The International Journal of Advanced Manufacturing Technology
https://doi.org/10.1007/s00170-025-16888-8

ORIGINAL ARTICLE

Predictive modeling of hardness in bainitic steel forging processes using artificial intelligence

Peterson Duarte Diehl[1] · **André Rosiak**[1] · **Roderval Marcelino**[2] · **Lirio Schaeffer**[1]

Received: 26 May 2025 / Accepted: 22 October 2025

Abstract
This work presents a practical, cost-effective pipeline to predict the hardness of hot-forged bainitic steels directly from process parameters, bringing AI techniques closer to industrial deployment. We combine wedge-forging experiments with inverse FEM to generate a database from which thermomechanical descriptors (equivalent strain, strain rate, and initial and final forging temperatures) and the corresponding hardness values were extracted. We benchmarked Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Random Forests (RF). The ANN, configured with three hidden layers (64, 128, and 256 neurons), achieved the best performance, with a Mean Absolute Error (MAE) of 1.3518 HV, a Mean Squared Error (MSE) of 8.8448 HV², and a coefficient of determination (R^2) of 0.994. Inference tests on forged connecting rods demonstrated robustness and transfer to real components, underscoring the model's potential as a soft sensor for decision support and supervisory process adjustments. By coupling minimal, physically meaningful inputs with high predictive accuracy, the study lowers barriers to AI adoption in hot forming and offers a tool with potential for early drift detection, parameter optimization, and improved quality and operational efficiency in industrial forging.

Keywords Artificial intelligence · Forging · Bainitic steel · Machine learning

1 Introduction

Significant efforts have been made to enhance the mechanical properties of forged steels intended for critical safety components. Among these developments are bainitic steels [1]. Bainitic steels combine the ideal properties of quenched and tempered (Q&T) steels with a shorter production route. Their versatility and cost-effective manufacturing have made them the preferred choice for structural components in automotive engineering. The mechanical strength and toughness values of bainitic steels fall within the range achieved by Q&T steels such as DIN 42CrMo4. Fatigue resistance analyses also indicate superior dynamic properties compared to Q&T steels [2].

The thermomechanical processing route for bainitic steels involves heating to temperatures around 1200 °C, forging, and continuous cooling. The use of continuous cooling, as opposed to prolonged isothermal treatments or conventional quenching and tempering, can significantly reduce energy consumption and, consequently, manufacturing costs [3].

The final properties of forged components are determined by both chemical composition and processing conditions. In particular, forming and cooling parameters are highly sensitive and exert a direct influence on production robustness. Differences in thermal history between the surface and the core of the part can result in distinct microstructures, leading to gradients in mechanical properties [4]. The ability to correlate real-time production data with final product quality—enabling early deviation detection and providing agile, precise decision-making support—makes Artificial Intelligence (AI) tools highly promising for the dynamic adjustment of the bainitic steel forging process.

The application of AI tools has shown great potential to transform the industry, driving significant advances in efficiency, quality, and innovation. In the context of mechanical forming, AI-based models have already been developed to optimize processing routes [5, 6], predict failures and defects [7–11], and estimate material behavior under complex loading conditions, capturing nonlinear relationships

✉ Peterson Duarte Diehl
peterson.diehl@gmail.com

[1] Metal Forming Innovation Center (CBCM), Federal University of Rio Grande do Sul, Porto Alegre, Brazil

[2] Federal University of Santa Catarina, Araranguá, Brazil

Published online: 03 November 2025

between stress and strain [12–16]. AI has also been applied in real-time control of forming processes, especially in quality-monitored systems, where neural networks are trained to automatically adjust process parameters based on sensor data [17–19].

This study investigates the application of machine learning techniques for quality control of bainitic forged products, through the prediction of part hardness based on experimental process data. The analyzed process involves a low-carbon microalloyed steel with the chemical composition 0.18 C–1.38Mn–0.36Si–0.47Cr–0.03Mo–0.05Al–0.04Nb–0.01Ti, which is heated to high temperatures, forged, and continuously air-cooled. Forging is performed on wedge-shaped specimens, allowing the generation of different combinations of thermomechanical parameters along the part. Data obtained through sensing and numerical simulation result in a dataset with 2,955 records, used for training and calibrating Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Random Forests (RF) models.

Defining a model capable of understanding the complex correlations between final hardness data and thermomechanical parameters constitutes an effective tool for process control in forging, enabling the identification of production inconsistencies and the implementation of real-time corrections.

2 Materials and methods

This study investigates the use of machine learning techniques to predict the hardness of forged products based on experimental data. Figure 1 schematically illustrates the methodology employed. The work is developed in three main phases, consisting of: (i) dataset creation; (ii) prediction and analysis; and (iii) inference testing. Each of these steps is detailed below.

2.1 Dataset creation

The hardness of a hot-forged bainitic component is determined by the conditions of the thermomechanical processing [20]. The development of artificial intelligence (AI)-based tools to predict the hardness of such components requires the construction of a robust dataset [21, 22], capable of capturing the interrelationships between various thermomechanical process parameters and the final hardness. To achieve this, it is necessary to generate a large number of combinations between process parameters and their corresponding hardness values.

An efficient strategy to obtain different combinations of thermomechanical parameters and their respective hardness values involves forging wedge-shaped samples between parallel flat dies [23]. As a result, the samples undergo varying degrees of deformation along their length. The combination of experimental procedures with numerical analyses yields a wide range of results without the need for a large number of tests. This is a cost-effective and agile method for evaluating the effects of strain and strain rate on the microstructural changes that occur during hot deformation.

This approach was employed to generate the dataset used for training and calibrating the artificial intelligence models. The wedge forging process is schematically represented in Fig. 2. The material investigated—a microalloyed steel with the chemical composition 0.18 C–1.38Mn–0.36Si–0.47Cr–0.03Mo–0.05Al–0.04Nb–0.01Ti—was forged at three different temperatures, after being heated in a resistive furnace to 1000 °C, 1100 °C, and 1200 °C. Conducting the tests at different temperatures aimed to increase the diversity of the dataset by incorporating multiple combinations of cooling conditions.

The selected range encompasses the typical operational window for heating during hot forging of microalloyed steels, enabling adequate capture of temperature effects on critical metallurgical phenomena—precipitate dissolution, dynamic recrystallization, and austenite grain refinement (or growth). This range defines the applicability domain of the developed models. Data points outside this domain require new experiments and recalibration to ensure reliable predictions. In industrial practice, operating outside this window is uncommon. Forging below 1000 °C tends to result in incomplete austenitization and demands higher loads, whereas temperatures above 1200 °C promote excessive

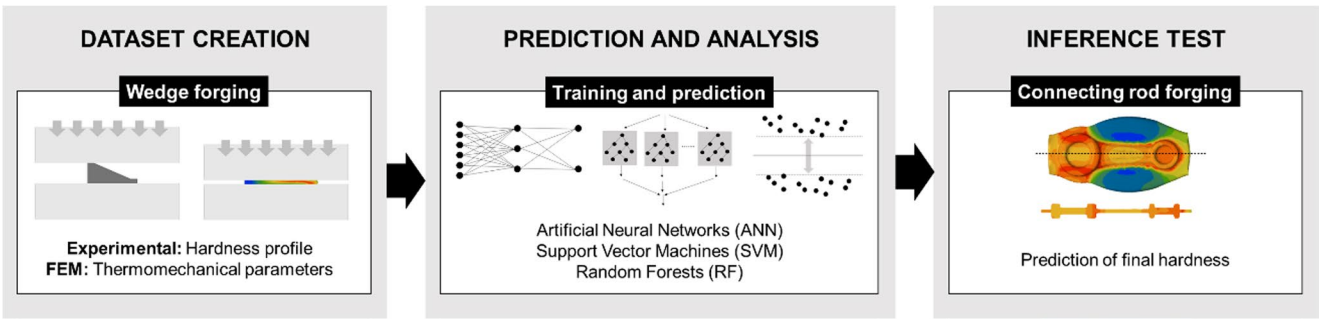

Fig. 1 Methodology employed

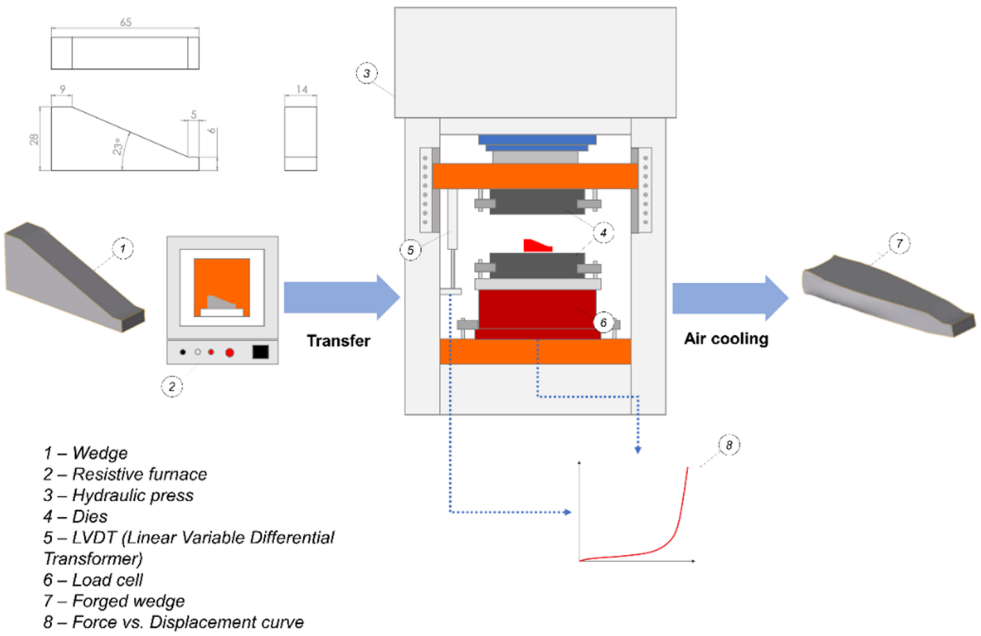

Fig. 2 Forging of wedge-shaped samples

Table 1 Detailed technical information on the wedge forging process

Material	0,18 C–1,38Mn–0,36Si–0,47Cr–0,03Mo–0,05Al–0,04Nb–0,01Ti
Material Temperature	1000 °C 1100 °C 1200 °C
Heating Method	Resistive furnace
Heating Time	30 min
Tool Temperature	150 °C
Lubrication	Graphite-based solution
Forging Equipment	Hydraulic press ($v = 5mm/s$)
Height Reduction	82%
Post-Forging Cooling	Continuous air cooling

austenitic grain growth. Such conditions do not favor the formation of refined microstructures or optimized mechanical properties.

Detailed technical information on the wedge forging process, including relevant thermomechanical and operational parameters, is presented in Table 1. Further details on the wedge tests are available in [20].

During the tests, force and displacement data were acquired. The press displacement was monitored using a Linear Variable Differential Transformer (LVDT), and the force was measured by a load cell mounted on the lower table of the press. The temperature of the samples was also monitored using a type K thermocouple inserted into the core of each specimen. The signals were acquired using an HBM Spider 8 device and processed with the Catman Express software.

The force, displacement, and temperature data were used as input variables in the inverse analysis of the experimental procedure, performed using the QForm UK software. Applying the finite element method, three-dimensional (3D) simulations of the forging tests were carried out. The experimental and numerical data were complemented by the corresponding hardness values obtained for each wedge. The forged samples were sectioned, and the microhardness profile was determined along the longitudinal section of each wedge. Each processing condition was executed in triplicate, and each microhardness profile was measured three times. Hardness measurements showed standard deviations $\leq$ 9 HV and a coefficient of variation $<$ 3%, with no statistically significant differences among replicates. The instrumentation met industrial precision criteria, with load cell and LVDT errors $\leq$ 0.5% FS, a Type K thermocouple (IEC 60584, Class 1) with $\pm$1.5 °C or 0.4%·T ($\approx \pm$2 °C in the range of interest), and an acquisition rate sufficient to avoid aliasing. Uncertainty propagation indicates typical prediction intervals $\leq \pm$10 HV (95%) within the process domain, consistent with quality control in hot forging.

The dataset consisted of information on hardness, equivalent strain (ϕ), strain rate ($\dot{\phi}$), initial temperature (ϑ_i)—corresponding to the heating temperature of the wedges—and final temperature (ϑ_f), referring to the temperature at the end of forging, immediately after the press opened. In total, 2,955 records were generated. The relationship between the thermomechanical parameters and the hardness values is presented in Fig. 3. It is observed that hardness varies nonlinearly with all analyzed parameters, highlighting the complexity of the interactions involved in the hot forging process.

Table 2 presents the descriptive statistics of the dataset. It is important to note that the artificial intelligence (AI) models developed will be capable of predicting the properties

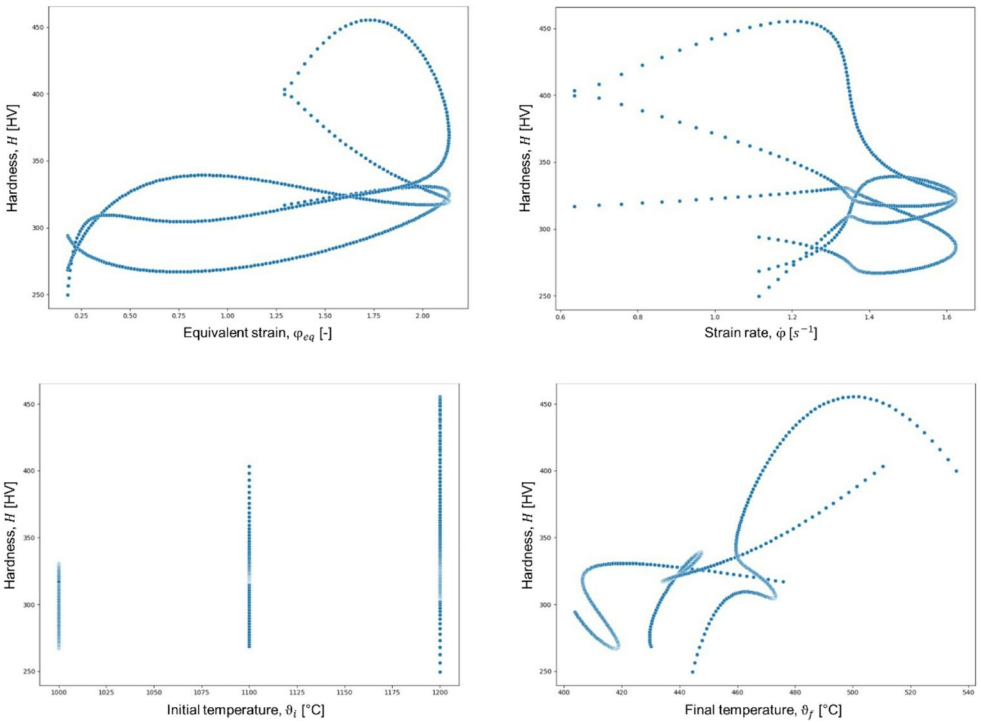

Fig. 3 Relationship between thermomechanical parameters and hardness values

Table 2 Descriptive statistics of the dataset

Variable	Mean	Standard Deviation	Minimum	Variable
Strain, ϕ [-]	1,549	0,657	0,180	2,135
Strain rate, $\dot{\phi}$ $\left[s^{-1}\right]$	1,395	0,178	0,636	1,623
Initial temperature, ϑ_i [° C]	1100 °C	81,718	1000	1200
Final temperature, ϑ_f [° C]	443,7 °C	26,695	43,9	535,8
Hardness, H [HV]	319,99	40,223	249,7	455,5

of forged components produced under thermomechanical routes whose parameters fall within the ranges represented in the dataset. Manufacturing processes involving different conditions will require the generation of new data to expand the model's coverage and, consequently, improve its accuracy and generalization capability.

The same reasoning applies to steel composition. Chemical variations modify the kinetics of metallurgical phenomena occurring during hot forging, thereby affecting the properties obtained after processing. In scenarios with batch-to-batch compositional fluctuations, this variable should be incorporated into model calibration/inputs. In the present approach, an effectively constant composition was assumed; therefore, explicit inclusion of these inputs is left as a future extension to enhance robustness across heats.

The database construction approach was designed to be efficient and cost-effective. However, the limited number of tests imposes constraints: it does not capture batch-to-batch variations and does not explicitly include cooling rate, microstructural descriptors, or physics-informed parameters, which limits transferability to new conditions. These limitations delineate the model's applicability domain and, for broader use, motivate the incorporation of such descriptors and calibration with additional data.

2.2 Prediction and analysis

The prediction and analysis stage was carried out through the development of Artificial Neural Network (ANN) models using the TensorFlow and Keras libraries, which provide a flexible and efficient platform for building, training, and validating deep learning models.

Prior to training the AI models, the dataset underwent a preprocessing step to ensure consistency and efficiency in the modeling process. Initially, all continuous attributes were normalized. Normalization is a fundamental step in data preprocessing, especially in problems involving machine learning algorithms. It consists of adjusting the variable values so that they fall within the same scale, preventing attributes with larger magnitudes from dominating the training process [24, 25].

In this study, Min-Max normalization was used, which transforms each input variable value into the [0, 1] range, according to the following formula:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Where X is the original value and X_{norm} is the normalized value. X_{min} and X_{max} correspond, respectively, to the minimum and maximum values observed for the variable. This procedure was implemented using the MinMaxScaler from the scikit-learn library, which automates the calculation and application of this transformation to all selected attributes.

The normalized dataset was randomly divided, with 70% allocated for training and 30% for model validation. The random split, repeated across different runs, aims to increase the statistical robustness of the results, reducing the risk that specific patterns or anomalous distributions in a single partition negatively affect the performance and generalization capability of the ANN.

To optimize the network weights, the Adam (Adaptive Moment Estimation) algorithm was chosen, widely recognized for its efficiency and stability, especially in problems involving complex datasets and potentially non-convex loss functions. Adam combines the advantages of stochastic gradient methods with individual adaptation of learning rates for each parameter, which favors faster and more reliable convergence [26].

The loss function chosen was the Mean Squared Error (MSE), due to its suitability for continuous regression problems, such as hardness prediction, and its property of more heavily penalizing larger errors, encouraging more precise model adjustments.

To avoid overfitting—a risk inherent in deep learning models, especially when the number of network parameters is high—an early stopping mechanism was implemented. This procedure monitors the model's performance on the validation set and automatically stops training if no improvement is observed after a predefined number of consecutive epochs, preventing the model from excessively fitting the noise in the training data. The maximum number of epochs was set at 300, as a prudent limit that balances the need for adequate convergence of the ANN without incurring excessive computational cost or overfitting risk.

The predictive performance of the network was evaluated using three complementary metrics: the Mean Absolute Error (MAE), which provides a direct measure of the average absolute errors made by the model; the Mean Squared Error (MSE), which quantifies the average of squared errors, giving greater weight to more significant deviations; and the Coefficient of Determination (R^2), which expresses the proportion of the total variance in the data explained by the model, serving as a global indicator of fit quality and predictive capability.

To assess the robustness and predictive capability of the developed Artificial Neural Network (ANN), two other models widely used in regression tasks were also implemented and tested: Support Vector Machines (SVM) and Random Forest (RF).

Support Vector Machines (SVM) are models based on margin maximization theory, originally designed for classification problems but adaptable to regression through the technique known as Support Vector Regression (SVR). SVM seeks to find a function that approximates the data within an acceptable tolerance margin while minimizing model complexity [27, 28]. In this work, the RBF (Radial Basis Function) kernel was used, a function widely employed for its ability to model nonlinear relationships, flexibly adapting to the characteristics of the dataset [29]. The SVM model was developed with prior data normalization and hyperparameter tuning through cross-validation, aiming to optimize the balance between bias and variance [30, 31].

Random Forest (RF), in turn, is a model based on the principle of ensemble learning, which consists of combining multiple decision trees to produce a more robust result that is less susceptible to overfitting [32–34]. In this study, the RF was implemented with 100 decision trees, configured with depth and minimum sample parameters adjusted through cross-validation, aiming to optimize the model's generalization.

2.3 Inference tests

To broaden the evaluation of the predictive and generalization capability of the developed model, inference tests were conducted using data not included in the original training and validation set of the Artificial Neural Network. This approach aims to simulate real-world application conditions, in which the model is exposed to new scenarios different from those previously known.

In the inference tests, the model was applied to predict the hardness of the investigated material after the hot forging of a connecting rod. The experimental process, schematically represented in Fig. 4, consisted of heating the material to 1200 °C, followed by its transfer to a hydraulic press, where the forging operation was performed. After the part was extracted, cooling was carried out continuously in air until room temperature.

Detailed technical information on the connecting rod forging process, including relevant thermomechanical and operational parameters, is presented in Table 3.

After forging, the connecting rods were sectioned, and the microhardness profile was determined along the longitudinal section, as illustrated in Fig. 5.

This methodology enabled a direct comparison between the predictions generated by the AI model and the actual hardness values obtained experimentally, constituting an essential step to validate the effectiveness and practical

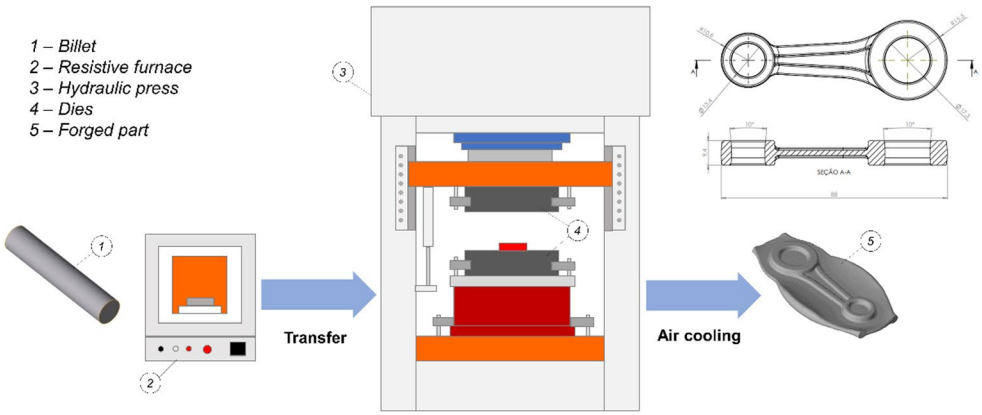

Fig. 4 Forging of the connecting rod

Table 3 Detailed technical information on the connecting rod forging process

Material	0,18 C–1,38Mn–0,36Si–0,47Cr–0,03Mo–0,05Al–0,04Nb–0,01Ti
Material temperature	1200 °C
Heating Method	Resistive furnace
Heating Time	30 min
Tool Temperature	150 °C
Lubrication	Graphite-based solution
Forging Equipment	Hydraulic press ($v = 5mm/s$)
Post-Forging Cooling	Contínuous air cooling

applicability of the developed model outside the specific training domain.

3 Results and discussion

3.1 Database analysis

Figure 5 presents the heat map of coefficients of determination (R^2), highlighting the sensitivity of hardness to the thermomechanical parameters of hot forging. The results indicate that the final forging temperature (ϑ_f) has the highest univariate explanatory power for hardness ($R^2 \approx 0.49$), followed by the initial temperature (ϑ_i) (≈ 0.24) and the equivalent strain (≈ 0.22); the strain rate shows the lowest average association (≈ 0.12). The univariate screening confirms this pattern. When the four inputs are

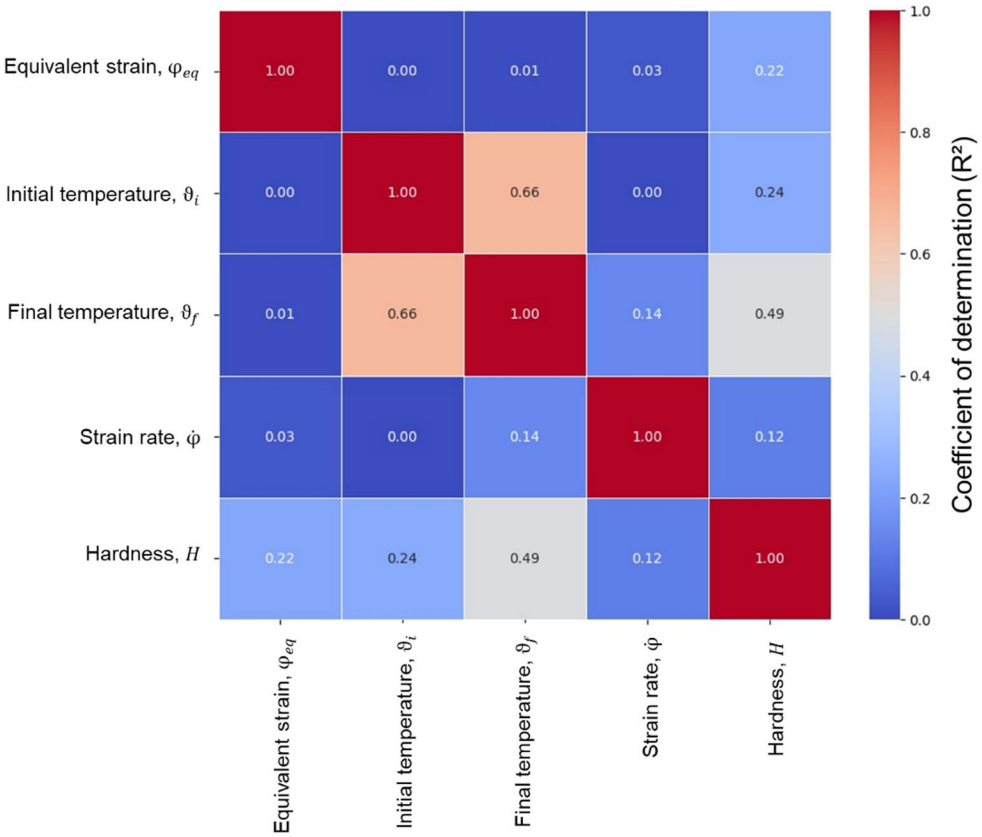

Fig. 5 Heat map of coefficients of determination (R^2), highlighting the sensitivity of hardness to the thermomechanical parameters of the hot forging process

considered simultaneously in a linear regression, the global fit explains about 70% of the variance in hardness, and the standardized coefficients indicate positive predominance of ϑ_f ($\beta \approx +0.55$) and strain ($\beta \approx +0.47$), with a negative effect of strain rate ($\beta \approx -0.22$) and residual influence of ϑ_i ($\beta \approx +0.04$).

Stratification by initial temperature (Fig. 6) reveals context dependence. At $\vartheta_i = 1000°$ C, the model explains 93.8% of the variance and hardness increases primarily with strain ($\beta z \approx +0.93$), while strain rate and ϑ_f act to reduce it ($\beta z \approx -0.76$ and -0.32, respectively). At $\vartheta_i = 1100°$ C, in contrast, ϑ_f becomes the principal positive determinant ($\beta z \approx +1.16$) and strain rate also appears positive ($\beta z \approx +0.43$), suggesting that, within the sampled intermediate thermal window, the process combinations favored a regime in which higher ϑ_f did not degrade hardness on average; strain exhibited a near-zero mean effect ($\beta z \approx +0.02$). At $\vartheta_i = 1200°$ C, strain again plays a relevant positive role ($\beta z \approx +0.64$), ϑ_f retains a moderate positive impact ($\beta z \approx +0.34$), and strain rate returns to a negative effect ($\beta z \approx -0.38$).

From a metallurgical standpoint, the findings are consistent with hot working of microalloyed bainitic steels. Hardness is strongly influenced by refinement of the bainitic microstructure, which depends on the prior austenite grain size and deformation-induced recrystallization. Because recrystallization is a thermally activated process, ϑ_i conditions the austenitization state and the prior grain size, while increasing strain promotes refinement. Strain rate tends to reduce hardness by shortening the time available for recrystallization. Despite these principles, the data show that the magnitude and even the sign of each parameter's effect vary with the processing condition, reinforcing the need for conditional (multivariate) analyses to guide operational adjustments based on evidence.

3.2 Performance of the artificial neural network

To maximize predictive capability, the ANN architecture was defined via hyperparameter optimization, systematically evaluating combinations of the number of hidden layers, the number of neurons per layer, and the activation function. Controlled searches (grid and random) were conducted with cross-validation, a dedicated validation set, and early stopping to mitigate overfitting, using MAE, MSE, and R^2 as selection criteria. The final topology (depth and width), as well as the activation function, were chosen based on the best validated average performance among the tested configurations. The resulting optimized architecture comprises three hidden layers with 64, 128, and 256 neurons, respectively (Fig. 7). ReLU (Rectified Linear Unit) was used in all hidden layers due to its computational simplicity and its ability to alleviate vanishing-gradient issues, promoting efficient and stable training.

The choice of this number of neurons and the multilayer structure reflects a balance between model complexity—necessary to capture the nonlinear interrelationships between thermomechanical parameters and hardness—and the need to avoid overfitting, especially considering the size of the available dataset.

This architecture proved to be suitable for the proposed problem, allowing the model to efficiently learn the underlying patterns in the data and to perform well in both validation and inference stages.

The input layer of the Artificial Neural Network consists of the four main thermomechanical parameters: equivalent strain, strain rate, initial temperature, and final forging temperature. The selection of thermomechanical input parameters was guided by the intended application of the models. By limiting the feature vector to four variables, we aim to facilitate integration with consistently available data from

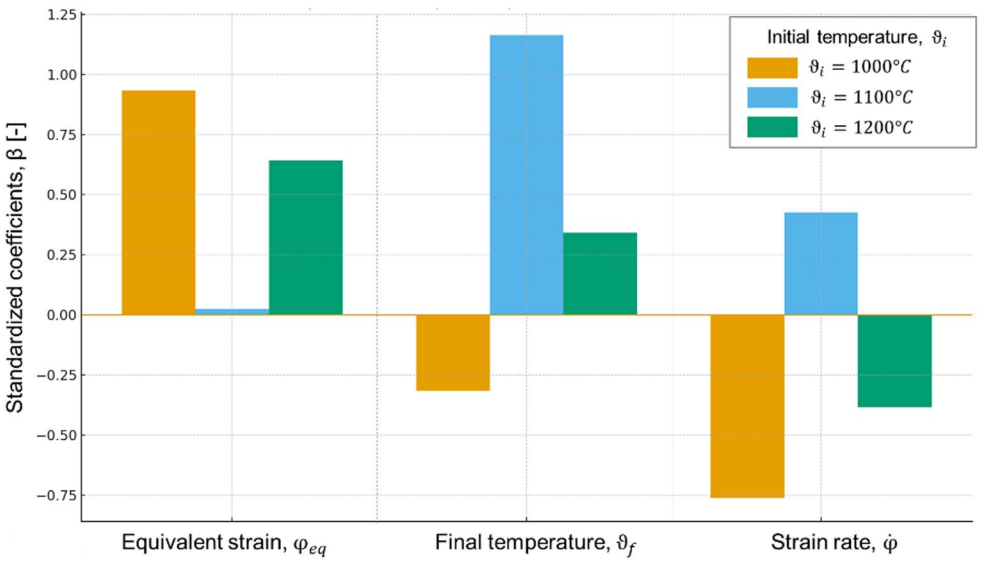

Fig. 6 Standardized coefficients (β) from the multivariate linear regression of hardness as a function of strain, final temperature, and strain rate, stratified by initial temperature

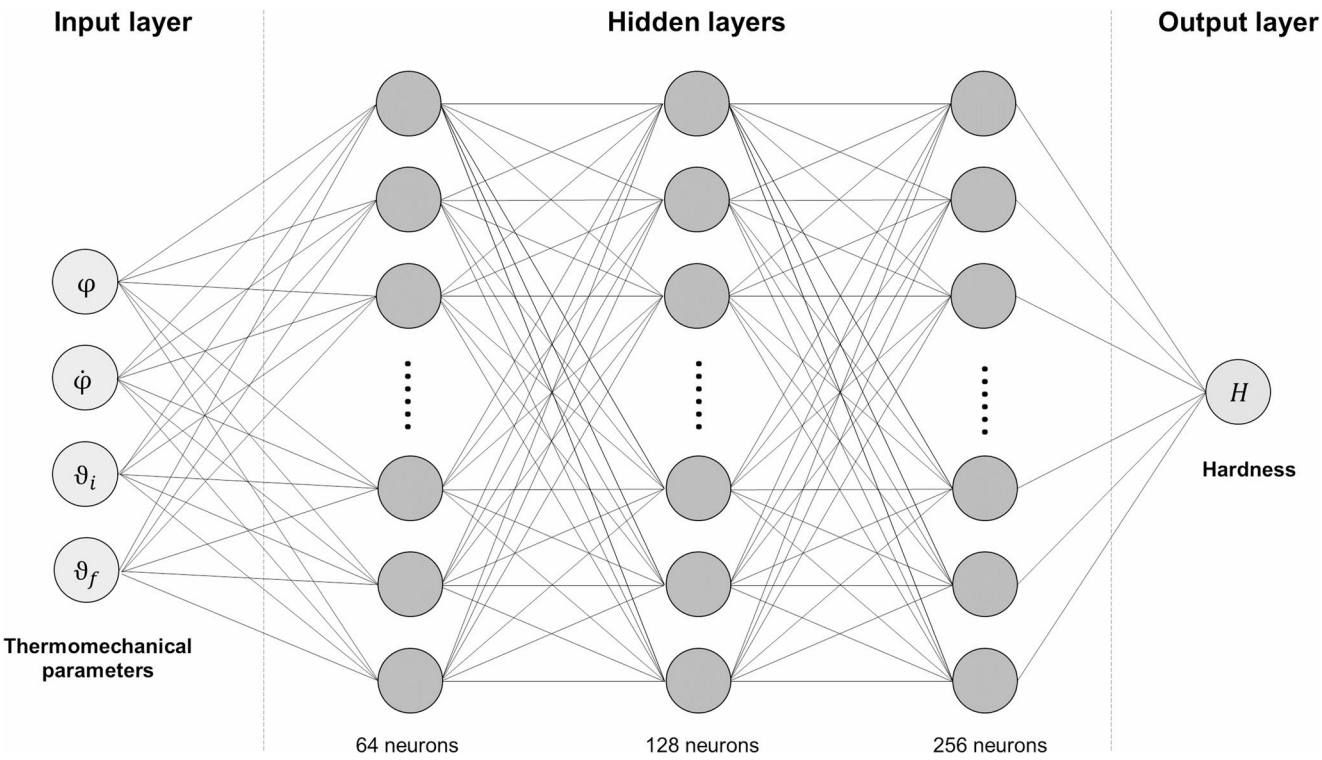

Fig. 7 Architecture of the proposed artificial neural network

numerical simulations and/or in-process measurements. Including more complex information—such as full cooling-rate profiles and metallurgical variables—besides not being uniformly available, could reduce model robustness and introduce information leakage between simulation and measurement stages.

The output layer consists of a single neuron, whose function is to predict the corresponding hardness value for each specific combination of input parameters. This configuration characterizes a typical regression model, suitable for predicting continuous variables.

The complexity of the network topology plays a fundamental role in its performance. The complexity study showed that shallower architectures exhibited higher bias (systematically higher errors), whereas deeper architectures increased cross-fold variance without robust gains in accuracy. Experiments with L2 regularization and dropout did not provide consistent additional benefits; therefore, we adopted the more parsimonious configuration with early stopping as the primary mechanism for overfitting mitigation. However, this significant increase in capacity also implies the need for greater computational resources, including processing time, hardware power, and often larger volumes of input data to ensure effective training [34].

In developing the model, an appropriate balance was sought between the depth and width of the network, in order to optimize predictive performance without excessively compromising computational efficiency. This approach aims to maximize the model's effectiveness, ensuring robust and accurate learning while avoiding unnecessary complexity that could lead to overfitting or the practical infeasibility of its application.

The model's convergence analysis indicated a consistent reduction in Mean Absolute Error (MAE) and Mean Squared Error (MSE) values throughout the training process, as illustrated in Fig. 6. Initially, a progressive reduction in MAE and MSE values was observed as the model adjusted its parameters to fit the patterns present in the training data. This behavior highlights the network's learning process, which seeks to minimize discrepancies between predictions and actual values. As the iterations progressed, the errors converged to relatively low values, indicating that the ANN was learning effectively and approaching an optimal solution. Subsequently, the graphs in Fig. 8 show a stabilization in the error metrics, with minimal variation over the final epochs. This pattern suggests that the model reached a saturation point, where further adjustments no longer resulted in significant performance improvements, characterizing the convergence of the training process and confirming the adequacy of the adopted architecture and hyperparameters.

The results obtained for the developed model demonstrate highly satisfactory predictive performance. After 300 iterations, the Mean Absolute Error was 1.3518 HV and the Mean Squared Error was 8.8448 HV^2. The MAE value of

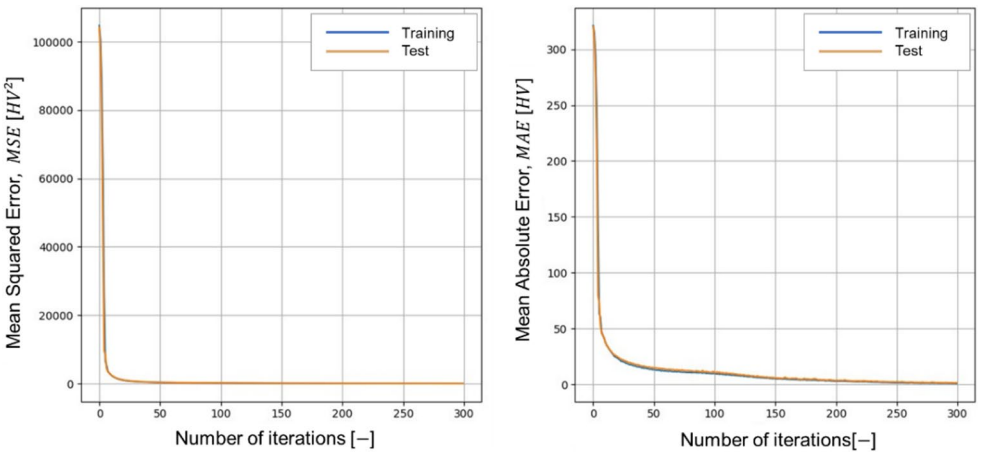

Fig. 8 Model convergence analysis through the evolution of mean squared error (MSE) and mean absolute error (MAE) as a function of the number of iterations

1.3518 HV indicates that, on average, the predictions of the Artificial Neural Network differ by less than 2 Vickers hardness units from the actual values. This level of error is extremely low, especially considering the typical variability of hot forging processes and the complexity of the nonlinear relationships between thermomechanical parameters and hardness. Complementarily, the MSE value of 8.8448 HV^2 confirms the model's accuracy, showing that significant deviations were rare and of low magnitude. The reduced MSE reinforces the model's ability to faithfully capture the underlying patterns in the data, minimizing both systematic errors and random fluctuations.

These indicators attest to the effectiveness of the defined ANN architecture, as well as the adequacy of the dataset and the preprocessing and training procedures adopted. Moreover, the low error values suggest that the model has a high generalization capability.

The value obtained for the Coefficient of Determination (R^2) was 0.994, indicating that the developed model is capable of explaining approximately 99.4% of the variance present in the hardness data, which demonstrates excellent predictive capability. This result confirms that the Artificial Neural Network was highly effective in modeling the complex nonlinear relationships between thermomechanical parameters and the resulting hardness of the forged material.

The graphical analysis of the relationship between predicted and actual hardness values (Fig. 9) reinforces this conclusion. It is observed that most points are strongly concentrated along the reference line (red dashed line), which represents the ideal match between prediction and reality. The distribution close to this line, with small dispersions, indicates that the model exhibits low systematic error and good generalization, even in the face of natural process variations.

The presence of a few slightly scattered points suggests occasional occurrences of higher error, possibly associated with regions of the parameter space with lower data density

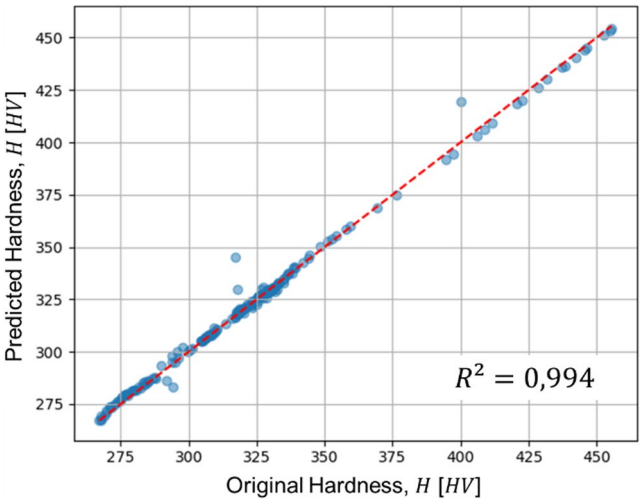

Fig. 9 Hardness prediction as a function of the corresponding actual value

in the training set or with inherent variabilities in the forging process. However, these deviations are few and of small magnitude, as already evidenced by the low MAE and MSE values.

Figure 10 shows the analysis of the error distribution of the Artificial Neural Network (ANN) and reveals an approximately symmetric behavior centered around zero, which is characteristic of a well-fitted model. The high concentration of errors near zero indicates that, in most predictions, the discrepancies between predicted and actual values were very small, confirming the model's accuracy and consistency.

The presence of a smoothed curve resembling a normal distribution further reinforces the model's suitability for the problem, suggesting that there are no systematic tendencies toward overestimation or underestimation in the predictions. Moreover, the occurrence of a few outliers, with more significant errors located at the extremes, is natural in industrial processes—especially when considering the intrinsic variability associated with hot forging and possible regions of lower density in the data space.

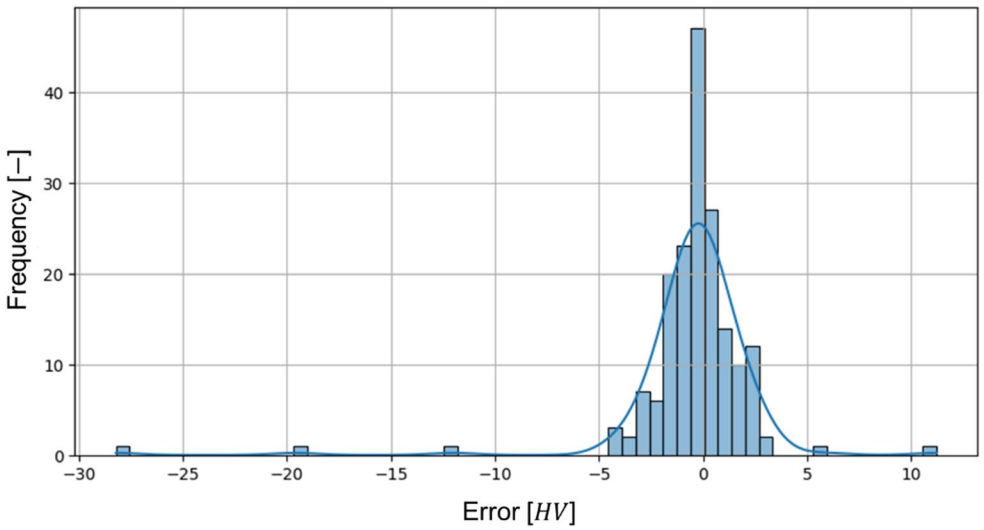

Fig. 10 Error distribution analysis of the proposed ANN

Table 4 Number of values found by error magnitude range

Error magnitude [HV]	Number of values found
<−10	3
−10 a −5	0
−5 a −2	3
−2 a 0	70
0 a 2	84
2 a 5	16
5 a 10	2
>10	0

The errors were classified into value ranges to quantify their frequency across different intervals, allowing for a more granular analysis of the model's behavior. The ranges were defined as: < −10, −10 to −5, −5 to −2, −2 to 0, 0 to 2, 2 to 5, 5 to 10, and > 10. Table 4 presents the distribution of prediction errors by interval.

The low dispersion of errors and reduced asymmetry highlight the robustness of the training process and the ability of the defined architecture to generalize well across different combinations of thermomechanical parameters. These results complement the evidence provided by the evaluation metrics and the coefficient of determination (R^2), confirming the efficiency and reliability of the developed model.

3.3 Performance comparison with other machine learning models

The evaluation of alternative models revealed satisfactory performance, although inferior to that of the Artificial Neural Network, as shown in Fig. 11. The Support Vector Machine (SVM) presented a Mean Absolute Error (MAE) of 2.891 HV and a Mean Squared Error (MSE) of 27.287 HV^2, with a Coefficient of Determination (R^2) of 0.984. These results indicate good predictive capability, but with relatively greater error dispersion, especially considering the MSE value, which reflects higher sensitivity to more significant deviations.

The Random Forest (RF) model, in turn, achieved an MAE of 1.654 HV and an MSE of 34.929 HV^2, with an R^2 of 0.980. Although the mean absolute error was close to that of the Neural Network, the high MSE suggests a greater occurrence of extreme errors, which negatively impact the model's overall accuracy. Furthermore, the R^2, although high, was the lowest among the three models, indicating a reduced ability to explain the variability present in the data.

Although both SVM and RF proved to be suitable models for predicting hardness based on thermomechanical parameters, the ANN stood out as the most efficient and robust option, combining low error margins with high predictive

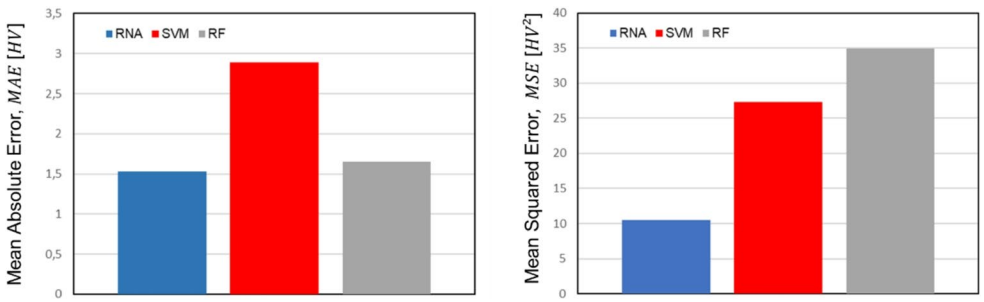

Fig. 11 Performance comparison of the investigated models in terms of mean squared error (MSE) and mean absolute error (MAE)

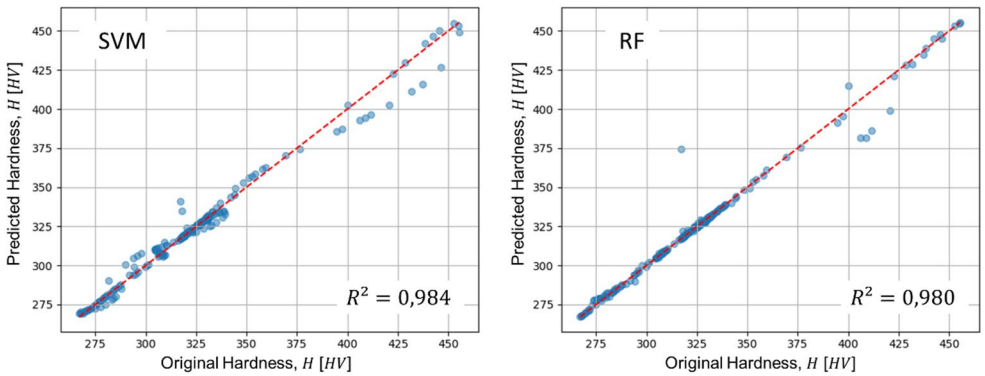

Fig. 12 Hardness prediction as a function of the corresponding actual value for the SVM and RF models

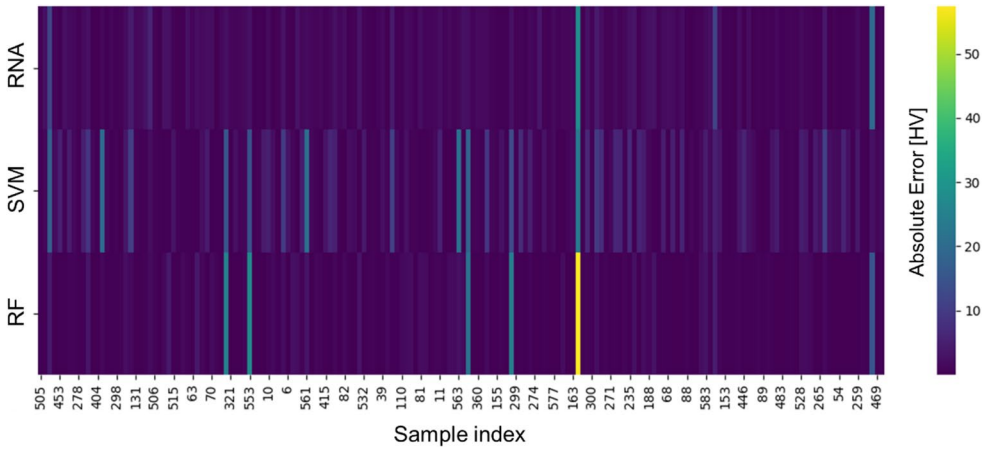

Fig. 13 Heatmap of absolute error per sample

fidelity, and is therefore the most recommended for applications requiring high levels of reliability and precision.

The analysis of the relationship between predicted and actual values for the Support Vector Machine (SVM) and Random Forest (RF) models (Fig. 12) shows that both methods exhibited good predictive performance, with a strong alignment of points along the identity line (red dashed line), which represents the ideal match between prediction and reality. However, it is observed that the SVM model shows greater dispersion of points, especially in the upper hardness range, from approximately 375 HV. In this region, there is a tendency to underestimate higher hardness values, with predictions deviating further from the ideal line. This behavior contributes to the higher mean and squared errors observed for the SVM, reflecting a lower generalization capability compared to the other models, particularly at the extremes of hardness.

On the other hand, the Random Forest (RF) model showed a more concentrated distribution of points along the reference line, with smaller systematic deviations. Although some isolated outliers exist, mainly in the intermediate and upper hardness ranges, the overall dispersion is lower than that observed in the SVM, which is reflected in superior performance in terms of Mean Absolute Error (MAE). Nevertheless, the greater sensitivity to squared errors, as evidenced by the MSE, indicates that RF had difficulty controlling more significant errors, possibly related to specific combinations of thermomechanical parameters with low representation in the dataset.

The absolute error heatmap per sample (Fig. 13) provides a detailed comparative view of the error distribution across all predictions made by the three models. It is observed that the ANN generally presents a more homogeneous distribution of errors, with the vast majority of samples associated with low absolute error values, indicated by darker shades (deep purple). Only a few isolated points show moderate errors, which reinforces the robustness and predictive stability of the model across different regions of the input space.

The SVM model, in turn, shows a higher frequency of moderate errors across several samples, with a more irregular and dispersed distribution. The lighter shades indicate the presence of multiple cases with higher errors, which supports the previous observation that SVM exhibited greater sensitivity to deviations and difficulty in generalizing adequately, especially in more complex regions of the parameter space.

The Random Forest (RF) model shows intermediate behavior: although most samples are also associated with low errors, there is an occurrence of extreme errors, highlighted by the presence of an intense yellow band corresponding to an error greater than 50 HV in a specific sample. This isolated occurrence, although singular, significantly

increases the model's mean squared error (MSE), indicating a vulnerability to outliers or to cases that are underrepresented in the training set.

Overall, the heatmap reinforces the superiority of the ANN in the hardness prediction task, showing lower variability in errors and greater consistency compared to the other models. Furthermore, it highlights that although both SVM and RF are capable of producing acceptable predictions, each presents specific weaknesses: SVM with greater error dispersion, and RF with sensitivity to extreme outliers.

Computational complexity is a crucial factor in selecting the most appropriate model, especially when considering practical application in industrial or research environments. In the present study, the Artificial Neural Network (ANN) required a longer training time, totaling approximately 1 min and 25 s. This longer duration is directly related to the complexity of the adopted architecture, with multiple layers and a significant number of neurons, as well as the need for intensive iterative adjustments to optimize weights and minimize the loss function.

However, this higher computational cost is offset by the model's high accuracy, which showed the best predictive performance among the tested approaches, both in terms of low error margins and high explanatory power. Thus, the ANN is fully justified in contexts where precision is critical and training time is not a limiting factor.

On the other hand, the Support Vector Machine (SVM) and Random Forest (RF) models showed significantly shorter training times, approximately 10 s and 30 s, respectively. This computational efficiency makes these techniques advantageous in situations where rapid model deployment is essential or when computational resources are limited.

Nevertheless, this speed comes with lower performance compared to the ANN, with higher mean and squared errors, as well as reduced generalization capability. Therefore, the choice between models involves a classic trade-off between computational time and predictive performance: while the ANN offers greater accuracy at a higher computational cost, SVM and RF provide faster solutions, albeit less precise.

This balance must be carefully considered according to the specific demands of the application: if the goal is maximum precision and robustness, the ANN stands out; if, on the other hand, rapid implementation with acceptable accuracy is required, SVM and RF may be viable alternatives.

3.4 Inference tests

To validate the generalization capability and robustness of the developed models, inference tests were conducted using data that were not part of the training or prior validation. The performance of the models was analyzed through direct comparison of the predictions with the actual hardness values of forged connecting rods, as well as by evaluating the distribution of deviations from the expected values.

Figure 14 presents the comparison between the first 10 predicted values and the actual values. It is observed that, in general, all models—Artificial Neural Network (ANN), Support Vector Machine (SVM), and Random Forest (RF)—were able to approximate the actual values with reasonable accuracy. However, it is noted that the ANN maintains greater regularity, with predictions systematically closer to the actual values compared to the other methods, while SVM and RF show more noticeable deviations in some samples, especially in cases with higher hardness values.

The analysis of deviation distribution was carried out using two complementary methods: the boxplot (Fig. 15) and the violin plot (Fig. 16). The boxplot shows that the ANN has a smaller dispersion of deviations compared to the other models, with most errors concentrated near zero and few low-magnitude outliers. In contrast, both SVM and RF exhibit wider distributions, with a greater number of outliers and more pronounced deviations, especially in the case of SVM, where systematically higher errors are observed in some cases.

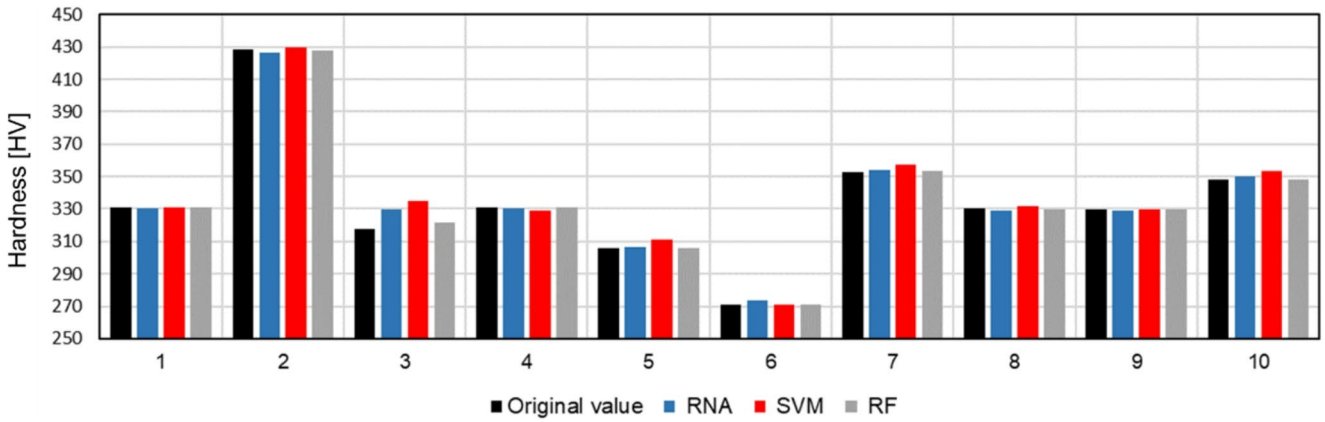

Fig. 14 Comparison between predicted and actual values for the analyzed models

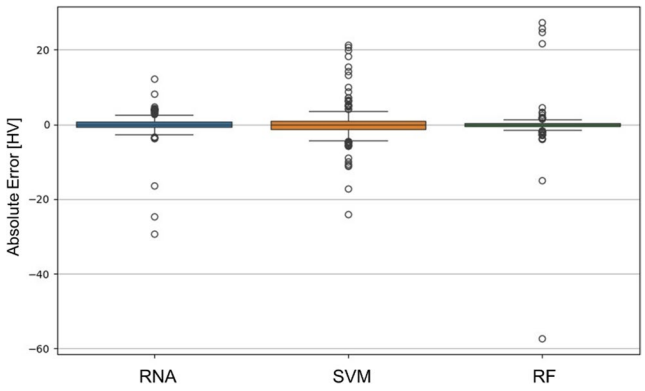

Fig. 15 Deviation distribution analysis using boxplot

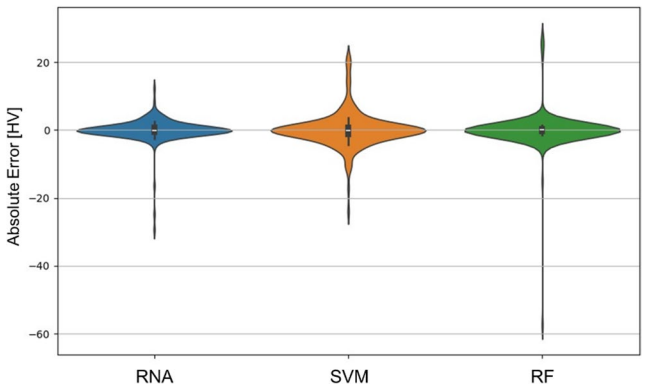

Fig. 16 Analysis of the deviation distribution using the violin plot

The violin plot complements this analysis by illustrating the density of the deviation distributions for each model. The ANN shows a narrower and more concentrated distribution, reinforcing its ability to produce predictions with lower variability and higher accuracy. On the other hand, SVM and RF exhibit wider distributions, indicating greater variability in errors and a higher tendency toward significant deviations, especially at the extremes, with isolated cases of large errors.

These results consistently demonstrate, in line with previous analyses, that the Artificial Neural Network offers the best predictive performance, combining low variability, reduced incidence of outliers, and smaller average deviations from the actual values. While the SVM is efficient, it proves to be more sensitive to errors in certain scenarios, whereas the Random Forest tends to exhibit more dispersed errors and, occasionally, extreme deviations, as evidenced by the identified outliers.

4 Conclusions

This study investigated the feasibility and effectiveness of applying artificial intelligence techniques to predict the hardness of forged bainitic steels, based on thermomechanical parameters obtained from experimental data and numerical simulations.

The Artificial Neural Network (ANN), with an architecture composed of three hidden layers (64, 128, and 256 neurons), showed the best predictive performance, achieving a Mean Absolute Error (MAE) of 1.3518 HV, a Mean Squared Error (MSE) of 8.8448 HV², and a Coefficient of Determination (R^2) of 0.994.

The alternative models, Support Vector Machine (SVM) and Random Forest (RF), although advantageous in terms of shorter training times, showed inferior performance, with higher mean and squared errors and lower explanatory power of the data.

The inference tests, conducted with data from the forging of connecting rods, validated the robustness and practical applicability of the developed model, demonstrating that the ANN maintained superior performance even when exposed to scenarios different from those used for training.

The results obtained in this study represent a relevant contribution to the intelligent design and control of bainitic steel forging processes, demonstrating the potential of machine learning techniques as effective tools to support decision-making and dynamic adjustment of industrial processes. The developed model can assist engineers in defining the ideal thermomechanical parameters during process design, providing a robust predictive basis for optimizing operational conditions. Furthermore, its direct application in real-time process control enables early identification of production inconsistencies and agile implementation of corrections, contributing to continuous improvement in quality and efficiency.

In this context, it is also worth highlighting the possibility of expanding the dataset by incorporating real production data, which will allow the models to become even more accurate and better adapted to specific operating conditions, promoting continuous evolution in the predictive capability and robustness of AI-based solutions in the metal forming sector.

Acknowledgements The authors would like to thank CNPq (National Council for Scientific and Technological Development) and CAPES (Coordination for the Improvement of Higher Education Personnel) for their financial support, as well as the Metalforming Laboratory (LdTM) for providing technical assistance.

Declarations

Ethical approval Not applicable.

Conflict of interest The authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest, or non-financial interest in the subject matter or materials discussed in this manuscript.

References

1. Hui W, Zhang Y, Zhao X, Xiao N, Hu F (2016) High cycle fatigue behavior of V-microalloyed medium carbon steels: a comparison between bainitic and ferritic-pearlitic microstructures. Int J Fatigue 91:232–241.
2. Buchmayr B (2016) Critical assessment 22: bainitic forging steels. Mater Sci Technol. https://doi.org/10.1080/02670836.2015.1114272
3. Gramlich A, Lange R, Zitz U, Büßenschütt K (2022) Air-hardening die-forged con-rods—achievable mechanical properties of bainitic and martensitic concepts. Metals. https://doi.org/10.3390/met12010097
4. Wirths W, Wagener R, Bleck W, Melz T (2014) Bainitic forging steels for cyclic loading. Advanced Materials Research 922:813–818
5. Park H-S, Anh T-V (2011) Optimization of bending sequence in roll forming using neural network and genetic algorithm. J Mech Sci Technol 25:2127–2136
6. Hartmann C, Opritescu D, Volk W (2019) An artificial neural network approach for tool path generation in incremental sheet metal free-forming. J Intell Manuf 30(2):757–770
7. Di Lorenzo R, Ingarao G, Micari F (2006) On the use of artificial intelligence tools for fracture forecast in cold forming operations. J Mater Process Technol 177(1–3):315–318
8. Stoerkle DD, Seim P, Thyssen L, Kuhlenkoetter B (2016) Machine learning in incremental sheet forming. Proceedings of ISR 2016: 47st International Symposium on Robotics, VDE, p 1–7
9. Yatkin MA, Kõrgesaar M (2024) Machine learning enabled identification of sheet metal localization. Int J Solids Struct 288:112592
10. Di S, Thomson PF (1997) Neural network approach for prediction of wrinkling limit in square metal sheet under diagonal tension. J Test Eval 25(1):74–81
11. Spathopoulos SC, Stavroulakis GE (2020) Springback prediction in sheet metal Forming, based on finite element analysis and artificial neural network approach. Applied Mechanics
12. Jaremenko C, Ravikumar N, Affronti E, Merklein M, Maier A (2019) Determination of forming limits in sheet metal forming using deep learning. Materials 12(7):1051
13. Derogar A, Djavanroodi F (2011) Artificial neural network modeling of forming limit diagram. Mater Manuf Processes 26(11):1415–1422. https://doi.org/10.1080/10426914.2010.544818
14. Mozafar M, Bostanabad R, Chen W, Ehmann K, Cao J, Bessa M (2019) Deep learning predicts path-dependent plasticity. Proc Natl Acad Sci 116(52):26414–26420
15. Rosiak A, Schmeling M, Marcelino R et al (2024) Machine learning applied to predict the flow curve of steel alloys. Int J Adv Manuf Technol 134:5481–5492. https://doi.org/10.1007/s00170-024-14472-0
16. Chen L et al (2021) Hybrid deep learning model for simultaneous prediction of hardness and grain size in titanium alloys. Materials & Design
17. Wu D, Jennings C, Terpenny J, Gao RX, Kumara SA (2017) Comparative study on machine learning algorithms for smart manufacturing: tool wear prediction using random forests. J Manuf Sci Eng. https://doi.org/10.1115/1.4036350
18. Mrzygłód B, Hawryluk M, Janik M, Olejarczyk-Wożeńska I (2020) Sensitivity analysis of the artificial neural networks in a system for durability prediction of forging tools to forgings made of C45 steel. Int J Adv Manuf Technol. https://doi.org/10.1007/s00170-020-05641-y
19. Wang J, Li Y, Zhao R, Gao R (2020) Physics guided neural network for machining tool wear prediction. J Manuf Syst. https://doi.org/10.1016/j.jmsy.2020.09.005
20. ROSIAK A, DE LUCCA DE COSTA L, GOMES DOS SANTOS T, SCHAEFFER L (2025) Analysis of hot forging of a bainitic steel hardened by continuous cooling using wedge test. J Mater Eng Perform. https://doi.org/10.1007/s11665-025-11091-y
21. Cao J et al (2024) Artificial intelligence in metal forming. CIRP Ann 73(2):561–587
22. Pelt DM, Sethian JA (2018) A mixed-scale dense convolutional neural network for image analysis. Proc Natl Acad Sci U S A 115:254–259
23. Dieter GE, Kuhn HA, Semiatin SL (2003) Handbook of workability and process design. ASM International
24. Agrawal A, Deshpande PD (2014) Exploration of data science techniques to predict fatigue strength of steel from composition and processing parameters. Integr Mater Manuf Innov 3:90–108
25. Wang Y, Wu X, Li X, Xie Z, Liu R, Liu W, Zhang Y, Xu Y, Liu C (2020) Prediction and analysis of tensile properties of austenitic stainless steel using artificial neural network. Metals 10:234. https://doi.org/10.3390/met10020234
26. Okewu E, Misra S, Fernandez-Sanz L (2020) Parameter tuning using adaptive moment estimation in deep learning neural networks. In: Computational science and its applications – ICCSA 2020. Conference paper, p 261–272. Publicado online em 30 set
27. Wang L, Mu Z, Guo H (2006) Application of support vector machine in the prediction of mechanical property of steel materials. J Univ Sci Technol Beijing Mineral Metall Mater 13(6):512–515. https://doi.org/10.1016/S1005-8850(06)60104-7
28. Lu W-C, Ji X-B, Li M-J, Liu L, Yue B-H, Zhang L-M (2013) Using support vector machine for materials design. Adv Manuf 1:151–159
29. Yan C, Shen X, Guo F, Zhao S, Zhang L (2019) A novel model modification method for support vector regression based on radial basis functions. Struct Multidiscip Optim 60:983–997
30. Brown T et al (2020) Language models are few-shot learners. Adv Neural Inf Process Syst 33:1877–1901
31. Smith S et al (2022) Using Deepspeed and megatron to train megatron-turing Nlg 530b, a large-scale generative language model arXiv preprint arXiv:2201.11990.
32. Kwak S, Kim J, Ding H, Xu X, Chen R, Guo J, Fu H (2022) Machine learning prediction of the mechanical properties of γ-TiAl alloys produced using random forest regression model. J Mater Res Technol 18:520–530
33. Schubert M, Luković M, Christen H (2020) Prediction of mechanical properties of wood fiber insulation boards as a function of machine and process parameters by random forest. Wood Sci Technol 54:703–713
34. Akbari P, Zamani M, Mostafaei A (2024) Machine learning prediction of mechanical properties in metal additive manufacturing. Addit Manuf 91:104320. https://doi.org/10.1016/j.addma.2024.104320