Computational Mechanics
https://doi.org/10.1007/s00466-024-02590-1

ORIGINAL PAPER

ANN material modeling for SMA fibers enhanced with a physical constraint and its application to FE computations

Patrick Weber[1] · **Werner Wagner**[1] · **Steffen Freitag**[1]

Received: 20 September 2024 / Accepted: 20 December 2024

Abstract
So-called shape memory alloys (SMAs) show intriguing multi-physical and history-dependent behavior. This includes most prominently the recovery of their initial shape after inelastic deformation, if the temperature is increased afterwards. This is known as the shape memory effect. The precise and reliable description of this and other SMA phenomena is crucial for industrial applications. Therefore, in addition to the wide range of analytical material models for SMA, we want to apply the material modeling strategy with artificial neural networks (ANN) to SMAs. We define an ANN material model in order to represent the SMA behavior with a feedforward ANN. Therefore, the correct setup of input and output vectors for rate-independent material behavior is investigated. The training is done based on synthetic data. The resulting SMA ANN material model is able to represent the SMA strain–stress behavior generally, for arbitrary strain and temperature fields. The resulting one-dimensional ANN material model is used within finite element computations. This increases the accuracy requirements due to the need for a material tangent. Therefore, we improve the performance of the ANN material model in terms of numerical stability by enforcing a material tangent related constraint during the ANN training process. In order to evaluate the performance of ANN material models during training reliably for these accuracy requirements, in depth studies on different target variables during the training process are done.

Keywords Artificial neural networks · Shape memory alloys · Constrained training · Finite element method · Material modeling

1 Introduction

In order to define a structural mechanics model, different sets of equations are mandatory. Physical conservation equations, like static equilibrium, are basic requirements we as engineers set on our models. These equations depend on assumptions, like quasi static or isothermal conditions, and have to be defined in advance. Kinematic equations describe the movement or deformation of bodies and are purely mathematical. They only differ in accuracy and can as well be chosen based on application-related assumptions. Constitutive equations on the other hand are mostly based on observation of real world material behavior. For the associated mathematical modeling, i.e., the functions and corresponding material parameters, there is no *right* solution, only suitable ones. Therefore, material modeling as a field of research on its own, is confronted among other things with the question of how the behavior observed from real world experiments can be transferred into mathematical equations suitable to be used within mechanical or structural models. The following literature review regarding shape memory alloys and ANN material modeling does not claim to be exhaustive. It is a subjective summary of important contributions to the respective fields of research.

Shape memory alloys (SMAs) are a unique class of materials that exhibit the remarkable ability to recover their original shape upon heating after deformation. This is called shape memory effect and is illustrated in Fig. 1. This actually results from various stress and temperature-induced lattice transformation processes within the material. There are various industrial applications for SMAs, ranging from biomedical

✉ Werner Wagner
werner.wagner@kit.edu

Patrick Weber
patrick.weber@kit.edu

Steffen Freitag
steffen.freitag@kit.edu

[1] Institut für Baustatik, KIT, Kaiserstr. 12, 76131 Karlsruhe, Germany

Published online: 06 January 2025

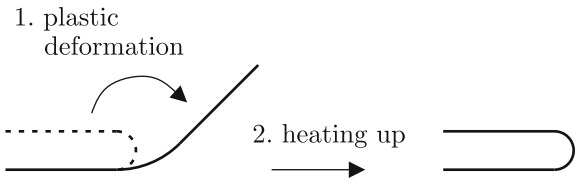

Fig. 1 Shape memory effect (simplified): Plastic deformation can be undone by increasing the temperature

devices like stents or surgical tools [27, 33] to actuators in aerospace or automotive industry [36] and robotics [34]. An overview is given in [24]. In civil engineering, SMAs can be used for example as seismic dampers, due to their ability to dissipate energy without inelastic deformations, see [16, 32, 39] for example. The constitutive modeling of SMA highly depends on the corresponding application and thus the effects one wants to incorporate into the mathematical framework. There is already a large amount of literature on how to model SMAs mathematically, see for example [3–5, 7, 12, 13, 15, 23] and the references therein.

One different approach is to use data obtained from real world experiments directly to train an artificial neural network (ANN). The latter serves afterwards as the material model, see the early works from [9–11] for example. In the past years, the ANN material modeling approach has gained a lot more popularity, as physical knowledge improves the otherwise purely mathematical ANN models. Subsequently, less data is needed and the corresponding ANN material models are robust enough to be incorporated into structural analyzes, e.g., with the finite element method. With a combination of internal variables and potentials, [28, 29] define so-called thermodynamics-based artificial neural networks (TANN) and enforce thermodynamic consistency via a constraint. In [51], the Cholesky factors of the material tangent are defined as input variables to ensure its positive definiteness. By redefining the ANN output as energy function and setting several conditions to the input vector and the weights, the ANN material models from [18, 19] fulfill the polyconvexity conditions exactly in order to model anisotropic hyperelastic material models. Different approaches to incorporate physical restrictions into inelastic ANN material models are computed in [35]. In [46, 47], several physical and mathematical constraints are enforced weakly during the ANN training process in order to enhance the performance of elastic and plastic ANN material models within finite element calculations.

ANNs have already been used in the context of SMA. In [1, 17, 25] ANNs are trained for parameter identification of analytical SMA material models. ANNs to approximate the behavior of whole SMA systems, like actuators, are described in [26, 38, 44]. While describing different effects and dependencies, the ANN models from [2, 14, 31, 37, 42, 43] do not provide a general purpose material model for arbitrary strain and stress paths, which could be used for example within a finite element computation. In [52], a ANN material model is generated, but only evaluated strain-controlled. This application has fewer requirements to the accuracy of the material model, because no material tangent is needed.

In the present work, we want to expand on the described existing progress and define a one-dimensional ANN material model stable enough and in such a way, that it can be used within finite element computations. In order to stay focused on the difficulties arising with the increased accuracy requirements, the training data is generated with an analytical rate-independent material model and applied to a standard feedforward architecture. The resulting ANN material model is able to calculate arbitrary loading and unloading scenarios with different temperatures and can reproduce pseudoelastic, pseudoplastic and shape memory effects. The contributions of the paper are summarized as follows:

- A suitable definition for the feedforward ANN input and output vectors in order to perform arbitrary loading and unloading paths under different temperatures and to be used within materially nonlinear finite element computations.
- Introduction of an inequality constraint enforced during the training process to enhance the numerical stability within equilibrium iteration schemes. The latter is especially important for finite element computations.
- Discussion and suggestion on how to monitor the ANN material model performance during the training process in order to achieve a numerically stable model for finite element computations.
- Application of the one-dimensional ANN material model to SMA fibers and calculation of two and three-dimensional composite structures.

The paper is organized as follows. In Sec. 2, we describe shortly the analytical material model for SMA fibers, which is used to generate the ANN training data. The ANN material model, generation of training data and the training with a physical constraint is shown in Sec. 3. Statistical investigations of the ANN training process, the discussion about different error measures and a simple one-dimensional rod example is given in Sect. 4. In Sect. 5, two- and three-dimensional structures with embedded SMA fibers are computed with the finite element method.

2 Material behavior of SMA fibers

This section describes the main characteristics of a nickel-titanium (NiTi) shape memory alloy (SMA) and the one-dimensional analytical model for the corresponding SMA

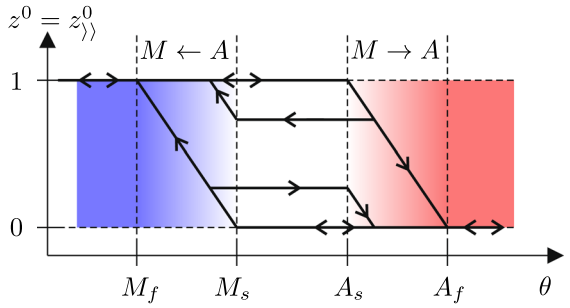

Fig. 2 Evolution of the martensite volume fraction z^0, depending on the temperature θ and its direction sign($\dot{\theta}$), if no additional load is applied. Therefore, $z^0 = z^0_{\rangle\rangle}$ applies

fibers. The latter is used eventually to generate training data for the ANN material model. This section is limited to very few fundamentals, which are required for the generation of the ANN training data and the interpretation of the corresponding results. Additional background information can be found in a variety of literature, for example [30]. It should be noted, that there exists already a lot of literature concerned with analytical material models for SMA, as has been already written in Sect. 1. However, the analytical material model is only a means to an end within this paper to generate training data for the ANN material model.

2.1 Transformation between martensite and austenite

The characteristic properties of SMA stem from the presence of two distinct crystal structures below its melting point, martensite and austenite. The current crystal state is characterized by the total martensite volume fraction $z \in [0, 1]$, with a purely martensitic state ($z = 1$), a purely austenitic state ($z = 0$) and intermediate configurations ($z \in (0, 1)$).

Figure 2 illustrates the evolution of the martensite volume fraction z^0 depending on a change in temperature θ. The Index 0 indicates the absence of additional stresses. We begin with $\theta \in [M_s, A_s]$, where, depending on the history, every combination of martensite and austenite is possible. As the temperature rises, austenite formation initiates at the austenite start temperature A_s. If the temperature surpasses the austenite finish temperature A_f, the material is purely austenitic with $z^0 = 0$. Conversely, as the temperature decreases, martensite formation starts at the martensite start temperature M_s. For temperatures below the martensite finish temperature M_f, the material is purely martensitic with $z^0 = 1$. A change in the martensite volume fraction z^0 only occurs when $\theta \in [M_f, M_s]$ and $\theta \in [A_s, A_f]$. In the following, $M_f < M_s < A_s < A_f$ is assumed.

An overview over the transformation processes under loading/unloading and heating/cooling conditions is depicted in Fig. 3. Under unloaded conditions, the martensite is completely in its *twinned* form. However, if stress is applied, both twinned martensite and austenite can be transformed to *oriented* martensite. Hence, the total martensite volume fraction

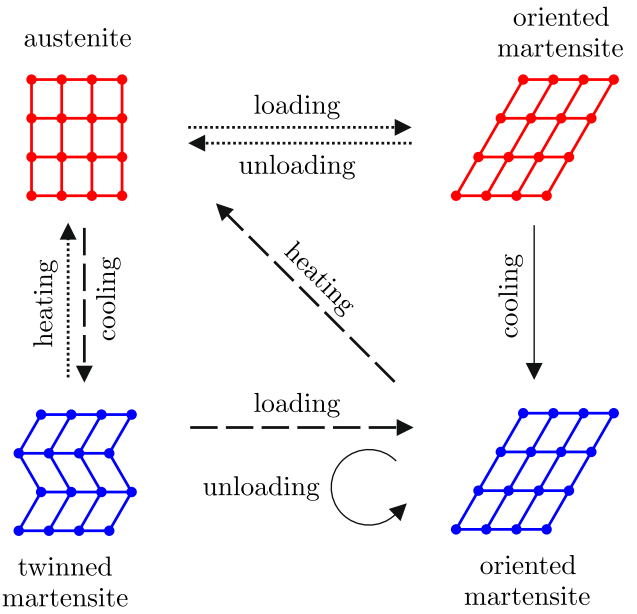

Fig. 3 Overview of load and temperature induced transformation processes of SMAs. The dashed arrows indicate the shape memory effect, the dotted arrows correspond to the pseudoelastic behavior at comparatively high temperatures

$$z = z_{\rangle\rangle} + z_{//} \tag{1}$$

is split into the fractions of twinned martensite $z_{\rangle\rangle}$ and oriented martensite $z_{//}$, each bounded with $0 \leq z_{\rangle\rangle} \leq 1$ and $0 \leq z_{//} \leq 1$. This split has been suggested in [6]. These load- and temperature dependent transformations between three different crystal lattice states result in the well known properties of SMAs. Mainly two different behaviors can be distinguished, depending on the temperature.

- At comparatively low temperatures, twinned martensite undergoes a stress-induced transformation into oriented martensite. The latter is stable at low temperatures, resulting in residual strains after unloading. Thus, this behavior is called *pseudoplasticity*. However, if the oriented martensite is heated up, austenite forms, which itself transforms to twinned martensite after cooling down. This results in the *shape memory effect*, because any remaining strains caused by the former loading process disappear and the initial state is restored. In Fig. 3, this behavior is indicated with the dashed arrows.
- At comparatively high temperatures, austenite undergoes a stress-induced transformation into oriented martensite. The latter is not stable at these temperatures. Thus, the

austenite structure is restored after unloading. No strains remain. Therefore, this behavior is called *pseudoelasticity*, which is indicated in Fig. 3 with the dotted arrows.

Upon completion of the phase transition to oriented martensite, the material exhibits elastic behavior over remarkably wide strain ranges before displaying plastic deformations due to dislocation movements within the lattice, a phenomenon referred to as *superelasticity*.

2.2 Analytical material model

The one-dimensional material model described below is state-of-the-art and in this case taken from [8, 21] and the associated references therein. It is one of many possibilities to model SMA fibers and is comparatively simple: Unlike more sophisticated models, it does not distinguish between material parameters of the austenitic and martensitic phases, such as the Young's modulus. Additionally, it lacks a two-way coupling mechanism between mechanical and thermal fields, and does not incorporate a strain rate dependency, as demonstrated in [16] for instance. Nonetheless, this material model is sufficient to demonstrate the approximation capabilities of feedforward ANNs with respect to SMA behavior. The principle can easily be extended to other models.

Figure 4 provides an overview over the temperature dependent, cyclic behavior of the one-dimensional SMA material model, which is described in the following. Generally, like in small strain plasticity models, the total strain

$$\varepsilon = \varepsilon^e + \varepsilon^i \tag{2}$$

is split into an elastic part ε^e and an inelastic one ε^i. The elastic strains are coupled to the stresses

$$\sigma = C\,\varepsilon^e = C\,(\varepsilon - \varepsilon^i)\,, \tag{3}$$

with Young's modulus C.

At first, we consider the pseudoplastic case with $\theta < \Delta u_0/\Delta\eta_0$ and the loading and unloading scenario described in Fig. 4. Within this material model, the difference of the inner energy Δu_0 and the difference in entropy $\Delta\eta_0$ between the martensitic and austenitic phases are material parameters. If the initial 'yield' stress σ_y is reached, the phase transition from twinned to oriented martensite starts. Kinematic hardening behavior is assumed with the back stresses

$$p = H\,\varepsilon^i \tag{4}$$

beeing linearly dependent on the inelastic strains via the hardening modulus H. The oriented martensite volume ratio $z_{//}$ is modeled linearly dependent on the absolute inelastic strain via

$$z_{//} = \frac{|\varepsilon^i|}{\beta} \in [0, 1]\,. \tag{5}$$

Thus, the material parameter β corresponds to the maximum width of the hysteresis. By definition, the stress-dependent twinned martensite volume ratio

$$z_{\rangle\rangle} = z^0_{\rangle\rangle}\,(1 - z_{//}) \in [0, 1] \tag{6}$$

decreases linearly from the stress-free ratio of twinned martensite $z^0_{\rangle\rangle}$ to zero. Once the phase transformation into oriented martensite has been finished, with $z = z_{//} = 1$, the material behavior exhibits the superelastic branch. In order to capture the described phase transition, the condition

$$\overrightarrow{\Phi} = \left|\sigma - p - \frac{\langle\Delta\Psi\rangle}{\beta}\,\mathrm{sign}(\sigma - p)\right| - \sigma_y \leq 0 \tag{7}$$

is sufficient and can be treated like a yield criterion. The driving force

$$\Delta\Psi = \Delta u_0 - \theta\,\Delta\eta_0 \tag{8}$$

depends on the current temperature θ and shifts the set of admissible strain- and stress configurations with $\langle\Delta\Psi\rangle/\beta$. This is also depicted in Fig. 4. Its influence on the transitions starts if the temperature is above a certain threshold, which is modeled with the Macaulay brackets

$$\langle\Delta\Psi\rangle = \begin{cases} \Delta u_0 - \theta\,\Delta\eta_0 & \theta \geq \Delta u_0/\Delta\eta_0 \\ 0 & \theta < \Delta u_0/\Delta\eta_0 \end{cases}\,. \tag{9}$$

In the purely pseudoplastic case, it has no effect.

However, if the temperature is increased above the threshold $\Delta u_0/\Delta\eta_0$, the material behavior changes. The set of admissible stress–strain pairs shifts optically, as depicted in Fig. 4, along three straight lines. The center of the half pseudoplastic hysteresis is marked with a red dot and is characterized by the stress $\langle\Delta\Psi\rangle/\beta$. The purely pseudoelastic case is reached if $\langle\Delta\Psi\rangle/\beta = \sigma_y$. The dimension of the maximum hysteresis always stays the same with $\beta \times 2\sigma_y$ in the strain- and stress direction. In the loading case, the condition (9) is not sufficient anymore. This can be realized for example with $p = 0, \sigma \ll 1$ and $\langle\Delta\Psi\rangle/\beta > \sigma_y$: The criterion (9) would indicate a phase transition to oriented martensite right at the beginning, even if hardly any stress is present. Therefore, an additional criterion is needed in the case $\langle\Delta\Psi\rangle > 0$ for the loading condition, which is

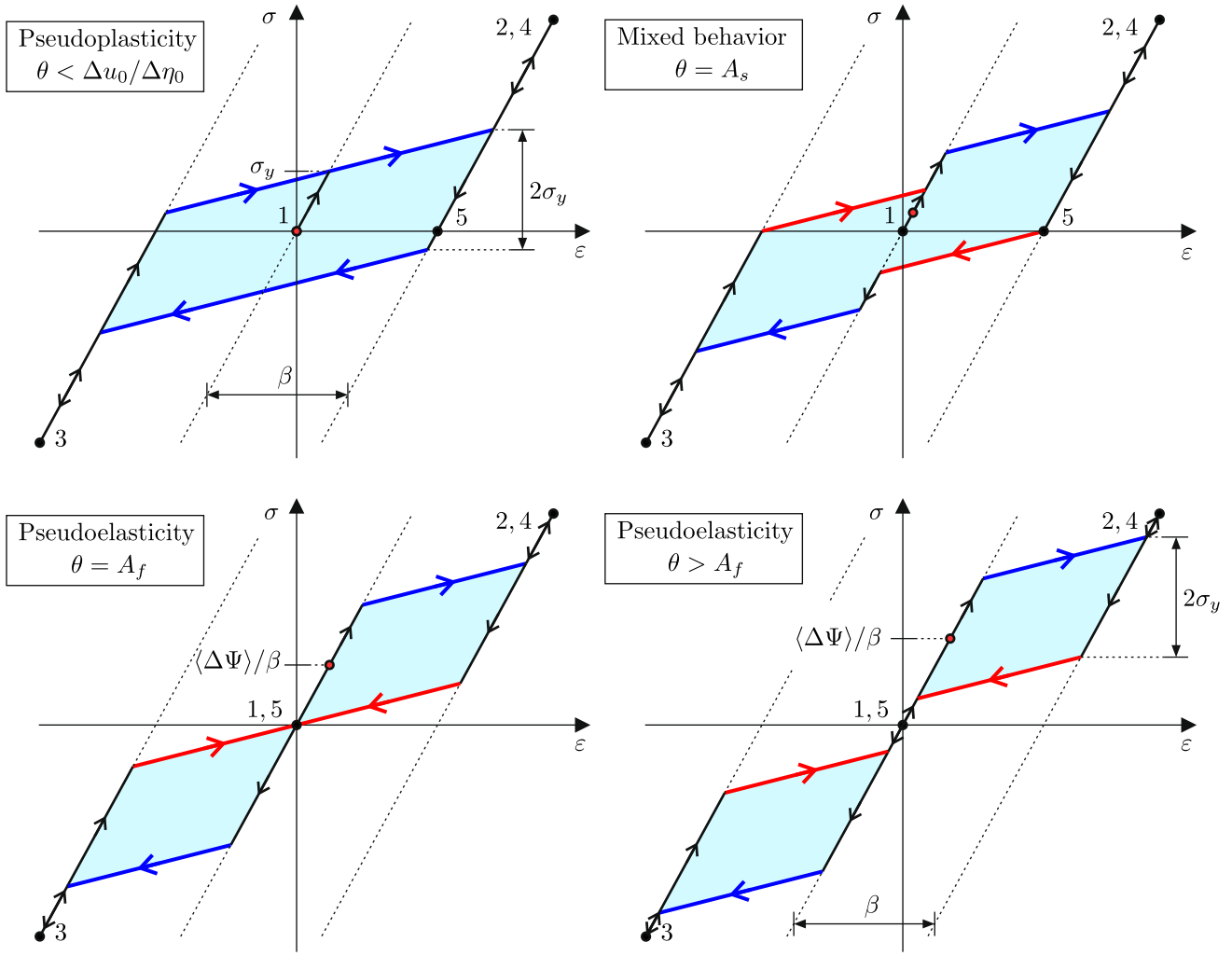

Fig. 4 Cyclic behavior of the one-dimensional SMA material model depending on different temperatures. Points 1 to 4 are accessed after another, before unloading to point 5 occurs. Up from a threshold temperature $\Delta u_0/\Delta\eta_0$, the space of admissible strain–stress-pairs shifts linearly depending on the temperature. The size of this maximum hysteresis is $\beta \times 2\sigma_y$. The stress σ_θ as the respective center is marked with a red dot. For $\theta = A_s$, the residual strains after unloading are just about β. For $\theta = A_f$, no residual strains occur after unloading. Edges with loading condition $\overrightarrow{\Phi} = 0$ are highlighted blue, edges with unloading condition $\overleftarrow{\Phi} = 0$ are highlighted red

$$\sigma_c = |\sigma - p| - \langle\Delta\Psi\rangle/\beta \overset{!}{\geq} 0\,. \tag{10}$$

The points on the maximum hysteresis, which fulfill these loading conditions, are marked blue in Fig. 4. On the other hand, condition (9) cannot be used for the pseudoelastic or mixed behavior transition back to austenite. These points are marked red in Fig. 4. A distinction has to be made between loading [L] and unloading [U], which is done with

$$\overleftrightarrow{\varepsilon} := \varepsilon^i \operatorname{sign}\left(\sigma - p - \frac{\langle\Delta\Psi\rangle}{\beta}\operatorname{sign}(\varepsilon^i)\right) \begin{cases} \geq 0, [L] \\ < 0, [U]. \end{cases} \tag{11}$$

In the loading case [L], condition (9) has to be used combined with (10). In the unloading case [U], another condition

$$\overleftarrow{\Phi} = \left|\sigma - p - \frac{\langle\Delta\Psi\rangle}{\beta}\operatorname{sign}(\varepsilon^i)\right| - \sigma_y \leq 0 \tag{12}$$

has to be applied. The necessity for this differentiation can be comprehended by comparing loading and unloading condition (9) and (12) at point 5 for case $\theta = A_s$ in Fig. 4: With $\sigma = \langle\Delta\Psi\rangle/\beta + H\beta + \sigma_y - 2\sigma_y = 0$, $p = H\beta > 0$ and therefore $\operatorname{sign}(\sigma - p) = -1$ follows $\overrightarrow{\Phi} = |2\langle\Delta\Psi\rangle/\beta - \sigma_y| - \sigma_y$, which is not zero. On the other hand, with $\operatorname{sign}(\varepsilon^i) = \operatorname{sign}(\beta) = 1$ is $\overleftarrow{\Phi} = |-\sigma_y| - \sigma_y = 0$, which starts the phase transition correctly.

The introduced loading and unloading conditions in combination with the limit value for the inelastic strains fully describe the underlying material model. It is specified with the six introduced material parameters, Young's modulus C, hardening modulus H, maximum inelastic strain β, half the height of the maximum hysteresis σ_y and the differences in inner energy and entropy Δu_0 and $\Delta \eta_0$ of the martensitic and austenitic phases. The threshold temperatures M_s, M_f, A_f and A_s, see Fig. 2, depend on these six parameters as follows.

$$M_s = \frac{\Delta u_0}{\Delta \eta_0} + \frac{\sigma_y \beta}{\Delta \eta_0} \qquad M_f = \frac{\Delta u_0}{\Delta \eta_0} + \frac{\sigma_y \beta}{\Delta \eta_0} + \frac{H \beta^2}{\Delta \eta_0}$$
$$A_f = \frac{\Delta u_0}{\Delta \eta_0} - \frac{\sigma_y \beta}{\Delta \eta_0} \qquad A_s = \frac{\Delta u_0}{\Delta \eta_0} - \frac{\sigma_y \beta}{\Delta \eta_0} + \frac{H \beta^2}{\Delta \eta_0} \tag{13}$$

This is done by definition. Graphically, the transition temperatures M_s, M_f, A_s and A_f are determined if the upper left, upper right, lower right and lower left points of the maximum size hysteresis touches the ε-axis, respectively. In Fig. 4, this is shown for $\theta = A_s$ and $\theta = A_f$.

For the evolution of the inelastic strain, an associated 'flow' rule is considered with

$$\dot{\varepsilon}^i = \lambda \frac{\partial \overrightarrow{\Phi}}{\partial \sigma} = \lambda \operatorname{sign}\left(\sigma - p - \frac{\langle \Delta \Psi \rangle}{\beta} \operatorname{sign}(\sigma - p)\right) \tag{14}$$

in the loading case and

$$\dot{\varepsilon}^i = \lambda \frac{\partial \overleftarrow{\Phi}}{\partial \sigma} = \lambda \operatorname{sign}\left(\sigma - p - \frac{\langle \Delta \Psi \rangle}{\beta} \operatorname{sign}(\varepsilon^i)\right) \tag{15}$$

in the unloading case. λ is the 'plastic' multiplier. While the inelastic strains increase or decrease, the analytical tangential stiffness is

$$C_T = \frac{CH}{C + H}\,, \tag{16}$$

which can easily be confirmed with Eqs. (3), (4) and the corresponding consistency condition $\overrightarrow{\Phi} = 0$ or $\overleftarrow{\Phi} = 0$.

2.3 Implicit integration

The previously given equations are treated here numerically within an implicit time integration scheme in combination with a return mapping algorithm. We start at pseudo time t_n, with given temperature θ_n, inelastic strain ε^i_n and volume ratios of the oriented martensite $z_{//n}$ and stress-free twinned martensite $z^0_{\rangle\rangle n}$. The temperature θ_{n+1} and strain ε_{n+1} of the next pseudo time step t_{n+1} are given. We aim to calculate the new stresses σ_{n+1} and all associated history variables. At first, Eqs. (14) and (15) are integrated and combined to

$$\varepsilon^i_{n+1} = \varepsilon^i_n + \gamma \operatorname{sign}\left(\xi_{n+1}\right), \tag{17}$$

with the unknown relative stress variable

$$\xi_{n+1} := \begin{cases} \sigma_{n+1} - p_{n+1} - \frac{\langle \Delta \Psi_{n+1} \rangle}{\beta} \operatorname{sign}(\sigma_{n+1} - p_{n+1}) & \text{[L]} \\ \sigma_{n+1} - p_{n+1} - \frac{\langle \Delta \Psi_{n+1} \rangle}{\beta} \operatorname{sign}(\varepsilon^i_{n+1}) & \text{[U]} \end{cases} \tag{18}$$

including the loading [L] and unloading [U] case. The algorithmic plastic multiplier

$$\gamma := \lambda \Delta t \tag{19}$$

includes the pseudo time increment $\Delta t = t_{n+1} - t_n$ and the plastic multiplier λ.

We consider a trial state in which the inelastic strain is not changing with $\varepsilon^i_{n+1} = \varepsilon^i_n$. The trial stresses and back stresses are

$$\sigma^{tr}_{n+1} = C(\varepsilon_{n+1} - \varepsilon^i_n) \tag{20}$$

and

$$p^{tr}_{n+1} = H \varepsilon^i_n\,, \tag{21}$$

respectively. The trial relative stress ξ^{tr}_{n+1} is calculated analogously with Eq. (18) and used within the applying loading or unloading transition condition $\Phi^{tr} := \Phi(\xi^{tr}_{n+1})$ to evaluate if the trial state is valid or not. In the latter case, assuming the consistency condition, the algorithmic plastic multiplier can be calculated with

$$\gamma = \frac{\Phi^{tr}}{C + H}\,. \tag{22}$$

Afterwards, ε^i_{n+1}, $z_{//n+1}$, σ_{n+1} and p_{n+1} can be calculated with the given equations, evaluated at pseudo time step t_{n+1}. The algorithmic tangent corresponds to the analytical one, see Eq. (16).

3 ANN material model for SMA fibers

The aim of this section is to describe how feedforward artificial neural networks (ANN) can be used to model the stress–strain behavior of shape memory alloy (SMA). The application within this paper is limited to the assumptions of rate-independence and a given temperature field, which itself is not influenced by deformation. It should be noted, that ANN material modeling is itself state-of-the-art, as has been written in Sect. 1. However, the application to SMAs in order to provide a general purpose material model and the enrichment of the corresponding training process with physical constraints are new to the best of our knowledge.

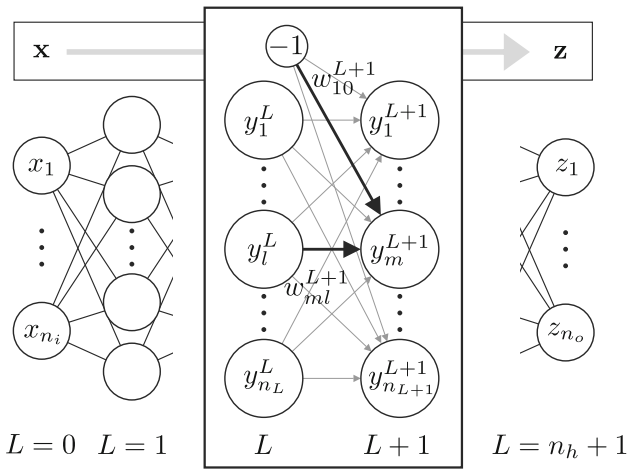

Fig. 5 Topology of a fully connected feedforward ANN: the input vector is processed through hierarchically sorted hidden layers into the output vector **z**. The information transmission between two consecutive layers L and $L+1$ is given as well

3.1 Feedforward ANN

A feedforward ANN is a function

$$\mathbf{z} = \mathbf{f}^{\mathrm{ANN}}(\mathbf{x}, \mathbf{w}) \tag{23}$$

with the input vector $\mathbf{x} \in \mathbb{R}^{n_i}$ and the output vector $\mathbf{z} \in \mathbb{R}^{n_o}$. The n_w-dimensional vector $\mathbf{w}$ collects the weights, which are the free functional parameters. A feedforward ANN is composed of hierarchical layers, the input layer $L = 0$, n_h hidden layers $L = 1, ..., n_h$ and the output layer $L = n_h + 1$, see Fig. 5. Each layer L consists of n_L neurons, between which the information is transferred. The number of hidden layers n_h and the number of neurons per layer $n_1, ..., n_{n_h}$ are hyperparameters, which have to be chosen. A specific topology is labeled as follows: [n_i-n_1-...-n_{n_h}-n_o]. In general, the more neurons and layers are selected, the more complex is the possible mapping. In our case, this has to be adjusted to the material behavior one wants to approximate and is done via trial and error.

In the following, we describe briefly the mapping $\mathbf{f}^{\mathrm{ANN}}$ from Eq. (23). Later, it is used to calculate the stresses of the ANN material function. We consider two arbitrary consecutive layers L and $L+1$, see Fig. 5. Both layers are fully connected, with scalar weights w_{ml}^{L+1} assigned to each connection. The neuron output values $y_1^L, ..., y_{n_L}^L$ of layer L have already been calculated. In order to calculate the neuron outputs of layer $L+1$, at first the weighted sum

$$s_m^{L+1} = (-1)\, w_{m0}^{L+1} + \sum_{l=1}^{n_L} y_l^{L+1}\, w_{ml}^{L+1} = \sum_{l=0}^{n_L} y_l^{L+1}\, w_{ml}^{L+1} \tag{24}$$

has to be calculated for each neuron $m = 1, ..., n_{L+1}$. The so called bias value w_{m0}^{L+1} is considered as weight to an additional input neuron, which is $y_0^L = -1$. The neuron output

$$y_m^{L+1} = g(s_m^{L+1}) = \ln\left(1 + e^{s_m^{L+1}}\right) \tag{25}$$

is calculated with the activation function $g(s)$. Throughout this paper, we use the Softplus-function for the hidden layers. Thus, the nonlinearity of the mapping $\mathbf{f}^{\mathrm{ANN}}$ results from combining and chaining these simple functions (24) and (25). At the input layer $y_l^{[0]} = x_l$ applies and at the output layer $y_m^{[n_h+1]} = z_m$. For the latter, the activation function $g_o(s) = s$ is used. Subsequently, by applying the weighted sum (24) and the activation function (25) successively from the input to the output layer, the overall ANN function $\mathbf{f}^{\mathrm{ANN}}$ can be calculated.

The ANN function (23) is differentiable, which follows from the choice of the activation function. Therefore, the Jacobian

$$\mathbf{J} = \frac{\mathrm{d}\mathbf{f}^{\mathrm{ANN}}(\mathbf{x}, \mathbf{w})}{\mathrm{d}\mathbf{x}} \tag{26}$$

can be calculated analytically as derivative with respect to the input vector **x**. Later, it is used for the material tangent stiffness C_T.

The weights are determined during the training process on the basis of a set of given training data

$$T = \{(\mathbf{x}_k, \mathbf{t}_k)\}\,, \quad k = 1, ..., P\,. \tag{27}$$

They consist of known target vectors $\mathbf{t}(\mathbf{x})$ to specified input vectors **x**. The goal of the training process is to determine a suitable set of weights $\hat{\mathbf{w}}$, such that the ANN provides a good approximation

$$\mathbf{z}_k = \mathbf{z}(\mathbf{x}_k, \hat{\mathbf{w}}) \approx \mathbf{t}_k \quad \forall\, k = 1, ..., P\,. \tag{28}$$

This is done by minimizing the error function

$$\mathcal{L}_D(\mathbf{w}) = \frac{1}{2P} \sum_{k=1}^{P} \left|\left|\mathbf{z}_k - \mathbf{t}_k\right|\right|^2, \tag{29}$$

considering the deviation from the training data in a mean square sense. We use the Quasi Newton method with a Wolfe condition line search strategy [48, 49] as minimization algorithm, as it shows superior convergence behavior compared to other solvers. This is especially important in the case of ANN material modeling, because a high level of accuracy is needed due to the necessity of the Jacobian (26).

The ANN itself, the data generation and the training process are programmed in Matlab [41] without using the

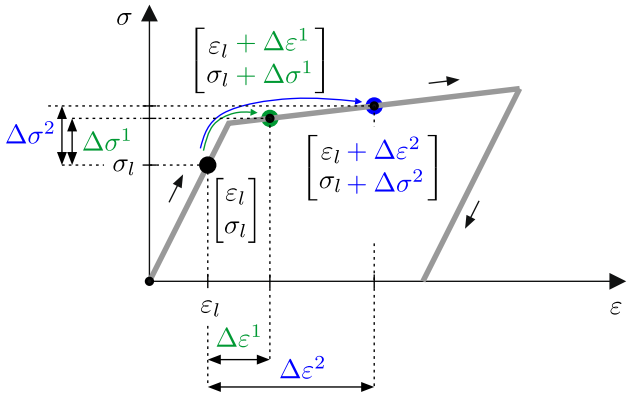

Fig. 6 Motivation for one-dimensional incremental formulation for ANN material model. Starting from an equilibrium state $(\varepsilon_l, \sigma_l)$ and taking a finite step $\Delta\varepsilon$ to the next one

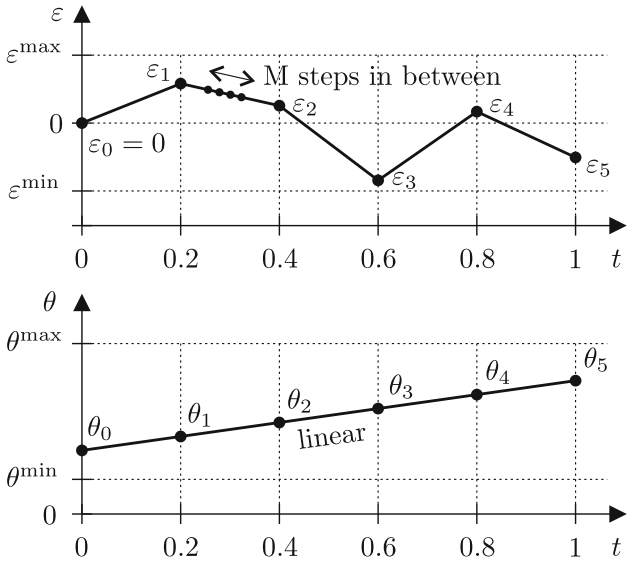

Fig. 7 Definition of paths to calculate incremental training samples from. With parameters $\varepsilon^{\min}$, $\varepsilon^{\max}$, $\theta^{\min}$, $\theta^{\max}$ and M a family of random loading paths is defined

corresponding toolboxes. In [47] for example, detailed algorithms are given on how to implement the ANN functions and the corresponding derivatives analytically.

3.2 Feedforward ANN as SMA material model

The input and output vectors $\mathbf{x}$ and $\mathbf{z}$ have to be defined in such a way, that the desired material behavior can be represented as a function. This task has neither a unique solution, nor is it straight forward. It must be ensured that there is no ambiguity, i.e., no different output vectors for a given input vector.

Within this paper we use an incremental definition for the strains and stresses, which already has been proposed in [11] for example and used in [46, 47] for rate-independent plasticity, see Fig. 6. For SMA, the current temperature θ and the temperature increment $\Delta\theta$ have to be incorporated into the input vector as well, leading to the following input vector definition with five input variables.

$$\mathbf{x} = \begin{bmatrix} \Delta\varepsilon \\ \Delta\theta \\ \varepsilon_l \\ \sigma_l \\ \theta \end{bmatrix} \in \mathbb{R}^5 \tag{30}$$

The strains ε_l and stresses σ_l are history variables, which have to be saved from the last equilibrium state. The temperature increment $\Delta\theta = \theta - \theta_l$ must also be calculated with the temperature θ_l of the last state. In the first step of each loading history, $\Delta\theta$ is always assumed to be zero. The output vector is the scalar stress increment, with

$$z = \Delta\sigma \in \mathbb{R}. \tag{31}$$

For the calculation of the current stress

$$\sigma^{\text{ANN}} = \sigma_l + \Delta\sigma = \sigma_l + f^{\text{ANN}}(\mathbf{x}, \mathbf{w}) \,, \tag{32}$$

the stress of the last equilibrium state is needed. The material tangent

$$C_T^{\text{ANN}} = \frac{\mathrm{d}\sigma}{\mathrm{d}\varepsilon} = \frac{\mathrm{d}\Delta\sigma}{\mathrm{d}\Delta\varepsilon} = \frac{\partial f^{\text{ANN}}(\mathbf{x}, \mathbf{w})}{\partial x_1} = \frac{\partial z(\mathbf{x}, \mathbf{w})}{\partial x_1} \tag{33}$$

can be calculated analytically. A detailed description on the ANN algorithms needed for implementation is given in [47].

3.3 Data generation with SMA material model

The training data will be generated with the analytical SMA material model from Sect. 2. Therefore, a family of random paths will be specified as depicted in Fig. 7. Starting at $\varepsilon_0 = 0$, one strain path consists of five randomly generated strain nodes ε_1 to ε_5, which are defined as uniformly distributed random numbers bounded by $[\varepsilon^{\min}, \varepsilon^{\max}]$. These five strain nodes are defined at the fifth points of the pseudo time interval $t \in [0, 1]$. While the strain paths are polygons, the temperature paths are always linear. Given two random temperatures θ_0 and θ_5 at $t = 0$ and $t = 1$, the whole temperature path is linearly interpolated in between. The temperatures are defined as uniformly distributed random numbers bounded by $[\theta^{\min}, \theta^{\max}]$. Within each of the five intervals $[(\varepsilon_i, \theta_i), (\varepsilon_{i+1}, \theta_{i+1})]$, $i = 0, ..., 4$, the strain and the temperature is interpolated linearly with M steps in between. Thus, a family of random paths can be defined by five parameters, namely $\varepsilon^{\min}$, $\varepsilon^{\max}$, $\theta^{\min}$, $\theta^{\max}$ and M. One specific path, a realization of the random path, is given by seven uniformly distributed random numbers for ε_1 to ε_5, θ_0 and θ_5.

There are two main reasons, why we have chosen this kind of sampling strategy. First, with five changes in direction a lot of possible loading and unloading scenarios are considered, for example reloading, while previously being in a phase transition to martensite, see the red lines in Fig. 4. And second, parameter studies are possible while changing the number of random path realizations. For further investigations, like ANN training with real world data, other strategies must be chosen which are more compatible with experimental setups. However, this is not the topic of this paper.

After defining the strain-temperature paths, the corresponding stresses are calculated with the analytical SMA material model from Sect. 2. Consequently, for every path, a set

$$\mathcal{P} = \{(\varepsilon_1, \sigma_1, \theta_1), ..., (\varepsilon_n, \sigma_n, \theta_n), ..., (\varepsilon_N, \sigma_N, \theta_N)\} \tag{34}$$

of equilibrium points is given and can be used for the ANN training data generation, with $N = 5M + 1$ here. At this point it should be noted, that these paths could be gathered by experiments, numerical homogenization of micro structures or, as in this paper, by analytical benchmark material models. It is important to mention, that all further calculations are based on pure strain, stress and temperature information, which could be, in principle, obtained by real experiments.

However, for the calculation of those paths with the analytical SMA material model from Sect. 2, an additional assumption has to be made. Giving the starting temperature θ_0 only, the martensite volume fraction z^0 is not unique, see Fig. 2. Therefore, we assume always the maximum martensite ratio possible, i.e.,

$$z^0 = \begin{cases} 1 & \theta_0 \leq A_s \\ 1 - \dfrac{\theta_0 - A_s}{A_f - A_s} & A_s < \theta_0 \leq A_f \\ 0 & \theta_0 > A_f \end{cases}, \tag{35}$$

for the generation of the training paths (34).

The given paths (34) of equilibrium points have to be transformed with the introduced input and output vector definition (30) of the ANN material model in order to be used within the training process. The increments $\Delta\varepsilon$, $\Delta\sigma$, and $\Delta\theta$ can be obtained by subtracting pairs of equilibrium points from each other, while keeping in mind the right order. In the first step of each loading history, $\Delta\theta$ is always assumed to be zero. Eventually, this leads to ANN data samples with the input and output vectors

$$\mathbf{x} = \begin{bmatrix} \varepsilon_n - \varepsilon_{n-\Delta n} \\ \theta_n - \theta_{n-\Delta n} \\ \varepsilon_{n-\Delta n} \\ \sigma_{n-\Delta n} \\ \theta_n \end{bmatrix} \quad \text{and} \quad z = \sigma_n - \sigma_{n-\Delta n}\,. \tag{36}$$

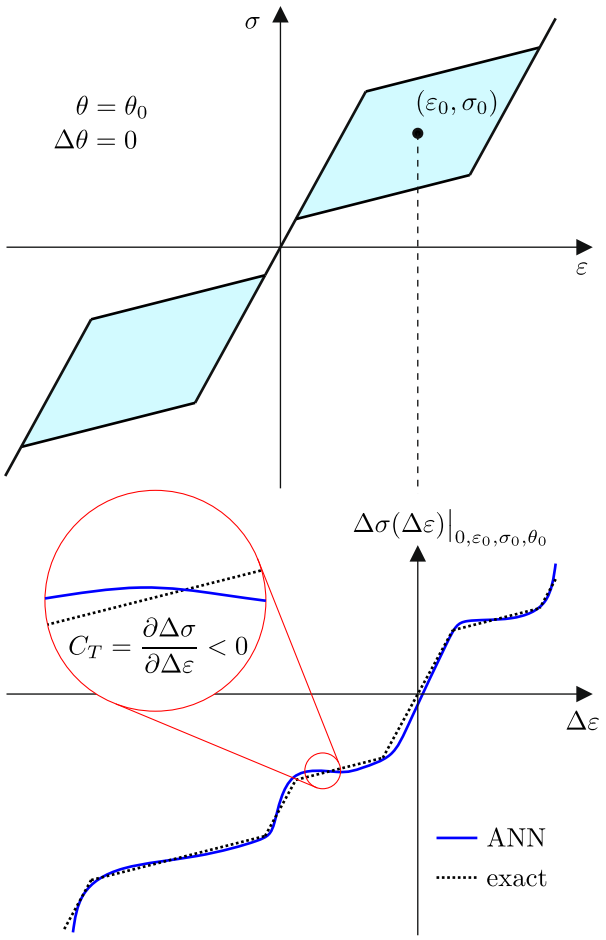

Fig. 8 A locally unstable ANN material model results from an apparently good approximation of the underlying input–output mapping. This is shown on an exemplary cut-function $\Delta\sigma(\Delta\varepsilon)$ with constant values for $\Delta\theta = 0$, $\varepsilon_l = \varepsilon_0$, $\sigma_l = \sigma_0$ and $\theta = \theta_0$

The index delay Δn must be greater or equal to zero in order to maintain the right sequential order. It is advantageous to allow a large set of index delays, the maximum one being $\Delta n \in \{0, 1, 2, ..., M\}$ in our case. This is due to the fact, that increments should only be made within one single loading direction in order to get thermodynamically valid samples. Based on our sampling strategy, we decided to only calculate increments within each of the five different loading directions, see Fig. 7. This strategy has also been used for two- or three-dimensional inelastic ANN material modeling, see [47].

3.4 Training constraint for ANN material model

In principle, it is possible for an ANN to learn the mapping $\mathbf{x} \mapsto z, \quad \mathbb{R}^5 \to \mathbb{R}$, which is assumed for the material behavior. However, this functional approximation is not restricted by physical or numerical properties. This is shown in Fig. 8, where a fictive cut-function $\Delta\sigma(\Delta\varepsilon)$ is shown with constant

values for $\Delta\theta = 0$, $\varepsilon_l = \varepsilon_0$, $\sigma_l = \sigma_0$ and $\theta = \theta_0$. The blue line indicates the ANN approximation on top of the dashed analytical solution. While apparently the fitting is suitable, the corresponding ANN material model shows local instabilities with $C_T < 0$. Within finite element calculations this could lead to diverging equilibrium iterations and thus a termination of the whole computation.

In order to prevent this, we enforce the material tangent to be higher than a given threshold $C_T^{\min}$:

$$C_T^{\text{ANN}} - C_T^{\min} = \frac{\partial z(\mathbf{x}, \mathbf{w})}{\partial x_1} - C_T^{\min} \overset{!}{\geq} 0 \,. \tag{37}$$

This is done with the penalty method, thus weakly. However, enforcing only $C_T^{\text{ANN}} > 0$ would lead locally to negligible small material tangents. Within an equilibrium iteration scheme, this would result in very large strain increments, far outside of the training space and subsequently to divergence. The respective value for $C_T^{\min}$, i.e., the minimum of all possible tangent values, depends on the specific example and can be estimated by the given data, especially for a one-dimensional material model.

Applying the classical penalty method, the data error function $\mathcal{L}_D(\mathbf{w})$ from Eq. (29) is extended with the error function

$$\begin{aligned} \mathcal{L}_C(\mathbf{w}) &= \frac{\epsilon}{2P_C} \sum_{k=1}^{P_C} \left(\min \left\{ 0, C_T(\mathbf{x}_k^C, \mathbf{w}) - C_T^{\min} \right\} \right)^2 \\ &= \frac{\epsilon}{2P_C} \sum_{k=1}^{P_C} \left(\min \left\{ 0, \frac{\partial z(\mathbf{x}_k^C, \mathbf{w})}{\partial x_1} - C_T^{\min} \right\} \right)^2 . \end{aligned} \tag{38}$$

The penalty parameter ϵ controls the hardness of penalization and has to be chosen by the user. The set of constraint samples $\{\mathbf{x}_k^C\}_{k=1}^{P_C}$ can be generated randomly within the training space defined by the input samples $\mathbf{x} \in T$, see Eq. (27). We use the convex hull strategy introduced in [47] to randomly generate P^C constraint samples. It should be noted, that these constraint samples do not have to be part of the set of the training data and they do not need any information about the target values t. The ANN training process subsequently is the minimization of the total error function

$$\begin{aligned} \mathcal{L}(\mathbf{w}) &= \mathcal{L}_D(\mathbf{w}) + \mathcal{L}_C(\mathbf{w}) \\ &= \frac{1}{P} \sum_{k=1}^{P} (z(\mathbf{x}_k, \mathbf{w}) - z_k)^2 \\ &\quad + \frac{\epsilon}{2P_C} \sum_{k=1}^{P_C} \left(\min \left\{ 0, \frac{\partial z(\mathbf{x}_k^C, \mathbf{w})}{\partial x_1} - C_T^{\min} \right\} \right)^2 . \end{aligned} \tag{39}$$

Further information, including algorithms for the implementation of inequality constraints into the ANN training process and background information of constraint ANN training in total can be found in [46, 47].

3.5 Remark to other ANN architectures or material models

The ANN material model formulation described above is not unique. One could redefine the input and output vectors directly with the strains ε and stresses σ of the unknown equilibrium point, providing the same information in another fashion. It should be noted, that also the temperature of the last step θ_l could be used within the input vector instead of the current temperature θ, see Eq (30). Instead of using a feedforward ANN, which naturally is not directly suitable for time dependent processes, a recurrent neural network could be used as well. Furthermore, physical or mathematical restrictions must not be incorporated via weak enforcement, but could be build into the ANN architecture itself. This has been done in [35] by defining the ANN to approximate a free energy function and not directly the stresses. Thus, we want to emphasize that the approach suggested in this paper is by far not the only one possible and that it could be interesting to compare different ANN approaches later on data from real experiments. Nonetheless, the conclusion of this paper regarding the improvement of numerical stability, the definition of reliable error measures and the application within FE computations, are applicable to all mentioned alternatives above.

3.6 Incorporation into FE scheme

The ANN material model is incorporated into the material library of an extended version of the finite element program FEAP [40]. It will be used within a geometrically nonlinear truss element in 3D space. The stress σ^{ANN} is part of the element vector of inner forces

$$\mathbf{F}_e = \int_{L^e} \mathbf{B}^T \sigma^{\text{ANN}} A \mathrm{d}x \,, \tag{40}$$

with element length L^e and cross section area A. Vector $\mathbf{B}$ contains the derivatives of the linear ansatz functions and maps the virtual element nodal displacement vector $\delta\mathbf{v}_e$ to the element strains $\delta\boldsymbol{\varepsilon}_e = \mathbf{B}\delta\mathbf{v}_e$. The material tangent C_T^{ANN} belongs to the material part of the tangent stiffness matrix

$$\mathbf{K}_{TMe} = \int_{L^e} \mathbf{B}^T C_T^{\text{ANN}} \mathbf{B} \, A \mathrm{d}x \,. \tag{41}$$

It should be noted, that ANN material models are generally nonlinear, which makes a linearization of the underlying virtual work equation necessary. More details on the implementation of a history dependent ANN material model are

Table 1 Material parameters for the analytical SMA material model of Sect. 2. The lower four are dependent on the upper six, see Eqs. (13)

parameter	symbol	unit	value
Young's modulus	E	N/mm^2	60 000
hardening modulus	H	N/mm^2	2 000
half hystereris height	σ_y	N/mm^2	120
half hystereris width	β	–	0.05
energy difference	Δu_0	N/mm^2	−204
entropy difference	$\Delta \eta_0$	N/mm^2/K	−0.8
martensite finish	M_f	K	241.25
martensite start	M_s	K	247.50
start mixed behavior	$\Delta u_0/\Delta \eta_0$	K	255.00
austenite start	A_s	K	256.25
austenite finish	A_f	K	262.50

Table 2 Parameters for random path generation, see Sect. 3.3

$\varepsilon^{\min}$	$\varepsilon^{\max}$	$\theta^{\min}$	$\theta^{\max}$	M
−0.07	0.07	230 K	300 K	40

provided in [47]. A general description of the finite element method can be found in a variety of literatures, see, e.g., [50] for example, where also the nonlinear truss element is presented.

4 ANN training and investigation of SMA effects

In this section, we investigate the performance of the introduced ANN material model for SMA fibers. We will demonstrate that it can reliably represent all the effects incorporated in the underlying analytical material model. Section 4.1 describes the training process and illustrates the impact of the material tangent constraint. We place particular emphasis on defining meaningful strain- and stress-based error measures to evaluate the quality of the ANN material model. It should be emphasized, that this study goes beyond the usual criterion of hyperparameter searches, where only the regression against the given data is monitored. In Sect. 4.2, we will show that the ANN material model can accurately represent pseudoplasticity, pseudoelasticity, and the shape memory effect within a one-dimensional finite element example. The material parameters provided in Table 1 will be used throughout Sects. 4 and 5 for the analytical SMA material model. These parameters have also been utilized in [8] and [21].

4.1 Training of the ANN material model

In several preliminary studies, a topology of [5-25-25-25-1] with a total of 1476 weights has been found sufficient for the ANN material model to represent the five-to-one mapping of SMA behavior. This has been done mostly by trial and error. However, more sophisticated optimization-based hyperparamter searches are possible. Additionally, 50 randomly generated paths, as described in Sect. 3.3, provide enough information to set the weights during the training process. These paths are generated using the parameters from Table 2. With the given topology, generating more random paths does not result in significant improvement. It is important to consider the mutual influence of data and unknowns (weights) carefully. More weights allow for the approximation of more complex functional behavior but require significantly more information. Conversely, if the topology is fixed, an infinite amount of data does not lead to infinite precision. With well-distributed data, convergence behavior is observed. For more information, see [45]. In the future, the number of paths could be significantly reduced by defining specific loading and unloading scenarios rich in information, thus minimizing the number of real-world experiments required to gather data. However, this is beyond the scope of this paper.

In order to calculate the incremental samples, all possible index delays $\Delta n \in \{0, 1, ..., M = 40\}$ are used, as discussed in Sect. 3.3. This results in a large amount of training data, so the total number of all possible incremental samples is randomly truncated to $P = 20\,000$ samples, which has proven sufficient for training the ANN topology described above. Increasing the number of incremental samples does not significantly affect the ANN model's performance. The material tangent constraint from Sect. 3.4 is enforced on $P_C = 10\,000$ randomly generated constraint samples, with the threshold set at $C_T^{\min} = 1000$ N/mm^2. The impact of the penalty parameter ϵ is investigated subsequently. Each training process is terminated after 10 000 epochs, i.e., 10 000 optimization iteration steps using the Quasi-Newton method.

To accurately evaluate the training processes, meaningful error measures must be defined. While the data error function (29) or the total error function (39) are minimized during the training process, they are not sufficient to rate the resulting ANN material models because they lack information about the generalization behavior of the ANN material model. Overfitting could not be observed. Furthermore, a test error in a classical sense, meaning

$$\mathcal{L}_T(\mathbf{w}) = \sqrt{\frac{1}{P}\sum_{k=1}^{P} \left|\left| z(\mathbf{x}_k, \mathbf{w}) - t_k \right|\right|^2}, \tag{42}$$

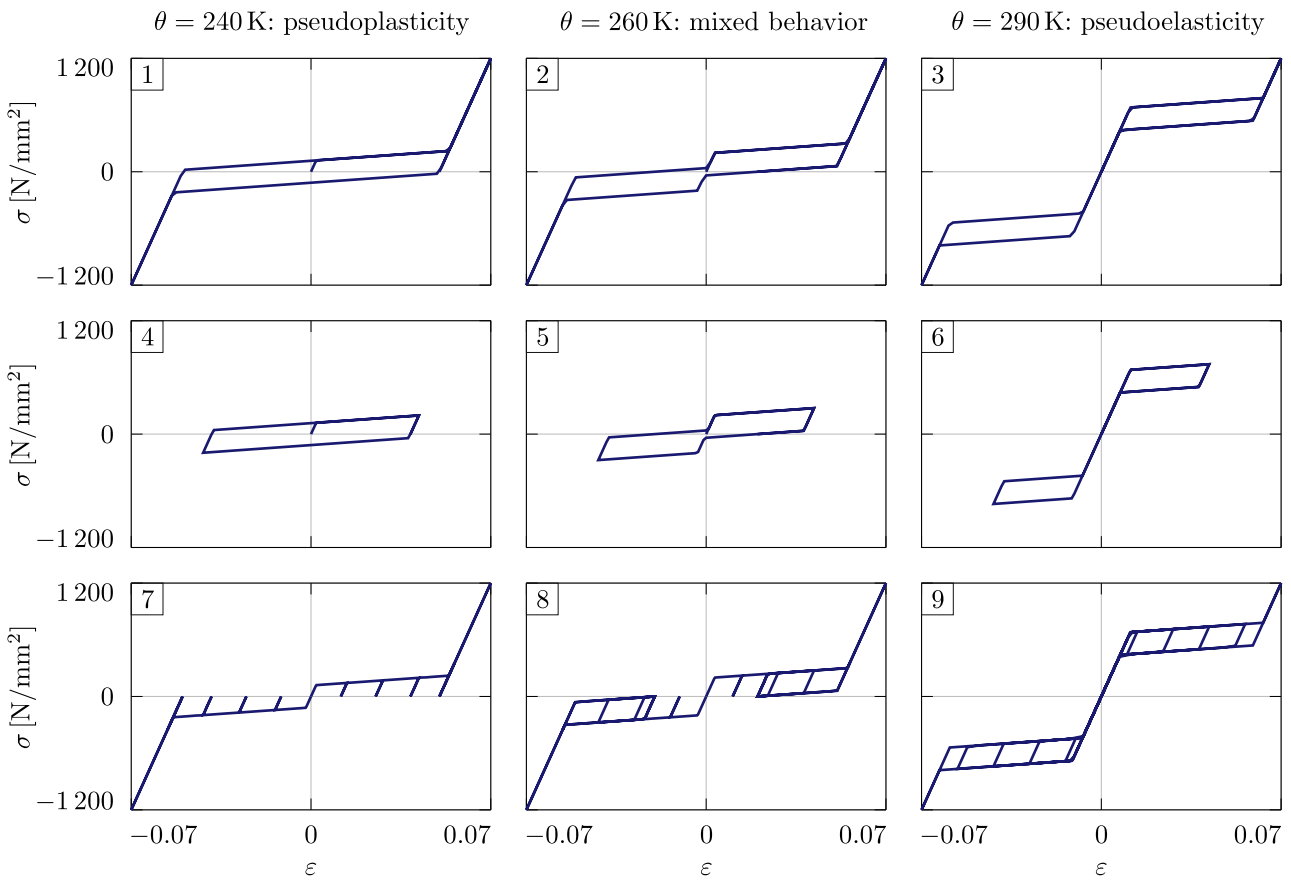

Fig. 9 Nine paths to calculate the stress-based error measure $\mathcal{L}_\sigma$ and strain-based error measure $\mathcal{L}_\varepsilon$ from Eqs. (43) and (44) for each ANN material model. The reference solution depicted is calculated with the analytical SMA material model. They include six cyclic paths with strain-controlled loading and a stress-controlled unloading at the and, as well as three paths consisting of consecutive loading and unloading parts

with samples $\{\mathbf{x}_k, t_k\}$, which have not been seen by the ANN during the training process, is also not sufficient. While it is suitable for normal function approcimation or classification tasks, it lacks information about the performance of the ANN material model within an equilibrium iteration, mainly due to the need for the material tangent. Consequently, other error measures are proposed to evaluate the performance of ANN material models during the training process reliably. These error measures could also be implemented in a hyperparameter search in order to find the best suitable model, which is out of the scope of this paper.

Nine paths are defined, which consist of strain- and stress-controlled parts. They are depicted in Fig. 9. The shown reference solution is calculated with the analytical SMA material model. Paths 1 to 3 consist of consecutive strain-controlled loadings to $\varepsilon^{\max} = 0.07$, $\varepsilon^{\min} = -0.07$ and back to $\varepsilon^{\max}$. Afterwards, a stress controlled unloading to $\sigma = 0$ is performed. Paths 4 to 6 are similar to paths 1 to 3, but with only 60 % strain amplitude. In paths 7 to 9, five consecutive strain-controlled loadings and stress-controlled unloadings are performed, with the latter always unloading to $\sigma = 0$. The turnaround points occur at the fifth of the maximum strain points. Paths 1, 4 and 7 are calculated with a constant temperature of $\theta = 240\,\mathrm{K}$, which leads to pseudoplastic behavior. With $\theta = 260\,\mathrm{K}$, mixed behavior is observed within paths 2, 5 and 8. Pseudoelasticity is achieved for a temperature of $\theta = 290\,\mathrm{K}$, which is valid for paths 3, 6 and 9. Each of the three loading and one unloading directions of the cyclic paths 1 to 6 is discretized with $M_T = 80$ steps. The consecutive loading and unloading directions of paths 7 to 9 are discretized with $M_T = 40$ steps, respectively. Eventually, this leads to $N_i = 321$ points for paths 1 to 6 and $N_i = 802$ points for paths 7 to 9.

The strain- and stress-controlled parts differ in how the stresses and strains of the ANN material model solution has to be calculated. In each of the strain-controlled parts, the stress $\sigma^{\mathrm{ANN}}(\varepsilon_n, \mathbf{w})$ for any load step n is calculated explicitly with Eq. (32) and the given strain ε_n. No material tangent is needed. The strain $\varepsilon^{\mathrm{ANN}}(\sigma_n, \mathbf{w})$ cannot be calculated directly with the ANN in the stress-controlled parts, but has to be

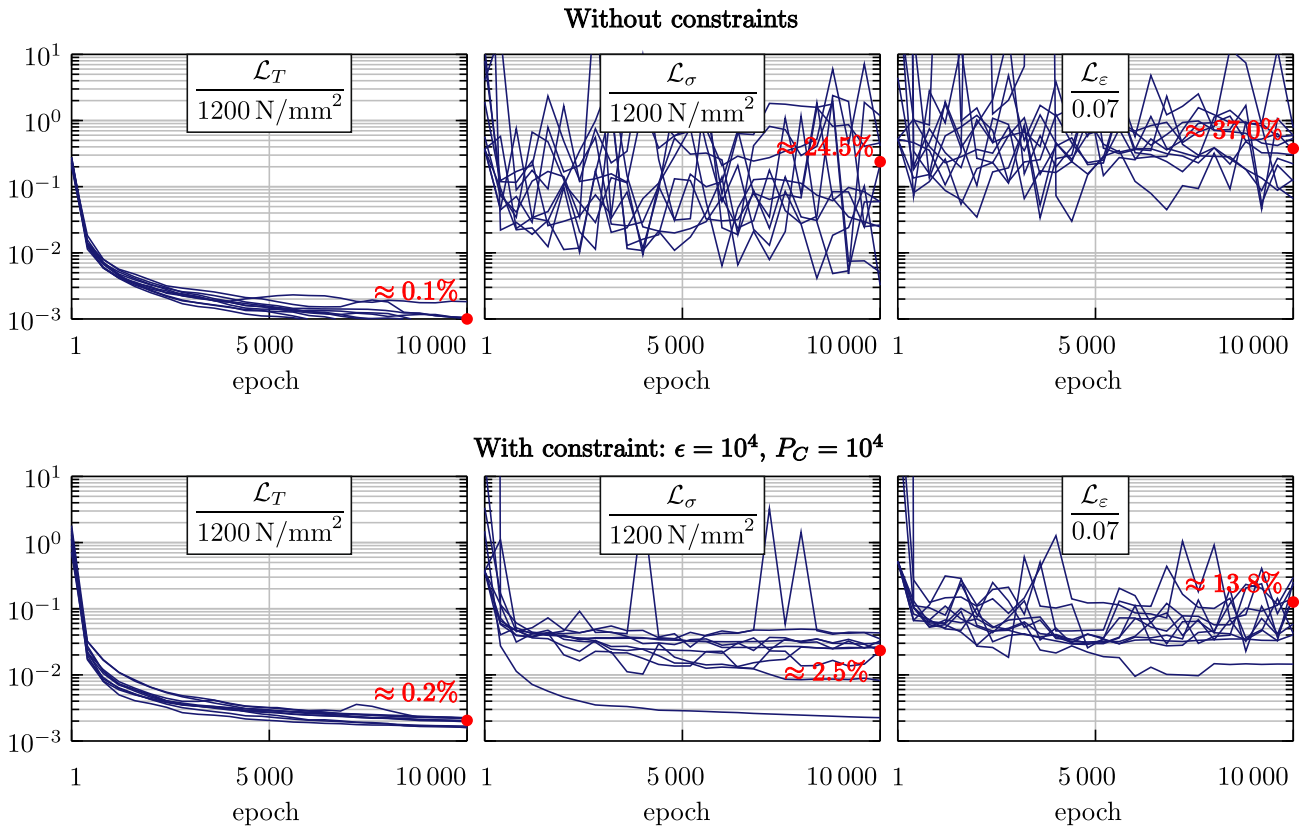

Fig. 10 Ten training runs without constraints and ten runs with constraints. Comparison of three different error measures from Eqs. (42). (43) and (44), each normalized to the maximum stress (1200 N/mm^2) or strain (0.07) respectively. The classical test error $\mathcal{L}_T$ decreases smoothly and indicates a good training. The stress and strain errors from the predefined test scenarios, see Fig. 9, behave more unsteady, while being more reliable in terms of usability as ANN material model. The averaged error values at 10000 epochs is marked red

calculated within an equilibrium iteration, by solving the equality equation $\sigma^{\mathrm{ANN}}(\varepsilon^{\mathrm{ANN}}, \mathbf{w}) - \sigma_n = 0$ for $\varepsilon^{\mathrm{ANN}}$ with given stress σ_n. This is done with Newton's method. Therefore, the material tangent of Eq. (33) is needed. Calculating the strains implicitly through an equilibrium iteration is significantly more complex than explicitly evaluating the ANN material model. This complexity must be considered when defining reliable error measures.

We propose two error measurements, considering the deviation of the stresses

$$\mathcal{L}_\sigma := \sqrt{\frac{\sum_{i=1}^{9}\sum_{n=1}^{N_i}\left(\sigma^{\mathrm{ANN}}(\varepsilon_n, \mathbf{w}) - \sigma_n\right)^2}{\sum_{i=1}^{9} N_i}} \tag{43}$$

and the strains

$$\mathcal{L}_\varepsilon := \sqrt{\frac{\sum_{i=1}^{9}\sum_{n=1}^{N_i}\left(\varepsilon^{\mathrm{ANN}}(\sigma_n, \mathbf{w}) - \varepsilon_n\right)^2}{\sum_{i=1}^{9} N_i}}\,. \tag{44}$$

as the root mean squared error values between ANN material model values and the analytical ones. Another big difference between the error measures from Eqs. (43) and (44) to the classical test error (42) is the evaluation over a whole process. The ANN input $\mathbf{x}_n$ of step n depends on the ANN output value z_{n-1}. Thus, errors can propagate and multiply. This is the reason, why the values of the test error $\mathcal{L}_T$ and the stress-based path error $\mathcal{L}_\sigma$ differ by magnitudes, which is shown in the following example.

In order to get an impression of the different error measures, they are compared in Fig. 10 within ten training runs without constraints and ten with constraints. For the latter a penalty parameter $\epsilon = 10^4$ is chosen. The classical test error $\mathcal{L}_T$ from Eq. (42) is calculated with 5% of the given incremental samples, which have not been used during the training process. All three error measures are normalized to the maximum stress 1200 N/mm^2 or strain 0.07 respectively, in order to compare them with each other. In both cases, with and without the constraint, the test error decreases smoothly, indicating a successful training process. There is also no significant variation among the ten training realizations. Additionally, the constraint does not significantly affect

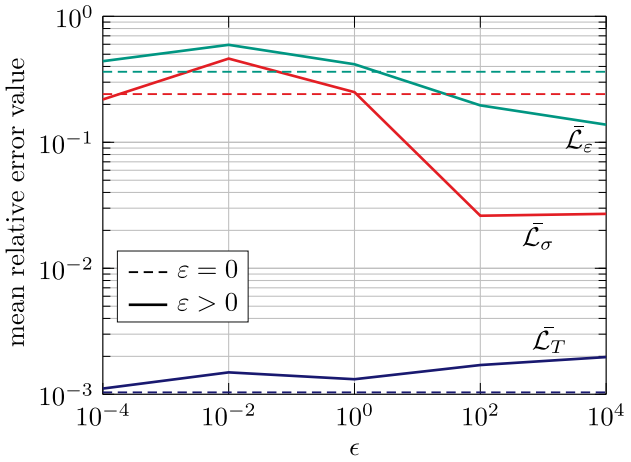

Fig. 11 Relative error measures averaged over ten training runs

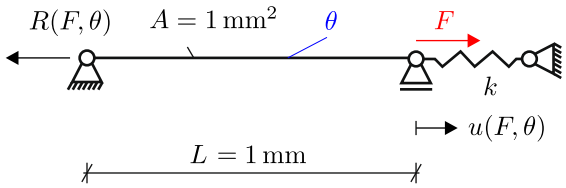

Fig. 12 Unit rod: system, parameter and quantity of interest definitions

Table 3 Relative error values of three ANNs trained with constraints and without constraints from Sect. 4.1 which are used for the unit rod calculations of Sect. 4.2

ANN	$\mathcal{L}_T/1200$ [−]	$\mathcal{L}_\sigma/1200$ [−]	$\mathcal{L}_\varepsilon/0.07$ [−]
Without constraints			
1	0.0010	1.1828	0.5960
2	0.0008	0.0032	0.5399
3	0.0010	0.0591	0.3024
With constraints			
1	0.0017	0.0259	0.1507
2	0.0023	0.0023	0.0255
3	0.0016	0.0264	0.0145

the test error, as previously investigated in [46]. However, the stress- and strain-based error measures $\mathcal{L}_\sigma$ and $\mathcal{L}_\varepsilon$ draw a different picture. First, there is a lot more scatter between different training runs and also between different epochs of the same training run. It should be noted, that this scatter makes these error measures pretty much unsuitable for an early stopping stategy. The constraint has a positive impact on the average values at the last epoch 10 000, which are highlighted in red. The stress-based path error $\mathcal{L}_\sigma$ with the applied constraint is ten times lower than without the constraint. Also, the scatter decreases a lot, indicating a more resilient ANN material model resulting from different training processes. Nevertheless, error magnitudes of the relative test error cannot be reached, which comes from the pathwise calculation which has been described above. Even in the constraint case, an absolute path-based strain error of $\mathcal{L}_\varepsilon = 0.138 \cdot 0.07 \approx 0.01$ seems pretty high. This is a bit misleading. Consider an unloading scenario: a comparatively low stress error value at the beginning of unloading can lead automatically to a constantly high error in the strains during unloading, even if the relative path is exact. Therefore, all these errors should always be interpreted from a relative point of view.

In Fig. 11, the relative error measures $\mathcal{L}_T/1200\,\mathrm{N/mm}^2$, $\mathcal{L}_\sigma/1200\,\mathrm{N/mm}^2$ and $\mathcal{L}_\varepsilon/0.07$ from Fig. 10 are shown for different values of the penalty parameter ϵ. They are evaluated at epoch 10 000 respectively and averaged over 10 training runs each. This averaging process is indicated with a bar: $\bar{\mathcal{L}}_T$, $\bar{\mathcal{L}}_\sigma$ and $\bar{\mathcal{L}}_\varepsilon$. It can be seen, that up from a penalty parameter of $\epsilon = 1$, the constraint has a positive effect on the path-dependent stress and strain error measures, while the test set error approximately stays the same. The highest improvement can be seen for the stress error measure. For the finite element examples in the following sections, it is to be expected that the ANN material models trained with the material tangent constraint show a significantly better performance than the one without constraint.

4.2 Basic effects tested on a unit rod

The system of a unit rod with cross section area $A = 1\,\mathrm{mm}^2$ and length $L = 1\,\mathrm{mm}$ is depicted in Fig. 12. The uniform temperature θ and the load F are given and the horizontal displacement u at the right support and the reaction force R at the left support are measured. A spring with stiffness k adds additional resistance to the system and will be used to show the two-way shape memory effect. The rod is calculated with FEAP [40] and consists of a single geometrically linear truss element. It should be noted, that the force induced loading results in a calculation with an equilibrium iteration for both loading and unloading, which requires the material tangent in every loading step.

In Sect. 4.1, ten ANN material models have been trained for each setting of a penalty parameters ϵ. Now, we take three of these models with $\epsilon = 0$ (no constraint) and three models with $\epsilon = 10^4$. They are part of the error measure visualization of Fig. 10. The corresponding relative error values for the chosen six ANN material models are given in Table. 3. The error values have been evaluated at the end of the training process, respectively. The numbering 1 to 3 is arbitrary. ANN 1 without constraints is not directly related to ANN 1 with constraint. Without constraint, the test error $\mathcal{L}_T$ is lower all the time. On average, the path-based stress and strain errors

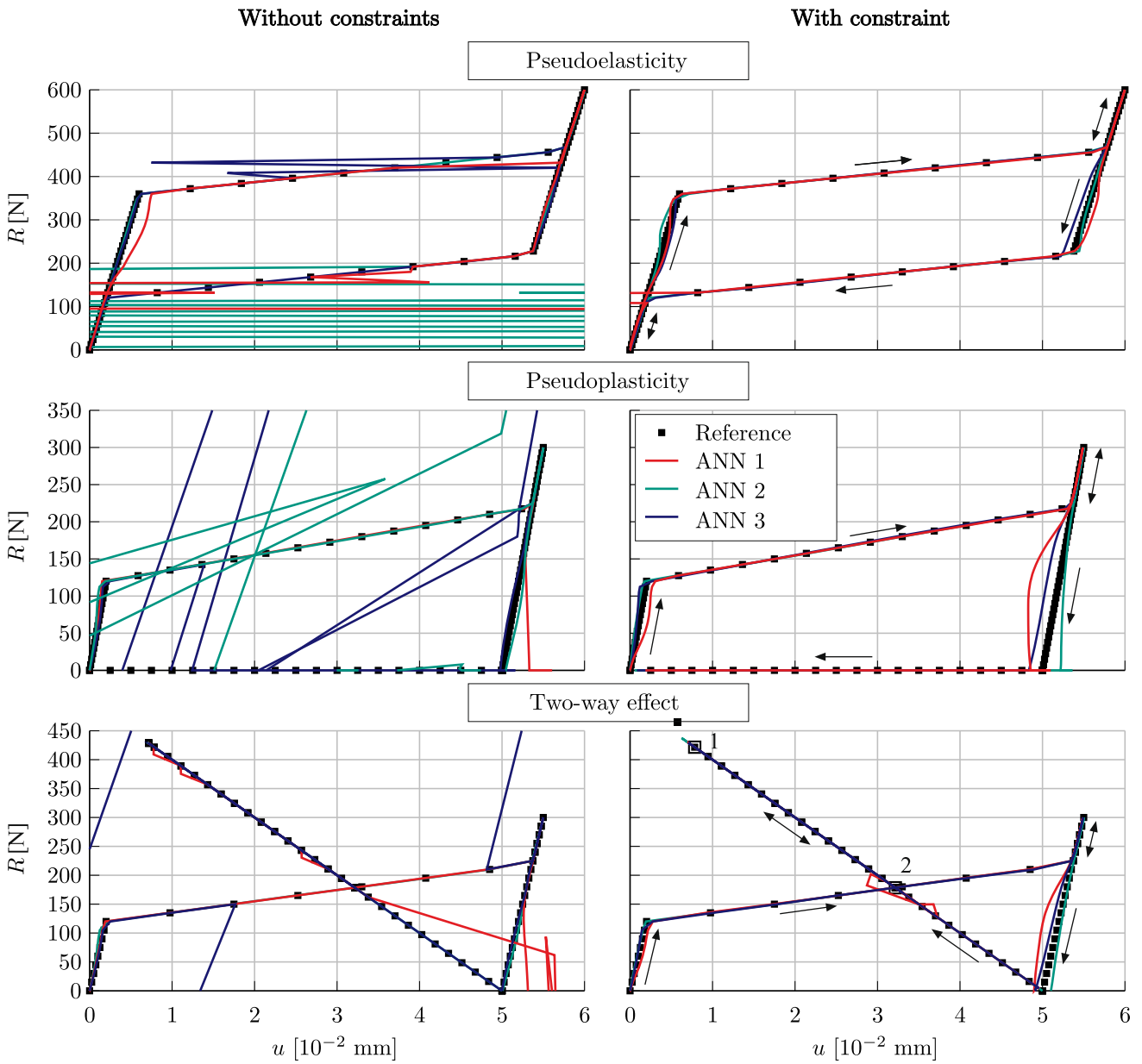

Fig. 13 Unit rod: pseudoplasticity, pseudoelasticity and two-way effect calculated with three ANN material models with and three without constraint, respectively. The ANN models sharing a number are not related. The analytical reference solution indicates the step size

$\mathcal{L}_\sigma$ and $\mathcal{L}_\varepsilon$ are lower in the constraint-case. This has already been discussed in Sect. 4.1.

The performance of the described ANN material models is compared within three standard scenarios: pseudoelasticity, pseudoplasticity and the two-way effect. The corresponding solution graphs showing the reation force R with respect to the displacement u are given in Fig. 13. They are described in detail in the following subsections. It should be noted, that the results are depicted as long as the corrresponding equilibrium iteration had found a solution. This leads partially to very chaotic behavior, especially in the case without constraint.

4.2.1 Unit rod: pseudoelasticity

The spring stiffness k is constantly zero in this case, leading to the reaction force R being equivalent to the load F. In order to achieve pseudoelastic behavior, the temperature is held constant at $\theta = 270$ K, which is above the austenite finish temperature of $A_f = 262.5$ K, see Table. 1. The load is increased from $F = 0$ N to $F = 600$ N in 50 steps. This can be checked with the filled squares indicating the reference solution with the analytical SMA material model. Phase transitioning starts at $F = 360$ N from austenite to oriented martensite. This and all other significant points during

loading and unloading can be calculated analytically with the SMA material model of Sect. 2.2. As soon as the maximum inelastic strain $\varepsilon^i = 0.05$ is reached at $F = 460\,\text{N}$, superelastic behavior can be observed. After reaching its maximum value, the load is decreased again in 50 steps to $F = 0$. At $F = 220\,\text{N}$ austenite starts to form until $F = 120\,\text{N}$, when the elastic branch is reached again. There are no remaining inelastic strains after unloading.

It is evident, that the three ANN material models trained with the tangent constraint achieve a much more stable and reliable prediction of the SMA wire behavior than the three training realizations without constraint. During numerical tests, it has been observed that the two phase transition regions pose the most significant numerical challenges. This can also be seen in Fig. 13 at the upper left figure. Additionally, an oscillating behavior of the ANN 2 without constraints (green curve) can be seen. Although the equilibrium iteration terminates at a mathematical solution, it is not useful from a physical point of view. While the ANN material models trained with the constraint show an overall better behavior, a non-physical jump can be observed for ANN 1 at the end of the back transitioning process. This shows, that the weakly enforced inequality constraint only improves the performance of the ANN material models. Exact fulfillment of physical or mathematical restrictions cannot be guaranteed.

4.2.2 Unit rod: pseudoplasticity

The spring stiffness k is constantly zero in this case, leading to the reaction force R being equivalent to the load F. Pseudoplastic behavior is accomplished by setting the temperature to $\theta = 250\,\text{K}$, which is below the start temperature of mixed behavior at $\Delta u_0/\Delta\eta_0 = 255\,\text{K}$, see Table. 1. The load is increased from $F = 0\,\text{N}$ to $F = 300\,\text{N}$ in 40 steps. This can be checked with the filled squares indicating the reference solution with the analytical SMA material model. Phase transitioning from twinned martensite to oriented martensite starts at $F = 120\,\text{N}$. As soon as the maximum inelastic strain $\varepsilon^i = 0.05$ is reached at $F = 220\,\text{N}$, superelastic behavior can be observed. After reaching its maximum value, the load is decreased again in 40 steps to $F = 0$. In the pseudoelastic case, no back transitioning to twinned martensite occurs, which leads to remaining strains of $\varepsilon^i = 0.05$. Once the structure is unloaded, temperature is increased from $\theta = 250\,\text{K}$ to the austenite finish temperature $\theta = A_f = 262.5\,\text{K}$ in 40 steps. Up from $\theta = 255\,\text{K}$, the area of admissible strain–stress combination, i.e., the maximum size hysteresis, shifts upwards. This can be comprehended with Fig. 4. When temperature increases above the austenite start temperature $\theta = A_s = 256.25\,\text{K}$, there is no permissible material equilibrium state for $\sigma = 0\,\text{N/mm}^2$ and $\varepsilon = 0.05$. Thus, in order to maintain the equilibrium of forces, strain must decrease, which can be observed in Fig. 13. When reaching the austenite finish temperature, all remnant strains have been removed and the initial structure of the system has been restored. This is known as the shape memory effect.

Similar to the case of pseudoelasticity, the ANN material models trained with constraint behave well in contrast to the ones trained without constraint. The elastic and superelastic branches are slightly off the reference solution, but still get it qualitatively. When temperature is increased, the strain should be constantly $\varepsilon = 0.05$ until $\theta = 256.25\,\text{K}$ is reached. The ANN material models represent this behavior also only approximately. At $\theta = 262.5\,\text{K}$, none of the graphs finish at $\varepsilon = 0$, i.e., they do not restore the initial shape of the structure completely.

4.2.3 Unit rod: two-way effect

The first part of this experiment is similar to the pseudoplastic case: spring stiffness is zero, temperature is 250 K and the load is increased to $F = 300$ before decreasing it again to zero. The step size is doubled, leading to 20 steps in the loading and 20 steps in the unloading direction. Now, before increasing the temperature as in the pseudoplastic case, the spring is activated with $k = 10\,000\,\text{N/mm}$. Its stress free configuration is therefore at $u = 0.05$. The spring provides resistance to further displacements, i.e., the contracting of the SMA material due to increasing temperature. This leads, up from $\theta = 255\,\text{K}$, to an increasing reaction force R during the contraction. The temperature is now increased in 40 steps to 300 K. The transition process from oriented martensite to austenite is finished, when the equilibrium point (ε, σ) reaches the elastic branch again. This state is marked as "1" in Fig. 13 on the lower right picture. The corresponding strain ε_1, stress σ_1 and temperature θ_1 can be calculated analytically and are given below.

$$\varepsilon_1 = \frac{k\beta}{C+k} = 0.00714 \tag{45}$$

$$\sigma_1 = \frac{Ck\beta}{C+k} = 428.57\,\text{N/mm}^2 \tag{46}$$

$$\theta_1 = \frac{\Delta u_0}{\Delta\eta_0} - \frac{Ck\beta^2}{(C+k)\Delta\eta_0} - \frac{\sigma_y\beta}{\Delta\eta_0} = 289.29\,\text{K} \tag{47}$$

Increasing the temperature above θ_1 has no effect on the system, which is already fully austenitic. When the temperature is decreased again to 240 K in 48 steps, austenite starts to transform into both oriented and twinned martensite, due to the acting stress. This transformation process ends at point "2" in Fig. 13. Decreasing the temperature further has no effect on the system, because the austenite is fully transformed into martensite.

The ANN material models trained with the constraint are able to represent the two-way effect. There is a local nonphysical part in the graph of ANN 1, but nonetheless the overall

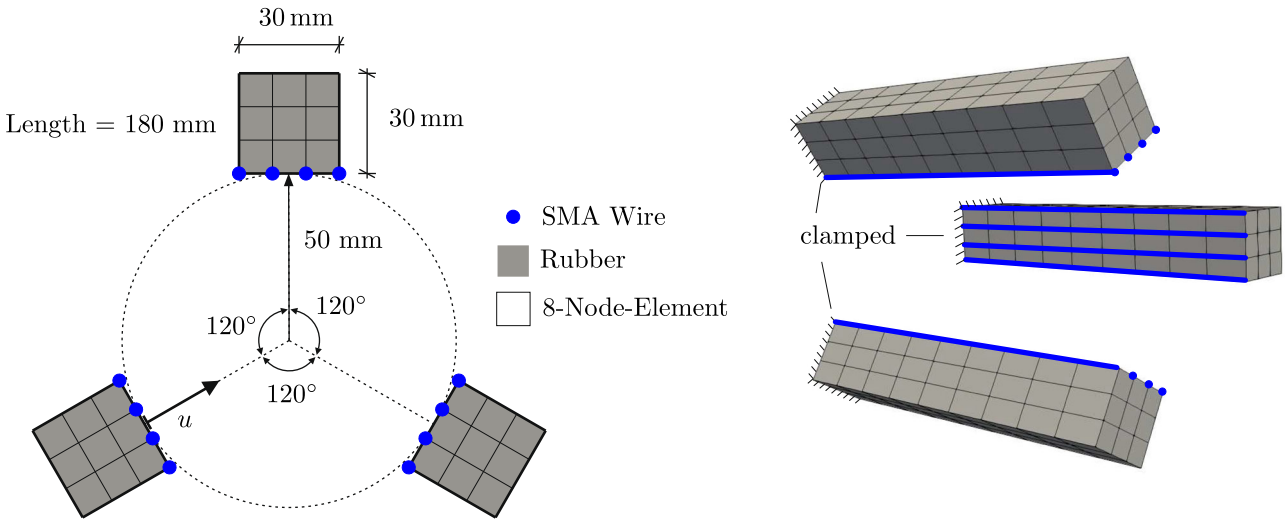

Fig. 14 Rubber grip element with SMA wires: geometry, boundary conditions, FE mesh and position of wires

behavior is met relatively accurate. Similar to the pseudoplastic case, the points "1" and "2", marking the end of the transformation processes, are not exactly met.

4.2.4 Unit rod: conclusion

In general, the ANN material model proposed in Sect. 3.2 is able to represent all relevant features of the underlying analytical SMA material model. When applying the inequality constraint, the numerical stability increases considerably, although local nonphysical behavior cannot be excluded in any case. The experiments showing pseudoelastic and pseudoplastic effects are even more difficult numerically for the ANN material model, because the rod is the only structural element providing stiffness to the system. Thus, if the material model shows local instabilities, the whole computation fails.

5 Numerical examples

In this section, we present finite element computations involving structures coated with SMA fibers in both two and three dimensions. In Sect. 5.1, we demonstrate a three-dimensional grip element where the gripping mechanism is activated through the shape memory effect. In Sect. 5.2, we conduct a numerical four-point bending test of a thermoplast coated with SMA under plain stress conditions.

5.1 Rubber-like grip element

In this section, we compute the gripping mechanism from the rubber-like grip element depicted in Fig. 14. It has also been calculated in [22]. For the arms, a linear elastic material model with Young's modulus $C = 100\,\mathrm{N/mm^2}$ and $\nu = 0$ is assumed, which has also been used in [22]. For the sake of comparability, we stick with these material parameters for the rubber-like material, even though they are not very realistic. The grip element consists of three rectangular bars with area $30 \times 30\,\mathrm{mm^2}$ and length 180 mm. All three arms are fixed on one side. They are placed 120° relative to each other, while their inner faces touch a circle with radius 50 mm. 4 SMA wires with each an area of $1\,\mathrm{mm^2}$ are coated to the inner surfaces of each grip arm. To prepare the shape memory effect, the wires have to be pre-stretched, here to $\varepsilon_0 = 0.05$. This pre-stretching could be achieved by a previously performed pseudoplastic loading and unloading process. The resulting residual strain corresponds to the given pre-stretch, which transfers no stress into the structure. The FE discretization of the rubber-like volumes consists of $3 \times 3 \times 9$ eight or 27 node displacement based isoparametric finite elements, see Fig. 14. Therefore, each SMA wire consists of nine truss elements. These settings match the numerical model of [22], where only 8 node elements have been used, and are adopted for the sake of comparability as well.

The analytical SMA material model from Sect. 2.2 is used with the parameters of Table. 1 to calculate the reference solution. As ANN material model, we choose one from Sect. 4, where the corresponding training processes and numerical studies on a unit rod have already been described. More precisely, we use the ANN material model labeled as "ANN 1" in Fig. 13 for the constraint case. The corresponding data generation and training process is described in Sect. 4.1. The resulting relative strain and stress error measures can also be seen in Table. 3 for "ANN 1" in the constraint case. In order to apply the pre-stretch of $\varepsilon_0 = 0.05$ for the ANN material model numerically, the third input variable of load step $n+1$ is adjusted with

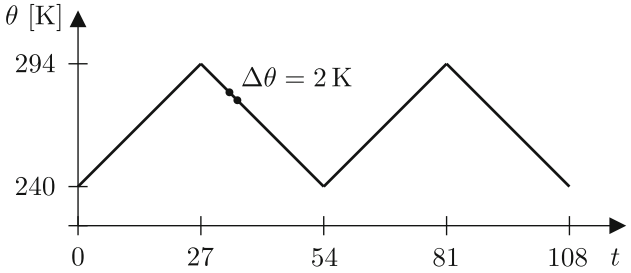

Fig. 15 Rubber grip element with SMA wires: temperature θ is increased and decresed two times between 240 K and 294 K. The temperature step is 2 K

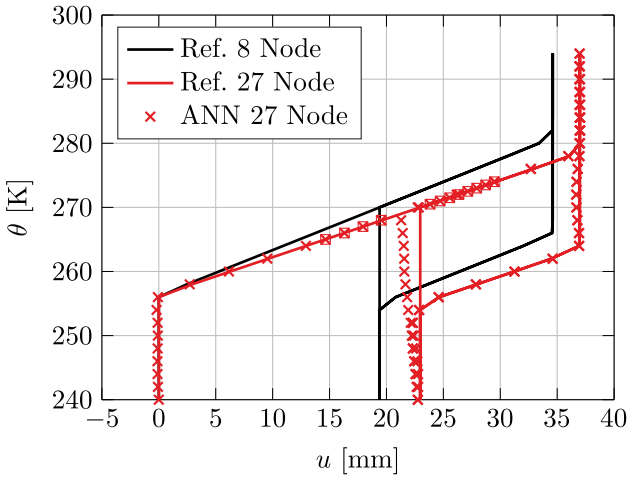

Fig. 16 Rubber grip element with SMA wires: radial displacement u depending on the temperature θ. The reference solution from [22] with the analytical SMA material model is compared to a reference solution with 27 node solid elements and the results with the ANN material model applied on the 27 node FE model. Load steps where the step size had to be adjusted are marked with a square

$$x_3 = \varepsilon_n + 0.05\,, \tag{48}$$

see definition (36). This pre-stretch is mandatory such that the pure temperature loading is able to start the gripping mechanism. At this point it should be noted, that it is comparatively easy to implement this kind of pre-stretch to a feedforward ANN architecture, because all history information is included in the input variables. They can easily be manipulated by the user. On the other hand, a recurrent architecture could not be prepared like this. The pseudoplastic pre-stretching would have to be simulated first in order to obtain the corresponding hidden neuron states.

Starting from $\theta = 240$ K, the temperature is increased to 294 K and decreased again to the initial temperature. This cycle is perfomed two times, see Fig. 15. The numerical temperature steps are 2 K each, leading to 108 steps in total. In the ANN case, we had to adapt the step size two times, which is indicated in Fig. 16. We observe the radial displacement with respect to the given temperature as quantity of interest, see Fig. 14. In Fig. 16, both results from the analytical SMA model and the ANN material model are illustrated for the case of tri-quadratic 27 node elements. The reference solution from [22] is also given, utilizing the analytical material model.

We start with notes on the choice of elements: By comparing both reference results, it is obvious that the given bending problem suffers from a locking phenomenon. The application of 27 node elements reduces this issue, resulting in larger deformations at times of maximum and minimum temperatures during the gripping process. It should be noted, that the lower number of degrees of freedom in the case of 8 node elements plays a minor role compared to the linear ansatz functions. The latter are mainly responsible for the locking effect when calculating this pure bending problem. This can be demonstrated, if the 27 node elements are replaced by 8 node solid shell elements, see e.g. [20]. The corresponding results are identical to the ones obtained by the 27 node elements. Continuing with the ANN and the overall structural behavior: As we can see, the ANN material model is able to predict the structural response correctly, even if there are local deviations. At first, the temperature change has no influence on the radial displacement. The alloy consists of pure twinned martensite. The ANN material model actually shows low initial deflections, which are comparatively low and therefore not harmful. This emphasizes the approximative character of the ANN, as it has to learn all temperature and strain dependent behaviors implicitly from the given data. When the austenite start temperature $A_s = 256.25$ K is reached, see Fig. 4 top right, the SMA wires start to contract, resulting in the grip elements initialize bending. Due to the additional stress in the SMA wires, austenite formation does not stop at the austenite finish temperature $A_f = 262.5$ K, but continues to approximately $\theta = 281$ K, when the bending process terminates. This is similar to the two-way-effect described in Sect. 4.2.3. Subsequent heating up has no further impact. When cooling down, reformation to martensite starts at approximately $\theta = 266$ K and finishes at ca. 255 K. Now, a mixture of twinned and oriented martensite is present. When reheating, the closed circle on the right of Fig. 16 is passed again, never reaching the purely twinned martensite structure. Therefore, the working range of the tip deflection of this rubber-like grip element is between approximately 19 mm and 35 mm. The deviation of the ANN material model solution to the reference one is highest when increasing the temperature again during the second loop. This could be due to accumulating errors, which lead to more and more deviating input vectors. This assumption is opposed by the fact that the points of the first and second loop are congruent after returning to the initial path.

A bending deflection of more than 35 mm compared to the initial length of 180 mm is comparatively large. Therefore, geometrical nonlinearity is necessary to be taken into

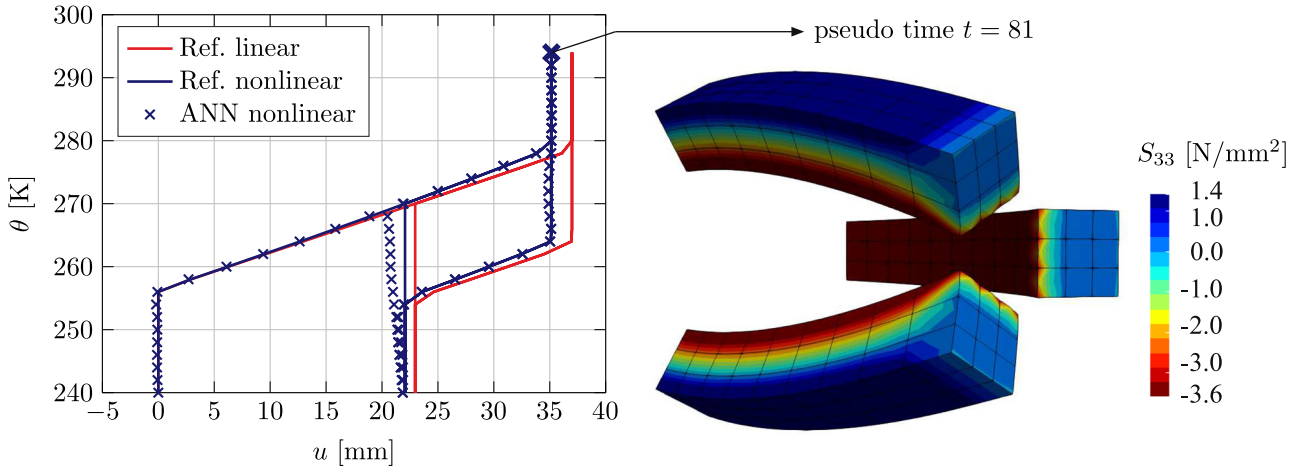

Fig. 17 Left: comparison of the previous geometric linear solution using the SMA reference model to the geometric nonlinear results utilizing the analytical and the ANN material model, respectively. All depicted graphs are calculated with 27 node elements. Right: deflected configuration at pseudo time 81, with longitudinal Piola-Kirchhoff stresses S_{33}

account. Considering small strains, which is valid in this case, geometrical nonlinearity can be included by replacing the linear strains $\boldsymbol{\varepsilon}$ and stresses $\boldsymbol{\sigma}$ by the Green Lagrange strains $\mathbf{E}$ and the 2. Piola-Kirchhoff stresses $\mathbf{S}$. Thus, the rubber-like solid is calculated with a St. Venant material model. The strains and stresses of the one-dimensional SMA fiber model are replaced as well, leading to the new ANN material model input vector

$$\mathbf{x} = \begin{bmatrix} \Delta E \\ \Delta\theta \\ E_l \\ S_l \\ \theta \end{bmatrix} \in \mathbb{R}^5 \tag{49}$$

and output vector $z = \Delta S \in \mathbb{R}$. Considering a local orthogonal coordinate system in three-dimensional Euclidean space, where the x-direction is alongside the SMA fiber axis and the y- and z axes are perpendicular to it. The corresponding displacements are defines as u_x, u_y and u_z. Then, the Green-Lagrangian strain for a SMA fiber is calculated via

$$E = u_{x,x} + \frac{1}{2}\left(u_{x,x}^2 + u_{y,x}^2 + u_{z,x}^2\right). \tag{50}$$

The linear strains $\varepsilon = u_{x,x}$ only contain the first part. It's worth noting that the same ANN material model can be utilized with both strain and stress measures. This is because the information regarding how displacements are used to compute corresponding strains and how stresses are further employed is not included in the ANN's input or output vectors. Moreover, the exchange of ansatz functions and calculation of strains and stresses has nothing to do directly with the defined ANN model. However, as the range of values of the strains and stresses may change, an additional check of the input and output space could be necessary. In this particular case, in no load step and no iteration step was the need for calculating input vectors outside of the initial training space. It should be emphasized again that this interchangeability of strains ε and E, as well as stresses σ and S, is only applicable for small strains.

In Fig. 17, the geometrical nonlinear results calculated with both the analytical and the ANN material model are compared to the linear reference solution. All three computations have been done with 27 node elements. Taking geometric non-linearity into account makes the system softer during the bending process. This leads to the maximum and minimum displacement getting smaller again. In the case of this example, both effects, i.e., the locking effect and the utilization of only linear strain measures, work contrary to each other. As in the geometrical linear case, the ANN is able to represent the structural behavior approximately. In this case, no special treatment of temperature step sizes had to be done.

5.2 Four-point bending test of an aluminum structure reinforced with SMA fibers

In this example, we demonstrate the computation of a four-point bending test under plain stress conditions. This test utilizes the shape memory effect of SMA fibers located in the mechanical tension zone of an aluminum beam to counteract inelastic deformations. This example, not previously presented in the literature, reveals interesting interactions between the SMA fibers and the beam matrix. The system, considering symmetry with respect to the 2-axis, is depicted in Fig. 18. The beam has a total length of $2L = 1000$ mm, a

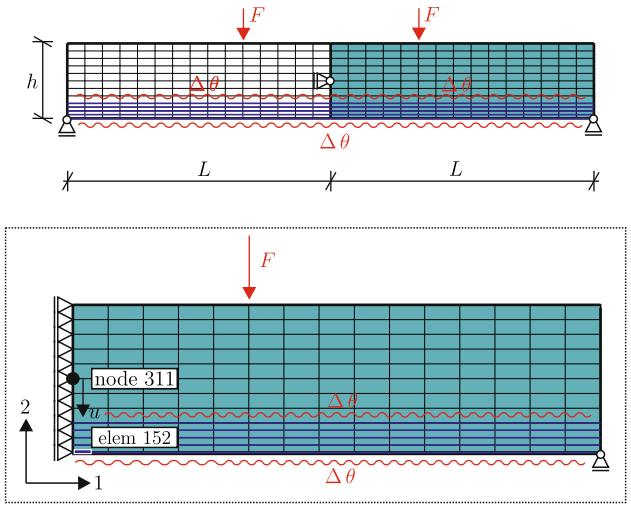

Fig. 18 Symmetrized system, boundary conditions and FE discretization of the four-point bending test. Five SMA-fibers are modeled in the tension zone of the beam. A purely mechanical loading and unloading cycle is followed by heating up and cooling down of the SMA fibers

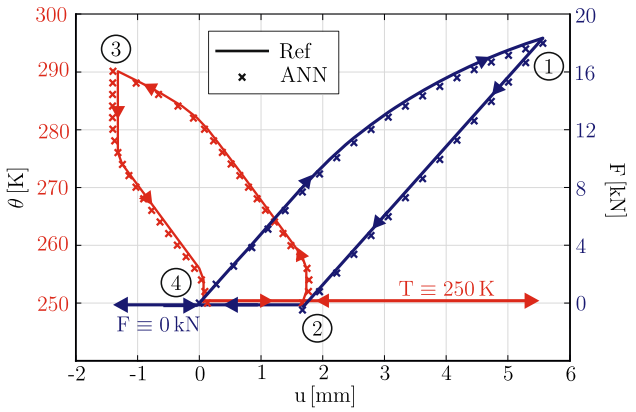

Fig. 19 Global behavior of the aluminum structure reinforced with SMA fibers: temperature θ and load F depending on the vertical displacement u. A purely mechanical loading and unloading process at constant temperature $\theta = 250$ K is followed by a purely thermal cycle at $F = 0$ kN. Inelastic deformations are reduced due to the shape memory effect. The ANN material model is able to describe the behavior similar to the reference solution

Table 4 Material parameters for the aluminum matrix material of the four-point bending test of Sect. 5.2

Parameter	Symbol	Unit	Value
Young's modulus	E	N/mm^2	70 000
Poisson's ratio	ν	–	0.34
Initial Yield strength	σ_y	N/mm^2	120
hardening modulus	H	N/mm^2	1000

height of $h = 140$ mm and a thickness of $b = 10$ mm. The matrix material is made of aluminum with the corresponding material parameters for the elastoplastic von Mises model provided in Table. 4. Five equidistantly spaced SMA fibers are located in the tension zone of the beam. The FE discretization of the matrix material in the symmetrized model consists of 15×10 nine node displacement based isoparametric finite elements, see also Fig. 18. The truss elements for the SMA fibers are orientated on the FE mesh of the matrix material. They are placed in the lower five node rows and match with the discretization of the matrix in 1-direction. The section area of the SMA fibers is $A = 20$ mm^2. The quantities of interest to be observed are the force F [kN], the mid beam displacement u [mm] at node 311, see Fig. 18, and the SMA fiber temperature θ [K]. Furthermore, the fiber stress σ_{152} [N/mm^2] and fiber strain ϵ_{152} of element 152 are observed, see Fig. 18.

The analytical model for SMA fibers of Sect. 2 along with the material parameters in Table. 1 are used to calculate the reference solution. To utilize the shape memory effect, the fibers are pre-stretched with $\varepsilon_0 = 0.05$, similar to the fibers in the grip element example of Sect. 5.1. For the ANN solution, a new ANN is trained. Apparently, the number of weights and the number of training paths provided in Sect. 4 are not sufficient for this example, where more complex mutual influences especially in the mechanical domain are present. This gets particularly evident, when observing local strains or stresses of the fibers. Therefore, an ANN with topology [5-40-30-20-1] is trained with the path generation strategy from Sect. 3.3. While the strain and temperature boundaries of Table. 2 are maintained, the number of steps in each interval is increased to $M = 60$. In total, 100 paths are generated like this. The calculation of incremental samples is additionally restricted to $\Delta\varepsilon = 0.01$. As in the previous examples, 20 000 incremental samples are randomly chosen as training data. Enforcement of the constraint is done similarly to the previous example. The training is truncated after 10 000 epochs and the weights of the last epoch are saved for the ANN material model. To numerically apply the pre-stretch of $\varepsilon_0 = 0.05$ in the ANN material model, the third input variable is adjusted as shown in Eq. (48). This pre-stretch is essential for initiating the subsequent shape memory effect.

The global behavior of the composite system is shown in Fig. 19. The SMA fibers initial temperature is $\theta = 250$ K. While maintaining a constant temperature, the mechanical load F is increased. This is done with an arclength method with a displacement control of the lower mid beam displacement u, see Fig. 18. Starting at $u = 0$ mm and $F = 0$ kN, 20 steps are performed, increasing the displacement to $u = 5.5$ mm. The SMA fibers act fully elastic, because the pre-stretch $\varepsilon_0 = \beta = 0.05$ simulates a previously performed pseudoplastic loading scenario, where oriented martensite has formed. Thus, they are already in the superelastic branch. The composite structure, which consists of the superelastic fibers and the elastoplastic aluminum, exhibits inelastic behavior starting from approximately $F = 10$ kN. While

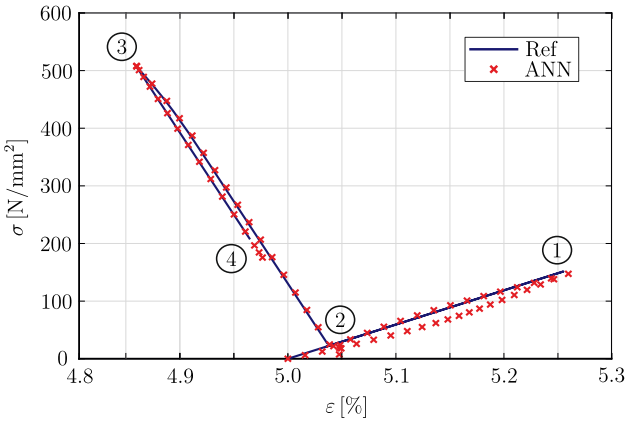

Fig. 20 SMA fiber stress–strain relationship of element 152 (highlighted in Fig. 18) during the mechanical and temperature dependent loading and unloading. The four distinct points refer to the same ones as indicated in Fig. 19

the aluminums hardening modulus is rather small, the elastic SMA fibers cause a further increase in force. The unloading process consists of 14 steps with increments of -0.275 mm. Both, the fibers and the aluminum act elastically. At a remanent displacement of $u = 1.65$ mm, the force F is approximately 0 kN. Now, the temperature is increased from $\theta = 250$ K to $\theta = 290$ K in 20 steps of 2 K and decreased to $\theta = 250$ with 20 steps of -2 K afterwards. This is done with Newton's method. During heating, oriented martensite is transformed to austenite, which comes along with a reduction in the SMA strains. The aluminum matrix restricts this, resulting in large compression forces in the previous beams tension zone and corresponding tension forces in the SMA fibers. This is similar to the two-way effect described before, leading to a backward deflection of the whole structure. As the temperature decreases, some of the austenite transforms back to twinned martensite. The corresponding start temperature is approximately $\theta = 275$ K and depends on the interaction between the aluminums resistance and the ratio of austenite and oriented martensite in the SMA fibers. This process stops at the austenit start temperature $\theta = 256.25$ K. As the temperature reaches the initial value of $\theta = 250$ K again, the global inelastic deflection is nearly gone. This higly depends on the applied mechanical and thermal loading and on the stiffnesses of the materials. As can be seen in Fig. 19, the ANN material model accurately predicts this complex behavior, although the stiffness of the composite structure is slightly underestimated, leading to larger displacements at specific load levels compared to the reference solution.

In Fig. 20, the stress–strain relationship of element 152 is shown. Element 152 corresponds to the bottom left SMA fiber in the center of the yield zone and is highlighted in Fig. 18. Due to the pre-stretch the process starts at $(\varepsilon, \sigma)_{152} = (0.05, 0)$. During the mechanical loading and unloading, the fiber stays in the superelastic region and behaves correspondingly. The relative error of the ANN approximation is comparatively high in this region. In fact, this region could only be approximated with the increased amount of training data. After unloading, there are still residual stresses in the system, i.e., in the matrix and also in the SMA fibers. When the temperature is increased, the two-way effect can be observed: starting at the austenite start temperature $\theta = 256.25$ K, the stresses increase heavily due to the restrained contraction of the fibers caused by the aluminum matrix. The subsequent decrease of the temperature leads to a delayed reduction of stresses starting at approximately 275 K. The significant temperatures are the same as described for the global structural behavior above. Worth mentioning is that the stress–strain points of the increasing and the decreasing temperatures are not on the same line. This results from the mechanical composite behavior, i.e., the change in the mechanical resistance of the elastoplastic matrix material depending on the loading direction.

In Fig. 19, four distinctive points during the mechanical and thermal loading and unloading processes are marked: (1) the point of maximal mechanical loading, (2) the point of complete mechanical unloading, (3) the point of maximal temperature, and (4) the point after final cooling down of the temperature. The fields of the von Mises equivalent stress

$$\sigma_e = \sqrt{\sigma_{11}^2 + \sigma_{22}^2 + 3\sigma_{12}^2}\,, \tag{51}$$

normalized with the initial yield stress $\sigma_y = 120$ N/mm^2 of the aluminum matrix, is shown in Fig. 21 for the mentioned points. They provide insight into the stress distribution within the composite structure throughout the different stages of loading and temperature change. The results obtained with the ANN material model for the SMA fibers are compared to the reference results, using the analytical SMA model. Qualitatively, the shapes of the individual fields are barely distinguishable, indicating an excellent approximation by the ANN model. It should be noted, that the aluminum matrix is calculated with an analytical material model and the corresponding results are influenced indirectly by the ANN SMA fiber model. At point 1, the constant plastic zone left of the load is visible, which is partially restored at point 2. When the maximum temperature is reached at point 3, a uniform plastic zone develops at the bottom of the beam due to the high compression forces induced into the matrix. As the temperature decreases again, a stronger remanent stress field can be observed compared to the situation after unloading. The reduction of the inelastic displacements, therefore, corresponds to higher residual stresses in the structure.

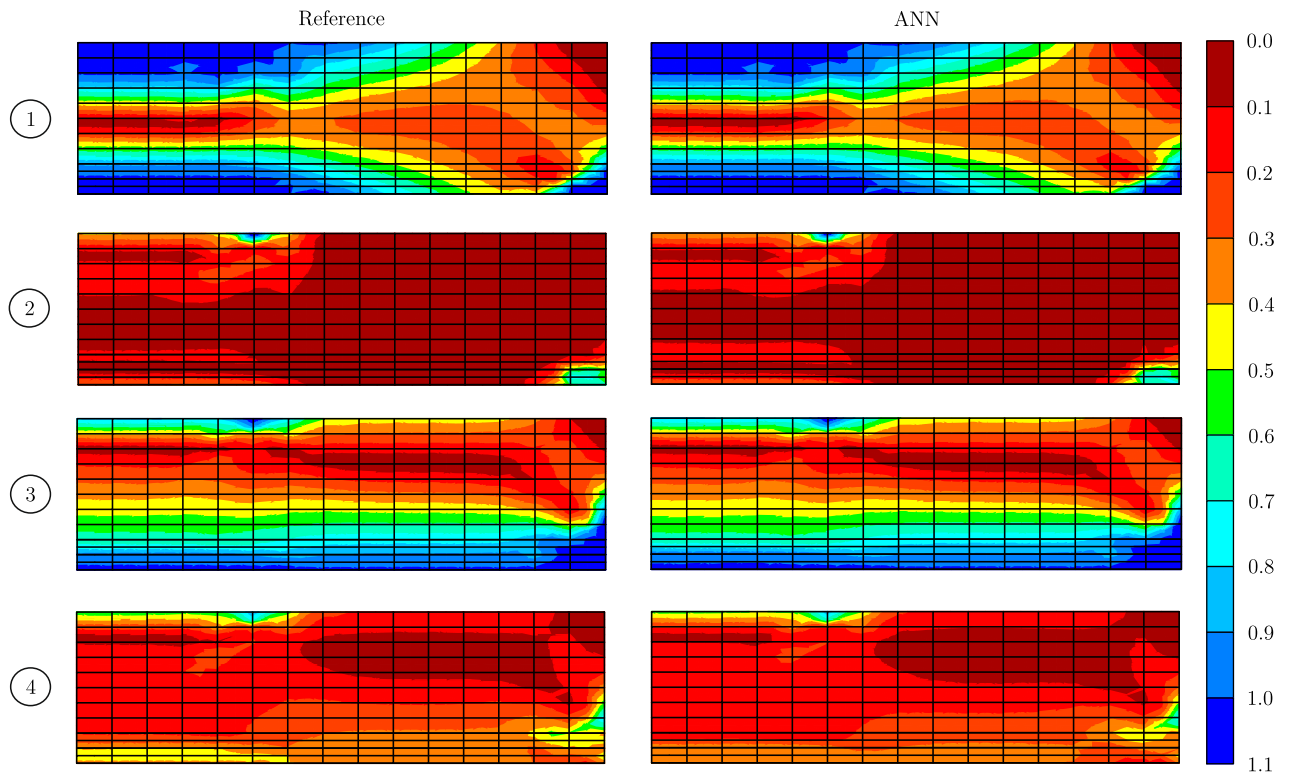

Fig. 21 Relative equivalent stresses σ_e/σ_y of the aluminum matrix. The ANN results are compared to the ones of the reference solution at four distinct points: 1) maximum load, 2) after unloading, 3) maximum temperature and 4) final configuration. The four points are also indicated in Fig. 19

6 Conclusion

The application of material modeling with artificial neural networks (ANN) has been presented for shape memory alloy (SMA) fibers. Although a lot of literature exists on this topic, an approach resulting in a material model applicable to arbitrary strains, stresses, and temperatures within finite element (FE) computations has been lacking. By properly defining the input and output vectors of a feedforward ANN and providing a sufficient amount of training data based on a given analytical SMA material model, it is possible to train a general-purpose one-dimensional ANN material model and perform materially and geometrically nonlinear FE analyses. The required numerical reliability is significantly improved by utilizing an inequality constraint of the material tangent incorporated in the ANN training process. The impact of this training enhancement is demonstrated on small examples. Furthermore, valuable insights on how to correctly evaluate the performance of ANN material models are provided in the presented context. In the subsequent FE examples, the capability of the introduced ANN SMA material model is demonstrated.

ANN material modeling applied to the multiphysical behavior of SMA offers a promising symbiosis between data-driven engineering and ANN material modeling. While this paper adds valuable first insights to the overall literature in this field, i.e. SMA modeling for arbitrary loading paths and the corresponding application to FE modeling, there are many potential future improvements to fully exploit its potential. Further investigations with real experimental data should be done in order to study how this or other ANN architectures perform in case of small sample sizes and data uncertainties. Furthermore, the restriction to rate independent behavior could be removed in the future, e.g., for structural damping applications. Last but not least, the concept of physically or numerically motivated restrictions should be expanded by the definition of additional constraints. This would equalize the downsides of the incremental model used, like the high amount of training data, without loosing its universality. Alternatively, methods of incorporating physics exactly by defining energy density functions and incorporating so-called Sobolev training are also recommended. It should be noted, that the conclusions regarding the training target values and the influence of the constraints to the numerical stability within incremental iteration schemes is also valid for other ANN architectures, i.e., not only for the feedforward ANNs used in this paper.

Acknowledgements The support of Maximilian Pawlik within the preparation of examples is greatfully acknowledged.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

1. Abedi H, Baghbaderani KS, Alafaghani A, Nematollahi M, Kordizadeh F, Attallah MM, Qattawi A, Elahinia M (2022) Neural network modeling of nitihf shape memory alloy transformation temperatures. J Mater Eng Performance 31(12):10258–10270. https://doi.org/10.1007/s11665-022-06995-y
2. Adarsh S, Sampath V (2020) Prediction of high temperature deformation characteristics of an fe-based shape memory alloy using constitutive and artificial neural network modelling. Mater Today Commun 22:100841. https://doi.org/10.1016/j.mtcomm.2019.100841
3. Auricchio F, Sacco E (1997) A one-dimensional model for superelastic shape-memory alloys with different elastic properties between austenite and martensite. Int J Non-Linear Mech 32(6):1101–1114. https://doi.org/10.1016/S0020-7462(96)00130-8
4. Auricchio F, Taylor RL, Lubliner J et al (1997) Shape-memory alloys: macromodelling and numerical simulations of the superelastic behavior. Comput Methods Appl Mech Eng 146(3–4):281–312
5. Auricchio F, Marfia S, Sacco E (2003) Modelling of sma materials: training and two way memory effects. Comput Struct 81(24):2301–2317. https://doi.org/10.1016/S0045-7949(03)00319-5
6. Brinson L (1993) One-dimensional constitutive behavior of shape memory alloys: thermomechanical derivation with non-constant material functions and redefined martensite internal variable. J Intell Mater Syst Struct 4(2):229–242. https://doi.org/10.1177/1045389x9300400213
7. Christ D (2009) Thermomechanical modelling of shape memory alloy structures in medical applications. PhD thesis, Technical University of Braunschweig
8. Christ D, Reese S (2009) A finite element model for shape memory alloys considering thermomechanical couplings at large strains. Int J Solids Struct 46(20):3694–3709. https://doi.org/10.1016/j.ijsolstr.2009.06.017
9. Furukawa T, Yagawa G (1998) Implicit constitutive modelling for viscoplasticity using neural networks. Int J Numerical Methods Eng 43(2):195–219
10. Ghaboussi J, Sidarta DE (1998) New nested adaptive neural networks (nann) for constitutive modeling. Comput Geotech 22(1):29–52. https://doi.org/10.1016/S0266-352X(97)00034-7
11. Ghaboussi J, Garrett J, Wu X (1991) Knowledge-based modeling of material behavior with neural networks. J Eng Mech 117(1):132–153. https://doi.org/10.1061/(asce)0733-9399(1991)117:1(132)
12. Haupt P (2002) Continuum Mechanics and Theory of Materials. Springer, Berlin Heidelberg,. https://doi.org/10.1007/978-3-662-04775-0
13. Helm D (2007) Thermomechanics of martensitic phase transitions in shape memory alloys, i: constitutive theories for small and large deformations. J Mech Mater Struct 2(1):87–112. https://doi.org/10.2140/jomms.2007.2.87
14. Hmede R (2023) Modelling and control by artificial intelligence of a shape memory alloy actuator for neurosurgical environment. PhD thesis, Université Clermont Auvergne
15. Kaup A, Ding H, Wang J, Altay O (2020) Strain rate dependent formulation of the latent heat evolution of superelastic shape memory alloy wires incorporated in multistory frame structures. J Intell Mater Syst Struct 32(11):1198–1214. https://doi.org/10.1177/1045389x20975473
16. Kaup A, Altay O, Klinkel S (2021) Strain amplitude effects on the seismic performance of dampers utilizing shape memory alloy wires. Eng Struct 244:112708. https://doi.org/10.1016/j.engstruct.2021.112708
17. Kaup A, Lenzen N, Altay O (2021) Neural network parameter identification based constitutive modeling of superelastic shape memory alloys. PAMM. https://doi.org/10.1002/pamm.202100251
18. Klein DK, Fernández M, Martin RJ, Neff P, Weeger O (2022) Polyconvex anisotropic hyperelasticity with neural networks. J Mech Phys Solids 159:104703. https://doi.org/10.1016/j.jmps.2021.104703
19. Klein DK, Roth FJ, Valizadeh I, Weeger O (2023) Parametrized polyconvex hyperelasticity with physics-augmented neural networks. Data-Centric Eng. https://doi.org/10.1017/dce.2023.21
20. Klinkel S, Gruttmann F, Wagner W (2006) A robust non-linear solid shell element based on a mixed variational formulation. Comput Methods Appl Mech Eng 195(1–3):179–201. https://doi.org/10.1016/j.cma.2005.01.013
21. Kohlhaas B, Klinkel S (2014) An FE^2 model for the analysis of shape memory alloy fiber-composites. Computational Mech 55(2):421–437. https://doi.org/10.1007/s00466-014-1112-3
22. Kohlhaas B, Klinkel S et al (2012) Constitutive modeling of prestressed shape memory fiber-matrix compounds. Blucher Mech Eng Proc 1(1):1139–1154
23. Lagoudas DC (2008) Shape memory alloys: modeling and engineering applications. Springer
24. Lecce L et al (2014) Shape memory alloy engineering: for aerospace, structural and biomedical applications. Elsevier
25. Lenzen N, Altay O (2022) Machine learning enhanced dynamic response modelling of superelastic shape memory alloy wires. Materials 15(1):304. https://doi.org/10.3390/ma15010304
26. Ma N, Song G, Lee HJ (2004) Position control of shape memory alloy actuators with internal electrical resistance feedback using neural networks. Smart Mater Struct 13(4):777–783. https://doi.org/10.1088/0964-1726/13/4/015
27. Machado L, Savi M (2003) Medical applications of shape memory alloys. Brazilian J Med Biol Res 36:683–691
28. Masi F, Stefanou I (2022) Multiscale modeling of inelastic materials with thermodynamics-based artificial neural networks (tann). Comput Methods Appl Mech Eng 398:115190. https://doi.org/10.1016/j.cma.2022.115190
29. Masi F, Stefanou I, Vannucci P, Maffi-Berthier V (2021) Thermodynamics-based artificial neural networks for constitutive modeling. J Mech Phys Solids 147:104277
30. Otsuka K, Wayman CM (1999) Shape memory materials. Cambridge University Press
31. Owusu-Danquah JS, Bseiso A, Allena S (2021) Artificial neural network models to predict the response of 55niti shape memory alloy under stress and thermal cycles. Neural Comput Appl 34(5):3829–3842. https://doi.org/10.1007/s00521-021-06643-x

32. Ozbulut OE, Hurlebaus S, Desroches R (2011) Seismic response control using shape memory alloys: a review. J Intell Mater Syst Struct 22(14):1531–1549. https://doi.org/10.1177/1045389x11411220
33. Petrini L, Migliavacca F (2011) Biomedical applications of shape memory alloys. J Metall 2011:1–15. https://doi.org/10.1155/2011/501483
34. Rodrigue H, Wang W, Han MW, Kim TJ, Ahn SH (2017) An overview of shape memory alloy-coupled actuators and robots. Soft Robotics 4(1):3–15. https://doi.org/10.1089/soro.2016.0008
35. Rosenkranz M, Kalina KA, Brummund J, Kästner M (2023) A comparative study on different neural network architectures to model inelasticity. Int J Numerical Methods Eng 124(21):4802–4840. https://doi.org/10.1002/nme.7319
36. Seelecke S, Müller I (2004) Shape memory alloy actuators in smart structures: modeling and simulation. Appl Mech Rev 57(1):23–46. https://doi.org/10.1115/1.1584064
37. Shu X, Lu S, Wang K, Li G (2015) A comparative study on constitutive equations and artificial neural network model to predict high-temperature deformation behavior in nitinol 60 shape memory alloy. J Mater Res 30(12):1988–1998. https://doi.org/10.1557/jmr.2015.144
38. Song G, Chaudhry V, Batur C (2003) A neural network inverse model for a shape memory alloy wire actuator. J Intell Mater Syst Struct 14(6):371–377. https://doi.org/10.1177/1045389x03034628
39. Song G, Ma N, Li HN (2006) Applications of shape memory alloys in civil structures. Eng Struct 28(9):1266–1274. https://doi.org/10.1016/j.engstruct.2005.12.010
40. Taylor RL (2024) FEAP - finite element analysis program. http://projects.ce.berkeley.edu/feap/
41. The MathWorks Inc (2021) Matlab version: 9.11.0 (r2021b). https://www.mathworks.com
42. Viet N, Zaki W (2023) Artificial neural network model of the mechanical behaviour of shape memory alloy schwartz primitive lattice architectures. Mech Mater 183:104680. https://doi.org/10.1016/j.mechmat.2023.104680
43. Wang F (2023) Neural network model for hysteretic characteristic of shape memory alloy. Mater Today Commun 35:105963. https://doi.org/10.1016/j.mtcomm.2023.105963
44. Wang H, Song G (2014) Innovative narx recurrent neural network model for ultra-thin shape memory alloy wire. Neurocomputing 134:289–295. https://doi.org/10.1016/j.neucom.2013.09.050
45. Weber P (2023) Physikalisch konsistente Modellierung von elastischem und plastischem Materialverhalten mit künstlichen neuronalen Netzen. PhD thesis, Karlsruhe Institute of Technology
46. Weber P, Geiger J, Wagner W (2021) Constrained neural network training and its application to hyperelastic material modeling. Computational Mech 68(5):1179–1204. https://doi.org/10.1007/s00466-021-02064-8
47. Weber P, Wagner W, Freitag S (2023) Physically enhanced training for modeling rate-independent plasticity with feedforward neural networks. Computational Mech 72:827–857. https://doi.org/10.1007/s00466-023-02316-9
48. Wolfe P (1969) Convergence conditions for ascent methods. SIAM Rev 11(2):226–235
49. Wolfe P (1971) Convergence conditions for ascent methods. II: Some corrections. SIAM Review 13(2):185–188
50. Wriggers P (2010) Nonlinear Finite Element Methods. Springer, Berlin Heidelberg
51. Xu K, Huang DZ, Darve E (2021) Learning constitutive relations using symmetric positive definite neural networks. J Computational Phys 428:110072. https://doi.org/10.1016/j.jcp.2020.110072
52. Zhan M, Liu J, Wang D, Chen X, Zhang L, Wang S (2021) Optimized neural network prediction model of shape memory alloy and its application for structural vibration control. Materials 14(21):6593. https://doi.org/10.3390/ma14216593