REVIEW ARTICLE

Neural Networks for Constitutive Modeling: From Universal Function Approximators to Advanced Models and the Integration of Physics

Johannes Dornheim[1] · Lukas Morand[2] · Hemanth Janarthanam Nallani[2] · Dirk Helm[2]

Abstract

Analyzing and modeling the constitutive behavior of materials is a core area in materials sciences and a prerequisite for conducting numerical simulations in which the material behavior plays a central role. Constitutive models have been developed since the beginning of the 19th century and are still under constant development. Besides physics-motivated and phenomenological models, during the last decades, the field of constitutive modeling was enriched by the development of machine learning-based constitutive models, especially by using neural networks. The latter is the focus of the present review paper, which aims to give an overview of neural networks-based constitutive models from a methodical perspective. The review summarizes and compares numerous conceptually different neural networks-based approaches for constitutive modeling including neural networks used as universal function approximators, advanced neural network models and neural network approaches with integrated physical knowledge. The upcoming of these methods is in-turn closely related to advances in the area of computer sciences, what further adds a chronological aspect to this review. We conclude the review paper with important challenges in the field of learning constitutive relations that need to be tackled in the near future.

1 Introduction and Background

1.1 Motivation

In continuum mechanics, we distinguish between universal relations (e.g. balance laws) and constitutive relations that describe the behavior of a specific class of materials. Using the theory of materials (see for example [1]), useful principles can be formulated that form the basis for development of these constitutive relations. Within this framework, constitutive models are developed based on phenomenological observations and/or physical knowledge, often related to physically motivated state variables like dislocation density in metallic materials. In addition, constitutive relations can be formulated in accordance to the second law of thermodynamics which enforces thermodynamically consistent models. Independent of the modeling approach, material models contain material dependent parameters, which are to be calibrated using experiments, potentially accompanied by numerical simulations like virtual testing techniques [2]. Typically, the advantage of physically motivated models are the reduced number of material dependent parameters that need to be determined from experiments.

Over the past few decades, a new approach based on supervised machine learning techniques for modeling nonlinear material behavior has emerged, specifically using neural networks. In general, compared to complex physics-based models, machine learning-based models can be executed in near real time, which enables extensive accelerated numerical simulations. This in-turn can lead to a breakthrough in engineering, as it allows for more accurate and detailed numerical simulations on component and process level. Neural networks-based approaches offer a significant advantage over other supervised learning models by virtue of their ability to represent any continuous functional relation [3, 4] and thus model arbitrary complex material behavior. However,

✉ Johannes Dornheim
johannes.dornheim@iav.de

✉ Lukas Morand
lukas.morand@iwm.fraunhofer.de

Hemanth Janarthanam Nallani
hemanth.janarthanam.nallani@iwm.fraunhofer.de

Dirk Helm
dirk.helm@iwm.fraunhofer.de

1   Institute for Applied Materials, Karlsruhe Institute of Technology, Karlsruhe, Germany

2   Fraunhofer Institute for Mechanics of Materials IWM, Freiburg, Germany

the large number of parameters that need to be calibrated during the neural network training is a shortcoming.

Besides constitutive modeling, the potential of using machine learning models is actively being investigated across multiple fields of computational material science. Among others, these methods are used in computational design of materials [5–9], in design of processes [10–12], in development of digital twins [13] and soft sensors [14], and in multi-scale simulations and homogenization schemes. Numerical prediction in these applications often involve frequent execution of simulations with variations of model parameters [15]. The central limiting factors of classical modeling approach are the complexity of constitutive models and computational performance of the underlying simulations. In order to improve the computational performance, surrogate models (i.e. learned simulations or constitutive models) can be set up on various levels.

Surrogate neural networks are often proposed at structural/component level [16–19], at continuum level across multiple length scales [15] to replace conventional constitutive models, or at atomic and molecular level [20]. As opposed to structural surrogates, learned constitutive models, once trained, are in principle applicable across different structures. Methods to train such constitutive neural network models are proposed by various authors. We outline and classify these training approaches in Sect. 1.2.

The idea of using neural networks to learn constitutive relations dates back to the pioneering work of Ghaboussi et al. [21] in the early 90s. A first short but general review summarizing the use of neural networks in computational mechanics, including constitutive modeling, was published in 1996 by Yagawa and Okuda [22]. Due to the recent heightened interest in this research field, a variety of review and survey papers have been published. Notably reviews focusing on the use of machine learning methods for materials discovery and design [23, 24], for engineering of materials, processes, and structures [25], for multi-scale modeling [26] and for meta materials design [27] have been presented. A broad review on the application of machine learning methods in continuum mechanics has been put forth by Bock et al. [28]. Pertaining to constitutive models, reviews focused on specific material classes like composites [29], soils [30], alloys [31], and sheet metals specific to forming processes [32].

This contribution, on the contrary, is a general characterization and classification of neural network methods for learning constitutive behavior. This review also has a chronological aspect, as the development of these methods evolved with advancements in the field of machine learning, such as the development of convolutional neural networks (CNNs) and recurrent neural networks (RNNs), see [33] for both, as well as physics-informed neural networks (PINNs), see [34] (although such approaches have already been

proposed in the late 90s, for example in [35]). The purpose of PINNs in [34] however, is basically to solve differential equations on a domain for given boundary values, which is primarily not directly transferable for learning constitutive relations. However, the combination of the physics-informed part of PINNs (which can be understood as an encoding of a physical law described by a differential equation) and the neural network part (which predicts the quantities of interest being fully differentiable with respect to all input variables by using automatic differentiation [36]), can be adjusted and leveraged for learning constitutive relations.

The field of research on the application of machine learning to constitutive modeling is very active and constantly expanding. This is reflected in the increase of annually published papers as shown in Fig. 1. As novel papers are published almost on a weekly basis, we limit this review to works that have been published before 2022 with a few important contributions from 2022. From the methodic viewpoint, we focus solely on approaches that utilize neural networks for learning constitutive relations. Methods beyond our scope are structure-level surrogates [37–40], data-driven solvers [41–45], constitutive model free approaches [46–48], and constitutive modeling using symbolic machine learning [49], random forests [50] and spline interpolation [51–53] methods.

1.2 Classification of Neural Networks-Based Constitutive Modeling Approaches

The numerous neural networks-based constitutive modeling approaches proposed in literature can be distributed along
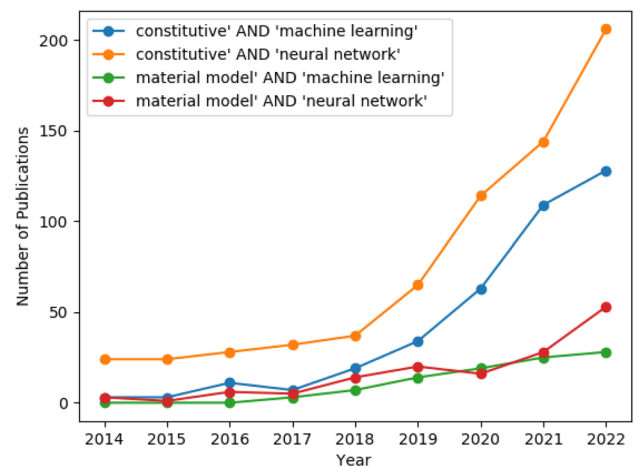
Fig. 1 Number of annual publications (articles, preprints, chapters or proceedings) that contain the combination of the following keywords within the title or the abstract: constitutive and machine learning, constitutive and neural network, material model and neural network, material model and machine learning. The data is gathered from app. dimensions.ai

an axis, as depicted in Fig. 2, from purely physics-based to universal function approximation models. The classical feedforward neural networks (FFNNs) is placed at one extremum of the spectrum. FFNNs can be used to learn constitutive models purely from data without any knowledge of the underlying physical relations. Using the universal function approximation theorem [3, 4], it has been shown that, given a sufficient amount of model parameters, FFNNs can in principle be parameterized to represent any continuous functional relation. In this regard, neural networks can be used to learn functional relations that typically originate from material model formulations. However, the need for a large set of labeled data and the lack of interpretability are few of the drawbacks of FFNNs in its basic form. Other types of neural networks fill the gap between the purely data-driven neural network models and the conventional physical constitutive modeling approaches. As depicted in Fig. 2, we associate five general characteristics to the methods with in the spectrum.

1. The basic requirement for building a physical constitutive model is sufficient domain knowledge, more specifically, physical knowledge about the material behavior.
2. In contrast, for a data-driven modeling approach there is a need for data, covering the full spectrum of the material behavior.
3. While, due to imperfections and simplifying assumptions, physical models are usually biased, trained FFNNs, under the assumption of sufficient data and enough model capacity, are unbiased. A universal estimation of the true material behavior is possible.
4. Purely data-driven models learn by minimizing the difference between predictions and ground-truth data and
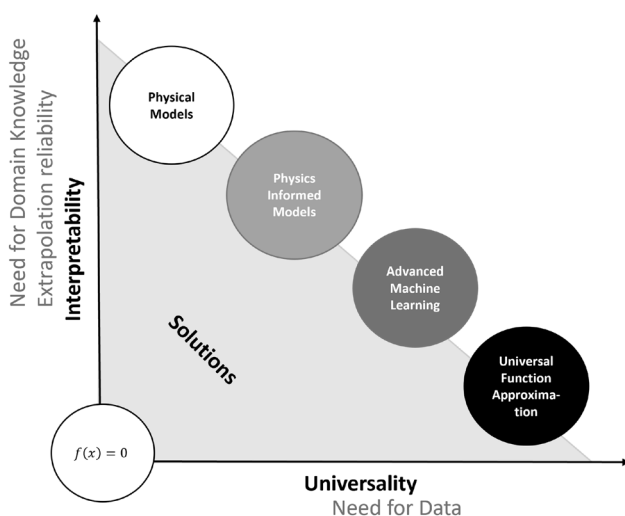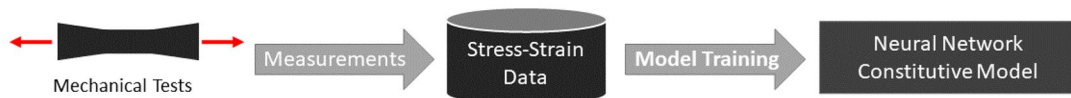
usually have no incentive to extrapolate beyond the ground truth. Combined with the non-linearity of such models, this typically leads to a highly non-linear model behavior outside of the training region. The incorporation of physical knowledge can be used to constrain the model and thereby improves model robustness, especially when extrapolating.

5. Another important aspect is the interpretability of the model. Machine learning models for universal approximation of any non-trivial function involve a large set of parameters with no physical meaning and complicated influence on the approximations. Such models have to be seen as black boxes. Physical models, on the other hand, are characterized by the fact that the parameters and their impact are well-defined.

There are advanced techniques available in neural networks-based constitutive modeling that surpass the use of simple FFNNs. In the following, we distinguish between two classes of approaches that incorporate domain knowledge into the neural networks to improve data efficiency, explainability, and extrapolation capabilities of the learned constitutive model. The first class of approach is based on the use of certain pre-structured layers and advanced neural network architectures to integrate knowledge about the temporal or geometric structure of the problem and data at hand. This includes the use of recurrent and time-convolutional neural network models to incorporate knowledge about the temporal structure within the data and the use of graph convolution and euclidean convolution to incorporate knowledge about spatial dependencies in the data. The second class includes methods to directly incorporate physical knowledge into the neural network in the form of differentiable layers that avoid violating physics constraints and thereby enable extrapolation in certain limits.

Based on the application mode of neural networks-based constitutive modeling, two distinct use cases of learning constitutive relations with neural networks can be categorized, as illustrated in Fig. 3. These are:

(1) Direct learning is attributed to neural networks, which are trained directly on stress–strain data. Direct learning can be subdivided further into learning from experimental data and learning from synthetic data (constitutive surrogate modeling):

   (a) Direct learning from global experimental stress–strain data offers means to establish unidentified constitutive relations directly from the data. This approach is particularly valuable when insufficient knowledge about the specific material is available or when purely physical constitutive models are challenging to acquire. This approach has a long

Fig. 2 Methodical spectrum of named constitutive modeling approaches and an outline of the central trade off of these models

(1) Direct learning

(a) Direct learning from experiments

(b) Surrogate constitutive NN training
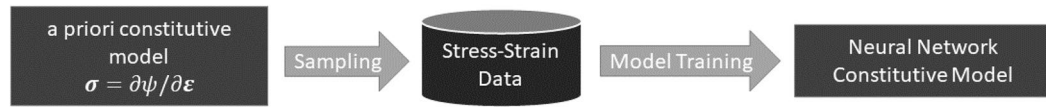
(2) Indirect learning
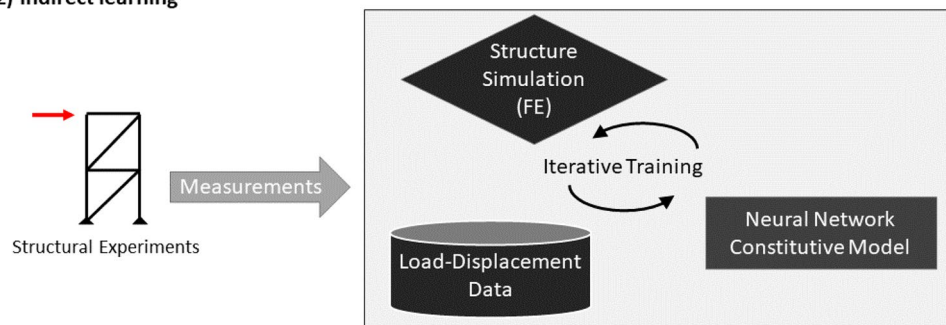
Fig. 3 Applications of neural networks in constitutive modeling: (1) direct learning of constitutive neural networks based on **a** experimental stress–strain data and **b** a priori constitutive models, (2) indirect learning of constitutive neural networks in an iterative training scheme

standing tradition and has been widely investigated since the early 90s [21]. However, a crucial prerequisite for direct learning is the availability of an adequate quantity of stress–strain data, commonly acquired from simple mechanical tests that yield one data sample per test run. In addition, these tests often assume stress–strain uniformity along a single direction, thereby limiting the data generation to one-dimensional measurements [54].

(b) Learning constitutive surrogates is based on data sampled from reference classical constitutive models and aims to replace them to facilitate real time computations. Both direct learning from global experimental data and the learning of constitutive surrogates can be methodically categorized as standard supervised regression problems. However, the underlying motivation for these scenarios differ in that the former seeks to establish a constitutive relation, while the latter aims to replace a pre-existing, a priori given model in a computationally efficient framework. In addition, a neural network is classified as a constitutive surrogate if it is a function of a history variable that can be obtained only using a classical constitutive

relation, and not from experiments. Although this method's ability to uncover novel material mechanisms is constrained by the physics-based model used to generate the training data at a particular length scale. Utilizing this approach effectively to bridge multiple length scales holds the potential to discover mechanisms at a higher scale.

(2) To leverage the rich local constitutive information from structural tests, indirect learning utilizes numerical simulations in conjunction with experimental observation to train the constitutive neural network, in often an incremental manner. To facilitate the learning process, an error measure is established by comparing simulation results with experimental data. Gathering experimental data is, however, challenging and time-consuming, leading to a prevalent reliance on simulations alone as proof-of-concept in literature.

1.3 Paper Structure

The paper is structured as follows. In the remainder of this section we give a brief overview of machine learning and materials modeling background, which is assumed in the following sections. In Sects. 2 to 5, we introduce and discuss

the reviewed work. Wherein, we categorized the work according to the methodical spectrum from Fig. 2, where applications of classical FFNNs (category universal function approximation in Fig. 2) are reviewed in Sect. 2. In Sects. 3 and 4, we discuss works that use advanced machine learning techniques that consider load history information and spatial information respectively. In Sect. 5, we review work on physics-informed neural networks. Finally, in Sect. 6, we give a summary, briefly discuss the current state of the field and give an outlook.

1.4 Aspects of Machine Learning and Data Science

In the following, we briefly describe the types of neural networks and dimensionality reduction techniques that are most important to methodically contextualize the literature reviewed throughout this paper. We refer to [33] for a more comprehensive and in-depth overview of the field of deep learning.

1.4.1 Neural Networks-Based Function Approximation

In this Section, we give an overview of basic neural network types used for constitutive modeling in the reviewed publications, including FFNNs, RNNs and CNNs. The outlined neural network types are depicted in Fig. 4. FFNNs are general mappings

$$\hat{y} = \tilde{f}(x, \theta) \tag{1}$$

with model parameters θ, which can be trained for non-linear approximation of typically real-valued functions

$$y = f(x), \tag{2}$$

with $x \in \mathbb{R}^m$ and $y, \hat{y} \in \mathbb{R}^n$. Gradient-based methods and the backpropagation algorithms are used to fit the parameters $\theta \in \mathbb{R}^l$ from sample data to approximate f.

An FFNN consists of several consecutive neural layers

$$\bar{y} = \tilde{f}^{(i)}(\bar{x}, \theta_i) \tag{3}$$

Fig. 4 Basic architectures for the neural network types discussed throughout the paper. **a** While the connections in FFNNs form an acyclic graph, **b** RNNs contain cyclic connections (depicted in blue color), which cause the output of the respective neurons affecting the subsequent input of the same neurons. **c** CNNs are designed to process structured data, like images, time series or graphs

(a)

(b)

(c)
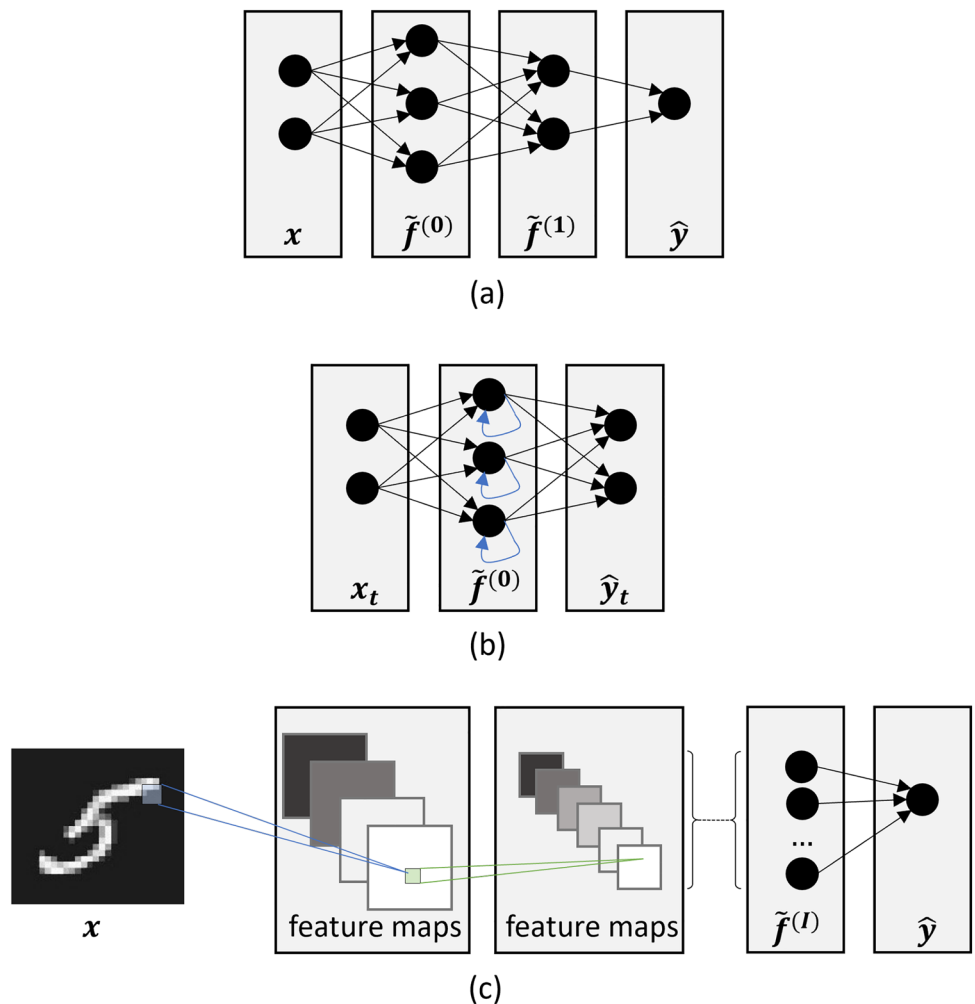
with parameters θ_i. The information flow is directed from the input layer $\tilde{f}^{(0)}(\boldsymbol{x}, \theta_0)$ through the hidden layers to the output layer $\tilde{f}^{(I)}$. The FFNN can then be seen as a nested function

$$\tilde{f}(\boldsymbol{x}, \theta) = \tilde{f}^{(I)}(\tilde{f}^{(I-1)}...(\tilde{f}^{(0)}(\boldsymbol{x}, \theta_0))) \tag{4}$$

with $\theta_i \subset \theta$ for $i \in [0, ..., I]$. In the case of FFNNs, each layer $f^{(i)}$ consists of a set of processing units, so-called neurons, with a subset of θ_i as parameters. The amount of neurons per layer is referred to as the width of the layer, while the amount I of layers in a neural network is referred to as its depth. The training of neural networks is typically conducted by mini-batch-wise application of a gradient descent-based algorithm [55].

While FFNNs are generally applicable, more advanced neural network models take advantage of the data structure at hand to improve accuracy and data efficiency. Three important classes of such structured data are sequential data, image data, and graph data. Pertaining to sequential input data (like for predicting nonlinear microstructure evolution [56]), which is often present in materials modeling as the material behavior is typically history-dependent, recurrent or time-convolution neural networks can be used. RNNs [57, 58] have special feedback connections, which makes them differ from the hierarchical structure of an FFNN. While the input layer of an RNN takes data from the current time-step, the recurrent connections enable information flow in between the time steps, and thereby the current output $\hat{\boldsymbol{y}}_t$ depends not only on the current input $\boldsymbol{x}_t$, but also on all previous inputs $\boldsymbol{x}_{t-i}$ for $i \in [0, t]$.

For long sequences, these dependencies often lead to problems with the flow of the gradients through the model. During backpropagation, the gradients flow from the output neurons, where the loss is calculated based on $\hat{\boldsymbol{y}}_t$, through the unfolded recurrent connections to the input neurons. For longer sequences, this chain of calculations gets bigger and the gradients tend to vanish (tend to zero) or to explode (tend to infinity). While in the first case, the network cannot learn long-term dependencies, in the second case, the learning stability is often diminished. Due to this problem, initially proposed RNNs are considered impractical for complex tasks involving sequential data. The long short-term memory network (LSTM) [59] is an attempt to solve the vanishing gradient problem based on learned input and output gates, which guard the flow of information. The gated recurrent unit (GRU) [60] is a simplified LSTM variant with a reduced amount of parameters. Both, GRUs and LSTMs, have been shown to often perform better than conventional RNNs in the case of small datasets, while there is no clear superiority of GRUs and LSTMs over the other [61].

Another important type of deep neural network model is the CNN [62, 63]. In addition to the conventional fully connected layers, CNN models are composed of a range of specialized layers, with the convolutional layers being the most significant among them. In these layers, convolution kernels with trainable weight parameters efficiently learn abstractions and features of the structured input. Especially in the area of computer vision, CNNs led to radical progress in terms of accuracy and sample efficiency. In materials modeling, CNN models are often used in the context of multi-scale modeling, where image or graph data on the crystallographic scale are to be processed. Besides images, which are processed by two-dimensional convolution operations, one-dimensional convolution layers can be applied to sequential data [64, 65]. Graph convolution, as a generalization of the classical convolution, enables the application of convolutional layers on data that have no regular (Euclidean) structure but are instead structured in a more general form of a graph, see [66] for an example in materials science.

1.4.2 Dimensionality Reduction, Representations and Symmetries

When working with results from experiments and simulations, often a huge amount of data is extracted, in which the relevant information is covered by a large amount of highly intercorrelating features. Moreover, material data, for example crystallographic texture data, often hold certain symmetries. In the following, we give a brief overview of methods to deal with these characteristics of data and to reduce the data dimensionality.

Dimensionality reduction techniques aim to embed high-dimensional data into a lower-dimensional embedding space, which is easier to process and interpret while preserving as much information as possible. The most prominent dimensionality reduction method is the principal component analysis (PCA), which projects the data linearly into a new coordinate system of lower dimension while preserving most of the variation within the data. In contrast, non-linear dimensionality reduction such as manifold learning methods (e.g. Isomap [67], multidimensional scaling [68] or local linear embedding [69]) are based on the assumption that the dimensionality of datasets is artificially inflated. Therefore, the data is extracted from a low dimensional manifold that is integrated within the high dimensional space of the dataset [70].

A prominent approach to reduce the dimensionality of a dataset nowadays is to use so-called autoencoder neural networks [71]. This method has already shown to be useful for compressing microstructure information, see [8, 72]. Such networks consist of an encoder and a decoder part. The encoder in-turn consists of fully connected layers only or a combination of convolutional layers with pooling and fully connected layers, depending on the input representation. In both cases, the output of each layer is usually of lower dimension than the input, and the aim of the encoder is to

embed the data in a low-dimensional latent feature space, which is represented by a neural network layer, called the bottleneck layer. The decoder part consists again of fully connected layers, which can be combined with deconvolutional layers and map the bottleneck layer to the network's output layer, which is of the same dimension as the encoder input. The encoder and decoder parts are trained simultaneously, by a loss that punishes the dissimilarity of the encoder input and the decoder output and thereby drives the autoencoder to encode as much of the data information as possible into the low dimensional bottleneck layer.

As in other domains, such as image processing, data in material sciences often holds symmetries. Mathematical operations on the data are often equivariant or invariant with respect to certain transformations. The efficiency of machine learning methods often relies on how these invariances and equivariances are dealt with. The success of CNNs on image data, for example, depends on the consideration of shift-invariance or equivariance by the application of a sliding convolution kernel, where the convolution parameters are learned independently of the location of the kernel. Recent research on deep learning through the lens of geometric symmetries can be found in the very active field of geometric deep learning [73]. Another approach to consider the assumed symmetries within the data is the augmentation of training data, e.g. by duplicating samples and manually manipulating the input, which is state-of-the-art in machine learning.

1.5 Aspects of Materials Modeling

For a prolonged period, materials science has been dominated by empirical, model-based theoretical, and computational research. However, in the last two decades, it has been extended to incorporate (big) data-driven science [74]. For now, material models have been developed mainly based on empirical investigations and theories derived from physics relations. In the second half of the last century, the field of computational mechanics evolved, which enabled model-based evaluations for example by conducting simulations on different length scales, [75].

On a continuum scale, following the theory of continuum mechanics and thermodynamics, universal balance equations exist (the balance of mass, linear momentum, angular momentum, energy, and entropy), which form the basis to calculate unknown mechanical field quantities (for motion, temperature, etc.). However, for solving mechanical problems, these equations are underdetermined. To determine all unknowns, material-related constitutive relations (e.g. stress–strain relationships) are needed. These constitutive relations are not universal and have to be formulated depending on the material at hand and its individual deformation behavior (see for example [76] for an introduction to material behavior and deformation mechanisms).

In classic continuum mechanics, different theories of materials behavior have been developed that describe elastic and viscous fluids as well as elastic, visco-elastic, plastic and visco-plastic solids, see [1]. A fundamental characteristic of elastic material behavior is that the stress state depends only on the actual deformation state. Inelastic material behavior, in contrast, depends on the entire deformation history. In such material models, the deformation history is typically represented by internal variables, such as the accumulated plastic strain.

For developing constitutive models, several approaches exist. Driven by experimental investigations, phenomenological material models describe phenomenons observed in experiments, such as the Hall–Petch relation describing the grain size dependent plastic yielding [77], the Armstrong-Frederick model describing the Bauschinger effect in plasticity, and rheological models describing linear visco-elastic material behavior on the macroscopic scale [1].

Material models on different length scales can also be formulated in the framework of thermodynamics principles. In such models, the modeling perspective focuses on energy storage and release, as well as dissipation effects. Therefore, thermodynamic potentials must be introduced based on the physical understanding of the underlying mechanisms and experimental observations. In addition, dissipation phenomena are often addressed by using internal variables theory. If the second law of thermodynamics is fulfilled, the model is called thermodynamically consistent. Basic approaches for developing thermodynamically consistent material models can be found in [1].

As an alternative to formulating macroscopic material models from scratch and calibrating them to experimental measurements, the material microstructure and its behavior can be modeled directly in a so-called representative volume element (RVE). By using this technique, the initial boundary value problem for representing the microstructural behavior is typically solved numerically. By virtual testing procedures, the deformation behavior can be analyzed on the basis of the Finite Element Method (FEM) [78] or fast Fourier transform (FFT) [79]. The advantage of using RVEs is that arbitrary load cases can be applied, which are often hard to apply in real experiments, [80–82]. The approach to calibrate macroscopic material models to RVE data can be called virtual laboratory [2]. The virtual laboratory can also be used to model scale transitions in whole process chains, as is shown in [83] on the example of dual-phase steel production.

Furthermore, RVE-based models can be used directly in Finite Element (FE) component simulations, which is typically called FE^2 [84, 85]. FE^2 corresponds to the case where on both scales, the micro and the macro-scale, coupled FE

simulations are performed to include the microstructural behavior in the simulated component. This in-turn allows for microstructure-driven design and optimization of manufacturing processes and components, [86]. However, FE^2 is usually too time-consuming for the use in engineering. To overcome this issue, machine learning techniques can be applied, as will be presented in the following sections of this review. Instead of using FEM to determine the stress response of an RVE under external load, for example, supervised learning can be used to learn the stress responses for given deformation paths [87]. The learned model forms an RVE surrogate model acting on the macroscopic scale, which is, in contrast to phenomenological models, sensitive to the underlying microstructure.

2 Universal Function Approximators: Fully Connected Neural Networks for Learning Constitutive Relations

In this section, we discuss work that uses fully connected neural networks to learn constitutive models, so-called constitutive neural networks (also called neural constitutive laws [88] or neural network constitutive models [89]). The section is structured into two subsections according to the application cases mentioned in Sect. 1.2. The use of FFNNs to directly learn the constitutive behavior of materials from global experimental stress–strain data as well as from synthetic simulation data is presented in Sect. 2.1, and work on indirect training schemes with the goal to learn constitutive neural networks from structural experiments is introduced in Sect. 2.2. The presented approaches are discussed briefly in Sect. 2.3.

2.1 Direct Learning

Ghaboussi et al. proposed the use of FFNNs for constitutive modeling in a series of early works [21, 90]. In [21], the application of FFNNs on experimental data of concrete is proposed. The experimental datasets of biaxial monotonic loading from [92] and uniaxial cyclic loading from [93] is used for training and evaluation of FFNNs. For the monotonic biaxial loading case, both, stress-controlled and strain-controlled models, are put forth. In case of stress-controlled models, the neural network predicts the strain increment $\Delta\boldsymbol{\varepsilon}$ based on stress–strain states $\boldsymbol{\sigma}^{(t-i)}, \boldsymbol{\varepsilon}^{(t-i)}$ sampled from previous time steps and the current stress increment $\Delta\boldsymbol{\sigma}$:

$$(\Delta\varepsilon_1, \Delta\varepsilon_2) = f(\sigma_1, \sigma_2, \varepsilon_1, \varepsilon_2, \Delta\sigma_1, \Delta\sigma_2, \theta_\sigma). \tag{5}$$

In strain-controlled models, increments of strain are provided as inputs and increments of stresses are obtained as outputs, given by the relation

$$(\Delta\sigma_1, \Delta\sigma_2) = f(\sigma_1, \sigma_2, \varepsilon_1, \varepsilon_2, \Delta\varepsilon_1, \Delta\varepsilon_2, \theta_\varepsilon), \tag{6}$$

where θ_σ and θ_ε are the neural network parameters. The strain-controlled model has the advantage of being directly usable in FE simulations. For uniaxial cyclic loading, a two-layer constitutive neural network which takes in two previous points on the stress–strain curve, in addition to the current point is proposed. The history of the stress–strain curve is important to capture the cyclic behavior. In the results, the authors show the ability of the neural networks to generalize to unseen proportional and non proportional stress paths for monotonic loading, but it falls short for unseen low stress cyclic loading path.

As a methodic extension to previous work, Ghaboussi et al. [90] proposed the nested adaptive neural network (NANN). The NANN is an FFNN, with a special structure of successively learned nested stages. The NANN, as depicted in Fig. 5a, consists of a base module, which corresponds to a strain-controlled constitutive FFNN

$$\Delta\sigma^{(t)} = f(\Delta\varepsilon^{(t)}, \sigma^{(t)}, \varepsilon^{(t)}, \theta). \tag{7}$$

The basis module is extended by gradually trained additional modules

$$\Delta\sigma^{(t)} = f(\sigma^{(t-i)}, \varepsilon^{(t-i)}, \theta) \tag{8}$$

of path-dependent deformation data for discrete time steps $(t-i)$ to form higher level NANNs, where θ are the associated neural network parameters. The method is tested on experimental data of triaxial compression tests of sand, for which the authors show how the prediction quality increases depending on the NANN level. Although the proposed architecture enables training constitutive neural networks from data describing variable length paths, the model complexity (i.e. the number of trainable parameters) grows linearly with the maximum level of the NANN and the method is therefore limited to short stress–strain paths. Due to the variable length input, the proposed method can be seen as a precursor of the later proposed applications of RNNs, which are discussed in Sect. 3.

Unlike in [21] and [90], where the internal material state is assumed to be captured sufficiently by discrete samples from the deformation history, Furukawa et al. [91] proposed to represent the material state in a state space model and to train a so-called implicit constitutive neural networks. Inspired by control theory, the implicit constitutive neural network acts as a surrogate of a dynamical system

$$\dot{\boldsymbol{x}} = f(\boldsymbol{x}, \boldsymbol{u}). \tag{9}$$

Where, for visco-plastic material models, Furukawa et al. defines the state
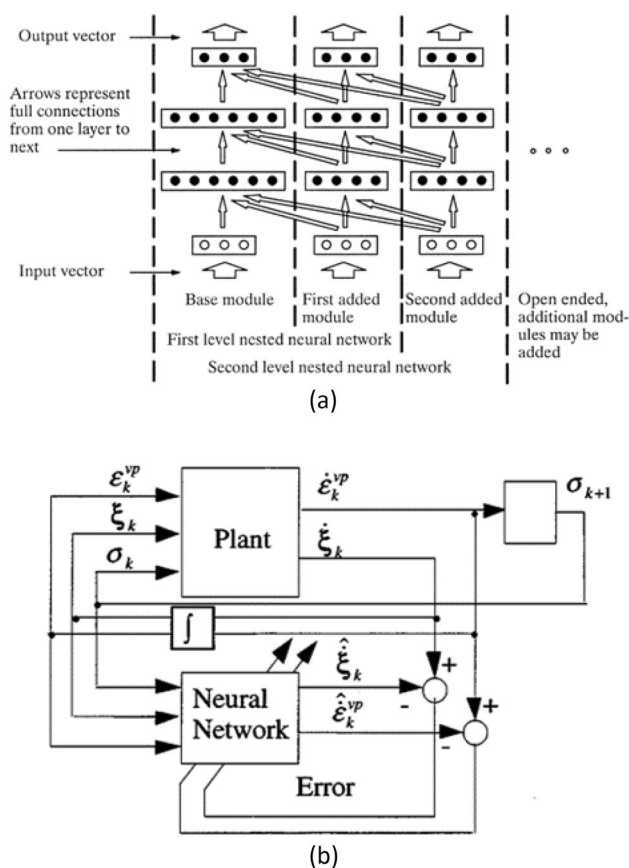
Fig. 5 Early approaches to constitutive neural networks. **a** The nested adaptive neural network architecture introduced in [90] Copyright© 1998, Elsevier. **b** Embedding graph of the implicit constitutive neural network introduced in [91] Copyright© 1998, John Wiley and Sons

$$\boldsymbol{x} = (\varepsilon_{vp}, \zeta) \tag{10}$$

as a combination of the visco-plastic strain ε_{vp} and internal variables ζ, such as the back stress and drag stress. The stress plays the role of the control input $\boldsymbol{u}$. As shown in Fig. 5b, the neural network models the influence of stress on the internal variables and strain rate, while the stress for the next time step is derived analytically. The proposed framework is instantiated for visco-plasticity and tested on pseudo-experimental uniaxial cyclic-loading data, which are sampled from the Chaboche's constitutive model [94], and on experimental data of steel on elevated temperature under the assumption of fixed strain rates and known elastic responses. Although comparable results of the implicit constitutive neural network to that of 1D Chaboche's constitutive model have been shown, its validity is not proven for a higher dimensional case, and for non proportional complex stress states.

Further, Lefik and Schrefler [95] proposed history-dependent constitutive neural networks to model material behavior of super-conducting fibers under biaxial loading with hysteresis. Several points on obtained stress–strain curves are used to describe the material state. For training, data from numerical simulations are used, for which a rotation data augmentation technique is applied to cover invariances within the data and to preserve objectivity of the constitutive model. The rotation operation is performed without additional constraints in the cross section of the fiber, where the material model behaves isotropic. Therefore, common rotation matrices $\boldsymbol{R}$ were applied to the relevant quantities, like for example to the stress tensor

$$\boldsymbol{\sigma}_{\text{rot}} = \boldsymbol{R}^{\top}\boldsymbol{\sigma}\boldsymbol{R}, \quad \text{with } \boldsymbol{R}^{\top}\boldsymbol{R} = \mathbf{1}. \tag{11}$$

The elasto-plastic constitutive neural network is applied to reproduce the behavior simulated in one dimension, as well as the homogenized behavior of super-conducting fibers from a two-dimensional simulation. The former results are compared to experimental results.

Al-Haik et al. [96] proposed the use of FFNNs to predict the relaxation of polymeric matrix composites depending on constant strain and temperature conditions. The model maps from strain, temperature, and process time to the relaxation stress and is trained on data from stress relaxation tests of a carbon fiber epoxy composite. The authors highlight that their model is in general more accurate than an explicit conventional visco-elastic model, in particular for temperatures near the glass transition temperature.

In more recent work, FFNNs are applied to learn from direct stress–strain data from various metallic materials [97, 98] and polypropylene [99]. By Gorji et al. [97], FFNN models are learned on pseudo-experimental data of sheet metal generated by using (a) the Zerilli-Armstrong model for temperature and strain rate-dependent hardening and (b) a J2 plasticity model used to simulate biaxial monotonic loading. The results show a good agreement of the FFNN response with responses of the J2 model for uniaxial notched tensile tests. Furthermore, Jordan et al. [99] used FFNNs with Bayesian regularization [100] to model the temperature and strain rate-dependent behavior of polypropylene based on experimental uniaxial loading data with varying temperature and strain rates. The FFNN learns the mapping from the viscous strain, viscous strain rate, and temperature to true stress. The comparison with a state-of-the-art conventional thermo-elastic visco-plastic model, which is calibrated to the experimental data, reveals a predominance of the FFNN in terms of accuracy and computational performance.

Also, for the direct learning setting, du Bos et al. [98] proposed and evaluated an alternative procedure, in which neural networks are trained to map from strain paths to corresponding stresses. The estimated stress paths are interpolated afterwards to reconstruct the stress–strain curve. In difference to the majority of the methods referenced above, which perform predictions of the material behavior at each

time step based on current and previous conditions, the proposed model aims to predict the global stress–strain relation for a given strain path in one pass. To reach a good model performance with a reduced input and output space, the sampling technique is optimized by minimizing the interpolation error. The method is evaluated on strain paths from reverse loading cases applied to isotropic and rate-independent elasto-plastic solids, for which it is shown to reach an acceptable prediction quality in the low data regime (in particular, 100 samples were used). Model training and evaluation was carried out on pseudo-experimental data.

When it comes to metallic materials, neural networks were used quite early for directly learning the hardening model from experimental high-temperature deformation data. In hot forming processes, many interrelated and non-linear hardening and softening phenomena such as work hardening, dynamic recovery, and recrystallization determine the constitutive flow behavior. The underlying mechanisms are significantly affected by temperature and strain rate. In this situation, conventional models with empirically fitted constants are often inaccurate or the applicable temperature and strain rate range is very limited [102, 103]. The parameters that define thermomechanical treatment also play a significant role for the microstructure and mechanical features of the hot formed product [104]. Therefore, accurate hardening models are an important requirement for process optimization and process analysis [105–107]. In several works, neural networks are proposed as black box models for modeling the flow behavior of various types of steels [101–103, 106, 108–112], aluminium alloys [113–116], titanium alloys [107, 117], magnesium alloys [104] and pure aluminium [105].

Already in the mid 90s, Rao and Prasad [101] trained a constitutive neural network on experimental data of medium carbon steel deformations under a variation of constant strain rates and temperatures. As depicted in Fig. 6, the neural network learns to map directly from the strain and the respective processing conditions to the flow stress. The model is shown to outperform a semi-empirical constitutive model. Similarly, Li et al. [107] trained and evaluated a constitutive neural network to predict the flow stress of a titanium alloy at elevated temperatures based on experimental data. Chun et al. [113] also proposed constitutive neural networks to predict the flow stress during hot compression and rolling of aluminium alloys. Hodgson et al. [108] proposed a constitutive neural network that includes additional input parameters such as the work hardening and evaluated it on experimental hot torsion data of 304 stainless steel. It is shown that the model with additional parameters clearly outperforms both, a constitutive neural network without the additional parameters and a conventional phenomenological model.

Moreover, Mandal et al. [109] proposed the use of a neural network to predict the flow stress of austenitic
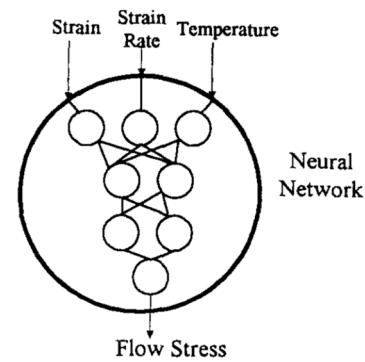
Fig. 6 Early general network structure for flow stress prediction introduced in [101] Copyright© 1995, Elsevier

stainless steels from hot compression test data. Unlike the above mentioned publications, in [109], the constitutive neural network also takes into account the alloy composition and can thereby be used to optimize the composition in addition to the hot forming parameters. Data from various grades of austenitic stainless steels are used and the composition is represented by the normalized amount of the most common elements. To be able to generalize across the composition, the amount of samples used for training and testing lies above 2000, which is about one magnitude above the amount of data used for training in the above mentioned works dealing with flow stress prediction. In addition to a quantitative evaluation of the model, a sensitivity analysis of the neural network was carried out to estimate the relative importance of the input parameters, especially the chemical composition.

Furthermore, Lin et al. [102] trained a neural network to predict the constitutive flow behavior of 42CrMo steel on data from hot compression tests. Sun et al. [117] also trained a constitutive neural network on data of compression tests but of a titanium alloy. The model predictions show good agreement with the experimental data and the model is shown to clearly outperform a conventionally used hyperbolic sine model. Ji et al. [103] trained a similar constitutive neural network on isothermal hot compression data of Aermet100 steel. The trained neural network is compared with a conventional Arrhenius-type constitutive model with strain compensation, which it outperforms in terms of accuracy. The measured mean absolute percentage error (MAPE, describing the mean over the relative regression errors) achieved by the conventional model is 7.62 while it is 2.58 for the constitutive neural network. At this point, however, we want to remark that while MAPE is a popular metric in studies of high-temperature deformation neural networks, it is known to be biased when used for model comparison as it assigns higher penalties to negative errors than to positive errors [118]. Therefore, and due to issues with close-to-zero values MAPE is not

commonly used in typical machine learning tasks and regression analysis nowadays.

Anyhow, regarding the learning of the behavior of Aermet100 steel, in [103], it is shown that, in contrast to a trained constitutive neural network, the conventional model performance collapses in the instability regimes, where the physical mechanisms differ from the ones in the stable regime. The constitutive neural network manages to accurately predict the constitutive flow behavior over the whole experimental range of temperatures and strain rates. In a similar manner, in [110] the constitutive neural network approach is compared to a modified Zerilli-Armstrong and a strain-compensated Arrhenius-type model on hot compression experiments of T24 steel. Again, the result of the comparative study is that the constitutive neural network outperforms the conventional approaches (with a MAPE of 0.45 for the constitutive neural network, 2.72 for the Arrhenius-type model and 5.22 for the Zerilli-Armstrong model). However, it is pointed out that this comes at the cost of interpretability and the requirement of high-quality data.

Bobbili et al. [114] trained a constitutive neural network to model the flow behavior under a high strain rate of 7017 aluminum alloy on data from the split Hopkinson pressure bar test [119]. In contrast to the majority of the publications in this section (where compression tests with strain rates up to 50 s^{-1} are considered), in [114], the experimental data covers strain rates between 1500 s^{-1} and 4500 s^{-1}. The trained constitutive neural network is compared to a Johnson-Cook model [120], and it is shown that the constitutive neural network outperforms the conventional baseline (with a MAPE of 2.58 for the constitutive neural network and 10.62 for the Johnson-Cook model).

More recent work on hot deformation constitutive neural networks include [104–106, 116]. Sani et al. [104] proposed a neural networks-based approach for modeling the behavior of cast magnesium (Mg-Al-Ca) alloy and compare it with a conventional hyperbolic sine-based model on hot compression experimental results. It is shown that the hyperbolic sine functions accurately predict the stress–strain curves for cases in the high temperature and low strain regime, but are biased for cases with low temperature and high strains, where twinning effects play an important role. The trained neural network on the other hand is shown to accurately predict the relationship over the whole strain-temperature range. Li et al. [116] compared the neural network approach with a phenomenological Arrhenius-type model and a physics-based model for work hardening and dynamic recovery on data from hot compression tests of 6082 aluminum alloy. As in the already mentioned comparative studies, the neural networks-based approach outperforms the conventional approaches according to the statistical metrics used, while the conventional methods have advantages in terms of model interpretability.

Moreover, Rezaei Ashtiani and Shayanpoor [105] proposed to use the initial grain size as an additional feature, besides strain, strain rate, and temperature, to predict the aluminum flow behavior under hot working conditions. Neural networks are trained based on data from isothermal hot compression tests of AA1070 specimens with varying initial grain sizes in the range 50–450 µm. Within the application study, the trained neural networks are used to determine processing maps and stable process regions depending on the initial grain size. While analyzing the relative sensitivity of the input parameters to the outputs, it is shown that temperature and initial grain size are most significant for the learned mapping, which is in accordance to well-known experimental observations, e.g. the Hall–Petch relation. Of course, this outcome mainly confirms that these effects are represented by the underlying data. As in [105], Kumar et al. [106] trained a neural network model and used it to create various processing maps, such as strain rate sensitivity maps, for 9CR-1Mo steel. Particularly, flow stress neural networks are proposed and compared to a linear interpolation stress correction scheme. In the results, the predominance of the neural network model is shown.

In general, the approaches to directly learn from global experimental stress-strain data heavily rely on idealized laboratory experiments, where deformation is kept homogeneous, and loading paths are nearly proportional. However, at the component level, deformations are heterogeneous, and the stress state and strain path often fall outside the experimental range. While the neural networks demonstrate promising results and occasionally outperform classical constitutive relations, their applicability beyond the training domain remains unproven. This is where the classical constitutive relation's advantage lies, as it can extrapolate predictions effectively. It's worth noting that a trained constitutive neural network has rarely been applied to predict at the component scale. Including the established constitutive relationship would help mitigate this issue, in addition to accelerating the current modeling strategies. This however comes at the cost of loss of universality of the neural network, as this is constrained by the constitutive equations. In the following, we review works on learning constitutive surrogates with the help of classical constitutive equations.

For building constitutive surrogates, neural networks are trained on data sampled from simulation results by using a reference constitutive model. The neural network surrogate model which is thereby obtained is used to replace the reference model in future simulations with the motivation to accelerate the simulation. Early works that use FFNNs to act as surrogate constitutive models includes [121–123]. In recent years, some works propose and discuss FFNNs as surrogate constitutive models [6, 124–127]. This revival of the usage of neural networks for constitutive modelling was, however, initialized by the work of Le et al. in 2015

[128]. Therein, an FFNN-based homogenization approach was proposed for hyper-elastic materials.

In some of the previously discussed works in indirect learning, the integration of constitutive neural networks into FE simulations via user-defined material subroutines was utilized [129, 130]. Hashash et al. [121] addressed numerical implementation issues in this context. Therein, an explicit formulation of the material stiffness matrix is proposed, which is derived from constitutive neural networks and leads to a more efficient convergence behavior of FE simulations incorporating constitutive neural networks. However, Yang et al. [131] pointed out, among other scenarios, that loading–unloading sequences are not considered in training the model proposed in [121].

Jung and Ghaboussi [122] trained a rate-dependent iso-tropic constitutive FFNN, which maps from strains, stresses, and strain rates of the current and previous discrete time step and the previous rate of stresses to the current rate of stresses to model visco-elastic behavior. The network is trained from synthetic creep and relaxation test data, sampled for various step sizes, and then evaluated on data from concrete beams. It is pointed out that the approach performs well with variable time steps, in opposition to previous rate-dependent constitutive neural networks (e.g. [91]). Yun et al. [123] proposed the use of various internal material state variables in combination with the stress–strain information of the last time step as additional neural network inputs to predict the cyclic hysteresis behavior of materials. The proposed neural network is trained on simulated cyclic loading data of concrete and steel parts, implemented into the FE material model, and tested on structural simulations. The model showed superior prediction performance compared to earlier developed neural network constitutive models. The approach was evaluated on various uniaxial cyclic experiments and it is pointed out that the model showed the capability to learn post-limit material behavior, e.g. buckling, tearing, and yielding. Though, the influence of varying time step increments on the neural network prediction is not addressed.

A study on the performance of FFNNs is presented by Bessa et al. [6] at the example of learning the material behavior of a 2D hyper-elastic composite. Particularly, this study investigates both, learning the material behavior on the basis of a constant RVE (including a comparison with the performance of a Gaussian process model [132]) and for varying microstructure features. Moreover, this publication highlights the importance of the design of experiments, the importance of modeling uncertainty and the benefits of such machine learning-based surrogate models. In another work, Stoffel et al. [124] compared two approaches to predict the highly dynamic behavior of shock-wave loaded plates. Both approaches involve the use of FFNNs. While in the first approach a (structural) FFNN is trained on experimental data to predict the structural deformation directly, in the

second approach a constitutive neural network that aims to model the visco-plastic behavior is trained on simulation data and is implemented into FE code. The proposed constitutive surrogate maps from stress, backstress, and plastic strain tensor components to the plastic strain rate and backstress rate tensor components and eliminates the need for an iterative solution of the constitutive behavior. The direct comparison shows that the results of the neural networks-based constitutive surrogate are much more accurate than the results of the structural FFNN and reduce the computational effort of the FE simulation. In contrast, the effort for implementing the structural FFNN is much lower as it can be trained on experimental data only, and does not require a numerical simulation model.

Huang et al. [125] proposed a combination of the proper orthogonal decomposition and per-component constitutive neural networks to model hyper-elasticity and plasticity. The constitutive neural networks are trained on simulation results of 2D and 3D unit cell data with the accumulated absolute strain as a history variable to represent the material state. In the 3D case, data is sampled from a homogeneous cubic specimen under triaxial loading-unloading conditions, where the end-point A' of each loading path is sampled randomly from a unit sphere as depicted in Fig. 7a. To highlight the generalization abilities of the trained neural networks, the models are applied to various 2D and 3D structural FE simulations by using the automatic differentiation and derivation toolbox AceGen [133] to derive the tangent matrix. As shown for a 3D bar necking example in Fig. 7b, the results of these simulations are in good overall agreement with the results obtained when using the reference constitutive model.

Furthermore, Zhang and Mohr [126] proposed a constitutive neural network for von Mises plasticity with isotropic hardening. The neural network maps from the current stress and plastic work to the elasto-plastic tangent matrix and the Young's modulus without making a priori assumptions about the yield surface, flow rule, or hardening law and is combined with modified algorithms for uniaxial loading and plane stress loading. The approach is shown to be able to sufficiently reproduce predictions of a J2 plasticity model, including large deformation responses for complicated multi-axial loading-unloading paths. However, the transition between elastic and plastic domain had to be artificially smoothed, which can lead to over or underestimation of initial yield stress. Also, Jang et al. [127] proposed a neural networks-based surrogate of a J2 constitutive model, in which linear elastic loading and unloading is covered by a conventional physics-based model, while a nonlinear plastic correction is covered by the neural network. Due to this decoupling, the neural network can be trained efficiently based on one-element simulations and can be applied to a wide range of simulations. The trained model is verified on

Fig. 7 **a** Training data generation strategy and **b** exemplary results, utilizing the general constitutive surrogate described in [125] in an FE simulation of a tensile bar Copyright© 2020, Elsevier. **c** Comparison of constitutive neural networks-based FE simulation results of a cup deep drawing process with the results of the original analysis based on a J2 plasticity model, from [127] Copyright© 2021, Elsevier

(a)

(b)

(c)
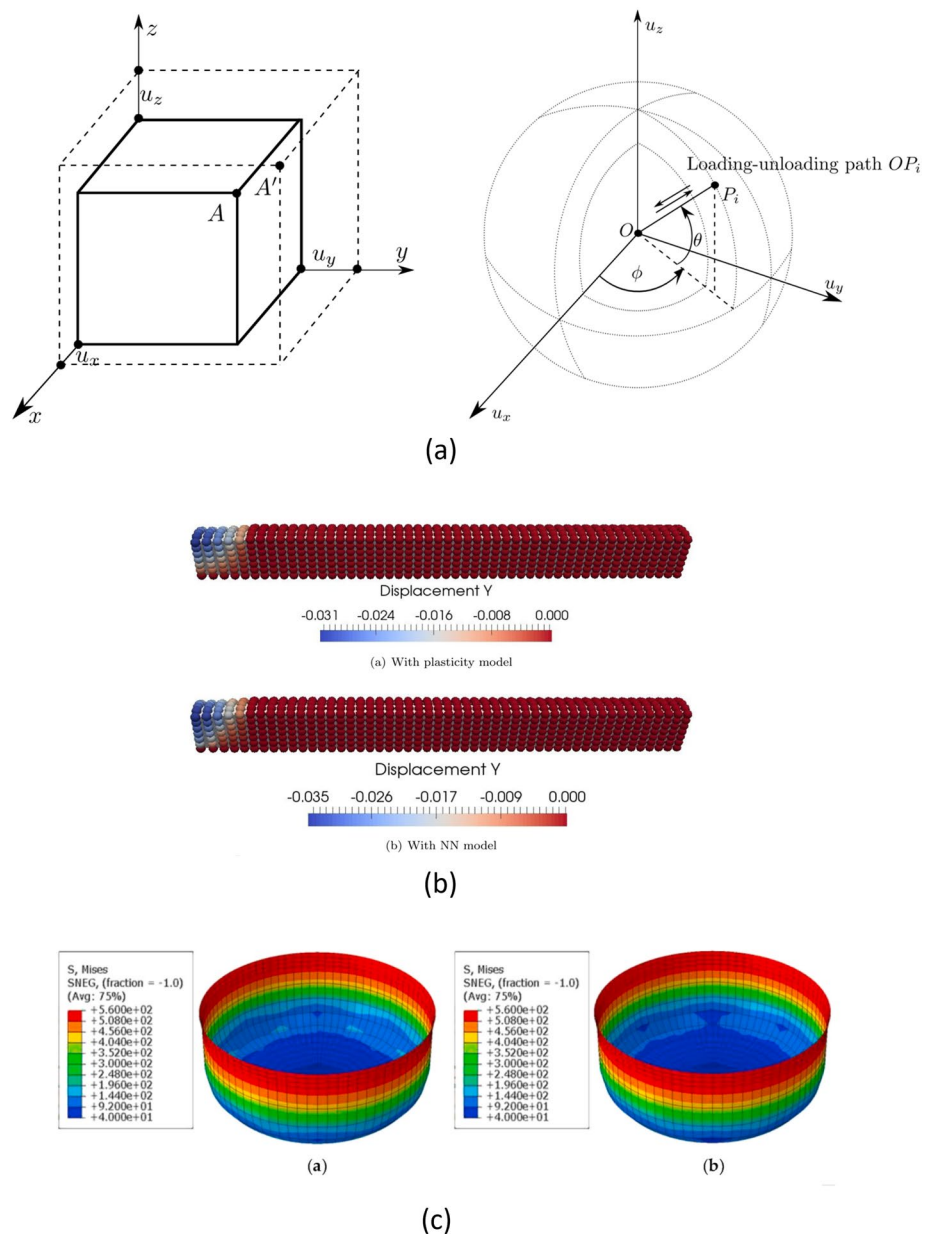
single-element simulations and tensile simulations of a dog-bone structure. Finally, a cup drawing simulation, depicted in Fig. 7c, is conducted based on the neural network model.

The usage of FFNNs as surrogates for conventional constitutive models is often accompanied by the need for a high amount of data and does not always lead to the aimed acceleration of the simulation speed. In their study, Zhang and Mohr [126] showed that ten to hundred thousand data points have to be sampled for an accurate J2 surrogate model for strains up to 20%. In the 3D case described above, Huang et. al [125] trained a model on samples from 8100 loading paths and emphasized that only loading is considered in the 3D case since the amount of data that would be needed to include unl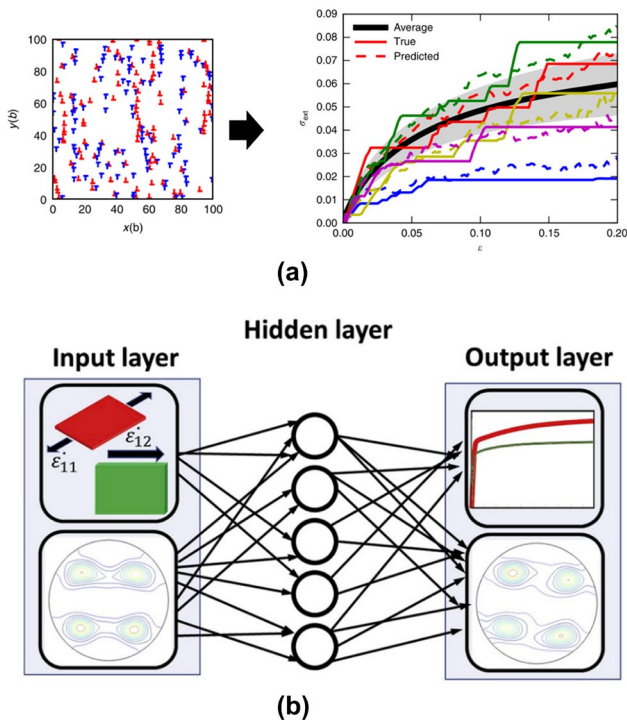oading is too high. When it comes to simulation acceleration, Jang et al. [127] reported a speed increase of 11% when using the neural network-based constitutive surrogate in comparison to the reference model. In [125], even a decrease in computational performance is reported. Stoffel et al. [124] reported a halving of the simulation time due to the proposed integration scheme. However, in computationally more complex scenarios, such as plasticity calculations on the microstructure level of polycrystals, constitutive neural networks are reported to provide a massive acceleration of simulations [134, 135].

On the micro-scale of polycrystalline materials, FFNNs are applied in the context of constitutive modeling also. Salmenjoki et al. [136] proposed neural networks to learn a model of the behavior of micro-scale crystalline structures

from two-dimensional discrete dislocation dynamics simulations. The work considers solids plastically deformed in a sequence of strain bursts, forming a staircase-like stress–strain curve, as shown on the right of Fig. 8a. The neural networks are trained to predict these curves from features originating from the initial dislocation configuration of the crystalline solid, which is depicted on the left of Fig. 8a, where red and blue symbols represent positive and negative Burgers vectors, respectively. Numerical experiments are conducted to quantify the approximation quality in various settings. The quality is found to be highly affected by the system size and, according to the authors, is surprisingly good for large-strain deformation dynamics.

Furthermore, Ali et al. [134] proposed FFNNs as fast surrogate models for the computationally expensive crystal plasticity Finite Element Method (CPFEM), where the homogenized response of the RVE is of interest. As depicted in 8b, the neural networks learn the stress–strain relation together with the texture evolution model. The neural networks are trained and evaluated based on data from experimentally validated rate-dependent CPFEM simulations of single crystals and polycrystal aluminum alloy AA6063-T6 under uniaxial tension and shear. The results show a
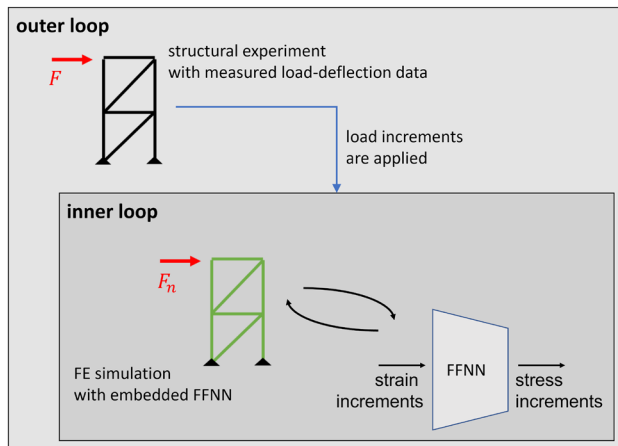
good agreement with the CPFEM results and an immense decrease in calculation time (of up to 99.9%).

Ling et al. [137] compared two neural network approaches for constitutive modeling in the case of known symmetry and invariance properties. While the first approach is based on an augmentation of the dataset to exploit the knowledge about the symmetries, the second approach is based on a proposed invariant representation of the input data. Besides the application of the approach to the modeling of a turbulent flow, the proposed methods are evaluated on a crystal-elasticity case study, with cubic crystal symmetry. The approach described in [137] is generalize by Jones et al. [138] to embed further constraints and invariances for stress and plastic flow and show that this can reduce the amount of training required.

2.2 Indirect Learning

Learning constitutive neural networks directly from global experimental stress–strain data is a straightforward and easy-to-implement way towards a machine learning constitutive model, if some preconditions are fulfilled. To create a dataset that is sufficient for neural network training, the constitutive relationships of interest have to be identifiable in sufficient quality, and a sufficient quantity of experiments to measure the data must be carried out. As stress fields in structural components are typically not measurable, training relies on the global material response measured. This means that each sample of the training set corresponds to one experiment, conducted usually under the assumption of uniform loading and homogeneous material behavior. Consequently, the direct learning approach is often restricted to simplified scenarios, such as the modeling of uniaxial and biaxial loading or the modeling of hardening behavior only. Compared to such tests, where each experiment is supposed to produce one stress–strain sample, results from structural experiments are rich in implicit constitutive information. The main goal of indirect learning of constitutive neural networks is to utilize this information and thereby enable neural network constitutive modeling in more complex scenarios.

Besides the direct usage of FFNNs on experimental data for implicit constitutive modeling, Ghaboussi et al. [129] proposed an early framework for indirect learning. Within the training framework, global load-deflection data from structural tests (i.e. by applying load on a truss structure) are used, combined with an FE simulation of the structural test with an embedded FFNN as constitutive model. After pre-training the FFNN on data from a linear elastic model in the first step, the FFNN is trained in an iterative bootstrapping procedure, which is called autoprogressive training by the authors in [129]. As depicted in Fig. 9, each iteration of the autoprogressive training consists of two nested loops to generate training data:

(a)

(b)

Fig. 8 a The mapping (learned in [136]) from the dislocation configuration of a crystalline solid to the staircase-like stress–strain curve (reassembled from [136], CC-BY). **b** Plot of the crystal constitutive neural network mapping introduced in [134], which learns the stress–strain relation together with the crystal evolution model Copyright© 2019, Elsevier

Fig. 9 Illustration of training data generation for the autoprogressive training procedure following [129]

- In the outer loop, load increments from the global load-deflection curve are applied to the FE model of the structural component.
- For each applied load increment, in an inner loop, the FE model with the embedded FFNN is used to compute the actual deflection. To meet the measured deflection, displacement corrections (the difference between actual and target deflection increment) are applied in an iterative procedure. As training data for the FFNN, stress–strain data at local material points of the structure are extracted, which follow from this iterative procedure. The inner loop ends, when the stress–strain data converges and the displacement correction is sufficiently small.

In [129], it is stated by the authors that the procedure described above may need to be performed several times of the full range of the applied load in order to train the neural network material model satisfactorily. The autoprogressive framework is introduced on a simple truss structure example and tested based on a more complex structural setting of graphite epoxy lamina, however with the assumption that the behavior of epoxy lamina is not path dependent. Follow-up works that build upon the autoprogressive algorithm apply it to sand under non-uniform triaxial compression [130] and to model soil behavior from field measurements excavations [139]. Methodical extensions of the autoprogressive algorithm were proposed by Yun et al. [140], who focused on hysteretic material behavior and by Shin et al. [89] to improve the overall robustness of the training framework. The autoprogressive framework was expanded by Yun et al. for applications with cyclic loading and hysteretic behavior [140]. To model such behavior, a neural networks-based material model using a novel algorithmic tangent stiffness formulation was proposed.
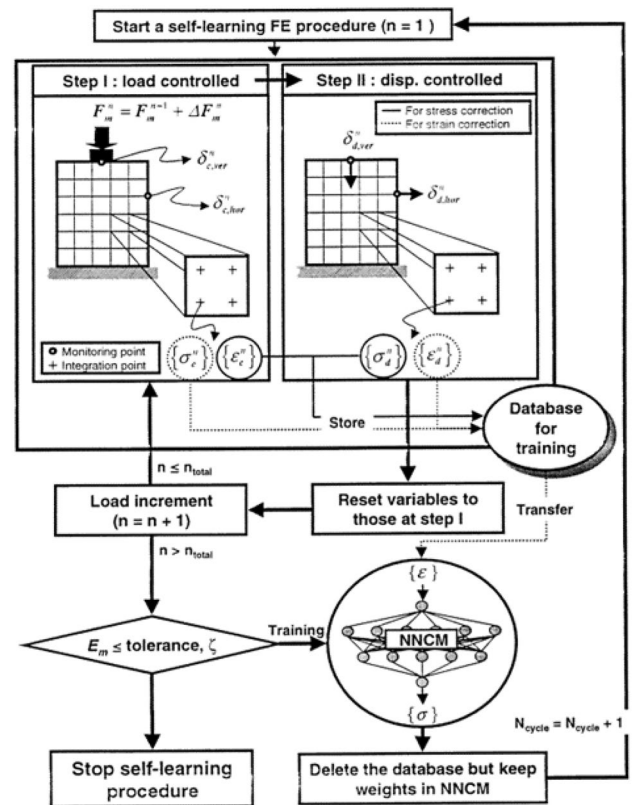
Fig. 10 Model architecture of the robust indirect approach introduced in [89] Copyright© 2000, Elsevier

The adopted autoprogressive procedure was applied to a structure of hysteretic beam-column connections [141] and it was shown that the trained neural network model has a superior learning capability compared to the previous direct neural networks-based material models. Moreover, the neural network was shown to be able to successfully extract the local cyclic behavior from the global responses measured in synthetic and real experiments and is capable to generalize to unseen cyclic motions. In this context, Shin and Pande [89] proposed a more robust variant of autoprogressive training. The proposed training scheme is depicted in Fig. 10. A central difference to the original autoprogressive training ([129]) is that the neural network is retrained only once per load pass by using only data from the current load pass. Also, the strain correction in the original autoprogressive training is complemented with a stress correction scheme to enable the successful training of strain softening. The authors of [89] investigated furthermore how the location and number of global monitoring points influence data efficiency and prediction accuracy. The autoprogressive framework relies on a proper initialization of the neural network parameters for fast convergence [142], e.g. based on an approximate linear elasticity matrix [89]. Moreover, as shown in Fig. 10, two FE simulations are needed per

iteration. In the so-called stress correction scheme (which is contrary to the strain correction scheme), the first simulation is load controlled and yields a displacement field, which is the basis for the second displacement controlled simulation yielding a stress field. The neural network is trained on the resulting strain–stress pairs until the observed and predicted displacement field match. However, such coupled simulations potentially lead to computational issues for applications with complex structural simulations [54].

In the context of dual phase steels [143], in multi-scale modeling [144–146] and laminates [142, 147], several methods for indirect learning were proposed in recent years. However, as [144, 146] involve the explicit formulation of energy principles, we classify them as physics-integrated surrogate models, which are discussed in Sect. 5.2. The other approaches mentioned are discussed in the following. Li et al. [143] trained a neural network as part of a modified Johnson-Cook model that reflects the non-monotonic strain rate and temperature effects on strain hardening of dual phase steel from experimental tensile test data. As before, the training of a constitutive neural network is accomplished in a bootstrapping process. First, the neural network is initialized by pre-training on user estimations of the material behavior under different temperatures and strain rates. In the second step, the pre-trained network is embedded into a 3D FE simulation and trained in the bootstrapping process, where the cost function is derived from the difference between the predicted and experimentally measured total forces, assuming a linear variation of axial stresses over cross-sectional areas. Although the authors measure the strain distribution on the surface experimentally, this has not been included in the error function for neural network training. The rich local information of strain distribution is lost here, which if used would increase the accuracy and generalization of the FFNN to a heterogeneous field. Based on the same approach, Pandya et al. [148] trained a model to describe the rate-dependent plastic behavior of aluminum 7075 in a hot stamping process.

For various examples from lamination, Liu et al. [142, 147] learned constitutive relations based on experimental data by combining neural networks with lamination theory, which is implemented into differentiable FE code. The goal thereby is not to learn a general stress–strain model, but instead more context-dependent models such as shear constitutive relations and the failure initiation criterion [142] or the damage accumulation law [147] of laminates. Although the results are very convincing, the requirement of differentiability of the used FE code leads to a very elaborated development process in the case of more complex FE problems [54]. In [54], a method is proposed to replace the custom FE code with the commercial Abaqus FEM solver. The training procedure of the so-called Abaqus Deep Neural Network is outlined in Fig. 11. Within this approach, a neural network
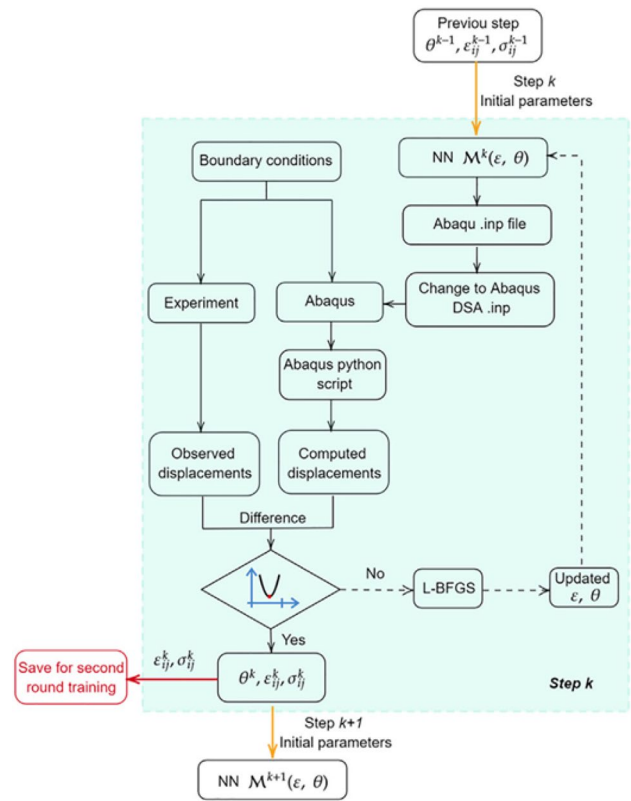

Fig. 11 Model architecture of the Abaqus Deep Neural Network approach introduced in [54] Copyright© 2021, Elsevier

(with parameters $\boldsymbol{\theta}$) that maps from strains to the Jacobian of $\boldsymbol{\sigma}(\boldsymbol{\varepsilon})$

$$\tilde{f}(\boldsymbol{\varepsilon}, \boldsymbol{\theta}) = \frac{\partial \boldsymbol{\sigma}}{\partial \boldsymbol{\varepsilon}} \qquad (12)$$

is included into the Abaqus FE simulation and is trained based on the difference between displacements observed in experiments and displacements computed by the FE simulation. The central contribution of [54] is the utilization of the so-called Abaqus design sensitivity analysis functionality to compute the gradient of the computed displacements with respect to the constitutive neural network and the integration of this functionality into backpropagation. The proposed method was applied to learn the progressive damage constitutive law of a fiber-reinforced composite and the linear constitutive laws of its constituents based on structural-level data for a specific loading scenario.

The indirect learning approach using FFNNs has rarely been employed to tackle path-dependent elasto-plastic or plasticity-driven ductile damage problems. The auto-progressive method, while intriguing, demands significant computational resources and necessitates expertise in both, FEM and constitutive modeling. In their work, Li et al. [143], present an framework for adaptively identifying the flow stress

but make the assumption that the plastic flow behavior is already known. It would be interesting to explore the entire elasto-plastic flow behavior in this context.

2.3 Discussion

As shown in this section, the use of FFNNs to learn constitutive models from experimental or simulated data is very popular. However, due to long-term effects, the state and behavior (e.g. hysteretic) of materials often depend on the load history. Various methods to represent the current material state for FFNNs were proposed and discussed. Examples, as discussed, include the use of past values from sampled isolated time-steps [21, 95, 121], the use of aggregated values [125] and the use of additional internal hardening variables [91] or the use of processing time-step [96].

However, neural networks are limited to fixed input mappings, while the load history is of variable length. Methods that face the problem by aggregating the history through subsampling [21, 95, 121] or using accumulated values [125] typically accept some degree of information loss by making assumptions about the history of internal variables. A first attempt in the constitutive modeling context to solve this problem by using a more flexible neural network architecture is the NANN approach. However, the NANN approach is limited to short sequences, as the number of model parameters highly depends on the sequence length. In contrast, advanced modeling techniques like RNNs and time convolution approaches are promising alternatives to solve these issues as in the case of RNNs the number of model parameters is independent from the sequence length and in the case of time convolution, the number of parameters required for learning temporal patterns is greatly reduced.

Furthermore, among the papers reviewed in this section that focus on direct learning, many address the learning of flow stress curves for different materials. Learning such curves is a 1D regression problem and can actually be solved by comparatively simple neural networks. While applying such is a reasonable proof-of-concept for the applicability of using neural networks in materials modeling, it, however, raises the general question of when applying machine learning is advantageous to classic models. In this specific case, an advantage of neural networks is their characteristic of being able to be fitted to any flow curve. However, a huge advantage in execution time is not to be expected. In contrast, fitting one out of a plethora of already developed analytical functions should result in similarly good curve approximation while at the same time forming derivatives is much more straightforward. In other cases, like in higher-dimensional regression problems, in contrast, the use of machine learning is more beneficial as the effort to find analytical functions increases with an increase in dimensions,

see for example the modeling of complex yield functions in [149] and [150].

3 Advanced Neural Networks for Constitutive Modeling: Considering the load history by learning from time-series

In this section, we outline and discuss work that utilizes and introduces advanced neural network models for processing sequential data in the context of constitutive modeling. As we are not aware of published indirect learning approaches that utilize such advanced neural networks, we present approaches for direct learning of advanced neural networks for processing sequential data in the following Sect. 3.1 only. In Sect. 3.2, a brief discussion of the presented approaches is given.

3.1 Direct Learning

In an early series of works by Oeser and Freitag [151], Graf, Freitag et al. [152, 153], and Freitag et al. [154], RNNs are applied to constitutive surrogate modeling tasks. In [151], RNNs are proposed as part of fractional material models to learn the history-dependent stress state of rheological materials with fading memory, where the effects of applied stress states on structural behavior gradually diminish over time. The RNN model is applied to synthetic creep test data, for which the better run time behavior in comparison to the exact solution of a fractional differential equation is highlighted. Based on this work, in [152, 153, 155], RNNs are combined with fuzzy structural analysis and the so-called α-level optimization [156] for history-dependent structural and constitutive models, which reflect measurement and process uncertainties. In [152], an RNN is used to map from loads and environmental influences of a reinforced concrete plate to structural responses. Both, inputs and outputs, of the RNN are fuzzy variables to reflect the uncertainty of the measurements, in the loading process itself and in the environmental influences. The approach is evaluated on the long-term responses of a reinforced concrete plate under dynamic loading.

While in [152] models reflect the material behavior on the structural level, in the follow-up work [153, 155], the approach is applied to learn constitutive surrogates for fuzzy FE models [157]. The approach is applied to learn a surrogate of a linear elastic constitutive model in [155] and from a fractional Newton element to simulate visco-elastic material behavior as well as from a fuzzy FE analysis of a three-dimensional structure under long-term loading in [153].

For various single-scale applications, different types of RNNs have been proposed [158–161]. Gorji et al. [158]

proposed and evaluated GRUs as a substitution for conventional physics-based plasticity models in the large deformation regime. The model is trained based on single-element simulations using a Yld2000-2d yield model with homogeneous anisotropic hardening and evaluated in various settings, including arbitrary multi-axial loading paths. Besides the homogeneous studies, the GRU-based model is trained and evaluated on results from a unit cell analysis of a two-dimensional foam. Zhang et al. [159] proposed an LSTM-based approach to model the cyclic behavior of sand under drained and undrained conditions. The proposed model consists of two separate LSTM networks. One network mimics strain-controlled and the other mimics stress-controlled soil behavior. The input and output variables of the proposed models are depicted in Fig. 12a, where $p^i, q^i, \epsilon_v^i, \epsilon_d^i$ are the mean and deviatoric stresses, volumetric and axial strains at time step i and additional variables $L1$, $L2$, $L3$ describe the current loading stage. Further, the features m and e_0 were introduced, to describe the drainage and the initial void ratio of sand. In the presented study, the model is trained on synthetic data of drained and undrained sand and is employed to simulate the behavior of real sands under cyclic loading. The model is evaluated on experimental data. As part of the

evaluation, the trained model is shown to be able to accurately predict history-dependent effects such as shear strain accumulation and densification.

More recently, Zopf and Kaliske [162] proposed to combine constitutive neural networks and RNNs with the so called micro-sphere description [163] to model the finite strain behavior of rubber-like materials. The micro-sphere model enables the reduction of the stress–strain dependency of polymer chains to only one dimension and thereby allows to train constitutive models with data from uniaxial loading tests to drastically reduce the experimental effort. The proposed model consists of both, an FFNN for pure elastic behavior and an RNN for inelastic behavior. The model is implemented into an FE model which is evaluated on experimental data of uncured elastomers.

Bonatti and Mohr [160] proposed a special recurrent architecture for constitutive sequence models, which is designed to combine a high model capacity with an arbitrarily small material state representation. The recurrent architecture is used to learn a general stress–strain model and consists of special quadratic layers followed by a simplified LSTM layer, which learns an implicit state representation of the material state. The state variables are connected to

Fig. 12 Sequential data processing constitutive neural networks: **a** LSTM-based models for describing cyclic behavior of granular materials [159] Copyright© 2020, John Wiley and Sons, **b** so-called smart constitutive law approach introduced in [166] Copyright© 2021, Elsevier
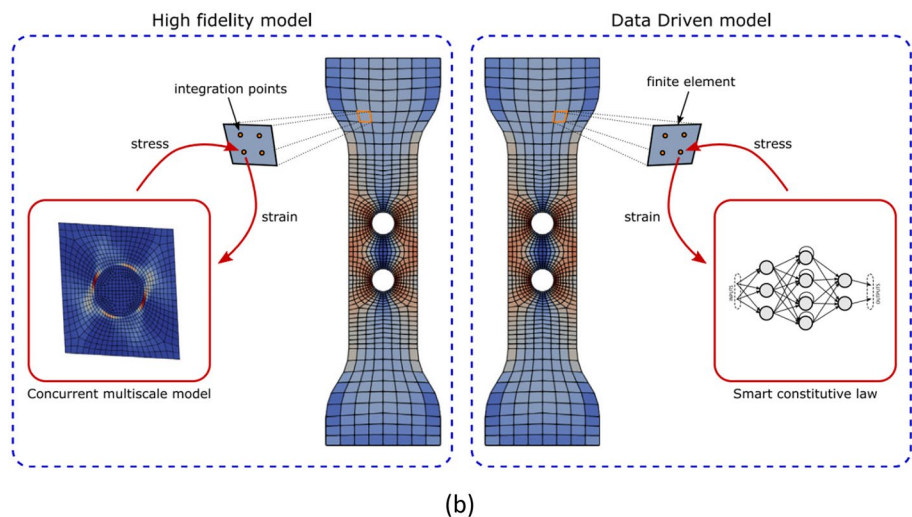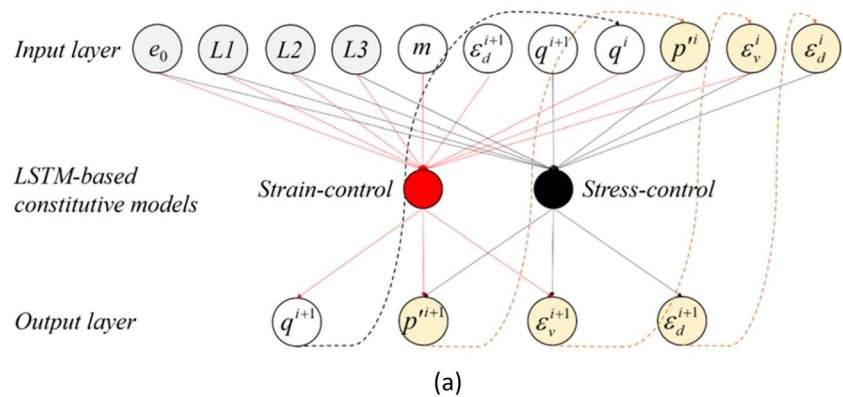
the next time step input and gated layer. Various separate constitutive surrogates are trained on the basis of the proposed architecture on data from FE simulations of single elements. Besides isotropic-hardening and mixed-hardening elasto-plastic models, the evaluation studies include models of crushable foam and hyper-elastic rubber with internal damage. Besides the evaluation of the reproduction quality, a study of the correlation between the implicitly learned state variables and the state variables of the conventional models shows the ability of the LSTM-based model to learn physically meaningful state spaces. Recently, Bonatti et al. [164] published an enhanced version of the model and showed that it can be used as an efficient and accurate constitutive surrogate for FE simulations.

Abueidda et al. [161] studied and compared various sequence models to predict the strain and temperature history-dependent behavior in two different applications. LSTM, GRU, and time CNNs are trained on data from an elasto-plastic cellular periodic material and a more complex thermo-visco-plastic steel solidification problem. In contrast to other sequence constitutive models, which are supposed to be implemented into an incremental FE solution procedure, in [161], the neural networks are used to predict the entire response sequence in one pass.

Up to this point, we discussed methods that learn path-dependent material models at the macro-scale, where the model represents the material in a homogenized sense. In the following, however, we discuss a second group of approaches namely models that learn the behavior of the material from RVEs, wherein the microstructure is spatially discretized. In computational homogenization, the computational bottleneck is the evaluation of the micro-scale model per integration point [165]. Due to various microscopic effects, the stress–strain relationship on the macro-scale is typically history-dependent [166]. RNN-based constitutive surrogate modeling approaches are recently proposed to solve these issues in a data-driven manner [15, 165–167]. Wang and Sun [15] integrated LSTM networks on different length scales into numerical models and investigated the computational efficiency of this hybrid approach using a multi-scale hydro-mechanical simulation of a multi-permeability porous material.

A popular approach is to utilize RNNs to learn a surrogate for homogenizing RVEs [87] and use the trained model as constitutive law in macro-scale simulations [165–167]. Mozaffar et al. [87] proposed this approach first and used RNNs to learn homogenization models for elasto-plastic composite materials. Therein, the RNN model is trained to map from the deformation path and microstructural descriptors to homogenized stresses and a quantity that the authors called plastic energy. The method is applied to two types of composites, for which it is shown to be accurate and efficient. Ghavamian and Simone [165] trained LSTM-based RVE surrogates to accelerate FE^2 simulations of history-dependent materials and applied it to virtual strain-softening experiments with a Perzyna-type visco-plasticity description. In contrast, Wu et al. [167] proposed GRU-based models as basis for meso-scale surrogates for an FE^2 simulation of a fiber reinforced composite under random cyclic loading.

Further, Logarzo et al. [166] introduced a so-called smart constitutive law approach, in which an LSTM is trained as a surrogate model on RVEs consisting of an elasto-plastic material matrix with hard elastic inclusions. Besides the homogenized stress components, the LSTM maps to the localized (maximum) plastic strain and von Mises stress. This localized information can be used in the application context, e.g. for predicting premature failure. The trained surrogate model is used as constitutive model in a structural FE analysis, as is depicted in Fig. 12b. Using the FE simulation, the machine learning-based approach is compared to a concurrent multi-scale scheme, in which the microstructural boundary value problem is solved explicitly for each integration point. The surrogate model was shown to be very accurate, as the coefficient of determination (R2-score) at the last of the 200 applied load increments was relatively high (above 0.989) for all stress components. Despite parallelization of the concurrent microstructure level calculations over 36 computing cores, the multi-scale scheme needed over three days to complete, while it took the machine learning-based approach less than 19 min to finish on a high-end GPU.

To model path-dependent elasto-plastic crystal behavior, Heider et al. [168] proposed special RNNs combined with a reference frame invariant formulation of the loss function. Based on synthetic data generated from crystal plasticity simulations, it was shown that the model results are highly dependent on the choice of the input representation and the loss function. It was furthermore shown that accurate results can be achieved with the right choice of the proposed frame-invariant loss in combination with some of the also proposed graph representations of the input data. Albeit the model was proven for a geometrically linear small strain case.

3.2 Discussion

Generally, for training RNN models, a sufficient amount of various non-linear strain paths and corresponding stress values is necessary. All of the reviewed papers present their own way of generating training data, which is sufficient for proving the introduced approaches. However, it is challenging to determine the superiority of one method over another. Therefore, it is crucial to establish benchmarking problems to evaluate the effectiveness of each method accurately. As the possibilities to sample strain paths is unlimited, appropriate approaches need to be

developed in order to provide comparable datasets. These can for example be to limit sampling on sub-spaces for applications in certain domains (e.g. metal forming) or to apply adaptive data generation and model training. Such sampling approaches furthermore enable the generation of standardized datasets that facilitate benchmarking. In contrast, another possibility for training models is the application of indirect learning. However, we have not found an indirect learning approach for RNN models in the reviewed literature. Learning indirectly from experiments would be an important extension to existing approaches, also tackling the data generation issue mentioned above.

In addition for path-dependent problems, the time step in FEM plays a crucial role in determining the accuracy and stability of the numerical simulation. Generally, irrespective of the path dependency, a small time step allows for better accuracy, but increases the computational cost of the simulation. On the contrary, a large time step can lead to numerical instability and loss of accuracy. In addition, in process simulations for example, variations in contact conditions necessitate different time step requirements. The path dependency amplifies this issue further, and hence is of importance. Commercially available FE solvers have the capability to automatically adjust the time step based on the system's state to optimize computational costs and maintain model accuracy. Except for [160, 164], there is a lack of research on how dynamically changing the time step affects RNNs. There are not many robust methods available to train RNNs in a way that mitigates the influence of the time step. This is a crucial aspect in driving the integration of neural networks into FEM and structural analysis beyond the academic community.

Moreover, while RRNs have been widely used for processing sequential data due to their ability to capture temporal dependencies, they suffer from several caveats. One issue is the vanishing or exploding gradient problem, which hinders the effective propagation of information across long sequences, leading to difficulty in learning long-term dependencies. Additionally, RNNs are computationally expensive and challenging to parallelize. Transformer models which utilize the attention mechanism [169] have emerged as a powerful alternative to address these shortcomings. Attention allows the model to selectively focus on relevant parts of the input sequence, enabling better capture of long-range dependencies without the mentioned gradient problems. Also, transformers process the entire input sequence in parallel, making them highly efficient and scalable. These properties have made transformer-based models already state-of-the-art in various machine learning domains, e.g. for natural language processing tasks [170] and subsequently in other time series processing applications [171].

4 Advanced Neural Networks for Constitutive Modeling: Considering spatial information by learning from images

Neural networks-based constitutive modeling approaches discussed so far operate typically without spatially integrating neighborhood relationships. However, in some applications it is crucial to consider the local neighborhood (or the sub-scale structure of a material) to accurately predict local quantities. CNNs are typically used for this purpose, hence we primarily discuss work presenting CNN-based models for constitutive modeling taking into account spatial information in the following. As we are not aware of published indirect learning approaches in this context, we review only approaches for direct learning in the following Sect. 4.1. In Sect. 4.2, the presented approaches are discussed briefly.

4.1 Direct Learning

Neural networks with convolutional layers have been proposed, e.g. to map RVE images in 2D [135, 172] or 3D [173, 174] or from graph representations [175] to homogenized stresses [172, 173], localized stresses [135], or the elastic energy functional [175]. In the following, we discuss these works in chronological order.

In his work, Frankel et al. [173] proposed an RNN with convolutional layers for the first time in order to predict the homogenized stress of oligocrystals from the strain history and a $20 \times 20 \times 20$ voxel image of initial grain orientations. In another work, Vlassis et al. [175] proposed a graph convolution-based neural network to predict the elastic energy functional for polycrystals under deformation. The polycrystal is represented by a graph, in which each node encodes a single crystal with its features, such as e.g. orientation and volume, and the edges encode the crystal connectivity. While the representation of crystallographic data as voxelized images is sensitive to the grid resolution and noise, the embedding into graphs followed by graph convolution enables to encode rotational invariances and frame indifference priors into the model. Thereby, convolution-based deep learning can be applied efficiently.

As shown in Fig. 13, the geometric encoder consists of graph convolution layers and dense layers and transforms the graph into a feature vector, which is concatenated with the deformation tensor in Voigt notation and processed by dense layers to finally predict the energy functional. The proposed loss term, furthermore, incorporates the gradients of the reference function with respect to the input
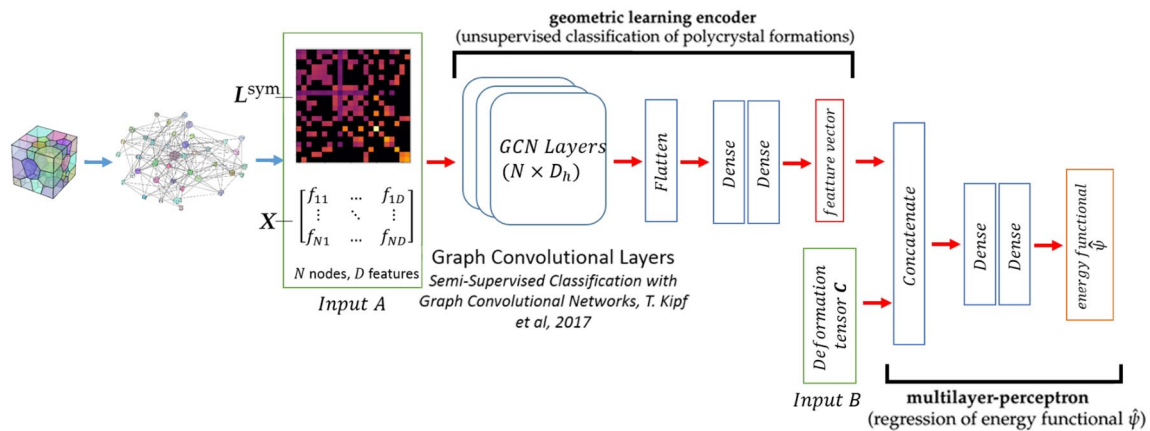
Fig. 13 Architecture of the graph convolution model for energy functional prediction of polycrystals introduced in [175] Copyright© 2020, Elsevier

based on so-called Sobolev training [176] in order to not only approximate the constitutive functions output but also its derivatives. This provides additional supervision, which leads to more sample-efficient training. Besides, the accurate stress tensor is obtained, as it is the gradient of the function with respect to the deformation tensor. In [175], it was shown that the proposed graph-encoding and neural architecture can leverage the crystallographic information in a reference frame indifferent manner and in consistency with the known thermodynamics principles. The usage of Sobolev training to guarantee being consistent with thermodynamics can, however, be seen as incorporating physical knowledge into the neural network learning process, which is a central point of many models that are summarized in Sect. 5.

Yang et al. [172] proposed a CNN to predict the homogenized stress–strain curve for a composite microstructure under uniaxial loading. The microstructure is represented by a 2D binary image, which encodes the distribution of a soft and a hard phase. The network is trained to predict principal components, from which the stress–strain curve can be obtained by reverting the a priori applied PCA transform. Mianroodi et al. [135] proposed a U-net type CNN as a surrogate of spectral simulations of inhomogeneous microstructures. The proposed neural network maps from 2D patches of local material properties (Young's modulus, Poisson ratio, yield strength) to 2D patches of according local von Mises stresses. When trained on 950 samples, the reported MAPE is 3.8% in the case of elastic and 6.4% in the case of elasto-plastic materials. Thereby, an immense performance gain compared to the spectral simulations can be observed. In this particular use case, the constitutive neural network approach is 103 times faster than the reference solver for elastic materials and 8300 times faster for elasto-plastic materials.

4.2 Discussion

As in Sect. 3, we did not come across any approach that uses indirect learning in the context of CNNs. However, it cannot be concluded that indirect learning is not applicable to the approaches described in this section. For instance, indirect learning may be accomplished in a 2D setting using in-situ EBSD or SEM measurements, followed by a simulation involving a CNN that attempts to match the surface deformation. However, for 3D microstructures, the only viable means of monitoring the 3D deformation field would be through high-energy accelerators, which is seldom accessible.

Nevertheless, as a first step towards the application of such models to real data, experimental data can be used to evaluate models that are directly learned from synthetic data. However, it is worth noting that all the methods discussed in the reviewed literature pertain exclusively to academic synthetic data. The validation of CNN models using real-world data remains limited. Achieving this is indeed challenging, as synthetic training data needs to be very similar to experimental data (including realistic microstructure representations). However it may be feasible to accomplish this at the surface level using advanced CPFEM in combination with in-situ EBSD measurements.

In general, CNNs are designed to work well on grid-like data, and can therefore be applied to 2D or 3D images of materials. However, there are certain limitations and challenges associated with the usage of CNNs. First, CNNs alone cannot be employed for sequence data processing. This necessitates the integration of other neural network layers, such as RNNs (e.g. for an effective analysis in the case of path dependencies). Another critical aspect of using CNNs for materials data is the computational expense associated with 3D convolutions, particularly with high-resolution 3D data. The data has to be represented as voxelized data, which

typically increases the computational burden and memory requirements. To address this challenge, graph-based CNNs have emerged as a promising alternative, particularly when values can be aggregated at the crystal level. By exploiting the local and global connectivity patterns within the crystal, graph-based CNNs can overcome some of the limitations of traditional CNNs. However, it is essential to carefully consider the specific requirements and characteristics of the data at hand when selecting the appropriate model.

5 Physics Integrated Neural Networks for Constitutive Modeling

This section deals with constitutive modeling based on integrating physics knowledge into the neural network learning process. The advantage of integrating physics is twofold: On the one hand, physics-informed models typically need fewer data to train than pure black-box approaches and, on the other hand, the models can extrapolate in a certain manner (e.g. extrapolation is consistent with integrated physics relations). Direct learning methods are presented in the following Sect. 5.1 and indirect learning methods are presented in Sect. 5.2. A brief discussion on the presented approaches is given in Sect. 5.3.

5.1 Direct Learning

From all the literature reviewed, we found only one work that introduces a model with integrated physics relations that was trained on experimental data (however, it can also to be trained on simulation data) [177]. In the approach described therein, the neural network loss function is customized such that predictions do not violate specific physical relations. This is achieved by integrating physical constraints into the learning process by adding penalty terms to the loss function. In the example of learning hyper-elastic material behavior, four constraints are applied particularly, which ensure that (i) the stress components remain zero when no strain is applied, (ii) the conservation of energy holds, (iii) the symmetry of the stiffness matrix for linear elasticity is retained and (iv) the material isotropy for nonlinear elasticity is hold.

In contrast to [177], many works exist that train constitutive surrogates. In this regard, we start with the approach described in [178], which is similar to the one introduced in [177] but evaluated using simulation data alone. In [178], a loss term is added to the neural network loss function to ensure being consistent with thermodynamics by punishing violations of the conservation of energy and the entropy inequality. The presented model was analyzed by training the network on simulation data of an Oldroyd-B-fluid and a hyper-elastic tyre. A different approach was described in

Haghighat et al. [179], which uses PINNs [180] as the basis for a framework to solve material-related differential equations. A boundary value problem solver incorporating elasto-plastic material behavior and damage formulations was developed and applied to a plate deformation problem. In [179], the authors stated that their approach can be applied to stress–strain data from meso- or micro-mechanical as well as molecular dynamics simulations. However, the PINN-based framework is only applicable to homogeneous stress and strain distributions.

Other approaches aim to set up models that learn thermodynamically consistent constitutive relations by mapping strain states to corresponding energies and deriving the stress states downwards. Consistency with thermodynamics is commonly obtained by predicting thermodynamic quantities (e.g. energies) instead of predicting stress components, as is done, for example in [146]. Earlier works already account for this while applying neural networks to learn hyper-elastic materials behavior, see [181, 182]. In general, hyper-elasticity is suited well for learning, as it is state-dependent and thermodynamics quantities can be approximated using standard FFNNs (without the need for storing internal history variables).

The first method we draw attention to was recently proposed by Masi et al. [183, 184], who introduced so-called thermodynamics-based neural networks. These are special neural networks that incorporate thermodynamics principles to learn the constitutive behavior of strain-rate dependent processes at the material point. The thermodynamics principles, namely the balance of energy and the dissipation inequality, can be written in the form of the following two equations [184]:

$$\boldsymbol{\sigma} = \frac{\partial \psi}{\partial \boldsymbol{\epsilon}} \tag{13}$$

and

$$D = -\sum_{i=1}^{N} \frac{\partial \psi}{\partial \zeta_i} \dot{\zeta}_i, \tag{14}$$

with ψ describing the energy potential, D describing the rate of mechanical dissipation and ζ_i describing kinematic variables.

Basically, thermodynamics-based neural networks consist of two connected FFNNs. The inputs for the first network are values of stress, strain, strain increment, and internal variables at a certain time increment, which are mapped to the increment of the internal variables. These and the strain increment are fed to the second neural network, which predicts the energy potential of the subsequent time increment. On this basis, the dissipation rate, as well as the stress increments (Eqs. 13 and 14), are evaluated from the network output using automatic differentiation. The dissipation rate is

forced to be positive or zero (depending on the material) during training to be consistent with the first and second laws of thermodynamics. In his work, Masi et al. [184] showed that the approach can be used to model 1D elasto-plastic material behavior with kinematic softening. In the example problem in [184], it was shown that thermodynamics-based neural networks perform better than pure black-box neural network approaches, such as the ones described in Sect. 2. Moreover, the results showed that the trained thermodynamics-based neural networks are accurate even for data points that are relatively far from the training range.

Similar to thermodynamics-based neural networks, but with a different neural network structure, Linka et al. introduced so-called constitutive neural networks in [185]. The inputs for the constitutive neural networks are the right Cauchy-Green tensor and (if necessary) non-kinematic information, such as microstructure or process conditions. On this basis, several sub-neural networks learn to predict invariants of the right Cauchy-Green tensor and the structure tensor. These are used to predict the strain energy in a subsequent neural network. The strain energy is in-turn used to derive the stress and stiffness tensor via automatic differentiation (as is done for the thermodynamics-based neural networks mentioned above). The applicability of constitutive neural networks to engineering problems is furthermore discussed in [185]. Therein, two use cases are analyzed, which are (i) predicting effective material properties of a matrix-inclusion composite RVE and (ii) accelerating macroscopic FE simulations of hyper-elastic material behavior. It was shown that the implementation of the trained constitutive neural network models into FE software is simple and computationally efficient. Moreover, it was shown that constitutive neural networks are data efficient and the trained models are robust against noise.

Another approach that is based on thermodynamics formulations is the so-called thermodynamics-informed neural network, introduced by Vlassis and Sun [186]. In contrast to thermodynamics-based neural networks and constitutive neural networks, only a simple FFNN is used therein, however, Sobolev training [176] is applied to integrate knowledge about derivatives in the loss function of the neural network (similar to [175], which is already mentioned in Sect. 4). The loss function is customized, such that a mapping from strain values to hyper-elastic energy functionals can be learned while including constraints for stress and stiffness values. Furthermore, in [186], it was shown that yield functions and plastic flow can be learned from simulation data of polycrystal RVEs using thermodynamics-informed neural networks. The thermodynamics-informed neural network approach has been developed further for learning rate-dependent, pressure-sensitive plastic behavior in [187]. It is pointed out, in contrast to the models presented in Sect. 3, that this method is better interpretable and allows for a modular design of the material models by for example treating elastic and plastic behavior separately in different networks. As an exemplary engineering use-case, an FE simulation is conducted on the macro-scale that incorporates a constitutive surrogate on the micro-scale trained on FFT simulation data of a polycrystal RVE.

Another neural networks-based approach to mention in this section is the approach described by Fernandez et al. [188], which maps the right Cauchy-Green tensor on potential energies to predict hyper-elastic material behavior (similar to the approaches described in [181, 182]). In [188], the approach was compared against a neural network that maps the right Cauchy-Green tensor on the corresponding stress components in a deformation problem of a meta material. By comparing the approaches, it was concluded that the approach mapping on potential energies has the best generalization capabilities, which is in accordance with the findings in the works presented above.

A very different approach, compared to the physics-informed approaches mentioned above, are the so-called deep material networks (DMN) [189]. DMNs are neural networks that learn homogenization operations to obtain the effective stiffness or compliance tensor from micro-scale simulation data (experimental data is also possible but cumbersome to gather). In contrast to thermodynamically consistent models, DMNs incorporate geometric information from RVEs. In its original version, DMNs consist of building blocks that have a tree-based structure, see Fig. 14. Each building block is a two-layered neural network which gets the micro-scale stiffness or compliance tensor for every phase as input and transforms it to the homogenized quantity, including a rotation operation to incorporate orientation dependence.
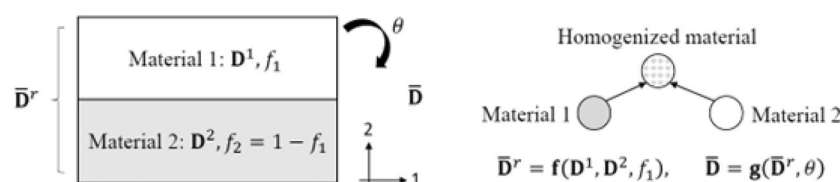
Fig. 14 Building block of a DMN and homogenization scheme as is introduced in [189]. D denotes the compliance matrix of the individual materials (with volume fraction f_1 and f_2) and $\bar{D}$ the homogenized compliance matrix for a rotation θ. Copyright© 2019, Elsevier

In [189], Liu et al. applied DMNs to homogenizing simple two-phase 2D-microstructures. The DMNs used therein were constructed out of binary tree building blocks. By training the DMNs on data from linear elastic homogenization simulations, it was shown that the approach is able to extrapolate to nonlinear plasticity and finite strain hyperelasticity. For solving 3D problems, Liu and Wu [190] introduced novel building blocks that include 3D rotation operations and applied them to crystal plasticity among others. Further extensions to DMNs have been introduced in [191] to incorporate strain localization enabled by cell division operations and in [192] for including interfacial effects via so-called cohesive layers. The ability of DMNs to enable fast multi-scale FE simulations has been shown in [193] and [194], both using examples of fiber reinforced composite.

5.2 Indirect Learning

In this section, we focus on approaches that learn constitutive relations indirectly with integrated physics knowledge. In particular, all of the models presented in the following integrate physics knowledge aiming to ensure being consistent with thermodynamics. In this regard, in [146], an RNN-based approach is described that infers the Cholesky factor of the tangent stiffness matrix of a deformed material. The prediction of the tangent stiffness matrix guarantees the approach to be consistent with the second order work criterion, that in-turn guarantees a more numerically stable model. Training can be done directly using pairs of stress–strain values or indirectly on the basis of measurements by evaluating balance equations during training. However, it is pointed out that measuring comprehensive stress–strain data only from experiments is challenging. The RNN-based model has been shown to successfully approximate and replace elasto-plastic and hyper-elastic material models in 1D and 2D FE simulations.

An alternative approach based on FFNNs that learns constitutive relations indirectly is described in [144]. In the approach described therein, neural network models are used to learn constitutive relations by evaluating balance equations (the loss function equals zero when the external forces and the relevant internal quantities are in balance), similar to [146]. However, in contrast to [146], the approach described in [144] does not require special physics relation to be incorporated in general. In this regard, in the application example, a neural network model is trained to relate principal stretches of a hyper-elastic material model to the first Piola-Kirchhoff stresses. It is pointed out that the neural network can be made thermodynamically consistent by predicting the hyper-elastic energy (instead of predicting the stress components) and use automatic differentiation to derive the stress components. However, this alternative has not been applied in the paper.

Besides [146] and [144], an early approach, in which FFNNs are trained indirectly using an energy-based loss function has been described already in 2011 by Man & Furukawa [195]. Therein, neural networks are supposed to learn the relation between stress and strain on a continuum basis. The learning process and the obtained constitutive models are evaluated at FE simulations of structural components that substitute real experiments.

5.3 Discussion

One advantage in learning directly from global experimental stress strain data and indirect learning of constitutive equations is that modeling assumptions and simplifications usually done in classic material modeling do not constrain the learned constitutive model and, hence, make the model more generally applicable. Therefore, constitutive relations need to be learned on the basis of experimental data. Gathering 'enough' experimental data in materials science is, however, demanding, which led to the idea of incorporating physical knowledge in the learning process, aiming to make models more reliable and data efficient. By doing so, the original advantage of universality vanishes depending on how much the machine learning models are constrained by the incorporated physical relations. Consequently, there is a trade-off between improving prediction accuracy and constraining learned constitutive models by incorporating physical knowledge. In this regard, however, we cannot assert our ability to provide recommendations on the appropriate level of physics to be integrated into neural network constitutive models. This depends on the quality and amount of available training data, which is to be assessed by the user.

Additionally, incorporating physical constraints while learning from experimental data is challenging. Although these learning scenarios come from the original intention of learning constitutive models (Fig. 3), only few works exist that describe corresponding approaches. The most straightforward approach described in the reviewed papers (and the only one that has been tested on experimental data in a direct learning scenario) is by modifying the neural network loss function in order to punish violations of physical constraints. Indirect learning, in contrast, opens more possibilities for incorporating physical knowledge into the learning process, as the neural network models can learn various relations out of which the material model is composed of (instead of just relating stress and strain). In this regard, few more elaborate approaches have been introduced in the reviewed papers, like inferring the Cholesky factor of the tangent stiffness matrix or using automatic differentiation. It can, however, be expected that there are many more possibilities to decompose the material model and incorporate physical constraints.

6 Summary and Outlook

The modeling of the constitutive behavior of materials using supervised learning methods, especially neural networks, is a fast-growing research field. As is shown in this review paper, many groundbreaking works have been published in the last years. For example, the modeling of path-dependent material behavior using recurrent neural networks, homogenization using convolutional neural networks, and integrating physics knowledge into the learning process. From a current point of view, the developed approaches are impressive and show the potential of learning algorithms in materials modeling. However, apart from the recent modeling advances, several challenges remain, which in our opinion are:

- *Benchmarking*: There is a need for standardized benchmarks to test and compare machine learning approaches for learning constitutive relations against each other. Benchmarks, which are common practice in machine learning (e.g. the MNIST and ImageNet datasets in computer vision [196, 197]), are designed to measure the efficiency of learning approaches and thereby enable comparability. This is necessary for engineers to make the right model decision for their application scenario. Although some of the discussed papers are published with associated code (e.g. in [144, 147, 160, 161, 188]) and some related benchmark datasets exist (such as the Mechanical MNIST [198] or MatBench [199]), to the authors knowledge there is no widely used benchmark dataset for the comparative evaluation of constitutive neural networks.
- *Sampling*: The space of possible loading conditions for material models is generally very large. Therefore, and, as performing numerical simulations and experiments to obtain information about material behavior is often time-consuming, sampling is crucial, see for example [6]. To tackle this issue, intelligent sampling algorithms, e.g. from the field of active learning [200], can be used for data generation and model training, see [201, 9] and [82] for application examples.
- *Measuring prediction quality*: Another effect of the large space of possible loading conditions is that learned constitutive models cannot be fitted accurately for any imaginable loading path. Hence, measures to estimate the prediction quality are essential for engineers to assess simulation results. Such measures are given intrinsically when applying probabilistic models, such as Gaussian processes, see for example [6, 201]. For descriptive machine learning models, however, such measures have to be still established. One approach can be found in [202], where a separate

machine learning model is used to estimate the error of a previously trained constitutive surrogate.
- *Data accessibility*: Having reliable data is the basic requirement for learning adequate constitutive relations. However, data accessibility and management (in particular respecting FAIR-principles [203]) is still an open challenge in materials sciences [204]. Database solutions are already under development (see for example [205–207] or relevant websites, such as www.materialscloud.org, www.nomad-lab.eu or www.materials-marketplace.eu) but far away from common use in engineering and industry. In this context, also the fusion of multi-fidelity data sources (e.g. experiments and numerical simulations) is an important research topic, see [208].
- *Knowledge integration*: Integrating knowledge into the learning process is a prominent way to tackle data sparsity and to improve extrapolation capabilities of machine learning models. Besides the presented ways for knowledge integration, in general, many ongoing research focuses on integrating knowledge into learning processes and thus to create so-called grey-box models, [209]. Regarding the learning of constitutive equations, additional knowledge can originate from all kinds of balance and conservation relations or physical constraints, such as convexity or monotonicity. First solution approaches that incorporate convexity have been introduced for hyper-elastic material behavior in [210]. While for monotonicity no such approaches are known by the authors in the materials modeling domain, however, in applied machine learning, approaches have already been proposed, such as in [211].
- *Extrapolation*: One of the main limiting factors in machine learning in general is the inability of trained models to extrapolate. The problem of extrapolation can, however, be addressed by methods mentioned in the above presented bullet points. These are: an efficient sampling to prevent from the need to extrapolate, measures for the prediction quality of models to detect and prevent from extrapolation, and, finally, the integration of physical knowledge into the learning process to enable model extrapolation within certain limits.
- *Interpretability*: A further general issue for using machine learning methods (especially neural networks) is interpretability [212]. The lack of interpretability makes it hard for engineers and scientists to analyze and understand model behavior. However, typically, integrating knowledge into machine learning algorithms improves interpretability. Besides knowledge integration, methods are already under development that improve the interpretability and explainability of neural network constitutive model predictions [174, 213].
- *Countering error accumulation*: In numerical simulations, the system response from an applied load is calcu-

lated incrementally (stress increments follow from strain increments, etc.). Therefore, errors in stress predictions lead to errors in subsequent increments and so on. The effect of error accumulation is studied in several constitutive neural network papers, including [166, 173, 214]. In [214], it was shown that error accumulation can be mitigated by regularizing constitutive neural networks. A further way to reduce error accumulation is to incorporate physical knowledge into the machine learning model to force predictions to be consistent with physics constraints.

Finally, we want to conclude this review paper by emphasizing that the reviewed works are only the beginning of the era of learned constitutive models. While already plenty of approaches exist for building surrogates of classic constitutive models, recently developed approaches that learn from experimental data are rare. Especially for applying deep learning and integrating physics, there is still room for further developments. However, currently, machine learning (and artificial intelligence) research receives a great deal of attention and machine learning methods quickly become replaced by new state-of-the-art methods. At the same time, the model complexity and demand of computational resources is growing. Consequently, this leads to an increased demand for interdisciplinary teams when adapting such models in specific scientific domains such as materials sciences. Further, in materials sciences, as we have seen throughout this paper, model evaluation is particularly difficult, as conducting experiments is often time-consuming and expensive.

Due to these facts, the adaption of new machine learning methods to problems in the materials science domain takes time. Recent examples are transformer architectures, which are based on the attention mechanism introduced in 2017 [169] and can consider long-range dependencies in sequential data without the need for recurrence. In constitutive modeling, first results were very recently published showing advantages of transformer models to recurrent neural networks [215]. Therefore, and in order to push forward the state-of-the-art neural networks-based constitutive modeling and to accelerate development, standard ways for machine learning adaption and implementation into commercial software codes need to be established aiming to facilitate interdisciplinary work. Nevertheless, the speed with which the field of learning constitutive models has evolved in the last decade makes us look forward to new developments in this regard as well as to developments tackling the above mentioned challenges in the near future.

Declarations

References

1. Haupt P (2013) Continuum mechanics and theory of materials. Springer Science & Business Media, Berlin
2. Zhang H, Diehl M, Roters F, Raabe D (2016) A virtual laboratory using high resolution crystal plasticity simulations to determine the initial yield surface for sheet metal forming operations. Int J Plast 80:111–138
3. Cybenko G (1989) Approximation by superpositions of a sigmoidal function. Math Control Signals Syst 2(4):303–314
4. Leshno M, Lin VY, Pinkus A, Schocken S (1993) Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. Neural Netw 6(6):861–867
5. Liu R, Kumar A, Chen Z, Agrawal A, Sundararaghavan V, Choudhary A (2015) A predictive machine learning approach for microstructure optimization and materials design. Sci Rep 5(1):1–12
6. Bessa MA, Bostanabad R, Liu Z, Hu A, Apley DW, Brinson C et al (2017) A framework for data-driven analysis of materials under uncertainty: countering the curse of dimensionality. Compu Methods Appl Mech Eng 320:633–667
7. Tran A, Mitchell JA, Swiler LP, Wildey T (2020) An active learning high-throughput microstructure calibration framework for solving inverse structure-process problems in materials informatics. Acta Materialia 194:80–92
8. Iraki T, Morand L, Dornheim J, Link N, Helm D (2023) A multi-task learning-based optimization approach for finding diverse sets of material microstructures with desired properties and its application to texture optimization. J Intell Manuf (accepted)
9. Morand L, Link N, Iraki T, Dornheim J, Helm D (2022) Efficient exploration of microstructure-property spaces via active learning. Front Mater 8:824441
10. Dornheim J, Link N, Gumbsch P (2020) Model-free adaptive optimal control of episodic fixed-horizon manufacturing processes using reinforcement learning. Int J Control Autom Syst 18(6):1593–1604
11. Dornheim J, Link N (2018) Multiobjective reinforcement learning for reconfigurable adaptive optimal control of manufacturing

processes. In: International Symposium on Electronics and Telecommunications (ISETC). IEEE 2018:1–5

12. Liu S, Shi Z, Lin J, Li Z (2020) Reinforcement learning in free-form stamping of sheet-metals. Procedia Manuf 50:444–449

13. Chinesta F, Cueto E, Abisset-Chavanne E, Duval JL, Khaldi FE (2020) Virtual, digital and hybrid twins: a new paradigm in data-based engineering and engineered data. Arch Comput Methods Eng 27(1):105–134

14. Kadlec P, Gabrys B, Strandt S (2009) Data-driven soft sensors in the process industry. Comput Chem Eng 33(4):795–814

15. Wang K, Sun W (2018) A multiscale multi-permeability poroplasticity model linked by recursive homogenizations and deep learning. Comput Methods Appl Mech Eng 334:337–380

16. Zhang R, Chen Z, Chen S, Zheng J, Büyüköztürk O, Sun H (2019) Deep long short-term memory networks for nonlinear structural seismic response prediction. Comput Struct 220:55–68

17. Koeppe A, Bamer F, Markert B (2019) An efficient Monte Carlo strategy for elasto-plastic structures based on recurrent neural networks. Acta Mechanica 230(9):3279–3293

18. Im S, Lee J, Cho M (2021) Surrogate modeling of elasto-plastic problems via long short-term memory neural networks and proper orthogonal decomposition. Comput Methods Appl Mech Eng 385:114030

19. Stoffel M, Gulakala R, Bamer F, Markert B (2020) Artificial neural networks in structural dynamics: a new modular radial basis function approach vs. convolutional and feedforward topologies. Comput Methods Appl Mech Eng 364:112989

20. Noé F, Tkatchenko A, Müller KR, Clementi C (2020) Machine learning for molecular simulation. Ann Rev Phys Chem 71:361–390

21. Ghaboussi J, Garrett J Jr, Wu X (1991) Knowledge-based modeling of material behavior with neural networks. J Eng Mech 117(1):132–153

22. Yagawa G, Okuda H (1996) Neural networks in computational mechanics. Arch Comput Methods Eng 3(4):435–512

23. Liu Y, Zhao T, Ju W, Shi S (2017) Materials discovery and design using machine learning. J Mater 3(3):159–177

24. Guo K, Yang Z, Yu CH, Buehler MJ (2021) Artificial intelligence and machine learning in design of mechanical materials. Mater Horizons 8(4):1153–1172

25. Dimiduk DM, Holm EA, Niezgoda SR (2018) Perspectives on the impact of machine learning, deep learning, and artificial intelligence on materials, processes, and structures engineering. Integr Materi Manuf Innov 7(3):157–172

26. Peng GC, Alber M, Tepole AB, Cannon WR, De S, Dura-Bernal S et al (2021) Multiscale modeling meets machine learning: what can we learn? Arch Compu Methods Eng 28(3):1017–1037

27. Jiao P, Alavi AH (2021) Artificial intelligence-enabled smart mechanical metamaterials: advent and future trends. Int Mater Rev 66(6):365–393

28. Bock FE, Aydin RC, Cyron CJ, Huber N, Kalidindi SR, Klusemann B (2019) A review of the application of machine learning and data mining approaches in continuum materials mechanics. Front Mater 6:110

29. Liu X, Tian S, Tao F, Yu W (2021) A review of artificial neural networks in the constitutive modeling of composite materials. Compos Part B 224:109152

30. Zhang P, Yin ZY, Jin YF (2021) State-of-the-art review of machine learning applications in constitutive modeling of soils. Arch Comput Methods Eng 28(5):3661–3686

31. Hart GL, Mueller T, Toher C, Curtarolo S (2021) Machine learning for alloys. Nature Rev Mater 6(8):730–755

32. Lourenço R, Andrade-Campos A, Georgieva P (2022) The use of machine-learning techniques in material constitutive modelling for metal forming processes. Metals 12(3):427

33. Goodfellow I, Bengio Y, Courville A, Bengio Y (2016) Deep learning. MIT Press, Cambridge

34. Raissi M, Perdikaris P, Karniadakis GE (2017) Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations. arXiv:1711.10561

35. Lagaris IE, Likas A, Fotiadis DI (1998) Artificial neural networks for solving ordinary and partial differential equations. IEEE Trans Neural Netw 9(5):987–1000

36. Baydin AG, Pearlmutter BA, Radul AA, Siskind JM (2018) Automatic differentiation in machine learning: a survey. J March Learn Res 18:1–43

37. Samaniego E, Anitescu C, Goswami S, Nguyen-Thanh VM, Guo H, Hamdia K et al (2020) An energy approach to the solution of partial differential equations in computational mechanics via machine learning: Concepts, implementation and applications. Comput Methods Appl Mech Eng 362:112790

38. Nguyen-Thanh VM, Zhuang X, Rabczuk T (2020) A deep energy method for finite deformation hyperelasticity. European J Mech -A/Solids 80:103874

39. Fuhg JN, Bouklas N (2021) The mixed deep energy method for resolving concentration features in finite strain hyperelasticity. arXiv:2104.09623

40. Fernández M, Rezaei S, Mianroodi JR, Fritzen F, Reese S (2020) Application of artificial neural networks for the prediction of interface mechanics: a study on grain boundary constitutive behavior. Adv Modeling Simul Eng Sci 7(1):1–27

41. Kirchdoerfer T, Ortiz M (2016) Data-driven computational mechanics. Comput Methods Appl Mech Eng 304:81–101

42. Ibañez R, Borzacchiello D, Aguado JV, Abisset-Chavanne E, Cueto E, Ladeveze P et al (2017) Data-driven non-linear elasticity: constitutive manifold construction and problem discretization. Comput Mech 60(5):813–826

43. Ibanez R, Abisset-Chavanne E, Aguado JV, Gonzalez D, Cueto E, Chinesta F (2018) A manifold learning approach to data-driven computational elasticity and inelasticity. Arch Comput Methods Eng 25(1):47–57

44. Eggersmann R, Stainier L, Ortiz M, Reese S (2021) Model-free data-driven computational mechanics enhanced by tensor voting. Comput Methods Appl Mech Eng 373:113499

45. Karapiperis K, Stainier L, Ortiz M, Andrade J (2021) Data-driven multiscale modeling in mechanics. J Mech Phys Solids 147:104239

46. Asteris PG, Roussis PC, Douvika MG (2017) Feed-forward neural network prediction of the mechanical properties of sandcrete materials. Sensors 17(6):1344

47. Nguyen LTK, Keip MA (2018) A data-driven approach to nonlinear elasticity. Comput Struct 194:97–115

48. Wang K, Sun W, Du Q (2021) A non-cooperative meta-modeling game for automated third-party calibrating, validating and falsifying constitutive laws with parallelized adversarial attacks. Comput Methods Appl Mech Eng 373:113514

49. Versino D, Tonda A, Bronkhorst CA (2017) Data driven modeling of plastic deformation. Comput Methods Appl Mech Eng 318:981–1004

50. Reimann D, Nidadavolu K, Vajragupta N, Glasmachers T, Junker P, Hartmaier A et al (2019) Modeling macroscopic material behavior with machine learning algorithms trained by micromechanical simulations. Front Mater 6:181

51. Crespo J, Latorre M, Montáns FJ (2017) WYPIWYG hyperelasticity for isotropic, compressible materials. Comput Mech 59(1):73–92

52. Romero X, Latorre M, Montáns FJ (2017) Determination of the WYPiWYG strain energy density of skin through finite element analysis of the experiments on circular specimens. Finite Elem Anal Desig 134:1–15

53. Latorre M, Montáns FJ (2020) Experimental data reduction for hyperelasticity. Comput Struct 232:105919

54. Tao F, Liu X, Du H, Yu W (2021) Learning composite constitutive laws via coupling Abaqus and deep neural network. Composite Struct 272:114137

55. Ruder S (2016) An overview of gradient descent optimization algorithms. arXiv:1609.04747

56. de Oca Montes, Zapiain D, Stewart JA, Dingreville R (2021) Accelerating phase-field-based microstructure evolution predictions via surrogate models trained by machine learning methods. Comput Mater 7(1):3

57. Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. Nature 323(6088):533–536

58. Elman JL (1990) Finding structure in time. Cognit Sci 14(2):179–211

59. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780

60. Cho K, Van Merriënboer B, Bahdanau D, Bengio Y (2014) On the properties of neural machine translation: Encoder-decoder approaches. arXiv:1409.1259

61. Chung J, Gulcehre C, Cho K, Bengio Y (2014) Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv:1412.3555

62. Fukushima K (1988) Neocognitron: a hierarchical neural network capable of visual pattern recognition. Neural Netw 1(2):119–130

63. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444

64. Cecen A, Dai H, Yabansu YC, Kalidindi SR, Song L (2018) Material structure-property linkages using three-dimensional convolutional neural networks. Acta Materialia 146:76–84

65. Durmaz AR, Müller M, Lei B, Thomas A, Britz D, Holm EA et al (2021) A deep learning approach for complex microstructure inference. Nature Commun 12(1):1–15

66. Thomas A, Durmaz A, Alam M, Gumbsch P, Sack H, Eberl C (2023) Materials fatigue prediction using graph neural networks on microstructure representations. preprint (Version 1) available at Research Square

67. Tenenbaum JB, Silva Vd, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. Science 290(5500):2319–2323

68. Cox T, Cox M (2001) Multidimensional Scaling. Monographs on Statistics and Applied Probability

69. Saul LK, Roweis ST (2003) Think globally, fit locally: unsupervised learning of low dimensional manifolds. J Mach Learn Res 4(Jun):119–155

70. Cayton L (2005) Algorithms for manifold learning. University of California at San Diego Tech Report. 12(1–17):1

71. Rumelhart DE, Hinton GE, Williams RJ (1985) Learning internal representations by error propagation. California Univ San Diego La Jolla Inst for Cognitive Science

72. Jung J, Yoon JI, Park HK, Jo H, Kim HS (2020) Microstructure design using machine learning generated low dimensional and continuous design space. Materialia 11:100690

73. Bronstein MM, Bruna J, Cohen T, Veličković P (2021) Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. arXiv:2104.13478

74. Agrawal A, Choudhary A (2016) Perspective: Materials informatics and big data: realization of the "fourth paradigm" of science in materials science. APL Mater 4(5):053208

75. Raabe D (1998) Computational materials science: the simulation of materials microstructures and properties. Wiley-VCH, Weinheim

76. Callister WD, Rethwisch DG (2018) Materials science and engineering: an introduction, vol 9. Wiley, New York

77. Hansen N (2004) Hall-Petch relation and boundary strengthening. Scripta Materialia 51(8):801–806

78. Roters F, Eisenlohr P, Hantcherli L, Tjahjanto DD, Bieler TR, Raabe D (2010) Overview of constitutive laws, kinematics, homogenization and multiscale methods in crystal plasticity finite-element modeling: Theory, experiments, applications. Acta Materialia 58(4):1152–1211

79. Eisenlohr P, Diehl M, Lebensohn RA, Roters F (2013) A spectral method solution to crystal elasto-viscoplasticity at finite strains. Int J Plast 46:37–53

80. Butz A, Wessel A, Pagenkopf J, Helm D (2019) Parameter identification of 3D yield functions based on a virtual material testing procedure. In: IOP Conference Series: Materials Science and Engineering. vol. 651. IOP Publishing; p. 012078

81. Wessel A, Morand L, Butz A, Helm D, Volk W (2021) A new machine learning based method for sampling virtual experiments and its effect on the parameter identification for anisotropic yield models. In: IOP Conference Series: Materials Science and Engineering. vol. 1157. IOP Publishing; p. 012026

82. Wessel A, Morand L, Butz A, Helm D, Volk W (2022) Machine learning-based sampling of virtual experiments within the full stress state to identify parameters of anisotropic yield models. arXiv:2211.00090

83. Butz A, Lossau S, Springub B, Roters F (2010) On the modeling of dual phase steels: microstructure-based simulation from the hot rolled sheet to the deep drawn component. Int J Mater Form 3(1):73–76

84. Smit RJ, Brekelmans WM, Meijer HE (1998) Prediction of the mechanical behavior of nonlinear heterogeneous systems by multi-level finite element modeling. Comput Methods Appl Mech Eng 155(1–2):181–192

85. Feyel F, Chaboche JL (2000) FE2 multiscale approach for modelling the elastoviscoplastic behaviour of long fibre SiC/Ti composite materials. Comput Methods Appl Mech Eng 183(3–4):309–330

86. Olson GB (1997) Computational design of hierarchically structured materials. Science 277(5330):1237–1242

87. Mozaffar M, Bostanabad R, Chen W, Ehmann K, Cao J, Bessa M (2019) Deep learning predicts path-dependent plasticity. Proc National Acad Sci 116(52):26414–26420

88. Pernot S, Lamarque CH (1999) Application of neural networks to the modelling of some constitutive laws. Neural Netw 12(2):371–392

89. Shin H, Pande G (2000) On self-learning finite element codes based on monitored response of structures. Comput Geotech 27(3):161–178

90. Ghaboussi J, Sidarta D (1998) New nested adaptive neural networks (NANN) for constitutive modeling. Comput Geotech 22(1):29–52

91. Furukawa T, Yagawa G (1998) Implicit constitutive modelling for viscoplasticity using neural networks. Int J Numer Methods Eng 43(2):195–219

92. Kupfer H, Hilsdorf HK, Rusch H (1969) Behavior of concrete under biaxial stresses. J Am Concr Inst 66:656–666

93. Sinha B, Gerstle KH, Tulin LG (1964) Stress-strain relations for concrete under cyclic loading. J Am Concr Inst 61:195–212

94. Chaboche JL (1989) Constitutive equations for cyclic plasticity and cyclic viscoplasticity. Int J Plast 5(3):247–302

95. Lefik M, Schrefler BA (2003) Artificial neural network as an incremental non-linear constitutive model for a finite element code. Comput Methods Appl Mech Eng 192(28–30):3265–3283

96. Al-Haik M, Hussaini M, Garmestani H (2006) Prediction of nonlinear viscoelastic behavior of polymeric composites using an artificial neural network. Int J Plast 22(7):1367–1392

97. Gorji MB, Mohr D (2019) Towards neural network models for describing the large deformation behavior of sheet metal. IOP Conf Series 651:012102 (**IOP Publishing**)

98. du Bos ML, Balabdaoui F, Heidenreich JN (2020) Modeling stress-strain curves with neural networks: a scalable alternative to the return mapping algorithm. Comput Mater Sci 178:109629

99. Jordan B, Gorji MB, Mohr D (2020) Neural network model describing the temperature-and rate-dependent stress-strain response of polypropylene. Int J Plast 135:102811

100. MacKay DJ (1992) Bayesian interpolation. Neural Comput 4(3):415–447

101. Rao K, Prasad Y (1995) Neural network approach to flow stress evaluation in hot deformation. J Mater Proc Technol 53(3–4):552–566

102. Lin Y, Zhang J, Zhong J (2008) Application of neural networks to predict the elevated temperature flow behavior of a low alloy steel. Comput Mater Sci 43(4):752–758

103. Ji G, Li F, Li Q, Li H, Li Z (2011) A comparative study on Arrhenius-type constitutive model and artificial neural network model to predict high-temperature deformation behaviour in Aermet100 steel. Mater Sci Eng 528(13–14):4774–4782

104. Sani SA, Ebrahimi G, Vafaeenezhad H, Kiani-Rashid AR (2018) Modeling of hot deformation behavior and prediction of flow stress in a magnesium alloy using constitutive equation and artificial neural network (ANN) model. J Magnes Alloys 6(2):134–144

105. Rezaei Ashtiani H, Shayanpoor A (2021) Hot deformation characterization of pure aluminum using artificial neural network (ANN) and processing map considering initial grain size. Metals Mater Int 27(12):5017–5033

106. Kumar S, Karmakar A, Nath SK (2021) Construction of hot deformation processing maps for 9Cr-1Mo steel through conventional and ANN approach. Materi Today Commun 26:101903

107. Li M, Liu X, Wu S, Zhang X (1998) Approach to constitutive relationships of a Ti-5AI-2Sn-2Zr-4Cr-4Mo alloy by artificial neural networks. Mater Sci Technol 14(2):136–138

108. Hodgson PD, Kong LX, Davies CH (1999) The prediction of the hot strength in steels with an integrated phenomenological and artificial neural network model. J Mater Proc Technol 87(1–3):131–138

109. Mandal S, Sivaprasad P, Venugopal S, Murthy K (2006) Constitutive flow behaviour of austenitic stainless steels under hot deformation: artificial neural network modelling to understand, evaluate and predict. Model Simul Mater Sci Eng 14(6):1053

110. Li HY, Wang XF, Wei DD, Hu JD, Li YH (2012) A comparative study on modified Zerilli-Armstrong, Arrhenius-type and artificial neural network models to predict high-temperature deformation behavior in T24 steel. Mater Sci Eng 536:216–222

111. Han Y, Qiao G, Sun J, Zou D (2013) A comparative study on constitutive relationship of as-cast 904L austenitic stainless steel during hot deformation based on Arrhenius-type and artificial neural network models. Comput Mater Sci 67:93–103

112. Gupta AK, Krishnamurthy HN, Singh Y, Prasad KM, Singh SK (2013) Development of constitutive models for dynamic strain aging regime in Austenitic stainless steel 304. Mater Design 45:616–627

113. Chun M, Biglou J, Lenard J, Kim J (1999) Using neural networks to predict parameters in the hot working of aluminum alloys. J Mater Proc Technol 86(1–3):245–251

114. Bobbili R, Ramakrishna B, Madhu V, Gogia A (2015) Prediction of flow stress of 7017 aluminium alloy under high strain rate compression at elevated temperatures. Defence Technol 11(1):93–98

115. Ashtiani HR, Shahsavari P (2016) A comparative study on the phenomenological and artificial neural network models to predict hot deformation behavior of AlCuMgPb alloy. J Alloys Compd 687:263–273

116. Li K, Pan Q, Li R, Liu S, Huang Z, He X (2019) Constitutive modeling of the hot deformation behavior in 6082 aluminum alloy. J Mater Eng Perform 28(2):981–994

117. Sun Y, Zeng W, Zhao Y, Qi Y, Ma X, Han Y (2010) Development of constitutive relationship model of Ti600 alloy using artificial neural network. Comput Mater Sci 48(3):686–691

118. Tofallis C (2015) A better measure of relative prediction accuracy for model selection and model estimation. J Operational Res Soc 66(8):1352–1362

119. Hopkinson BX (1914) A method of measuring the pressure produced in the detonation of high, explosives or by the impact of bullets. Philoso Trans Royal Soc London Series Contain Papersa Math Phys Character 213(497–508):437–456

120. Johnson GR (1983) A constitutive model and data for materials subjected to large strains, high strain rates, and high temperatures. Proc 7th Inf Sympo Ballistics. 541–547

121. Hashash Y, Jung S, Ghaboussi J (2004) Numerical implementation of a neural network based material model in finite element analysis. Int J Numeri Methods Eng 59(7):989–1005

122. Jung S, Ghaboussi J (2006) Neural network constitutive model for rate-dependent materials. Comput Struct 84(15–16):955–963

123. Yun GJ, Ghaboussi J, Elnashai AS (2008) A new neural network-based model for hysteretic behavior of materials. Int J Numer Methods Eng 73(4):447–469

124. Stoffel M, Bamer F, Markert B (2018) Artificial neural networks and intelligent finite elements in non-linear structural mechanics. Thin-Walled Struct 131:102–106

125. Huang D, Fuhg JN, Weißenfels C, Wriggers P (2020) A machine learning based plasticity model using proper orthogonal decomposition. Comput Methods Appl Mech Eng 365:113008

126. Zhang A, Mohr D (2020) Using neural networks to represent von Mises plasticity with isotropic hardening. Int J Plast 132:102732

127. Jang DP, Fazily P, Yoon JW (2021) Machine learning-based constitutive model for J2-plasticity. Int J Plast 138:102919

128. Le B, Yvonnet J, He QC (2015) Computational homogenization of nonlinear elastic materials using neural networks. Int J Numeri Methods Eng 104(12):1061–1084

129. Ghaboussi J, Pecknold DA, Zhang M, Haj-Ali RM (1998) Auto-progressive training of neural network constitutive models. Int J Numer Methods Eng 42(1):105–126

130. Sidarta D, Ghaboussi J (1998) Constitutive modeling of geomaterials from non-uniform material tests. Comput Geotech 22(1):53–71

131. Yang H, Guo X, Tang S, Liu WK (2019) Derivation of heterogeneous material laws via data-driven principal component expansions. Comput Mech 64(2):365–379

132. Rasmussen C, Williams C (2006) Gaussian processes for machine learning. MIT Press, Cambridge

133. Korelc J, Wriggers P (2016) Automation of Finite Element Methods. Springer, Berlin

134. Ali U, Muhammad W, Brahme A, Skiba O, Inal K (2019) Application of artificial neural networks in micromechanics for polycrystalline metals. Int J Plast 120:205–219

135. Mianroodi JR, Siboni NH, Raabe D (2021) Teaching solid mechanics to artificial intelligence-a fast solver for heterogeneous materials. Comput Mater 7(1):1–10

136. Salmenjoki H, Alava MJ, Laurson L (2018) Machine learning plastic deformation of crystals. Nature Commun 9(1):1–7

137. Ling J, Jones R, Templeton J (2016) Machine learning strategies for systems with invariance properties. J Comput Phys 318:22–35

138. Jones R, Templeton J, Sanders C, Ostien J (2018) Machine learning models of plastic flow based on representation theory. Comput Modeling Eng Sci 117(3):309–342

139. Hashash Y, Marulanda C, Ghaboussi J, Jung S (2003) Systematic update of a deep excavation model using field performance data. Comput Geotech 30(6):477–488

140. Yun GJ, Ghaboussi J, Elnashai AS (2008) Self-learning simulation method for inverse nonlinear modeling of cyclic behavior of connections. Comput Methods Appl Mech Eng 197(33–40):2836–2857

141. Yun GJ, Ghaboussi J, Elnashai AS (2008) A design-variable-based inelastic hysteretic model for beam-column connections. Earthquake Eng Struct Dynam. 37(4):535–555

142. Liu X, Tao F, Du H, Yu W, Xu K (2020) Learning nonlinear constitutive laws using neural network models based on indirectly measurable data. J Appl Mech 87(8):081003

143. Li X, Roth CC, Mohr D (2019) Machine-learning based temperature-and rate-dependent plasticity model: application to analysis of fracture experiments on DP steel. Int J Plast 118:320–344

144. Huang DZ, Xu K, Farhat C, Darve E (2020) Learning constitutive relations from indirect observations using deep neural networks. J Comput Phys 416:109491

145. Xu K, Tartakovsky AM, Burghardt J, Darve E (2020) Inverse modeling of viscoelasticity materials using physics constrained learning. arXiv:2005.04384

146. Xu K, Huang DZ, Darve E (2021) Learning constitutive relations using symmetric positive definite neural networks. J Comput Phys 428:110072

147. Liu X, Tao F, Yu W (2020) A neural network enhanced system for learning nonlinear constitutive law and failure initiation criterion of composites using indirectly measurable data. Compos Struct 252:112658

148. Pandya KS, Roth CC, Mohr D (2020) Strain rate and temperature dependent fracture of aluminum alloy 7075: experiments and neural network modeling. Int J Plast 135:102788

149. Hartmaier A (2020) Data-oriented constitutive modeling of plasticity in metals. Materials 13(7):1600

150. Shoghi R, Hartmaier A (2022) Optimal data-generation strategy for machine learning yield functions in anisotropic plasticity. Front Mater 9:868248

151. Oeser M, Freitag S (2009) Modeling of materials with fading memory using neural networks. Int J Numer Methods Eng 78(7):843–862

152. Graf W, Freitag S, Kaliske M, Sickert JU (2010) Recurrent neural networks for uncertain time-dependent structural behavior. Comput-Aided Civil Infrastruct Eng 25(5):322–323

153. Graf W, Freitag S, Sickert JU, Kaliske M (2012) Structural analysis with fuzzy data and neural network based material description. Comput-Aided Civil Infrastruct Eng 27(9):640–654

154. Freitag S, Graf W, Kaliske M (2011) Recurrent neural networks for fuzzy data. Integr Comput-Aided Eng 18(3):265–280

155. Freitag S, Graf W, Kaliske M (2013) A material description based on recurrent neural networks for fuzzy data and its application within the finite element method. Comput Struct 124:29–37

156. Möller B, Graf W, Beer M (2000) Fuzzy structural analysis using α-level optimization. Comput Mech 26(6):547–565

157. Rao SS, Sawyer JP (1995) Fuzzy finite element approach for analysis of imprecisely defined systems. AIAA J 33(12):2364–2370

158. Gorji MB, Mozaffar M, Heidenreich JN, Cao J, Mohr D (2020) On the potential of recurrent neural networks for modeling path dependent plasticity. J Mech Phys Solids 143:103972

159. Zhang P, Yin ZY, Jin YF, Ye GL (2020) An AI-based model for describing cyclic characteristics of granular materials. Int J Numer Anal Methods Geomechanics 44(9):1315–1335

160. Bonatti C, Mohr D (2021) One for all: universal material model based on minimal state-space neural networks. Sci Adv 7(26):eabf3658

161. Abueidda DW, Koric S, Sobh NA, Sehitoglu H (2021) Deep learning for plasticity and thermo-viscoplasticity. Int J Plast 136:102852

162. Zopf C, Kaliske M (2017) Numerical characterisation of uncured elastomers by a neural network based approach. Comput Struct 182:504–525

163. Miehe C, Göktepe S, Lulei F (2004) A micro-macro approach to rubber-like materials-part I: the non-affine micro-sphere model of rubber elasticity. J Mech Phys Solids 52(11):2617–2660

164. Bonatti C, Mohr D (2022) On the importance of self-consistency in recurrent neural network models representing elasto-plastic solids. J Mech Phys Solids 158:104697

165. Ghavamian F, Simone A (2019) Accelerating multiscale finite element simulations of history-dependent materials using a recurrent neural network. Comput Methods Appl Mech Eng 357:112594

166. Logarzo HJ, Capuano G, Rimoli JJ (2021) Smart constitutive laws: inelastic homogenization through machine learning. Comput Methods Appl Mech Eng 373:113482

167. Wu L, Kilingar NG, Noels L et al (2020) A recurrent neural network-accelerated multi-scale model for elasto-plastic heterogeneous materials subjected to random cyclic and non-proportional loading paths. Comput Methods Appl Mech Eng 369:113234

168. Heider Y, Wang K, Sun W (2020) SO (3)-invariance of informed-graph-based deep neural network for anisotropic elastoplastic materials. Comput Methods Appl Mech Eng 363:112875

169. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN et al (2017) Attention is all you need. Adv Neural Inf Proc Syst. https://doi.org/10.48550/arXiv.1706.03762

170. Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, et al (2020) Transformers: State-of-the-art natural language processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations; p. 38–45

171. Wen Q, Zhou T, Zhang C, Chen W, Ma Z, Yan J, et al (2022) Transformers in time series: a survey. arXiv preprint arXiv:2202.07125

172. Yang C, Kim Y, Ryu S, Gu GX (2020) Prediction of composite microstructure stress-strain curves using convolutional neural networks. Mater Design 189:108509

173. Frankel AL, Jones RE, Alleman C, Templeton JA (2019) Predicting the mechanical response of oligocrystals with deep learning. Comput Mater Sci 169:109099

174. Henkes A, Caylak I, Mahnken R (2021) A deep learning driven pseudospectral PCE based FFT homogenization algorithm for complex microstructures. Comput Methods Appl Mech Eng 385:114070

175. Vlassis NN, Ma R, Sun W (2020) Geometric deep learning for computational mechanics part i: anisotropic hyperelasticity. Comput Methods Appl Mech Eng 371:113299

176. Czarnecki WM, Osindero S, Jaderberg M, Swirszcz G, Pascanu R (2017) Sobolev training for neural networks. Adv Neural Inf Proc Syst. 30

177. Weber P, Geiger J, Wagner W (2021) Constrained neural network training and its application to hyperelastic material modeling. Comput Mech 68(5):1179–1204

178. Hernandez Q, Badias A, Gonzalez D, Chinesta F, Cueto E (2021) Deep learning of thermodynamics-aware reduced-order models from data. Comput Methods Appl Mech Eng 379:113763

179. Haghighat E, Abouali S, Vaziri R (2022) Constitutive model characterization and discovery using physics-informed deep learning. arXiv:2203.09789

180. Raissi M, Perdikaris P, Karniadakis GE (2019) Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J Comput Phys 378:686–707

181. Shen Y, Chandrashekhara K, Breig W, Oliver L (2004) Neural network based constitutive model for rubber material. Rubber Chem Technol 77(2):257–277

182. Liang G, Chandrashekhara K (2008) Neural network based constitutive model for elastomeric foams. Eng Struct 30(7):2002–2011

183. Masi F, Stefanou I, Vannucci P, Maffi-Berthier V (2020) Material modeling via Thermodynamics-based artificial neural networks. In: Workshop on Joint Structures and Common Foundations of Statistical Physics, Information Geometry and Inference for Learning. Springer; p. 308–329

184. Masi F, Stefanou I, Vannucci P, Maffi-Berthier V (2021) Thermodynamics-based artificial neural networks for constitutive modeling. J Mech Phys Solids 147:104277

185. Linka K, Hillgärtner M, Abdolazizi KP, Aydin RC, Itskov M, Cyron CJ (2021) Constitutive artificial neural networks: a fast and general approach to predictive data-driven constitutive modeling by deep learning. J Comput Phys 429:110010

186. Vlassis NN, Sun W (2021) Sobolev training of thermodynamic-informed neural networks for interpretable elasto-plasticity models with level set hardening. Comput Methods Appl Mech Eng 377:113695

187. Vlassis NN, Sun W (2022) Component-based machine learning paradigm for discovering rate-dependent and pressure-sensitive level-set plasticity models. J Appl Mech 89(2):021003

188. Fernández M, Jamshidian M, Böhlke T, Kersting K, Weeger O (2021) Anisotropic hyperelastic constitutive models for finite deformations combining material theory and data-driven approaches with application to cubic lattice metamaterials. Comput Mech 67(2):653–677

189. Liu Z, Wu C, Koishi M (2019) A deep material network for multiscale topology learning and accelerated nonlinear modeling of heterogeneous materials. Comput Methods Appl Mech Eng 345:1138–1168

190. Liu Z, Wu C (2019) Exploring the 3D architectures of deep material network in data-driven multiscale mechanics. J Mech Phys Solids 127:20–46

191. Liu Z (2021) Cell division in deep material networks applied to multiscale strain localization modeling. Comput Methods Appl Mech Eng 384:113914

192. Liu Z (2020) Deep material network with cohesive layers: multi-stage training and interfacial failure analysis. Comput Methods Appl Mech Engi. 363:112913

193. Liu Z, Wei H, Huang T, Wu C (2020) Intelligent multiscale simulation based on process-guided composite database. In: 16th International LS-DYNA Users Conference

194. Gajek S, Schneider M, Böhlke T (2021) An FE-DMN method for the multiscale analysis of short fiber reinforced plastic components. Comput Methods Appl Mech Eng 384:113952

195. Man H, Furukawa T (2011) Neural network constitutive modelling for non-linear characterization of anisotropic materials. Int J Numer Methods Eng 85(8):939–957

196. Deng L (2012) The mnist database of handwritten digit images for machine learning research [best of the web]. IEEE Signal Proc Mag 29(6):141–142

197. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE; p. 248–255

198. Lejeune E (2020) Mechanical MNIST: a benchmark dataset for mechanical metamodels. Extreme Mech Lett 36:100659

199. Dunn A, Wang Q, Ganose A, Dopp D, Jain A (2020) Benchmarking materials property prediction methods: the matbench test set and automatminer reference algorithm. Comput Mater 6(1):1–10

200. Settles B (2009) Active learning literature survey. University of Wisconsin-Madison Department of Computer Sciences

201. Rocha I, Kerfriden P, van der Meer F (2021) On-the-fly construction of surrogate constitutive models for concurrent multiscale mechanical analysis through probabilistic machine learning. J Comput Phys 9:100083

202. Fritzen F, Fernández M, Larsson F (2019) On-the-fly adaptivity for nonlinear twoscale simulations using artificial neural networks and reduced order modeling. Front Mater 6:75

203. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A et al (2016) The FAIR guiding principles for scientific data management and stewardship. Sci Data 3(1):1–9

204. Kimmig J, Zechel S, Schubert US (2021) Digital transformation in materials science: a paradigm change in material's development. Adv Mater 33(8):2004940

205. Himanen L, Geurts A, Foster AS, Rinke P (2019) Data-driven materials science: status, challenges, and perspectives. Adv Sci 6:1900808

206. Ekaputra F, Sabou M, Serral Asensio E, Kiesling E, Biffl S (2017) Ontology-based data integration in multi-disciplinary engineering environments: a review. Open J Inf Sys 4(1):1–26

207. Alam M, Dittmann F, Niebel M, Lehmann J, Dessì D, Morgado JF, et al (2020) Towards digitizing physical entities in materials science. FIZ Karlsruhe, white paper

208. Batra R, Pilania G, Uberuaga BP, Ramprasad R (2019) Multi-fidelity information fusion with machine learning: a case study of dopant formation energies in hafnia. ACS Appld Mater Interfaces. 11(28):24906–24918

209. Von Rueden L, Mayer S, Garcke J, Bauckhage C, Schuecker J (2019) Informed machine learning–towards a taxonomy of explicit integration of knowledge into machine learning. arXiv:1903.12394

210. Klein DK, Fernández M, Martin RJ, Neff P, Weeger O (2022) Polyconvex anisotropic hyperelasticity with neural networks. J Mech Phys Solids 159:104703

211. von Kurnatowski M, Schmid J, Link P, Zache R, Morand L, Kraft T et al (2021) Compensating data shortages in manufacturing with monotonicity knowledge. MDPI Algorithms 14(12):345

212. Du M, Liu N, Hu X (2019) Techniques for interpretable machine learning. Commun ACM 63(1):68–77

213. Koeppe A, Bamer F, Selzer M, Nestler B, Markert B (2021) Explainable artificial intelligence for mechanics: physics-informing neural networks for constitutive models. arXiv:2104.10683

214. Rocha I, Kerfriden P, van der Meer F (2020) Micromechanics-based surrogate models for the response of composites: a critical comparison between a classical mesoscale constitutive model, hyper-reduction and neural networks. Eur J Mech-A/Solids 82:103995

215. Li QJ, Cinbiz MN, Zhang Y, He Q, Beausoleil G II, Li J (2023) Robust deep learning framework for constitutive relations modeling. Acta Materialia 254:118959