Mechanics Research Communications 155 (2026) 104707

Contents lists available at ScienceDirect

Mechanics Research Communications

journal homepage: www.elsevier.com/locate/mechrescom

A note on constitutive artificial neural networks for finite strain plasticity ☆

Tim van der Velden [a], Birte Boes [b], Tim Brepols [a], Ellen Kuhl [c,d], Hagen Holthusen [c,*]

[a] RWTH Aachen University, Institute of Applied Mechanics, Aachen, Germany
[b] University of Wuppertal, Computational Applied Mechanics, Wuppertal, Germany
[c] Friedrich-Alexander-Universität Erlangen-Nürnberg, Institute of Applied Mechanics, Erlangen, Germany
[d] Stanford University, Department of Mechanical Engineering, Stanford, CA, United States

ARTICLE INFO

Dataset link: https://doi.org/10.5281/zenodo.16900782

Keywords:
Recurrent neural networks
Constitutive neural networks
Plasticity
Finite strains
General hardening

ABSTRACT

Data-driven approaches have led to structured constitutive artificial neural network formulations for inelastic phenomena such as viscoelasticity and growth, while the systematic extension to finite-strain plasticity remains an active area of research. In this communication, we extend the previously introduced inelastic constitutive artificial neural network framework to finite-strain elasto-visco-plasticity with combined nonlinear kinematic and isotropic hardening. The proposed formulation ensures thermodynamic consistency within a potential-based architecture by consistently integrating isotropic hardening and viscoplastic evolution mechanisms. In particular, a shared inelastic potential is employed for linear kinematic and isotropic hardening, while two additional potentials govern nonlinear kinematic hardening and the plastic yield function, enabling the modeling of non-associative plasticity. A JAX implementation of the extended framework and the associated simulation results are publicly available at Zenodo.org.

1. Introduction

The accurate prediction of plastic material behavior is crucial for the design of modern forming processes and for the investigation of structures subject to loads beyond the ultimate limit state. The finite element formulations employed in contemporary simulation tools yield reliable and robust results, whereas the corresponding material models still remain questionable.

Nevertheless, the automated material characterization based on experimental data provides an objective solution for the formulation of the material models. Early neural-network-based constitutive models, e.g. [1], incorporated mechanical knowledge at the input level through invariant-based strain measures, while retaining standard network architectures. More recent structure-preserving approaches embed continuum-mechanical principles directly into the architecture to enforce properties such as objectivity and stability by construction, see e.g. [2]. The present work follows this paradigm and extends it to finite-strain elasto-visco-plasticity with combined nonlinear kinematic and isotropic hardening within a potential-based framework. Related works on automated model discovery for plasticity at finite strains include [3,4].

Constitutive Artificial Neural Networks (CANNs) have been proposed as a systematic framework for data-driven constitutive model discovery within mechanically admissible model classes [5]. Within the paradigm of CANNss, model discovery refers to the data-driven identification of constitutive energy and potential functions within a thermodynamically constrained architectural class, rather than to unrestricted architecture search.

In this communication, we extend the previously introduced inelastic constitutive artificial neural network (iCANN) framework to finite-strain elasto-visco-plasticity with combined nonlinear kinematic and isotropic hardening. The focus lies on the consistent thermodynamic integration of the additional hardening and viscosity mechanisms within the established potential-based architecture [3]. A JAX implementation is provided in [6], and the numerical examples serve as proof-of-concept demonstrations of the proposed extension.

2. Constitutive framework for elasto-visco-plasticity

Kinematics. In accordance with Fig. 1, the deformation gradient is multiplicatively decomposed into elastic and plastic parts $\boldsymbol{F} = \boldsymbol{F}_e \boldsymbol{F}_p$ [7, 8] with an additional split of the plastic part into $\boldsymbol{F}_p = \boldsymbol{F}_b \boldsymbol{F}_a$ [9]. Next, we define the stretch-measures in the intermediate configuration

$$\boldsymbol{C}_e = \boldsymbol{F}_e{}^{\mathrm{T}} \boldsymbol{F}_e, \tag{1}$$

$$\boldsymbol{B}_p = \boldsymbol{F}_p \boldsymbol{F}_p{}^{\mathrm{T}}, \tag{2}$$

$$\boldsymbol{B}_b = \boldsymbol{F}_b \boldsymbol{F}_b{}^{\mathrm{T}}, \tag{3}$$

☆ This article is part of a Special issue entitled: ‘SciML in Mechanics’ published in Mechanics Research Communications.

* Corresponding author.
E-mail address: hagen.holthusen@fau.de (H. Holthusen).

https://doi.org/10.1016/j.mechrescom.2026.104707
Received 28 August 2025; Received in revised form 7 April 2026; Accepted 16 April 2026
Available online 21 April 2026

a	scalar	$\cosh(\bullet)$	hyperbolic cosine	$\mathrm{I}_1^{(\bullet)}$	$\mathrm{tr}(\bullet)$
$\boldsymbol{a}$	first order tensor	$\det(\bullet)$	determinant	$\mathrm{I}_2^{(\bullet)}$	$\left(\mathrm{tr}^2(\bullet) - \mathrm{tr}\left((\bullet)^2\right)\right)/2$
$\boldsymbol{A}$	second order tensor	$\mathrm{dev}(\bullet)$	deviatoric part	$\mathrm{I}_3^{(\bullet)}$	$\det(\bullet)$
$(\bullet)^{\mathrm{T}}$	transpose	$\exp(\bullet)$	exponential	$\tilde{\mathrm{I}}_1^{(\bullet)}$	$\mathrm{I}_1(\bullet)/\det^{1/3}(\bullet)$
$(\bullet)^{-1}$	inverse	$\ln(\bullet)$	logarithm	$\tilde{\mathrm{I}}_2^{(\bullet)}$	$\mathrm{I}_2(\bullet)/\det^{2/3}(\bullet)$
$\dot{(\bullet)}$	time derivative	$\mathrm{sym}(\bullet)$	symmetric part	$\mathrm{J}_2^{(\bullet)}$	$\mathrm{tr}\left(\mathrm{dev}^2(\bullet)\right)/2$
$\langle \bullet \rangle$	Macaulay bracket	$\mathrm{tr}(\bullet)$	trace	$\mathrm{J}_3^{(\bullet)}$	$\mathrm{tr}\left(\mathrm{dev}^3(\bullet)\right)/3$
$\lvert \bullet \rvert$	absolute value				
$\cdot$	single contraction				
$:$	double contraction				

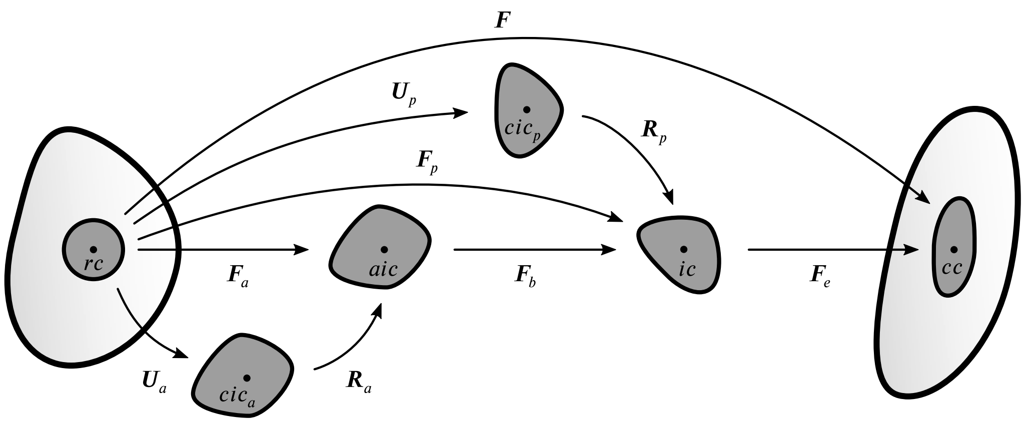


Fig. 1. Illustration of the multiplicative decomposition of the deformation gradient. The abbreviations denote the reference configuration (rc), the current configuration (cc), the intermediate configuration (ic), the additional intermediate configuration (aic), the co-rotated intermediate configuration related to $\boldsymbol{F}_p$ (cic_p), and the co-rotated intermediate configuration related to $\boldsymbol{F}_a$ (cic_a) (cf. [10]).

where $\boldsymbol{C}_e$ denotes the elastic right Cauchy–Green tensor and $\boldsymbol{B}_p$ and $\boldsymbol{B}_b$ denote the inelastic left Cauchy–Green tensors. To ensure that the formulation is independent of rigid body rotations of the reference and of the current configuration (objectivity), we assume the Helmholtz free energy $\psi = \hat{\psi}(\boldsymbol{C}_e, \boldsymbol{B}_p, \boldsymbol{B}_b, \kappa)$ to be an isotropic function of the stretch tensors $\boldsymbol{C}_e$, $\boldsymbol{B}_p$, and $\boldsymbol{B}_b$ as well as the accumulated plastic strain κ.

Co-rotated intermediate configuration. The stretch tensors in the intermediate configuration ic still suffer from the problem of rotational non-uniqueness. Therefore, we follow the concept of [10] and pull the quantities from the arbitrary intermediate configuration ic to the uniquely defined co-rotated intermediate configuration cic_p. The similarity transformation $\bar{(\bullet)} = \boldsymbol{R}_p{}^{\mathrm{T}}(\bullet)\boldsymbol{R}_p$ with the proper orthogonal plastic rotation tensor $\boldsymbol{R}_p$ yields the unique stretch tensors

$$\bar{\boldsymbol{C}}_e = \boldsymbol{R}_p{}^{\mathrm{T}}\boldsymbol{C}_e\boldsymbol{R}_p = \boldsymbol{U}_p{}^{-1}\boldsymbol{C}\boldsymbol{U}_p{}^{-1}, \tag{4}$$

$$\bar{\boldsymbol{B}}_p = \boldsymbol{R}_p{}^{\mathrm{T}}\boldsymbol{B}_p\boldsymbol{R}_p \equiv \boldsymbol{C}_p, \tag{5}$$

$$\bar{\boldsymbol{B}}_b = \boldsymbol{R}_p{}^{\mathrm{T}}\boldsymbol{B}_b\boldsymbol{R}_p = \boldsymbol{U}_p\boldsymbol{C}_a{}^{-1}\boldsymbol{U}_p \tag{6}$$

that are similar to their counterparts in the intermediate configuration. Therefore, we can equivalently define the Helmholtz free energy to be an isotropic function of the stretch tensors in the co-rotated intermediate configuration $\psi = \bar{\psi}(\bar{\boldsymbol{C}}_e, \boldsymbol{C}_p, \bar{\boldsymbol{B}}_b, \kappa)$ now depending on unique arguments.

Clausius–Planck inequality. To ensure a physically sound model formulation, the Clausius–Planck inequality has to be fulfilled for arbitrary processes and reads

$$-\dot{\psi} + \frac{1}{2}\boldsymbol{S} : \dot{\boldsymbol{C}} \geq 0 \tag{7}$$

where $\boldsymbol{S}$ denote the second Piola–Kirchhoff stress tensor and $\boldsymbol{C}$ the right Cauchy–Green stretch tensor. The rate of the Helmholtz free energy follows from the rates of its arguments as

$$\dot{\psi} = \frac{\partial \psi}{\partial \bar{\boldsymbol{C}}_e} : \dot{\bar{\boldsymbol{C}}}_e + \frac{\partial \psi}{\partial \boldsymbol{C}_p} : \dot{\boldsymbol{C}}_p + \frac{\partial \psi}{\partial \bar{\boldsymbol{B}}_b} : \dot{\bar{\boldsymbol{B}}}_b + \frac{\partial \psi}{\partial \kappa}\dot{\kappa}. \tag{8}$$

Next, with the definition from [10] for the co-rotated inelastic velocity gradients $\bar{\boldsymbol{L}}_i := \dot{\boldsymbol{U}}_i\boldsymbol{U}_i{}^{-1}$ and their symmetric parts $\bar{\boldsymbol{D}}_i := \mathrm{sym}\left(\bar{\boldsymbol{L}}_i\right) = \frac{1}{2}\boldsymbol{U}_i{}^{-1}\dot{\boldsymbol{C}}_i\boldsymbol{U}_i{}^{-1}$ with $i \in (p, a)$, the rates of the co-rotated stretch tensors follow as

$$\dot{\bar{\boldsymbol{C}}}_e = \boldsymbol{U}_p{}^{-1}\dot{\boldsymbol{C}}\boldsymbol{U}_p{}^{-1} - \bar{\boldsymbol{L}}_p{}^{\mathrm{T}}\bar{\boldsymbol{C}}_e - \bar{\boldsymbol{C}}_e\bar{\boldsymbol{L}}_p{}^{\mathrm{T}}, \tag{9}$$

$$\dot{\boldsymbol{C}}_p = \bar{\boldsymbol{L}}_p\boldsymbol{C}_p + \boldsymbol{C}_p\bar{\boldsymbol{L}}_p{}^{\mathrm{T}}, \tag{10}$$

$$\dot{\bar{\boldsymbol{B}}}_b = -\boldsymbol{U}_p\boldsymbol{C}_a{}^{-1}\dot{\boldsymbol{C}}_a\boldsymbol{C}_a{}^{-1}\boldsymbol{U}_p + \bar{\boldsymbol{L}}_p\bar{\boldsymbol{B}}_b + \bar{\boldsymbol{B}}_b\bar{\boldsymbol{L}}_p{}^{\mathrm{T}}. \tag{11}$$

Inserting the rates from Eqs. (8)–(11) into Eq. (7) and applying the procedure of [11], the constitutive state law for the second Piola–Kirchhoff stress tensor is obtained as

$$\boldsymbol{S} = 2\boldsymbol{U}_p{}^{-1}\frac{\partial \psi}{\partial \bar{\boldsymbol{C}}_e}\boldsymbol{U}_p{}^{-1}. \tag{12}$$

Moreover, with the definition of the stress-like quantities $\bar{\boldsymbol{\Gamma}}$, $\breve{\boldsymbol{\Theta}}$, and R as

$$\bar{\boldsymbol{\Gamma}} := 2\bar{\boldsymbol{C}}_e\frac{\partial \psi}{\partial \bar{\boldsymbol{C}}_e} - 2\frac{\partial \psi}{\partial \bar{\boldsymbol{B}}_p}\bar{\boldsymbol{B}}_p - 2\frac{\partial \psi}{\partial \bar{\boldsymbol{B}}_b}\bar{\boldsymbol{B}}_b, \tag{13}$$

$$\breve{\boldsymbol{\Theta}} := 2\boldsymbol{U}_a{}^{-1}\boldsymbol{U}_p\frac{\partial \psi}{\partial \bar{\boldsymbol{B}}_b}\boldsymbol{U}_p\boldsymbol{U}_a{}^{-1}, \tag{14}$$

$$R := -\frac{\partial \psi}{\partial \kappa}, \tag{15}$$

the reduced dissipation inequality yields

$$\mathcal{D}_{red} := \bar{\boldsymbol{\Gamma}} : \bar{\boldsymbol{D}}_p + \breve{\boldsymbol{\Theta}} : \bar{\boldsymbol{D}}_a + R\dot{\kappa} \geq 0. \tag{16}$$

For the details of the derivation, we kindly refer to the works of e.g. [10,12,13].

Thermodynamically consistent evolution equations. For the case of elasto-plasticity with nonlinear kinematic hardening at finite strains, the choice of thermodynamically consistent evolution was presented in [3]. Here, we generalize the framework in order to account

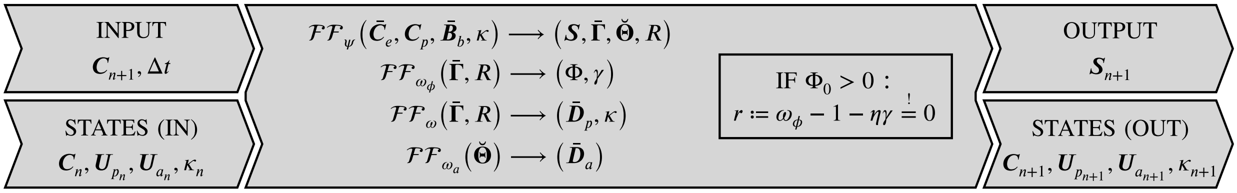


Fig. 2. Illustration of recurrent network architecture for every time step.

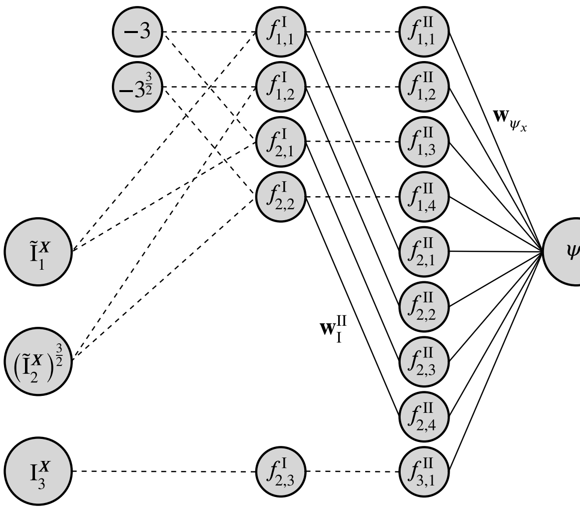


Fig. 3. Feed-forward network structure for the elastic energy and the plastic energies related to kinematic hardening with $\boldsymbol{X} \in (\bar{\boldsymbol{C}}_e, \boldsymbol{C}_p, \bar{\boldsymbol{B}}_b)$ and $\psi_x \in (\psi_e, \psi_p, \psi_b)$. The activation functions read $f^{\mathrm{I}} \in (x, x^2)$ and $f^{\mathrm{II}} \in (x, \exp(x) - 1, x^{w_{vol}} - 1 - \ln(x^{w_{vol}}))$. The dashed lines indicate fixed unity weights.

for non-associative elasto-visco-plasticity with nonlinear kinematic and nonlinear isotropic hardening at finite strains. Following [14,15], we assume potential-based evolution equations and postulate the following zero-valued, convex, and non-negative inelastic potentials based on the invariants of the thermodynamic driving forces $\bar{\boldsymbol{\Gamma}}$ and $\breve{\boldsymbol{\Theta}}$ as well as the thermodynamic driving force for isotropic hardening R as

$$\omega = \omega \left(\mathrm{I}_1^{\bar{\boldsymbol{\Gamma}}}, \sqrt{\mathrm{J}_2^{\bar{\boldsymbol{\Gamma}}}}, \sqrt[3]{\mathrm{J}_3^{\bar{\boldsymbol{\Gamma}}}}, R \right), \tag{17}$$

$$\omega_a = \omega_a \left(\mathrm{I}_1^{\breve{\boldsymbol{\Theta}}}, \sqrt{\mathrm{J}_2^{\breve{\boldsymbol{\Theta}}}}, \sqrt[3]{\mathrm{J}_3^{\breve{\boldsymbol{\Theta}}}} \right) \tag{18}$$

and define the corresponding evolution equations as

$$\bar{\boldsymbol{D}}_p = \gamma \frac{\partial \omega}{\partial \bar{\boldsymbol{\Gamma}}}, \quad \bar{\boldsymbol{D}}_a = \gamma \frac{\partial \omega_a}{\partial \breve{\boldsymbol{\Theta}}}, \quad \dot{\kappa} = \gamma \frac{\partial \omega}{\partial R}. \tag{19}$$

For covering the general case of non-associative plasticity and for modeling the onset of plasticity, we introduce a further zero-valued, convex, and non-negative inelastic potential

$$\omega_\phi = \omega_\phi \left(\mathrm{I}_1^{\bar{\boldsymbol{\Gamma}}}, \sqrt{\mathrm{J}_2^{\bar{\boldsymbol{\Gamma}}}}, \sqrt[3]{\mathrm{J}_3^{\bar{\boldsymbol{\Gamma}}}}, R \right) \tag{20}$$

that serves for a general definition of the yield function

$$\Phi = \omega_\phi - \sigma_{Y0}. \tag{21}$$

The plastic multiplier is formulated in a rate-dependent approach (cf. [16])

$$\gamma = \frac{1}{\eta} \frac{\langle \Phi \rangle}{\sigma_{Y0}} \geq 0 \tag{22}$$

with the plastic relaxation time η and the initial yield stress σ_{Y0}. Thereafter, we introduce the vector $\boldsymbol{z}$ containing the arguments of the potential ω

$$\boldsymbol{z} := \begin{pmatrix} \mathrm{I}_1^{\bar{\boldsymbol{\Gamma}}} \\ \sqrt{\mathrm{J}_2^{\bar{\boldsymbol{\Gamma}}}} \\ \sqrt[3]{\mathrm{J}_3^{\bar{\boldsymbol{\Gamma}}}} \\ R \end{pmatrix} \tag{23}$$

and, in analogy to the derivation in [17], we utilize that

$$\bar{\boldsymbol{\Gamma}} : \bar{\boldsymbol{D}}_p + R\dot{\kappa} = \frac{\partial \omega}{\partial \boldsymbol{z}} \cdot \boldsymbol{z} \geq \omega(\boldsymbol{z}) \geq 0 \tag{24}$$

holds, if ω is constructed convex with respect to its arguments. Hence, by inserting Equation (24) into Eq. (16), thermodynamic consistency for the elasto-visco-plastic model with nonlinear kinematic and nonlinear isotropic hardening is proven, viz.

$$\mathcal{D}_{red} = \underbrace{\gamma}_{\geq 0} \underbrace{\breve{\boldsymbol{\Theta}} : \frac{\partial \omega_a}{\partial \breve{\boldsymbol{\Theta}}}}_{\geq 0} + \underbrace{\gamma}_{\geq 0} \underbrace{\boldsymbol{z} \cdot \frac{\partial \omega}{\partial \boldsymbol{z}}}_{\geq 0} \geq 0. \tag{25}$$

Here, $\breve{\boldsymbol{\Theta}} : \partial \omega_a / \partial \breve{\boldsymbol{\Theta}} \geq 0$ holds due to the same reasoning as in Eq. (24).

3. Network architecture

Analogously to inelastic constitutive artificial neural networks for other material phenomena (see e.g. [18] for viscoelasticity and [19] for growth), the architecture of the present framework for plasticity embeds multiple feed-forward networks into an recurrent architecture as depicted schematically in Fig. 2.

The employed feed-forward networks are deliberately structured in accordance with the Constitutive Artificial Neural Network (CANN) paradigm and reflect classical constitutive modeling principles, rather than fully unconstrained universal approximation. The architecture remains systematically extensible by increasing the number of neurons and activation functions, see e.g. [20].

Feed-forward networks – energies. The Helmholtz free energy $\psi(\bar{\boldsymbol{C}}_e, \boldsymbol{C}_p, \bar{\boldsymbol{B}}_b, \kappa)$ is additively decomposed into four separate energies

$$\psi(\bar{\boldsymbol{C}}_e, \boldsymbol{C}_p, \bar{\boldsymbol{B}}_b, \kappa) = \psi_e(\bar{\boldsymbol{C}}_e) + \psi_p(\boldsymbol{C}_p) + \psi_b(\bar{\boldsymbol{B}}_b) + \psi_\kappa(\kappa) \tag{26}$$

where ψ_e is related to the elastic energy, ψ_p to the plastic energy with respect to linear kinematic hardening, ψ_b to the plastic energy with respect to nonlinear kinematic hardening, and ψ_κ to the plastic energy with respect to isotropic hardening. In line with [5,21], feed-forward networks based on a volumetric-deviatoric split (see Fig. 3) are employed for the elastic energy as well as for the plastic energies related to kinematic hardening (cf. [3]).

To account for isotropic hardening, we introduce a new network that maps the accumulated plastic strain κ on the associated plastic energy related to isotropic hardening ψ_κ as depicted in Fig. 4.

The additive decomposition of the free energy into elastic and hardening contributions follows standard finite-strain plasticity formulations and allows for a transparent separation of elastic storage and

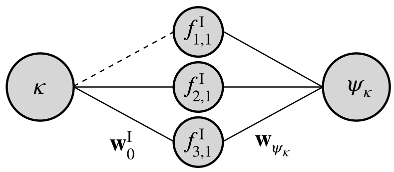


Fig. 4. Feed-forward network structure for the plastic energy related to isotropic hardening ψ_κ. The activation functions read $f^{\mathrm{I}} \in \left(x^2, \ln(\cosh(|w_{ln.cosh}x|)), \cosh(|w_{cosh}x|) - 1\right)$. The dashed line indicates a fixed unity weight.

internal variable evolution. This choice assumes energetic decoupling of the individual hardening mechanisms and excludes cross-coupling terms between isotropic and kinematic hardening. While sufficient for the present proof-of-concept extension, more general formulations could incorporate non-separable energy contributions, see [20].

Feed-forward networks – potentials. Analogously to the works in [3,17–19,22], we utilize feed-forward networks to identify the relation between the thermodynamic driving forces and the inelastic potentials. According to Eqs. (17), (18), and (20), three feed-forward networks for the three inelastic potentials are employed (see Fig. 5). Here, only ω and ω_ϕ require the thermodynamic driving force R conjugate to the isotropic hardening rate $\dot{\kappa}$. Moreover, the hidden layers feature flexible formulations with respect to the number of neurons per activation function in the potential related networks, while the number of neurons is fixed for the energy related networks.

Time discretization and Newton–Raphson iteration. For the evolution of the internal tensorial variables $\boldsymbol{C}_p$ and $\boldsymbol{C}_a$ the explicit exponential map integrator scheme is applied

$$\boldsymbol{C}_i = \boldsymbol{U}_{i,n} \exp\left(2\, \Delta t\, \gamma\, \frac{\partial \omega}{\partial \boldsymbol{X}_n}\right) \boldsymbol{U}_{i,n} \tag{27}$$

with $\boldsymbol{X} = \bar{\boldsymbol{\Gamma}}$ for $i = p$ and $\boldsymbol{X} = \check{\boldsymbol{\Theta}}$ for $i = a$ and the forward Euler method for the evolution of the accumulated plastic strain κ. In our experience, the explicit integration scheme has not led to noticeable stability issues during training. Avoiding differentiation through a Newton iteration results in a simpler computational graph and was observed to be advantageous for gradient-based optimization. Although conditional stability is a general characteristic of explicit schemes, it did not appear to be the limiting factor in the present experiments. We note that recent studies have demonstrated successful training with implicit integration schemes in related viscoelastic settings (e.g. [23]). Such approaches, however, typically increase the computational cost per training step. A systematic comparison between explicit and implicit integration strategies in the present plasticity setting is beyond the scope of this work.

Furthermore, the modeling of plasticity requires to distinguish between an elastic and an inelastic step. Thus, the yield function from Eq. (21) has to be evaluated at the beginning of every time step to determine whether plastic flow occurs (see Fig. 2 and Algorithm 1). In case of an inelastic step, the identification of the plastic multiplier from Eq. (22) requires an iterative solution using the Newton–Raphson method for the residual

$$r := \hat{\omega}_\phi - 1 - \gamma\eta \overset{!}{=} 0 \tag{28}$$

where $\hat{\omega}_\phi = \omega_\phi / \sigma_{Y0}$ is employed. For the Jacobian, a numerical tangent has been implemented. The entire procedure to determine the plastic multiplier is presented in Algorithm 1.

Algorithm 1: Determination of plastic multiplier

Input: $\boldsymbol{U}_{p,n}$, $\boldsymbol{U}_{a,n}$, κ_n, Δt, $\dfrac{\partial \omega}{\partial \bar{\boldsymbol{\Gamma}}_n}$, $\dfrac{\partial \omega}{\partial R_n}$, $\dfrac{\partial \omega_a}{\partial \check{\boldsymbol{\Theta}}_n}$, $\boldsymbol{C}$, ϵ_Φ, ϵ_r, ϵ_γ

$\gamma \leftarrow 0$
if $\Phi > \epsilon_\Phi$ **then**
  for $i = 1$ **to** 30 **do**
    $\kappa \leftarrow \kappa_n + \Delta t \dfrac{\partial \omega}{\partial R_n}$
    $\boldsymbol{C}_p \leftarrow \boldsymbol{U}_{p,n} \exp\left(2\, \Delta t\, \gamma\, \dfrac{\partial \omega}{\partial \bar{\boldsymbol{\Gamma}}_n}\right) \boldsymbol{U}_{p,n}$
    $\boldsymbol{U}_p \leftarrow +\sqrt{\boldsymbol{C}_p}$
    $\bar{\boldsymbol{C}}_e \leftarrow \boldsymbol{U}_p^{-1} \boldsymbol{C} \boldsymbol{U}_p^{-1}$
    $\boldsymbol{C}_a \leftarrow \boldsymbol{U}_{a,n} \exp\left(2\, \Delta t\, \gamma\, \dfrac{\partial \omega_a}{\partial \check{\boldsymbol{\Theta}}_n}\right) \boldsymbol{U}_{a,n}$
    $\boldsymbol{U}_a \leftarrow +\sqrt{\boldsymbol{C}_a}$
    $\bar{\boldsymbol{B}}_b \leftarrow \boldsymbol{U}_p \boldsymbol{C}_a^{-1} \boldsymbol{U}_p$
    $\bar{\boldsymbol{\Gamma}} \leftarrow 2\bar{\boldsymbol{C}}_e \dfrac{\partial \psi}{\partial \bar{\boldsymbol{C}}_e} - 2\dfrac{\partial \psi}{\partial \boldsymbol{C}_p} \boldsymbol{C}_p - 2\dfrac{\partial \psi}{\partial \bar{\boldsymbol{B}}_b} \bar{\boldsymbol{B}}_b$
    $R \leftarrow -\dfrac{\partial \psi}{\partial \kappa}$
    $r \leftarrow \hat{\omega}_\phi - 1 - \gamma\eta$
    if $|r| < \epsilon_r$ **then**
      break
    $r' \leftarrow (r(\gamma + \epsilon_\gamma) - r(\gamma))/\epsilon_\gamma$
    $\gamma \leftarrow \gamma - r/r'$
return γ

4. Numerical example

The training procedure combines a pretraining stage with the subsequent optimization phase (see Algorithm 2). This strategy is motivated by our empirical observations during training, where the loss landscape in the plasticity setting appeared to be less smooth and more sensitive to initialization compared to previous iCANN applications to viscoelasticity and growth, which did not employed an `if`-statement in the implementation.[1]

The pretraining stage is therefore used to identify a suitable starting point by performing multiple evaluations of the loss function with randomly perturbed weights and biases within a predefined range. During the subsequent training phase, the occurrence of NaN values in either the parameters or the loss function is handled by applying a small random perturbation to the previously best parameter state in order to re-enter a descending region of the loss landscape.

These heuristics are not intended as a general remedy but were found to improve robustness in our experiments. For implementation details, the interested reader is referred to the *execute_training.py* file in [6].

The following example is intended as a proof-of-concept demonstration of the extended constitutive framework. While the deformation levels considered here remain moderate, the formulation itself is fully finite-strain consistent due to the multiplicative kinematics and invariant-based potential structure introduced in Section 2.

The loss for the numerical example is defined as the mean squared error between the network model's stress response $\boldsymbol{S}(\boldsymbol{C}_t)$ and the stress training data $\hat{\boldsymbol{S}}_t$.

$$\mathcal{L} = \frac{1}{n_{\text{data}}} \sum_{t=1}^{n_{\text{data}}} \left(\boldsymbol{S}(\boldsymbol{C}_t) - \hat{\boldsymbol{S}}_t\right)^2 \tag{29}$$

The implementation also features the options to include L_1 and L_2 regularization for which we again kindly refer to [6] and suggest regularization studies as an interesting supplement for future works.

The numerical example in Figs. 6 and 7 is trained on synthetic data generated from the finite-strain plasticity model described in Appendix.

[1] We tested a Fischer–Burmeister reformulation of the yield condition to avoid an explicit conditional activation. However, this did not lead to a substantial improvement in training robustness in our experiments.

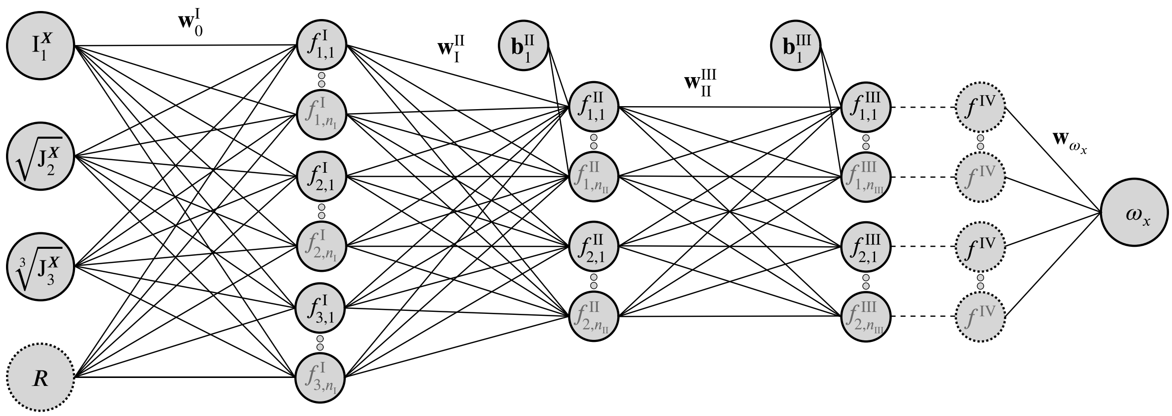
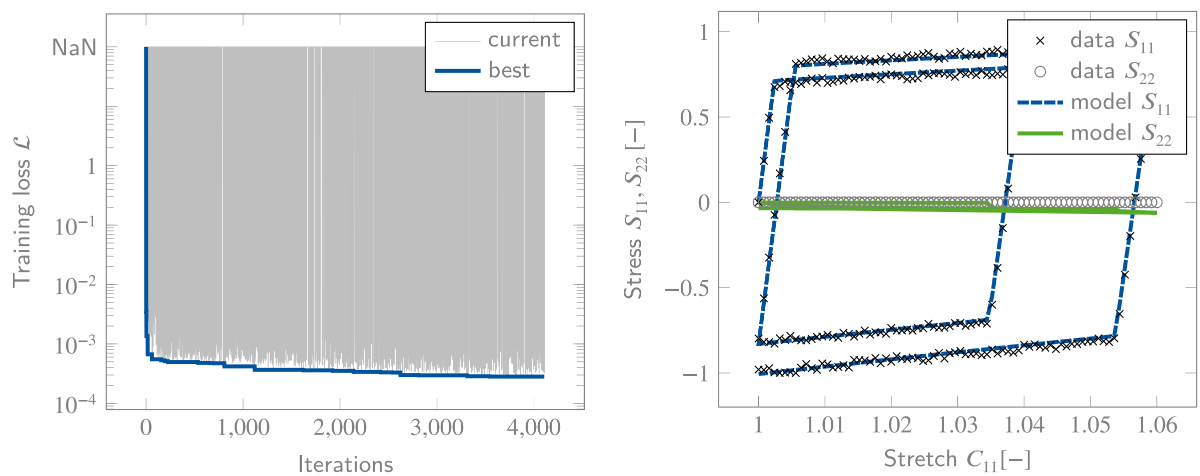


Fig. 5. Feed-forward network structure for the inelastic potentials with $\boldsymbol{X} \in (\bar{\boldsymbol{\Gamma}}, \check{\boldsymbol{\Theta}})$ and $\omega_x \in (\omega, \omega_a, \omega_\phi)$. Only the potentials ω and ω_ϕ utilize the driving force R as input. The activation functions read $f^{\mathrm{I}} \in (x, |x|^{w_{pow1}+1}, \ln(\cosh(|x^{w_{pow2}+1}|)))$, $f^{\mathrm{II}}, f^{\mathrm{III}} \in (\max(x,0), \exp(x)-1)$ and $f^{\mathrm{IV}} \in (\max(x,0))$. The dashed lines indicate fixed unity weights.



(a) Training loss (b) Data and model stresses

Fig. 6. (a) Evolution of the training loss $\mathcal{L}$. (b) Stress components S_{11} (loading direction) and S_{22} (transverse direction) under uniaxial deformation, comparing data and model prediction after training. The model accurately reproduces the stress in loading direction while maintaining negligible transverse stresses, correctly capturing the uniaxial stress condition without explicitly constraining the transverse response. Results are shown for the potential configuration $n_{\mathrm{I}} = n_{\mathrm{II}} = n_{\mathrm{III}} = 3$ with 100 pretraining iterations.

The training data are fully specified through explicitly defined energy functions, inelastic potentials, material parameters, and loading conditions, ensuring reproducibility of the results. The dataset is obtained from a cyclic uniaxial loading test with increasing strain amplitude. The constitutive equations are integrated using an explicit exponential integration scheme, and the resulting stress response is recorded at each time step. To mimic measurement uncertainty and assess robustness, the axial stress component is additionally perturbed by multiplicative uniform noise, as detailed in Appendix.

The inelastic potential networks employ three neurons per activation function ($n_{\mathrm{I}} = n_{\mathrm{II}} = n_{\mathrm{III}} = 3$), resulting in a total of 459 trainable parameters (weights and biases) for the complete architecture.

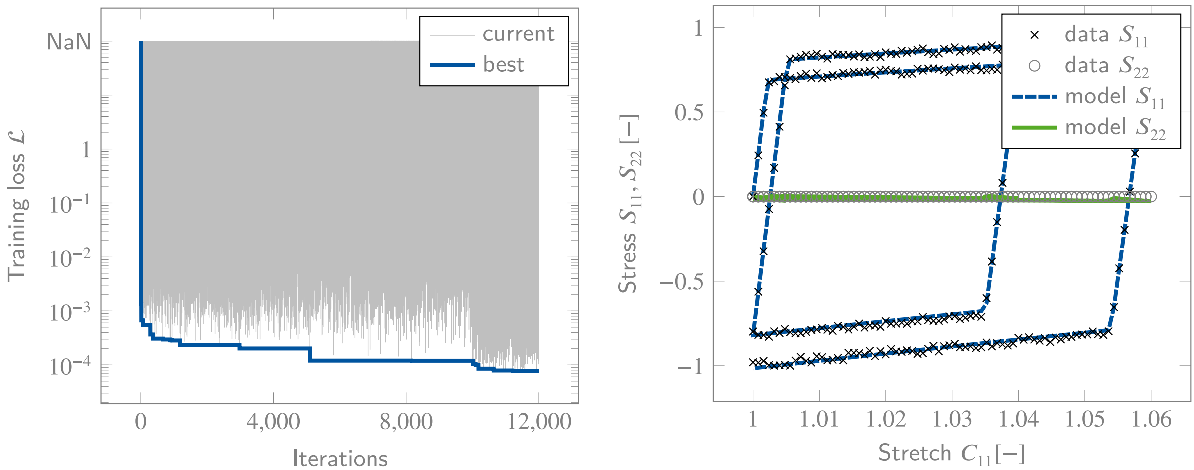
In Fig. 6, the number of pretraining iterations is $n_{\mathrm{pre}} = 100$, with $n_{\mathrm{pert}} = 500$ perturbation steps during training. The final loss amounts to $\mathcal{L} = 2.82 \times 10^{-4}$ (Fig. 6(a)). While the stress component in loading direction (S_{11}) is accurately captured, the transverse stress component (S_{22}) exhibits small but noticeable nonzero values (Fig. 6(b)). This behavior is reflected in the corresponding mean squared errors reported in Table 1, where the error in S_{22} is significantly larger than that in S_{11}, indicating an incomplete identification of the parameters governing lateral contraction.

Importantly, the transverse stress condition $S_{22} = 0$ was not explicitly enforced during training, e.g., by constraining lateral stresses in the loss function. Its approximate recovery therefore results from the learned constitutive response, indicating that the network identifies physically consistent material parameters governing the transverse behavior.

A straightforward remedy is provided by adapting the parameters of the solution strategy, e.g. by increasing the number of pretraining iterations. Setting $n_{\mathrm{pre}} = 10{,}000$ yields a final loss of $\mathcal{L} = 7.79 \times 10^{-5}$ (Fig. 7(a)) and a markedly improved stress response in S_{22} (Fig. 7(b)). As quantified in Table 1, the MSE of S_{22} decreases from 1.31×10^{-3} to 2.42×10^{-4}, demonstrating a substantially improved identification of the material parameters governing the transverse response, while the error in S_{11} is reduced more moderately.

5. Conclusion

We presented a thermodynamically consistent constitutive framework for finite strain plasticity with combined isotropic and nonlinear



(a) Training loss (b) Data and model stresses

Fig. 7. (a) Evolution of the training loss $\mathcal{L}$ for 10000 pretraining iterations, illustrating improved initialization and faster convergence compared to Figs. 6(a) and 6(b). (b) Stress components S_{11} (loading direction) and S_{22} (transverse direction) under uniaxial deformation, comparing data and model prediction. The increased number of pretraining iterations leads to a more stable identification of the constitutive parameters, resulting in an accurate stress prediction in loading direction while consistently preserving negligible transverse stresses. Results are shown for the potential configuration $n_{\mathrm{I}} = n_{\mathrm{II}} = n_{\mathrm{III}} = 3$.

Algorithm 2: Solution strategy

 # pre-training
$\mathcal{L}_{\mathrm{best}} \leftarrow \inf, \mathbf{w}_{\mathrm{best}} \leftarrow$ none
for $i = 1$ **to** n_{pre} **do**
  $\mathbf{w}_{\mathrm{rand}} \leftarrow \mathrm{random}(\mathbf{w}_{\min}, \mathbf{w}_{\max})$
  $\mathcal{L}_{\mathrm{pre}} \leftarrow \mathcal{L}(\mathbf{w}_{\mathrm{rand}})$
  if $\mathcal{L}_{\mathrm{pre}} < \mathcal{L}_{\mathrm{best}}$ **then**
    $\mathcal{L}_{\mathrm{best}}, \mathbf{w}_{\mathrm{best}} \leftarrow \mathcal{L}_{\mathrm{pre}}, \mathbf{w}_{\mathrm{rand}}$
if contains_nan($\mathcal{L}_{\mathrm{best}}, \mathbf{w}_{\mathrm{best}}$) **then**
  error stop: *'no starting solution found'*
 # training
$\mathbf{w} \leftarrow \mathbf{w}_{\mathrm{best}}$
for $i = 1$ **to** n_{epoch} **do**
  $\mathcal{L}, \mathbf{w} \leftarrow$ training_step($\mathbf{w}$)
  if ($\mathcal{L} < \mathcal{L}_{\mathrm{best}}$) **and** (**not** contains_nan($\mathcal{L}, \mathbf{w}$)) **then**
    $\mathcal{L}_{\mathrm{best}}, \mathbf{w}_{\mathrm{best}} \leftarrow \mathcal{L}, \mathbf{w}$
  else if contains_nan($\mathcal{L}, \mathbf{w}$) **then**
    # parameter perturbation
    $\mathcal{L}_{\mathrm{pert\_best}} \leftarrow \inf, \mathbf{w}_{\mathrm{pert\_best}} \leftarrow$ none
    for $p = 1$ **to** n_{pert} **do**
      $\mathbf{w}_{\mathrm{pert}} \leftarrow \mathrm{perturbation}(\mathbf{w}_{\mathrm{best}})$
      $\mathcal{L}_{\mathrm{pert}} \leftarrow \mathcal{L}(\mathbf{w}_{\mathrm{pert}})$
      if ($\mathcal{L}_{\mathrm{pert}} < \mathcal{L}_{\mathrm{pert\_best}}$) **and** (**not** contains_nan($\mathcal{L}_{\mathrm{pert}}$)) **then**
        $\mathcal{L}_{\mathrm{pert\_best}}, \mathbf{w}_{\mathrm{pert\_best}} \leftarrow \mathcal{L}_{\mathrm{pert}}, \mathbf{w}_{\mathrm{pert}}$
    if $\mathcal{L}_{\mathrm{pert\_best}} < \mathcal{L}_{\mathrm{best}}$ **then**
      $\mathcal{L}_{\mathrm{best}}, \mathbf{w}_{\mathrm{best}} \leftarrow \mathcal{L}_{\mathrm{pert\_best}}, \mathbf{w}_{\mathrm{pert\_best}}$
    else
      error stop: *'$\mathcal{L}$, $\mathbf{w}$ still contain NaN after perturbation'*

Table 1
Mean squared error (MSE) between model prediction and experimental data for different numbers of pretraining iterations.

Pretraining iterations	MSE S_{11}	MSE S_{22}
100	4.05×10^{-4}	1.31×10^{-3}
10000	2.65×10^{-4}	2.42×10^{-4}

kinematic hardening. The continuum mechanical model additionally accounts for non-associative plasticity and a viscoplastic Perzyna-type evolution. Thermodynamic consistency is ensured by the introduction of a shared inelastic potential for isotropic and linear kinematic hardening.

Moreover, we introduced a recurrent neural network architecture in the spirit of inelastic constitutive artificial neural networks for the plastic model discovery.

Although the illustrative example involves moderate deformation levels, the proposed framework is formulated entirely within a multiplicative finite-strain setting. A systematic investigation under pronounced large-deformation loading paths constitutes a natural continuation of the present work.

Further work could address an enhancement of the loss function by the plastic dissipation to improve the identification of elastic and plastic regimes during training. Additionally, the effect of regularization, the hyperparameters, and the optimizer along with the solution strategy should be investigated.

CRediT authorship contribution statement

Tim van der Velden: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Conceptualization. **Birte Boes:** Writing – review & editing, Software, Methodology, Conceptualization. **Tim Brepols:** Writing – review & editing, Methodology, Funding acquisition. **Ellen Kuhl:** Writing – review & editing, Funding acquisition. **Hagen Holthusen:** Writing – review & editing, Writing – original draft, Software, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Funding granted by the German Research Foundation (DFG) is gratefully acknowledged by T. Brepols for projects 453596084 (B05) and 453715964, and by H. Holthusen for project 417002380 (A01).

Furthermore, funding granted by the NSF CMMI Award 2320933 Automated Model Discovery for Soft Matter and by the ERC Advanced Grant 101141626 DISCOVER is gratefully acknowledged by E. Kuhl.

Appendix. Constitutive material model

We generate the training data for the numerical example using the following constitutive model. In accordance with Eq. (26), the Helmholtz free energy is additively decomposed into elastic, plastic, backstress, and isotropic hardening contributions,

$$\psi_{\mathrm{el}}(\bar{\boldsymbol{C}}_e) = \frac{\mu_e}{2}\left(\frac{\mathrm{tr}(\bar{\boldsymbol{C}}_e)}{\det(\bar{\boldsymbol{C}}_e)^{1/3}} - 3\right) + \frac{K_e}{4}\left(\det(\bar{\boldsymbol{C}}_e) - 1 - \ln\det(\bar{\boldsymbol{C}}_e)\right), \tag{30}$$

$$\psi_{\mathrm{p}}(\boldsymbol{C}_p) = c\left(\frac{\mathrm{tr}(\boldsymbol{C}_p)}{\det(\boldsymbol{C}_p)^{1/3}} - 3\right), \tag{31}$$

$$\psi_{\mathrm{b}}(\boldsymbol{B}_b) = a\left(\frac{\mathrm{tr}(\boldsymbol{B}_b)}{\det(\boldsymbol{B}_b)^{1/3}} - 3\right), \tag{32}$$

$$\psi_{\kappa}(\kappa) = e\left(\kappa + \frac{\exp(-s\,\kappa) - 1}{s}\right) + \frac{1}{2}H\kappa^2. \tag{33}$$

Here, $\mathrm{tr}(\cdot)$ and $\det(\cdot)$ denote the trace and determinant of a tensor, respectively.

The material parameters are chosen as $\mu_e = 7500.0$ MPa, $K_e = 10000.0$ MPa, $c = 10.0$ MPa, $a = 10.0$ MPa, $e = 0.0$ MPa, $s = 10.0$ [-], and $H = 100.0$ MPa.

The dissipation potentials are defined as

$$\omega(\bar{\boldsymbol{\Gamma}}, R) = \sqrt{\frac{3}{2}}\sqrt{\mathrm{tr}\left(\mathrm{dev}(\bar{\boldsymbol{\Gamma}})^2\right)} + R, \tag{34}$$

$$\omega_{\phi}(\bar{\boldsymbol{\Gamma}}, R) = \sqrt{\frac{3}{2}}\sqrt{\mathrm{tr}\left(\mathrm{dev}(\bar{\boldsymbol{\Gamma}})^2\right)} + R, \tag{35}$$

$$\omega_a(\breve{\boldsymbol{\Theta}}) = b\sqrt{\mathrm{tr}\left(\mathrm{dev}(\breve{\boldsymbol{\Theta}})^2\right)}, \tag{36}$$

where $\mathrm{dev}(\cdot)$ denotes the deviatoric part of a tensor. The parameter is chosen as $b = 10.0$ [-]. Furthermore, the viscosity parameter is set to $\eta = 1.0\ MPa{\cdot}$s and the initial yield stress to $\sigma_{Y0} = 20.0$ MPa.

Data generation. The training data are generated from a uniaxial loading test with cyclic deformation. The total simulation time is chosen as $T = 10.0$ s and discretized into $N = 250$ time steps, resulting in a constant time increment of $\Delta t = T/N = 0.04$ s. The model is implemented according to Algorithm 1.

The loading path consists of two symmetric strain cycles with increasing amplitude. The first cycle comprises 100 time steps with a maximum Green–Lagrange strain of $E_{11}^{\max,1} = 0.02$. The second cycle comprises 150 time steps with an increased maximum strain of $E_{11}^{\max,2} = 0.03$. In each cycle, the strain is increased linearly to its maximum value and subsequently decreased symmetrically, resulting in a triangular loading path.

Only the axial strain component E_{11} is prescribed, while all remaining strain components are determined such that a uniaxial stress state is fulfilled, i.e., $S_{22} = S_{33} = S_{12} = S_{13} = S_{23} = 0$. This constraint is enforced by a Newton iteration at each time step.

For each time step, the following quantities are stored: the right Cauchy–Green tensor $\boldsymbol{C}$ and the second Piola–Kirchhoff stress tensor $\boldsymbol{S}$.

Noisy data. In order to mimic measurement noise and assess the robustness of the learning approach, an additional dataset with perturbed stresses is generated. To this end, the axial stress component is multiplicatively perturbed according to

$$S_{11}^{\mathrm{scat}}(t) = S_{11}(t)\left(1 + \delta(t)\right), \tag{37}$$

where $\delta(t)$ is a uniformly distributed random variable in the interval $[-\alpha, \alpha]$ with $\alpha = 0.025$. This corresponds to a maximum perturbation of $\pm 2.5\%$. The scattered data are subsequently normalized using the maximum absolute value of the perturbed stress signal, i.e.,

$$S_{11,\mathrm{scat}}^{\max} = \max_t \left|S_{11}^{\mathrm{scat}}(t)\right|. \tag{38}$$

Data availability

The source code and data are made accessible to the public via https://doi.org/10.5281/zenodo.16900782.

References

[1] Y. Shen, K. Chandrashekhara, W.F. Breig, L.R. Oliver, Neural network based constitutive model for rubber material, Rubber Chem. Technol. 77 (2) (2004) 257–277, http://dx.doi.org/10.5254/1.3547822.

[2] F. As'ad, P. Avery, C. Farhat, A mechanics-informed artificial neural network approach in data-driven constitutive modeling, Internat. J. Numer. Methods Engrg. 123 (12) (2022) 2738–2759, http://dx.doi.org/10.1002/nme.6957, URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/nme.6957, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/nme.6957.

[3] B. Boes, J.-W. Simon, H. Holthusen, Accounting for plasticity: An extension of inelastic constitutive artificial neural networks, Eur. J. Mech. A Solids 117 (2026) 105998, http://dx.doi.org/10.1016/j.euromechsol.2025.105998, URL: https://www.sciencedirect.com/science/article/pii/S0997753825004322.

[4] A.A. Jadoon, K.A. Meyer, J.N. Fuhg, Automated model discovery of finite strain elastoplasticity from uniaxial experiments, Comput. Methods Appl. Mech. Engrg. 435 (2025) 117653.

[5] K. Linka, M. Hillgärtner, K.P. Abdolazizi, R.C. Aydin, M. Itskov, C.J. Cyron, Constitutive artificial neural networks: A fast and general approach to predictive data-driven constitutive modeling by deep learning, J. Comput. Phys. 429 (2021) 110010.

[6] T. van der Velden, B. Boes, J.W. Simon, T. Brepols, E. Kuhl, H. Holthusen, A note on constitutive artificial neural networks for finite strain plasticity (source code and data), Zenodo (2025) http://dx.doi.org/10.5281/zenodo.16900782.

[7] C. Eckart, The thermodynamics of irreversible processes. IV. The theory of elasticity and anelasticity, Phys. Rev. 73 (4) (1948) 373–382.

[8] E. Kröner, Allgemeine kontinuumstheorie der versetzungen und eigenspannungen, Arch. Ration. Mech. Anal. 4 (1) (1959) 273–334.

[9] A. Lion, Constitutive modelling in finite thermoviscoplasticity: a physical approach based on nonlinear rheological models, Int. J. Plast. 16 (5) (2000) 469–494.

[10] H. Holthusen, C. Rothkranz, L. Lamm, T. Brepols, S. Reese, Inelastic material formulations based on a co-rotated intermediate configuration—Application to bioengineered tissues, J. Mech. Phys. Solids 172 (2023) 105174.

[11] B.D. Coleman, W. Noll, Foundations of linear viscoelasticity, Rev. Modern Phys. 33 (1961) 239–249.

[12] I.N. Vladimirov, M.P. Pietryga, S. Reese, On the modelling of non-linear kinematic hardening at finite strains with application to springback—Comparison of time integration algorithms, Internat. J. Numer. Methods Engrg. 75 (1) (2008) 1–28.

[13] T. Brepols, S. Wulfinghoff, S. Reese, A gradient-extended two-surface damage-plasticity model for large deformations, Int. J. Plast. 129 (2020) 102635.

[14] J. Kestin, J.R. Rice, Paradoxes in the application of thermodynamics to strained solids, in: E.B. Stuart, B. Gal-Or, A.J. Brainard (Eds.), A Critical Review of Thermodynamics, Mono Book Corp., Baltimore, 1970, pp. 275–298.

[15] P. Germain, Q.S. Nguyen, P. Suquet, Continuum thermodynamics, J. Appl. Mech. 50 (1983) 1010–1020.

[16] P. Perzyna, Fundamental problems in viscoplasticity, Adv. Appl. Mech. 9 (1966) 243–377.

[17] H. Holthusen, K. Linka, E. Kuhl, T. Brepols, A generalized dual potential for inelastic constitutive artificial neural networks: A JAX implementation at finite strains, 2025, arXiv, URL: https://arxiv.org/abs/2502.17490.

[18] H. Holthusen, L. Lamm, T. Brepols, S. Reese, E. Kuhl, Theory and implementation of inelastic constitutive artificial neural networks, Comput. Methods Appl. Mech. Engrg. 428 (2024) 117063.

[19] H. Holthusen, T. Brepols, K. Linka, E. Kuhl, Automated model discovery for tensional homeostasis: Constitutive machine learning in growth and remodeling, Comput. Biol. Med. 186 (2025) 109691.

[20] H. Holthusen, E. Kuhl, A complement to neural networks for anisotropic inelasticity at finite strains, Comput. Methods Appl. Mech. Engrg. 450 (2026) 118612, http://dx.doi.org/10.1016/j.cma.2025.118612, URL: https://www.sciencedirect.com/science/article/pii/S0045782525008849.

[21] K. Linka, E. Kuhl, A new family of constitutive artificial neural networks towards automated model discovery, Comput. Methods Appl. Mech. Engrg. 403 (2023) 115731.

[22] H. Holthusen, L. Lamm, T. Brepols, S. Reese, E. Kuhl, Polyconvex inelastic constitutive artificial neural networks, PAMM 24 (3) (2024) e202400032.

[23] K.A. Kalina, J. Brummund, M. Kästner, A physics-augmented neural network framework for finite strain incompressible viscoelasticity, 2025, URL: https://arxiv.org/abs/2511.02959, arXiv:2511.02959.