CYBERNETICS AND PHYSICS, VOL. 14, NO. 3, 2025, 255–266

CONSTITUTIVE EQUATIONS BASED ON NEURAL NETWORK FOR HSLA-65 WITH SHAP

Ruimin Ma
St. Petersburg State University
Russia
mrm19990206@gmail.com

Xinran Hao
St. Petersburg State University
Russia
haoxinran@yandex.ru

Dongfang Qi
St. Petersburg State University
Russia
tsidongfang@gmail.com

Yuyi Zhang
St. Petersburg State University
Russia
lesliezhang0825@gmail.com

Shixiang Zhao
St. Petersburg State University
Russia
zhaoshixiang@yandex.ru

Yuri V. Petrov
Institute for Problems in Mechanical Engineering of the Russian Academy of Sciences
Russia
y.v.petrov@spbu.ru

Article history:
Received 03.10.2025, Accepted 20.11.2025

Abstract

Constitutive models are fundamental for analyzing the mechanical behavior of structural steels. Classical phenomenological or physics-based formulations provide explicit equations but often struggle to capture strong nonlinearity. Consequently, neural-network-based (NN) constitutive models have attracted growing attention for their ability to approximate complex response surfaces efficiently. In this work, we develop data-driven NN models for HSLA-65 steel across a wide range of temperatures and strain rates. We train a Multilayer Perceptron (MLP) and a Recurrent Neural Network (RNN), and evaluate generalization with five different metrics. To further interpret the models, we apply SHAP (SHapley Additive exPlanations) to quantify input-feature contributions. The results show that temperature (thermal softening) and strain rate (rate sensitivity) are dominant drivers, while plastic strain provides a positive cumulative contribution (work hardening) whose global weight is modest in the MLP but approaches that of temperature and strain rate in the RNN. Compared with the MLP, the RNN achieves higher accuracy and demonstrates superior robustness and generalization on unseen data. Overall, the study advances NN-based constitutive modeling by combining high-fidelity predictions with physics-consistent explanations and offering a reproducible workflow applicable to other alloys and loading conditions.

Key words

Constitutive equation, Neural network models, Multilayer Perceptron (MLP), Recurrent Neural Network (RNN), SHapley Additive exPlanations (SHAP)

1 Introduction

In the field of solid mechanics and materials science, constitutive equations serve as the mathematical backbone for modeling material behavior under mechanical loading. These equations define the relationship between stress, strain, strain rate, and temperature, enabling the prediction of material responses during forming, impact, and fatigue processes. Accurate constitutive models are essential for finite element simulations, damage assessment, and material design in sectors such as aerospace, automotive, and structural engineering ([Anand, 1982]).

Traditional constitutive models are typically phenomenological or physically based, such as the Johnson–Cook model ([Johnson and Cook, 1983]), the Zerilli–Armstrong model ([Zerilli and Armstrong, 1987]), or the Bergström model ([Bergström, 1970]). These models are often constructed based on simplified ideas regarding empirical assumptions and/or microstructural behavior and are expressed in closed-form mathematical equations. While widely used due to their physical interpretability, such models have several limitations: they often fail to accurately capture strong nonlinearities, particularly at extreme conditions such as high strain rates and elevated temperatures. They require manual calibration of parameters, which can be time-consuming and material-type-specific. They are typically not adaptable

to new materials or loading paths without modifications ([Dou et al., 2021]).

These limitations are especially critical in modern manufacturing and aerospace applications, where materials are subjected to dynamic and multi-axial stress states. For example, High-Strength Low-Alloy (HSLA) steels, such as HSLA-65, are often exposed to thermal–mechanical coupling effects that are difficult to model with conventional Johnson–Cook approaches.

Moreover, one of the important phenomena under high-rate loading conditions is adiabatic temperature rise. As plastic deformation progresses, the mechanical work is converted into heat, raising the local temperature and leading to a reduction in flow stress. This effect is especially relevant in high strain-rate or adiabatic conditions and has been widely observed in dynamic tests such as Split Hopkinson Pressure Bar experiments ([Zhao et al., 2024]). To model this, constitutive equations must incorporate temperature sensitivity, and neural networks offer a natural way to do so through data integration ([Zhang et al., 2024]; [Zhang et al., 2025]; [Yuyi Zhang, 2025]).

With the recent development of machine learning techniques, especially deep neural networks, researchers have begun exploring data-driven constitutive modeling as a promising alternative. Unlike traditional approaches, neural networks can directly learn the mapping from strain history, strain rate, and temperature to stress responses, offering greater flexibility in capturing complex nonlinearities ([Hornik et al., 1989]). This paradigm shift has received significant attention in materials science, mechanics, and computational modeling communities.

Early studies using artificial neural networks (ANNs) demonstrated their ability to capture the thermo-mechanical behavior of alloys such as Ti–22Al–25Nb ([Zhu et al., 2011]), 7050 aluminum ([Quan et al., 2016]), and H13 tool steel ([Pasco et al., 2022]). More recently, MLP and RNN have been applied to model path-dependent plasticity in metals with promising results ([Gorji et al., 2020]). RNNs, due to their memory-based architecture, are particularly suitable for capturing sequence-dependent behavior in materials. Recent research also highlighted the issue of increment-size dependence in RNNs, raising concerns over their self-consistency and stability ([Bonatti and Mohr, 2022]).

In addition to nonlinear elasticity and hardening effects, metals exhibit significant thermal softening, especially under high-rate deformation. This phenomenon is driven by the temperature rise due to plastic work, leading to a decrease in flow stress. Zhao et al. ([Zhao et al., 2024]) quantitatively investigated this effect in HSLA steels using ANN-enhanced constitutive formulations, emphasizing the need to embed temperature sensitivity explicitly in neural-based models.

This study aims to develop a data-driven constitutive model for HSLA-65 steel that incorporates strain, strain rate, and temperature as key input features. Two types of neural network architectures are employed: MLP model, which learns the direct mapping from input parameters to flow stress. RNN model, which captures the historical dependency and path-dependent behavior of materials.

In addition to performance comparison between MLP and RNN models, we incorporate SHAP to enhance interpretability and identify the relative influence of each input factor. This approach addresses a key limitation of black-box neural models, which typically lack transparency and physical consistency. By using SHAP, we are able to provide physically meaningful insights into the contributions of strain, strain rate, and temperature to the model predictions, especially in capturing thermal softening and strain rate sensitivity. This strategy aligns with recent efforts to apply explainable machine learning techniques in constitutive modeling, where SHAP has been used to interpret neural models of alloys under thermo-mechanical loading and to validate whether the learned behavior matches physical expectations such as hardening or rate dependence ([Li et al., 2022]).

In this study, our main contributions are threefold. First, we develop data-driven constitutive models for HSLA-65 steel based on both MLP and RNN that operate across a wide range of temperatures and strain rates. Second, we integrate SHAP into the modeling pipeline to quantify the relative influence of temperature, strain rate, and plastic strain, thereby linking the learned representations to classical constitutive mechanisms. Third, we demonstrate that the RNN yields smoother, more stable stress-strain trajectories and more balanced feature attributions, making it particularly suitable for engineering-oriented simulations. The remainder of the paper is organized as follows: Section 2 describes the thermo-mechanical background, neural network architectures, and SHAP methodology. Section 3 presents the numerical results and SHAP-based interpretations. Section 4 concludes the paper.

2 Methodology

2.1 Thermal softening

Metal thermal softening refers to the phenomenon where the material's mechanical properties, specifically its strength, decrease with increasing temperature. This phenomenon is crucial in the analysis of materials under high-temperature conditions, where the deformation behavior changes significantly as the metal heats up. Typically, as the temperature increases, the atomic vibrations intensify, leading to a reduction in yield strength and an increased rate of plastic deformation([Le and Piao, 2019]). This is especially important in processes such as forging, welding, and casting, where temperature fluctuations play a significant role in the material's behavior. Thermal softening can be attributed to various factors, including the increased movement of dislocations and the weakening of atomic bonds ([Le and Piao, 2019]). In the context of constitutive modeling, understanding

this effect is essential for accurate predictions of the material's response to thermal and mechanical loads.

During plastic deformation, a portion of the mechanical work is converted into thermal energy, which may be dissipated to the surrounding environment or retained within the material itself, leading to a temperature increase. When the rate of heat generation exceeds the rate of heat dissipation, the temperature of the material rises. Under high-strain rate conditions, the deformation process is often considered adiabatic, wherein the majority of the heat generated by rapid plastic working is retained within the material. This rapid temperature elevation induces thermal softening, resulting in a concomitant reduction in the flow stress. The temperature rise is commonly evaluated using equation (1), which simplifies to the one-dimensional form given in equation (2) ([Mason et al., 1994]; [Hodowany et al., 2000]; [Rogers, 1979]).

$$\dot{T} = \frac{\eta}{\rho C_v} \boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}}^{\boldsymbol{p}} \tag{1}$$

$$\Delta T = \frac{\eta}{\rho C_v} \int \sigma \, d\varepsilon_p \tag{2}$$

Here, equation (1) describes the rate of temperature change during plastic deformation in a differential form. In this equation, $\dot{T}$ represents the temperature rate, η is the plastic work-heat conversion factor, ρ denotes the material density, C_v is the specific heat capacity at constant volume, $\boldsymbol{\sigma}$ is the stress tensor, and $\dot{\boldsymbol{\varepsilon}}^{\boldsymbol{p}}$ is the plastic strain rate tensor. This formulation establishes a transient relationship between the plastic power and the temperature rise rate from the perspective of energy conservation, making it suitable for incremental constitutive integration. And equation (2) provides an explicit expression for the accumulated temperature rise over a finite plastic deformation process. This form is particularly applicable under uniaxial stress states or approximately homogeneous deformation conditions, offering a practical basis for experimental calibration of the parameter η and theoretical estimation of adiabatic temperature rise.

2.2 Neural network models for constitutive equation

The goal is to construct a data-driven constitutive equation using neural networks, which can effectively capture the complex relationship between strain, stress, and temperature, especially accounting for the metal thermal softening effect. We propose two neural network architectures for this purpose: MLP and RNN. These models are chosen for their ability to capture nonlinearities and temporal dependencies in the data.

2.2.1 MLP The MLP ([Karlik and Olgac, 2011]; [Desai and Shah, 2021]) is a type of feedforward neural network composed of multiple layers of neurons. The structure consists of an input layer, one or more hidden layers, and an output layer. The MLP is well-suited for learning mappings from input features (such as strain, strain rate and temperature) to output values (such as stress). The network structure of MLP is shown in Figure 1.

The MLP consists of multiple layers, each composed of multiple neurons (also called. The layers are fully connected through weight matrices. In this study, we use an MLP network with $l(l = 2)$ hidden layers, each containing $n(n = 256)$ neurons. LeakyReLU activation with a negative slope of $\alpha = 0.1$ is applied after each hidden layer. The output layer contains a single neuron with linear activation, producing the predicted stress value.

Input Layer: The input layer contains an input vector of $m(m = 3)$ features $[x_1, x_2, x_3]$, corresponding to plastic strain, temperature, and strain rate. In this layer, each input node x_i is connected to the corresponding weight w_i.

Hidden Layers: The hidden layers of the MLP consist of 2 layers, each with 256 neurons. Each neuron receives the output from the previous layer, performs a weighted sum, and then applies the LeakyReLU activation function to generate the output.

For the l-th layer, the output of the j-th neuron is computed as:

$$h_l^j = f\left(\sum_{i=1}^{n_{l-1}} w_{l-1,i} h_{l-1}^i + b_l^j\right) \tag{3}$$

, where h_l^j is the output of the j-th neuron in the l-th layer, $w_{l-1,i}$ is the weight from the i-th neuron of layer $l-1$ to the j-th neuron of layer l, b_l^j is the bias term of the j-th neuron in the l-th layer, $f(\cdot)$ is the activation function (such as ReLU, Sigmoid, etc.).

Output Layer: The output layer computes the weighted sum of the neurons in the last hidden layer and applies the activation function to produce the final output. Assuming the output layer has 1 nodes (representing predicted stress values), the calculation is:

$$o = f\left(\sum_{i=1}^{n_l} w_l^i h_l^i + b_l\right) \tag{4}$$

, where o is the output (predicted stress value), w_l^i is the weight from the i-th neuron in the l-th layer to the output neuron, h_l^i is the output of the i-th neuron in the l-th layer, b_l is the bias term of the output layer, $f(\cdot)$ is the activation function (such as Sigmoid, Softmax, LeakyReLU, etc.).

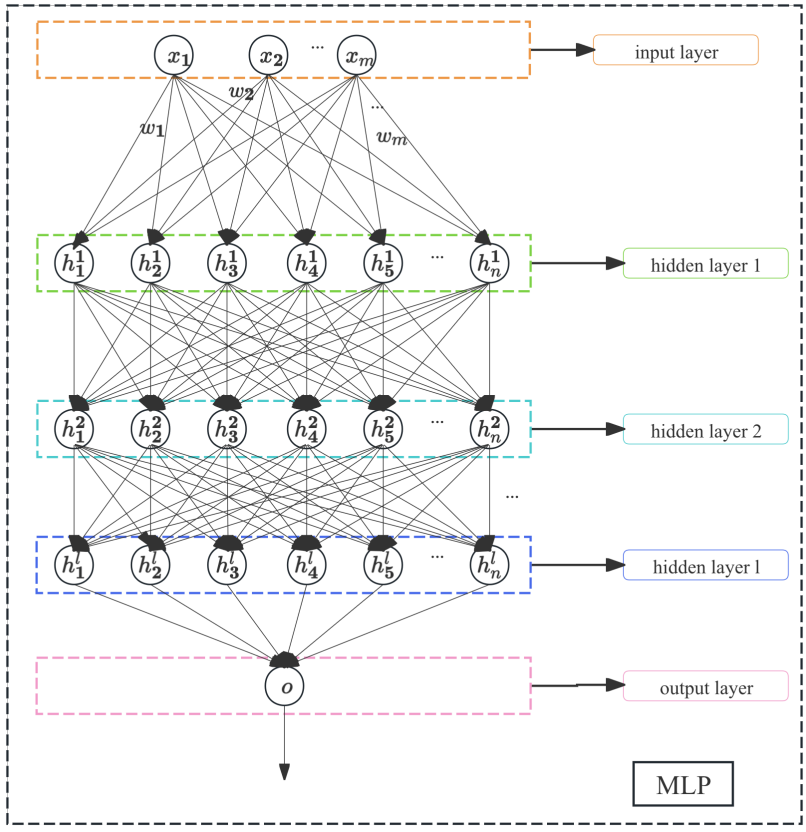

Figure 1. The network structure of MLP model, showing the input layer with 3 features (plastic strain, temperature, strain rate), 2 hidden layers with 256 neurons each using LeakyReLU activation, and an output layer with a single neuron for stress prediction.

2.2.2 RNN The RNN ([Bengio et al., 1993]; [Sutskever and Hinton, 2010]) is a type of neural network designed for sequence prediction tasks. Unlike traditional feedforward networks, RNNs have connections that allow information to persist over time, making them suitable for tasks where the input data has temporal dependencies. The RNN consists of an input layer, one or more recurrent hidden layers, and an output layer. At each time step, the hidden state is updated based on the current input and the previous hidden state using a tanh activation function, and the final hidden state is mapped to a scalar stress value through a linear output layer. The structure of the RNN is shown in Figure 2.

The RNN consists of multiple layers, each containing multiple neurons. In each hidden layer, each neuron is recurrently connected to itself at the next time step, creating a loop. The layers are fully connected through weight matrices, and the hidden state from the previous time step influences the current time step. This structure allows the RNN to learn temporal dependencies between input sequences and output. In this study, we use a stacked vanilla RNN with $l(l = 2)$ hidden layers, each containing $n(n = 128)$ neurons. Tanh activation is applied after each hidden layer. The output layer contains a single neuron with linear activation, producing the predicted stress value.

Input Layer: The input layer contains a vector of $m(m = 3)$ features $[x_1, x_2, x_3]$, corresponding to plastic strain, temperature, and strain rate. In this layer, each input node x_i is connected to the corresponding weight w_i. The input layer provides the input sequence at each time step.

Hidden Layers: The hidden layers of the RNN consist of 2 layers, each with 128 neurons, using tanh as the activation. Each neuron receives input not only from the previous time step's hidden state but also from the current input. This enables the network to maintain memory over time. The hidden state of the RNN is updated at each time step.

For the l-th hidden layer at time step t, the output h_t^l is calculated based on the previous hidden state h_{t-1}^l and the current input x_t. The mathematical expression for the calculation of the hidden state is:

$$h_t^l = f\left(\sum_{i=1}^{n_{l-1}} w_{l-1,i} h_{t-1}^{l-1} + \sum_{j=1}^{m} w_{l,j} x_{t,j} + b_l^j\right) \tag{5}$$

, where h_t^l is the output of the l-th hidden layer at

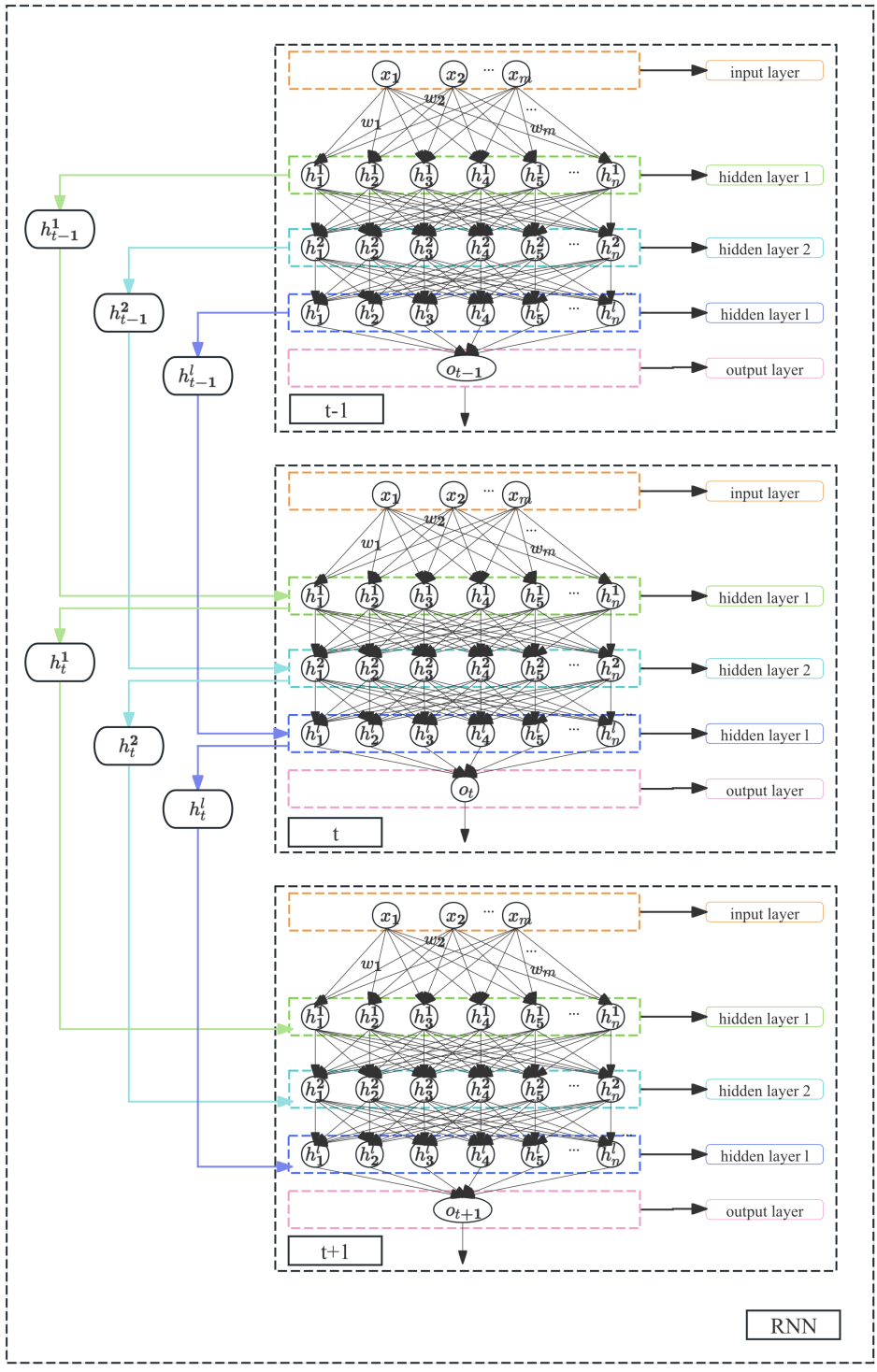

Figure 2. The network structure of RNN model, illustrating the input layer with 3 features, two recurrent hidden layers with 128 neurons each using tanh activation, and an output layer for stress prediction at each time step, emphasizing its sequential memory mechanism.

time step t, h_{t-1}^{l-1} is the hidden state from the previous time step and layer $l-1$, x_t is the input at time step t, including all features, $w_{l-1,i}$ are the weights between the previous layer's neurons h_{t-1}^{l-1} and the current layer's neurons, $w_{l,j}$ are the weights from the input x_t to the neurons in the current layer, b_l^j is the bias term for the j-th neuron in the l-th hidden laye, $f(\cdot)$ is the activation function (such as ReLU, tanh, sigmoid, etc.).

Each layer receives its input from both the previous time step's hidden states and the current time step's

input, which allows RNNs to capture temporal dependencies.

Output Layer: The output layer computes the weighted sum of the neurons in the final hidden layer and applies the activation function to produce the final output. Assuming the output layer has 1 nodes (representing predicted stress values.The mathematical expression for the output at time step t is:

$$o_t = f\left(\sum_{i=1}^{n_l} w_{o,i} h_t^l + b_o\right) \tag{6}$$

, where o_t is the output at time step t (representing predicted stress value), $w_{o,i}$ are the weights from the i-th neuron in the final hidden layer l to the output, h_t^l is the output of the last hidden layer at time step t, b_o is the output layer's bias term, $f(\cdot)$ is the activation function for the output layer (typically sigmoid or softmax).

2.2.3 SHAP Although data-driven models such as neural networks possess certain advantages in modelling constitutive relationships, their lack of interpretability continues to constrain further development. This work therefore employs SHAP ([Lundberg and Lee, 2017]; [Nohara et al., 2022]; [García and Aznarte, 2020]; [Dwivedi et al., 2023]) as an interpretative framework to quantify the contribution of strain, temperature, and rate to stress prediction outcomes, thereby mitigating trust issues arising from interpretability challenges to a significant degree. Grounded in cooperative game theory, SHAP represents a model's output for a given instance as the sum of a baseline term - defined by the model's expectation over a background distribution - and additive contributions from each input feature. The sign of a SHAP value indicates whether a feature increases or decreases the prediction relative to the baseline; its magnitude reflects the strength of that effect. The attributions satisfy additivity and completeness: the baseline plus the sum of all feature SHAP values exactly equals the model's output for that instance, enabling auditable local explanations.

To attribute predictions of nonlinear models, we treat features as players in a coalition game and evaluate their contributions via $val_x(S)$. The Shapley value of feature i equals the average of its marginal contributions over all permutations of the features, which is equivalent to the weighted sum over all subsets in equation (7). This aligns with the fair allocation principle in cooperative games.

$$\phi_i(val) = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} \times \left[val(S \cup \{i\}) - val(S)\right] \tag{7}$$

, where $F = \{1, \ldots, F\}$ is the full set of features, and S is any subset of features excluding the i-th feature. The contribution ϕ_i for the i-th feature is obtained by calculating its expected marginal contribution across all possible coalitions. The weight term $\frac{|S|!(|F|-|S|-1)!}{|F|!}$ compensates for the difference in the number of combinations across subsets of different sizes.

The coalition value $val_x(S)$ quantifies the prediction attributable to the feature set S for instance x by integrating out all complementary features. As shown in equation (8), it is written as a conditional–expectation–type functional that averages $\hat{f}(x_S, X_{\overline{S}})$ over the distribution of $X_{\overline{S}}$, and subtracts the unconditional baseline $\mathbb{E}[\hat{f}(X)]$, the baseline term is evaluated as the empirical average of the model predictions over a randomly sampled background subset of the training data, which serves as a reference state for the SHAP attributions..

$$val_x(S) = \int \hat{f}(x_S, X_{\overline{S}}) d\mathbb{P}(X_{\overline{S}}) - E_X(\hat{f}(X)) \tag{8}$$

3 Experimental results and analysis

The HSLA-65 Steel dataset used in this study covers compression conditions across multiple temperatures and strain rates. Following common practices in the experimental literature, the data are organized into four representative temperatures (77 K, 296 K, 500 K, and 700 K) and three strain-rate (0.001/s, 3000/s, and 8500/s), yielding 12 combinations of operating conditions in total. The raw data consist of pointwise records of plastic strain, strain rate, temperature, and the corresponding true stress.

3.1 Stress-strain curve predictions

Figures 3 and 4 illustrate the stress-strain descriptive performance of the MLP and RNN models on the training dataset, while Figures 5 and 6 present their predictive performance on the test dataset at a fixed temperature of 500 K under three strain rates (0.001/s, 3000/s, and 8500/s). In these figures, the horizontal axis denotes plastic strain, and the vertical axis represents stress (MPa). Solid triangular markers correspond to experimental data, dashed lines indicate model performance, and different colors distinguish the combinations of strain rate and temperature conditions.

Figures 3 and 4 show the stress-strain curve fits of the MLP and RNN models on the training dataset, respectively. At the level of global mechanical trends, both models behave consistently: increasing strain leads to work hardening, higher temperature produces thermal softening, and higher strain rate elevates the overall strength. Their principal difference lies in the treatment of local features. The MLP is more sensitive to small-scale fluctuations and closely follows sharp knees in the small-strain regime, which can introduce additional

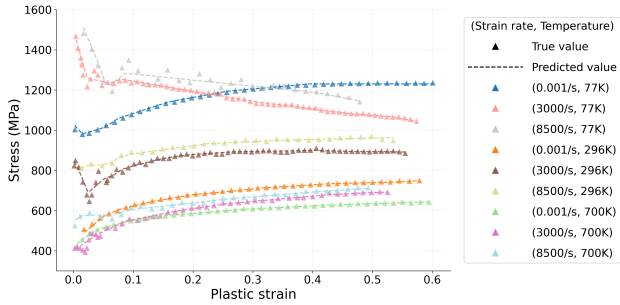

Figure 3. Stress–strain curves of HSLA-65 steel under various temperatures (77 K, 296 K, 700 K) and strain rates (0.001/s, 3000/s, 8500/s), comparing experimental data with predictions from the MLP model on the training dataset. The model reproduces the overall trends of strain hardening and thermal softening present in the experimental data. However, the predicted curves exhibit minor oscillations and adhere closely to individual data points, suggesting a tendency to learn fine-grained noise and indicating a potential risk of overfitting.

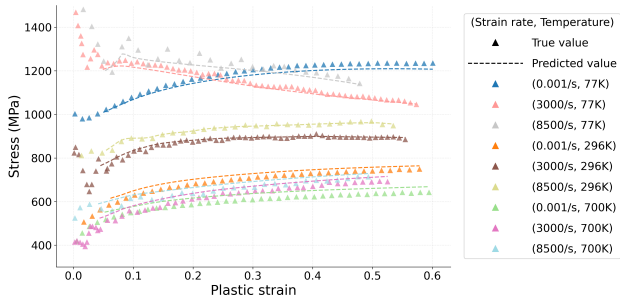

Figure 4. Stress–strain curves of HSLA-65 steel under multiple temperatures and strain rates on the training dataset, showing a comparison between experimental data and predictions from the RNN model. The RNN generates smoother and more continuous predictive curves, which reflects its inherent strength in learning the underlying path-dependent and temporally evolving material response. This smoothness indicates a robust capture of the fundamental constitutive behavior rather than overfitting to local variations.

local undulations under certain conditions. By contrast, the RNN produces smoother response sequences and effectively suppresses high-frequency noise. Under low-temperature and high-rate conditions (77 K, 3000–8500/s), the experimental curves typically exhibit a transient peak at small strains, followed by a drop and re-hardening. The MLP closely follows these local peaks and troughs, whereas the RNN, while preserving the overall strength level and the ordering of the curves, mitigates excessively sharp peaks and suppresses noise-induced fluctuations. At the intermediate temperature (296 K), both models show monotonic hardening that approaches saturation; the RNN displays a more continuous evolution of the hardening modulus, a more stable saturation segment, and a more uniform, clearly separated spacing among strain-rate curves. At the high temperature (700 K), where the overall strength drops and rate sensitivity weakens, the MLP predictions exhibit multiple inflection points in the small-strain range, whereas the RNN maintains uniformly separated curves and a smooth manifestation of thermal softening, reducing the influence of minor undulations on global interpretation. In summary, on the training set the RNN provides robust suppression of high-frequency fluctuations and yields more consistent, smoother curve families, which facilitates mechanical interpretation of hardening/softening processes and of the coupled effects of strain rate and temperature.

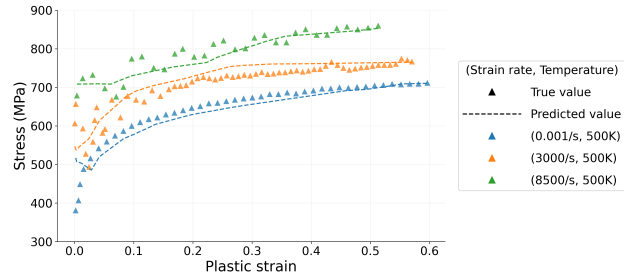

Figure 5. Predictive performance of the MLP model on the test dataset at a fixed temperature of 500 K under three strain rates (0.001/s, 3000/s, and 8500/s). The comparison with experimental data reveals systematic deviations and oscillatory predictions, particularly at 8500/s. This behavior, combined with its high accuracy on the training set, suggests potential overfitting, limiting its generalization to unseen conditions.

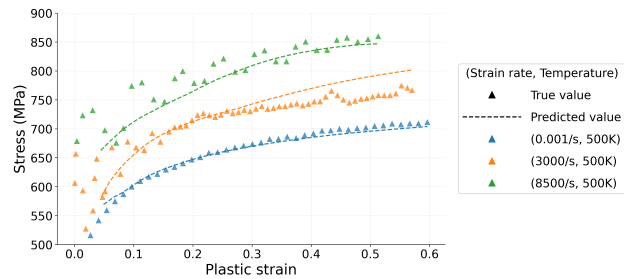

Figure 6. Predictive performance of the RNN model on the test dataset under the same conditions (500 K, three strain rates). The RNN generates smoother and more stable stress-strain curves that closely align with the experimental data across all strain rates. This demonstrates its superior ability to learn the underlying physical trends without overfitting, leading to robust generalization.

Figures 5 and 6 show the corresponding responses of the MLP and RNN models on the test dataset. For the MLP (Figure 5), the curves in the small-strain regime are sensitive to minor experimental fluctuations and often exhibit multiple turning points and zigzag patterns. As strain increases, the slope (hardening rate) tends to change in a piecewise manner, and the spacing among

the three strain-rate curves locally narrows or widens. These features indicate that the MLP tends to follow fine-scale variations and, under the new temperature condition, shows signs of overfitting to local details. By contrast, the RNN (Figure 6) produces curve families that harden monotonically and gradually level off; knees are rounded and short-lived peaks/valleys at small strain are markedly attenuated. Across the entire strain range, the three curves largely preserve their ordering with relatively uniform spacing, yielding a more consistent expression of rate effects. Such smooth behavior suggests that the RNN captures the temporal dependence in the strain–stress sequence more effectively, thereby exhibiting greater robustness and extrapolation capability under unseen conditions.

A synthesis of the training- and test-set results shows that both models reproduce the principal trends-work hardening with increasing strain, thermal softening with increasing temperature, and rate-induced strengthening. Nevertheless, the RNN offers greater engineering utility in curve morphology and mechanism display: across temperature-rate conditions it yields smooth, order-preserving, and relatively uniformly spaced curve families, attenuating short-lived peaks and small fluctuations in the small-strain regime and presenting a continuous hardening-to-plateau trajectory that clearly conveys the joint influence of strain rate and temperature. This shape stability facilitates cross-condition extrapolation and seamless coupling with numerical solvers (e.g., finite-element integration), supporting parameter identification and mechanistic interpretation. By contrast, the MLP is more detail-sensitive, tending to follow small-amplitude fluctuations at small strains and to exhibit piecewise changes in slope and local variations in curve spacing, it thus serves as a complementary tool for inspecting brief peaks and minor inflections observed experimentally. Overall, the RNN is better suited for constitutive modeling of HSLA-65 steel and further underscores the effectiveness of data-driven neural-network approaches for characterizing nonlinear material responses.

3.2 Quantitative performance evaluation

To further evaluate and compare the predictive accuracy and robustness of the two models, several statistical metrics (equation (9), (10), (11), (12) and (13)) were computed on the test dataset, including Mean Relative Error (MRE), Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). The results are summarized in Table 1.

$$MRE = \frac{1}{m}\sum_{i=1}^{m}\frac{|pred_i - true_i|}{|true_i|} \tag{9}$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|pred_i - true_i| \tag{10}$$

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(pred_i - true_i)^2 \tag{11}$$

$$RMSE = \sqrt{MSE} \tag{12}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(pred_i - true_i)^2}{\sum_{i=1}^{m}(true_i - \overline{true})^2} \tag{13}$$

Table 1. Model Performance Comparison

Metric	MLP	RNN
MRE	0.0251	0.0175
MAE	17.1139	13.0811
MSE	468.4130	335.5691
RMSE	21.6429	18.3185
R^2	0.9278	0.9311

Both models achieved high predictive accuracy, with R^2 values exceeding 0.9, confirming strong agreement between predicted and experimental stress values. However, when examining the metrics in detail, the MRE emerges as the most representative measure of performance. Unlike absolute error metrics, MRE normalizes the prediction error with respect to the magnitude of the measured stress, thereby providing a scale-independent evaluation of accuracy.

The results show that the RNN achieved an MRE of 0.0175 (= 1.75%), significantly lower than the MLP's MRE of 0.0251 (= 2.51%). This reduction corresponds to an improvement of approximately 30%, demonstrating that the RNN's predictions deviate, on average, by less than 2% from the experimental data. Such a low relative error is particularly noteworthy in the context of constitutive modeling, where prediction errors must remain small across a wide range of temperatures and strain rates.

Supporting metrics are consistent with the superiority of the RNN: the model reduced MAE by 25% and RMSE by 16%, while increasing R^2 from 0.9278 to 0.9311, further indicating improved goodness-of-fit. Nevertheless, the reduction in MRE is the most compelling evidence of the RNN's enhanced ability to generalize across diverse thermo-mechanical conditions.

Taken together, these findings highlight the RNN's stronger capacity to capture the inherent nonlinear dependencies in HSLA-65 steel's strain-stress response.

When combined with the qualitative curve comparisons, it can be concluded that the RNN not only achieves better alignment with experimental observations but also provides quantitatively more reliable predictions, with MRE as the key indicator of accuracy. This underscores the promise of RNN-based approaches for data-driven constitutive modeling, especially in scenarios where precise prediction across multiple regimes is critical.

3.3 Model explainability with SHAP

To evaluate whether the models have learned input-output mappings consistent with the underlying constitutive mechanisms of the material, we compute Shapley values in the standardized input space.

Figures 7 and 8 present the relationships between input feature values and their corresponding SHAP contributions for the MLP and RNN models on the training dataset. In each subplot, the horizontal axis denotes the raw values of a specific input feature, while the vertical axis indicates its marginal contribution to the predicted stress (SHAP value). Scatter points are color-coded by feature magnitude (ranging from purple for low values, through blue, to yellow for high values). The bold curve represents the median trend, and the shaded area indicates the interquartile range (IQR). It should be emphasized that the positive or negative sign of the SHAP value reflects the marginal influence relative to the background output, rather than the absolute stress magnitude.

Figure 7 illustrates the SHAP feature contribution trends of the MLP model on the training dataset. Subfigure (a) shows a clear monotonic negative correlation between temperature and SHAP values: low-temperature samples (purple, 77 K) generally yield positive contributions, which decrease toward zero at intermediate temperatures (blue, 296 K) and become strongly negative at high temperatures (yellow, 700 K). The median curve decreases approximately linearly, with a narrow IQR, indicating that the temperature effect is both strong and stable. This behavior aligns with the thermal softening mechanism: increasing temperature reduces material strength, resulting in a systematic downward adjustment of the predicted stress. Subfigure (b) demonstrates a monotonic positive correlation between strain rate and SHAP values: at low strain rates (purple, 0.001/s), the median SHAP values are negative, crossing zero near the intermediate strain rate (blue, 3000/s), and becoming strongly positive at the high strain rate (yellow, 8500/s). The IQR also narrows as the rate increases. This result clearly reflects strain-rate sensitivity: higher strain rates lead to upward adjustments of the predicted stress. Subfigure (c) reveals a non-monotonic evolution with plastic strain: at very small strains (purple, < 0.1), SHAP values are predominantly negative; with increasing strain, SHAP values cross zero and become moderately positive at medium-to-high strains (green/yellow, ≈ 0.3). The IQR is wider at small strains but converges as strain increases, indicating variations in early yield and hardening onset across temperature-strain-rate combinations. Overall, the trend is consistent with the physical picture of progressive strain hardening that strengthens and saturates with increasing deformation. In summary, temperature exhibits the largest and most stable influence (strongly negative, reflecting thermal softening), followed by strain rate with a clear positive effect (rate hardening), and plastic strain contributing a moderate positive effect associated with strain hardening. The directions, magnitudes, and dispersions of these trends are in agreement with classical constitutive behavior, demonstrating that the MLP not only achieves accurate fitting but also implicitly learns the coupled physical mechanisms of temperature, strain rate, and strain.

Figure 8 illustrates the SHAP feature contribution trends of the RNN model on the training dataset. Subfigure (a) shows a monotonic decrease in SHAP values with increasing temperature: the low-temperature regime (purple, 77 K) corresponds to strong positive contributions, while the high-temperature regime (yellow, 700 K) exhibits strongly negative contributions. This trend clearly reveals the thermal softening effect, whereby low temperatures enhance stress levels whereas high temperatures reduce material strength. Subfigure (b) demonstrates a typical strain-rate sensitivity pattern characterized by a transition from negative to positive contributions. At low strain rates (purple, 0.001/s), SHAP values are negative, indicating that slow deformation lowers the predicted stress level. At high strain rates (yellow, 8500/s), SHAP values become strongly positive, reflecting the rate-hardening effect. Notably, the RNN curve is steeper than that of the MLP, suggesting that the RNN captures more pronounced nonlinearities in strain-rate sensitivity. Subfigure (c) exhibits an approximately linear upward trend. As plastic strain increases (colors shifting from purple to yellow), SHAP values progressively transition to positive contributions, reflecting the strain-hardening effect during sustained plastic deformation. Compared with the MLP, the RNN's plastic-strain contribution curve is more concentrated and less noisy, indicating that the RNN places greater emphasis on capturing the overall evolutionary trend rather than local fluctuations.

Taken together, the SHAP interpretation of the RNN not only confirms the three principal physical mechanisms of deformation (thermal softening, strain-rate sensitivity, and strain hardening),but also underscores the advantage of RNNs in modeling temporal dependencies. Compared with the MLP, the RNN reveals more explicit monotonic relationships between input features and stress responses, particularly in the cases of strain rate and plastic strain, thereby demonstrating stronger physical consistency.

Figure 9 quantifies the global importance of each feature based on the mean absolute SHAP values, thereby measuring their relative contributions to stress prediction. The vertical axis lists the three input features, while the horizontal axis denotes their average contribution

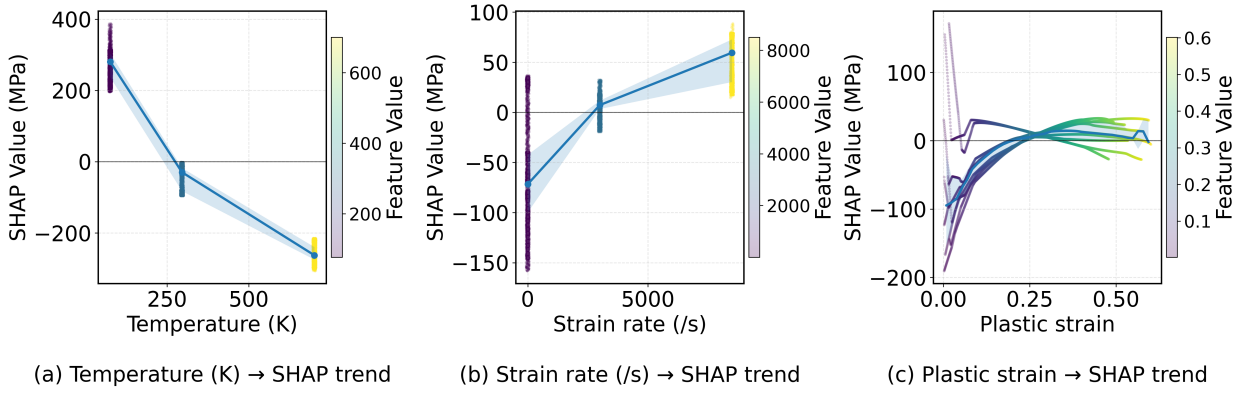

Figure 7. SHAP analysis of feature contributions for the MLP model, showing the marginal effects of temperature (thermal softening), strain rate (rate sensitivity), and plastic strain (strain hardening) on stress predictions, with trends consistent with physical mechanisms such as monotonic decrease with temperature and increase with strain rate.

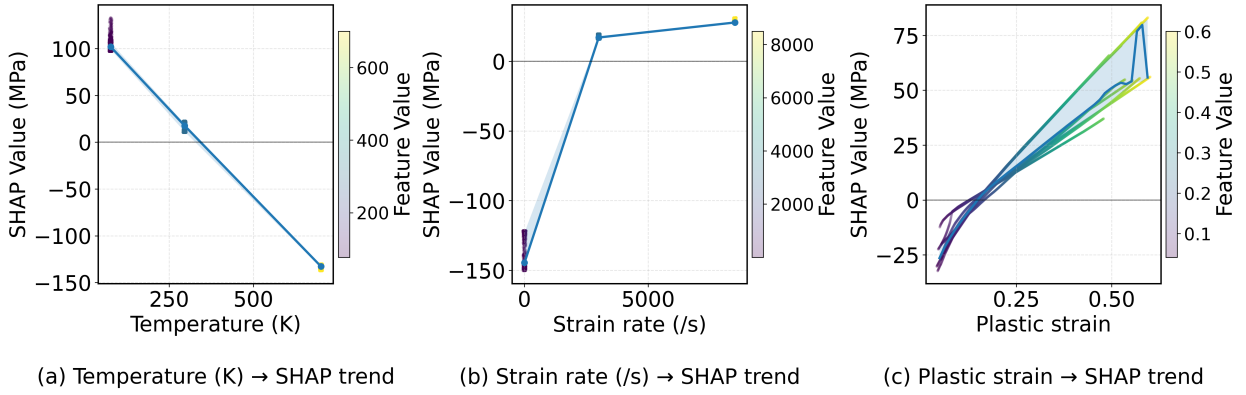

Figure 8. SHAP analysis of feature contributions for the RNN model, illustrating the relationships between input features (temperature, strain rate, plastic strain) and their SHAP values, revealing monotonic trends for thermal softening and rate sensitivity, and a linear strain hardening effect, with reduced noise compared to the MLP.

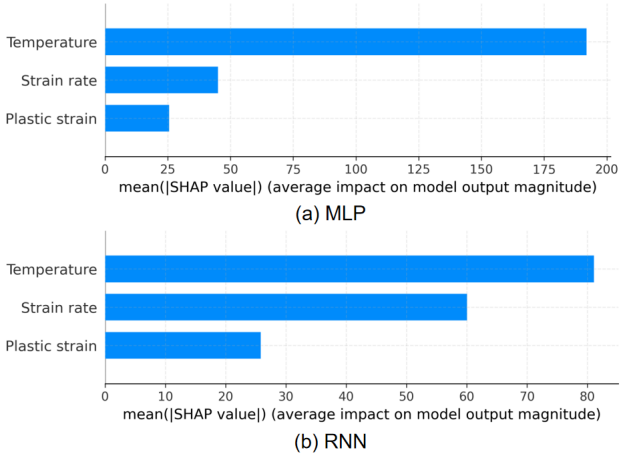

Figure 9. Global feature importance for the MLP (a) and RNN (b) model based on mean absolute SHAP values, indicating that temperature is the dominant feature with the largest contribution to stress predictions, followed by strain rate and plastic strain, reflecting the model's reliance on thermal softening as the primary driver.

magnitudes. Conceptually, the mean (|SHAP value|) represents the average marginal effect of a given input feature on the model output across all samples. Larger values signify that the feature plays a more critical role in the model's decision-making process.

Figure 9(a) presents the global feature importance of the MLP model based on SHAP analysis. The results indicate that temperature overwhelmingly dominates the feature contributions, with an average SHAP magnitude, far exceeding those of strain rate and plastic strain. This finding suggests that, within the static mapping mechanism of the MLP, temperature is regarded as the primary driver controlling stress levels, with the model's predictions relying almost exclusively on temperature to distinguish the baseline stress across different conditions. This conclusion is consistent with fundamental principles of solid mechanics, where thermal softening directly reduces the flow stress level. By contrast, the contribution of strain rate is approximately half that of temperature, while plastic strain contributes the least, indicating that the MLP captures dynamic loading effects and strain hardening behavior only to a limited extent.

Figure 9(b) illustrates a markedly different feature utilization pattern for the RNN model on the test dataset. Although temperature remains the most important feature, its relative dominance is substantially reduced; strain rate exhibits a contribution nearly comparable to temperature. This suggests that the recurrent neural network, by leveraging its sequential memory structure, effectively captures the persistent modulation of material response by strain rate. This highlights the RNN's unique advantage in handling dynamic loading conditions: it not only learns the baseline effect associated with temperature but also incorporates strain-rate effects as temporally dependent regulatory factors. Moreover, the importance of plastic strain in the RNN is significantly enhanced, with SHAP values approaching those of temperature and strain rate. This indicates that the RNN is able to identify and utilize the cumulative and nonlinear evolution of strain, thereby providing a more accurate representation of the material's strain-hardening process.

Overall, the differences in feature utilization between MLP and RNN reflect the correspondence between their architectural characteristics and the underlying physical mechanisms. The MLP, as a feedforward static network, tends to extract predictive signals from a single dominant feature (temperature), with limited sensitivity to history dependence and dynamic effects. In contrast, the RNN, by virtue of its recurrent structure, is capable of jointly considering temperature, strain rate, and plastic strain when characterizing the stress–strain evolution, yielding a more balanced and physically consistent pattern of feature contributions. These findings further emphasize that, under high strain-rate and complex thermomechanical conditions, deep learning frameworks based on sequence modeling (such as the RNN) exhibit clear advantages in capturing constitutive behavior and providing explanations that align more closely with the physical mechanisms of material deformation.

The SHAP analysis corroborates three key physical trends: (1) temperature exerts a dominant negative influence on stress, consistent with thermal softening; (2) strain rate contributes positively to stress in a strongly nonlinear fashion, reflecting rate-dependent strengthening; and (3) accumulated plastic strain makes a positive contribution associated with strain hardening. Compared with the MLP, the RNN distributes importance more evenly across these three features, indicating a more balanced and physically coherent use of the input variables.

4 Conclusion

This study investigates the deformation behavior of HSLA-65 steel under complex thermo-mechanical loading paths by developing data-driven constitutive models using both MLP and RNN architectures. The internal decision mechanisms and physical consistency of the models are systematically interpreted using the SHAP framework.

First, both neural network models effectively reproduce the experimentally observed stress-strain responses, validating the applicability of data-driven approaches for constitutive modeling. Although the MLP and RNN exhibit distinct fitting behaviors during training, they consistently identify temperature, strain rate, and plastic strain as the dominant variables governing the material response, aligning with established principles in solid mechanics.

Second, comparative analysis reveals that the MLP closely follows the experimental data points within the training regime but exhibits localized oscillations, such as jagged fluctuations and segmented slope variations-particularly in small-strain regions-indicating sensitivity to data noise and a tendency toward overfitting, which constrains its extrapolation stability. In contrast, the RNN, leveraging its sequential modeling capability, generates smoother, more continuous stress–strain curves with stable spacing and monotonic transitions, reflecting a superior ability to capture path-dependent deformation. Consequently, the RNN demonstrates more robust predictive performance under unseen conditions, especially under high strain rates and extreme temperatures, making it more suitable for engineering-level numerical simulations. SHAP analysis further reveals that the MLP relies predominantly on temperature, whereas the RNN integrates the effects of strain rate history and accumulated plastic strain in a more balanced manner, enabling a more effective representation of nonlinear multi-field couplings that aligns with the smooth and monotonic characteristics of the macroscopic curves.

Finally, this study underscores the importance of interpretable machine learning in constitutive modeling. The low-noise, high-stability output characteristics of the RNN better meet the requirements of engineering computations for trajectory smoothness and predictive reliability, offering a more practical data-driven framework for constitutive modeling. This work provides a valuable reference for the development of physically consistent, morphologically stable, and explainable material constitutive models.

Acknowledgements

This work was financially supported by the Russian Science Foundation (RSF 22-11-00091).

References

Anand, L. (1982). Constitutive equations for the rate-dependent deformation of metals at elevated temperatures. *Journal of Engineering Materials and Technology-transactions of The Asme*, **104**, pp. 12–17.

Bengio, Y., Frasconiy, P., and Goriy, M. (1993). Recurrent neural networks for adaptive temporal processingyoshua bengio.

Bergström, Y. (1970). A dislocation model for the stress-strain behaviour of polycrystalline α-fe with special emphasis on the variation of the densities of mobile

and immobile dislocations. *Materials Science and Engineering*, **5** (4), pp. 193–200.

Bonatti, C. and Mohr, D. (2022). On the importance of self-consistency in recurrent neural network models representing elasto-plastic solids. *Journal of the Mechanics and Physics of Solids*, **158**, pp. 104697.

Desai, M. and Shah, M. (2021). An anatomization on breast cancer detection and diagnosis employing multilayer perceptron neural network (mlp) and convolutional neural network (cnn). *Clinical eHealth*, **4**, pp. 1–11.

Dou, W., Xu, Z., Hu, H., and Huang, F. (2021). A generalized plasticity model incorporating stress state, strain rate and temperature effects. *International Journal of Impact Engineering*, **155**. Publisher Copyright: © 2021.

Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., Qian, B., Wen, Z., Shah, T., Morgan, G., et al. (2023). Explainable ai (xai): Core ideas, techniques, and solutions. *ACM Computing Surveys*, **55** (9), pp. 1–33.

García, M. V. and Aznarte, J. L. (2020). Shapley additive explanations for no2 forecasting. *Ecological Informatics*, **56**, pp. 101039.

Gorji, M. B., Mozaffar, M., Heidenreich, J. N., Cao, J., and Mohr, D. (2020). On the potential of recurrent neural networks for modeling path dependent plasticity. *Journal of the Mechanics and Physics of Solids*, **143**, pp. 103972.

Hodowany, J., Ravichandran, G., Rosakis, A. J., and Rosakis, P. (2000). Partition of plastic work into heat and stored energy in metals. *Experimental Mechanics*, **40** (2), pp. 113–123.

Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, **2** (5), pp. 359–366.

Johnson, G. R. and Cook, W. H. (1983). A constitutive model and data for metals subjected to large strains, high strain rates and high temperatures. *Engineering Fracture Mechanics*, **21**, pp. 541–548.

Karlik, B. and Olgac, A. V. (2011). Performance analysis of various activation functions in generalized mlp architectures of neural networks. *International Journal of Artificial Intelligence and Expert Systems*, **1** (4), pp. 111–122.

Le, K. and Piao, Y. (2019). Thermal softening during high-temperature torsional deformation of aluminum bars. *International Journal of Engineering Science*, **137**, pp. 1–7.

Li, X., Roth, C., Bonatti, C., and Mohr, D. (2022). Counterexample-trained neural network model of rate and temperature dependent hardening with dynamic strain aging. *International Journal of Plasticity*.

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, **30**.

Mason, J., Rosakis, A., and Ravichandran, G. (1994). On the strain and strain rate dependence of the fraction of plastic work converted to heat: an experimental study using high speed infrared detectors and the kolsky bar. *Mechanics of Materials*, **17** (2), pp. 135–145.

Nohara, Y., Matsumoto, K., Soejima, H., and Nakashima, N. (2022). Explanation of machine learning models using shapley additive explanation and application for real data in hospital. *Computer Methods and Programs in Biomedicine*, **214**, pp. 106584.

Pasco, J., Mccarthy, T., Parlee, J., Nazri, N. A., Padmajan, S., Rodrigues, S., and Aranas, C. (2022). Constitutive modeling of modified-h13 steel. *MRS Communications*, **12**, pp. 343 – 351.

Quan, G.-z., Wang, T., Li, Y.-l., Zhan, Z.-y., and Xia, Y.-f. (2016). Artificial neural network modeling to evaluate the dynamic flow stress of 7050 aluminum alloy. *Journal of Materials Engineering and Performance*, **25** (2), pp. 553–564.

Rogers, H. (1979). Adiabatic plastic deformation. *Annual Review of Materials Science*, **9**, pp. 283–311.

Sutskever, I. and Hinton, G. E. (2010). Temporal-kernel recurrent neural networks. *Neural networks : the official journal of the International Neural Network Society*, **23 2**, pp. 239–43.

Yuyi Zhang, Andrey Logachov, A. S. N. K. (2025). Artificial neural network surrogates for impact problem simulations. *Cybern. Phys*, **14**, pp. 200–213.

Zerilli, F. J. and Armstrong, R. W. (1987). Dislocation-mechanics-based constitutive relations for material dynamics calculations. *Journal of applied physics*, **61** (5), pp. 1816–1825.

Zhang, Y., Logachev, A., Smirnov, A., and Kazarinov, N. (2025). Artificial neural networks for impact strength prediction of composite barriers. *Materials*, **18** (13), pp. 3001.

Zhang, Y., Zhao, S., Kazarinov, N., and Petrov, Y. V. (2024). Neural networks with iterative parameter generation for determining parameters of constitutive models. *Cybern. Phys*, **13**, pp. 334–345.

Zhao, S., Petrov, Y. V., Zhang, Y., Volkov, G., Xu, Z., and Huang, F. (2024). Modeling of the thermal softening of metals under impact loads and their temperature–time correspondence. *International Journal of Engineering Science*, **194**, pp. 103969.

Zhu, Y., Zeng, W., Sun, Y., Feng, F., and Zhou, Y. (2011). Artificial neural network approach to predict the flow stress in the isothermal compression of as-cast tc21 titanium alloy. *Computational Materials Science*, **50** (5), pp. 1785–1790.