Mn, Cu, and Fe, thereby undesirably influencing the mechanical properties of the aluminium products [8–12]. The typical aluminium strip process chain for beverage can production consists of several processing steps, namely melting and casting, preheating, hot rolling, and cold rolling. Parameters of each processing step, like hot rolling temperature, cold reduction, recrystallisation, work hardening and softening, microstructure, texture, etc., influence the properties of the aluminium strip which may ultimately cause anisotropy in the strip [13–15]. When deep drawing the beverage can from aluminium strips having higher anisotropy, significant earing formation can occur on the formed can body, resulting in uneven edges [14, 16, 17]. Since these uneven edges must be trimmed away, they generate scrap at the site of a can

✉ Nilesh Thakare
nilesh.thakare@ibf.rwth-aachen.de

1 Institute of Metal Forming, RWTH Aachen University, Intzestrasse 10, 52072 Aachen, Germany

2 Speira GmbH, Research & Development, Georg-von-Boeselager-Straße 21, 53117 Bonn, Germany

3 Aluminium Norf GmbH, Koblenzer Str. 120, 41468 Neuss, Germany

Published online: 13 January 2026

manufacturer. Furthermore, earing formation, if not considered during the deep drawing process layout, can lead to the stoppage of the ultra-fast beverage can production facilities, resulting in increased downtime and reduced productivity [18]. Hence, aluminium strip producers are obliged to deliver high-quality material fulfilling the customer requirements in terms of strip geometry, isotropy, and strength. To test the formability of the aluminium mill products, strip producers conduct cup drawing tests and measure the cup heights to obtain the earing profile and earing characteristics, which serve as a quality indicator of the produced coil.

With the advances in digitalization and simulation methods, it is now possible to acquire large amounts of production data and use numerical as well as fast analytical process models that can simulate individual processing steps to calculate non-measurable or hard to measure product properties like temperature, strain, grain size distributions within the workpiece, recrystallized fraction, strip strength [19–23]. Thus, production and simulation data together help process experts to gain a better understanding to optimise their processes. However, looking at each processing step separately does not capture the different interdependencies between each processing step and the final strip properties, thus hindering a robust process optimisation. Moreover, optimising a process chain is a repetitive, expensive, and error-prone task involving data acquisition, pre-processing, setting up process models, running simulations, extracting results, and finally analysing all data. This is usually conducted manually by process experts. Hence, modelling the complete process chain by coupling production data and simulation models in an automated sequence is imperative. Since data is stored in different formats, units, and structures, it must be pre-processed before it can be used as an input to the next process model. Nevertheless, the lack of a suitable tool that can allow efficient handling and management of all process models and the data that is generated or captured along the entire process chain poses a technological challenge.

To overcome these limitations, a novel approach to modelling a coupled process chain is demonstrated in this work. For coupling the production data and simulation models along the process chain, the Microsoft Azure Databricks platform is used which is flexible to implement custom scripts to check and correct erroneous data, convert units, initiate process models, run process simulations, extract simulation results and store them together to provide the aggregate data containing production as well as simulation data to the user and finally analyse it to obtain better process insights and determine optimal processing route. Moreover, the proposed approach is also suitable for processing new production data from daily production, where the newly ingested data can be processed batchwise and included in the analysis to derive process insights. The benefits of implementing the proposed approach are demonstrated by the automated processing of industrial production and simulation data to generate aggregate data that is visualised and analysed to obtain process insights. Finally, an optimum processing route fulfilling an artificially generated pseudo-requirements, including a processing window leading to minimum earing formation, is identified by using an evaluation function.

Within the Microsoft Azure platform of Speira GmbH the measured process- and quality data of Alunorf GmbH are continuously updated and linked to the Databricks platform, developed in this paper. In addition, two simulation models are coupled, which belong to the proprietary "rolling simulation environment (RoSE)" of Speira GmbH [24]. The first model provides the numerical re-simulation of the hot rolling schedule with an integrated prediction of the microstructure development, and thus it provides non-measured material property data. The second model takes the measured earing profile of the final cold rolled strip as an input and back-calculates the earing profile of the strip after hot rolling, while taking the cold rolling deformation into account. The model results and the process data are then aggregated and made available in the Azure platform by newly developed data management and filtering algorithms. This includes the concepts for linking data from several sources which belong to one coil and connectors for the communication between the data platform and the models. Filters are implemented to analyse and detect the peak earing locations in the measured cup profiles. And finally, a dashboard allowing for user interactions is constructed to facilitate advanced data analytics procedures and automated data visualizations.

Approach

The approach consists of a data management platform that ingests the measured production data, applies necessary filters, and generates input decks for the simulation algorithms. The platform triggers the simulations and takes in the computed results. Production and simulation data are aggregated and are made accessible to suitable advanced analysis and visualisation tools using interactive dashboards. All data operations work in an automated manner on the data of a long production period, i.e., several hundred coils over the span of several months.

Analysis of an industry process chain

The principal steps for the can body strip production are shown in Fig. 1. A production schedule for alloy AA3104 is selected for this work. It consists of pre-heating of the cast

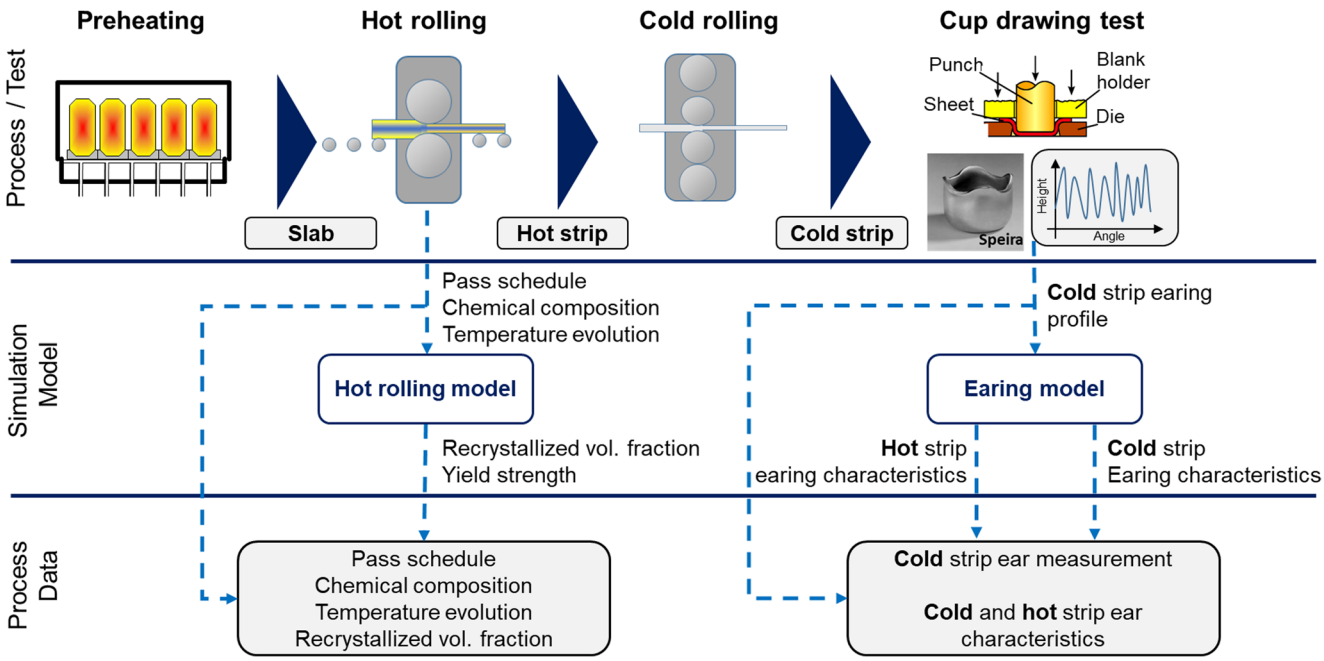

Fig. 1 Aluminium strip production process chain showing the flow and processing of material, flow of data and available process models

aluminium ingot, followed by hot and cold rolling. Finally, a cup drawing test is conducted on a Zwick/Roell P600 equipment at Speira R&D to measure the earing formation in the aluminium strip.

To test the formability of the aluminium mill product, strip producers conduct cup drawing tests and measure the cup heights along the cup circumference from 0° to 360° angle to obtain the earing profile, as shown in Fig. 2.

The earing profile is then used to calculate characteristic earing values, namely mean ear height (Z) and delta ear (ΔZ) according to the Eqs. 1–1 and 1–2. These scalar characteristic earing values serve as strip's quality indicators that can be used for optimizing the process chain to minimize the earing formation [25, 26].

$$mean\ ear\ height,\ Z\ [\%] = \frac{\overline{h}_{peak} - \overline{h}_{valley}}{\overline{h}_{valley}} \times 100 \qquad (1\text{-}1)$$

$$delta\ ear,\ \Delta Z\ [\%] = \frac{2 * \overline{h}_{45°} - (\overline{h}_{0°} + \overline{h}_{90°})}{\overline{h}_{0°} + \overline{h}_{90°}} \times 100 \qquad (1\text{-}2)$$

where,

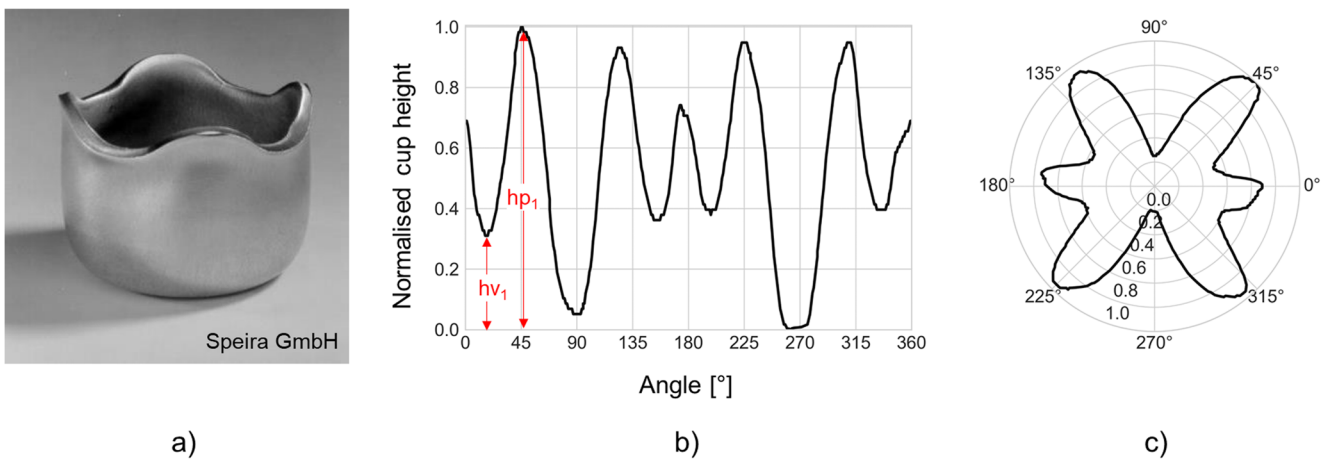

Fig. 2 (**a**) Deep drawn cup with significant earing formation, (**b**) Cup height measured along its circumference showing 6 ears at around 0°, 45°,135°, 180°, 225°, and 315° with 6 peaks and 6 valleys, (**c**) Butterfly-like structure when the measured earing profile is plotted in the polar coordinate system

$$\overline{h}_{peak} = Average\ peak\ height\ [mm] = \frac{h_{p1} + h_{p2} + h_{p3} + \ldots}{Number\ of\ peaks}$$

$$\overline{h}_{valley} = Average\ valley\ height\ [mm] = \frac{h_{v1} + h_{v2} + h_{v3} + \ldots}{Number\ of\ valleys}$$

$$\overline{h}_{0^\circ,45^\circ,90^\circ} = Average\ ear\ height\ at\ an\ angle\ w.r.t.\ rolling\ direction$$

During continuous production only the earing profile of the finished cold rolled material is measured since this is the decisive quality for a customer. However, the most important lever to influence and optimize this profile is during hot rolling. At this stage a continuous measurement is not possible and only limited validation data are available from internal R&D investigations. Hence, an earing model was developed by Speira GmbH which uses the Fourier method to inversely calculate hot strip earing profiles from the measured cold strip earing profile [26]. Using the simulated hot- or measured cold rolled earing profiles, corresponding earing characteristics are calculated according to Eqs. 1–1 and 1–2 that are part of the earing model. In order to exclude effects from the cold rolling process steps the calculated hot rolled cup can be directly correlated to the hot rolling conditions.

For hot rolling, a single-stand roughing mill followed by a 4-stand tandem line setup is used with intermediate cooling to control the hot strip temperature, while cold rolling is done using a single-stand rolling mill. A dislocation density-based fast analytical hot rolling model [27] is used to simulate the hot rolling processes, which takes as input the hot rolling production data (pass schedule), material parameters and chemical composition to calculate amongst other values the fraction recrystallized, average grain size, temperature and yield strength of the workpiece, rolling force and torque as shown in Fig. 3.

The fast analytical rolling model is developed in a modular way and consists of a temperature module that calculates temperature evolution using the finite difference method. The deformation module calculates the stress and strain state for each strip element along the roll gap. The core of this system is a dislocation-based material model, which calculates the evolution of distinct dislocation populations. From these, the flow stress can be derived, which is incrementally used by the deformation model. Furthermore, the dislocation densities directly after a rolling pass are the driving forces for recrystallization, which is then calculated using a modified JMAK model. Finally, the microstructure module calculates grain size and yield strength based on the dislocation densities.

In this work, the cold rolling process is not modeled to keep the process chain simple by assuming it to be static, where only strip strength increases due to strain hardening while thermally activated mechanisms like recovery and recrystallization are not present since the temperature levels are too low. The results of model validations are presented in Figure A-1 and Figure A-2 available in Appendix A.

Microsoft Azure databricks platform for storage and processing of the data

Microsoft Azure Databricks is a cloud-based big data and machine learning platform that provides all necessary features and capabilities for engineers to work and collaborate on data science projects [28]. It combines the advantages of Apache Spark, which can process a huge amount of data, provides data streaming capability and an interactive query engine, Microsoft Azure, which is a cloud computing platform and Databricks, which can ingest large amounts of data, clean up the data and apply machine learning algorithms [28]. Being a cloud computing-based platform, it provides its users with the necessary flexibility to work with different users at multiple locations without the need for powerful hardware on a local device. Moreover, the platform provides a graphical user interface that makes it easy to operate for users with all levels of programming skills. The platform has two main components - Workspaces and Clusters. In the Workspace, ingested data, notebooks, etc.,

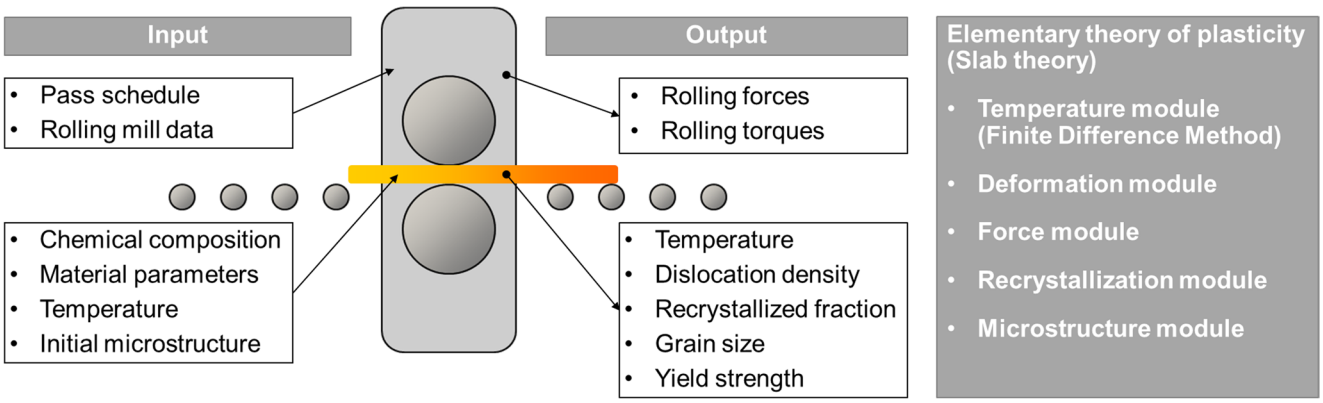

Fig. 3 Illustration of the fast analytical hot rolling model which is developed in modular fashion to calculate global as well as local variables

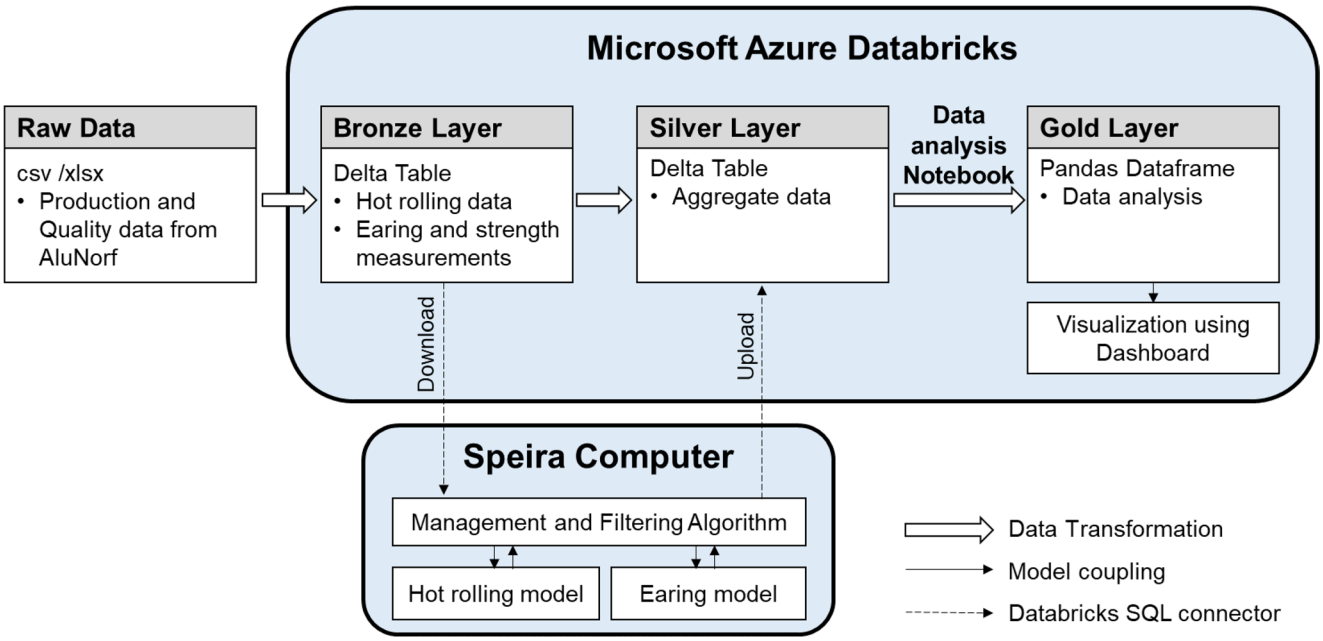

Fig. 4 Concept for storing, processing, and analysing the data along the process chain using Azure Databricks. Databricks ingests production and quality data and stores them in its data lake as a delta table from where the data can be used for further processing to prepare the aggregate data. Finally, the data analysis notebook running on Databricks performs feature importance analysis, identifies optimal processing route and visualizes the results

are located, while Clusters provide the computing power to fetch the data from the Workspace and run Jupyter-style notebooks written in Python, Scala, R, or SQL to process and analyse data. The results can be visualised within the notebook as well as in an interactive dashboard.

The concept for using the Azure Databricks platform for this work is shown in Fig. 4. The raw data from the production and quality testing, either from daily, weekly, or monthly production that is stored locally in.csv format, consisting of hot rolling data, earing measurements and tensile testing data, is uploaded to the platform as a bronze layer. The data from the bronze layer is then fetched by the Management and Filtering algorithm running on the Speira local computer, and the processed aggregate data is uploaded back as a silver layer in the platform using the Databricks SQL Connector for Python. Finally, a Python data analysis notebook running in Azure Databricks analyses the aggregate data for user user-defined period and its results are visualised in the interactive dashboard.

Concept for coupling of process model and production data

In Fig. 5a), a concept for coupling the production data and process models is presented. The input data, consisting of hot rolling data and quality data, is fetched from the Microsoft Azure Databricks platform. At the core of the presented approach is the management and filtering algorithm, whose workflow is shown in Fig. 5b), where all the logic and functions are coded using Python. The Algorithm runs iteratively over each production run by first loading the earing profile as a measurement of the cup heights from 0–360° along its circumference, checks it for missing data and corrects it, if needed, using a cubic spline interpolation method. The corrected earing profile is then analysed by a peak detection algorithm shown in Fig. 6, which identifies the number of peaks representing ears and the number of valleys representing troughs in the earing profile as shown in Fig. 7a).

The peak detection algorithm presented in Fig. 6 starts by adding extra signals towards the start and end of the corrected earing profile (E_c) to avoid abrupt changes in the smoothened profile. Next, the extended earing profile is smoothed to remove the noisy data with the help of a Savitzky-Golay filter. Using the findpeaks library [29] available in Python, all the peaks and valleys in the smoothened extended earing profile (E_s) are detected and duplicate peak or valley present at 0° and 360° is removed. These steps of smoothing, peak/valley identification and duplicate removal are repeated for different smoothing parameter combinations namely the order of the polynomial used in filtering and the length of the filtering window until the termination criteria of an even number of peaks and valleys as well as a minimum distance of 10° between 2 consecutive peaks or valleys is fulfilled. Finally, the location of peaks and valleys identified in the smoothed earing profile is then used to identify peaks and valleys in the corrected earing

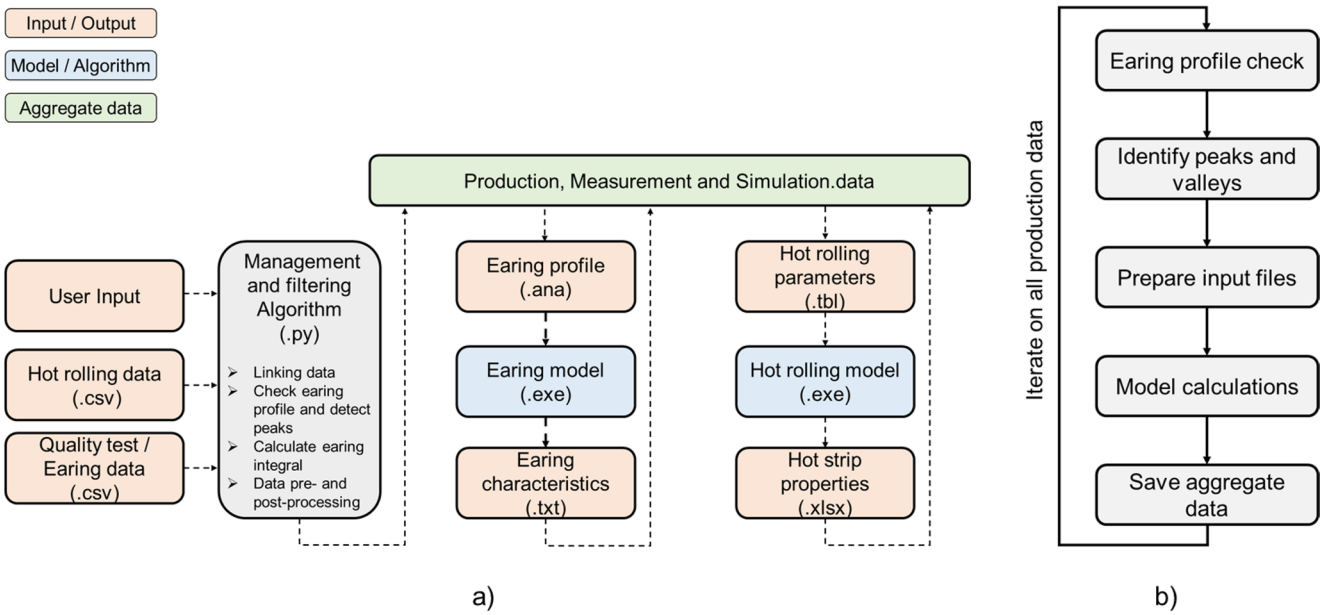

Fig. 5 (**a**) Concept for the coupling of the production data and process models in the aluminium strip process chain, (**b**) workflow of the management and filtering algorithm showing the logic of different steps and processing

1. Input: Corrected earing profile (E_c)
2. Extended earing profile (E_C') = Patch 0-10° signal at the end of E_c and 350-360° signal at the start of E_c
3. Initialize the smoothing parameter combination matrix (order, window)
4. Repeat
5. E_s = Smoothened extended earing profile
6. Identify peaks p and valleys v in E_s
7. Check for the duplicate peaks or valleys at 0° and 360°
8. If True: Remove the detected peak or valley at 360°
9. Until Termination (Even number of peaks and valleys, Distance between consecutive peaks > 10°)
10. Find exact values and location of peak and valley in E_c
11. Output: Peaks and valleys values, angle

Fig. 6 Peak detection algorithm to detect peaks and valleys in the corrected earing profile

profile to get the exact values of peak or valley heights and angles as shown in Fig. 7a).

If the peaks determined by the peak detection algorithm are located at the physically admissible peak positions, the corrected earing profile is used to calculate the earing integral as depicted in Fig. 7b) by using Eqs. 2−1.

$$Earing\ integral\ [\%] = \left(\frac{A_{total} - A_{eariing\ free}}{A_{total}}\right) \times 100 \qquad (2\text{-}1)$$

where,

$A_{total} = Area\ under\ the\ earing\ profile\ calculated\ using\ the\ Simpson\prime s\ rule$

$A_{eariing\ free} = Area\ below\ the\ lowest\ valley\ in\ the\ earing\ profile$

The corrected earing profile is then sent to the earing model to calculate earing characteristic values, namely $Z, \Delta Z$, Z_{0°, Z_{45° and Z_{90°. Using a unique identifier, the associated hot rolling production data is fetched and sent to the hot rolling model to simulate the hot rolling process and calculate difficult to measure variables like recrystallised volume fraction, grain size, yield strength of strip, etc. Next, all the quality data, hot rolling production data, results of hot

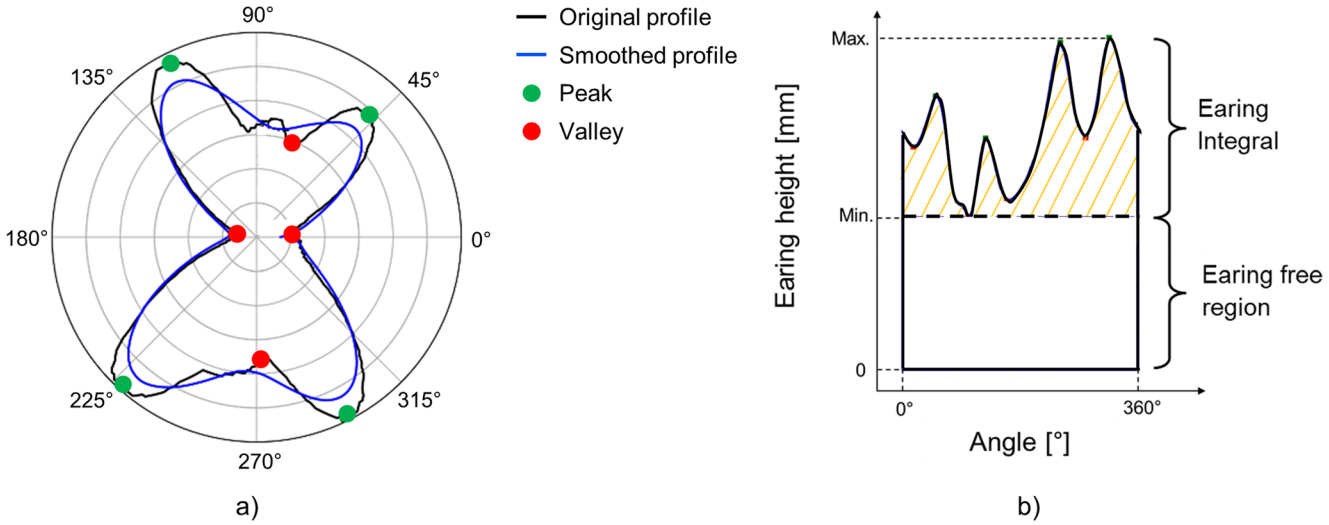

Fig. 7 Processing of earing profiles (**a**) peaks and valleys in the earing profile identified by the peak detection algorithm (**b**) calculation of earing integral using Simpson's rule

rolling and earing model are saved under a unique identifier as aggregate data. Finally, upon completion of the processing of the complete historical production data, the aggregate data is analysed in an automated way to determine parameter importance as well as the best processing route for minimizing earing formation.

Modelling of the coupled process chain

Study of the parameters influencing the earing formation.

The aggregate data is used to extract process knowledge and insights, thereby leading to better process understanding. For this purpose, different methods from exploratory data analysis, like histograms and scatter plots, are used. Since industrial process chains are extremely complex and have non-linear correlated process variables, a nonlinear regression technique from supervised learning called extreme gradient boosting (XGBoost) [30] is used. It is used to map the linear as well as nonlinear interactions between process variables and earing integral. One of the inherent features of XGBoost is that it can provide variance in the model output for each of the input features, thus serving as an indicator of the importance of each input feature (process parameters) in predicting the model output (earing integral).

Figure 8 shows the flowchart of the training process for the XGBoost model using the automatic hyperparameter optimization software framework named Optuna [31], which is available in Python. The objective function containing upper and lower limits for the hyperparameters is defined. For robust determination of the hyperparameters while reducing the influence of random splitting of the data into training and testing datasets, the mean cross validation

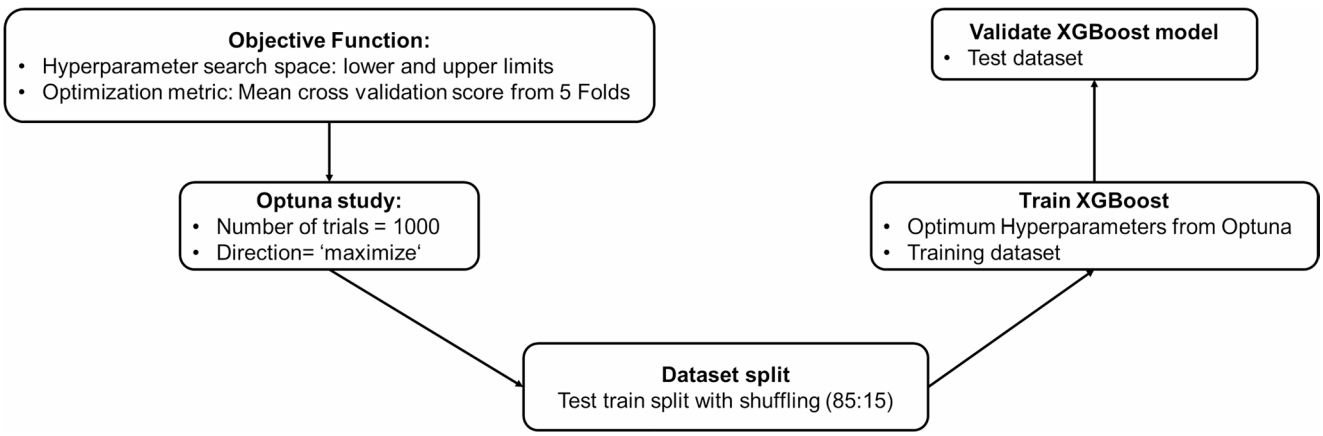

Fig. 8 Flowchart for training of the XGBoost model using Optuna hyperparameter optimization framework

score, calculated from a 5-fold cross validation, is used as a metric for the hyperparameter optimization. A random seed of 42 is used everywhere to ensure the reproducibility of the training process. Next, an Optuna optimization study is created for 1,000 iterations with the goal of maximizing the defined optimization metric. Using the optimum hyperparameters, the complete dataset is split (85/15) into training and testing datasets using the train_test_split function available in Scikit-learn [32], with the shuffle option activated. The XGBoost model is trained on the training data and finally validated using the test data. Finally, the local explanations of the gradient boosted tree model based on SHapley Additive exPlanation (SHAP) values [33] are used to obtain further insights into how the model uses input features to make predictions.

Identification of an optimum processing route.

The results of the data analysis in Sect. 2.4.1 helped to identify the influence of the different process parameters on the earing formation. However, they cannot determine which process parameters should be selected to produce strips fulfilling a given set of prescribed requirements. Therefore, a model-free and gradient-free approach using an evaluation function is necessary in this work to evaluate each individual production run and calculate a scalar evaluation value that specifies how closely all user-specified requirements are met. This approach identifies optimal production routes by analysing the available historical production data. For demonstration purposes, a set of pseudo-requirements is specified as follows.

I. Cold strip thickness ($h_{coldstrip,\ ref}$)=0.243 mm.
II. Cold strip strength ($R_{p02\ coldstrip,\ ref}$)=283 MPa.
III. Homogeneous earing properties at the centre and edge of the cold strip.
IV. Minimum earing formation ($Earing\ integral$).

An evaluation function is used to evaluate each production run for the above defined pseudo-requirements to identify an optimum processing route. The best process route is the one having the lowest value of the evaluation function, or in other words, it is the processing route that closest fulfils the defined pseudo-requirements. The evaluation function presented in Eqs. 2–2 has four components, each belonging to one of the pseudo-requirements.

$$\begin{aligned} Evaluation\ function = &\ E_{thickness} \cdot \\ & w_1 + E_{strength} \cdot\ w_2 \\ + &\ E_{homogenous\ properties} \cdot\ w_3 \\ & + E_{earing} \cdot\ w_4 \end{aligned} \tag{2-2}$$

where,

$$E_{thickness} = \begin{cases} 0.0, & (1-tol) \times h_{coldstrip,ref} \leq \\ & h_{coldstrip} \leq (1+tol) \times h_{coldstrip,ref} \\ 1.0, & otherwise \end{cases}$$

$$E_{strength} = \begin{cases} 0.0, & (1-tol) \times R_{p02\ coldstrip,ref} \leq R_{p02\ coldstrip} \\ & \leq (1+tol) \times R_{p02\ coldstrip,ref} \\ 1.0, & otherwise \end{cases}$$

$$E_{earing} = |Earing\ integral|$$

The parameter tol is the allowable tolerance in the cold strip thickness and strength within a pseudo customer specification, and w_{1-4} are weights to control the penalty associated with each individual component of the evaluation function. To test the suitability of the proposed approach, the tolerance (tol) of 2% and weights w_1= 25, w_2 = 25, w_3 = 1, w_4 = 1 are used. Defining tolerance helps in avoiding situations where the strip's strength and thickness do not exactly match with the pseudo specifications. If the tolerance is relaxed, then the evaluation function will allow cold strips with larger deviations in target thickness and strength. The weights in the evaluation function penalize the production routes if the corresponding property lies outside the allowable tolerance. Since the strip thickness and strength are the strictest customer specifications since they affect the cup making directly, they are given the highest penalty if outside the tolerance. Higher w_1 and w_2 are essential to be able to identify drastic changes in the evaluation values to identify areas in which the process route does not produce acceptable results. Significant variations in properties from the strip centre to the edges, as well as a significant earing formation, can lead to increased scrap generation at the beverage can maker and are therefore given equal weightage by defining $w_3 = w_4$. Thus, all coils within the tolerance window for strength and thickness will give 0.0 for $E_{thickness}$ and $E_{strength}$, whereas $E_{homogenous\ properties}$ and E_{earing} components of the evaluation function will favour coils with homogenous properties and low earing formation.

Results and discussion

Study of the parameters influencing the earing formation.

The production data of 1,972 coils, made available by Speira and AluNorf GmbH, are processed with the help of the implemented concept of coupling production data and process models presented in Sect. 2.3 to prepare aggregate data containing production, quality and simulation results. The coils are not pre-selected but come from a continuous time period and belong to a single alloy specification

used for beverage can production. They are rolled to a similar final thickness in the range of the customary industry standards for beverage cans. The most significant process variations occur since different customers require different widths and strengths and the alloy composition may vary within the allowed alloy specification due to usage of different scrap sources. In hot rolling, the thermal conditions of the equipment are important, which may change over time periods with higher or lower production loads.

The aggregate data is visualised using normalised histograms for the important process variables along with their distribution, as shown in Fig. 9. All the hot rolling parameters are normalised between 0.0 and 1.0 using the minimum and maximum value for each parameter across all the four passes. The spread in the histograms of the process parameters can largely be attributed to the deviations in their chemical composition, ingot geometry and temperature resulting in the different rolling and earing characteristics.

It can be observed that the hot rolling process operates stably and employs large thickness reductions during the first pass, where the strip temperature is highest due to lower material resistance to deformation. As expected from rolling process theory, the shape factor varies across the passes, exhibiting lower values during pass 1 when the strip is thicker compared to pass 4, where the strip is thinner and has a higher shape factor. This variation is attributed

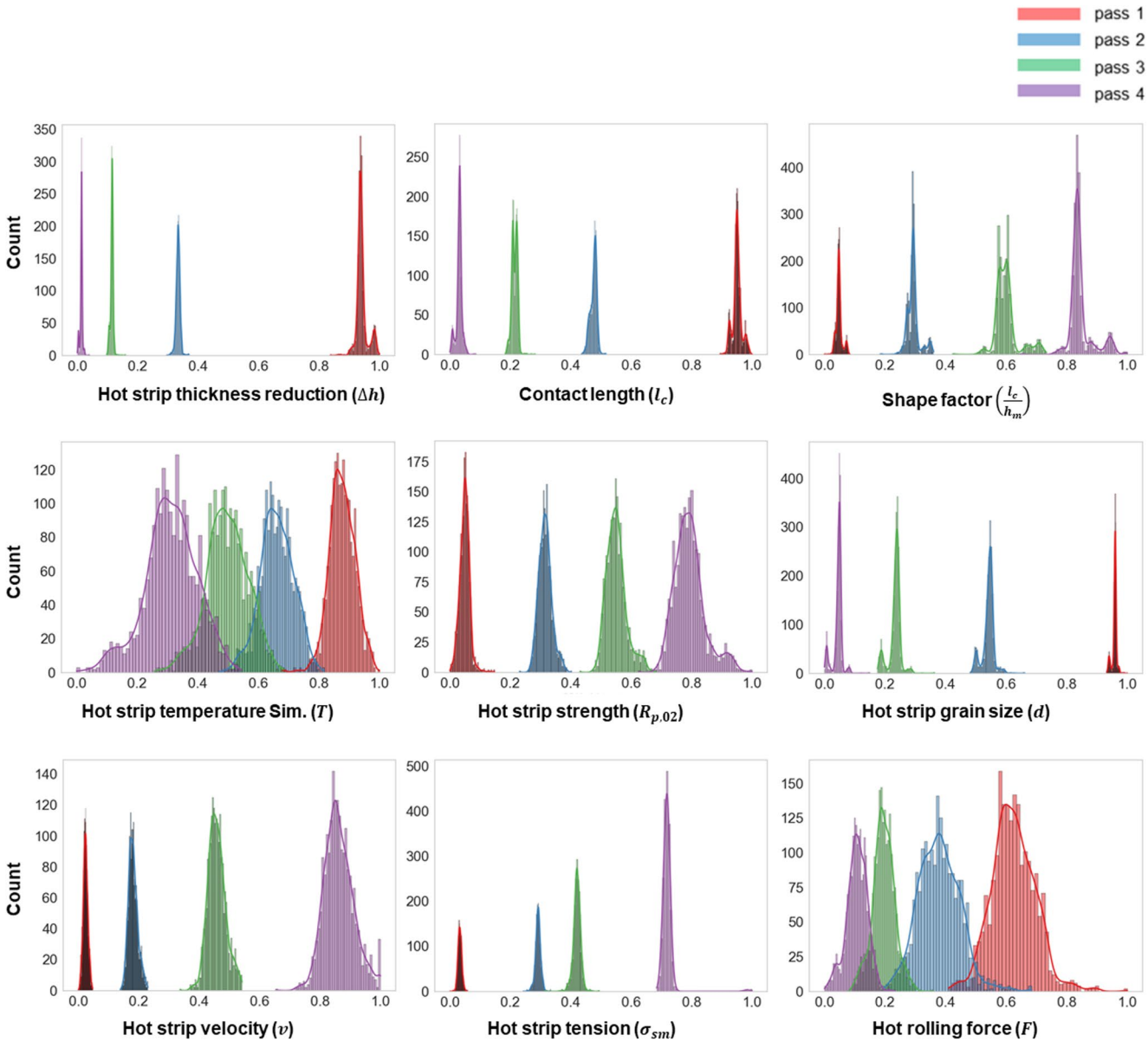

Fig. 9 Visualisation of the key process variables (normalised) providing a complete overview of the hot rolling process

to larger thickness reductions during initial passes, which result in increased projected contact length as well as higher mean strip thickness. In hot rolling, rolling force is one of the particularly important quantities used to analyse and design the hot rolling process because it is measured accurately and integrally reflects the state of the material under given thermomechanical loads. The relationship between different rolling process parameters and rolling force can be roughly estimated by using the elementary theory by Siebel [34] as given by Eqs. 3 − 1.

$$F = (1.15) \cdot l_c \cdot w \cdot k_{fm} \cdot Q_f - \sigma_{sm} \tag{3-1}$$

where,

$$l_c = contact\ length = \sqrt{r.\Delta h - \left(\frac{\Delta h}{2}\right)^2} \approx \sqrt{r \,.\, \Delta h}$$

$$r = roll\ radius$$

$$\Delta h = Height\ reduction = h_{initial} - h_{final}$$

$$w = strip\ width$$

$$k_{fm} = mean\ flow\ stress\ of\ the\ material$$

$$Q_f = Geometry\ factor = 1 + \left(0.5 \cdot \mu \cdot \frac{l_c}{h_m}\right)$$

$$h_m = mean\ strip\ height = 0.5 \cdot (h_{initial} + h_{final})$$

$$\mu = friction\ coefficient$$

$$\frac{l_c}{h_m} = shape\ factor$$

$$\sigma_{sm} = mean\ strip\ tension$$

This can help to interpret the more complex interactions of the full rolling model in the system and to check the plausibility. Some principal dependencies can be observed in the analysed production data, as shown in Fig. 10.

An increase in height reduction leads to an increase in contact length, both of which contribute to an increase in rolling force, while an increase in the shape factor results in a reduction of the required rolling force. Additionally, increasing strip tensions also causes a decrease in roll forces. As observable from Eqs. 3−1, an increase in strip temperature leads to a decrease in mean flow stress (k_{fm}) due to reduced material resistance to deformation, thereby resulting in a reduction of the rolling force. However, an opposite correlation can be observed from the subplot for

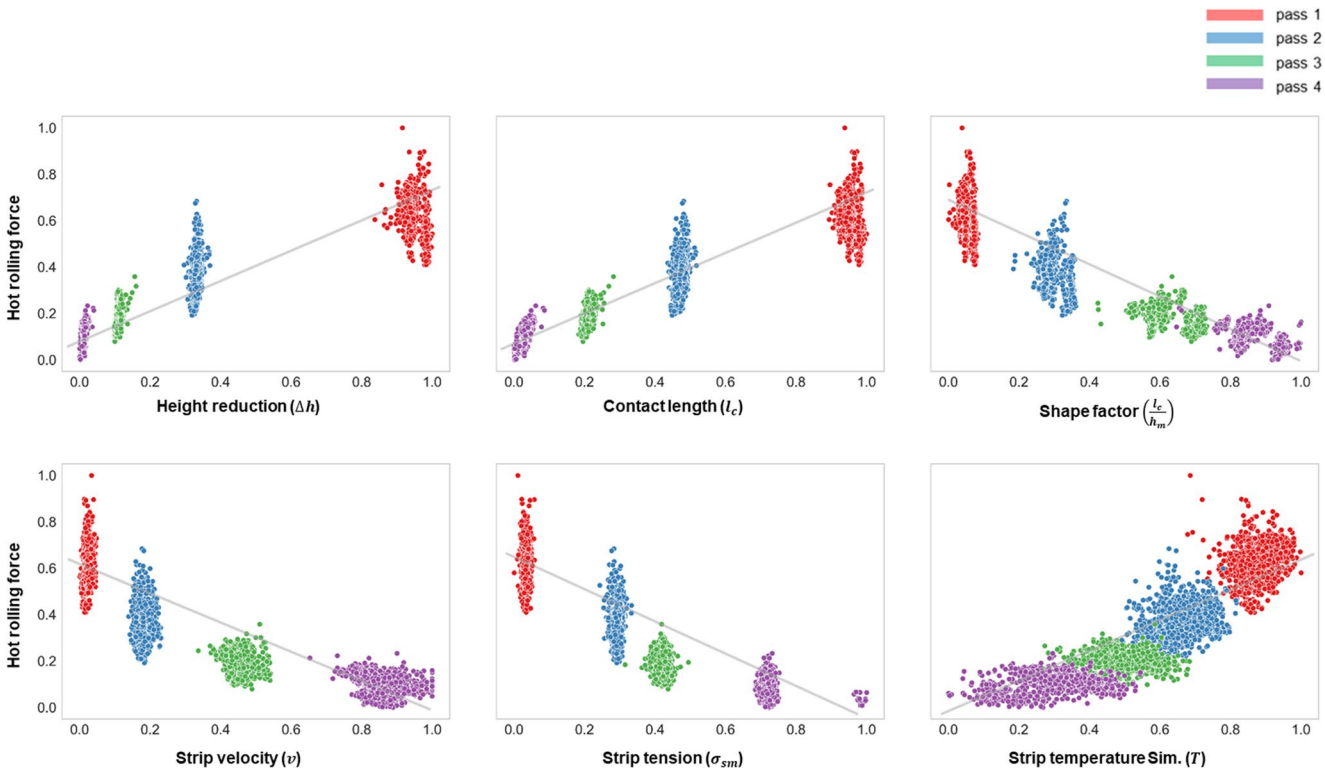

Fig. 10 Visualisation of key technological hot rolling process parameters in regards to their relationship with the rolling force for all 1972 production runs each consisting of a complete pass schedule with four finishing passes

strip temperature (sim.) against rolling force prepared by plotting the aggregate production data. This contradictory correlation can be attributed to the use of larger contact lengths (l_c) together with lower mean strip tensions (σ_{sm}) during the first pass resulting in larger rolling force as compared to the fourth pass, where relatively smaller contact lengths and higher mean strip tensions are employed. Considering the volume constancy and neglecting the spreading due to significantly less material flow in the width direction, increasing height reduction over each pass causes increase in the length of the strip. Hence the rolling velocity must increase over passes to be able to process a longer strip. This is evident from the subplot showing the rolling velocity where rolling velocity is lower in first pass when strip length is comparatively small and higher rolling velocity during fourth pass when strip is comparatively longer.

Next, the aggregate data from all 1,972 production runs is visualised using the cobweb plot shown in Fig. 11. This plot provides the overview of all the production routes (indicated by blue colour) as well as those with an earing integral below 0.1 (indicated by black colour), which indicates good processing routes resulting in lower earing formation.

The densely populated regions indicate the frequency of using a particular processing route. However, only linear interdependencies between any two successive plotted parameters can be observed through such a representation of the data. It can be seen that the majority of the good processes show high values of hot strip earing at a 45° angle to the rolling direction, Z_{45° (HS) and a low value of cold strip earing at a 45° angle, Z_{45° (CS). Furthermore, all good processes have earing peaks≥4 and a hot thickness reduction greater than 0.6, which implies lower cold thickness reduction is required to achieve a similar final strip thickness.

For conducting the feature importance analysis using the XGBoost regression technique, initially, the optimum hyperparameters are identified using the Optuna framework available in Python. The results of the hyperparameter optimisation conducted for 1000 iterations and the scores of the model trained with optimum hyperparameters are listed in Table 1.

A histogram of the true earing integral (Fig. 12a) indicates that the majority of production runs result in an earing integral between 0.0 and 0.75, following a normal distribution cantered around a mean of 0.35. The trained XGBoost model is validated using the test dataset, and the validation results are presented in Fig. 12b. The test data is well distributed across the whole range and the trained XGBoost model shows good agreement with the experimental data with a validation score of 0.864. The comparatively lower validation score can be explained by the limited amount of testing data.

The trained XGBoost regression model provides feature importance scores for each of the input features, explaining the variance in the output parameter, which in this case is the earning integral. Figure 13a) shows the most relevant input features, sorted in ascending order of their importance from bottom to top. The importance of each input feature is assessed based on its contribution to predicting the output feature. It can be observed that the top ten input features collectively contribute to approximately 50% of the model's prediction of the earing integral. Thus, 50% of the total

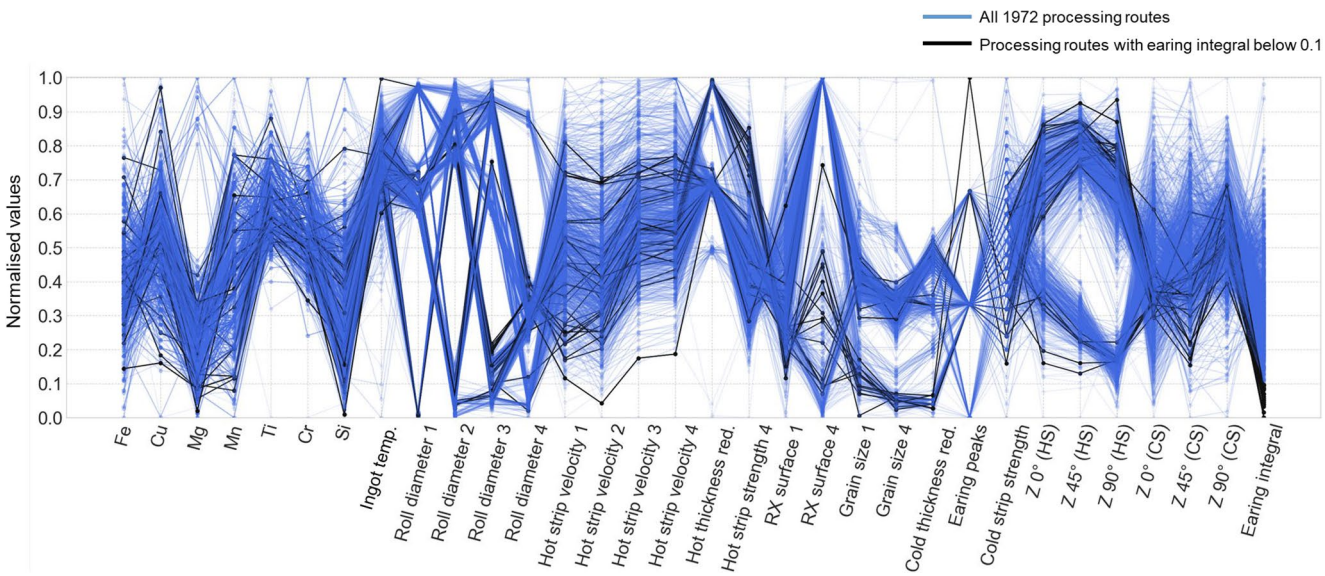

Fig. 11 Visualisation all the 1972 production runs (blue colour) and processing routes with earing integral less than 0.1 (black colour) using parallel coordinates plot. Lines connecting between two parameters represent parameters used during same production run. The densely populated areas represent the frequency of the production run with same or similar process parameters

Table 1 Hyperparameter optimisation of XGBoost regressor model using optuna framework, along with the training and testing scores of the model trained with the identified optimum hyperparameter values. The hyperparameter search space is defined based on prior experience and further modified through trial and error

Hyperparameter	Description	Lower limit	Upper limit	Optimum value	XGBoost training score	XGBoost testing score
n_estimators	Number of gradient boosted trees	10	2500	2229	0.982	0.864
max_depth	Maximum tree depth for base learners	1	1000	506		
learning_rate	Boosting learning rate	0.001	1.0	0.812		
subsample	Subsample ratio of the training instance	0.0	1.0	0.788		
colsample_bytree	Subsample ratio of columns when constructing each tree	0.0	1.0	0.682		
reg_alpha	L1 regularisation term on weights	0.0	10.0	0.203		
reg_lambda	L2 regularisation term on weights	1.0	500.0	452.01		

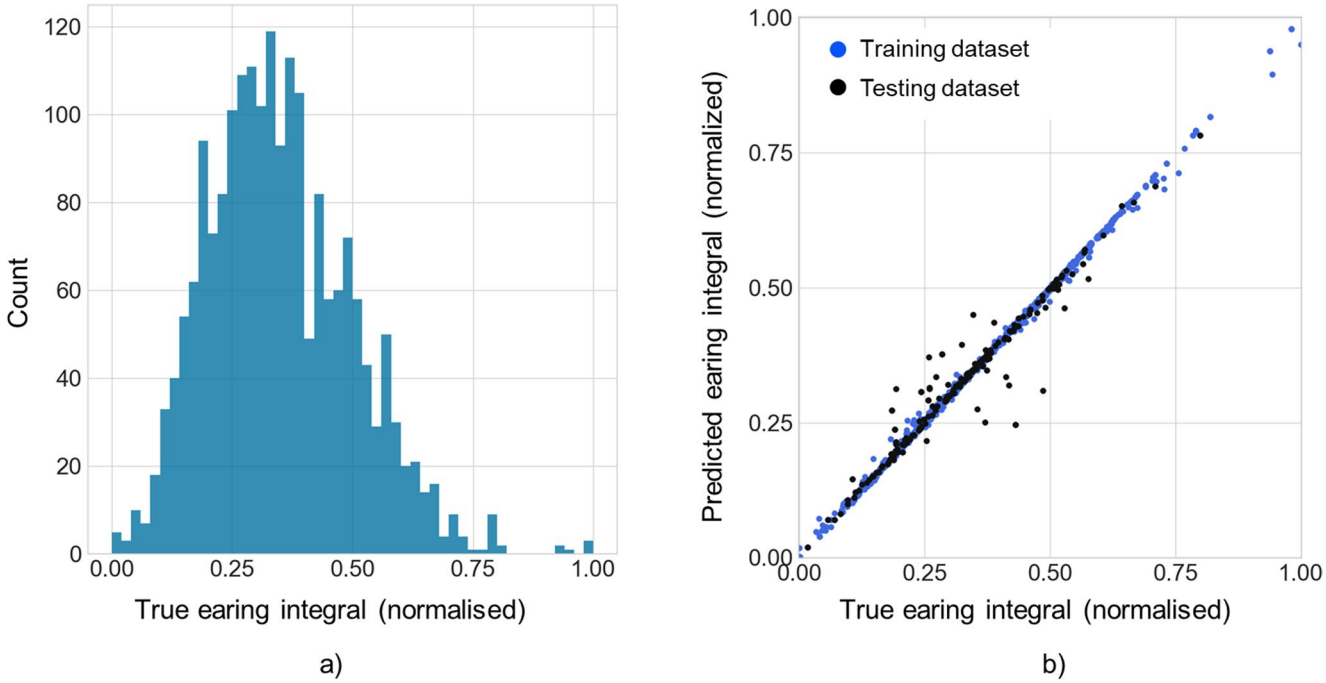

Fig. 12 (**a**) Histogram showing the distribution of earing integral calculated using the cold strip earing measurements; (**b**) Validation of the XGBoost model using the validation data

variance in the model output is explained by the ten most significant input features. However, the feature importance score does not provide additional information regarding the nature of the effect between input and output features. This makes it difficult to test the plausibility of the input-output mapping learned by the model by looking for the know or expected relationships.

In addition, SHapley Additive exPlanation (SHAP) values are utilized for further analysis, as they provide better insights into the learned input-output mappings. The SHAP values quantify and visualize how individual features impact model predictions, thereby enhancing transparency and trust in the model. The trained XGBoost regression model is employed to calculate the SHAP values for each input feature. All input features are sorted in order of their influence on earing integral, with the most influential feature at the top. An input feature with a positive SHAP value causes an increase in the earing integral, while a negative SHAP value results in a decrease in the earing integral and the magnitude of the SHAP value indicates the scale of influence.

From the SHAP summary plot shown in Fig. 13b), it is observed that roll diameter in pass 2 is the most influential factor, followed by silicon content, cold thickness reduction, and recrystallized fraction after pass 1. The use of rolls with a smaller diameter in pass 2 (indicated by the blue colour) can result in reduced earing formation, as the corresponding SHAP value is less than zero. In contrast, larger

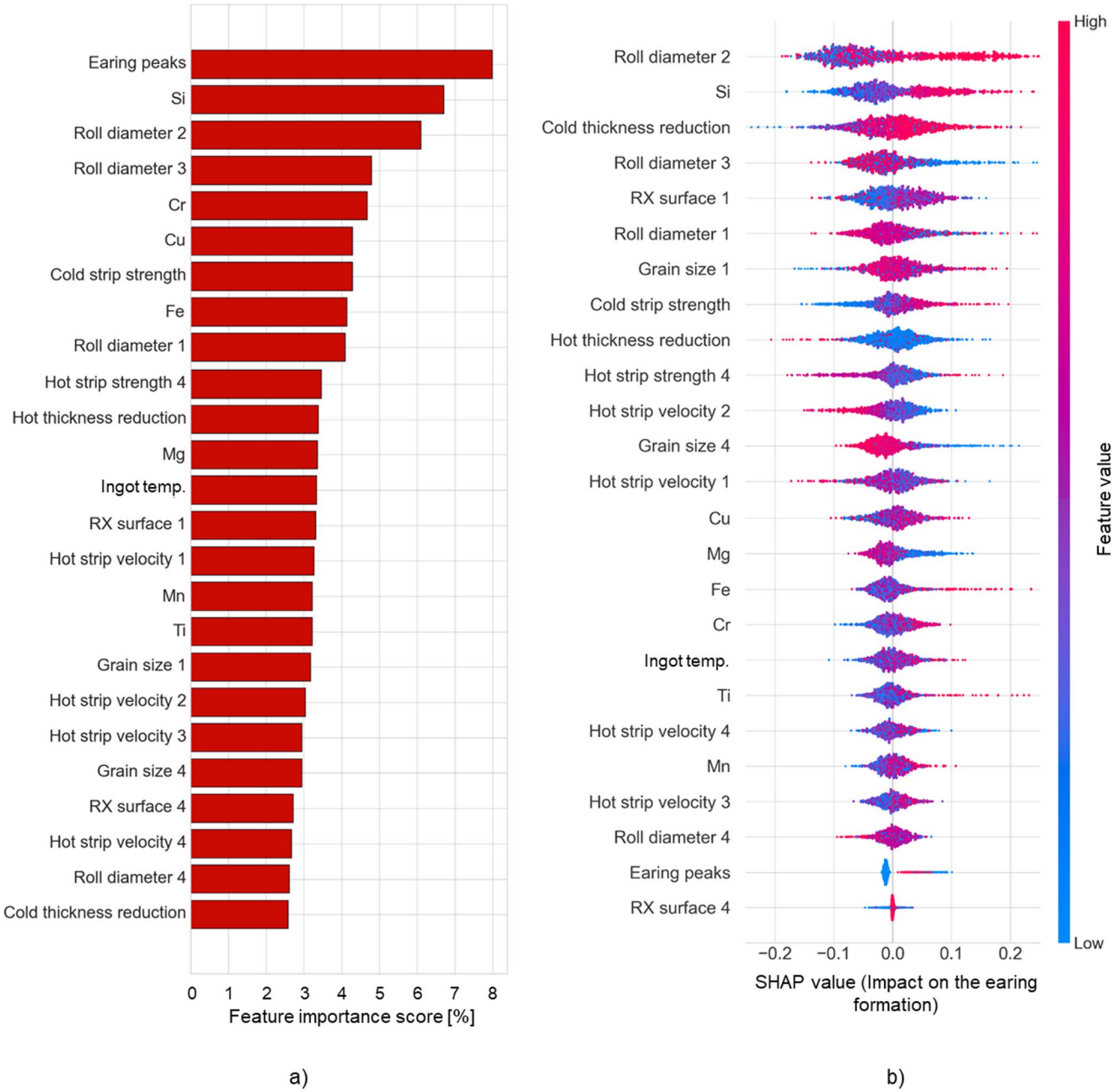

Fig. 13 Feature importance analysis with all input features sorted in ascending order using; (**a**) the feature importance score provided by the trained XGBoost regression model to identify most influential input features, (**b**) SHAP values showing the local impact of each input feature on the earing formation

roll diameters in pass 2 (indicated by the red colour) do not exhibit a clear distinction, as they are distributed across both positive and negative SHAP values. This suggests that a large roll diameter in pass 2, in combination with other process parameters, can lead to both reduced and increased earing formation. A physical explanation for this is, that the roll diameter changes the shear that the roll exerts on the strip, which influences the texture and thereby changes the earing. However, the magnitude of this effect coming from the data analysis is surprising. A lower cold thickness reduction (indicated by the blue colour) generally promotes reduced earing formation. However, the presence of a few blue dots with positive SHAP values suggests that, together with other process parameters, a lower cold thickness reduction can also lead to increased earing formation. Thus, a lower roll diameter in second pass, reduced lower silicon content and reduced cold thickness reduction are favourable for minimizing the earing formation, while high iron and

titanium contents are unfavourable. Additionally, a high hot thickness reduction is beneficial, and a large grain size in pass 4 which indicates complete recrystallisation, is advantageous for minimizing the earing formation.

Identification of an optimum processing route

Using the defined pseudo-requirements and the evaluation function, all the 1,972 production runs are analysed and sorted according to their evaluation values, as shown in Fig. 14. The steep jumps observed in the evaluation values are attributed to the high penalty factor applied to processes that do not meet the tolerances for cold strip thickness and yield strength. Violations are penalized by assigning a high weight to the respective component in the evaluation function (Eqs. 2–2). Thus, all processes with evaluation values between zero and the first jump satisfy the desired tolerances strip strength and thickness while exhibiting varying levels of earing characteristics.

The production run with a normalised evaluation value of zero represents the best processing route, as it closely fulfils all the pseudo-requirements according to the criteria set in the evaluation function. Similarly, the production run with a normalised evaluation value just at the beginning of the jump represents the worst processing route, while it meets the tolerances for cold strip thickness and strength but exhibits higher earing formation and greater inhomogeneity in the earing properties across the strip width. The characteristics of both the best and worst processing routes are summarized in Table 2. Both routes meet the final strip thickness and strength requirements, however, the best processing route results in strips with an earing integral of 1.151%, which is approximately 2.7 times lower than that of the worst processing route.

The earing profiles of both best and worst processing routes is shown in Fig. 15. It can be observed that the earing profile belonging to the best processing route (indicated by the blue colour) exhibits a relatively flat profile with smaller peaks and valleys, indicating lower earing formation. In contrast, the profile associated with the worst processing route (indicated by the red colour) shows significant fluctuations with larger peaks and valleys, indicating higher earing formation.

Figure 16 shows the best (indicated by green line) and worst (indicated by dotted red line) processing routes, along with limits of the top five best processing routes representing the best processing window (black lines). The magnesium, titanium, and silicon contents of the worst processing route are higher than those of the best route, and they also lie outside the best processing window. In accordance with the observations from the feature importance analysis as discussed in Sect. 3.1, the worst processing route uses larger work rolls during pass 2 compared to the best processing route. Additionally, the ingot temperatures and rolling velocities for the worst processing route are lower than those of the best route and they lie outside the best

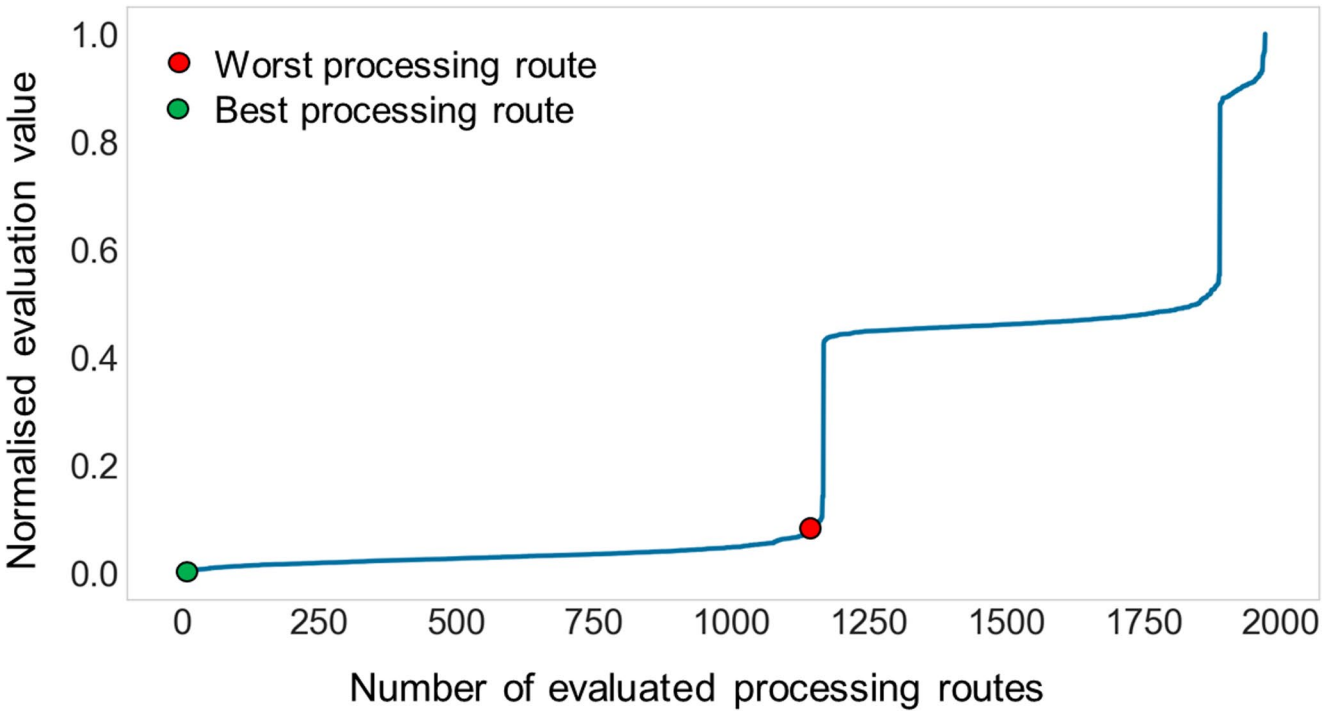

Fig. 14 Aggregate data of 1972 production runs evaluated using the evaluation function and sorted in ascending order along with the evaluation values of best (green circle) and worst (red circle) processing routes

Table 2 The cold strip properties of the best and the worst processing routes fulfilling the pseudo-requirements identified using the evaluation function

Processing route	Alu. strip thickness [mm]	Alu. strip strength [MPa]	Z_0 (Centre/Edge) [%]	Z_{45} (Centre/Edge) [%]	Z_{90} (Centre/Edge) [%]	Earing Integral [%]
Best	0.239	288	0.067/0.187	1.031/0.797	−0.541/−0.539	1.151
Worst	0.240	287	2.49/0.864	3.580/0.778	0.617/−1.610	3.058

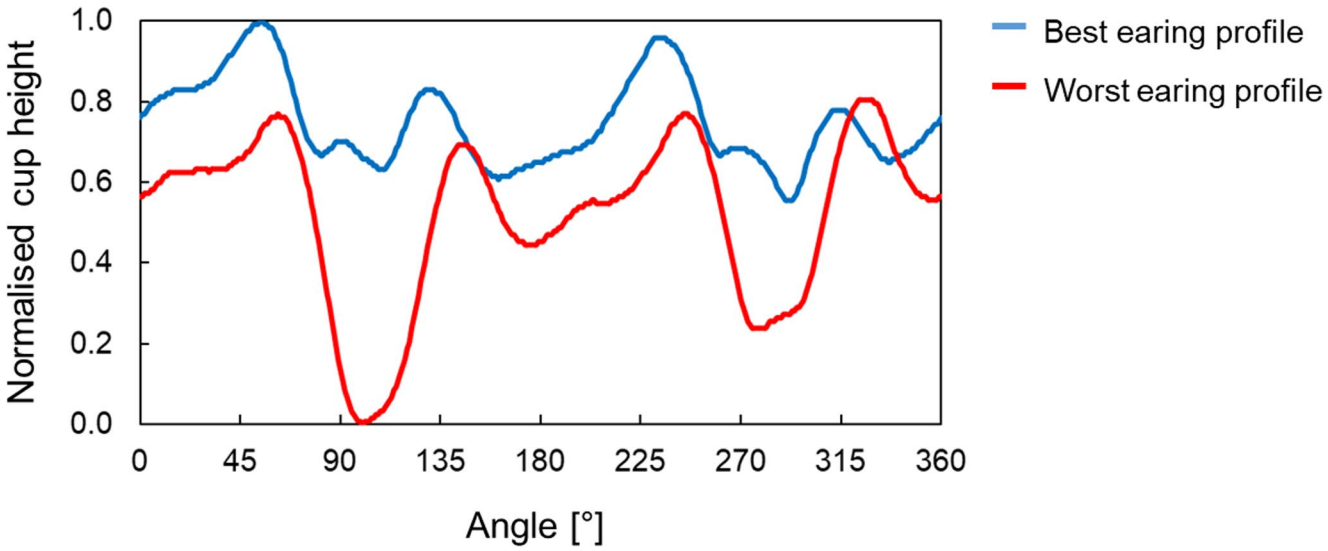

Fig. 15 Earing profile belonging to the best and worst processing route identified using the evaluation function

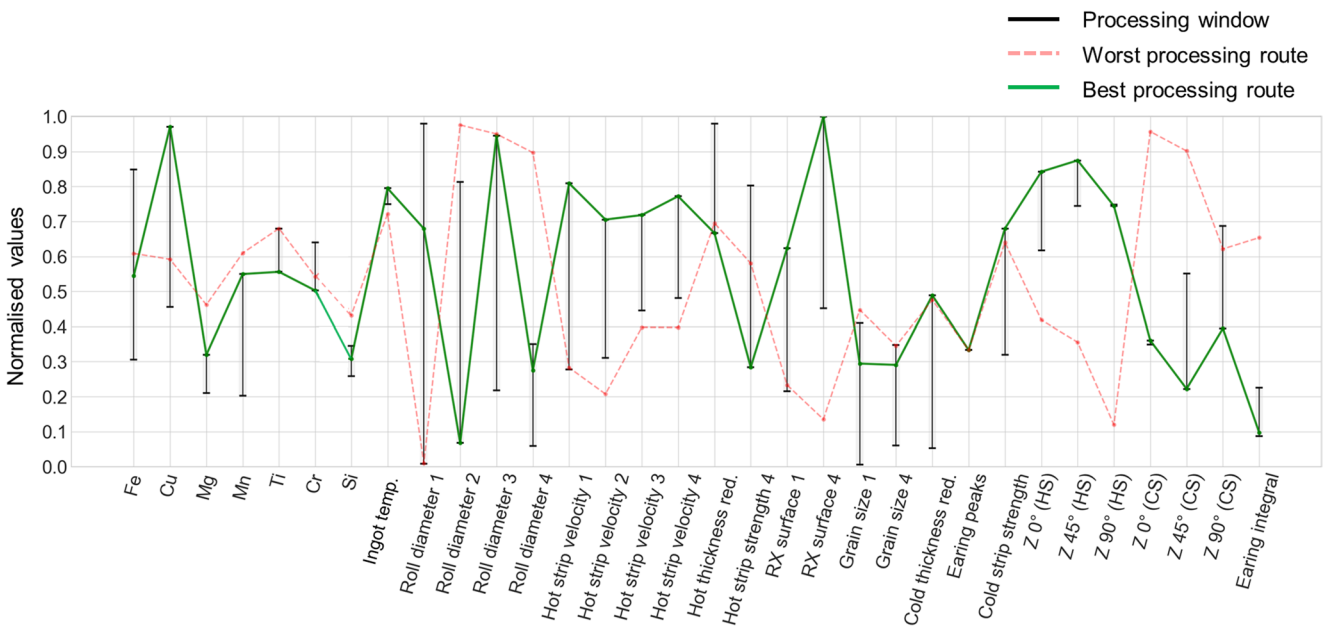

Fig. 16 Parallel coordinates plot showing the best (green colour) and worst (red colour) processing routes as well as the best processing window (black colour) identified using the evaluation function

processing window. The recrystallised fraction after pass 1 and pass 4 both indicate that higher or complete recrystallisation is favourable for minimizing earing formation. It can also be seen that the best processing window requires high hot strip earing characteristics values namely $Z_{0^\circ}\ (HS)$, $Z_{45^\circ}\ (HS)$ and $Z_{90^\circ}\ (HS)$ to meet the pseudo-requirements, while the worst processing route shows low values.

Conclusions

A concept for modelling and analysing an industrial aluminium strip production process chain has been successfully implemented by coupling production data with process models using the Microsoft Azure Databricks platform. The effectiveness of the proposed approach is demonstrated through the automated processing of data from 1,972 production coils, resulting in the generation of aggregate datasets containing both production and simulation data.

The aggregate data is initially visualised using histograms and scatter plots to explore correlations between hot rolling force and key process parameters such as thickness reduction, contact length, shape factor, rolling velocity, and strip tension. Subsequently, the dataset is used to train an XGBoost model, with optimal hyperparameters identified via the Optuna hyperparameter optimisation framework. The tuned model is then employed to identify the process parameters with the most significant influence on earing formation.

To evaluate process performance, pseudo-requirements are defined, and all production runs are assessed using a dedicated evaluation function. This analysis identifies the best processing route, achieving a minimum earing integral of 1.151%. Furthermore, a robust processing window is established by analysing the five best-performing routes. A comparative analysis between the best and worst processing routes for strips with identical specifications reveals that lower values of rolling diameter in pass 2, and reduced concentrations of silicon, magnesium, and titanium, combined with higher recrystallised fractions and rolling velocities, are favourable for minimizing earing formation.

This work demonstrates the feasibility and advantages of modelling an industrial process chain by integrating production data and process models in an automated, scalable, and efficient manner. The proposed approach not only facilitates deeper process understanding but also serves as a data-driven decision support system for process experts during production planning.

Limitations and future work

The demonstrated approach is implemented to model and analyse a coupled aluminium process chain for can body strip production. However, the implemented evaluation function operates on historical production data to determine an optimal processing route that fulfils a set of user-specified requirements. This means that the demonstrated approach can identify the best processing route from the available historical production data, but it cannot be used in scenarios such as adaptive processing where new process parameters should be generated in response to deviations in the incoming material properties. Therefore, in future work, an approach involving a mathematical optimization algorithm combined with a surrogate model will be investigated to determine an optimal processing route. Additionally, the availability of more production data will help reduce any imbalance in the dataset, thereby further improving the XGBoost model accuracy.

Appendix A

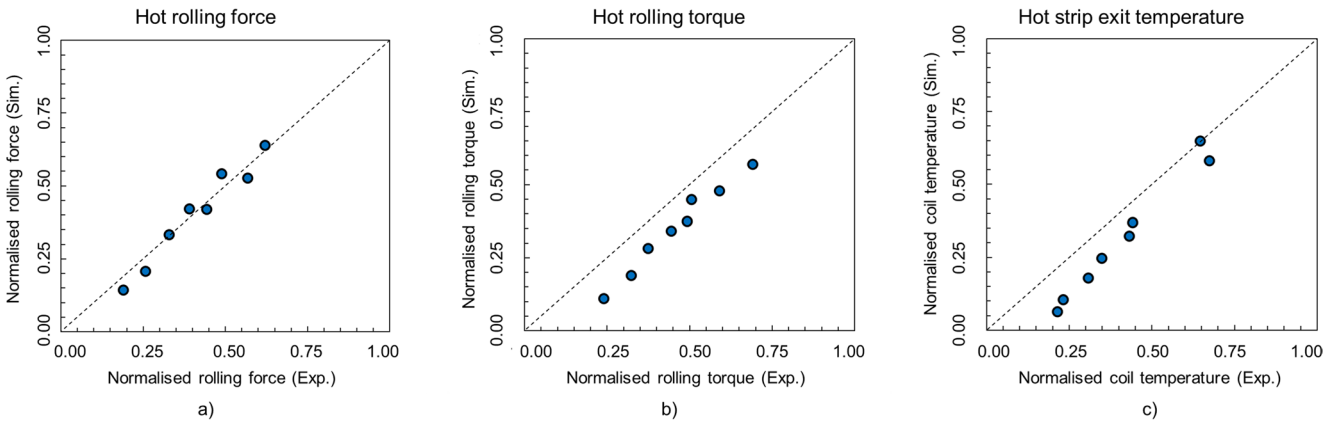

Fig. 17 Validation of the fast physical hot rolling model by comparing the normalised values of rolling force, rolling torque and hot strip temperature in the fourth pass with the normalised experiment values

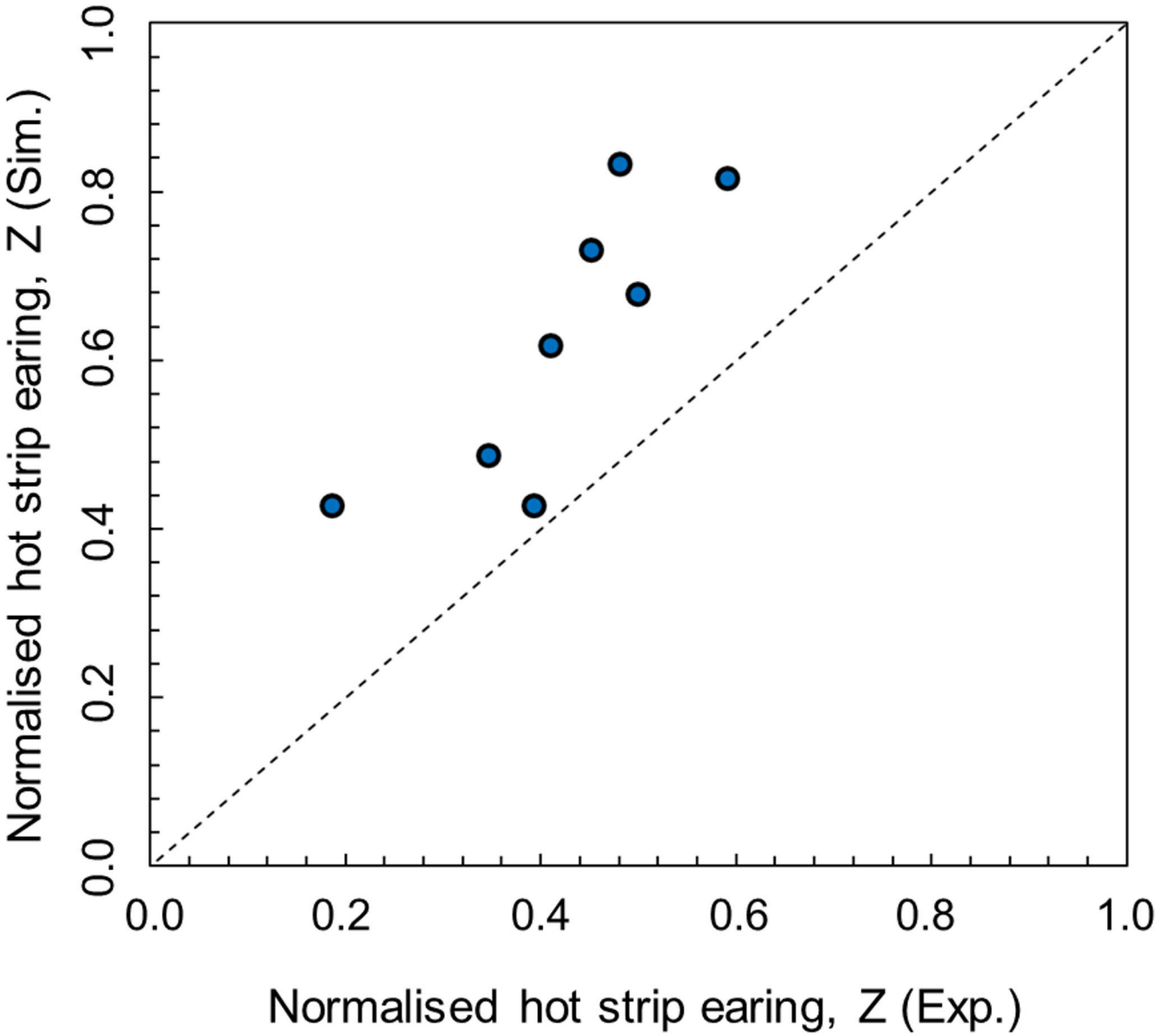

Fig. 18 Validation of the earing model by comparing the normalised earing characteristics calculated from measured hot strip earing profiles and inversely simulated hot strip earing profiles using the earing model

Acknowledgements This work is based upon the results generated within the IGF project 21884 N of the FOGI e.V. funded via the AiF within the framework of the programme for the promotion of joint industrial research (IGF) by the German Federal Ministry for Economic Affairs and Climate Action on the basis of a resolution of the German Parliament followed by a subsequent implementation project financed by Speira GmbH. The authors would like to thank Prof. Dr.-Ing. Gerhard Hirt for his guidance and insights that have been helpful in the completion of this work.

Author contributions N.T.: Conceptualization, Methodology, Data processing, Data visualization, Formal analysis, Writing-original draft. K.K: Writing-review& editing. H.M., E.S, D.B.: Review and Supervision.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

References

1. Georgantzia E, Gkantou M, Kamaris GS (2021) Aluminium alloys as structural material: a review of research. Eng Struct 227:111372. https://doi.org/10.1016/j.engstruct.2020.111372
2. Riedewald F, Wilson E, Patel Y et al (2022) Recycling of aluminium laminated pouches and Tetra Pak cartons by molten metal pyrolysis – pilot-scale experiments and economic analysis. Waste Manag 138:172–179. https://doi.org/10.1016/j.wasman.2021.11.049
3. Sun Y (2023) The use of aluminum alloys in structures: review and outlook. Structures 57:105290. https://doi.org/10.1016/j.istruc.2023.105290
4. Zambrano V, Mueller-Roemer J, Sandberg M et al (2022) Industrial digitalization in the industry 4.0 era: classification, reuse and authoring of digital models on Digital Twin platforms. Array 14:100176. https://doi.org/10.1016/j.array.2022.100176
5. Li S, Yue X, Li Q et al (2023) Development and applications of aluminum alloys for aerospace industry. J Mater Res Technol 27:944–983. https://doi.org/10.1016/j.jmrt.2023.09.274
6. Metal Packaging Europe and European Aluminium (2024) Aluminium beverage can recycling in 2021 at a new record level of 76%
7. Gaustad G, Olivetti E, Kirchain R (2012) Improving aluminum recycling: a survey of sorting and impurity removal technologies. Resour Conserv Recycl 58:79–87. https://doi.org/10.1016/j.resconrec.2011.10.010
8. Vicent Fanconi MM, Gil Fernández-Marcote I, Ruiz-Bustinza Í (2023) The challenge of impurities (Fe, Si) to recycling in the rolled aluminum industry in the coming years in relation to their influence on ultimate tensile strength. Metals 13:2014. https://doi.org/10.3390/met13122014
9. De Caro D, Tedesco MM, Pujante J et al (2023) Effect of recycling on the mechanical properties of 6000 series aluminum-alloy sheet. Materials 16:6778. https://doi.org/10.3390/ma16206778
10. Bell S, Davis B, Javaid A, Essadiqi E (2006) Final Report on Effect of Impurities in Aluminum
11. Schroettner H, Pabel T, Petkov T et al (2016) Effect of trace elements on the material properties of an aluminium casting alloy. In: European Microscopy Society (ed) European Microscopy Congress 2016: Proceedings, 1st ed. Wiley, pp 271–272
12. Van Linden JHL, Bachowski R, Miller RE (1988) Chemical impurities in aluminum. In: Katz S, Landefeld CF (eds) Foundry processes. Springer US, Boston, MA, pp 393–409
13. Zaidi MA, Sheppard T (1985) Control of earing quality in AA 5052 and AA 5454 aluminium alloys. Mater Sci Technol 1:593–599. https://doi.org/10.1179/mst.1985.1.8.593
14. Hutchinson WB, Oscarsson A, Karlsson Å (1989) Control of microstructure and earing behaviour in aluminium alloy AA 3004 hot bands. Mater Sci Technol 5:1118–1127. https://doi.org/10.1179/mst.1989.5.11.1118
15. Oscarsson A, Hutchinson WB, Ekström H-E (1991) Influence of initial microstructure on texture and earing in aluminium sheet after cold rolling and annealing. Mater Sci Technol 7:554–564. https://doi.org/10.1179/mst.1991.7.6.554
16. Naess SE (1991) Development of earing and texture during temper rolling of the aluminium alloys AA3005 and AA5050 / Zipfel- und texturbildung während des Warmwalzens der aluminiumlegierungen AA3005 und AA5050. Int J Mater Res 82:259–264. https://doi.org/10.1515/ijmr-1991-820402
17. Zhao Z, Mao W, Roters F, Raabe D (2004) A texture optimization study for minimum earing in aluminium by use of a texture component crystal plasticity finite element method. Acta Mater 52:1003–1012. https://doi.org/10.1016/j.actamat.2003.03.001
18. Engler O, Aegerter J, Calmer D (2020) Control of texture and earing in aluminium alloy AA 8011A-H14 closure stock. Mater Sci Eng A 775:138965. https://doi.org/10.1016/j.msea.2020.138965
19. Sellars CM, Davies GJ, Sheffield M (1980) Hot working and forming processes: proceedings of an International Conference on Hot Working and Forming Processes
20. Beynon JH, Sellars CM (1992) Modelling microstructure and its effects during multipass hot rolling. ISIJ Int 32:359–367. https://doi.org/10.2355/isijinternational.32.359
21. Yuen WYD (2003) On-line and off-line models for the rolling process. Scand J Metall 32:86–93. https://doi.org/10.1034/j.1600-0692.2003.20619.x
22. Lohmar J, Seuren S, Bambach M, Hirt G (2014) Design and application of an advanced fast rolling model with through thickness resolution for heavy plate rolling. ICRF Proceedings, 7–9 May 2014, Milan, Italy
23. Rout M, Pal SK, Singh SB (2017) Finite element modeling of hot rolling. Computational methods and production engineering. Elsevier, pp 83–124. https://doi.org/10.1016/B978-0-85709-481-0.00004-5
24. Karhausen KF (2008) Process Chain Modelling of Al Sheet Production. Wiley-VCH Ed. J.Hirsch :60–70
25. Kishor N, Ravi Kumar D (2002) Optimization of initial blank shape to minimize earing in deep drawing using finite element method. J Mater Process Technol 130:20–30. https://doi.org/10.1016/S0924-0136(02)00790-2
26. Europäisches Komitee für Normung (CEN) (1996) Aluminium und aluminiumlegierungen Prüfverfahren. - Zipfelprüfungan Blechen und Bändern
27. Karhausen K, Kopp R (1993) Application of FEM and Elelementary Theory of Plasticity to Prediction of Microstructure in Hot Rolling. theinstituteofmaterials1993 66–79
28. Etaati L (2019) Azure databricks. Machine learning with Microsoft technologies. A, Berkeley, CA, pp 159–171. https://doi.org/10.1007/978-1-4842-3658-1_10
29. Erdogan Taskesen (2020) Findpeaks is for the detection of peaks and valleys in a 1D vector and 2D array (image). https://erdogant.github.io/findpeaks
30. Chen T, Guestrin C (2016) XGBoost: A scalable tree boosting system. https://doi.org/10.48550/ARXIV.1603.02754
31. Akiba T, Sano S, Yanase T et al (2019) Optuna: A Next-generation Hyperparameter Optimization Framework
32. Pedregosa F, Varoquaux G, Gramfort A et al (2011) Scikit-learn: machine learning in python. J Mach Learn Res 12:2825–2830
33. Lundberg SM, Erion G, Chen H et al (2020) From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2:56–67. https://doi.org/10.1038/s42256-019-0138-9
34. Erich Siebel Werner Lueg Untersuchungen über die Spannungsverteilung im Walzspalt. 14 S: Ill - Mitteilungen aus dem Kaiser-Wilhelm-Institut für Eisenforschung zu Düsseldorf ; Bd 15, Lfg 1=Abh 218 Kaiser-Wilhelm-Institut für Eisenforschung. Mitteilungen aus dem Kaiser-Wilhelm-Institut für Eisenforschung zu Düsseldorf:Bd. 15, Lfg. 1=Abh. 218