

Theory and implementation of inelastic Constitutive Artificial Neural Networks

Hagen Holthusen^{a,*}, Lukas Lamm^a, Tim Brepols^a, Stefanie Reese^a, Ellen Kuhl^b

^a Institute of Applied Mechanics, RWTH Aachen University, Mies-van-der-Rohe-Str. 1, 52074 Aachen, Germany

^b Department of Mechanical Engineering, Stanford University, Stanford, CA 94305, United States

ARTICLE INFO

Dataset link: <https://doi.org/10.5281/zenodo.10066805>

Keywords:

Automated model discovery
Hyperelasticity
Viscoelasticity
Constitutive neural networks
Recurrent neural networks
Inelasticity

ABSTRACT

The two fundamental concepts of materials theory, pseudo potentials and the assumption of a multiplicative decomposition, allow a general description of inelastic material behavior. The increase in computer performance enabled us to thoroughly investigate the predictive capabilities of ever more complex choices for the potential and the Helmholtz free energy. Today, however, we have reached a point where materials and their models are becoming increasingly sophisticated. This raises the question: How do we find the best model that includes all inelastic effects to explain our complex data? Constitutive Artificial Neural Networks (CANN) may answer this question. Here, we extend the CANNs to inelastic materials (iCANN). Rigorous considerations of objectivity, rigid motion of the reference configuration, multiplicative decomposition and its inherent non-uniqueness, choice of appropriate stretch tensors, restrictions of energy and pseudo potential, and consistent inelastic evolution guide us towards the general architecture of the iCANN satisfying thermodynamics per design. We combine feed-forward networks of the Helmholtz free energy and pseudo potential with a recurrent neural network approach to take time dependencies into account. Specializing the general iCANN to visco-elasticity, we demonstrate that the iCANN is capable of autonomously discovering models for artificially generated data, the response of polymers at different stretch rates for cyclic loading as well as the relaxation behavior of muscle data. Since the design of the network is not limited to visco-elasticity, iCANNs might help to autonomously identify the inelastic phenomena of the material and subsequently select the most appropriate model. Here, our focus is on providing a thermodynamically consistent framework for inelastic material behaviors and how to incorporate this framework into neural networks in an architecture-based manner. Our source code, data, and examples are available at Holthusen et al. (2023a) (<https://doi.org/10.5281/zenodo.10066805>).

1. Motivation

In computational mechanics, traditional neural networks can be used to replace ‘hand-crafted’ material models. However, as they do not incorporate any thermodynamical knowledge into the network’s architecture, they may learn the relation between strains and stresses well, but often fail to predict the material behavior outside the training regime. Unphysical predictions of stress outside the training regime can already be observed in the case of uniaxial tensile tests of hyperelastic materials (see e.g. Linka and Kuhl [1]). In the latter mentioned work, the authors trained a traditional neural networks using Treloar’s data [2], where the loss is defined as

* Corresponding author.

E-mail address: hagen.holthusen@ifam.rwth-aachen.de (H. Holthusen).

Nomenclature

a, A	Scalar
a	First order tensor
A	Second order tensor
I	Identity tensor
$SO(\bullet)$	Special orthogonal group
A^T	Transpose of A
A^{-1}	Inverse of A
$\text{sym}(A)$	$\frac{1}{2}(A + A^T)$
$\text{tr}(A)$	Trace of A
$\det(A)$	Determinant of A
$\text{dev}(A)$	$A - \frac{\text{tr}(A)}{3}I$
$\exp(A)$	Exponential of A
I_1^A	$\text{tr}(A)$
I_2^A	$\frac{1}{2}(\text{tr}(A)^2 - \text{tr}(A^2))$
I_3^A	$\det(A)$
J_2^A	$\frac{1}{2}\text{tr}(\text{dev}(A)^2)$
J_3^A	$\frac{1}{3}\text{tr}(\text{dev}(A)^3)$

the difference between the stress predicted by the network and the experimentally observed one. The corresponding stretch serves as input for the network. Although the network is capable of adequately describing the stress within the training regime, an unphysical decrease in stiffness is predicted with further stretching beyond the training regime. In a similar manner, Wang et al. [3] studied the performance of traditional neural networks in the case of inelastic material behavior, in particular visco-elasticity. They observed unphysical predictions for several uniaxial relaxation tests. During relaxation, the stress is expected to monotonically decrease while the overall stretch is hold constant. However, the traditional neural network predicts a non-smooth stress during relaxation and even an increase in stress. This is in clear contradiction to thermodynamics for standard materials. Another issue of traditional neural networks is that they typically require a rather huge amount of solution data of the underlying problem for the training process. This can be an obstacle whenever the acquisition of data from real experiments or realistic micromechanical simulations is either not feasible or too costly.

For this reason, neural networks were developed in which — in one way or the other — prior knowledge about the underlying physics of the problem is incorporated. This is done with the aim of requiring less data and improving the predictive capabilities of the trained network. One prominent approach in this direction are Physics-Informed Neural Networks, also called PINNs [4]. PINNs and PINN-like approaches¹ incorporate physical constraints to the formulation by adding additional terms to the loss function. This strategy facilitates the training process by narrowing the admissible solution space and assists the optimizer in finding a network behaving physically more reasonably, i.e. showing better extrapolation capabilities, which strictly obey thermodynamics. Typically, PINNs and PINN-like approaches are used for solving partial differential equations, i.e. whole (initial-)boundary value problems, as an alternative to, e.g., the finite element method (for several recent examples in the context of solid mechanics, see Rao et al. [5], Zhang et al. [6], Amini Niaki et al. [7], Cai et al. [8], Haghighat et al. [9], Vahab et al. [10], Henkes et al. [11], Rezaei et al. [12], Harandi et al. [13], Niu et al. [14]). Much less frequently, PINNs are also used to replace constitutive models. For instance, Haghighat et al. [15] present a PINN framework which is able to characterize and discover yield surfaces and flow rules in geometrically linear elasto-plasticity, showing how to deal with complex inequality constraints in the loss function. Eghbalian et al. [16] develop a surrogate model approximating classical elasto-plastic constitutive relations that can be used for large deformations via integration into a hypoelastic-based plasticity framework. However, as mentioned e.g. by Weber et al. [17] PINNs and PINN-like approaches only weakly enforce physical laws during training, such as the strong form of linear momentum in case of boundary value problems or the second law of thermodynamics for constitutive material models. Thus, there is no guarantee that any prediction will satisfy these physical laws a priori.

In a completely different direction, so-called (material) model-free approaches have been developed that rely only on experimental data [18,19]. Compared to a combination of constitutive knowledge and machine learning algorithms, such purely data-driven methods have the clear advantage that only minimal and generally accepted laws (such as basic kinematic relationships and force equilibrium in a mechanical context) are assumed as given and no complicated (and possibly error-prone) constitutive equations need to be formulated at all. On the downside, the large amount of data required for the success of such methods can be disadvantageous – a property shared with traditional neural network approaches. These methods have already been extended to inelasticity (see

¹ We refer here to neural network approaches which also add physical constraints to the loss function, but are not explicitly referred to as PINNs in the corresponding papers.

e.g. Eggersmann et al. [20], Ibanez et al. [21]). Nevertheless, their successful applicability in the case of complex multi-dimensional and history-dependent problems or problems with non-unique solutions is not yet thoroughly investigated. Hence, they are subject of intensive research in various directions (see e.g. Eggersmann et al. [22], Ciftci and Hackl [23], Zschocke et al. [24], Kuang et al. [25], to name only a few).

As a very promising alternative to PINNs and PINN-like approaches described above, neural networks have been developed that embed physical constraints into the architecture of the networks. For instance, Thermodynamics-based Artificial Neural Networks (TANNs) developed by Masi et al. [26] and the Mechanics-Informed Artificial Neural Networks (MIANNs) by As'ad et al. [27] follow this approach. The latter approaches were already extended to account for inelastic materials by employing the concept of state variables. For instance, Masi and Stefanou [28] also include the transition between different scales while Masi and Stefanou [29] overcome the need of an incremental time-discretization scheme. In As'ad and Farhat [30], MIANNs were extended to account for visco-elasticity. For completeness, recent studies explored the capabilities of symbolic regression in constitutive modeling in more detail. As in the case of physically-guided neural networks such as the former mentioned ones, thermodynamics are taught to the symbolic regression algorithm. Thus, symbolic regression provides interpretable hyperelastic models in terms of stretch invariants (see, for instance, Abdusalamov et al. [31,32,33], and Hou et al. [34]. For a brief comparison of physics-embedded neural networks and symbolic regression, the interested reader is referred to the work of Hou et al. [34]. Another way of proceeding is to work with traditional neural networks in the first place but to enforce the constraints a posteriori, as suggested e.g. by Kalina et al. [35]. These types of networks obey fundamental physical laws, such as the second law of thermodynamics. Therefore, these approaches are very much in line with our goal of not wanting to design neural networks that completely discard all the knowledge of physical and thermodynamical principles having been accumulated over centuries. Instead, it is our strong belief that we should better teach networks the mentioned principles already beforehand. Consequently, as the specific expressions (so to say the active terms within the network) of the Helmholtz free energy combined with an inelastic potential give rise to the amount of (possibly inelastic) material phenomena, such networks may reveal all the effects for us in the data by finding the best thermodynamically consistent models.

Recently, these considerations have led to a novel development. The basic idea is as follows: instead of laboriously deducing which specific phenomena are actually taking place in a material by manually analyzing experimental data and setting up a corresponding biased material model (which is likely either too simplistic or too complex, depending on the developer's level of experience), a constitutive model is automatically discovered that optimally only reflects the material phenomena truly present in the data. One representative following this idea is the framework EUCLID developed by Flaschel et al. [36,37]. Basically, EUCLID is an intelligent algorithm using unlabeled data and sparse regression techniques to select and link only those candidates from a large pre-formulated library of constitutive models (generalized standard materials) which are actually necessary to describe the data. For an overview of particular choices for at least the Helmholtz free energy, we kindly refer the interested reader to Steinmann et al. [38]. In Flaschel et al. [39], it has been shown (for the case of small deformations) that such an approach can be used to successfully identify a model for a material whose material class was previously completely unknown.

The approach followed by EUCLID gives rise to the idea of networks based on constitutive knowledge. Instead of laboriously identifying the relations between the entire strain and stress tensor, we can employ constitutive relations. Thus, we only have to discover a scalar function, the Helmholtz free energy, and design our network in such a way that constitutive principles are strictly ensured. In a more general context, the idea of function learning networks, known as equation learner (EQL), was proposed by Martius and Lampert [40]. These networks provide a physically interpretable function (e.g. the Helmholtz free energy) by particular choices of activation functions, and thus, help us to understand the underlying phenomenon. The improved version of EQL networks by Sahoo et al. [41], called EQL⁺, includes division operation and enhances the model selection to increase the efficiency of the identification procedure. These networks have already found their way into the identification of partial differential equations, e.g. in Long et al. [42]. Since the scope of this contribution is on designing neural networks for inelastic material behavior at finite strains, we kindly refer the interested reader for a review of hybrid models generally combining machine learning and physics to Rai and Sahu [43], including EQL/EQL⁺.

A somehow related strategy to EQL/EQL⁺ networks in a mechanical context is pursued by Constitutive Artificial Neural Networks, also known as CANNs [44–46]. In contrast to most other neural networks, which discover the constitutive relations, CANNs aim to formulate a physically reasonable, yet general expression for the Helmholtz free energy by a particular choice of the network architecture. Here we take advantage of the fact that the Helmholtz free energy is considered to be a smooth and differentiable function. Thus, a Taylor series with respect to the invariants of the right Cauchy–Green tensor can be performed, which is less restrictive compared to pre-formulating a library of constitutive models. Similar to EQL/EQL⁺ networks, custom-designed activation functions ensure that, for example, convexity is guaranteed and at the same time expand the range of possible functional forms. However, the architecture of the network is specifically designed for mechanical problems, i.e. the network is not fully connected due to convexity requirements. Moreover, kinematic, thermodynamic, and physical constraints (such as those given by the second law of thermodynamics, material objectivity, material symmetry, etc.) are rigorously considered in the design of the network's inputs and outputs. Consequently, through appropriate training, the network will autonomously recognize which components of this parameterized energy are actually needed to describe the data. The main novelty is that, in special cases, the network will be able to exactly represent known constitutive models from the literature; in general, however, it will generate a previously completely unknown material model that describes the data even more accurately and leads to better predictions [1,47]. Important in this context is also the fact that the network learns a set of physically interpretable parameters, which is clearly advantageous. However, a critical missing link is to expand this concept to inelastic materials.

Today, there are nearly as many approaches to model inelastic material behavior as there are inelastic phenomena. Some of them are only useful for individual effects, such as visco-elasticity. Therefore, we should not ask ourselves how we can fuse all

these approaches into one model; instead, what is the most general approach? What are the essential steps to describe inelastic effects? Once these general mechanisms are found, the question of inelasticity, e.g. plasticity, is merely a question of specializing the approach and not a question of changing the model. For solids, we identify the concepts of internal state variables [48] and pseudo/dissipation potentials of stress [49] as universal and perfectly complementary approaches. The former concept allows to distinguish between elastic (reversible) strains associated with the amount of elastically stored energy and inelastic (irreversible) strains related to the dissipation. In the finite strain theory, the assumption of a multiplicative decomposition of the deformation gradient into elastic and inelastic parts is advantageous, well-known, widely accepted in the continuum mechanics community, and able to capture a broad range of material behaviors (see Kröner [50], Lee [51], Mandel [52], Sidoroff [53], Rodriguez et al. [54], Lion [55], Haupt and Sedlan [56], Menzel et al. [57], Bammann and Solanki [58], Latorre and Montáns [59], and Reina and Conti [60] among many others). It is worth noting that none of these parts generally satisfies the compatibility condition, and thus, cannot be derived from an elastic and/or inelastic motion field, so they should not be misunderstood as second-order gradient-like tensors such as the deformation gradient. The latter concept follows from sound thermodynamic considerations that are valid for rate-dependent (see e.g. Lion [61,62], Reese and Govindjee [63], and Besson [64]) and -independent (see e.g. Miehe [65], Badreddine et al. [66]) materials and can be easily extended to induced anisotropy (see e.g. Hill and Orowan [67], Benzerga and Besson [68]). Similar to the Helmholtz free energy, which gives us the relationship between strains and stresses, the pseudo potential reveals how the thermodynamic driving force influences the microstructure and, thus, determines the inelastic rate (see Maugin and Muschik [69]). The charm of this combination lies in the fact that the potential can be interpreted mechanically in a way similar to the volumetric–isochoric split [70], and further, that restricting the potential to be zero-valued, positive, and convex satisfies thermodynamical requirements a priori (see Germain et al. [71]). Actually, the approach is in line with that for generalized standard materials (see Halphen and Nguyen [72]), with the difference that the dissipation potential is formulated directly in terms of stress-like quantities.

Objective and outline. In this contribution, we present a new framework for model discovery that extends Constitutive Artificial Neural Networks (CANNs) to general inelastic material behavior (from now on called iCANNs). Our new iCANN assumes a multiplicative decomposition of the deformation gradient and is, thus, capable to capture finite deformations and deformation rates. To satisfy thermodynamics a priori, we derive the inelastic rates from a pseudo potential that depends on the thermodynamic driving force. In Section 2, we briefly review the fundamentals of thermodynamics for inelastic materials in general. To this end, we introduce universally valid restrictions on the Helmholtz free energy as well as the pseudo potential. Our considerations lead us to the architecture of the network in Section 3. In analogy to CANNs, we use a feed-forward network for the pseudo potential, the underlying concept of which is inherently generic and holds for any type of inelasticity. We subsequently embed the feed-forward networks of the energy and pseudo potential into a recurrent neural network environment. Thus, we obtain a modular structure for iCANNs, which we specialize to visco-elasticity for illustrative purposes in Section 4. In Section 5, we investigate the performance of our iCANN using artificially generated data, the visco-elastic behavior of polymers subjected to cyclic loading, and the relaxation behavior of muscle data. We continue with a discussion of our approach and mention its limitations in Section 6, before providing concluding remarks and an outlook in Section 7.

2. Generalization: Constitutive framework for inelastic materials

To ensure a thermodynamically consistent and objective formulation of our inelastic constitutive network, we will briefly recall the basic principles of continuum mechanics. Our two main model ingredients, from which all constitutive relations are derived, are the Helmholtz free energy, ψ_0 , and the pseudo potential, g_0 , which must satisfy all these principles. Therefore, we introduce a motion, φ , that relates material points with position vector $\mathbf{x}_0$ in the reference configuration defined at a fixed moment in time to points in the current configuration, $\mathbf{x} = \varphi(\mathbf{x}_0, t)$, at time t (see Marsden and Hughes [73]). Since φ is a bijective function, the motion is invertible, i.e., $\mathbf{x}_0 = \varphi^{-1}(\mathbf{x}, t)$. By linearizing the (inverse) motion, we obtain the deformation gradient as well as its inverse (see Holzapfel [74])

$$\mathbf{F} := \frac{\partial \varphi(\mathbf{x}_0, t)}{\partial \mathbf{x}_0}, \quad \mathbf{F}^{-1} = \frac{\partial \varphi^{-1}(\mathbf{x}, t)}{\partial \mathbf{x}} \quad \text{with} \quad I_3^F > 0. \quad (1)$$

Note that we may also include higher gradients of motion to describe the relative change between two points, but this is not considered in this work (see Bertram and Forest [75]).

Objectivity. The principle of objectivity or frame indifference restricts the choice of constitutive equations so that they must be independent of the observer. Mathematically, this is equivalent to a rigid body motion of the current configuration (see Ogden [76]). Thus, we can introduce a second motion, $\varphi^+(\mathbf{x}_0, t)$, and its inverse, $\varphi^{+^{-1}}(\mathbf{x}^+, t)$, as well as the translation and rigid rotation of the current position $\mathbf{x}^+(\mathbf{x}_0, t) = \mathbf{Q}^+(t)\mathbf{x}(\mathbf{x}_0, t) + \mathbf{c}^+(t)$ with $\mathbf{Q}^+ \in \text{SO}(3)$. Since the reference configuration remains the same for both motions, we conclude that $\varphi^{-1} = \varphi^{+^{-1}}$ and differentiate both with respect to $\mathbf{x}$

$$\underbrace{\frac{\partial \varphi^{+^{-1}}(\mathbf{x}^+, t)}{\partial \mathbf{x}^+}}_{=: \mathbf{F}^{+^{-1}}} \frac{\partial \mathbf{x}^+}{\partial \mathbf{x}} = \frac{\partial \varphi^{-1}(\mathbf{x}, t)}{\partial \mathbf{x}} \quad \Rightarrow \quad \mathbf{F}^+ = \mathbf{Q}^+ \mathbf{F} \quad (2)$$

which gives us the transformation of the deformation gradient when the observer changes.

Rigid motion of the reference configuration. Analogous to objectivity, a rigid motion of the reference configuration discusses under which alternative choices of reference configuration the response of the material must remain unchanged (see Chadwick [77]). If we assume a rigid body motion of the reference configuration and subsequently apply the same deformation, i.e., the relative distance between all points is changed the same as without the rigid motion, the behavior of the material must be invariant. Consequently, we can introduce a motion, $\varphi^\#(x_0^\#, t)$, which accounts for a translated and rotated reference configuration, i.e., $x_0^\# = Q^\# x_0 + c^\#$ with $Q^\# \in \text{SO}(3)$. Since both φ and $\varphi^\#$ yield to the same current configuration, we find the following (cf. [78])

$$\underbrace{\frac{\partial \varphi^\#(x_0^\#, t)}{\partial x_0^\#}}_{=: F^\#} \frac{\partial x_0^\#}{\partial x_0} = \frac{\partial \varphi(x_0, t)}{\partial x_0} \Rightarrow F^\# = F Q^{\#T}. \quad (3)$$

Multiplicative decomposition and rotational non-uniqueness. Regardless of whether the material exhibits visco-elastic behavior, undergoes elasto-plastic deformations, or grows and remodels itself, all these phenomena have in common that the (elastic) stresses arise from the elastically stored energy. Therefore, we utilize the multiplicative decomposition

$$F = F_e F_i, \quad \text{with} \quad I_3^{F_e}, I_3^{F_i} > 0 \quad (4)$$

of the deformation gradient into elastic, F_e , and inelastic, F_i , parts. This decomposition is applied in various fields of mechanics of inelastic materials, for instance, visco-elasticity (see Sidoroff [53], Lubliner [79]), elasto-(visco)plasticity (see Kröner [50], Eckart [80], and Lee [51]), damage (see e.g. Schütte and Bruhns [81], Dorn and Wulfinghoff [82]), phase transformation (see Christ and Reese [83], Evangelista et al. [84], and Sengupta and Papadopoulos [85]) as well as growth and remodeling (see Rodriguez et al. [54], Lubarda and Hoger [86]). However, since none of these parts can be derived from motion in general, they should not be misinterpreted as gradients. Unfortunately, due to the fact that the introduced intermediate configuration is fictitious, we find that the decomposition (4) suffers from an inherent rotational non-uniqueness, i.e.,

$$F = F_e Q^{*T} Q^* F_i =: F_e^* F_i^* \quad \text{with} \quad Q^* \in \text{SO}(3) \quad (5)$$

is equivalent to (4). As a consequence, the multiplicative decomposition into elastic and inelastic parts is not unique, which is illustrated in Fig. 1. Thus, neither F_e nor F_i can be obtained in general, as there is an infinite number of possible combinations. However, any constitutive law must be independent of this non-uniqueness (see, for instance, the brief but precise discussion by Casey [87]).

Elastic and inelastic stretch tensors. In the case of elastic materials, objectivity constrains the Helmholtz free energy to be a function of the right stretch tensor $U = +\sqrt{C}$, where $C = F^T F$ is the right Cauchy Green tensor, while a rigid motion of the reference configuration may lead to a dependence on the left stretch tensor $V = +\sqrt{B}$, where $B = F F^T$ is the left Cauchy Green tensor. If we superimpose both considerations, we find ψ_0 to be a scalar-valued isotropic function of its argument. In order to find suitable stretch measures in the case of inelasticity, we superimpose the findings of Eqs. (2), (3), and (5), i.e.,

$$F^{+*} = \underbrace{Q^+ F_e Q^{*T}}_{=: F_e^{+*}} \underbrace{Q^* F_i Q^{\#T}}_{=: F_i^{*#}}. \quad (6)$$

At this point, we assume ψ_0 to be a function of elastic and inelastic arguments. However, analogously to elasticity, it remains to find suitable elastic and inelastic stretch tensors. Therefore, we employ the polar decompositions of $F_e = R_e U_e$ and $F_i = V_i R_i$ with $R_e, R_i \in \text{SO}(3)$ into their rotational and symmetric stretch parts. We insert these relations into the definitions of F_e^{+*} and $F_i^{*#}$ and find the following (cf. [88])

$$U_e^{+*} = Q^* U_e Q^{*T}, \quad V_i^{*#} = Q^* V_i Q^{*T} \quad (7)$$

where U_e^{+*} and $V_i^{*#}$ are the stretch parts of F_e^{+*} and $F_i^{*#}$, respectively. Consequently, the Helmholtz free energy as an isotropic function of U_e and V_i satisfies the requirements of objectivity, rigid motion of the reference configuration as well as rotational non-uniqueness, since both U_e and V_i transform in a similar way if subjected to these three requirements. Nevertheless, contrary to their invariants, it is important to note that both tensors are still not unique.

Thermodynamically consistent evaluation. The above considerations have shown us how to design the Helmholtz free energy as a function of stretches. Now, in order to obtain a thermodynamically consistent material response when subjected to arbitrary loading, we must satisfy the Clausius–Planck inequality $\mathcal{D} := -\dot{\psi}_0 + \frac{1}{2} \mathcal{S} : \dot{C} \geq 0$. In the latter, $\mathcal{S}$ denotes the second Piola–Kirchhoff stress tensor. In line with the discussion led to Eq. (7), we can state the following general form of the Helmholtz free energy

$$\psi_0 = \psi(C_e, B_i). \quad (8)$$

For convenience, we expressed ψ in terms of the elastic right Cauchy Green tensor, $\mathbf{C}_e = \mathbf{U}_e^2$, and the inelastic left Cauchy Green tensor, $\mathbf{B}_i = \mathbf{V}_i^2$. We then evaluate the Clausius–Planck inequality for the Helmholtz free energy (8)

$$\left(\mathbf{S} - 2 \mathbf{F}_i^{-1} \frac{\partial \psi_0}{\partial \mathbf{C}_e} \mathbf{F}_i^{-T} \right) : \frac{1}{2} \dot{\mathbf{C}} + \underbrace{\left(2 \mathbf{C}_e \frac{\partial \psi_0}{\partial \mathbf{C}_e} - 2 \frac{\partial \psi_0}{\partial \mathbf{B}_i} \mathbf{B}_i \right)}_{=: \boldsymbol{\Gamma}} : \mathbf{D}_i \geq 0 \quad (9)$$

and, under consideration of the arguments of Coleman and Noll [89,90], Coleman and Gurtin [91], end up with the state law $\mathbf{S} = 2 \mathbf{F}_i^{-1} \frac{\partial \psi_0}{\partial \mathbf{C}_e} \mathbf{F}_i^{-T}$. Noteworthy, we employed the symmetry of the relative stress, $\boldsymbol{\Gamma}$, to reduce $\dot{\mathbf{F}}_i \mathbf{F}_i^{-1}$ to its symmetric part, $\mathbf{D}_i = \text{sym}(\dot{\mathbf{F}}_i \mathbf{F}_i^{-1})$ (see Svendsen [92]). In contrast to $\boldsymbol{\Gamma}$, neither the elastic Mandel-like stress, $\boldsymbol{\Sigma}$, nor the backstress, $\mathbf{X}$, are generally symmetric. Moreover, we recognize that then subjected to the three requirements mentioned, neither the thermodynamically consistent driving forces $\boldsymbol{\Sigma}^{+*} = \mathbf{Q}^* \boldsymbol{\Sigma} \mathbf{Q}^{*T}$, $\mathbf{X}^{+*} = \mathbf{Q}^* \mathbf{X} \mathbf{Q}^{*T}$, and $\boldsymbol{\Gamma}^{+*} = \mathbf{Q}^* \boldsymbol{\Gamma} \mathbf{Q}^{*T}$, nor the inelastic deformation rate $\mathbf{D}_i^{+*} = \mathbf{Q}^* \mathbf{D}_i \mathbf{Q}^{*T}$ are unique in the sense of Eq. (7).

Evolution equation. It remains to choose an evolution equation for $\mathbf{D}_i$ such that the remaining Inequality (9), $\boldsymbol{\Gamma} : \mathbf{D}_i \geq 0$, is fulfilled for arbitrary processes. To this end, we introduce a pseudo potential in terms of the thermodynamically consistent driving force, $g_0 = g(\boldsymbol{\Gamma})$, which is also a scalar-valued isotropic function. Next, the evolution equation is assumed to be (cf. [49])

$$\mathbf{D}_i = \gamma \frac{\partial g_0}{\partial \boldsymbol{\Gamma}} \quad (10)$$

where $\gamma \geq 0$ introduces a time-scale into the model. Depending on the inelastic phenomenon modeled, γ might be interpreted as a relaxation time (visco-elasticity) or Lagrange multiplier (elasto-plasticity). In case of growth, the concept of homeostatic surfaces (see Lamm et al. [93]) allows the interpretation of γ . For the time being, we constrain g_0 to be convex, non-negative and zero-valued at the origin, i.e., $g_0(\mathbf{0}) = 0$. Hence, the dissipation inequality is satisfied (see Germain et al. [71]), which will be discussed in more detail in Section 2.2. In addition, it is worth mentioning that many potentials found in literature assume g_0 to be a positive homogeneous function of degree $\alpha > 0$. Thus, it is straightforward to prove that $\mathcal{D} \geq 0$, since $\boldsymbol{\Gamma} : \mathbf{D}_i = \alpha \gamma g(\boldsymbol{\Gamma})$ (cf. [94]).

Inelastic phenomenon. The theoretical considerations so far hold for any inelastic material behavior that can be modeled by the multiplicative decomposition. For specializing this general framework to a specific inelastic phenomenon, e.g. elasto-plasticity, γ in Eq. (10) needs to be determined accordingly. In material modeling, we usually distinguish between constrained and unconstrained phenomena. In the former, γ acts as a Lagrange multiplier subjecting the material to an additional constraint, $\Phi(\boldsymbol{\Gamma})$, which usually depends on the relative stress. In case of elasto-plasticity, damage and phase transformation, we distinguish between elastic ($\Phi(\boldsymbol{\Gamma}) < 0$) and inelastic regimes ($\Phi(\boldsymbol{\Gamma}) = 0$).² The work of Lamm et al. [93] introduces the concept of homeostatic surfaces, i.e. $\Phi(\boldsymbol{\Gamma}) = 0$ must always be satisfied, allowing the interpretation of γ for growth and remodeling. In contrast, visco-elasticity is modeled as an unconstrained problem. Here we identify γ as the reciprocal of the relaxation time and consider it to be constant or most likely a function of stress. In this regard, Ricker et al. [95] provide a comprehensive overview of suitable functional choices.

Co-rotated configuration. We observed that none of the tensors defined with respect to the intermediate configuration are unique. A consequence of this is that none of the partial derivatives with respect to $\mathbf{C}_e$, $\mathbf{B}_i$, and $\boldsymbol{\Gamma}$ can be obtained using algorithmic differentiation. This is a pity, because in our neural network we only want to design general functions of both ψ_0 and g_0 that are capable of finding the best model to explain the experimental data without implementing additional pull-backs (see Simo and Miehe [96], Dettmer and Reese [97]). To this end, we use the co-rotated formulation introduced by Holthusen et al. [98], which has proven advantageous when combined with algorithmic differentiation. The fundamental idea is to rotate all non-unique tensors to their co-rotated counterparts, $(\bar{\bullet}) = \mathbf{R}_i^{-1}(\bullet)\mathbf{R}_i$ (see Fig. 1),

$$\bar{\mathbf{C}}_e = \mathbf{U}_i^{-1} \mathbf{C} \mathbf{U}_i^{-1}, \quad \bar{\mathbf{B}}_i \equiv \mathbf{C}_i, \quad \bar{\boldsymbol{\Sigma}} = 2 \bar{\mathbf{C}}_e \frac{\partial \psi_0}{\partial \bar{\mathbf{C}}_e}, \quad \bar{\mathbf{X}} = 2 \frac{\partial \psi_0}{\partial \mathbf{C}_i} \mathbf{C}_i, \quad \bar{\mathbf{D}}_i = \gamma \frac{\partial g_0}{\partial \bar{\boldsymbol{\Gamma}}}. \quad (11)$$

In the latter equation, $\mathbf{C}_i = \mathbf{F}_i^T \mathbf{F}_i = \mathbf{U}_i^2$ is the inelastic right Cauchy Green tensor and $\bar{\boldsymbol{\Gamma}} = \bar{\boldsymbol{\Sigma}} - \bar{\mathbf{X}}$ is the co-rotated relative stress. Employing the right polar decomposition of $\mathbf{F}_i = \mathbf{R}_i \mathbf{U}_i$, we obtain the inelastic right stretch tensor $\mathbf{U}_i$ (see Fig. 1). Since all tensors and their co-rotated counterparts are *similar*, both $\psi(\mathbf{C}_e, \mathbf{B}_i) = \psi(\bar{\mathbf{C}}_e, \mathbf{C}_i)$ and $g(\boldsymbol{\Gamma}) = g(\bar{\boldsymbol{\Gamma}})$ maintain the same functional structure. It is interesting to note that $\bar{\boldsymbol{\Sigma}}$ and the Kirchhoff stress, $\mathbf{F} \mathbf{S} \mathbf{F}^T$, share the same eigenvalues (see Dettmer and Reese [97]). Hence, the pseudo potential expressed in terms of $\bar{\boldsymbol{\Gamma}}$ is physically reasonable, which is considered an advantage. Finally, the state law reads $\mathbf{S} = 2 \mathbf{U}_i^{-1} \frac{\partial \psi_0}{\partial \bar{\mathbf{C}}_e} \mathbf{U}_i^{-1}$ (cf. [90,91], and [48]).

² For references, see the literature cited in the subsection ‘Multiplicative decomposition and rotational non-uniqueness’.

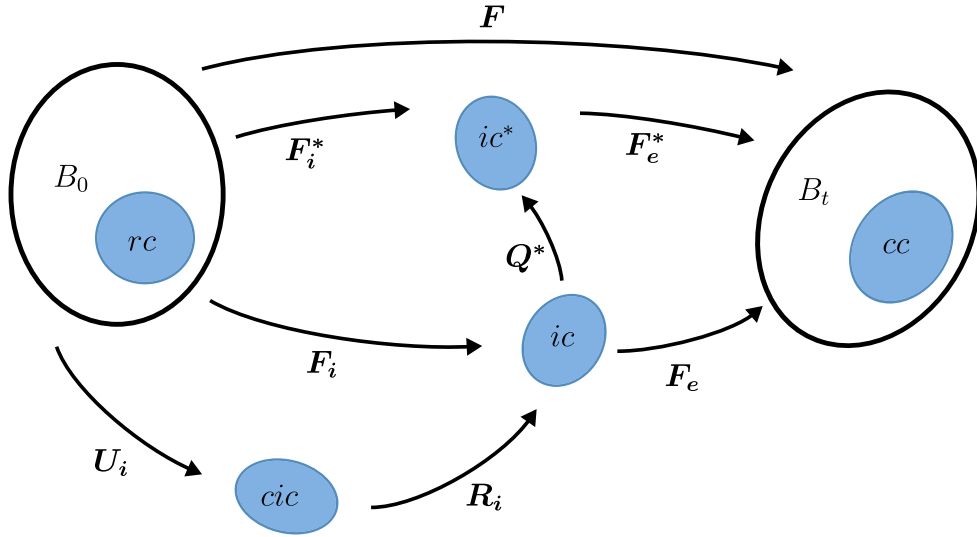


Fig. 1. Multiplicative decomposition of the deformation gradient, F , into elastic, F_e , and inelastic, F_i , parts as well as their arbitrarily rotated counterparts F_e^* and F_i^* . Further, $F_i = R_i U_i$ possesses its polar decomposition. Configurations: rc – reference configuration, ic – intermediate configuration, ic^* – arbitrarily rotated intermediate configuration, cic – co-rotated intermediate configuration, cc – current configuration.

2.1. Helmholtz free energy: How strains cause stresses

So far, we have seen that, in addition to kinematics, the Helmholtz free energy function is the essential function that tells us how to go from strains to all the stresses we need. Further, objectivity, rigid motion of the reference configuration, and rotational non-uniqueness led us to the conclusion that ψ_0 must be an isotropic function of $\bar{C}_e$ and C_i . Thus, we can express the Helmholtz free energy in terms of its irreducible integrity basis (see Spencer [99], Boehler [100], and Zheng [101])

$$\psi_0 = \psi(\bar{C}_e, C_i) = \psi(I_1^{\bar{C}_e}, I_2^{\bar{C}_e}, I_3^{\bar{C}_e}, I_4, I_5, I_6, I_7, I_1^{C_i}, I_2^{C_i}, I_3^{C_i}) \quad (12)$$

with the mixed invariants $I_4 = \text{tr}(\bar{C}_e C_i)$, $I_5 = \text{tr}(\bar{C}_e^2 C_i)$, $I_6 = \text{tr}(\bar{C}_e C_i^2)$, and $I_7 = \text{tr}(\bar{C}_e^2 C_i^2)$. The constitutive artificial neural network that we design will learn the energy that best explains the experimental data by identifying the most valuable terms of a generic form for the Helmholtz free energy (see Linka and Kuhl [1]). The goal of this section is therefore to design such a generic function satisfying thermodynamics.

Isotropy. Within this contribution, we will only consider *isotropic* materials, i.e., any directionally dependence is excluded. As kinematic hardening shifts the pseudo potential in the principal stress space, we neglect any influence of C_i , thus reducing the energy to $\psi(\bar{C}_e) = \psi(I_1^{\bar{C}_e}, I_2^{\bar{C}_e}, I_3^{\bar{C}_e})$. Furthermore, considering the energy as sufficiently smooth and infinitely differentiable, we can express it in terms of a Taylor series $\psi = \sum_{i,j,k=0}^{\infty} a_{ijk}(I_1^{\bar{C}_e} - 3)^i (I_2^{\bar{C}_e} - 3)^j (I_3^{\bar{C}_e} - 1)^k$ at the elastically undeformed state $\bar{C}_e = I$. Since the free energy is zero in the undeformed state, we can conclude that $a_{000} = 0$. However, within the network that we will design, it is not possible, from a numerical point of view alone, to consider an infinite number of terms. Unfortunately, this may violate fundamental conditions the energy must fulfill, namely

$$\lim_{I_3^{\bar{C}_e} \rightarrow \infty} \psi \rightarrow +\infty \quad \text{and} \quad \lim_{I_3^{\bar{C}_e} \searrow 0} \psi \rightarrow +\infty. \quad (13)$$

as well as

$$\lim_{I_1^{\bar{C}_e} \rightarrow \infty} \psi \rightarrow +\infty \quad \text{and/or} \quad \lim_{I_2^{\bar{C}_e} \rightarrow \infty} \psi \rightarrow +\infty. \quad (14)$$

Volumetric–isochoric split. To circumvent these issues, we exploit the volumetric–isochoric split (see Flory [70]) and assume the elastic energy to take the form $\psi = \psi^{\text{iso}}(\bar{I}_1^{\bar{C}_e}, \bar{I}_2^{\bar{C}_e}) + \psi^{\text{vol}}(I_3^{\bar{C}_e})$ where $\bar{I}_1^{\bar{C}_e} = I_1^{\bar{C}_e} / (I_3^{\bar{C}_e})^{1/3}$ and $\bar{I}_2^{\bar{C}_e} = I_2^{\bar{C}_e} / (I_3^{\bar{C}_e})^{2/3}$. This allows us to choose the following polynomial forms for the energy terms (cf. [102])

$$\psi^{\text{iso}} = \sum_{i=0}^n \sum_{j=0}^m c_{ij} (\bar{I}_1^{\bar{C}_e} - 3)^i (\bar{I}_2^{\bar{C}_e} - 3)^j, \quad \psi^{\text{vol}} = \sum_{k=0}^p d_k (I_3^{\bar{C}_e} - 1)^{2k} + W(I_3^{\bar{C}_e}) \quad (15)$$

where we introduced $W(I_3^{\bar{C}_e})$ in order to ensure the requirements (13). The isochoric response, ψ^{iso} , addresses the second requirement (14). Similar to a_{000} and for mechanical convenience, we find that $c_{00} = d_0 = 0$ as well as $W(1) = 0$. In addition

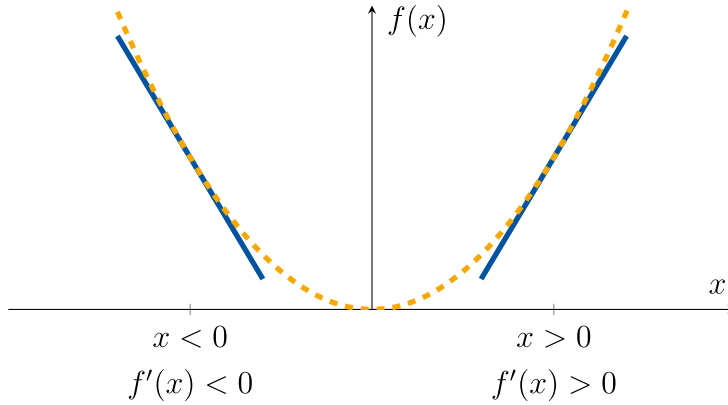


Fig. 2. If a scalar function $f(x)$ is convex, non-negative, and zero-valued, then it is always true that $x \frac{df(x)}{dx} \geq 0$.

to the Helmholtz free energy being zero in the undeformed state, mechanics tells us that the same must hold for the stress. For the isochoric response, this is naturally ensured by vanishing gradients of the modified invariants, i.e., $\left. \frac{\partial \bar{I}_1^{\bar{C}_e}}{\partial \bar{C}_e} \right|_{\bar{C}_e=I} = \left. \frac{\partial \bar{I}_2^{\bar{C}_e}}{\partial \bar{C}_e} \right|_{\bar{C}_e=I} = 0$. Since the first term of the volumetric response, $(\bar{I}_3^{\bar{C}_e} - 1)^{2k}$, satisfies this condition as well, we conclude that $\frac{\partial W(\bar{I}_3^{\bar{C}_e})}{\partial \bar{I}_3^{\bar{C}_e}}$ must be zero for $\bar{I}_3^{\bar{C}_e} = 1$.

Design of isochoric energy. Lastly, we pay attention to the isochoric part of the energy, $\psi^{\text{iso}}(\bar{I}_1^{\bar{C}_e}, \bar{I}_2^{\bar{C}_e})$. We choose a generalized Mooney–Rivlin approach (see Mooney [103], Rivlin and Rideal [104], and Bower [105])

$$\psi^{\text{iso}} = \sum_{i=0}^n c_{i0} (\bar{I}_1^{\bar{C}_e} - 3)^i + \sum_{j=0}^m c_{0j} (\bar{I}_2^{\bar{C}_e} - 3)^j \quad (16)$$

which is in line with the CANN framework by Linka and Kuhl [1], Linka et al. [44].³ A physical argument for this choice of invariants and the underlying mechanical assumptions can be found in Sansour [106], Pellicciari et al. [107]. Note that the additive split of Eq. (16) can also be motivated by convexity arguments, since it is easier to prove that the sum of convex functions is convex (cf. [108]). However, it should be noted that $(\bar{I}_1^{\bar{C}_e} - 3)^i$ is a polyconvex function, but $(\bar{I}_2^{\bar{C}_e} - 3)^j$ is generally not.

2.2. Pseudo potential: How elastic stresses lead to inelastic strains

The previous section discussed several conditions that the free energy must satisfy to describe materials thermodynamically consistent. Within our overall recurrent neural network, we incorporate both a feed-forward network for the energy and a feed-forward network for the pseudo potential. The aforementioned conditions for the energy must be embedded into the design of the associated architecture of the feed-forward network. Next, we will discuss the design of our second scalar function necessary to describe inelastic materials, the pseudo potential $g(\bar{\mathbf{F}})$. As an isotropic function, we can express the potential in terms of its integrity basis as well, i.e., $g(\bar{\mathbf{F}}) = g(I_1^{\bar{\mathbf{F}}}, J_2^{\bar{\mathbf{F}}}, J_3^{\bar{\mathbf{F}}})$.

Set of invariants. The choice of stress invariants is by no means unique. For instance, choosing the principal stresses, which is the case for the yield criterion by Tresca [109], is possible as well. However, as will be discussed below, we will observe that this choice of invariants is vividly reflected in the evolution of inelastic strains. For the subsequent discussions, it is useful to take the following into account

$$\begin{aligned} \frac{\partial I_1^{\bar{\mathbf{F}}}}{\partial \bar{\mathbf{F}}} &= \mathbf{I}, & \frac{\partial I_1^{\bar{\mathbf{F}}}}{\partial \bar{\mathbf{F}}} : \bar{\mathbf{F}} &= I_1^{\bar{\mathbf{F}}} \\ \frac{\partial J_2^{\bar{\mathbf{F}}}}{\partial \bar{\mathbf{F}}} &= \text{dev}(\bar{\mathbf{F}}), & \frac{\partial J_2^{\bar{\mathbf{F}}}}{\partial \bar{\mathbf{F}}} : \bar{\mathbf{F}} &= 2 J_2^{\bar{\mathbf{F}}} \\ \frac{\partial J_3^{\bar{\mathbf{F}}}}{\partial \bar{\mathbf{F}}} &= \text{dev}(\text{dev}(\bar{\mathbf{F}})^2), & \frac{\partial J_3^{\bar{\mathbf{F}}}}{\partial \bar{\mathbf{F}}} : \bar{\mathbf{F}} &= 3 J_3^{\bar{\mathbf{F}}}. \end{aligned} \quad (17)$$

In the latter equation, the first invariant accounts for the hydrostatic pressure, the second takes the amount of shear stresses into account, and the third contains information about the direction and state of stress.⁴ As for the free energy, we need to design a

³ Noteworthy, Abdusalamov et al. [31] follow a similar approach for model discovery by means of symbolic regression.

⁴ In materials theory, Lode's angle is often used instead of the third stress invariant (cf. [110]).

generic formulation of the pseudo potential. Since this is considered to be smooth and continuously differentiable, the Taylor series seems to be a suitable tool to inspire such a formulation. Note that we use the expansion to find a reasonable and mathematically suitable formulation of the potential rather than to identify the (possible) real potential. Therefore, we expand the pseudo potential by a Taylor series at the origin $\bar{\mathbf{F}} = \mathbf{0}$, i.e., $g = \sum_{i,j,k} r_{ijk} (I_1^{\bar{\mathbf{F}}})^i (J_2^{\bar{\mathbf{F}}})^j (J_3^{\bar{\mathbf{F}}})^k$. Expanding the Taylor series at $\bar{\mathbf{F}} = \mathbf{0}$ is by no means a unique choice. In contrast to the Taylor series of ψ , where it can be reasonably assumed that $\bar{\mathbf{C}}_e$ is close to identity, $\bar{\mathbf{F}}$ might be far from zero. Unfortunately, in this case a high number of Taylor series terms would be required. Although one might consider to expand g close to the actual value $\bar{\mathbf{F}}$ at the current time step, this is not possible in practice as we have to uniquely find the values of r_{ijk} . Regardless of whether we set up a 'hand-crafted' material model or discover a constitutive neural network, we therefore have to choose a single point of expansion for all time steps. In our case, it seems most reasonable to choose zero, since the value of r_{000} is based on thermodynamic considerations, see the discussion below. To reduce the issue in question, we could either use more Taylor terms, which is considered disadvantageous as the parameter identification would be more challenging, or broaden the functional space by applying custom-designed activation functions, which will be discussed in more detail in Section 3.

Thermodynamic consistency. As already mentioned in Section 2, the pseudo potential must be chosen such that the potential (cf. [71])

(i) is *zero-valued* at the origin.

Therefore, we choose $r_{000} = 0$ in the Taylor series of g . In practice, we could add any constant term to the potential and still get a thermodynamically consistent model. However, due to the fact that the potential can give rise to the dissipated energy, a zero-valued potential seems most reasonable and is a common choice.

(ii) is *convex*.

For similar argumentations as in the previous section, we therefore omit all mixed variant terms. Consequently, the Taylor series reduces to

$$g = \sum_{i=0}^{\alpha} r_{i00} (I_1^{\bar{\mathbf{F}}})^i + \sum_{j=0}^{\beta} r_{0j0} (J_2^{\bar{\mathbf{F}}})^j + \sum_{k=0}^{\kappa} r_{00k} (J_3^{\bar{\mathbf{F}}})^k =: g_1(I_1^{\bar{\mathbf{F}}}) + g_2(J_2^{\bar{\mathbf{F}}}) + g_3(J_3^{\bar{\mathbf{F}}}). \quad (18)$$

Hence, the feed-forward network architecture of the potential is *not fully connected* as well.

(iii) is *non-negative*.

Since $J_2^{\bar{\mathbf{F}}} \geq 0$, g_2 in Eq. (18) always satisfies this condition, and further, also the requirements (i) and (ii). Unfortunately, both $I_1^{\bar{\mathbf{F}}}$ and $J_3^{\bar{\mathbf{F}}}$ might be negative, and thus, all odd values of i and k can lead to negative sub-potentials g_1 and g_3 . Since we do not want to exclude odd values a priori, we need to ensure non-negativity as well as (i) and (ii) through appropriate custom-designed activation functions within the feed-forward network of our potential. The fulfillment will therefore be guaranteed by design. We will introduce our choice of functions in Section 3. Alternatively, one may use input convex neural networks (see Amos et al. [111]), as for instance used by As'ad and Farhat [112].

These three requirements together lead to a potential that has its global minimum at its origin. In order to briefly illustrate why this leads to thermodynamically consistent, inelastic materials in general, we recall the reduced dissipation inequality, $\bar{\mathbf{F}} : \bar{\mathbf{D}}_i \geq 0$, the evolution Eq. (10), the derivatives of the stress invariants (17), and the choice of our potential (18), viz.

$$\bar{\mathbf{F}} : \bar{\mathbf{D}}_i = \gamma \left(\frac{\partial g_1(I_1^{\bar{\mathbf{F}}})}{\partial I_1^{\bar{\mathbf{F}}}} I_1^{\bar{\mathbf{F}}} + 2 \frac{\partial g_2(J_2^{\bar{\mathbf{F}}})}{\partial J_2^{\bar{\mathbf{F}}}} J_2^{\bar{\mathbf{F}}} + 3 \frac{\partial g_3(J_3^{\bar{\mathbf{F}}})}{\partial J_3^{\bar{\mathbf{F}}}} J_3^{\bar{\mathbf{F}}} \right) \geq 0. \quad (19)$$

Since the derivative of the sub-potentials with respect to their corresponding arguments, e.g. $\frac{\partial g_1(I_1^{\bar{\mathbf{F}}})}{\partial I_1^{\bar{\mathbf{F}}}}$ and $I_1^{\bar{\mathbf{F}}}$, always have the same sign due to the requirements (i)-(iii), their product is always positive. Fig. 2 illustrates this property qualitatively. Hence, we generally fulfill thermodynamics.

Effect of stress invariants on inelastic evolution. To motivate the choice of invariants as well as sub-potentials also from a more mechanical point of view, we additionally consider the inelastic volume, $I_3^{F_i}$, or rather its evolution over time. In this context, we found the analytical expression $\overline{(I_3^{F_i})} = I_3^{U_i} \text{tr}(\bar{\mathbf{D}}_i)$, revealing that the evolution over time can be expressed in terms of the evolution Eq. (10). With this expression at hand, it becomes obvious that the inelastic volume ratio is related to the hydrostatic pressure, $I_1^{\bar{\mathbf{F}}}$, which seems intuitive. This means that as long as $\frac{\partial g_1(I_1^{\bar{\mathbf{F}}})}{\partial I_1^{\bar{\mathbf{F}}}} = 0$, the volume ratio must remain constant. For our specific choice made in Eq. (18), we found that $g_1(I_1^{\bar{\mathbf{F}}}) \neq \text{const.}$ is strongly related to the volume ratio over time. In addition, from a computational point of view, this guides our choice of time discretization schemes for the evolution equation, which will be discussed in Section 4.

3. Inelastic constitutive artificial neural networks

The previous sections have set the physical boundaries for the constitutive equations to obtain thermodynamically consistent results. In the following, we will design an iCANN architecture that inherently satisfies these requirements. We will discuss the inputs and outputs in each time step, the general network architecture as well as the feed-forward architectures of the energy and potential networks including their individual activation functions.

iCANN inputs and outputs. Within a time step, the overall inputs for both networks are $\mathbf{C}$ and the time increment Δt . Moreover, since our aim is to discover history-dependent materials, the inelastic stretches, $\mathbf{U}_i$, are passed through the hidden states. In addition,

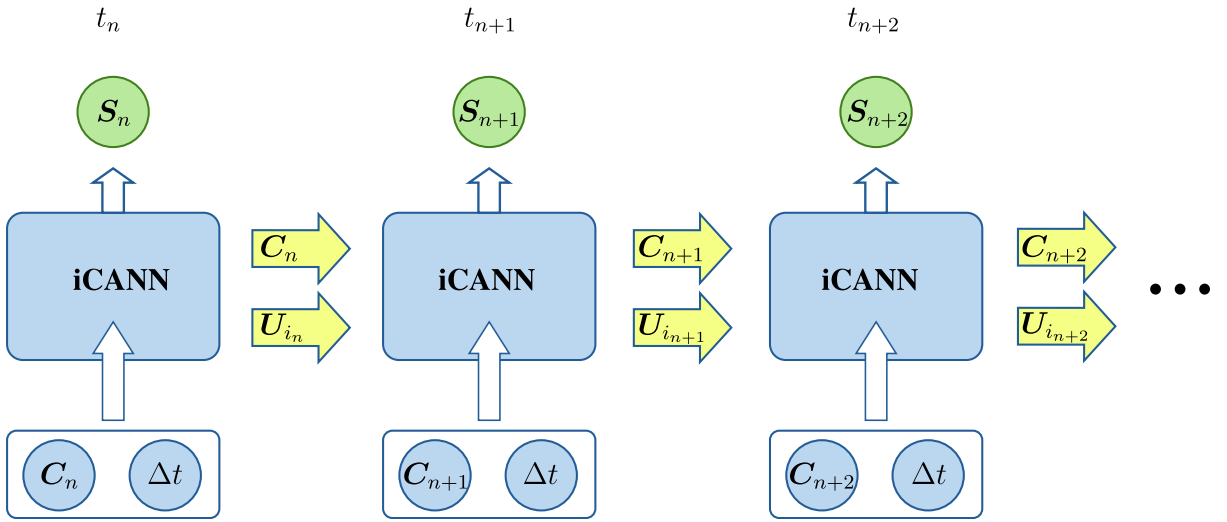


Fig. 3. Schematic illustration of recurrent neural network for iCANN. The history dependence of the material is taken into account by the hidden state of the recurrent neural network. Thus, the inelastic stretches, U_i , as well as the previous total stretches, C , are propagated through time. Fig. 4 illustrates the iCANN architecture in each time step.

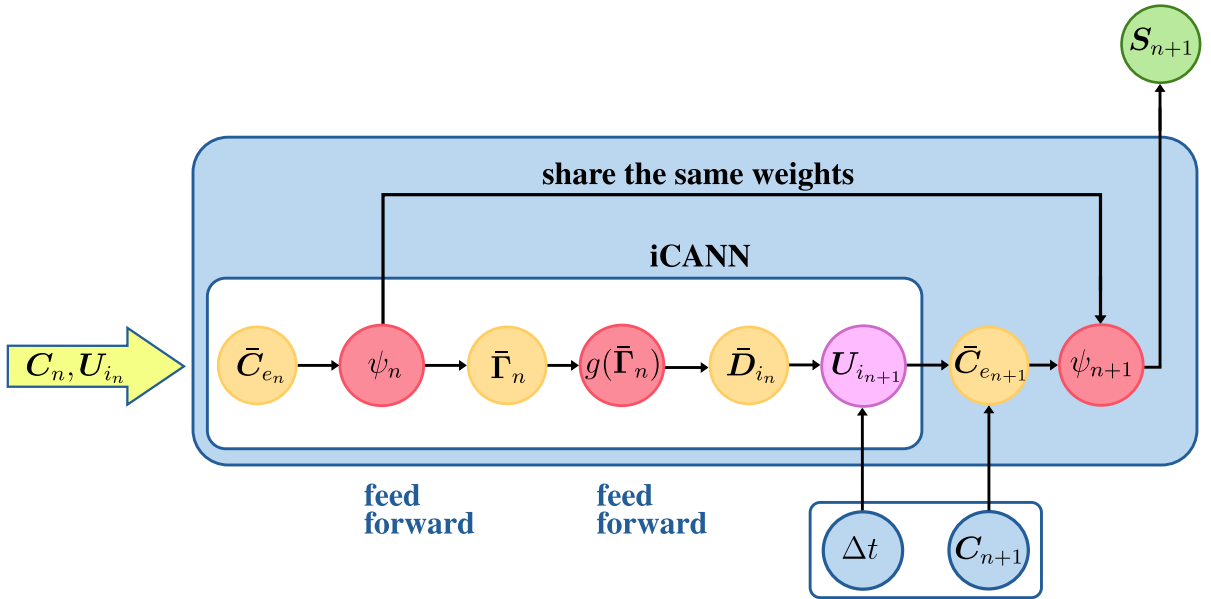


Fig. 4. Schematic illustration of the iCANN architecture. Color code: blue — current inputs; orange — basic calculation; red — feed-forward network; yellow — hidden state variables; purple — time discretization (Eq. (27)); green — current output. It is important to note that the feed-forward networks for both ψ_n and ψ_{n+1} have the same weights. Thus, there is no double set of weights for the energy. The architecture of the energy network is illustrated in Fig. 5, Fig. 7 shows the architecture of the feed-forward network of the potential. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

as we will discuss in detail in Section 4, we will use an explicit time integration scheme. Thus, C corresponding to the last time step is passed through the hidden states as well. While $\bar{C}_e$ serves as the input for the energy network, the input for the potential network is $\bar{\Gamma}$. Furthermore, considering the constitutive requirements of objectivity, rigid motion of the reference configuration, and rotational non-uniqueness, the basis for both networks are the invariants of these tensors.

Network architecture. The network architecture combines a recurrent neural network with two individual feed-forward networks for the Helmholtz free energy as well as the pseudo potential. Both feed-forward networks are inspired by the classical CANN approach (see Linka et al. [44]). As alluded in Linka and Kuhl [1], both networks are not fully connected. Fig. 3 illustrates schematically the recurrent architecture at sequential time steps. Noteworthy, Δt does not have to be constant rather it is the

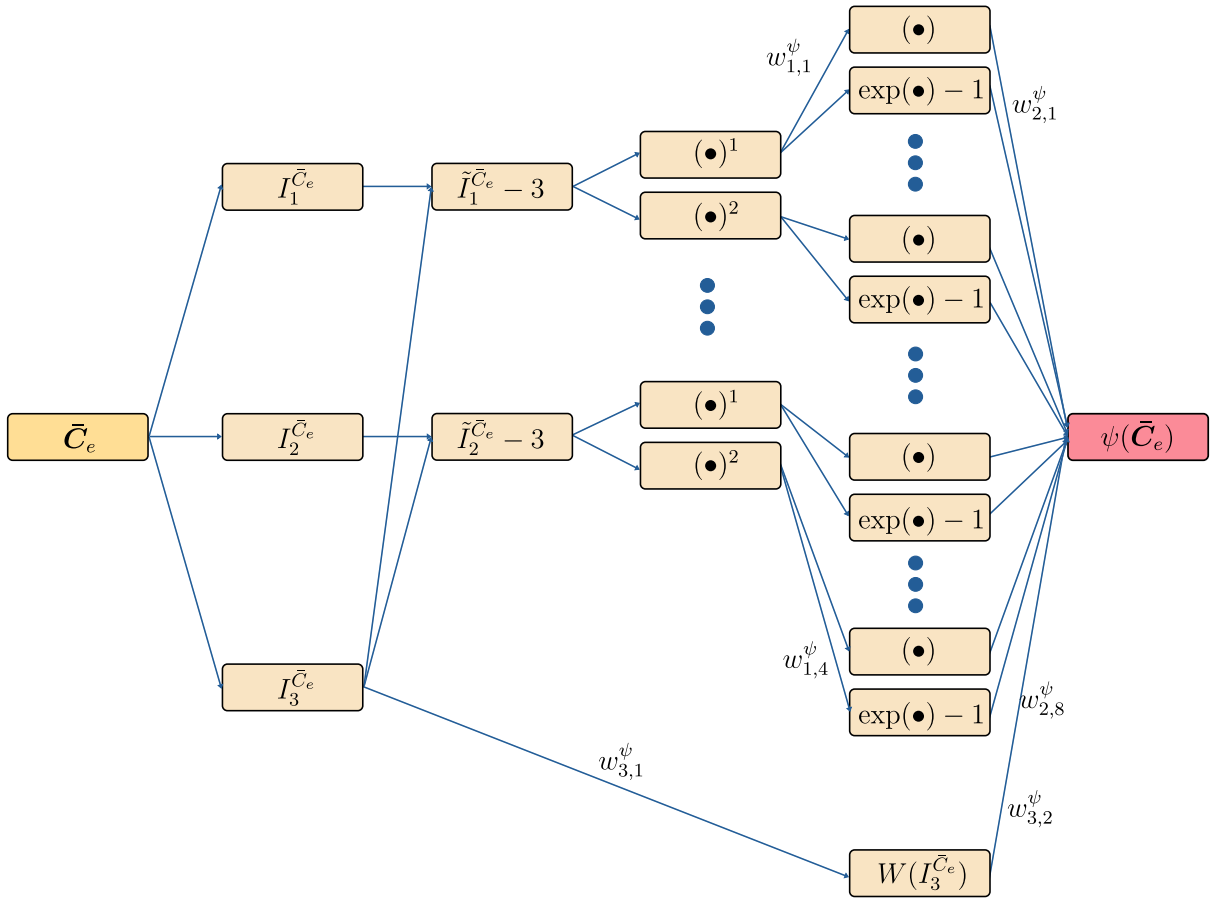


Fig. 5. Schematic illustration of the feed-forward network architecture of the elastic and compressible Helmholtz free energy, ψ , within the recurrent neural network. The first layer computes the invariants, the second employs the volumetric–isochoric split, the third generates the powers $(\bullet)^1$ and $(\bullet)^2$, and the fourth applies the custom-designed activation functions identity $(\bullet)$ and the exponential $\exp(\bullet) - 1$. Note that the network is not fully connected by design.

increment between the previous, t_n , and current time step, t_{n+1} . For instance, at t_{n+1} we find $\Delta t = t_{n+1} - t_n$. To solve the evolution Eq. (10), we integrate explicitly in time. If the evolution equation were solved implicitly, this would require additional iterations to solve for the inelastic stretch tensor, which would certainly increase the numerical effort during training. A more detailed illustration of the iCANN structure is provided in Fig. 4. First, from the previous values $\mathbf{C}_n := \mathbf{C}(t = t_n)$ and $\mathbf{U}_{i_n} := \mathbf{U}_i(t = t_n)$ we obtain the previous elastic right Cauchy Green tensor, $\bar{\mathbf{C}}_{e_n}$. We then evaluate the Helmholtz free energy, $\psi_n := \psi(t = t_n)$, to obtain the relative stress at the previous time step, which is required for an explicit integration of the evolution equation. Then, the pseudo potential at the previous time step, g_n , can be evaluated from which we obtain the right hand side of the explicit time-discretized evolution equation (cf. Eq. (10)) by means of algorithmic differentiation. The result of this time integration allows us to obtain the current value $\mathbf{U}_{i_{n+1}} := \mathbf{U}_i(t = t_{n+1})$. Finally, we can update the elastic right Cauchy Green tensor, $\bar{\mathbf{C}}_{e_{n+1}}$, and also evaluate our Helmholtz free energy at the current time step, ψ_{n+1} , from which we derive the current second Piola–Kirchhoff stress, $\mathcal{S}_{n+1}$, using algorithmic differentiation. For the numerical implementation it is crucial that ψ_n and ψ_{n+1} are evaluated with exactly the same set of weights. Otherwise we will not learn a unique set of weights for the Helmholtz free energy.

Helmholtz free energy feed-forward network. The feed-forward network for the energy that we are using is shown in Fig. 5. We extend the CANN structure of Linka and Kuhl [1] to compressible solids by decoupling the volumetric and isochoric response (cf. Eq. (15)).⁵ Due to this decoupling and the discussions in Section 2, our network is considered *not fully connected*. Within this contribution, the explicit form we choose for $W(I_3^{\bar{\mathbf{C}}_e})$ is taken from Ogden and Hill [113].⁶ Thus, the equation for the Helmholtz

⁵ We kindly refer the interested reader to the aforementioned publication for a more detailed explanation.

⁶ For an overview of alternative choices, the reader is referred to Hartmann and Neff [108].

free energy within our networks takes the form

$$\begin{aligned} \psi \left(\bar{I}_1^{\bar{C}_e}, \bar{I}_2^{\bar{C}_e}, \bar{I}_3^{\bar{C}_e} \right) &= w_{2,1}^w (\bar{I}_1^{\bar{C}_e} - 3) + w_{2,2}^w \left[\exp(w_{1,1}^w (\bar{I}_1^{\bar{C}_e} - 3)) - 1 \right] \\ &+ w_{2,3}^w (\bar{I}_1^{\bar{C}_e} - 3)^2 + w_{2,4}^w \left[\exp(w_{1,2}^w (\bar{I}_1^{\bar{C}_e} - 3)^2) - 1 \right] \\ &+ w_{2,5}^w (\bar{I}_2^{\bar{C}_e} - 3) + w_{2,6}^w \left[\exp(w_{1,3}^w (\bar{I}_2^{\bar{C}_e} - 3)) - 1 \right] \\ &+ w_{2,7}^w (\bar{I}_2^{\bar{C}_e} - 3)^2 + w_{2,8}^w \left[\exp(w_{1,4}^w (\bar{I}_2^{\bar{C}_e} - 3)^2) - 1 \right] \\ &+ w_{3,2}^w \underbrace{\left[\left(\bar{I}_3^{\bar{C}_e} \right)^{-w_{3,1}^w} - 1 + w_{3,1}^w \ln(\bar{I}_3^{\bar{C}_e}) \right]}_{w(\bar{I}_3^{\bar{C}_e})} + \dots \end{aligned} \quad (20)$$

As discussed in Section 2.1, we approximate the energy by a Taylor series within one layer of the network. The subsequent layer applies particularly designed activation functions, which can be thought of as custom-designed activation functions (see As'ad et al. [27]). This allows us to broaden the functional space, while using less high numbers of exponents of the power series in Eq. (16). In line with Linka and Kuhl [1], we utilize the following custom-designed activation functions: linear ($\bullet$) and exponential $\exp(\bullet) - 1$. These functions are monotonic, continuous, continuous differentiable, and zero at the origin.

As shown in Fig. 5, we introduce two types of weights here. The first, $w_{1,i}^w$ and $w_{3,1}^w$, control the shape of the activation function, while the second, $w_{2,i}^w$ and $w_{3,2}^w$, scale the individual contribution. Note that due to the linearity of the first activation function, we can reduce the weights to one single weight for the linear activation, e.g., $w_{2,1}^w$. In order to ensure physically reasonable results, all weights except $w_{3,1}^w$ are constraint to be greater or equal to zero. Noteworthy, $w_{3,1}^w = 0$ would violate Eq. (14), however, any solid material is considered to be resistant against volumetric deformations. Thus, this case can be excluded a priori. Nevertheless, to be sure about this, one could restrict $w_{3,1}^w$ to be greater than zero.

Pseudo potential feed-forward network. We use a similar structure to that of the CANN to design the feed-forward architecture of the pseudo potential. The underlying idea is to find a general expression for the potential that is physically interpretable and that also satisfies the requirements given in Section 2.2 by design. Therefore, we custom-design activation functions that satisfy a zero-value potential for zero stress, are convex for inputs of any power since we expand the potential by a Taylor series, and are non-negative for any positive and negative arguments. One possible set of activation functions we are choosing is shown in Fig. 6: absolute value $\text{abs}(\bullet)$, natural logarithm of the hyperbolic cosine $\ln(\cosh(\bullet))$, and the hyperbolic cosine $\cosh(\bullet) - 1$. Like the activation functions for the energy, these functions are continuous and continuously differentiable. Moreover, to guarantee convexity, we do not use a fully-connected feed-forward network, i.e., we do not include mixed products of the invariants (cf. Eq. (18)).

It remains that we introduce a set of possible weights. To this end, we take inspiration from the CANN design, which means we introduce two types of weights. On the one hand, the shape of each activation function is controlled by a weight, $w_{1,i}^g$, and, in addition, the individual contribution of each function is scaled by a second weight, $w_{2,i}^g$. Due to the properties of the absolute activation function, a weight controlling the shape is redundant. Thus, we can again reduce the set of weights, since only one weight is needed for this very activation function. We also restrict the range of values of the weights to be greater than or equal to zero for thermodynamic considerations. All of these considerations lead us to the design of the potential network shown schematically in Fig. 7, or its explicit form

$$\begin{aligned} g \left(\bar{I}_1^{\bar{F}}, \bar{I}_2^{\bar{F}}, \bar{J}_3^{\bar{F}} \right) &= w_{2,1}^g \text{abs}(\bar{I}_1^{\bar{F}}) + w_{2,2}^g \ln(\cosh(w_{1,1}^g \bar{I}_1^{\bar{F}})) + w_{2,3}^g \left[\cosh(w_{1,2}^g \bar{I}_1^{\bar{F}}) - 1 \right] \\ &+ w_{2,4}^g \text{abs}((\bar{I}_1^{\bar{F}})^2) + w_{2,5}^g \ln(\cosh(w_{1,3}^g (\bar{I}_1^{\bar{F}})^2)) + w_{2,6}^g \left[\cosh(w_{1,4}^g (\bar{I}_1^{\bar{F}})^2) - 1 \right] \\ &+ w_{2,7}^g \text{abs}(\bar{I}_2^{\bar{F}}) + w_{2,8}^g \ln(\cosh(w_{1,5}^g \bar{I}_2^{\bar{F}})) + w_{2,9}^g \left[\cosh(w_{1,6}^g \bar{I}_2^{\bar{F}}) - 1 \right] \\ &+ w_{2,10}^g \text{abs}((\bar{I}_2^{\bar{F}})^2) + w_{2,11}^g \ln(\cosh(w_{1,7}^g (\bar{I}_2^{\bar{F}})^2)) + w_{2,12}^g \left[\cosh(w_{1,8}^g (\bar{I}_2^{\bar{F}})^2) - 1 \right] \\ &+ w_{2,13}^g \text{abs}(\bar{J}_3^{\bar{F}}) + w_{2,14}^g \ln(\cosh(w_{1,9}^g \bar{J}_3^{\bar{F}})) + w_{2,15}^g \left[\cosh(w_{1,10}^g \bar{J}_3^{\bar{F}}) - 1 \right] \\ &+ w_{2,16}^g \text{abs}((\bar{J}_3^{\bar{F}})^2) + w_{2,17}^g \ln(\cosh(w_{1,11}^g (\bar{J}_3^{\bar{F}})^2)) + w_{2,18}^g \left[\cosh(w_{1,12}^g (\bar{J}_3^{\bar{F}})^2) - 1 \right] \\ &+ \dots \end{aligned} \quad (21)$$

which can be interpreted as the sum of convex potentials. In a more general context, this design is a generic convex potential (cf. [114])

$$g(\bar{F}) = \left(\sum_{i=1}^n \text{abs}(f_i(\bar{F}))^p \right)^{\frac{1}{p}} \quad (22)$$

with ' p -norm' equal to one. In the latter equation, f_i are convex functions of the stress $\bar{F}$. Noteworthy, since our custom-designed activation functions are all positive, one could exclude the $\text{abs}(\bullet)$ in Eq. (22). Remarkably, changing the ' p -norm' regularizes the corners of intersecting yield criteria/potentials. Furthermore, varying the ' p -norm' allows us to consider a wider range of pseudo potentials using the same set of activation functions, which is seen as an easy way to extend our approach. In case we would like to

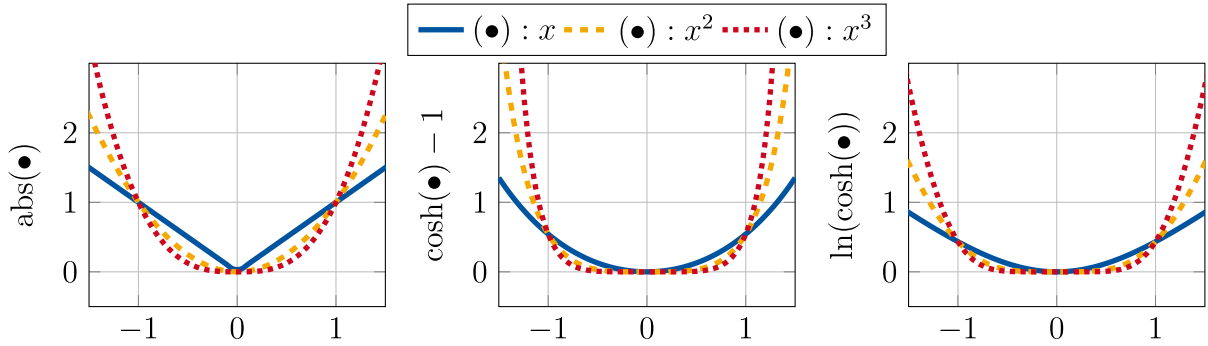


Fig. 6. Particular choices of custom-designed activation functions for the pseudo potential, g , with linear, quadratic, and cubic inputs. As required, all functions are convex, non-negative, and zero-valued.

implement the ‘ p -norm’ into the network of the pseudo potential, we would introduce an additional weight for it and restrict it to be greater than zero.

The purpose of CANNs is to provide a simple network architecture, which is also physically sound and interpretable. The extension made to iCANNs by the use of recurrent neural networks for the history dependence as well as the network of the pseudo potential follows exactly this way. All quantities within the ‘overall’ network are physically interpretable⁷ and can be clearly assigned to the influence of strains on stresses (Section 2.1) or to the influence of stresses on inelastic strains (Section 2.2). More precisely, the weights w^w reflect the resistance of the material to deformation (see Eq. (9) and the corresponding assumption on S), while the rate of inelastic flow is described by w^s (see Eq. (10)). The interested reader is referred to the work of Linka and Kuhl [1] where it is shown how to link CANNs and known constitutive models from the literature.

4. Specialization: Incompressibility, visco-elasticity and time discretization

In the previous chapters we have discussed the constitutive properties of general inelastic materials from a theoretical point of view as well as its particular implementation and incorporation into a recurrent neural network. In the following, we will now specify the material equations and architecture of the network for the application of visco-elastic material behavior.

Incompressibility and loadings. For the time being, we restrict ourselves to the case of incompressible material behavior, i.e., $I_3^C = 1$. In order to ensure this condition within our studies, we add a Lagrange term, $-p(I_3^C - 1) = -p(I_3^{C_e} I_3^{C_i} - 1)$, to the overall Helmholtz free energy ψ_0 . In the latter, p can be interpreted as the hydrostatic pressure and is calculated from the boundary conditions. It is important to note that although the overall material behavior is incompressible, the elastic response might not. Only if $I_3^{C_i} = 1$ remains constant, we obtain elastic incompressibility. Furthermore, we only investigate coaxial loads, i.e., uniaxial loading ($C = \text{diag}[C_{11}, 1/\sqrt{C_{11}}, 1/\sqrt{C_{11}}]$), equibiaxial loading ($C = \text{diag}[C_{11}, C_{11}, 1/C_{11}^2]$), and pure shear ($C = \text{diag}[C_{11}, 1, 1/C_{11}]$) (see Steinmann et al. [38]). Hence, the principal axes do not rotate during loading, and thus, both $\bar{C}_e$ and C_i are diagonal tensors as well (cf. [115]). Finally, we calculate p by the condition that $S_{33} = 0$ must hold for each of these loading cases.

Maxwell element. In terms of rheology, visco-elastic material behavior can be represented as a spring and dashpot element in serial connection, which is schematically illustrated in Fig. 8. Additionally, since we consider incompressible behavior, we find the following overall energy

$$\psi_0 = \psi^{\text{Neq}}(\bar{C}_e) - p(I_3^{C_e} I_3^{C_i} - 1) \quad (23)$$

where ψ^{Neq} belongs to the spring element. Noteworthy, having Eq. (11) in mind, we observe that $\bar{F} = \bar{S} - \bar{X} \equiv 2\bar{C}_e \frac{\partial \psi^{\text{Neq}}(\bar{C}_e)}{\partial \bar{C}_e} =: \bar{S}^{\text{Neq}}$. Thus, we realize that the driving force for inelastic effects, $\bar{S}^{\text{Neq}}$, is independent from the Lagrange multiplier, p , and express the reduced dissipation inequality as $\bar{S}^{\text{Neq}} : \bar{D} \geq 0$ for the case of visco-elastic, incompressible material behavior. In addition, we can rewrite the state law per Maxwell element as $S^{\text{Neq}} = 2U_i^{-1} \frac{\partial \psi^{\text{Neq}}(\bar{C}_e)}{\partial \bar{C}_e} U_i^{-1} - pC^{-1}$.

Moreover, as discussed in Section 2, for discovering the viscous response, we need to specify the evolution Eq. (10). For the time being, we consider γ to be constant.

Reduction of Maxwell element. For the studies we will conduct within this contribution, we design the feed-forward network of the Helmholtz free energy of the Maxwell element, $\psi^{\text{Neq}}(\bar{C}_e)$, to be the very same as shown in Eq. (20) without any further terms. Consequently, the weights of this network are: 4 (w_1^w) + 8 (w_2^w) + 2 (w_3^w) = 14 weights in total. In contrast, since the amount of experimental data is limited, we reduce the form of the potential shown in Eq. (21), which was for example also used by Taç et al.

⁷ For instance, they can be identified as the shear or bulk modulus of the material.

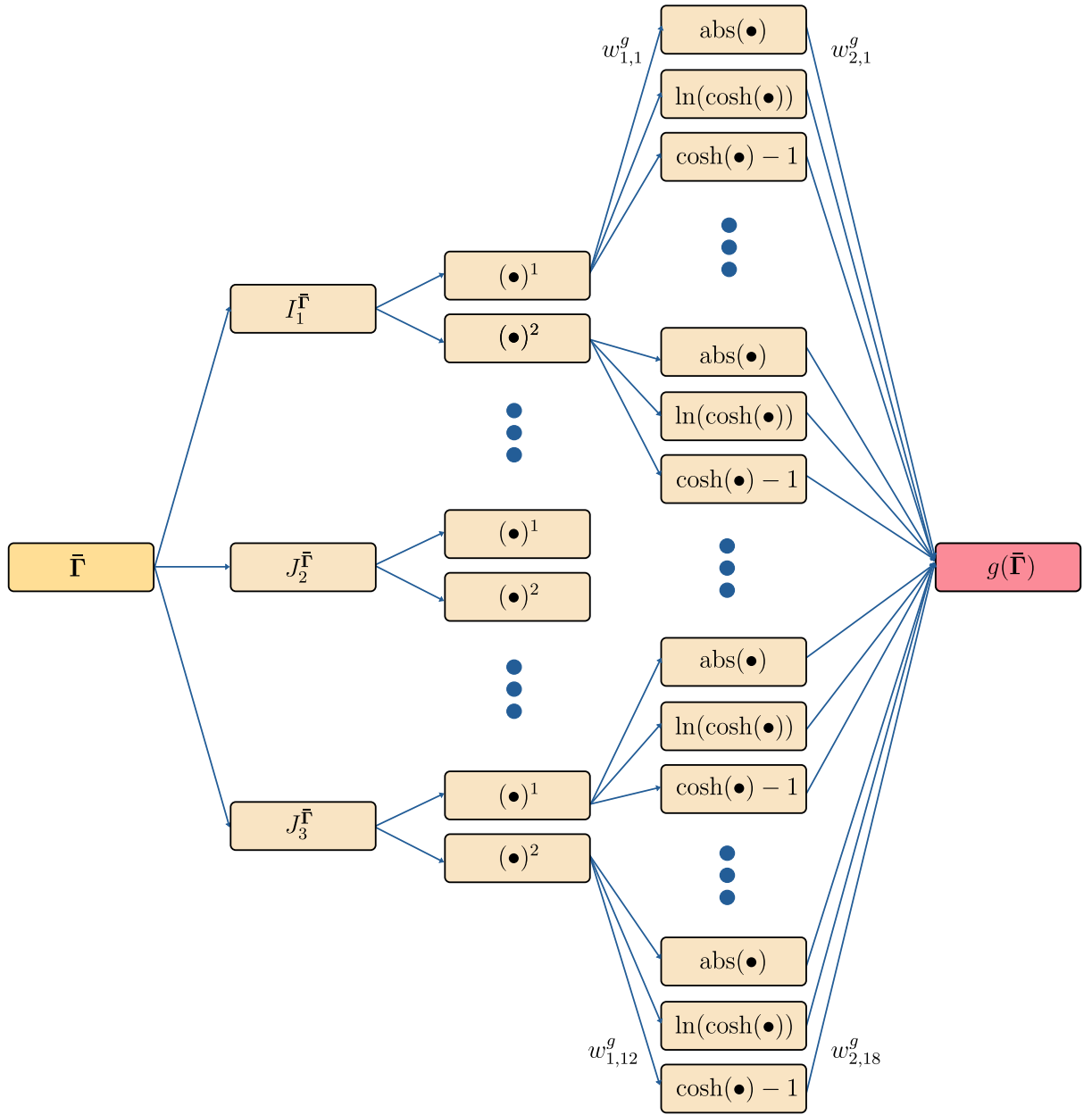


Fig. 7. Schematic illustration of the feed-forward network architecture of the pseudo potential, g , within the recurrent neural network. The first layer computes the invariants, the second generates the powers $(\bullet)^1$ and $(\bullet)^2$, and the third applies the custom-designed activation functions absolute value $\text{abs}(\bullet)$, logarithm/hyperbolic cosine $\ln(\cosh(\bullet))$, and the hyperbolic cosine $\cosh(\bullet) - 1$. To satisfy convexity a priori, the network is not fully connected by design.

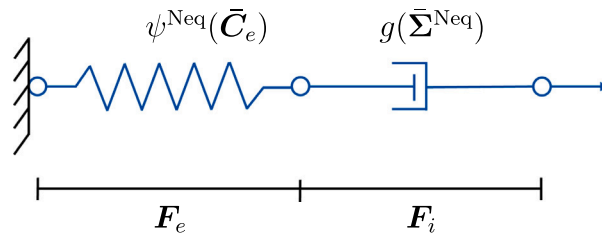


Fig. 8. Schematic illustration of a single Maxwell element. Due to the history dependence and the inelastic evolution, the energy associated with the spring element, ψ^{Neq} , is not instantaneously in equilibrium.

[116]. In addition, also in line with the latter mentioned work, we replace $J_2^{\tilde{\Sigma}^{\text{Neq}}}$ by a scaled version $\tilde{J}_2^{\tilde{\Sigma}^{\text{Neq}}} = 3 J_2^{\tilde{\Sigma}^{\text{Neq}}}$, yielding the following potential

$$\begin{aligned} g(I_1^{\tilde{\Sigma}^{\text{Neq}}}, \tilde{J}_2^{\tilde{\Sigma}^{\text{Neq}}}) &= w_{2,1}^g \text{abs}(I_1^{\tilde{\Sigma}^{\text{Neq}}}) + w_{2,2}^g \ln(\cosh(w_{1,1}^g I_1^{\tilde{\Sigma}^{\text{Neq}}})) \\ &+ w_{2,4}^g \text{abs}((I_1^{\tilde{\Sigma}^{\text{Neq}}})^2) + w_{2,5}^g \ln(\cosh(w_{1,3}^g (I_1^{\tilde{\Sigma}^{\text{Neq}}})^2)) \\ &+ \tilde{w}_{2,7}^g \text{abs}(\tilde{J}_2^{\tilde{\Sigma}^{\text{Neq}}}) + w_{2,8}^g \ln(\cosh(\tilde{w}_{1,5}^g \tilde{J}_2^{\tilde{\Sigma}^{\text{Neq}}})). \end{aligned} \quad (24)$$

Thus, for the potential we have in total: $3 (w_{1,1}^g) + 6 (w_{2,2}^g) = 9$ weights. Noteworthy, in comparison to Eq. (21), the relations $w_{2,7}^g = 3 \tilde{w}_{2,7}^g$ and $w_{1,5}^g = 3 \tilde{w}_{1,5}^g$ hold. At this point, an additional weight for the relaxation time, $\frac{1}{\gamma}$, would increase the overall number of weights in our network to 24. However, considering Eq. (10), we realize that this weight is obsolete if the relaxation time is constant, since the weights $w_{2,7}^g$ can be considered to have the dimension ‘stiffness divided by time’. Hence, the time-scale is included in the potential. Although this might be unintuitive in terms of mechanics, we can reduce the numerical effort within the training process of our network without changing the underlying constitutive framework.

Visco-elastic solid. A Maxwell element is not well suited to model visco-elastic solids. Therefore, to learn the materials’ responses of different solids in Section 5, we utilize a common approach in continuum mechanics. Here, we connect three Maxwell elements and an equilibrium spring in parallel, where the equilibrium spring behaves purely elastically. As a result, the Helmholtz free energy can be expressed as

$$\psi_0 = \sum_{\alpha=1}^3 \left(\psi_{\alpha}^{\text{Neq}}(\bar{C}_{e_{\alpha}}) - p_{\alpha}(I_3^{C_{e_{\alpha}}} I_3^{C_{i_{\alpha}}} - 1) \right) + \psi^{\text{Eq}}(I_1^C, I_2^C) - p(I_3^C - 1) \quad (25)$$

together with the potential

$$g_0 = \sum_{\alpha=1}^3 g_{\alpha} \left(I_1^{\tilde{\Sigma}^{\text{Neq}}}, \tilde{J}_2^{\tilde{\Sigma}^{\text{Neq}}} \right) \quad (26)$$

with α denoting the contribution of each Maxwell element. Fig. 9 schematically illustrates the iCANN used here. Note that the energy, ψ^{Eq} , depends solely on the right Cauchy Green tensor, C . Since the determinant of the latter is always equal to one here, we do not need to consider the invariant I_3^C within the equilibrium energy term. Thus, ψ^{Eq} is designed in analogy to Eq. (20), but without $W(I_3^C)$ (cf. [1]). In summary, our visco-elastic solid neural network contains 3×23 (Maxwell element) + 1×12 (equilibrium spring) = 81 weights and the state law reads $S = \sum_{\alpha=1}^3 \left(2 U_{i_{\alpha}}^{-1} \frac{\partial \psi_{\alpha}^{\text{Neq}}(C_{e_{\alpha}})}{\partial C_{e_{\alpha}}} U_{i_{\alpha}}^{-1} - p_{\alpha} C^{-1} \right) + 2 \frac{\partial \psi^{\text{Eq}}(I_1^C, I_2^C)}{\partial C} - p C^{-1}$.⁸

Time discretization. Regardless the material behavior of interest, we need to discretize the evolution equation within the time interval $t \in [t_n, t_{n+1}]$. As discussed in Section 2.2, we have to choose an algorithm that preserves the inelastic volume in absence of $I_1^{\tilde{\Sigma}}$. Therefore, we will use the class of exponential integrators, and further, we will integrate explicitly in time (cf. [78]). For our purpose within this contribution, this choice is sufficient and easier in terms of the numerical implementation. Exploiting that the exponential satisfies the identity $\exp(BAB^{-1}) = B \exp(A) B^{-1}$ if B is invertible and further algebraic reformulations (cf. [98]), we obtain the following

$$C_{i_{n+1}} = U_{i_n} \exp \left(\Delta t 2 \bar{D}_{i_n} \right) U_{i_n}, \quad U_{i_{n+1}} = + \sqrt{C_{i_{n+1}}} \quad (27)$$

with the time interval $\Delta t := t_{n+1} - t_n$. We recognize that the derivatives in Eq. (17) are all coaxial, and, due to $\exp(A+B) = \exp(A)\exp(B)$ for coaxial arguments, we can rewrite the exponential in Eq. (27) as

$$\begin{aligned} \exp(\Delta t 2 \bar{D}_i) &= \exp \left(\Delta t 2 \gamma \frac{\partial g_1(I_1^{\tilde{F}})}{\partial I_1^{\tilde{F}}} I \right) \exp \left(\Delta t 2 \gamma \frac{\partial g_2(J_2^{\tilde{F}})}{\partial J_2^{\tilde{F}}} \text{dev}(\tilde{F}) \right) \dots \\ &\exp \left(\Delta t 2 \gamma \frac{\partial g_3(J_3^{\tilde{F}})}{\partial J_3^{\tilde{F}}} \text{dev}(\text{dev}(\tilde{F})^2) \right) \end{aligned} \quad (28)$$

where the index n was omitted. Lastly, we take the identity $\det(\exp(A)) = \exp(\text{tr}(A))$ into account, and thus, prove that the exponential integrator satisfies $\det(C_{i_{n+1}}) = \exp \left(\Delta t 6 \gamma \left(\frac{\partial g_1(I_1^{\tilde{F}})}{\partial I_1^{\tilde{F}}} \right)_n \right) \det(C_{i_n})$.

5. Results

In the following, we apply our Maxwell iCANN (Eqs. (23)–(24)) as well as our visco-elastic iCANN (Eqs. (25)–(26)) to various experimental and artificially generated data sets. We implemented our iCANN into the open-source software library *TensorFlow* (see Abadi et al. [117]). Since this paper focuses on the design of thermodynamically consistent neural networks rather on questions

⁸ One could also combine all hydrostatic pressures, p_{α} and p , into one total pressure. However, since the Maxwell elements and the equilibrium spring are distinct from each other, it is convenient to separate these Lagrange functions then it comes to the numerical implementation.

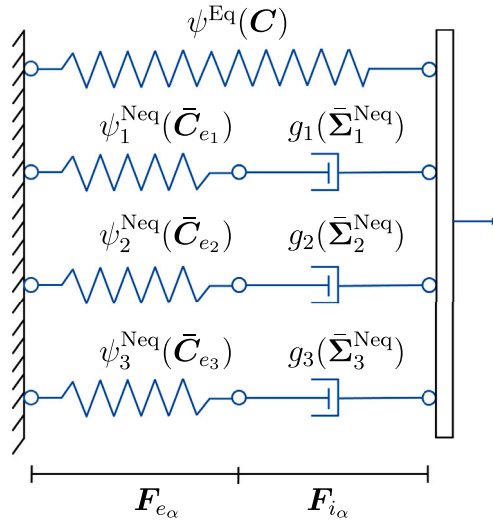


Fig. 9. Schematic illustration of a generalized Maxwell element describing a visco-elastic solid. While the overall deformation, $\mathbf{F}$, is the same for the equilibrium, ψ^{Eq} , and each of the non-equilibrium springs, ψ_a^{Neq} , the elastic parts, $\mathbf{F}_{e_\alpha}$, may vary.

related to the training procedure, we have set the L2-regularizer $L_2 = 0.001$ for all simulations. The relationship between potential overfitting and L2-regularization in the context of CANNs is explored, for example, by Wang et al. [3], St. Pierre et al. [118]. The loss is expressed in terms of least squares, i.e., the sum of the squared differences between the experimentally observed stresses and the stress predicted by the iCANN (cf. [3]). In addition, we use two error measures, i.e. the normalized root mean squared error, ϵ , and the coefficient of determination, R^2 , to evaluate the performance and accuracy of our iCANN

$$\epsilon := \frac{1}{|\bar{S}|} \sqrt{\frac{1}{n_{\text{data}}} \sum_{i=1}^{n_{\text{data}}} (S_{11_i} - \hat{S}_{11_i})^2}, \quad R^2 := \max \left(0, 1 - \frac{\sum_{i=1}^{n_{\text{data}}} (S_{11_i} - \hat{S}_{11_i})^2}{\sum_{i=1}^{n_{\text{data}}} (\bar{S} - \hat{S}_{11_i})^2} \right). \quad (29)$$

In the latter, S_{11} denotes the iCANN stress response, the experimentally observed stress is indicated by $\hat{S}_{11}$, while n_{data} is the number of data points. Further, $|\bar{S}| = \frac{1}{n_{\text{data}}} \sum_{i=1}^{n_{\text{data}}} |\hat{S}_{11_i}|$ denotes the sum of absolute values of the experimentally observed stress and $\bar{S} = \frac{1}{n_{\text{data}}} \sum_{i=1}^{n_{\text{data}}} \hat{S}_{11_i}$ is its mean. In all examples, the ADAM optimizer is used for training the network. The weights we discover during training for each example can be found in Appendix A.1–A.3.

5.1. Example 1: Artificially generated data

We start by creating artificial data using a Maxwell model. Consequently, the iCANN used corresponds to the Maxwell network (cf. Eqs. (23)–(24)). We use a continuum mechanical model to generate our artificial data. This constitutive framework is closed by introducing a Helmholtz free energy function of the elastic stretch, a pseudo potential, and the definition of the evolution equation

$$\psi^{\text{Neq}}(\bar{\mathbf{C}}_e) = \frac{\mu}{2} \left(\frac{\text{tr}(\bar{\mathbf{C}}_e)}{\det(\bar{\mathbf{C}}_e)^{1/3}} - 3 \right) + \frac{K}{25} (\det(\bar{\mathbf{C}}_e) - 1 - \ln(\det(\bar{\mathbf{C}}_e))) \quad (30)$$

$$g(\bar{\Sigma}^{\text{Neq}}) = \frac{1}{4\mu} \text{tr} \left(\text{dev}(\bar{\Sigma}^{\text{Neq}})^2 \right) + \frac{1}{18K} \text{tr}(\bar{\Sigma}^{\text{Neq}})^2 \quad (31)$$

$$\bar{\mathbf{D}}_i = \frac{1}{\tau} \frac{\partial g}{\partial \bar{\Sigma}^{\text{Neq}}}. \quad (32)$$

In the latter, $\mu = 12.5$ [kPa] is the shear modulus, $K = 25$ [kPa] is the material's bulk modulus, and $\tau = 10$ [s] denotes the relaxation time. We choose a compressible Neo-Hookean model, while the potential is taken from literature (see Reese and Govindjee [63]). For the numerical implementation, we use the algorithmic differentiation tool *AceGen* (see Korelc [119]) and calculate the matrix exponential by the closed-form expression provided by Korelc and Stupkiewicz [120].

We load the model monotonically using uniaxial tension (Fig. 10(a)), uniaxial compression (Fig. 10(b)), equibiaxial tension (Fig. 10(c)), pure shear (Fig. 10(d)), and uniaxial cyclic tension/compression (Fig. 10(e)). The first four loadings are relaxation tests, i.e., the load is applied within 0.5 [s] and then kept constant. For cyclic loading, we utilize three cycles: (i) increase stretch until $C_{11}^{\text{max}} = 1.2$ at $t = 0.4$ [s], (ii) increase stretch until $C_{11}^{\text{max}} = 2.1$ at $t = 1.2$ [s], decrease stretch until $C_{11}^{\text{max}} = 0.5$ at $t = 1.6$ [s].

The iCANN is simultaneously trained using the four monotonic relaxation tests (Fig. 10(a)–10(d)). Here, we used 10,000 epochs for training. We observe that the iCANN is able to discover a model that explains our training data almost exactly, which can be

Table 1

Normalized root mean squared error, ϵ , corresponding to the results shown in Fig. 11. The first row represents the results of training, the remaining three rows are the values of testing. In [121] no experimental data are provided for the middle rate with $C_{11}^{\max} = 9.0$ [–].

ϵ		$\dot{F}_{11}$ [1/s]		
		0.05	0.03	0.01
$C_{11}^{\max}$	9.0	0.02	–	0.04
	2.25	0.09	0.07	0.07
	4.0	0.04	0.05	0.14
	6.25	0.16	0.17	0.13

seen qualitatively in the stress–time curves and quantitatively in terms of ϵ and R^2 (see Fig. 10). We test the performance of the iCANN using the cyclic data set. The prediction is in almost perfect agreement with the artificially generated data. It is important to mention that the loading/unloading rate is not constant. This property is accurately reproduced by the model, although only relaxation tests are used for training.

5.2. Example 2: Discovering a model for the polymer VHB 4910 subjected to cyclic loading

Having validated our network on artificially generated data, we are interested in how well our approach can learn and predict experimental data. Therefore, we take experimental data from Hossain et al. [121] for the very-high-bond (VHB) 4910 polymer. Note that these data have already been used in a visco-elastic CANN by Abdolazizi et al. [45]. However, they use a generalized Maxwell approach to account for visco-elasticity. Contrary to our approach, their CANN is not based on pseudo potentials. Hence, it is not straightforward to extend it to further inelastic phenomena. Since polymers are considered visco-elastic, we use the visco-elastic network approach (Eqs. (25)–(26)).

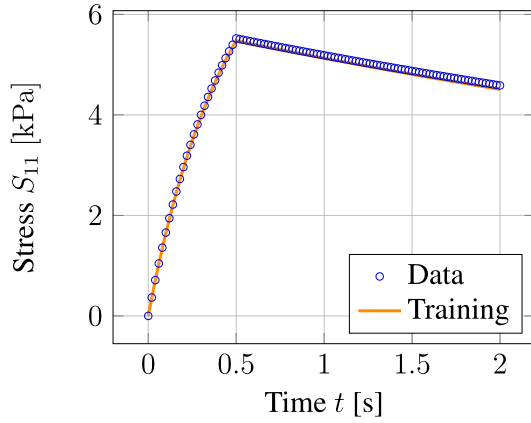
Within the experimental investigation, Hossain et al. [121] subjected the material to uniaxial loading–unloading cycles at different maximum stretches $C_{11}^{\max} = \{2.25, 4.0, 6.25, 9.0\}$ and three constant stretch rates $\dot{F}_{11} = \{0.01, 0.03, 0.05\}$ [1/s].⁹ For the maximum stretch $C_{11}^{\max} = 9.0$, the authors did not provide any experimental data for the rate $\dot{F}_{11} = 0.03$ [1/s]. The experimental data, which we transferred to the second Piola–Kirchhoff stress as well as the right Cauchy–Green tensor, are shown in Fig. 11. In addition, the authors derived a classical constitutive model using the material model of Arruda and Boyce [122]. In their ‘hand-crafted’ model, they studied the performance of both a model with four Maxwell elements and a model with five Maxwell elements. In order to identify the corresponding material parameters, additional multi-step relaxation tests were conducted by the authors.

Similar to Abdolazizi et al. [45], we are not using any of the multi-step relaxation tests to learn the weights of our iCANN for the polymer. We set the epochs to 6,000 for this example. However, contrary to the latter mentioned authors, we use less experimental data for training. Here, we only use the data up to the maximum of $C_{11}^{\max} = 9.0$ with the two loading rates $\dot{F}_{11} = \{0.01, 0.05\}$ [1/s] for training (Fig. 11(a)). In contrast, Abdolazizi et al. [45] use the data for $C_{11}^{\max} = 2.25$ as well, but exclude the rate $\dot{F}_{11} = 0.03$ [1/s]. Therefore, any data with a rate of $\dot{F}_{11} = 0.03$ [1/s] is not seen by the iCANN during training. The remaining nine experimental curves (Fig. 11(b)–11(d)) serve for testing.

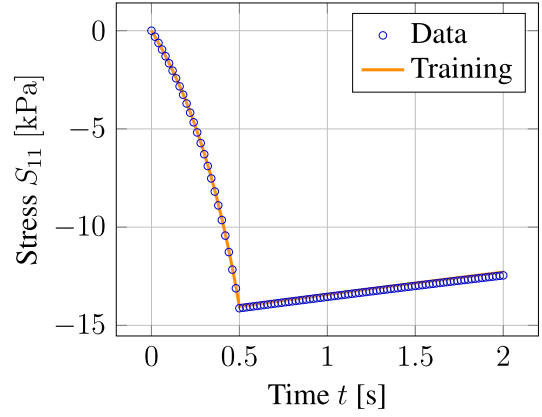
We observe reliable agreements with the results of the training process and the corresponding experimental data, not only qualitatively but also quantitatively. From a constitutive point of view, this indicates that our generic formulation of energy and potential is able to find a material model for real experimental data, not only for artificially generated data. The normalized root squared error for all experiments is provided in Table 1, the coefficient of determination is listed in Table 2. Furthermore, we recognize sound predictions for the unseen data at stretch levels $C_{11}^{\max} = \{2.25, 4.0\}$. This is particularly interesting since the data for the rate at which no training was performed is also accurately predicted. Despite these results, it should be noted that the data for the maximum stretch $C_{11}^{\max} = 6.25$ (Fig. 11(d)) as well as the lowest strain rate for a maximum stretch $C_{11}^{\max} = 4.0$ (Fig. 11(c)) are less closely matched. A plausible explanation for this mismatch was already given by Abdolazizi et al. [45]. Indeed, the experimental data are inconsistent here at least from a constitutive point of view: For constant loading rates but different maximum stretch levels applied, the stress–strain curves should be approximately the same. This is not the case here (see Appendix A.5). Thus, an artificial neural network obeying the laws of constitutive modeling is simply not capable of predicting such results. However, we see this not as a disadvantage, but rather as an advantage of iCANNs and CANNs, since inconsistencies in the experimentally measured data will not seriously affect the discovered model.

To evaluate the performance of our iCANN, Appendix A.5 shows the comparison between the iCANN and the ‘hand-crafted’ model of Hossain et al. [121] (see Fig. 15) as well as the comparison between the iCANN and the visco-elastic CANN developed by Abdolazizi et al. [45] (see Fig. 16). Overall, it can be said that the performance level of the iCANN is in a similar range to approaches from the literature. Noteworthy, in contrast to the model from Hossain et al. [121], we only used the cyclic experimental data for the model discovery analogously to the visco-elastic CANN by Abdolazizi et al. [45].

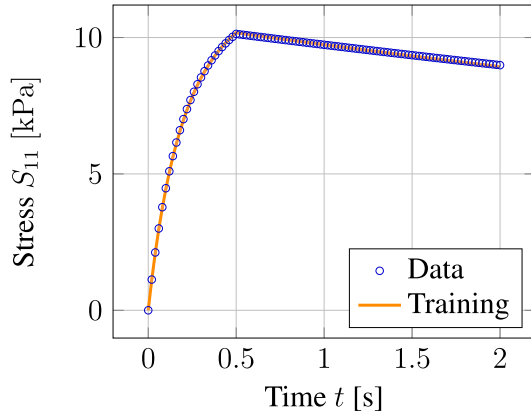
⁹ $C_{11} = F_{11}^2$ for uniaxial loading.



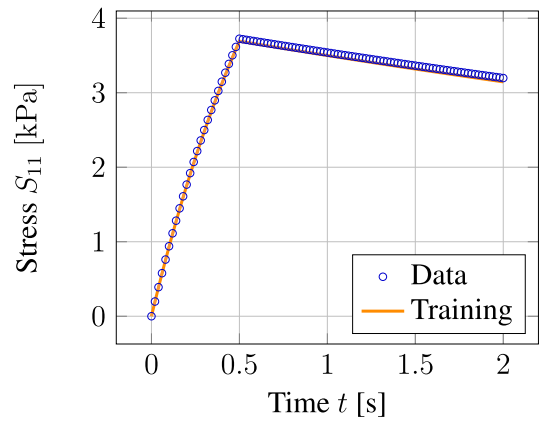
(a) Uniaxial tension ($\epsilon = 0.01$, $R^2 = 1.00$, $C_{11}^{\max} = 1.5$)



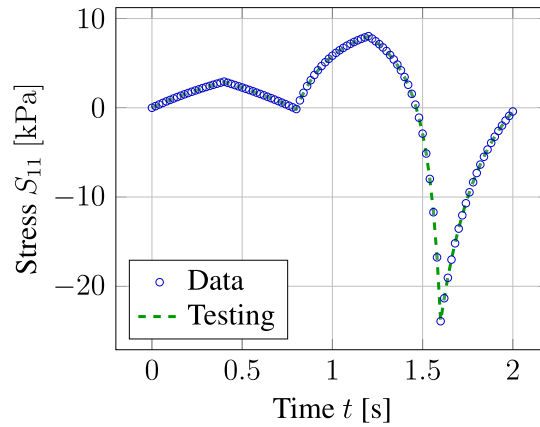
(b) Uniaxial compression ($\epsilon = 0.01$, $R^2 = 1.00$, $C_{11}^{\max} = 0.6$)



(c) Equibiaxial tension ($\epsilon = 0.00$, $R^2 = 1.00$, $C_{11}^{\max} = 1.8$)



(d) Pure shear ($\epsilon = 0.01$, $R^2 = 1.00$, $C_{11}^{\max} = 1.2$)



(e) Uniaxial cyclic loading ($\epsilon = 0.01$, $R^2 = 1.00$)

Fig. 10. Discovering a model for artificially generated data. For training, C_{11} is increased linearly within the first 0.5 seconds until $C_{11}^{\max}$ is reached and then kept constant for the rest of the simulation. For testing, the specimen is subjected to cyclic loading within five loading intervals.

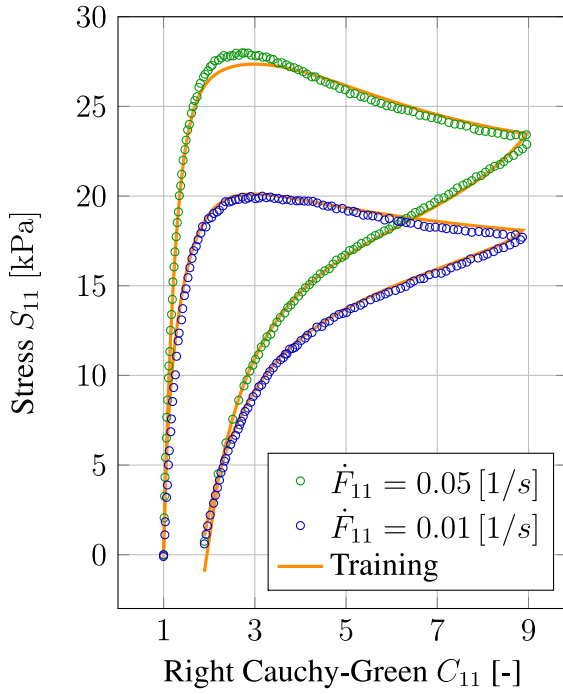
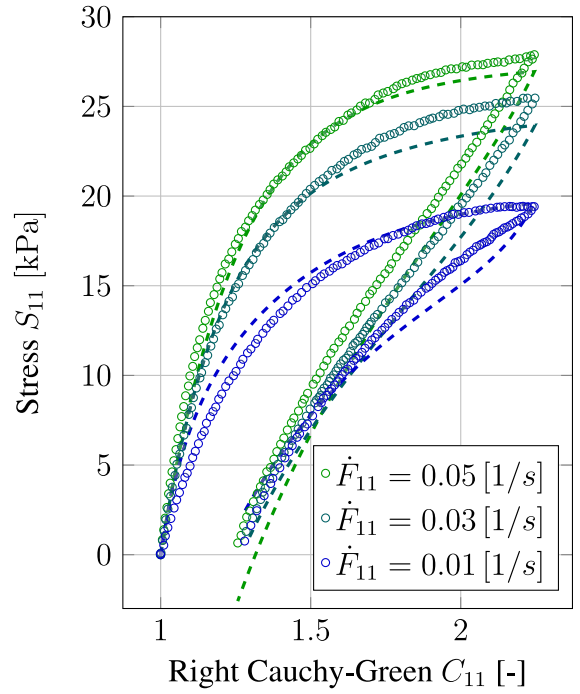
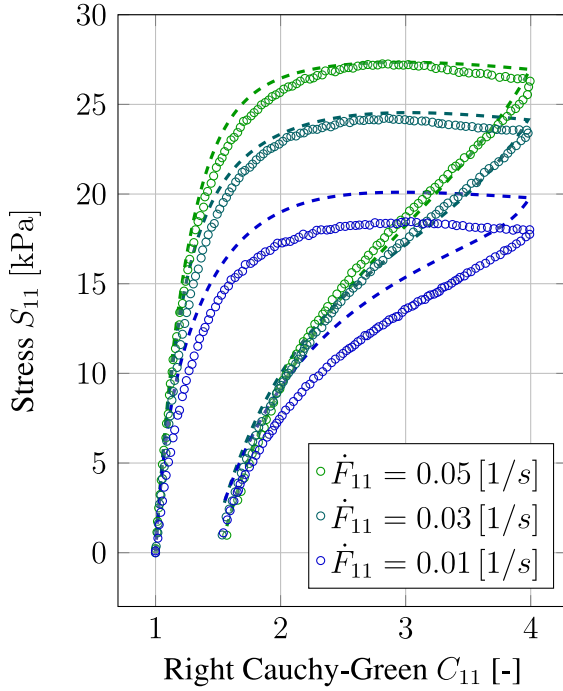
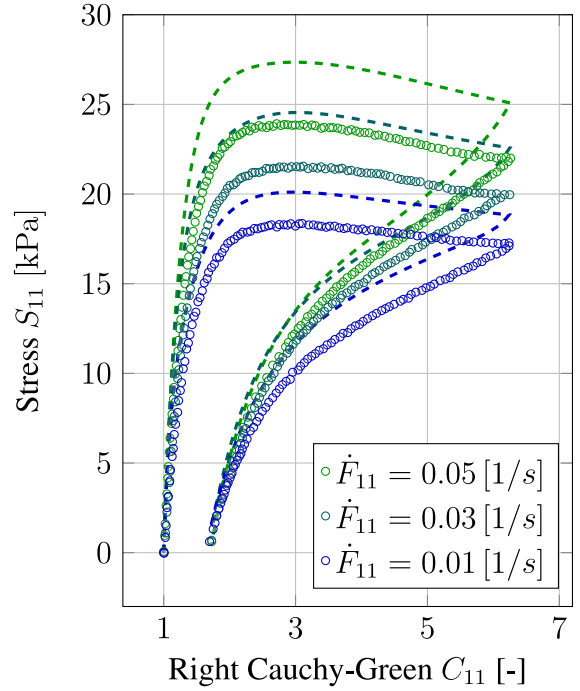
(a) Training with $C_{11}^{\max} = 9.0 [-]$ (b) Testing with $C_{11}^{\max} = 2.25 [-]$ (c) Testing with $C_{11}^{\max} = 4.0 [-]$ (d) Testing with $C_{11}^{\max} = 6.25 [-]$

Fig. 11. Discovering a model for the experimental data of the polymer VHB 4910 taken from Hossain et al. [121]. Each Fig. 11(a)–11(d) shows loading/unloading stress–stretch curves at constant deformation rates, $\dot{F}_{11}$, but different maximum applied stretch levels, $C_{11}^{\max}$. The experimental data are indicated by dots, the solid lines correspond to the training results of the network (Fig. 11(a)), and dashed lines represent the testing/prediction of the discovered model (Fig. 11(b)–11(d)).

Table 2

Coefficient of determination, R^2 , corresponding to the results shown in Fig. 11. The first row represents the results of training, the remaining three rows are the values of testing. In [121] no experimental data are provided for the middle rate with $C_{11}^{\max} = 9.0$ [–].

	R^2	$\dot{F}_{11}$ [1/s]		
		0.05	0.03	0.01
$C_{11}^{\max}$	9.0	1.00	–	0.99
	2.25	0.96	0.97	0.97
	4.0	0.99	0.99	0.88
	6.25	0.81	0.80	0.89

5.3. Example 3: Discovering a model for passive skeletal muscle subjected to relaxation

Finally, we are interested in discovering a model for passive skeletal muscle subjected to uniaxial compression. We take the data of Van Looke et al. [123], who performed relaxation tests. In total, they performed five different experiments on the same material, i.e. they varied both the maximum stretch $C_{11}^{\max} = \{0.81, 0.64, 0.49, 0.49, 0.49\}$ [–] applied for relaxation and the rate of deformation $\dot{F}_{11} = \{0.01, 0.01, 0.01, 0.005, 0.05\}$ [1/s]. The experimental results used here are the mean of six tests per data set.

Wang et al. [3] studied the performance of a Constitutive Artificial Neural Network combined with a Prony series approach to take visco-elasticity into account. In their investigations, Wang et al. [3] formulated the architecture of the CANN based on both principal stretches and the invariants of stretch as published in Linka and Kuhl [1]. For completeness, the results of the Prony-series CANN for both formulations are provided in Appendix A.6. In addition, the authors compared their CANN model with the performance of a ‘vanilla’ recurrent neural network, i.e., no physical or constitutive constraints are included in the design of the network. Although the performance of the training sets was more accurate compared to CANNs, the predictions generally provided non-physical results, which were in strong disagreement with thermodynamics. Thus, visco-elastic CANNs outperform classical neural network approaches because their results are most likely to violate physics. Another aspect investigated by the authors was the influence of increased data for training. One the one hand, they started by training the network using one experimental curve and test their discovered model using the remaining four experiments (see Fig. 17). On the other hand, training the network using four experimental data sets and just using one unseen data set for testing should generally increase the network’s performance (see Fig. 18). In the following, we will study the same using our iCANN approach. Regardless whether we train on one or on four experiments, we set the epochs to 5,000 in this example.

Train on one, test on four. To begin with, we train our iCANN on one experimental curve and test on four remaining data sets. Additionally, we vary the training data set with respect to the five different combination of maximum stretch and rate of deformation. The models we learned as well as the experimental data are shown in Fig. 12. Table 3 lists the normalized root mean squared error, while the coefficient of determination is given in Table 4. Generally, we can say that the training data sets are learned well using our iCANN approach, regardless of which experiment is used for training. Adding up ϵ per column in Table 3, we find the third column to perform the best ($\sum \epsilon = 1.21$). Nevertheless, both the ϵ and R^2 values of the tests in each column are relatively poor, suggesting that an optimal model has not yet been discovered. As possible causes we may consider either the limitation of the number of Maxwell elements to three or the insufficient amount of training data. Regarding the first reason, however, it can be assumed that no better model will be found if the training data remain the same, since the error measures are already accurate for the training data. Comparing our results to the Prony-series CANN by Wang et al. [3] in Fig. 17, a similar performance of our iCANN and Prony-series CANNs is observed.

Train on four, test on one. Consequently, we increase the data used for training. Our results are illustrated in Fig. 13 where we use four curves for training and test on the remaining experimental curve. Again, we calculate the normalized root mean squared error in Table 5 and the coefficient of determination in Table 6. As expected, we generally observe a more accurate learning performance compared to the setup ‘train on one, test on four’. Overall, both error measurements used here are in a relatively good range for training and testing. In contrast to the previous setup, where it was difficult to find a model that explained both training and testing, the results shown in the second and fifth columns agree well with the training data and especially the predictions are in line with the test data. The best model is discovered in the second column, where we have added up the normalized root mean squared error $\sum \epsilon = 0.68$. In view of the results obtained by Wang et al. [3] for both the principal stretch and invariant-based formulations (see Fig. 18), our iCANN is capable to predict the test data to a comparable degree of accuracy. As already mentioned above, we might obtain better results if we increase the number of Maxwell elements, however, this is out of the scope of this contribution.

Regardless whether we train with one or four experiments, our results infer that the choice of the training set is by no means unique or trivial. Different choices influence the accuracy of prediction capabilities. There are many reasons for this, including the quality of the measurement during the experimental investigation, the choice of optimizer, or how much information about inelastic effects is hidden in the data. Nonetheless, the results suggest that iCANNs are not only able to learn the training set well, as is also the case for ‘vanilla’ networks (cf. [3]), but also provide us with physically reasonable and accurate results when applied outside the training regime.

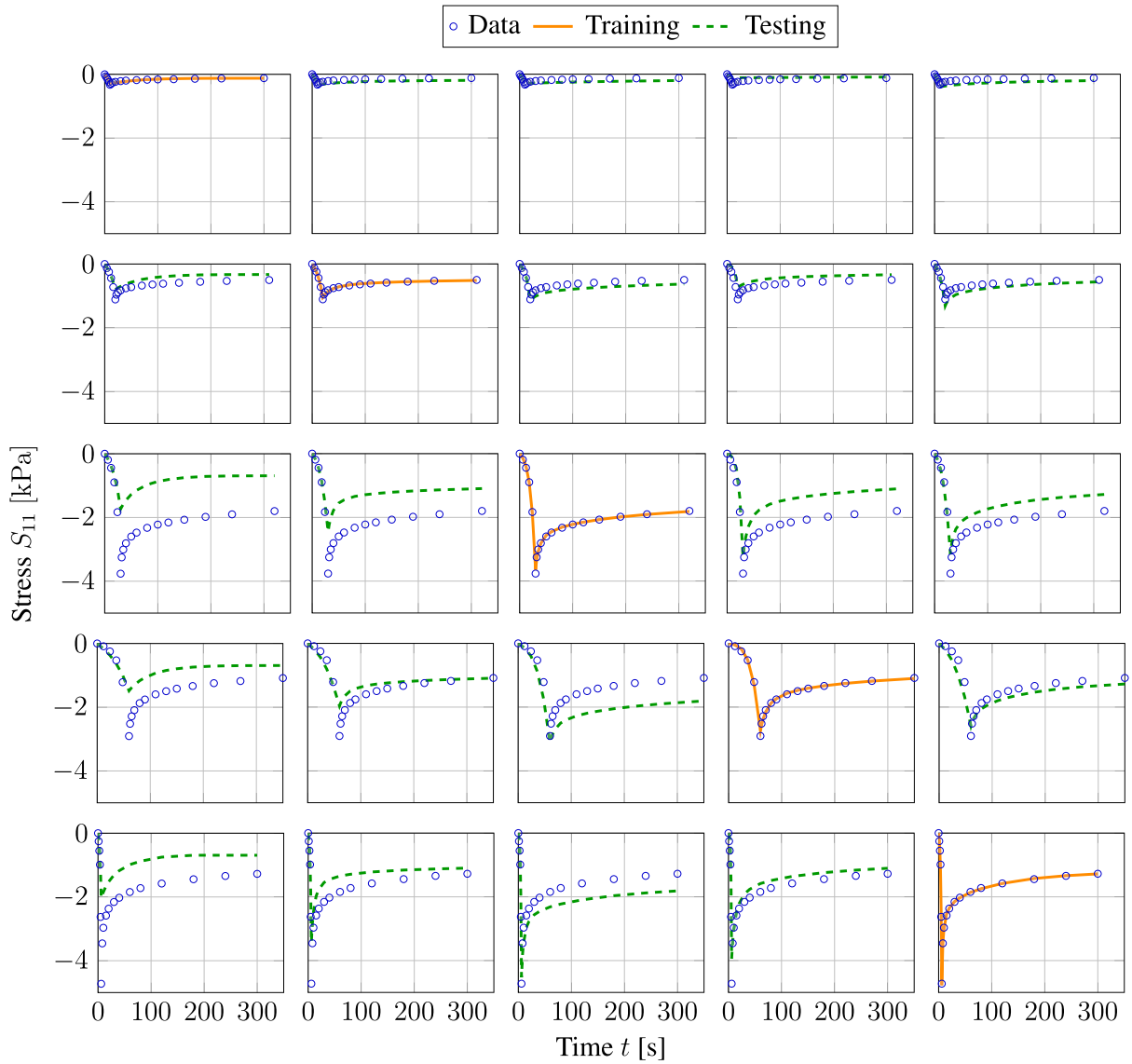


Fig. 12. Train on one, test on four. Experimental data and discovered model for passive skeletal muscle taken from Van Loocke et al. [123]. Each column represents a combination of training and testing, while each row represents one experimental setup. The experimental setup varies with respect to the maximum applied stretch $C_{11}^{\max} = \{0.81, 0.64, 0.49, 0.49, 0.49\}$ [–] and the rate of deformation $\dot{F}_{11} = \{0.01, 0.01, 0.01, 0.005, 0.05\}$ [s]. The columns vary with respect to the one experiment used for training.

Table 3

Normalized root mean squared error, ϵ , corresponding to the results shown in Fig. 12. Each column represents a combination of training and testing, while each row represents one experimental setup. The colored boxes indicate the experiment used for training, the remaining four tests are utilized for testing.

Stress S_{11} [kPa]	0.09	0.32	0.34	0.60	0.54
	0.28	0.04	0.22	0.35	0.26
	0.61	0.41	0.02	0.28	0.21
	0.48	0.25	0.41	0.04	0.17
	0.55	0.28	0.22	0.21	0.04
ϵ	Time t [s]				

Table 4

Coefficient of determination, R^2 , corresponding to the results shown in Fig. 12. Each column represents a combination of training and testing, while each row represents one experimental setup. The colored boxes indicate the experiment used for training, the remaining four tests are utilized for testing.

	0.96	0.47	0.40	0.00	0.00
Stress	0.60	0.99	0.76	0.39	0.66
S_{11} [kPa]	0.00	0.33	1.00	0.68	0.82
	0.29	0.79	0.49	0.99	0.91
	0.16	0.78	0.87	0.88	1.00
R^2	Time t [s]				

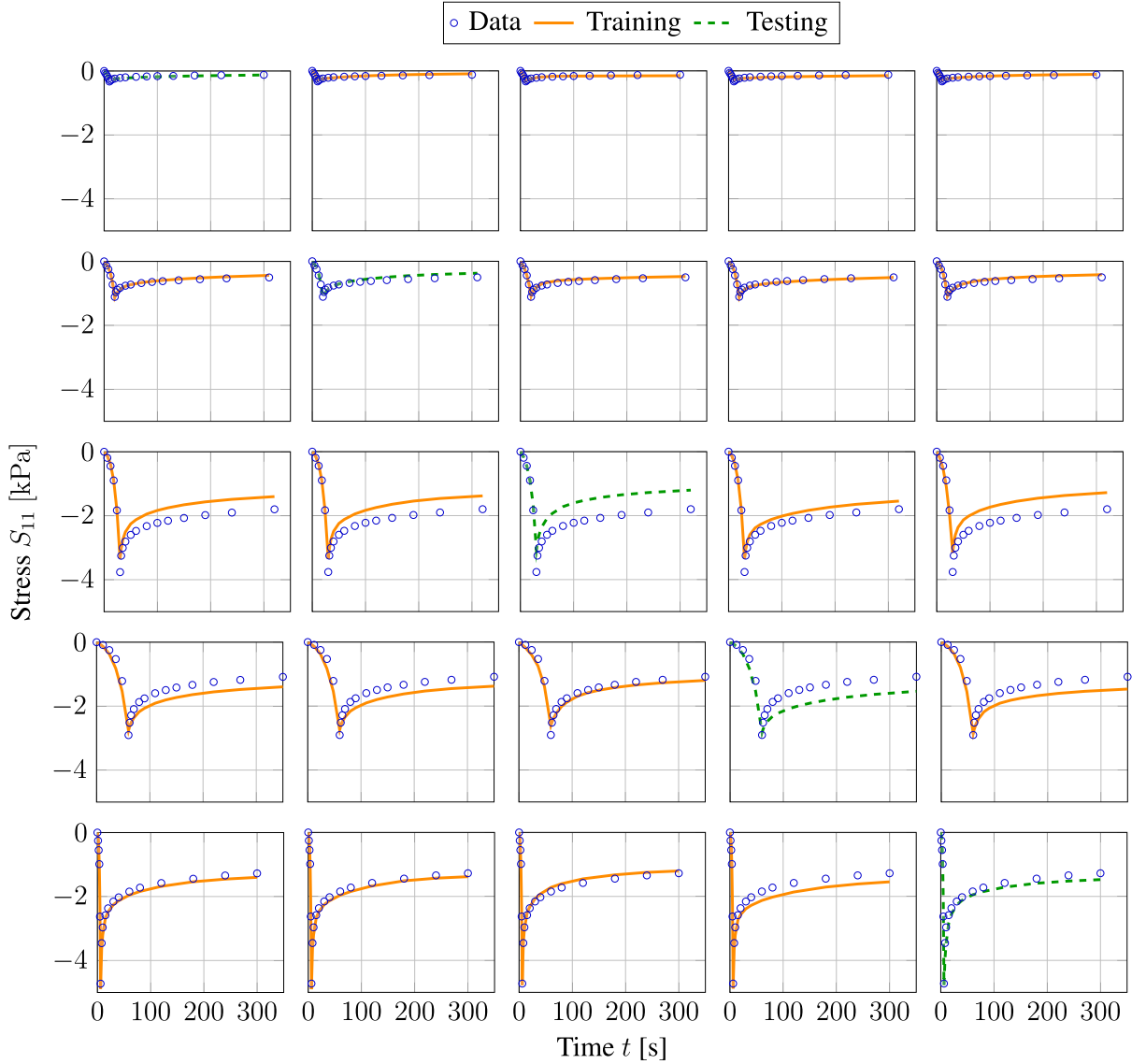


Fig. 13. Train on four, test on one. Experimental data and discovered model for passive skeletal muscle taken from Van Loocke et al. [123]. Each column represents a combination of training and testing, while each row represents one experimental setup. The experimental setup varies with respect to the maximum applied stretch $C_{11}^{\max} = \{0.81, 0.64, 0.49, 0.49, 0.49\}$ [-] and the rate of deformation $\dot{F}_{11} = \{0.01, 0.01, 0.01, 0.005, 0.05\}$ [s]. The columns vary with respect to the four experiments used for training.

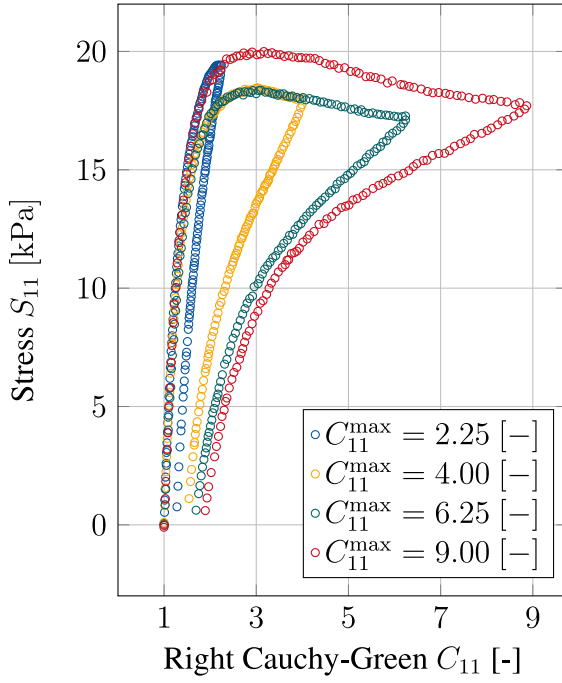
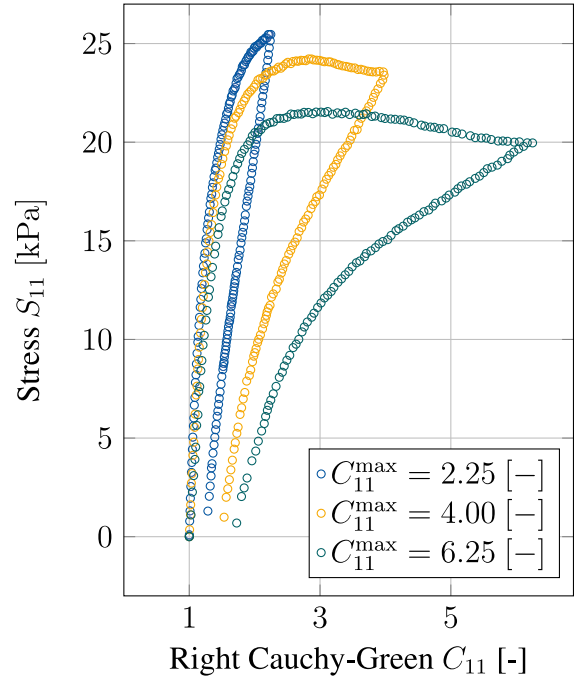
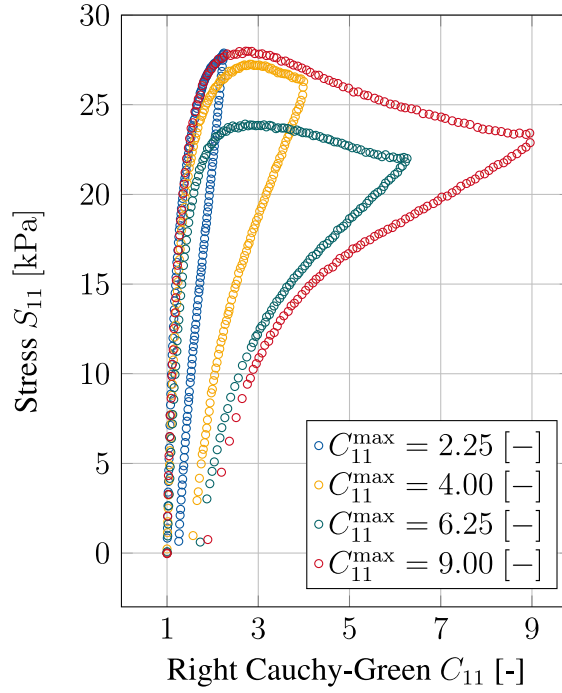
(a) $\dot{F}_{11} = 0.01$ [1/s](b) $\dot{F}_{11} = 0.03$ [1/s](c) $\dot{F}_{11} = 0.05$ [1/s]

Fig. 14. Experimental data of polymer VHB 4910 provided by [121]. The material is subjected to uniaxial loading–unloading at different maximum stretch levels, $C_{11}^{\max}$. For constant loading rates, $\dot{F}_{11}$, but different maximum stretch levels applied, the material's response during loading should be the same.

Table 5

Normalized root mean squared error, ϵ , corresponding to the results shown in Fig. 13. Each column represents a combination of training and testing, while each row represents one experimental setup. The colored boxes indicate the experiment used for testing, the remaining four tests are utilized for training.

Stress S_{11} [kPa]	0.15	0.14	0.07	0.17	0.13
	0.08	0.12	0.06	0.07	0.10
	0.16	0.16	0.25	0.10	0.14
	0.19	0.18	0.10	0.27	0.21
	0.08	0.08	0.05	0.12	0.12
ϵ	Time t [s]				

Table 6

Coefficient of determination, R^2 , corresponding to the results shown in Fig. 13. Each column represents a combination of training and testing, while each row represents one experimental setup. The colored boxes indicate the experiment used for testing, the remaining four tests are utilized for training.

Stress S_{11} [kPa]	0.89	0.90	0.97	0.85	0.91
	0.97	0.93	0.98	0.98	0.95
	0.90	0.90	0.76	0.96	0.93
	0.89	0.90	0.97	0.77	0.87
	0.98	0.98	0.99	0.96	0.96
R^2	Time t [s]				

6. Discussion and limitations

In this paper, we have designed a general inelastic constitutive artificial neural network that satisfies thermodynamics for all kinds of inelastic phenomena that can be accounted for by the multiplicative decomposition. To investigate our proposed approach, we limited our study to the effect of visco-elasticity. In addition, we focused on a thermodynamically sound and flexible theoretical description of the network rather than on aspects related to the machine learning algorithms themselves. Of course, this leads to some limitations of this work and leaves room for future investigations, some of which we would like to briefly touch upon.

iCANNs are able to discover a model for visco-elasticity. Our results show that our extension of CANNs by multiplicative decomposition and pseudo potentials is a suitable approach for viscoelastic materials. The sparse amount of data does not hinder the successful discovery. Since our pseudo potential, and especially the activation functions, are chosen in a general way that satisfies the mathematical requirements of thermodynamics, rather than inspired by existing models in the literature, it would be interesting to compare pseudo potentials used in constitutive modeling to describe certain materials with those discovered by iCANN. Furthermore, we have fixed the number of Maxwell elements to three. In future studies, it should be investigated how to design an iCANN that learns whether the number either can be decreased or must be increased.

Are our discovered iCANN models unique? Although we found several models that explain the experimental data well and also perform accurate in tests compared to standard neural networks, we did not investigate the uniqueness of our solution. Thus, it is possible that fewer terms in both the Helmholtz free energy and the pseudo potential are needed to explain the data with the same accuracy. To avoid possible overfitting, it should be investigated to what extent the optimizer's regularization can help us achieve this goal, as has already been done in the literature, e.g. by Wang et al. [3], St. Pierre et al. [118], and McCulloch et al. [124] for the Helmholtz free energy of CANNs.

Specialization to further inelastic phenomena. The network's architecture is formulated to accommodate a variety of inelastic phenomena. To investigate the performance of the iCANN, we focused ourselves to the case of visco-elasticity. Future works should study the ability to explain the data of, for instance, elasto-(visco)plasticity. Such a phenomenon includes some kind of yield criterion, which distinguishes the elastic from the inelastic regime. Having in mind that one might need multi-surface models, i.e., several parallel arrangements of iCANNs similar to using several Maxwell elements, we ask ourselves what is an appropriate implementation strategy. In order to keep the numerical effort during training low and the implementation as simple as possible, we would suggest to keep an explicit time integration, while the equality and/or inequality constraints are satisfied by Fischer–Burmeister approach (see e.g. Fischer [125]). For example, such an approach was successfully applied in constitutive modeling by Kiefer et al. [126], Brepols et al. [127].

Initial and induced anisotropy. We have not considered anisotropy. However, some materials exhibit a pronounced type of anisotropy, either intrinsic (e.g., fiber reinforced, collagen fibers, etc.) or induced (e.g., anisotropic damage). Since both phenomena can be constitutively modeled by structural tensors, future work should include this additional tensorial argument. It would be interesting to study whether the iCANN helps to identify the degree of anisotropy, e.g. the ratio between microvoids and microcracks. In addition, structural tensors combined with multiplicative decomposition require push-forward operations of the structural tensors

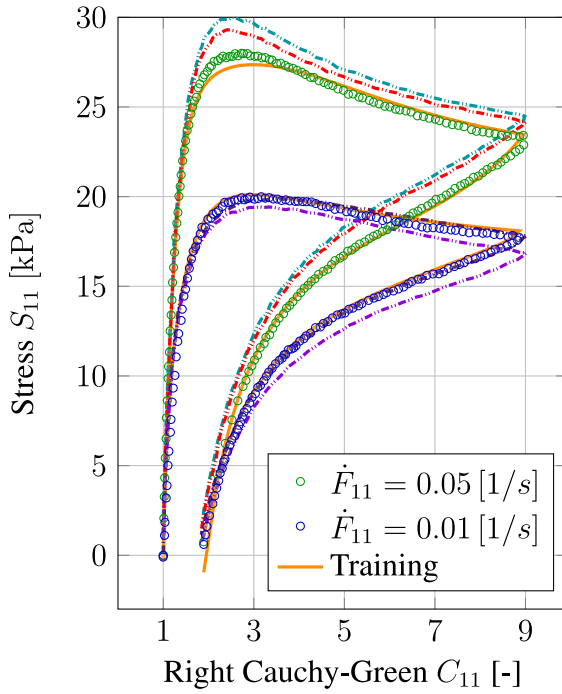
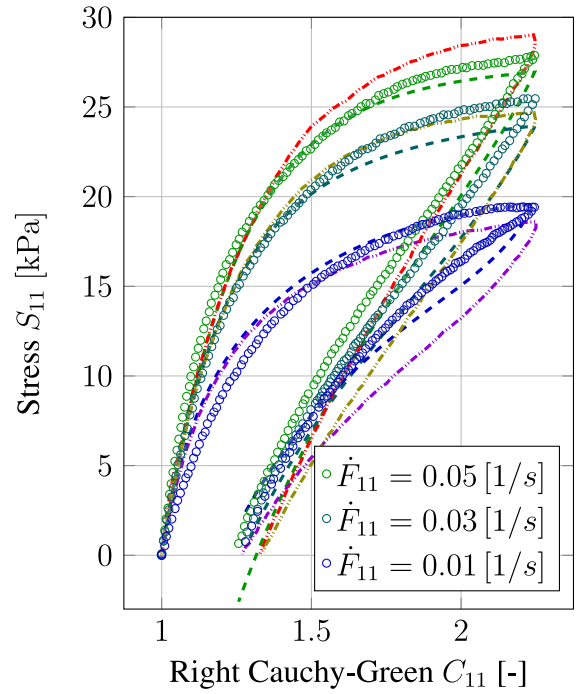
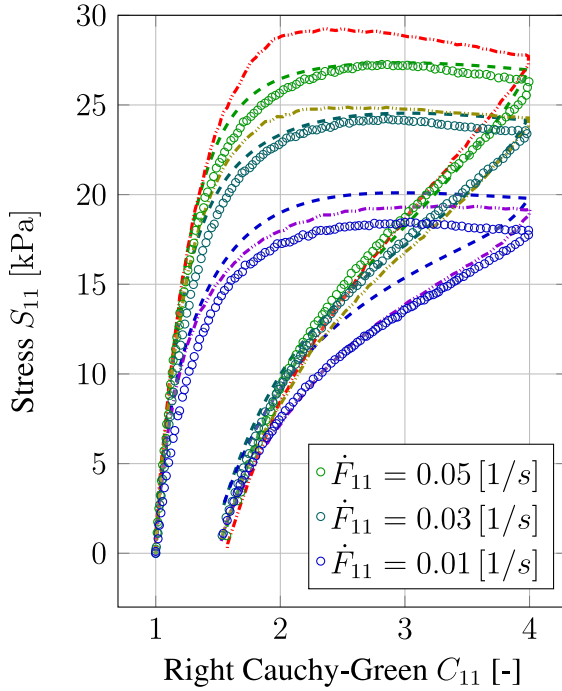
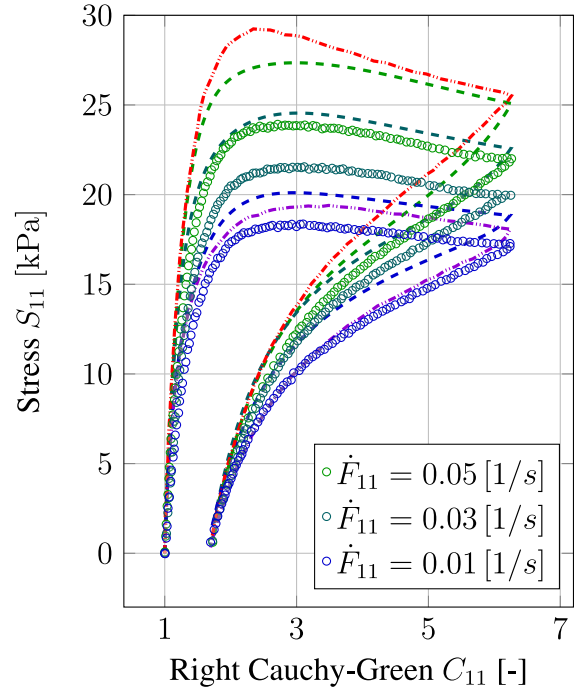
(a) Training with $C_{11}^{\max} = 9.0 [-]$ (b) Testing with $C_{11}^{\max} = 2.25 [-]$ (c) Testing with $C_{11}^{\max} = 4.0 [-]$ (d) Testing with $C_{11}^{\max} = 6.25 [-]$

Fig. 15. Comparison of the results using the ‘hand-crafted’ material model developed by [121] with the results of the iCANN. Hossain et al. [121] used five Maxwell elements (cyan, densely dashdotted, $\dot{F}_{11} = 0.05$, / burgundy, densely dashdotted, $\dot{F}_{11} = 0.01$, - - -) and four Maxwell elements (red, densely dashdotted, $\dot{F}_{11} = 0.05$, - - - / olive, densely dashdotted, $\dot{F}_{11} = 0.03$, - - - / purple, densely dashdotted, $\dot{F}_{11} = 0.01$, - - -). The results of the iCANN correspond to Fig. 11. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

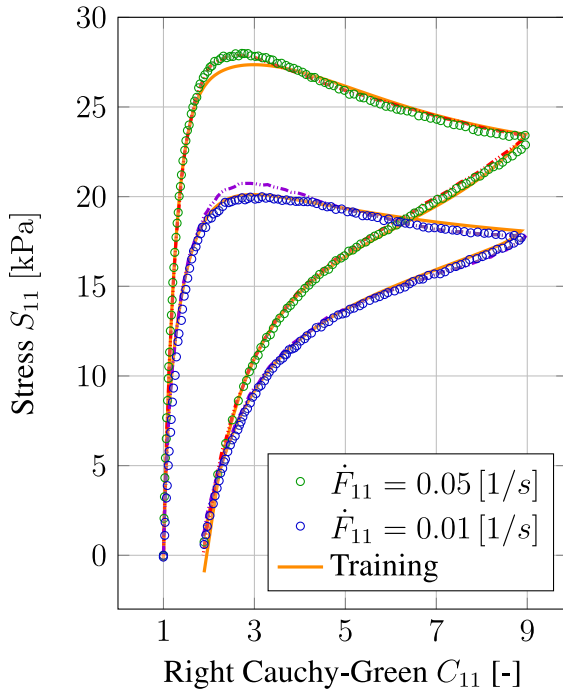
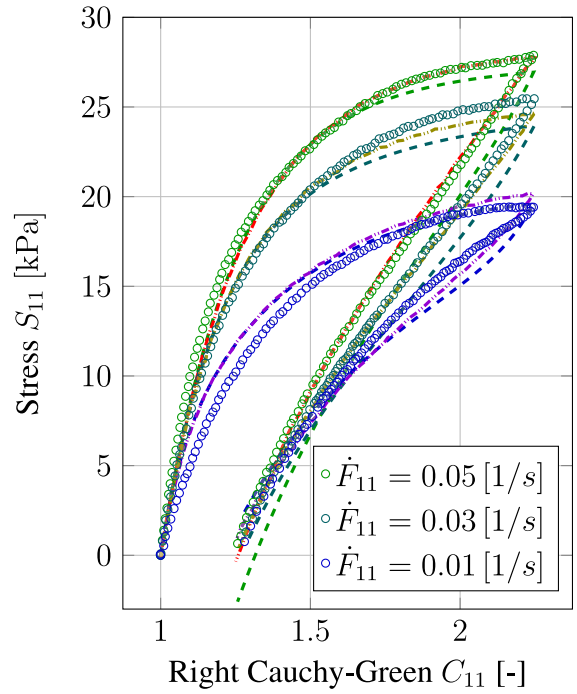
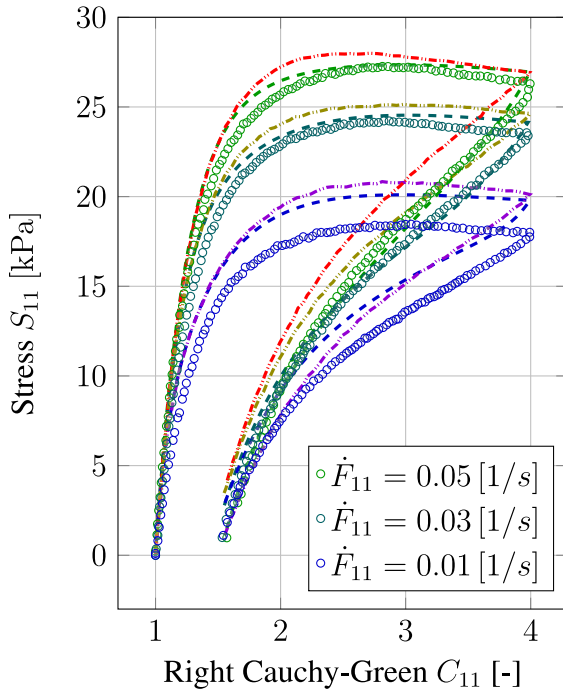
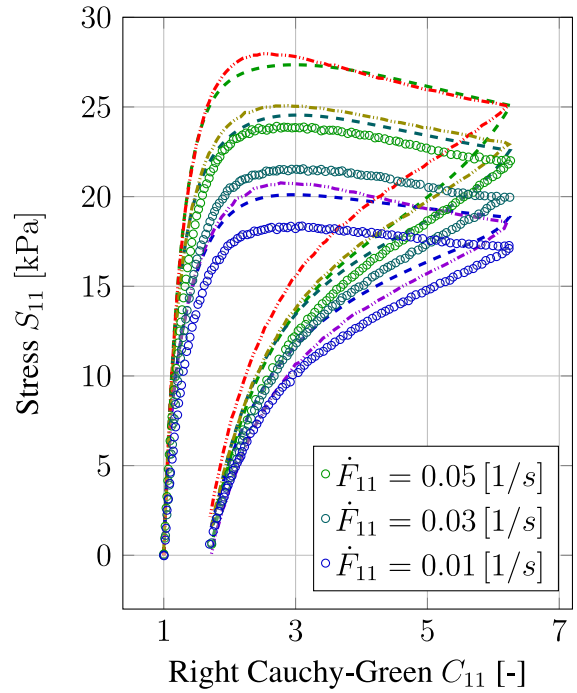
(a) Training with $C_{11}^{\max} = 9.0 [-]$ (b) Testing with $C_{11}^{\max} = 2.25 [-]$ (c) Testing with $C_{11}^{\max} = 4.0 [-]$ (d) Testing with $C_{11}^{\max} = 6.25 [-]$

Fig. 16. Comparison of the results using the visco-elastic CANN by [45] with the results of the iCANN. Abdolazizi et al. [45] ended up with two generalized Maxwell elements (red, densely dashdotted, $\dot{F}_{11} = 0.05$, / olive, densely dashdotted, $\dot{F}_{11} = 0.03$, / purple, densely dashdotted, $\dot{F}_{11} = 0.01$, / purple, densely dashdotted, $\dot{F}_{11} = 0.01$). Note that the approach used by [45] does not rely on the concept of pseudo potentials and is only applicable for Fig. 11. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

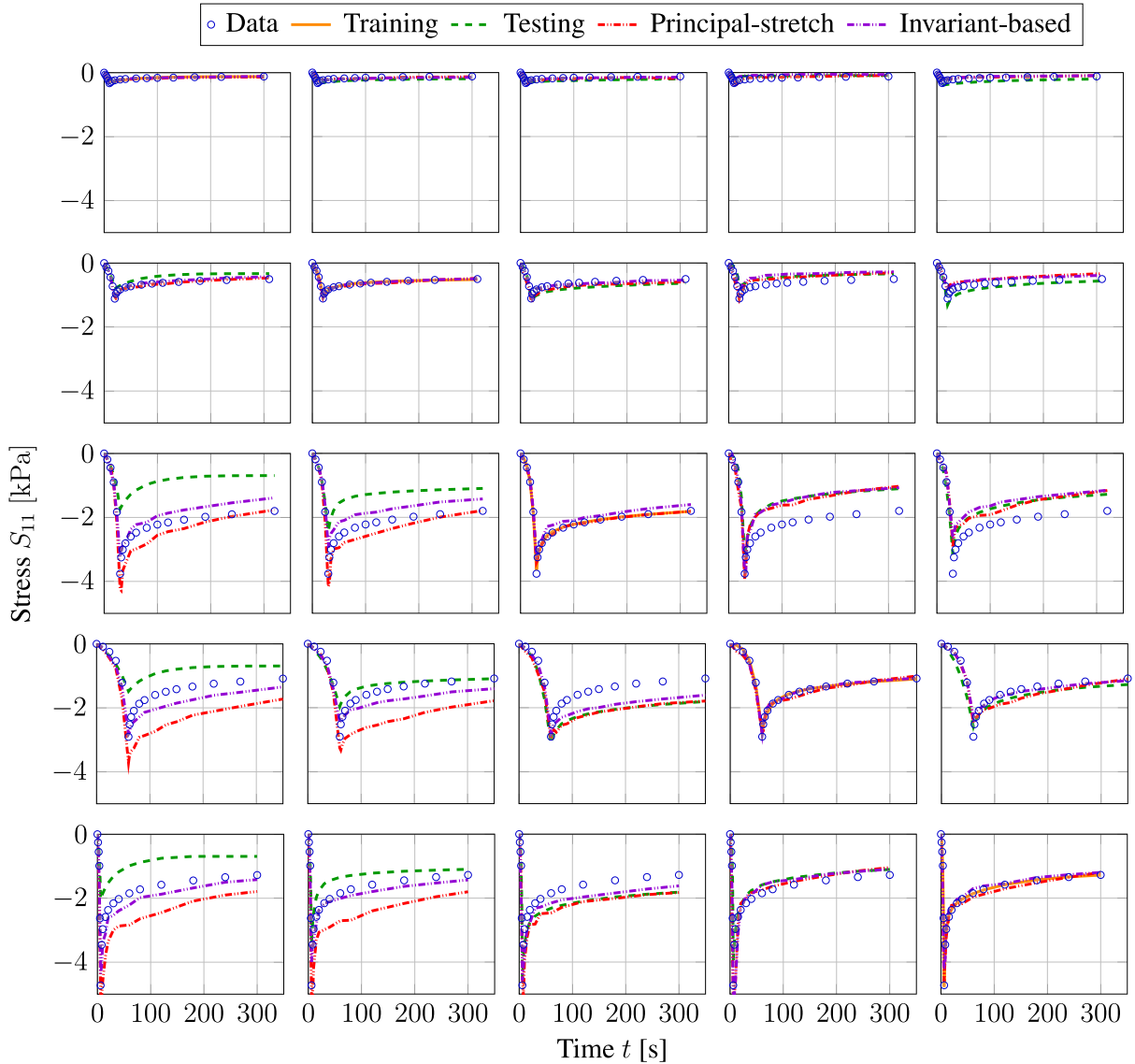


Fig. 17. Train on one, test on four. Comparison of the results using the visco-elastic CANN by Wang et al. [3] using a Prony-series approach with results of the iCANN. In Wang et al. [3], the formulation of the CANN relies either on the principal stretches or the invariants of $\mathbf{C}$. The results of the iCANN correspond to Fig. 12.

to the intermediate configuration. This mapping is not unique, see Holthusen et al. [98], Sansour et al. [128] for an overview of possible choices. This raises the question of whether iCANNs are capable of deciding which mapping is best for certain types of materials.

How to include multiphysics? Up to now, we only have taken mechanical influences into account. However, almost all engineering materials behave temperature dependent, e.g., by thermal expansion and/or thermal softening. To account for this phenomenon, we suggest to use a multiplicative decomposition of the deformation gradient into a thermal part and a mechanical part (see Stojanović et al. [129], Vujosevic and Lubarda [130]). In addition, the influence of electrochemical fields is of great importance in manufacturing (see Dorfmann and Ogden [131], Wulfinghoff and Dorn [132], and van der Velden et al. [133]). Moreover, the diffusion of hormones and nutrients plays an important role in the growth and remodeling of living organisms (see Manjunatha et al. [134]). Hence, it is highly desirable to extend the CANN and iCANN approaches to multiphysical problems. The question is how to find a generic formulation of our iCANN to account for the interactions between the different fields at the material point level.

How to set up experimental investigations? We have shown that we need less experimental data to discover a model explaining our data compared to the identification of material parameters used in constitutive modeling. The long-term perspective of iCANN is

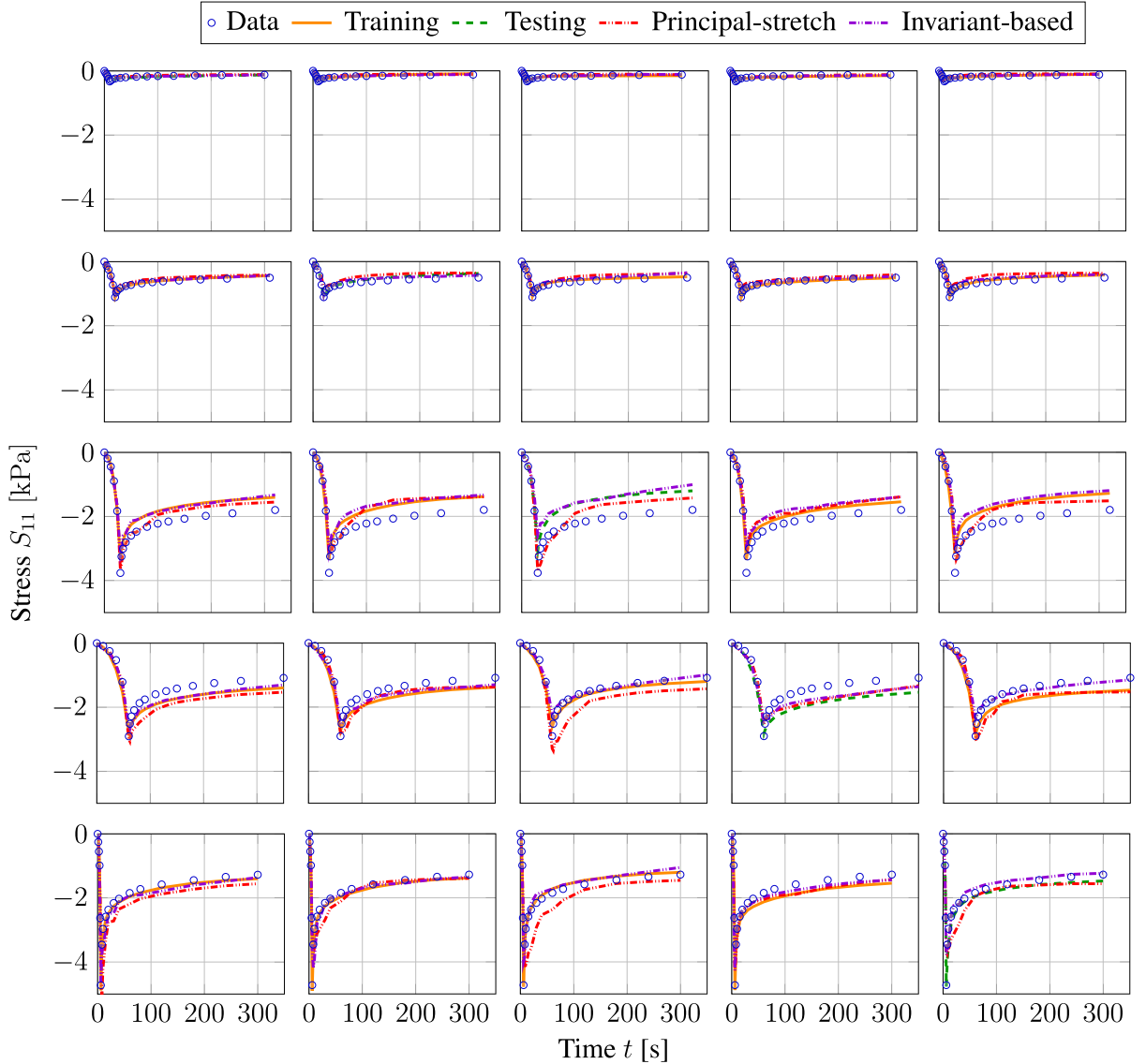


Fig. 18. Train on four, test on one. Comparison of the results using the visco-elastic CANN by Wang et al. [3] using a Prony-series approach with results of the iCANN. In Wang et al. [3], the formulation of the CANN relies either on the principal stretches or the invariants of $\mathbf{C}$. The results of the iCANN correspond to Fig. 13.

that by design several inelastic effects are included. This will allow the iCANN to automatically learn the inelastic effects hidden in the data and provide us with information about the micromechanical processes. Up to now, such an identification requires complex experimental investigations, e.g. several cyclic tests, relaxation tests etc. In view of our results, the question arises whether and how experiments have to be conducted in order to generate a maximum of information for the iCANN with as little effort as possible (see also the results and discussion in Linka et al. [47]).

7. Conclusion and outlook

The inelastic Constitutive Artificial Neural Networks (iCANN) architecture provides a new family of neural networks for the prediction of inelastic material behavior under finite deformations. This architecture represents a consistent extension of the original Constitutive Artificial Neural Network (CANN) design and inherits its advantages over classical neural network architectures being used for constitutive modeling. The overall design of the iCANN is based on sound kinematic and thermodynamic considerations. It combines well-established principles of continuum mechanics with the power of modern machine learning procedures to provide a neural network architecture that does not only offer thermodynamic consistency and satisfies principles of materials theory a priori,

Table 7

Discovered weights for the Helmholtz free energy feed-forward network. The weights belong to the results for the artificially generated data in Section 5.1. The order of the weights corresponds to the numerical implementation provided.

	ψ^{Neq}
$w_{1,1}^w$	0.16197191
$w_{1,3}^w$	0
$w_{1,2}^w$	0
$w_{1,4}^w$	0
$w_{3,1}^w$	4.4753417e-33
$w_{2,1}^w$	4.5402040e+00
$w_{2,5}^w$	9.2217451e-01
$w_{2,3}^w$	0
$w_{2,7}^w$	0
$w_{2,2}^w$	2.2400634e+00
$w_{2,6}^w$	0
$w_{2,4}^w$	1.5364066e-33
$w_{2,8}^w$	1.8214491e-33
$w_{3,2}^w$	0

Table 8

Discovered weights for the pseudo potential feed-forward network. The weights belong to the results for the artificially generated data in Section 5.1. The order of the weights corresponds to the numerical implementation provided.

	g
$w_{1,1}^g$	0
$w_{1,3}^g$	0
$w_{1,5}^g$	0
$w_{2,1}^g$	0
$w_{2,4}^g$	0
$w_{2,7}^g$	0.00107837
$w_{2,2}^g$	0
$w_{2,5}^g$	0
$w_{2,8}^g$	0

but also provides a flexible, modular and interpretable machine learning algorithm. Based on the multiplicative decomposition of the deformation gradient, this framework uses individual subnets to approximate both, the Helmholtz free energy as well as a pseudo potential function. This idea enables the network to flexibly adjust to the data provided, and therefore, represents a generic formulation to explain various kinds of inelastic phenomena. For illustrative purposes, we chose the example of finite viscoelasticity within this publication. We were able to demonstrate that the iCANN is capable to predict this kind of rate-dependent, inelastic material response properly, even though the amount of experimental data is sparse. Due to its general constitutive approach, the iCANN should be specialized for inequality-constrained material behavior (e.g., elasto-plasticity and damage), equality-constrained behavior (e.g., biological growth), multiphysics (e.g., thermoelasticity), and initial anisotropy (structural tensors) in future investigations to explore its ability to capture these phenomena as well (see Fig. 14).

Code availability

Our source code and examples are available at [135] (<https://doi.org/10.5281/zenodo.10066805>)

CRediT authorship contribution statement

Hagen Holthusen: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Lukas Lamm:** Data curation, Software, Writing – original draft, Writing – review & editing. **Tim Brepol:** Funding acquisition, Writing – original draft, Writing – review & editing. **Stefanie Reese:** Funding acquisition. **Ellen Kuhl:** Funding acquisition, Methodology, Writing – original draft, Writing – review & editing.

Table 9

Discovered weights for the Helmholtz free energy feed-forward network. The weights belong to the results for the VHB 4910 polymer in Section 5.2. The order of the weights corresponds to the numerical implementation provided.

	ψ^{Eq}	ψ_1^{Neq}	ψ_2^{Neq}	ψ_3^{Neq}
$w_{1,1}^{\psi}$	0.08989155	0.6867937	0.06954185	2.9327118
$w_{1,3}^{\psi}$	0.3871473	0.6118265	0.28366846	2.6145873
$w_{1,2}^{\psi}$	0	0.06561667	0	0.03622768
$w_{1,4}^{\psi}$	0.01116255	0.21422595	0.01034374	0.17840806
$w_{3,1}^{\psi}$	–	3.9390977e–33	–2.5810559e–33	–1.908444e–33
$w_{2,1}^{\psi}$	2.646479	1.0183624	2.3602471	1.0819899
$w_{2,5}^{\psi}$	3.1516218	0.70885456	2.8352766	1.2915097
$w_{2,3}^{\psi}$	0	0.35119042	0	0.21051936
$w_{2,7}^{\psi}$	0.6164215	0.35955966	0.1741371	0.2551231
$w_{2,2}^{\psi}$	1.4593441	1.894548	1.1949594	3.15422
$w_{2,6}^{\psi}$	2.9215317	1.354211	2.65872	2.8406792
$w_{2,4}^{\psi}$	0.0537339	0.1450241	0.08411078	0.09968195
$w_{2,8}^{\psi}$	0.24964914	0.32216933	0.2144338	0.23061126
$w_{3,2}^{\psi}$	–	0	0	0

Table 10

Discovered weights for the pseudo potential feed-forward network. The weights belong to the results for the VHB 4910 polymer in Section 5.2. The order of the weights corresponds to the numerical implementation provided.

	g_1	g_2	g_3
$w_{1,1}^g$	0	0	0
$w_{1,3}^g$	0	0	0
$\tilde{w}_{1,5}^g$	0	0	0
$w_{2,1}^g$	0	0	0
$w_{2,4}^g$	0	0	0
$\tilde{w}_{2,7}^g$	0.02321647	0.00174507	0.08019841
$w_{2,2}^g$	0	0	0
$w_{2,5}^g$	0	0	0
$w_{2,8}^g$	0	0	0

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.

Data availability

We shared the source code data via <https://doi.org/10.5281/zenodo.10066805>.

Acknowledgments

Hagen Holthusen would like to thank the Heinrich Hertz-Stiftung for a scholarship that enabled a research stay at Stanford University. The foundation had no role in study design, collection, analysis and interpretation of data, in writing the report, or in the decision to submit the article for publication. Tim Brepols and Stefanie Reese gratefully acknowledge financial support of the projects 453596084 and 453715964 by Deutsche Forschungsgemeinschaft. Ellen Kuhl acknowledges funding by NSF CMMI Award 2320933 Automated Model Discovery for Soft Matter.

Appendix A

A.1. Discovered weights for artificially generated data

See Tables 7 and 8.

A.2. Discovered weights for VHB 4910 polymer

See Tables 9 and 10.

Table 11

Discovered weights for the Helmholtz free energy feed-forward network. The weights belong to the best results for passive skeletal muscle in Section 5.3, i.e., the third column in Fig. 12. The order of the weights corresponds to the numerical implementation provided.

	ψ^{Eq}	ψ_1^{Neq}	ψ_2^{Neq}	ψ_3^{Neq}
$w_{1,1}^{\psi}$	0.10420806	0.01166395	0.00115344	0.01122695
$w_{1,3}^{\psi}$	0.1841228	0.01309146	0.00347792	0.01664283
$w_{1,2}^{\psi}$	0.02560516	0.00111786	0.00250525	0.00101571
$w_{1,4}^{\psi}$	0.21253704	0.0157883	0.6823113	0.01311576
$w_{3,1}^{\psi}$	—	5.497631e−34	−2.4178903e−14	−1.6557697e−05
$w_{2,1}^{\psi}$	0.06608981	0.04277665	0	4.2320956e−02
$w_{2,5}^{\psi}$	0.0637252	0.03449954	0	3.4831800e−02
$w_{2,3}^{\psi}$	0.06350973	0.01193809	0.02947876	1.5011105e−02
$w_{2,7}^{\psi}$	0.07190274	0.01611285	0.06952783	1.8784763e−02
$w_{2,2}^{\psi}$	0.09108008	0.01163169	0	1.1191454e−02
$w_{2,6}^{\psi}$	0.14969026	0.01267746	0	1.6345050e−02
$w_{2,4}^{\psi}$	0.02551797	0	0.0014553	7.9535537e−05
$w_{2,8}^{\psi}$	0.20174257	0	0.52773875	0
$w_{3,2}^{\psi}$	—	0	0	0

Table 12

Discovered weights for the pseudo potential feed-forward network. The weights belong to the best results for passive skeletal muscle in Section 5.3, i.e., the third column in Fig. 12. The order of the weights corresponds to the numerical implementation provided.

	g_1	g_2	g_3
$w_{1,1}^g$	0	0	0
$w_{1,3}^g$	0	0	0
$\tilde{w}_{1,5}^g$	0	0	0
$w_{2,1}^g$	4.5470802e−06	0	0
$w_{2,4}^g$	7.6929213e−13	0	0
$\tilde{w}_{2,7}^g$	1.7541333e−01	0.5692788	0.15919787
$w_{2,2}^g$	0	0	0
$w_{2,5}^g$	0	0	0
$w_{2,8}^g$	0	0	0

Table 13

Discovered weights for the Helmholtz free energy feed-forward network. The weights belong to the best results for passive skeletal muscle in Section 5.3, i.e., the second column in Fig. 13. The order of the weights corresponds to the numerical implementation provided.

	ψ^{Eq}	ψ_1^{Neq}	ψ_2^{Neq}	ψ_3^{Neq}
$w_{1,1}^{\psi}$	0.00138299	0	0	6.044407e−06
$w_{1,3}^{\psi}$	0.03045796	0	0	0
$w_{1,2}^{\psi}$	0.07063455	0	0	0
$w_{1,4}^{\psi}$	0.09885824	7.708453e−10	0.00558791	0
$w_{3,1}^{\psi}$	—	3.6285916e−33	3.733226e−33	3.766066e−33
$w_{2,1}^{\psi}$	0.02312493	3.51029038e−02	0.02298807	8.7625727e−02
$w_{2,5}^{\psi}$	0.04258661	1.12825185e−02	0.01470857	5.3523790e−02
$w_{2,3}^{\psi}$	0.11215075	2.31756195e−01	0.02367271	8.5964095e−04
$w_{2,7}^{\psi}$	0.12233058	5.45360267e−01	0.04010505	8.0012449e−04
$w_{2,2}^{\psi}$	0.00133567	0	0	5.9973318e−06
$w_{2,6}^{\psi}$	0.02934347	0	0	0
$w_{2,4}^{\psi}$	0.0702935	0	0	0
$w_{2,8}^{\psi}$	0.0983757	7.71025022e−10	0.00552198	0
$w_{3,2}^{\psi}$	—	0	0	0

A.3. Discovered weights for passive skeletal muscle

See Tables 11–14.

A.4. Experimental data of polymer VHB 4910

See Fig. 14.

Table 14

Discovered weights for the pseudo potential feed-forward network. The weights belong to the best results for passive skeletal muscle in Section 5.3, i.e., the second column in Fig. 13. The order of the weights corresponds to the numerical implementation provided.

	g_1	g_2	g_3
$w_{1,1}^g$	0	0	0
$w_{1,3}^g$	0	0	0
$\tilde{w}_{1,5}^g$	0	0	0
$w_{2,1}^g$	0	0	0
$w_{2,4}^g$	4.6914302e-11	0	0
$\tilde{w}_{2,7}^g$	7.3394459e-01	0.24998124	0.29360896
$w_{2,2}^g$	0	0	0
$w_{2,5}^g$	0	0	0
$w_{2,8}^g$	0	0	0

A.5. Comparisons for polymer data

See Figs. 15 and 16.

A.6. Comparisons for muscle data

See Figs. 17 and 18.

References

- [1] K. Linka, E. Kuhl, A new family of constitutive artificial neural networks towards automated model discovery, *Comput. Methods Appl. Mech. Engrg.* 403 (2023) 115731.
- [2] L.R.G. Treloar, Stress-strain data for vulcanised rubber under various types of deformation, *Trans. Faraday Soc.* 40 (1944) 59–70.
- [3] L.M. Wang, K. Linka, E. Kuhl, Automated model discovery for muscle using constitutive recurrent neural networks, *J. Mech. Behav. Biomed. Mater.* 145 (2023) 106021.
- [4] M. Raissi, P. Perdikaris, G. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *J. Comput. Phys.* 378 (2019) 686–707.
- [5] C. Rao, H. Sun, Y. Liu, Physics-informed deep learning for incompressible laminar flows, *Theor. Appl. Mech. Lett.* 10 (3) (2020) 207–212.
- [6] E. Zhang, M. Yin, G.E. Karniadakis, Physics-informed neural networks for nonhomogeneous material identification in elasticity imaging, *CoRR abs/2009.04525*, 2020.
- [7] S. Amini Niaki, E. Haghighat, T. Campbell, A. Poursartip, R. Vaziri, Physics-informed neural network for modelling the thermochemical curing process of composite-tool systems during manufacture, *Comput. Methods Appl. Mech. Engrg.* 384 (2021) 113959.
- [8] S. Cai, Z. Wang, S. Wang, P. Perdikaris, G.E. Karniadakis, Physics-informed neural networks for heat transfer problems, *J. Heat Transfer* 143 (6) (2021) 060801.
- [9] E. Haghighat, M. Raissi, A. Moure, H. Gomez, R. Juanes, A physics-informed deep learning framework for inversion and surrogate modeling in solid mechanics, *Comput. Methods Appl. Mech. Engrg.* 379 (2021) 113741.
- [10] M. Vahab, E. Haghighat, M. Khaleghi, N. Khalili, A physics informed neural network approach to solution and identification of biharmonic equations of elasticity, *CoRR abs/2108.07243*, 2021.
- [11] A. Henkes, H. Wessels, R. Mahnken, Physics informed neural networks for continuum micromechanics, *Comput. Methods Appl. Mech. Engrg.* 393 (2022) 114790.
- [12] S. Rezaei, A. Harandi, A. Moeineddin, B.-X. Xu, S. Reese, A mixed formulation for physics-informed neural networks as a potential solver for engineering problems in heterogeneous domains: Comparison with finite element method, *Comput. Methods Appl. Mech. Engrg.* 401 (2022) 115616.
- [13] A. Harandi, A. Moeineddin, M. Kaliske, S. Reese, S. Rezaei, Mixed formulation of physics-informed neural networks for thermo-mechanically coupled systems and heterogeneous domains, *Internat. J. Numer. Methods Engrg.* 125 (4) (2024) e7388.
- [14] S. Niu, E. Zhang, Y. Bazilevs, V. Srivastava, Modeling finite-strain plasticity using physics-informed neural network and assessment of the network performance, *J. Mech. Phys. Solids* 172 (2023) 105177.
- [15] E. Haghighat, S. Abouali, R. Vaziri, Constitutive model characterization and discovery using physics-informed deep learning, *Eng. Appl. Artif. Intell.* 120 (2023) 105828.
- [16] M. Eghbalian, M. Pouragha, R. Wan, A physics-informed deep neural network for surrogate modeling in classical elasto-plasticity, *Comput. Geotech.* 159 (2023) 105472.
- [17] P. Weber, W. Wagner, S. Freitag, Physically enhanced training for modeling rate-independent plasticity with feedforward neural networks, *Comput. Mech.* 72 (4) (2023) 827–857.
- [18] T. Kirchdoerfer, M. Ortiz, Data-driven computational mechanics, *Comput. Methods Appl. Mech. Engrg.* 304 (2016) 81–101.
- [19] R. Ibañez, D. Borzacchiello, J.V. Aguado, E. Abisset-Chavanne, E. Cueto, P. Ladeveze, F. Chinesta, Data-driven non-linear elasticity: constitutive manifold construction and problem discretization, *Comput. Mech.* 60 (2017) 813–826.
- [20] R. Eggersmann, T. Kirchdoerfer, S. Reese, L. Stainier, M. Ortiz, Model-free data-driven inelasticity, *Comput. Methods Appl. Mech. Engrg.* 350 (2019) 81–99.
- [21] R. Ibanez, E. Abisset-Chavanne, J.V. Aguado, D. Gonzalez, E. Cueto, F. Chinesta, A manifold learning approach to data-driven computational elasticity and inelasticity, *Arch. Comput. Methods Eng.* 25 (2018) 47–57.
- [22] R. Eggersmann, L. Stainier, M. Ortiz, S. Reese, Model-free data-driven computational mechanics enhanced by tensor voting, *Comput. Methods Appl. Mech. Engrg.* 373 (2021) 113499.
- [23] K. Cifci, K. Hackl, Model-free data-driven simulation of inelastic materials using structured data sets, tangent space information and transition rules, *Comput. Mech.* 70 (2) (2022) 425–435.

- [24] S. Zschocke, F. Leichsenring, W. Graf, M. Kaliske, A concept for data-driven computational mechanics in the presence of polymorphic uncertain properties, *Eng. Struct.* 267 (2022) 114672.
- [25] Z. Kuang, X. Bai, Q. Huang, J. Yang, W. Huang, S. Belouettar, H. Hu, Data-driven computational framework for snap-through problems, *Int. J. Solids Struct.* 269 (2023) 112226.
- [26] F. Masi, I. Stefanou, P. Vannucci, V. Maffi-Berthier, Thermodynamics-based artificial neural networks for constitutive modeling, *J. Mech. Phys. Solids* 147 (2021) 104277.
- [27] 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.
- [28] F. Masi, I. Stefanou, Multiscale modeling of inelastic materials with thermodynamics-based artificial neural networks (TANN), *Comput. Methods Appl. Mech. Engrg.* 398 (2022) 115190.
- [29] F. Masi, I. Stefanou, Evolution TANN and the identification of internal variables and evolution equations in solid mechanics, *J. Mech. Phys. Solids* 174 (2023) 105245.
- [30] F. As'ad, C. Farhat, A mechanics-informed deep learning framework for data-driven nonlinear viscoelasticity, *Comput. Methods Appl. Mech. Engrg.* 417 (2023) 116463.
- [31] R. Abdusalamov, M. Hillgärtner, M. Itskov, Automatic generation of interpretable hyperelastic material models by symbolic regression, *Int. J. Num. Methods Eng.* 124 (9) (2023) 2093–2104, First published: 08 January 2023.
- [32] R. Abdusalamov, J. Kaplunov, M. Itskov, Discovering asymptotic expansions using symbolic regression, 2023.
- [33] R. Abdusalamov, J. Weise, M. Itskov, Rediscovering the mullins effect with deep symbolic regression, 2024.
- [34] J. Hou, X. Chen, T. Wu, E. Kuhl, X. Wang, Automated data-driven discovery of material models based on symbolic regression: A case study on human brain cortex, 2024.
- [35] K.A. Kalina, L. Linden, J. Brummund, P. Metsch, M. Kästner, Automated constitutive modeling of isotropic hyperelasticity based on artificial neural networks, *Comput. Mech.* (2022) 1–20.
- [36] M. Flaschel, S. Kumar, L. De Lorenzis, Unsupervised discovery of interpretable hyperelastic constitutive laws, *Comput. Methods Appl. Mech. Engrg.* 381 (2021) 113852.
- [37] M. Flaschel, S. Kumar, L. De Lorenzis, Discovering plasticity models without stress data, *npj Comput. Mater.* 8 (1) (2022) 91.
- [38] P. Steinmann, M. Hossain, G. Possart, Hyperelastic models for rubber-like materials: consistent tangent operators and suitability for treloar's data, *Arch. Appl. Mech.* 82 (9) (2012) 1183–1217.
- [39] M. Flaschel, S. Kumar, L. De Lorenzis, Automated discovery of generalized standard material models with EUCLID, *Comput. Methods Appl. Mech. Engrg.* 405 (2023) 115867.
- [40] G. Martius, C.H. Lampert, Extrapolation and learning equations, 2016.
- [41] S. Sahoo, C. Lampert, G. Martius, Learning equations for extrapolation and control, in: J. Dy, A. Krause (Eds.), *Proceedings of the 35th International Conference on Machine Learning*, in: *Proceedings of Machine Learning Research*, vol. 80, PMLR, 2018, pp. 4442–4450.
- [42] Z. Long, Y. Lu, B. Dong, PDE-net 2.0: Learning PDEs from data with a numeric-symbolic hybrid deep network, *J. Comput. Phys.* 399 (2019) 108925.
- [43] R. Rai, C.K. Sahu, Driven by data or derived through physics? A review of hybrid physics guided machine learning techniques with cyber-physical system (CPS) focus, *IEEE Access* 8 (2020) 71050–71073.
- [44] 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.
- [45] K.P. Abdolazizi, K. Linka, C.J. Cyron, Viscoelastic constitutive artificial neural networks (vCANNs) – a framework for data-driven anisotropic nonlinear finite viscoelasticity, 2023.
- [46] K. Linka, S.R. St. Pierre, E. Kuhl, Automated model discovery for human brain using constitutive artificial neural networks, *Acta Biomater.* 160 (2023) 134–151.
- [47] K. Linka, A. Buganza Tepole, G.A. Holzapfel, E. Kuhl, Automated model discovery for skin: Discovering the best model, data, and experiment, *Comput. Methods Appl. Mech. Engrg.* 410 (2023) 116007.
- [48] J. Rice, Inelastic constitutive relations for solids: An internal-variable theory and its application to metal plasticity, *J. Mech. Phys. Solids* 19 (6) (1971) 433–455.
- [49] J. Kestin, J.R. Rice, Paradoxes in the application of thermodynamics to strained solids, 1969, pp. 275–298.
- [50] E. Kröner, Allgemeine kontinuumstheorie der versetzungen und eigenspannungen, *Arch. Ration. Mech. Anal.* 4 (1) (1959) 273.
- [51] E.H. Lee, Elastic-Plastic Deformation at Finite Strains, *J. Appl. Mech.* 36 (1) (1969) 1–6.
- [52] J. Mandel, Equations constitutives et directeurs dans les milieux plastiques et viscoplastiques, *Int. J. Solids Struct.* 9 (6) (1973) 725–740.
- [53] F. Sidoroff, UN modele viscoelastique non lineaire avec configuration intermediaire, *J. Mec.; FR.; DA* 13 (4) (1974) 679–713, *ABS. ANGL.; BIBL.18 REF..*
- [54] E.K. Rodriguez, A. Hoger, A.D. McCulloch, Stress-dependent finite growth in soft elastic tissues, *J. Biomech.* 27 (4) (1994) 455–467.
- [55] A. Lion, On the large deformation behaviour of reinforced rubber at different temperatures, *J. Mech. Phys. Solids* 45 (11) (1997) 1805–1834.
- [56] P. Haupt, K. Sedlan, Viscoplasticity of elastomeric materials: experimental facts and constitutive modelling, *Arch. Appl. Mech.* 71 (2) (2001) 89–109.
- [57] A. Menzel, M. Ekh, K. Runesson, P. Steinmann, A framework for multiplicative elastoplasticity with kinematic hardening coupled to anisotropic damage, *Int. J. Plast.* 21 (3) (2005) 397–434.
- [58] D. Bammann, K. Solanki, On kinematic, thermodynamic, and kinetic coupling of a damage theory for polycrystalline material, *Int. J. Plast.* 26 (6) (2010) 775–793.
- [59] M. Latorre, F.J. Montáns, Anisotropic finite strain viscoelasticity based on the sidoroff multiplicative decomposition and logarithmic strains, *Comput. Mech.* 56 (3) (2015) 503–531.
- [60] C. Reina, S. Conti, Incompressible inelasticity as an essential ingredient for the validity of the kinematic decomposition $f=f_{el}$, *J. Mech. Phys. Solids* 107 (2017) 322–342.
- [61] A. Lion, A constitutive model for carbon black filled rubber: Experimental investigations and mathematical representation, *Contin. Mech. Thermodyn.* 8 (3) (1996) 153–169.
- [62] A. Lion, A physically based method to represent the thermo-mechanical behaviour of elastomers, *Acta Mech.* 123 (1) (1997) 1–25.
- [63] S. Reese, S. Govindjee, A theory of finite viscoelasticity and numerical aspects, *Int. J. Solids Struct.* 35 (26) (1998) 3455–3482.
- [64] J. Besson, Damage of ductile materials deforming under multiple plastic or viscoplastic mechanisms, *Int. J. Plast.* 25 (11) (2009) 2204–2221.
- [65] C. Miehe, A constitutive frame of elastoplasticity at large strains based on the notion of a plastic metric, *Int. J. Solids Struct.* 35 (30) (1998) 3859–3897.
- [66] H. Badreddine, K. Saanouni, A. Dogui, On non-associative anisotropic finite plasticity fully coupled with isotropic ductile damage for metal forming, *Int. J. Plast.* 26 (11) (2010) 1541–1575.
- [67] R. Hill, E. Orowan, A theory of the yielding and plastic flow of anisotropic metals, *Proc. R. Soc. London. Series A* 193 (1033) (1948) 281–297.
- [68] A.A. Benzerga, J. Besson, Plastic potentials for anisotropic porous solids, *Eur. J. Mech. A Solids* 20 (3) (2001) 397–434.
- [69] G.A. Maugin, W. Muschik, Thermodynamics with internal variables. Part I. General concepts, *J. Non-Equilib. Thermodyn.* 19 (3) (1994).
- [70] P. Flory, Thermodynamic relations for high elastic materials, *Trans. Faraday Soc.* 57 (1961) 829–838.
- [71] P. Germain, Q.S. Nguyen, P. Suquet, Continuum thermodynamics, *J. Appl. Mech.* 50 (4b) (1983) 1010–1020.
- [72] B. Halphen, Q.S. Nguyen, Sur les matériaux standard généralisés, *J. Mécanique* 14 (1975) 39–63.

- [73] J. Marsden, T.J.R. Hughes, Mathematical foundations of elasticity, Dover Publications, Inc., ISBN: 0-486-67865-2, 1983.
- [74] G.A. Holzapfel, Nonlinear solid mechanics, Repr., Wiley, Chichester ; Weinheim [u.a.], 2010, XIV, 455 S., 0-471-82304-X and 0-471-82319-8 and 978-0-471-82304-9 and 978-0-471-82319-3.
- [75] A. Bertram, S. Forest (Eds.), Mechanics of strain gradient materials, first ed., in: CISM International Centre for Mechanical Sciences, Springer Nature, Cham, Switzerland, 2020.
- [76] R. Ogden, Non-linear elastic deformations, in: Ellis Horwood series in mathematics and its applications, E. Horwood, ISBN: 9780853122739, 1984.
- [77] P. Chadwick, Continuum mechanics, second ed., in: Dover Books on Physics, Dover Publications, Mineola, NY, 1998.
- [78] J.C. Simo, T.J. Hughes, Computational inelasticity, vol. 7, Springer Science & Business Media, 2006.
- [79] J. Lubliner, A model of rubber viscoelasticity, Mech. Res. Commun. 12 (2) (1985) 93–99.
- [80] C. Eckart, The thermodynamics of irreversible processes. IV. The theory of elasticity and anelasticity, Phys. Rev. 73 (1948) 373–382.
- [81] H. Schütte, O. Bruhns, On a geometrically nonlinear damage model based on a multiplicative decomposition of the deformation gradient and the propagation of microcracks, J. Mech. Phys. Solids 50 (4) (2002) 827–853.
- [82] C. Dorn, S. Wulfinghoff, A gradient-extended large-strain anisotropic damage model with crack orientation director, Comput. Methods Appl. Mech. Engrg. 387 (2021) 114123.
- [83] D. Christ, S. Reese, A finite element model for shape memory alloys considering thermomechanical couplings at large strains, Int. J. Solids Struct. 46 (20) (2009) 3694–3709.
- [84] V. Evangelista, S. Marfia, E. Sacco, A 3D SMA constitutive model in the framework of finite strain, Internat. J. Numer. Methods Engrg. 81 (6) (2010) 761–785.
- [85] A. Sengupta, P. Papadopoulos, A note on the use of multiplicative decomposition for displacive phase transformations, Internat. J. Engrg. Sci. 49 (8) (2011) 801–805.
- [86] V. Lubarda, A. Hoger, On the mechanics of solids with a growing mass, Int. J. Solids Struct. 39 (18) (2002) 4627–4664.
- [87] J. Casey, A convenient form of the multiplicative decomposition of the deformation gradient, Math. Mech. Solids 22 (3) (2017) 528–537.
- [88] A. Green, P. Naghdi, Some remarks on elastic-plastic deformation at finite strain, Internat. J. Engrg. Sci. 9 (12) (1971) 1219–1229.
- [89] B.D. Coleman, W. Noll, Foundations of linear viscoelasticity, Rev. Mod. Phys. 33 (1961) 239–249.
- [90] B.D. Coleman, W. Noll, The thermodynamics of elastic materials with heat conduction and viscosity, Arch. Ration. Mech. Anal. 13 (1) (1963) 167–178.
- [91] B.D. Coleman, M.E. Gurtin, Thermodynamics with internal state variables, J. Chem. Phys. 47 (2) (1967) 597–613.
- [92] B. Svendsen, On the modelling of anisotropic elastic and inelastic material behaviour at large deformation, Int. J. Solids Struct. 38 (52) (2001) 9579–9599.
- [93] L. Lamm, H. Holthusen, T. Brepols, S. Jockenhövel, S. Reese, A macroscopic approach for stress-driven anisotropic growth in bioengineered soft tissues, Biomech. Model. Mechanobiol. 21 (2) (2022) 627–645.
- [94] N. Hansen, H. Schreyer, A thermodynamically consistent framework for theories of elastoplasticity coupled with damage, Int. J. Solids Struct. 31 (3) (1994) 359–389.
- [95] A. Ricker, M. Gierig, P. Wriggers, Multiplicative, non-Newtonian viscoelasticity models for rubber materials and brain tissues: Numerical treatment and comparative studies, Arch. Comput. Methods Eng. (2023).
- [96] J. Simo, C. Miehe, Associative coupled thermoplasticity at finite strains: Formulation, numerical analysis and implementation, Comput. Methods Appl. Mech. Engrg. 98 (1) (1992) 41–104.
- [97] W. Dettmer, S. Reese, On the theoretical and numerical modelling of Armstrong–Frederick kinematic hardening in the finite strain regime, Comput. Methods Appl. Mech. Engrg. 193 (1) (2004) 87–116.
- [98] 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.
- [99] A. Spencer, Part III. Theory of invariants, Contin. Phys. 1 (1971) 239–353.
- [100] J.-P. Boehler, A simple derivation of representations for non-polynomial constitutive equations in some cases of anisotropy, ZAMM - J. Appl. Math. Mech. (Zeitschrift für Angewandte Mathematik und Mechanik) 59 (4) (1979) 157–167.
- [101] Q.-S. Zheng, Theory of representations for tensor functions— a unified invariant approach to constitutive equations, Appl. Mech. Rev. 47 (11) (1994) 545–587.
- [102] R.S. Rivlin, D.W. Saunders, E.N.D.C. Andrade, Large elastic deformations of isotropic materials VII. Experiments on the deformation of rubber, Philos. Trans. R. Soc. London. Series A 243 (865) (1951) 251–288.
- [103] M. Mooney, A theory of large elastic deformation, J. Appl. Phys. 11 (9) (1940) 582–592.
- [104] R.S. Rivlin, E.K. Rideal, Large elastic deformations of isotropic materials iv. further developments of the general theory, Philos. Trans. R. Soc. London. Series A 241 (835) (1948) 379–397.
- [105] A.F. Bower, Applied mechanics of solids, CRC Press, Boca Raton, FL, 2009.
- [106] C. Sansour, On the physical assumptions underlying the volumetric-isochoric split and the case of anisotropy, Eur. J. Mech. A Solids 27 (1) (2008) 28–39.
- [107] M. Pellicciari, S. Sirotti, A.M. Tarantino, A strain energy function for large deformations of compressible elastomers, J. Mech. Phys. Solids 176 (2023) 105308.
- [108] S. Hartmann, P. Neff, Polyconvexity of generalized polynomial-type hyperelastic strain energy functions for near-incompressibility, Int. J. Solids Struct. 40 (11) (2003) 2767–2791.
- [109] H. Tresca, Mémoire sur l'écoulement des corps solides, in: Mémoires de l'Académie des sciences de l'Institut de France, Imp. impériale, 1868.
- [110] D. Han, W. Chen, A nonuniform hardening plasticity model for concrete materials, Mech. Mater. 4 (3) (1985) 283–302.
- [111] B. Amos, L. Xu, J.Z. Kolter, Input convex neural networks, in: D. Precup, Y.W. Teh (Eds.), Proceedings of the 34th International Conference on Machine Learning, in: Proceedings of Machine Learning Research, vol. 70, PMLR, 2017, pp. 146–155.
- [112] F. As'ad, C. Farhat, A mechanics-informed neural network framework for data-driven nonlinear viscoelasticity, in: AIAA SCITECH 2023 Forum, American Institute of Aeronautics and Astronautics, 2023.
- [113] R.W. Ogden, R. Hill, Large deformation isotropic elasticity: on the correlation of theory and experiment for compressible rubberlike solids, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 328 (1575) (1972) 567–583.
- [114] F. Mollica, A.R. Srinivasa, A general framework for generating convex yield surfaces for anisotropic metals, Acta Mech. 154 (1) (2002) 61–84.
- [115] M. Itskov, On the application of the additive decomposition of generalized strain measures in large strain plasticity, Mech. Res. Commun. 31 (5) (2004) 507–517.
- [116] V. Taç, M.K. Rausch, F. Sahli Costabal, A.B. Tepole, Data-driven anisotropic finite viscoelasticity using neural ordinary differential equations, Comput. Methods Appl. Mech. Engrg. 411 (2023) 116046.
- [117] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: Large-scale machine learning on heterogeneous systems, 2015.
- [118] S.R. St. Pierre, K. Linka, E. Kuhl, Principal-stretch-based constitutive neural networks autonomously discover a subclass of ogden models for human brain tissue, Brain Multiphys. 4 (2023) 100066.
- [119] J. Korelc, Multi-language and multi-environment generation of nonlinear finite element codes, Eng. Comput. 18 (4) (2002) 312–327.

- [120] J. Korelc, S. Stupkiewicz, Closed-form matrix exponential and its application in finite-strain plasticity, *Internat. J. Numer. Methods Engrg.* 98 (13) (2014) 960–987.
- [121] M. Hossain, D.K. Vu, P. Steinmann, Experimental study and numerical modelling of VHB 4910 polymer, *Comput. Mater. Sci.* 59 (2012) 65–74.
- [122] E.M. Arruda, M.C. Boyce, A three-dimensional constitutive model for the large stretch behavior of rubber elastic materials, *J. Mech. Phys. Solids* 41 (2) (1993) 389–412.
- [123] M. Van Loocke, C. Lyons, C. Simms, Viscoelastic properties of passive skeletal muscle in compression: Stress-relaxation behaviour and constitutive modelling, *J. Biomech.* 41 (7) (2008) 1555–1566.
- [124] J.A. McCulloch, S.R.S. Pierre, K. Linka, E. Kuhl, On sparse regression, lp-regularization, and automated model discovery, 2023.
- [125] A. Fischer, A special newton-type optimization method, *Optimization* 24 (3–4) (1992) 269–284.
- [126] B. Kiefer, T. Bartel, A. Menzel, Implementation of numerical integration schemes for the simulation of magnetic SMA constitutive response, *Smart Mater. Struct.* 21 (9) (2012) 094007.
- [127] T. Brepols, S. Wulfinghoff, S. Reese, Gradient-extended two-surface damage-plasticity: Micromorphic formulation and numerical aspects, *Int. J. Plast.* 97 (2017) 64–106.
- [128] C. Sansour, I. Karšaj, J. Sorić, On a formulation for anisotropic elastoplasticity at finite strains invariant with respect to the intermediate configuration, *J. Mech. Phys. Solids* 55 (11) (2007) 2406–2426.
- [129] R. Stojanović, S. Djurić, L. Vujosević, On finite thermal deformations, *Arch. Mech. Stosow.* 16 (1964) 103–108.
- [130] L. Vujosevic, V. Lubarda, Finite-strain thermoelasticity based on multiplicative decomposition of deformation gradient, *Theor. Appl. Mech.* (2002) 379–399.
- [131] A. Dorfmann, R.W. Ogden, Nonlinear electroelasticity, *Acta Mech.* 174 (3) (2005) 167–183.
- [132] S. Wulfinghoff, C. Dorn, A continuum theory for stripe-shaped magnetic domains in thin films, *J. Magn. Magn. Mater.* 588 (2023) 171375.
- [133] T. van der Velden, B. Rommes, A. Klink, S. Reese, J. Waimann, A novel approach for the efficient modeling of material dissolution in electrochemical machining, *Int. J. Solids Struct.* 229 (2021) 111106.
- [134] K. Manjunatha, M. Behr, F. Vogt, S. Reese, A multiphysics modeling approach for in-stent restenosis: Theoretical aspects and finite element implementation, *Comput. Biol. Med.* 150 (2022) 106166.
- [135] H. Holthusen, L. Lamm, T. Brepols, S. Reese, E. Kuhl, Theory and implementation of inelastic constitutive artificial neural networks: Source code and data, 2023.