Acta Biomaterialia 160 (2023) 134–151

Contents lists available at ScienceDirect

Acta Biomaterialia

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

Full length article

Automated model discovery for human brain using Constitutive Artificial Neural Networks

Kevin Linka, Sarah R. St. Pierre, Ellen Kuhl*

Department of Mechanical Engineering, Stanford University, Stanford, California, USA

ARTICLE INFO

Article history:
Received 8 November 2022
Revised 24 January 2023
Accepted 26 January 2023
Available online 2 February 2023

Keywords:
Mechanics of the brain
Automated science
Constitutive artificial neural networks
Constitutive modeling
Thermodynamics
Machine learning

ABSTRACT

The brain is our softest and most vulnerable organ, and understanding its physics is a challenging but significant task. Throughout the past decade, numerous competing models have emerged to characterize its response to mechanical loading. However, selecting the best constitutive model remains a heuristic process that strongly depends on user experience and personal preference. Here we challenge the conventional wisdom to *first* select a constitutive model and *then* fit its parameters to data. Instead, we propose a new strategy that *simultaneously* discovers both model and parameters. We integrate more than a century of knowledge in thermodynamics and state-of-the-art machine learning to build a Constitutive Artificial Neural Network that enables automated model discovery. Our design paradigm is to reverse engineer the network from a set of functional building blocks that are, by design, a generalization of popular constitutive models, including the neo Hookean, Blatz Ko, Mooney Rivlin, Demiray, Gent, and Holzapfel models. By constraining input, output, activation functions, and architecture, our network a priori satisfies thermodynamic consistency, objectivity, symmetry, and polyconvexity. We demonstrate that–out of more than 4000 models–our network autonomously discovers the model and parameters that best characterize the behavior of human gray and white matter under tension, compression, and shear. Importantly, our network weights translate naturally into physically meaningful parameters, such as shear moduli of 1.82kPa, 0.88kPa, 0.94kPa, and 0.54kPa for the cortex, basal ganglia, corona radiata, and corpus callosum. Our results suggest that Constitutive Artificial Neural Networks have the potential to induce a paradigm shift in soft tissue modeling, from user-defined model selection to automated model discovery. Our source code, data, and examples are available at https://github.com/LivingMatterLab/CANN.

Statement of significance

Human brain is ultrasoft, difficult to test, and challenging to model. Numerous competing constitutive models exist, but selecting the best model remains a matter of personal preference. Here we automate the process of model selection. We formulate the problem of autonomous model discovery as a neural network and capitalize on the powerful optimizers in deep learning. However, rather than using a conventional neural network, we reverse engineer our own Constitutive Artificial Neural Network from a set of modular building blocks, which we rationalize from common constitutive models. When trained with tension, compression, and shear experiments of gray and white matter, our network simultaneously discovers both model and parameters that describes the data better than any existing invariant-based model. Our network could induce a paradigm shift from user-defined model selection to automated model discovery.

1. Motivation

Traumatic brain injury is a major cause of death and disability worldwide [1], with a global annual incidence of 69 million [2]. In the United States alone, 176 people die each day from traumatic brain injury, and every nine seconds, someone sustains a

* Corresponding author.
E-mail address: ekuhl@stanford.edu (E. Kuhl).

https://doi.org/10.1016/j.actbio.2023.01.055

new injury to the brain. Fortunately, not all concussions are life-threatening; yet, more than 5 million Americans are living with brain-injury-related disabilities and need long-term assistance in their everyday life [3]. Without a doubt, understanding the mechanics of brain injury is a challenging but significant task [4]. Throughout the past decade, scientists across the world have made significant strides in testing, modeling, and simulating the human brain [5–13]. However, because of its ultrasoft nature, the results vary greatly, both qualitatively and quantitatively [14]. This has resulted in a wide selection of competing constitutive models for gray and white matter tissue, without any real guidance which model to choose [7,15–17]. Throughout this manuscript, we ask how we can select the best constitutive model for the human brain, whether the current existing models are really the best, and if not, how we can systematically search and find a better model.

In machine learning, the process of finding relationships in complex data is known as *automated model discovery* [18–21]. The preface automated implies that model discovery can be performed entirely without human interaction [22,23]. Neural networks have emerged as a powerful strategy to discover constitutive models from large data, even in the complete absence of knowledge about the underlying physics [24]. However, classical neural networks ignore more than a century of research in constitutive modeling [25]: They violate thermodynamic constraints [26], neglect generally accepted physical principles [27], and fail to predict the behavior outside the training regime [20,28]. In essence, neural networks perform excellently at fitting a complex function to big data, but they are not interpretable; they teach us nothing about the underlying physics [29]. So, really, what we are looking for is a strategy to autonomously *discover a physically motivated model*.

Two successful but fundamentally different strategies have emerged to integrate physics into neural network models: *Physics-Informed Neural Networks* that add physics-based equations as additional terms to the loss function [30] and *Constitutive Artificial Neural Networks* that explicitly modify the network input, output, and architecture to hardwire physics-based constraints into the network design [31]. The first type of networks is more general and works well for ordinary [32] or partial [29] differential equations, whereas the second type is specifically tailored towards constitutive equations [33]. Constitutive Artificial Neural Networks, with strain invariants as input and free energy functions as output, were first proposed for rubber-like materials almost two decades ago [34], and have recently regained attention in the constitutive modeling community [27,35–37]. They are now also increasingly recognized in the soft tissue biomechanics community with applications to skin [38], blood clots [39], arteries [33,40], and myocardial tissue [39]. A common feature of all these neural networks is to use multiple hidden layers, generic activation functions, and several hundreds, if not thousands of unknowns. To no surprise, they perform well at interpolating non-linear stress-stretch relations from tension, compression, or shear experiments. However, one critical limitation remains: the lack of an intuitive interpretation of the model and its parameters [36].

Here, instead of using a generic neural network architecture, we *reverse-engineer* a new family of Constitutive Artificial Neural Networks from *constitutive building blocks* that are, by design, a generalization of widely used and commonly accepted constitutive models, including the neo Hookean [41], Blatz Ko [42], Mooney Rivlin [43,44], Demiray [45], Gent [46], and Holzapfel [47] models. As such, their network weights naturally translate into material parameters with standard physical units and a clear physical interpretation [20]. We train our network with tension, compression, and shear tests from the human cortex, basal ganglia, corona radiata, and corpus callosum [7,8,14] and demonstrate that it can simultaneously *discover both model and parameters* that best describe the data. Beyond automated model discovery, we show that we can also use our network for the *parameter identification* of existing constitutive models. By systematically comparing the goodness of fit of the different models, trained with the different experiments, we not only discover the model that best describes the experiments, but we also *discover the experiment* that best informs the models. Designing informative experiments is particularly significant for human brain tissue, for which fresh samples are rare, challenging to preserve, and difficult to mount and test [14].

2. Methods

2.1. Kinematics

To characterize the deformation of the sample we want to test, we introduce the deformation map $\boldsymbol{\varphi}$ that maps material particles $\boldsymbol{X}$ from the undeformed configuration to particles, $\boldsymbol{x} = \boldsymbol{\varphi}(\boldsymbol{X})$, in the deformed configuration [48]. We describe relative deformations within the sample using the deformation gradient $\boldsymbol{F}$, the gradient of the deformation map $\boldsymbol{\varphi}$ with respect to the undeformed coordinates $\boldsymbol{X}$, and its Jacobian J,

$$\boldsymbol{F} = \nabla_{\boldsymbol{X}} \boldsymbol{\varphi} \qquad \text{with} \qquad J = \det(\boldsymbol{F}) > 0\,. \tag{1}$$

Multiplying $\boldsymbol{F}$ with its transpose $\boldsymbol{F}^{\mathrm{t}}$ introduces the symmetric right Cauchy Green deformation tensor $\boldsymbol{C}$,

$$\boldsymbol{C} = \boldsymbol{F}^{\mathrm{t}} \cdot \boldsymbol{F}\,. \tag{2}$$

In the undeformed state, both tensors are identical to the unit tensor, $\boldsymbol{F} = \boldsymbol{I}$ and $\boldsymbol{C} = \boldsymbol{I}$, and the Jacobian is one, $J = 1$. A Jacobian smaller than one, $0 < J < 1$, denotes compression and a Jacobian larger than one, $1 < J$, denotes extension.

Isotropy. To characterize an *isotropic* material, we introduce the three principal invariants I_1, I_2, I_3 and their derivatives $\partial_{\boldsymbol{F}} I_1$, $\partial_{\boldsymbol{F}} I_2$, $\partial_{\boldsymbol{F}} I_3$,

$$\begin{aligned} I_1 &= \boldsymbol{F} : \boldsymbol{F} & & & \partial_{\boldsymbol{F}} I_1 &= 2\,\boldsymbol{F} \\ I_2 &= \frac{1}{2}\,[I_1^2 - [\boldsymbol{F}^{\mathrm{t}} \cdot \boldsymbol{F}] : [\boldsymbol{F}^{\mathrm{t}} \cdot \boldsymbol{F}]] & & \text{with} & \partial_{\boldsymbol{F}} I_2 &= 2\,[I_1 \boldsymbol{F} - \boldsymbol{F} \cdot \boldsymbol{F}^{\mathrm{t}} \cdot \boldsymbol{F}] \\ I_3 &= \det(\boldsymbol{F}^{\mathrm{t}} \cdot \boldsymbol{F}) = J^2 & & & \partial_{\boldsymbol{F}} I_3 &= 2\,I_3\,\boldsymbol{F}^{-\mathrm{t}}\,. \end{aligned} \tag{3}$$

In the undeformed state, $\boldsymbol{F} = \boldsymbol{I}$, the three invariants are equal to three and one, $I_1 = 3$, $I_2 = 3$, and $I_3 = 1$.

Perfect incompressibility. For *isotropic, perfectly incompressible* materials, the third invariant always remains identical to one, $I_3 = J^2 = 1$. This reduces the set of invariants to two, I_1 and I_2.

2.2. Constitutive equations

In solid mechanics, constitutive equations are tensor-valued tensor functions that define the relation between a stress measure, for example the Piola or nominal stress, $\boldsymbol{P} = \lim_{\mathrm{d}\boldsymbol{A} \to \boldsymbol{0}} (\mathrm{d}\boldsymbol{f}/\mathrm{d}\boldsymbol{A})$, the force $\mathrm{d}\boldsymbol{f}$ per undeformed area $\mathrm{d}\boldsymbol{A}$, and a deformation measure, for example the deformation gradient $\boldsymbol{F}$ [49,50],

$$\boldsymbol{P} = \boldsymbol{P}(\boldsymbol{F})\,. \tag{4}$$

At this point, we could use an arbitrary neural network to learn the functional relation between $\boldsymbol{P}$ and $\boldsymbol{F}$ and many neural networks in the literature do exactly that [26,51,52]. However, the functions $\boldsymbol{P}(\boldsymbol{F})$ that we learn through this approach generally violate widely-accepted thermodynamical constraints and their parameters have no physical meaning [35]. For moderate amounts of data, standard neural networks are also associated with a high risk of overfit-

ting [36]. Our objective is therefore to build a Constitutive Artificial Neural Network that a priori satisfies thermodynamic constraints and introduces parameters with a clear physical interpretation, while, at the same time, limiting the space of admissible functions to prevent overfitting when available data are sparse.

Thermodynamic consistency. First, we ensure *thermodynamic consistency* and guarantee that the Piola stress $\boldsymbol{P}$ inherently satisfies the second law of thermodynamics, the dissipation inequality [53], $\mathcal{D} = \boldsymbol{P} : \dot{\boldsymbol{F}} - \dot{\psi}(\boldsymbol{F}) \geq 0$, where $\mathcal{D}$ is the dissipation and ψ is the Helmholtz free energy with $\dot{\psi} = \partial\psi(\boldsymbol{F})/\partial\boldsymbol{F} : \dot{\boldsymbol{F}}$. For *hyperelastic* or Green-elastic materials with $\mathcal{D} \doteq 0$, the entropy inequality directly defines the Piola stress [65],

$$\boldsymbol{P} = \frac{\partial\psi(\boldsymbol{F})}{\partial\boldsymbol{F}} \,. \tag{5}$$

This implies that, rather than approximating the nine stress components $\boldsymbol{P}(\boldsymbol{F})$ as nine generic functions of the nine components of the deformation gradient $\boldsymbol{F}$, the network can simply approximate the scalar-valued free energy function $\psi(\boldsymbol{F})$ from which we derive the stress $\boldsymbol{P}$ in a post-processing step. Satisfying thermodynamic consistency according to Eq. (5) directly affects the *output* of the neural network.

Material objectivity and frame indifference. Second, we constrain the choice of the free energy function ψ to satisfy *material objectivity* or *frame indifference* and ensure that the constitutive laws do not depend on the external frame of reference [54]. To a priori satisfy this constraint, we require that the arguments of the free energy function are independent of rotations, and must be functions of the right Cauchy Green deformation tensor $\boldsymbol{C}$ [50],

$$\boldsymbol{P} = \frac{\partial\psi(\boldsymbol{C})}{\partial\boldsymbol{F}} = \frac{\partial\psi(\boldsymbol{C})}{\partial\boldsymbol{C}} : \frac{\partial\boldsymbol{C}}{\partial\boldsymbol{F}} = 2\,\boldsymbol{F}\cdot\frac{\partial\psi(\boldsymbol{C})}{\partial\boldsymbol{C}} \,. \tag{6}$$

This implies that, rather than using the nine independent components of the deformation gradient $\boldsymbol{F}$ as input, we constrain the input to the six independent components of the symmetric right Cauchy Green deformation tensor, $\boldsymbol{C} = \boldsymbol{F}^{\mathrm{t}}\cdot\boldsymbol{F}$. Satisfying material objectivity according to Eq. (6) directly affects the *input* of the neural network.

Material symmetry and isotropy. Third, we can further constrain the choice of the free energy function ψ to include *material symmetry* and assume that the material response remains unchanged under transformations of the reference configuration. Here we consider the special case of *isotropy*, for which the free energy function is a function of the strain *invariants*, $\psi(I_1, I_2, I_3)$, and the Piola stress takes the following explicit representation,

$$\begin{aligned}\boldsymbol{P} &= \frac{\partial\psi(I_1,I_2,I_3)}{\partial\boldsymbol{F}} = \frac{\partial\psi}{\partial I_1}\frac{\partial I_1}{\partial\boldsymbol{F}} + \frac{\partial\psi}{\partial I_2}\frac{\partial I_2}{\partial\boldsymbol{F}} + \frac{\partial\psi}{\partial I_3}\frac{\partial I_3}{\partial\boldsymbol{F}} \\ &= 2\left[\frac{\partial\psi}{\partial I_1} + I_1\frac{\partial\psi}{\partial I_2}\right]\boldsymbol{F} - 2\frac{\partial\psi}{\partial I_2}\boldsymbol{F}\cdot\boldsymbol{F}^{\mathrm{t}}\cdot\boldsymbol{F} + 2I_3\frac{\partial\psi}{\partial I_3}\boldsymbol{F}^{-\mathrm{t}} \,.\end{aligned} \tag{7}$$

This implies that, rather than using the six independent components of the symmetric right Cauchy Green deformation tensor $\boldsymbol{C}$ as input, we constrain the input to our set of three invariants I_1, I_2, I_3. Considering materials with known symmetry classes according to Eqs. (7) directly affects, and ideally reduces, the *input* of the neural network.

Perfect incompressibility. Fourth, we can further constrain the choice of the free energy function ψ for the special case of *perfect incompressibility* for which the Jacobian remains constant and equal to one, $I_3 = J^2 = 1$. The condition of perfect incompressibility implies that Eq. (7) simplifies to an expression in terms of only the first two invariants I_1 and I_2, corrected by the pressure term, $-p\boldsymbol{F}^{-\mathrm{t}}$, where $p = -\frac{1}{3}\boldsymbol{P} : \boldsymbol{F}$ is the hydrostatic pressure that we determine from the boundary conditions,

$$\begin{aligned}\boldsymbol{P} &= \frac{\partial\psi(I_1,I_2,I_3)}{\partial\boldsymbol{F}} = \frac{\partial\psi}{\partial I_1}\frac{\partial I_1}{\partial\boldsymbol{F}} + \frac{\partial\psi}{\partial I_2}\frac{\partial I_2}{\partial\boldsymbol{F}} - p\boldsymbol{F}^{-\mathrm{t}} \\ &= 2\left[\frac{\partial\psi}{\partial I_1} + I_1\frac{\partial\psi}{\partial I_2}\right]\boldsymbol{F} - 2\frac{\partial\psi}{\partial I_2}\boldsymbol{F}\cdot\boldsymbol{F}^{\mathrm{t}}\cdot\boldsymbol{F} - p\boldsymbol{F}^{-\mathrm{t}} \,.\end{aligned} \tag{8}$$

This implies that, rather than using the set of three invariants, I_1, I_2, I_3, as input, we reduce the input to a set of only two invariants, I_1 and I_2. Considering materials with perfect incompressibility according to Eq. (8) reduces the *input* of the neural network.

Physically reasonable constitutive restrictions. Fifth, we can further constrain the functional form of the free energy ψ by including additional constitutive restrictions that are both physically reasonable and mathematically convenient [48]: (i) The free energy ψ is *non-negative* for all deformation states$\boldsymbol{F}$,

$$\psi(\boldsymbol{F}) \geq 0 \quad \forall \quad \boldsymbol{F} \,. \tag{9}$$

(ii) The free energy ψ and the stress $\boldsymbol{P}$ are *zero* in the reference configuration, $\boldsymbol{F} = \boldsymbol{I}$,

$$\psi(\boldsymbol{F}) \doteq 0 \quad \text{and} \quad \boldsymbol{P}(\boldsymbol{F}) \doteq \boldsymbol{0} \quad \text{for} \quad \boldsymbol{F} = \boldsymbol{I} \,. \tag{10}$$

(iii) The free energy ψ is *infinite* for infinite compression, $J \to 0$, and infinite expansion, $J \to \infty$,

$$\psi(\boldsymbol{F}) \to \infty \quad \text{for} \quad J \to 0 \quad \text{or} \quad J \to \infty \,. \tag{11}$$

To facilitate a stress-free reference configuration according to Eq. (10), instead of using the invariants I_1, I_2, I_3 themselves as input, we use their deviation from the energy- and stress-free reference state, $[\,I_1 - 3\,]$, $[\,I_2 - 3\,]$, $[\,I_3 - 1\,]$, as input. In addition, from all possible activation functions, we select activation functions that a priori comply with conditions (i), (ii), and (iii). Satisfying physical considerations according to Eqs. (9), (10), and (11) directly affects the *activation functions* of the neural network.

Polyconvexity. Sixth, to guide the selection of the functional forms for the free energy function ψ, and ultimately the selection of appropriate activation functions, we consider *polyconvexity* requirements [55]. From the general representation theorem we know that in its most generic form, the free energy of an isotropic material can be expressed as an infinite series of products of powers of the invariants [56], $\psi(I_1, I_2, I_3) = \sum_{j,k,l=0}^{\infty} a_{jk}\,[I_1 - 3]^j[I_2 - 3]^k[I_3 - 1]^l$, where a_{jkl} are material constants. Importantly, mixed products of convex functions are generally not convex, and it is easier to show that the sum of specific convex subfunction usually is [57]. This motivates a special subclass of free energy functions in which the free energy is the sum of three individual polyconvex subfunctions ψ_1, ψ_2, ψ_3, such that $\psi(\boldsymbol{F}) = \psi_1(I_1) + \psi_2(I_2) + \psi_3(I_3)$, is polyconvex by design and the stresses take the following form,

$$\boldsymbol{P} = \frac{\partial\psi(I_1,I_2,I_3)}{\partial\boldsymbol{F}} = \frac{\partial\psi_1}{\partial I_1}\frac{\partial I_1}{\partial\boldsymbol{F}} + \frac{\partial\psi_2}{\partial I_2}\frac{\partial I_2}{\partial\boldsymbol{F}} + \frac{\partial\psi_3}{\partial I_3}\frac{\partial I_3}{\partial\boldsymbol{F}} \,. \tag{12}$$

This implies that we can either select polyconvex activation functions from a set of algorithmically predefined activation functions [20], or custom-design our own activations functions from known polyconvex subfunctions ψ_1, ψ_2, ψ_3 [27]. Here we select first and second powers of the invariants for the first hidden layer and linear, exponential, and logarithmic functions of these powers for the second hidden layer, all with *non-negative coefficients*. In

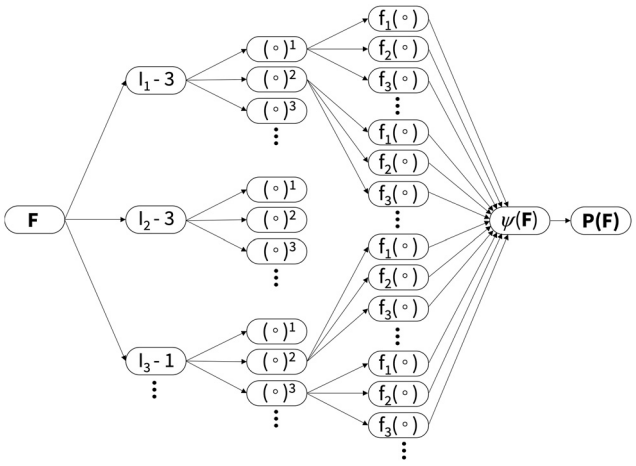

Fig. 1. Constitutive Artificial Neural Network. Family of a feed forward Constitutive Artificial Neural Networks with two hidden layers to approximate the single scalar-valued free energy function $\psi(I_1, I_2, I_3)$ as a function of the scalar-valued invariants I_1, I_2, I_3 of the deformation gradient $\boldsymbol{F}$. The first layer generates powers $(\circ)$, $(\circ)^2$, $(\circ)^3$ of the network input and the second layer applies thermodynamically admissible activation functions $f(\circ)$ to these powers. Constitutive Artificial Neural Networks are typically not fully connected by design to a priori satisfy the condition of polyconvexity.

addition, we abandon the fully-connected network architecture, in which mixed products of the invariants I_1, I_2, I_3 emerge naturally [31]. Instead, we decouple the inputs I_1, I_2, I_3 and only combine them additively in the free energy function, $\psi = \psi_1 + \psi_2 + \psi_3$ [58,59]. Satisfying polyconvexity, for example according to Eq. (12), can imply enforcing *non-negative network weights* [27,38], and directly affects the *architecture* of the neural network [36]. In practical applications, the constraints associated with polyconvexity may also affect training cost and expressivity [60].

2.3. Constitutive Artificial Neural Networks

Motivated by these considerations, we build a family of Constitutive Artificial Neural Networks that satisfy the conditions of thermodynamic consistency, material objectivity, material symmetry, incompressibility, constitutive restrictions, and polyconvexity by design. This guides our selection of network *input*, *output*, *architecture*, and *activation functions* to a priori satisfy the fundamental laws of physics. Special members of this family represent well-known constitutive models, including the neo Hookean [41], Blatz Ko [42], Mooney Rivlin [43,44], Demiray [45], Gent [46], and Holzapfel [47] models, for which the network weights gain a clear physical interpretation.

Constitutive Artificial Neural Network input and output. To ensure thermodynamical consistency, rather than directly approximating the stress $\boldsymbol{P}$ as a function of the deformation gradient $\boldsymbol{F}$, the Constitutive Artificial Neural Network approximates the scalar-valued free energy function ψ as a function of the scalar-valued invariants I_1, I_2, I_3. The Piola stress $\boldsymbol{P}$ then follows naturally from the second law of thermodynamics as the derivative of the free energy ψ with respect to the deformation gradient $\boldsymbol{F}$ according to Eq. (7). Fig. 1 illustrates a Constitutive Artificial Neural Network with the invariants I_1, I_2, I_3 as input and the the free energy ψ as output.

Constitutive Artificial Neural Network architecture. To model a hyperelastic *history-independent* material, we select a feed forward architecture in which information only moves in one direction, from the input nodes, without any cycles or loops, to the output nodes. To ensure polyconvexity, we choose a selectively connected architecture according to Eq. (12), such that the free energy

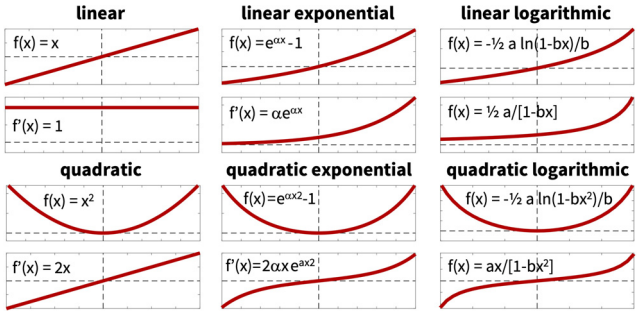

Fig. 2. Activation functions for Constitutive Artificial Neural Networks. We use custom-design activation functions $f(x)$ along with their derivatives $f'(x)$ that include linear and quadratic mappings, left, linear and quadratic exponentials, middle, and linear and quadratic logarithmic functions, right, to reverse engineer a free energy function that captures popular functional forms of constitutive terms.

function does not contain mixed terms in the invariants. Fig. 1 illustrates one possible network architecture that a priori decouples the individual invariants. Its free energy function,

$$\begin{aligned}\psi(I_1,I_2,I_3) = {} & w_{2,1}\ f_1(w_{1,1}\ [I_1-3]^1) + w_{2,2}\ f_2(w_{1,2}\ [I_1-3]^1) \\ & + w_{2,3}\ f_3(w_{1,3}\ [I_1-3]^1) + w_{2,4}\ f_1(w_{1,4}\ [I_1-3]^2) \\ & + w_{2,5}\ f_2(w_{1,5}\ [I_1-3]^2) + w_{2,6}\ f_3(w_{1,6}\ [I_1-3]^2) \\ & \vdots \\ & + w_{2,26}\ f_2(w_{1,26}\ [I_3-1]^3) + w_{2,27}\ f_3(w_{1,27}\ [I_3-1]^3)\,, \end{aligned} \tag{13}$$

introduces $3\times3\times3+3\times3\times3=54$ weights. The first set of 27 weights, $w_{1,1..27}$, weighs the powers of the invariants and the second set of 27 weights, $w_{2,1..27}$, weighs the contributions of the functions f_1, f_2, f_3.

Activation functions. To ensure that our network satisfies basic physically reasonable constitutive restrictions, rather than selecting from a set of pre-defined activation functions such as the binary step, soft step, hyperbolic tangent, inverse tangent, or soft plus functions, we custom-design our own activation functions to *reverse-engineer* a free energy function that captures popular forms of constitutive terms. Specifically, we select linear and quadratic powers of the first and second invariants for the first layer of the network, and linear, exponential, or logarithmic functions for the second layer.

Fig. 2 illustrates the six activation functions $f(x)$ along with their derivatives $f'(x)$ that we use throughout the remainder of this work. Notably, in contrast to the activation functions for classical neural networks, all six functions are not only *monotonic*, $f(x+\varepsilon) \geq f(x)$ for $\varepsilon \geq 0$, such that increasing deformations result in increasing stresses, but also *continuous* at the origin, $f(-0) = f(+0)$, *continuously differentiable* and *smooth* at the origin, $f'(-0) = f'(+0)$, and zero at the origin, $f(0) = 0$, to ensure an energy- and stress-free reference configuration according to Eq. (10), and *unbounded*, $f(-\infty) \to \infty$ and $f(+\infty) \to \infty$, to ensure an infinite energy and stress for extreme deformations according to Eq. (11).

Fig. 3 illustrates our isotropic, perfectly incompressible Constitutive Artificial Neural Network with two hidden layers and four and twelve nodes. The first layer generates powers $(\circ)^1$ and $(\circ)^2$ of the network inputs, $[I_1-3]$ and $[I_2-3]$, and the second layer applies the identity, $(\circ)$, the exponential function, $(\exp((\circ))-1)$, and the natural logarithm, $(-\ln(1-(\circ)))$, to these powers. The set

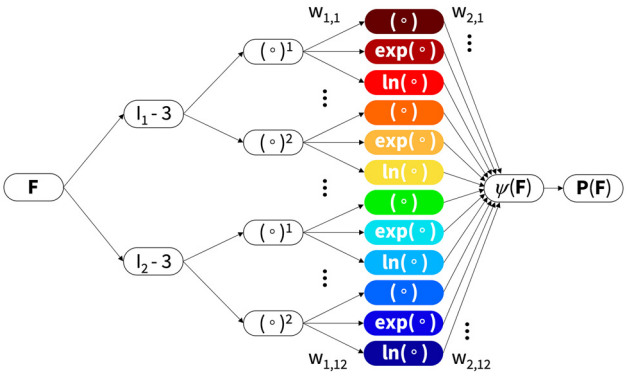

Fig. 3. Constitutive Artificial Neural Network. Isotropic, perfectly incompressible Constitutive Artificial Neural Network with with two hidden layers to approximate the single scalar-valued free energy function $\psi(I_1, I_2)$ as a function of the first and second invariants of the deformation gradient $\boldsymbol{F}$ using twelve terms. The first layer generates powers $(\circ)^1$ and $(\circ)^2$ of the network inputs, $[I_1 - 3]$ and $[I_2 - 3]$ and the second layer applies the identity $(\circ)$, the exponential function, $(\exp((\circ)) - 1)$, and the natural logarithm, $(-\ln(1 - (\circ)))$, to these powers. The networks is selectively connected by design to a priori satisfy the condition of polyconvexity.

of equations for this networks takes the following explicit form,

$$\begin{aligned}\psi(I_1,I_2) = {} & w_{2,1}\, w_{1,1}\, [I_1-3] + w_{2,2}\, [\exp(w_{1,2}\, [I_1-3]\,) - 1] \\ & - w_{2,3}\, \ln(1 - w_{1,3}\, [I_1-3]\,) \\ & + w_{2,4}\, w_{1,4}\, [I_1-3]^2 + w_{2,5}\, [\exp(w_{1,5}\, [I_1-3]^2) - 1] \\ & - w_{2,6}\, \ln(1 - w_{1,6}\, [I_1-3]^2) \\ & + w_{2,7}\, w_{1,7}\, [I_2-3] + w_{2,8}\, [\exp(w_{1,8}\, [I_2-3]\,) - 1] \\ & - w_{2,9}\, \ln(1 - w_{1,9}\, [I_2-3]\,) \\ & + w_{2,10} w_{1,10}\, [I_2-3]^2 + w_{2,11}\, [\exp(w_{1,11}\, [I_2-3]^2) - 1] \\ & - w_{2,12}\, \ln(1 - w_{1,12}\, [I_2-3]^2)\,.\end{aligned} \tag{14}$$

For this particular format, one of the first two weights of each row becomes redundant, and we can reduce the set of network parameters from 24 to 20, $\mathbf{w} = [\,(w_{1,1}w_{2,1}), w_{1,2}, w_{2,2}, w_{1,3}, w_{2,3}, (w_{1,4}w_{2,4}), w_{1,5}, w_{2,5}, w_{1,6}, w_{2,6}, (w_{1,7}w_{2,7}), w_{1,8}, w_{2,8}, w_{1,9}, w_{2,9}, (w_{1,10}w_{2,10}), w_{1,11}, w_{2,11}, w_{1,12}, w_{2,12}\,]$. Using the second law of thermodynamics, we can derive an explicit expression for the Piola stress from Eq. (5), $\boldsymbol{P} = \partial\psi/\partial\boldsymbol{F}$, or, more specifically, for the case of perfect incompressibility from Eq. (8), $\boldsymbol{P} = \partial\psi/\partial I_1 \cdot \partial I_1/\partial\boldsymbol{F} + \partial\psi/\partial I_2 \cdot \partial I_2/\partial\boldsymbol{F} - p\boldsymbol{F}^{-t}$,

$$\begin{aligned}\boldsymbol{P} = {} & \partial I_1/\partial\boldsymbol{F}\quad [w_{2,1}\, w_{1,1} + w_{2,2}\, w_{1,2}\quad \exp(w_{1,2}\, [I_1-3]\,) \\ & \qquad + w_{2,3}\, w_{1,3}\ /[\,1 - w_{1,3}\, [I_1-3]\,]\,] \\ & + 2\,[I_1-3]\,[w_{2,4}\, w_{1,4} + w_{2,5}\, w_{1,5}\quad \exp(w_{1,5}\, [I_1-3]^2)] \\ & \qquad + w_{2,6}\, w_{1,6}\ /[\,1 - w_{1,6}\, [I_1-3]^2]] \\ & + \partial I_2/\partial\boldsymbol{F}\quad [w_{2,7}\, w_{1,7} + w_{2,8}\, w_{1,8}\quad \exp(w_{1,8}\, [I_2-3]\,) \\ & \qquad + w_{2,9}\, w_{1,9}\ /[\,1 - w_{1,9}\, [I_2-3]\,]\,] \\ & + 2\,[I_2-3]\,[w_{2,10}w_{1,10} + w_{2,11}w_{1,11}\quad \exp(w_{1,11}\, [I_2-3]^2)] \\ & \qquad + w_{2,12}w_{1,12}/[\,1 - w_{1,12}\, [I_2-3]^2]],\end{aligned} \tag{15}$$

corrected by the pressure term, $-p\boldsymbol{F}^{-t}$, with $p = -\frac{1}{3}\boldsymbol{P} : \boldsymbol{F}$. The stress definition (15) suggests that our model is a *generalization* of many popular constitutive models for incompressible hyperelastic materials. It seems natural to ask whether and how its network parameters $w_{1..2,1..12}$ relate to the parameters of these models.

Special types of constitutive equations. To demonstrate that the family of Constitutive Artificial Neural Networks in Fig. 1 and its specific example in Fig. 3 are *generalizations* of popular constitutive models, we consider six widely used models and systematically compare their material parameters to our network weights. The *neo Hookean model* [41], the simplest of all models, has a free energy function that is a constant function of only the first invariant, $[I_1 - 3]$, scaled by the shear modulus μ,

$$\psi = \tfrac{1}{2}\,\mu\,[I_1 - 3] \qquad \text{where} \qquad \mu = 2\,w_{1,1}w_{2,1}\,. \tag{16}$$

The *Blatz Ko model* [42], has a free energy function that depends only the second and third invariants, $[I_2 - 3]$ and $[I_3 - 1]$, scaled by the shear modulus μ as $\psi = \frac{1}{2}\,\mu\,[I_2/I_3 + 2\sqrt{I_3} - 5]$. For perfectly incompressible materials, $I_3 = 1$, it simplifies to the following form,

$$\psi = \tfrac{1}{2}\,\mu\,[I_2 - 3] \qquad \text{where} \qquad \mu = 2\,w_{1,7}w_{2,7}\,. \tag{17}$$

The *Mooney Rivlin model* [43,44] is a combination of both (16) and (17) and accounts for the first and second invariants, $[I_1 - 3]$ and $[I_2 - 3]$, scaled by the moduli μ_1 and μ_2 that sum up to the overall shear modulus, $\mu = \mu_1 + \mu_2$,

$$\psi = \frac{1}{2}\mu_1\,[I_1 - 3] + \tfrac{1}{2}\,\mu_2\,[I_2 - 3] \quad \text{where} \quad \begin{aligned}\mu_1 &= 2\,w_{1,1}w_{2,1}\\ \mu_2 &= 2\,w_{1,7}w_{2,7}\,.\end{aligned} \tag{18}$$

The *Demiray model* [45] or *Delfino model* [61] uses linear exponentials of the first invariant, $[I_1 - 3]$, in terms of two parameters a and b,

$$\psi = \frac{1}{2}\,\frac{a}{b}\,[\exp(b\,[I_1 - 3\,]) - 1\,] \quad \text{where} \quad \begin{aligned}a &= 2\,w_{1,2}w_{2,2}\\ b &= w_{1,2}\,.\end{aligned} \tag{19}$$

The *Gent model* [46] uses linear logarithms of the first invariant, $[I_1 - 3]$, in terms of two parameters α and β,

$$\psi = -\frac{1}{2}\,\frac{\alpha}{\beta}\ln(1 - (\beta\,[I_1 - 3\,]) \quad \text{where} \quad \begin{aligned}\alpha &= 2\,w_{1,3}w_{2,3}\\ \beta &= w_{1,3}\,.\end{aligned} \tag{20}$$

The *Holzapfel model* [47] uses quadratic exponentials, typically of the fourth invariant, which we adapt here for the the first invariant, $[I_1 - 3\,]$, in terms of two parameters a and b,

$$\psi = \frac{1}{2}\,\frac{a}{b}[\exp(b[I_1 - 3\,]^2) - 1\,] \text{ where} \quad \begin{aligned}a &= 2\,w_{1,5}w_{2,5}\\ b &= w_{1,5}.\end{aligned} \tag{21}$$

These simple examples demonstrate that we can recover popular constitutive functions for which the network weights gain a well-defined physical meaning.

Loss function. The objective of our Constitutive Artificial Neural Network is to learn the network parameters $\boldsymbol{\theta} = \{w_{ij}\}$, the network weights of the first and second layers, by minimizing a loss function L that penalizes the error between model and data. Similar to classical neural networks, we characterize this error as the mean squared error, the L_2-norm of the difference between model $\boldsymbol{P}(\boldsymbol{F}_i)$ and data $\hat{\boldsymbol{P}}_i$, divided by the number of training points n_{trn},

$$L(\boldsymbol{\theta};\boldsymbol{F}) = \frac{1}{n_{\text{trn}}}\sum_{i=1}^{n_{\text{trn}}} ||\,\boldsymbol{P}(\boldsymbol{F}_i) - \hat{\boldsymbol{P}}_i\,||^2 \to \min\,. \tag{22}$$

To reduce potential overfitting, we also study the effects of Lasso or L1 regularization and L2 regularization,

$$L(\boldsymbol{\theta};\boldsymbol{F}) = \frac{1}{n_{\text{trn}}}\sum_{i=1}^{n_{\text{trn}}} ||\,\boldsymbol{P}(\boldsymbol{F}_i) - \hat{\boldsymbol{P}}_i\,||^2 + \alpha_1\,||W||_1 + \frac{1}{2}\,\alpha_2\,||W||_2^2 \to \min\,, \tag{23}$$

by enhancing the loss function by the weighted L1 norm, $||W||_1 = \sum_i \sum_j |\,w_{ij}\,|$, or the weighted Euclidian or L2 norm, $||W||_2^2 = \sum_i \sum_j w_{ij}^2$, where α_1 and α_2 are the weighting coefficients. We train the network by minimizing the loss function (22) or (23) and learn the network parameters $\boldsymbol{\theta} = \{w_{ij}\}$ using the ADAM optimizer, a robust adaptive algorithm for gradient-based first-order optimization, and constrain the weights to always remain non-negative, $w_{ij} \geq 0$.

2.4. Data

To demonstrate automated model discovery with our Constitutive Artificial Neural Network, we perform a systematic study using widely-used benchmark data for human brain tissue [7,8,14]. Specifically, we train our two-layer Constitutive Artificial Neural Network for isotropic, perfectly incompressible materials from Fig. 3, discover a material model and its parameters, and compare the model and parameters against six traditional constitutive models for soft biological tissues [41–47]. We consider two training scenarios, *single-mode training* and *multi-mode training*, for the homogeneous deformation modes of uniaxial tension, uniaxial compression, and simple shear.

Tension and compression. For the case of uniaxial tension and compression, we stretch the specimen in one direction, $F_{11} = \lambda_1 = \lambda$. For an *isotropic, perfectly incompressible* material with $I_3 = \lambda_1^2\lambda_2^2\lambda_3^2 = 1$, the stretches orthogonal to the loading direction are identical and equal to the square root of the stretch, $F_{22} = \lambda_2 = \lambda^{-1/2}$ and $F_{33} = \lambda_3 = \lambda^{-1/2}$. From the resulting deformation gradient, $\boldsymbol{F} = \text{diag}\{\lambda, \lambda^{-1/2}, \lambda^{-1/2}\}$, we calculate the first and second invariants and their derivatives,

$$\begin{matrix} I_1 = \lambda^2 + 2/\lambda \\ I_2 = 2\lambda + 1/\lambda^2 \end{matrix} \quad \text{with} \quad \begin{matrix} \partial_\lambda I_1 = 2\lambda - 2/\lambda^2 \\ \partial_\lambda I_2 = 2 - 2/\lambda^3 \,, \end{matrix} \tag{24}$$

to evaluate the nominal uniaxial stress P_{11} using the general stress-stretch relationship for perfectly incompressible materials, $P_{ii} = [\partial\psi/\partial I_1]\,[\partial I_1/\partial\lambda_i] + [\partial\psi/\partial I_2]\,[\partial I_2/\partial\lambda_i] - [1/\lambda_i]\,p$, for $i = 1, 2, 3$. Here, p denotes the hydrostatic pressure that we determine from the zero stress condition in the transverse directions, $P_{22} = 0$ and $P_{33} = 0$, as $p = [2/\lambda]\,\partial\psi/\partial I_1 + [2\lambda + 2/\lambda^2]\,\partial\psi/\partial I_2$. This results in the following explicit *uniaxial stress-stretch relation for perfectly incompressible, isotropic* materials,

$$P_{11} = 2\left[\frac{\partial\psi}{\partial I_1} + \frac{1}{\lambda}\frac{\partial\psi}{\partial I_2}\right]\left[\lambda - \frac{1}{\lambda^2}\right]. \tag{25}$$

Shear. For the case of simple shear, we shear the specimen in one direction, $F_{12} = \gamma$. For an *isotropic, perfectly incompressible* material with $F_{11} = F_{22} = F_{33} = 1$, we calculate the first and second invariants and their derivatives,

$$\begin{matrix} I_1 = 3 + \gamma^2 \\ I_2 = 3 + \gamma^2 \end{matrix} \quad \text{with} \quad \begin{matrix} \partial_\lambda I_1 = 2\,\gamma \\ \partial_\lambda I_2 = 2\,\gamma \,, \end{matrix} \tag{26}$$

to evalute the nominal shear stress P_{12} using the general stress-stretch relationship for perfectly incompressible materials. This results in the following explicit *shear stress-strain relation for perfectly incompressible, isotropic* materials,

$$P_{12} = 2\left[\frac{\partial\psi}{\partial I_1} + \frac{\partial\psi}{\partial I_2}\right]\gamma \,. \tag{27}$$

Testing and training data. Tables 1 and 2 summarize our benchmark data of human gray matter tissue from the cortex and basal ganglia and white matter tissue from the corona radiata and corpus callosum tested in tension, compression, and shear [7,8,14]. All tests were performed on cubic $5 \times 5 \times 5$ mm^3 samples, harvested from ten human brains, six male and four female, ages 54 to 81 years, tested within 60 hours post mortem [7]. Since brain tissue is most vulnerable to tensile loading, each sample was first tested in shear, then in compression, and then in tension, at maximum shear strains and stretches well within the elastic regime. We report each data set as 17 pairs of stretches and nominal uniaxial stresses, $\{\lambda, P_{11}\}$, or shear strains and nominal shear stresses, $\{\gamma, P_{12}\}$, where the stretches and shear strains range from $0.9 \le \lambda \le 1.1$ and $0.0 \le \gamma \le 0.2$, and the stresses are the means of the loading and unloading curves of n samples. We first conduct a general performance study, illustrate the convergence of the loss function, and split each data set into 75% training data and 25% test data to demonstrate that our model generally performs well at both interpolation and extrapolation. Throughout the remainder of this study, we then perform *single-mode training* with one mode used as training data and the remaining two modes as test data, and *multi-mode training* with all three modes used as training data. For convenience, all data sets from Tables 1 and 2 are available at https://github.com/LivingMatterLab/CANN.

3. Results

Fig. 4 illustrates the performance of our Constitutive Artificial Neural Network with two hidden layers and twelve nodes for single-mode training with the tension, compression, and shear data, and for multi-mode training with all three data sets combined. The top row documents the loss function from Eq. (22) plotted over the training epochs, and the bottom row shows the resulting nominal stress P from Eq. (15) as a function of the stretch λ and shear strain γ. The dots summarize the experimental data from the human cortex under tension, compression, and shear [7] from Table 1, and the color-coded areas highlight the twelve contributions to the discovered stress function (22) according to Fig. 3 with the discovered weights from Table 3. In each plot, we

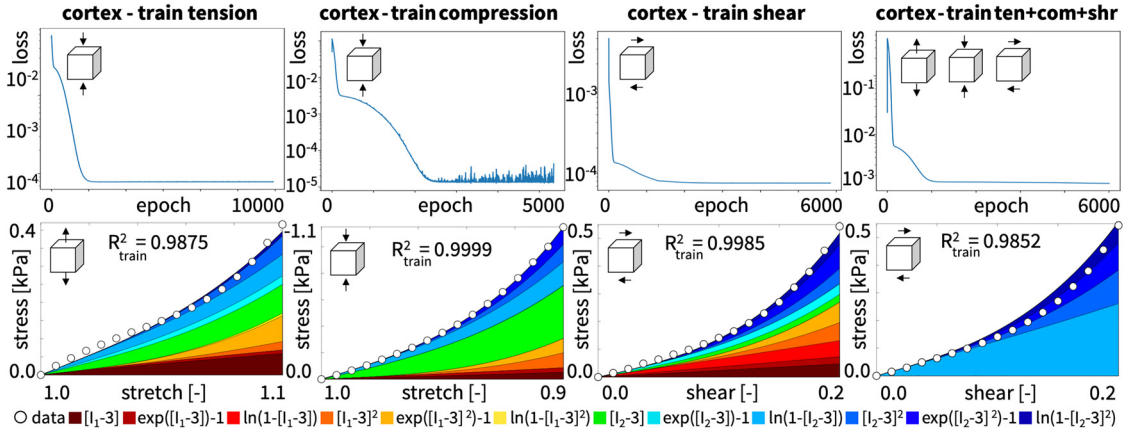

Fig. 4. Model and parameter discovery for gray matter. Loss over epochs, top, and nominal stress as a function of stretch and shear strain, bottom, for the isotropic, perfectly incompressible Constitutive Artificial Neural Network with two hidden layers and twelve nodes, for single-mode training with the tension, compression, and shear data, and for multi-mode training with all three data sets combined. Dots illustrate the tension, compression, and shear data of the human cortex [7] from Table 1; color-coded areas highlight the twelve contributions to the discovered stress function according to Fig. 3 from Table 3.

Table 1
Gray matter data. Cortex and basal ganglia tested in tension, compression, and shear; stresses are reported as means from the loading and unloading curves of n samples [7].

cortex tension $n = 15$		**cortex compression** $n = 17$		**cortex shear** $n = 35$		**basal ganglia tension** $n = 15$		**basal ganglia compression** $n = 15$		**basal ganglia shear** $n = 29$	
λ [-]	P_{11} [kPa]	λ [-]	P_{11} [kPa]	γ [-]	P_{12} [kPa]	λ [-]	P_{11} [kPa]	λ [-]	P_{11} [kPa]	γ [-]	P_{12} [kPa]
1.0000	0.0000	1.0000	0.0000	0.0000	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000	0.0000
1.0063	0.0251	0.9938	-0.0308	0.0125	0.0147	1.0063	0.0149	0.9938	-0.0174	0.0125	0.0070
1.0125	0.0462	0.9875	-0.0659	0.0250	0.0294	1.0125	0.0251	0.9875	-0.0358	0.0250	0.0140
1.0188	0.0666	0.9812	-0.1040	0.0375	0.0486	1.0188	0.0345	0.9812	-0.0534	0.0375	0.0210
1.0250	0.0838	0.9750	-0.1479	0.0500	0.0633	1.0250	0.0446	0.9750	-0.0778	0.0500	0.0305
1.0312	0.1010	0.9688	-0.1908	0.0625	0.0814	1.0312	0.0540	0.9688	-0.1021	0.0625	0.0397
1.0375	0.1175	0.9625	-0.2375	0.0750	0.0983	1.0375	0.0619	0.9625	-0.1265	0.0750	0.0488
1.0437	0.1324	0.9563	-0.2920	0.0875	0.1186	1.0437	0.0705	0.9563	-0.1479	0.0875	0.0579
1.0500	0.1488	0.9500	-0.3504	0.1000	0.1412	1.0500	0.0791	0.9500	-0.1752	0.1000	0.0703
1.0562	0.1661	0.9437	-0.4127	0.1125	0.1649	1.0562	0.0862	0.9437	-0.2102	0.1125	0.0805
1.0625	0.1856	0.9375	-0.4866	0.1250	0.1942	1.0625	0.0963	0.9375	-0.2414	0.1250	0.0930
1.0688	0.2091	0.9313	-0.5684	0.1375	0.2292	1.0688	0.1050	0.9313	-0.2842	0.1375	0.1088
1.0750	0.2366	0.9250	-0.6579	0.1500	0.2698	1.0750	0.1151	0.9250	-0.3270	0.1500	0.1257
1.0813	0.2710	0.9187	-0.7630	0.1625	0.3227	1.0813	0.1277	0.9187	-0.3776	0.1625	0.1449
1.0875	0.3125	0.9125	-0.8837	0.1750	0.3791	1.0875	0.1426	0.9125	-0.4321	0.1750	0.1686
1.0938	0.3650	0.9062	-1.0005	0.1875	0.4557	1.0938	0.1582	0.9062	-0.4905	0.1875	0.1969
1.1000	0.4151	0.9000	-1.1484	0.2000	0.5435	1.1000	0.1778	0.9000	-0.5528	0.2000	0.2262

Table 2
White matter data. Corona radiata and corpus callosum tested in tension, compression, and shear; stresses are reported as means from the loading and unloading curves of n samples [7].

corona radiata tension $n = 18$		**corona radiata compression** $n = 18$		**corona radiata shear** $n = 36$		**corpus callosum tension** $n = 19$		**corpus callosum compression** $n = 20$		**corpus callosum shear** $n = 39$	
λ [-]	P_{11} [kPa]	λ [-]	P_{11} [kPa]	γ [-]	P_{12} [kPa]	λ [-]	P_{11} [kPa]	λ [-]	P_{11} [kPa]	γ [-]	P_{12} [kPa]
1.0000	0.0000	1.0000	0.0000	0.0000	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000	0.0000
1.0063	0.0157	0.9938	-0.0193	0.0125	0.0079	1.0063	0.0078	0.9938	-0.0096	0.0125	0.0036
1.0125	0.0235	0.9875	-0.0387	0.0250	0.0159	1.0125	0.0149	0.9875	-0.0164	0.0250	0.0072
1.0188	0.0345	0.9812	-0.0543	0.0375	0.0238	1.0188	0.0196	0.9812	-0.0300	0.0375	0.0109
1.0250	0.0423	0.9750	-0.0800	0.0500	0.0318	1.0250	0.0251	0.9750	-0.0427	0.0500	0.0170
1.0312	0.0509	0.9688	-0.1040	0.0625	0.0409	1.0312	0.0298	0.9688	-0.0564	0.0625	0.0217
1.0375	0.0572	0.9625	-0.1305	0.0750	0.0488	1.0375	0.0337	0.9625	-0.0730	0.0750	0.0319
1.0437	0.0642	0.9563	-0.1674	0.0875	0.0601	1.0437	0.0376	0.9563	-0.0895	0.0875	0.0342
1.0500	0.0721	0.9500	-0.2024	0.1000	0.0681	1.0500	0.0415	0.9500	-0.1051	0.1000	0.0422
1.0562	0.0791	0.9437	-0.2453	0.1125	0.0817	1.0562	0.0454	0.9437	-0.1363	0.1125	0.0468
1.0625	0.0869	0.9375	-0.2959	0.1250	0.0964	1.0625	0.0486	0.9375	-0.1596	0.1250	0.0558
1.0688	0.0940	0.9313	-0.3543	0.1375	0.1133	1.0688	0.0533	0.9313	-0.1946	0.1375	0.0627
1.0750	0.1050	0.9250	-0.4127	0.1500	0.1347	1.0750	0.0580	0.9250	-0.2297	0.1500	0.0751
1.0813	0.1151	0.9187	-0.4827	0.1625	0.1596	1.0813	0.0634	0.9187	-0.2764	0.1625	0.0853
1.0875	0.1292	0.9125	-0.5723	0.1750	0.1878	1.0875	0.0697	0.9125	-0.3270	0.1750	0.1011
1.0938	0.1418	0.9062	-0.6657	0.1875	0.2227	1.0938	0.0775	0.9062	-0.3854	0.1875	0.1192
1.1000	0.1582	0.9000	-0.7591	0.2000	0.2611	1.1000	0.0862	0.9000	-0.4555	0.2000	0.1429

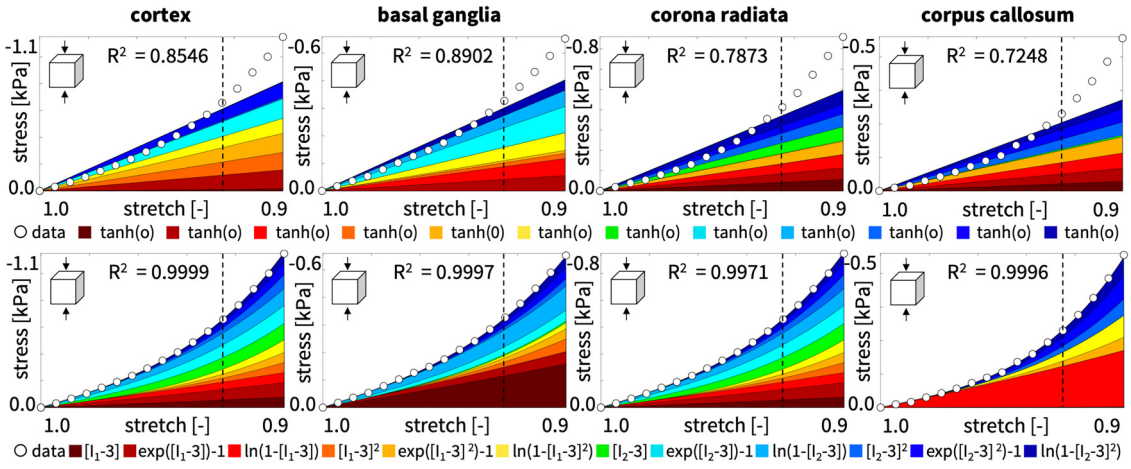

Fig. 5. Predictive potential of classical neural network vs. Constitutive Artificial Neural Network. Nominal stress as a function of stretch for a classical neural network with one hidden layer, twelve nodes, and 37 unknowns, top, and for the isotropic, perfectly incompressible Constitutive Artificial Neural Network with two hidden layers, twelve nodes, and 20 unknowns. Dots illustrate the compression data of the human cortex, basal ganglia, corona radiata, and corpus callosum [7] from Table 1, of which we used the left 75% for training and the right 25% for testing; color-coded areas highlight the twelve contributions to the discovered stress function of hyperbolic tangent type, top, and according to Fig. 3, bottom.

Table 3
Gray matter model. Cortex parameters learned for tension, compression, and shear data from Table 1 using the isotropic, perfectly incompressible Constitutive Artificial Neural Network with two hidden layers, and twelve nodes from Fig. 3. Summary of the 24 weights $w_{1:2.1:12}$ and the coefficient of determination R^2 for training with the three individual tests and for all three tests combined.

	cortex tension $n = 15$		**cortex compression** $n = 17$		**cortex shear** $n = 35$		**cortex ten+com+shr** $n = 15, 17, 35$	
	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]
$w_{\bullet,1}$	0.3135	0.3456	0.4027	0.1979	0.6628	0.1796	0.0000	0.0000
$w_{\bullet,2}$	0.1576	0.1094	0.0628	0.7898	0.2422	0.2599	0.0000	0.0000
$w_{\bullet,3}$	0.0000	0.0000	0.0000	0.0000	0.7662	0.1840	0.0000	0.0000
$w_{\bullet,4}$	1.1303	0.6813	2.3725	1.1085	1.4402	1.4200	0.0000	0.0000
$w_{\bullet,5}$	1.4721	1.5618	1.1856	2.1032	1.3360	1.7106	0.0000	0.0000
$w_{\bullet,6}$	0.5017	0.4345	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$w_{\bullet,7}$	0.9522	0.1690	1.8534	0.2897	0.3725	0.1899	0.0000	0.0000
$w_{\bullet,8}$	0.2275	0.2072	0.0587	0.0585	0.2607	0.3574	0.0000	0.0000
$w_{\bullet,9}$	0.6824	0.1727	1.9469	0.1144	0.0000	0.0000	0.9875	0.6339
$w_{\bullet,10}$	2.2641	0.8482	2.2740	1.1302	0.8798	1.9874	2.7738	1.3702
$w_{\bullet,11}$	0.0382	0.3571	1.2234	2.0668	1.7350	1.5506	1.6495	1.8880
$w_{\bullet,12}$	0.9325	0.4734	0.0000	0.0000	0.8817	1.4250	1.4026	1.6663
	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}
	0.9875	0.9282	0.0000	0.8176	0.6209	0.9985	0.3560	0.9852
	0.4366	0.7829	0.9999	0.8602	0.7297	0.8785	0.8972	0.9306

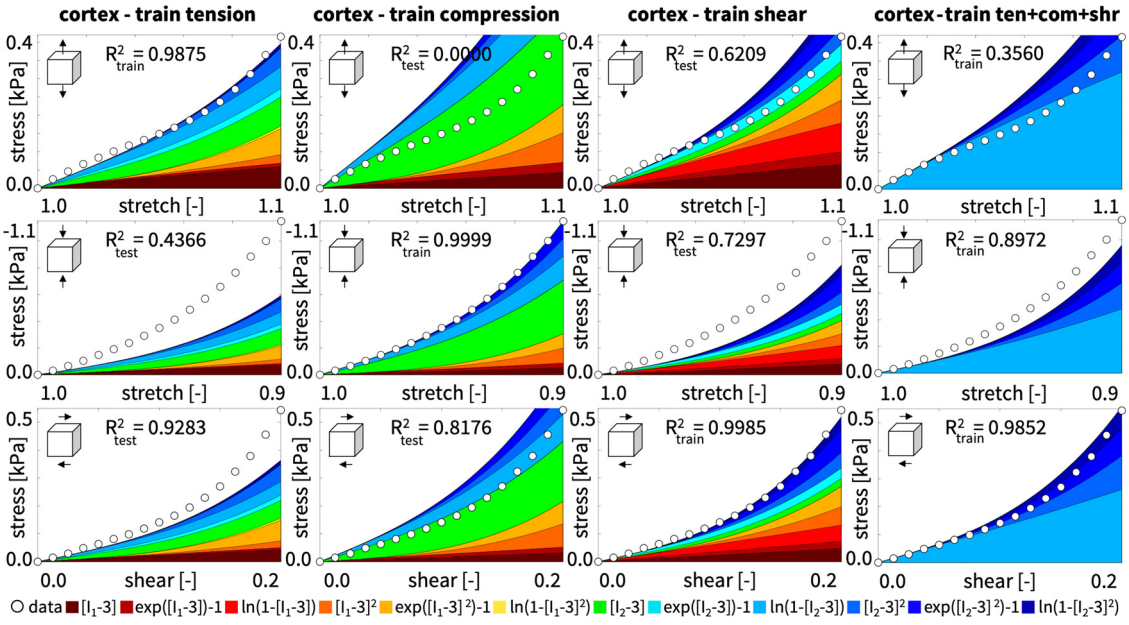

Fig. 6. Gray matter data vs. model. Nominal stress as a function of stretch and shear strain for the isotropic, perfectly incompressible Constitutive Artificial Neural Network with two hidden layers, and twelve nodes from Fig. 3. Dots illustrate the tension, compression, and shear data of the human cortex [7] from Table 1; color-coded areas highlight the twelve contributions to the discovered stress function according to Fig. 3 from Table 3.

report the coefficients of determination R^2 to quantify the goodness of fit between model and data. First, the rapidly dropping loss in all four graphs of the top row confirms that our network *trains robustly* and converges within less than 5,000 epochs. Second, the R^2 values of 0.99, 1.00, 1.00, 0.99 confirm that our network trains well for both single- and multi-mode training. Third, training our network is relatively *inexpensive*, with computational costs per training run varying between 2-3 minutes on a standard desktop computer. Fig. 5 compares the predictive potential of a classical neural network and our Constitutive Artificial Neural Network. The top row shows the nominal stress P as a function of stretch λ for a classical neural network with one hidden layer, twelve nodes, 24 weights, 13 biases, and a total of 37 unknowns. The bottom row shows the same stress P versus stretch λ response for our isotropic, perfectly incompressible Constitutive Artificial Neural Network with two hidden layers, twelve nodes, and 20 unknowns. The dots illustrate the compression data of the human cortex, basal ganglia, corona radiata, and corpus callosum [7] from Table 1. To assess the predictive potential of both networks, we used 75% of the data to the left of the dashed line for training and the remaining 25% to the right of the dashed line for testing. First, both networks train well and succeed in *interpolating* or *fitting* the dots to the left of the dashed line. Second, the classical neural network in the top row fails to extrapolate or pre-

Table 4
White matter model. Corona radiata parameters learned for tension, compression, and shear data from Table 2 using the isotropic, perfectly incompressible Constitutive Artificial Neural Network with two hidden layers, and twelve nodes from Fig. 3. Summary of the 24 weights $w_{1:2,1:12}$ and the coefficient of determination R^2 for training with the three individual tests and with all three tests combined.

	corona radiata tension $n = 18$		**corona radiata compression** $n = 18$		**corona radiata shear** $n = 33$		**corona radiata ten+com+shr** $n = 18, 18, 33$	
	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]
$w_{\bullet,1}$	0.0000	0.0000	1.7357	0.2807	0.3643	0.2492	0.0000	0.0000
$w_{\bullet,2}$	0.0000	0.0000	0.0000	0.0000	0.1032	0.2404	0.0000	0.0000
$w_{\bullet,3}$	0.0000	0.0000	0.0000	0.0000	0.0369	0.3070	0.0000	0.0000
$w_{\bullet,4}$	0.8932	0.1474	1.5473	1.0767	1.3942	0.6520	0.0000	0.0000
$w_{\bullet,5}$	0.3760	0.2325	1.1415	1.2150	1.3600	1.1027	0.0000	0.0000
$w_{\bullet,6}$	1.3081	0.4295	1.2115	1.1480	0.4401	0.8310	0.0000	0.0000
$w_{\bullet,7}$	1.0042	0.0717	0.0000	0.0000	0.0349	0.2945	1.3862	0.1598
$w_{\bullet,8}$	0.0867	0.0717	0.0029	0.0295	0.0550	0.3905	0.2398	0.4900
$w_{\bullet,9}$	0.8403	0.2065	0.0000	0.0000	0.7680	0.1179	0.0000	0.0000
$w_{\bullet,10}$	0.0000	0.0000	1.0083	1.4130	1.0552	0.8552	0.0000	0.0000
$w_{\bullet,11}$	0.0000	0.0000	1.2191	1.1331	0.9990	1.0740	1.8893	1.6859
$w_{\bullet,12}$	1.1048	0.0030	2.6478	0.8233	0.0000	0.0000	1.1789	1.9113
	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}
	0.9875	0.8250	0.0000	0.0620	0.0000	0.9989	0.0000	0.9349
	0.0460	0.5471	0.9998	0.6251	0.4837	0.7271	0.7898	0.8361

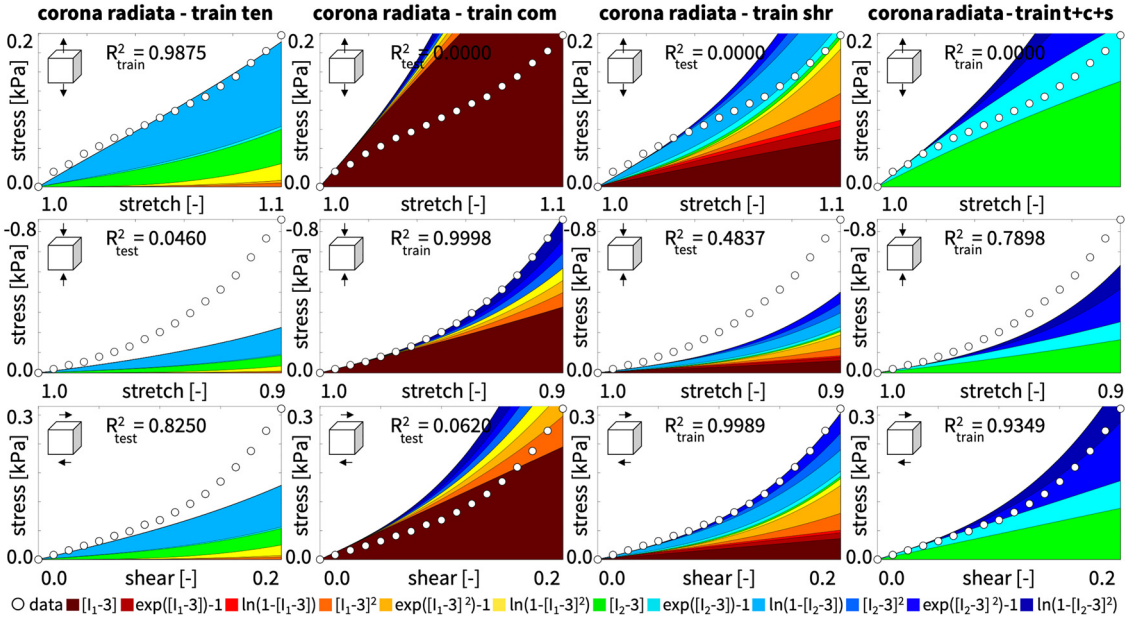

Fig. 7. White matter data vs. model. Nominal stress as a function of stretch and shear strain for the isotropic, perfectly incompressible Constitutive Artificial Neural Network with two hidden layers, and twelve nodes from Fig. 3. Dots illustrate the tension, compression, and shear data of the human corona radiata [7] from Table 1; color-coded areas highlight the twelve contributions to the discovered stress function according to Fig. 3 from Table 4.

dict the behaviorbeyond the training regime and the model deviates progressively from the data, as confirmed by the coefficients of determination R^2 of 0.85, 0.89, 0.79, 0.72. Third, our Constitutive Artificial Neural Network in the bottom row performs well at *extrapolating* or *predicting* the test data to the right of the dashed line, with coefficients of determination R^2 of 1.00, 1.00, 1.00, 1.00 across all four brain regions.

Table 3 and Fig. 6 summarize and illustrate the discovered models for the human cortex for the tension, compression, and shear data from Table 1 using our isotropic, perfectly incompressible Constitutive Artificial Neural Network from Fig. 3. Table 3 summarizes the 24 weights $w_{1:2,1:12}$ and the coefficient of determination R^2 for single-mode training with the three individual modes and for multi-mode training with all three modes combined. Fig. 6 directly compares the data and model in terms of the nominal stress as a function of the stretch and shear strain, where the dots indicate the measured data and the color-coded regions highlight the individual contributions of the twelve network nodes to the discovered free energy function ψ. First, for single-mode training with the individual modes, we note that the Constitutive Artificial Neural Network succeeds in *interpolating* or *fitting* the three individual sets of training data: The learned network parameters define stress functions that fit the individual tension, compression, and shear data excellently with R^2_{train} values of 0.99, 1.00, and 1.00. Second, for single-mode training, we observe that the network performs moderately at *extrapolating* or *predicting* data outside the train-

Table 5
Gray and white matter models. Cortex, basal ganglia, corona radiata, and corpus callosum parameters learned for combined tension, compression, and shear data from Tables 1 and 2 using the isotropic, perfectly incompressible Constitutive Artificial Neural Network with two hidden layers, and twelve nodes from Fig. 3. Summary of the 12 non-zero weights $w_{1:2,7:12}$ and the coefficient of determination R^2 for training with all three tests combined.

	cortex ten+com+shr $n = 15, 17, 35$		**basal ganglia** ten+com+shr $n = 15, 15, 29$		**corona radiata** ten+com+shr $n = 18, 18, 36$		**corpus callosum** ten+com+shr $n = 19, 20, 39$	
	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]
$w_{\bullet,7}$	0.0000	0.0000	1.7880	0.1927	1.3862	0.1598	0.5635	0.1067
$w_{\bullet,8}$	0.0000	0.0000	0.0000	0.0000	0.2398	0.4900	0.2363	0.1383
$w_{\bullet,9}$	0.9875	0.6339	0.0000	0.0000	0.0000	0.0000	0.8398	0.1135
$w_{\bullet,10}$	2.7738	1.3702	0.9396	0.8143	0.0000	0.0000	0.9210	1.1218
$w_{\bullet,11}$	1.6495	1.8880	1.6193	1.1867	1.8893	1.6859	1.0628	1.0185
$w_{\bullet,12}$	1.4026	1.6663	0.9666	0.7932	1.1789	1.9113	0.7282	1.2621
	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}
	0.3560 0.8972	0.9852 0.9306	0.0000 0.8646	0.9739 0.9135	0.0000 0.7898	0.9349 0.8361	0.0000 0.7847	0.9209 0.8303

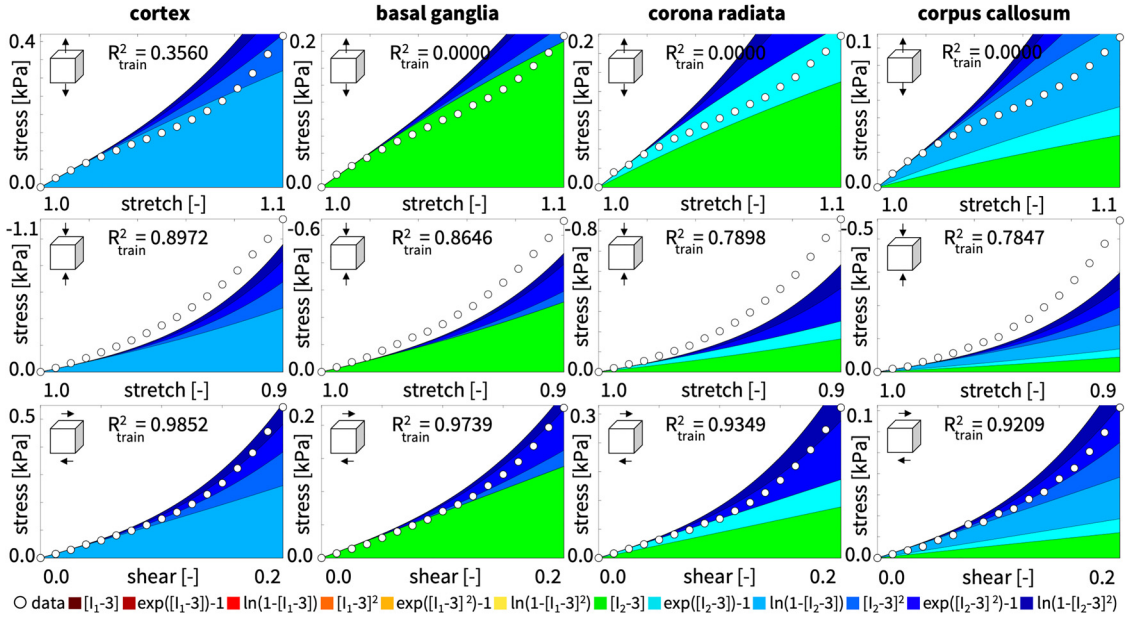

Fig. 8. Gray and white matter data vs. model. Nominal stress as a function of stretch and shear strain for the isotropic, perfectly incompressible Constitutive Artificial Neural Network with two hidden layers, and twelve nodes from Fig. 3. Dots illustrate the tension, compression, and shear data of all four brain regions [7] from Table 1; color-coded areas highlight the six contributions to the discovered stress function according to Fig. 3 from Table 5.

ing regime: The network parameters trained individually for each mode do not predict the other modes well, with R^2_{test} values ranging from 0.00 for the tension prediction with compression training to 0.93 for the shear prediction with tension training. Third, for multi-mode training with all data sets combined, the coefficient of determination R^2_{train} of the individual tests decreases to 0.36, 0.90, and 0.99, while the sum of the three R^2 values, the collective fit of all three tests, increases. Fourth, for single-mode training, all twelve terms of the model are activated as indicated through the full color spectrum in the first three columns. At the same time, for for multi-mode training, the Constitutive Artificial Neural Network discovers a model with only four terms, while the weights of the other terms train to zero. Strikingly, against our intuition, these four terms are all functions of the second invariant $[I_2 - 3]$ instead of the first $[I_1 - 3]$, as indicated through the cold blue-type colors.

Table 4 and Fig. 7 summarize and illustrate the discovered models for the human corona radiata for the tension, compression, and shear data from Table 2. The white matter results from the corona radiata confirm the trends of the gray matter results for the cortex in Table 3 and Fig. 6. First, for single-mode training, our neural network succeeds in *interpolating* or *fitting* the individual training data: The learned network parameters define stress functions that fit the individual tension, compression, and shear data excellently with R^2_{train} values of 0.99, 1.00, and 1.00. Second, the network performs moderately at *extrapolating* or *predicting* data outside the training regime: The network parameters trained for each individual mode fail to predict the other modes equally well, with R^2_{test} values ranging from 0.00 for the tension predictions with both compression and shear training to 0.83 for the shear prediction with tension training. Third, we find that, for all tests combined, the coefficient of determination R^2_{train} of the tensile test remains 0.00 and decreases to 0.79 and 0.93 for compression and shear, but the collective fit increases. Fourth, similar to the gray matter results in Fig. 6, the white matter model trained with the individual tests in Fig. 7 activates seven, eight, and eleven terms as indi-

Table 6
Gray and white matter models. Cortex, basal ganglia, corona radiata, and corpus callosum parameters learned for combined tension, compression, and shear data from Tables 1 and 2 using the isotropic, perfectly incompressible Constitutive Artificial Neural Network with two hidden layers, and twelve nodes from Fig. 3 with additional L2 regularization for the weights. Summary of the four non-zero weights $w_{1:2,8:9}$ and the coefficient of determination R^2 for training with all three tests combined.

	cortex ten+com+shr $n = 15, 17, 35$		**basal ganglia** ten+com+shr $n = 15, 15, 29$		**corona radiata** ten+com+shr $n = 18, 18, 36$		**corpus callosum** ten+com+shr $n = 19, 20, 39$	
	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]
$w_{\bullet,8}$	0.4957	0.4442	0.0000	0.0000	0.4560	0.4351	0.2409	0.2367
$w_{\bullet,9}$	0.9840	0.7064	0.6802	0.6250	0.5614	0.4987	0.4739	0.4551
	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}
	0.4590	0.9477	0.4778	0.9699	0.0000	0.9551	0.0000	0.9620
	0.7425	0.8788	0.6696	0.8588	0.5143	0.7418	0.5109	0.7276

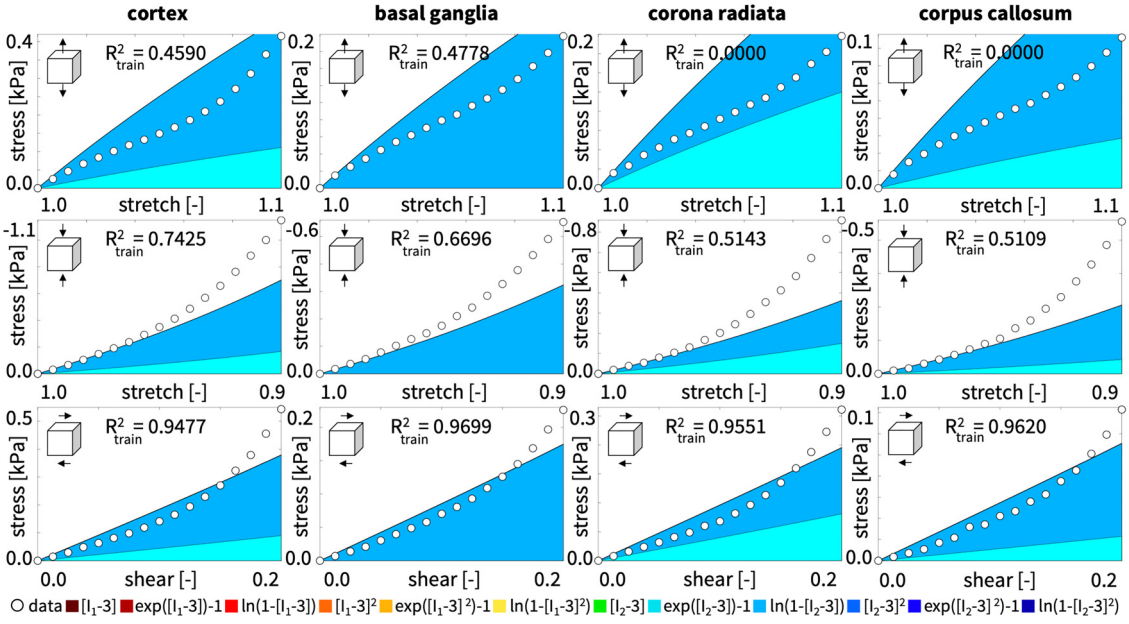

Fig. 9. Gray and white matter data vs. model. Nominal stress as a function of stretch and shear strain for the isotropic, perfectly incompressible Constitutive Artificial Neural Network with two hidden layers, and twelve nodes from Fig. 3 with additional L2 regularization for the weights. Dots illustrate the tension, compression, and shear data of all four brain regions [7] from Table 1; color-coded areas highlight the two contributions to the discovered stress function according to Fig. 3 from Table 6.

cated through the broad color spectrum in the first three columns. Interestingly, for all three tests combined, the Constitutive Artificial Neural Network discovers a model with only four terms, which are again all functions of the second invariant, $[I_2 - 3]$, as indicated through the cold blue-type colors.

Table 5 and Fig. 8 summarize and illustrate the discovered models for the human cortex, basal ganglia, corona radiata, and corpus callosum, all for multi-mode training with the tension, compression, and shear data from Tables 1 and 2. First and foremost, for multi-mode training, the fit of the shear data with R^2_{train} values of 0.99, 0.97, 0.93, and 0.92 is uniformly the best across all four brain regions. Second, the model universally *underestimates* the compressive stresses with R^2 values ranging from 0.78 to 0.90, and *overestimates* the tensile stresses with R^2 values from 0.00 to 0.36, indicating a poor fit of the tensile data. Third, and most importantly, the side-by-side comparison of all four brain regions confirms the trends of the cortex and the corona radiata: Our Constitutive Artificial Neural Network uniquely discovers a family of models that is parameterized in terms of the *second invariant only*, while the weights of the first invariant terms consistently train to zero. The blue color spectrum in Fig. 8 underscores this observation.

Table 6 and Fig. 9 highlight the effects of the L2 regularization according to Eq. (23). Compared to all oter examples without regularization, as expected, the regularization reduces the number of non-zero terms, in our case from six in Table 5 and Fig. 8, to two for the cortex, the corona radiata, and the corpus callosum, and only one for the basal ganglia. The associated non-zero weights, $w_{1,8}$, $w_{2,8}$, $w_{1,9}$, $w_{2,9}$, activate the linear exponential term, $\exp([I_2 - 3]) - 1$, and the linear logarithmic term, $\ln(1 - [I_2 - 3])$, which are highlighted in turquoise and light blue in Fig. 9. The general trends are the same for the discovered six-term model without regularization and two-term model with L2 regularization: Both models depend on the second invariant only and their fits are best for the shear data with R^2 values well above 0.90 and worst for the tension data with R^2 values ranging from 0.00 to 0.48.

Table 7 and Fig. 10 demonstrate an application of our Constitutive Artificial Neural Network beyond model discovery, the param-

Table 7
Special cases of neo Hookean, Blatz Ko, Mooney Rivlin, Demiray, Gent, and Holzapfel models. Cortex, basal ganglia, corona radiata, and corpus callosum parameters learned for combined tension, compression, and shear data from Tables 1 and 2. Summary of the non-zero weights, the physics parameters μ, μ_1, μ_2, a, b, α, β, and the coefficient of determination R^2 for training with all three tests combined.

	neo Hookean ten+com+shr $n = 15, 17, 35$		**Blatz Ko ten+com+shr** $n = 15, 17, 35$		**Mooney Rivlin ten+com+shr** $n = 15, 17, 35$		**Demiray ten+com+shr** $n = 15, 17, 35$		**Gent ten+com+shr** $n = 15, 17, 35$		**Holzapfel ten+com+shr** $n = 15, 17, 35$	
	cortex		**cortex**		**cortex**		**cortex**		**cortex**		**cortex**	
	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]
$w_{\bullet,1}$	0.7880	1.1522	–	–	0.0026	0.4128	–	–	–	–	–	–
$w_{\bullet,2}$	–	–	–	–	–	–	1.0529	0.8760	–	–	–	–
$w_{\bullet,3}$	–	–	–	–	–	–	–	–	1.8399	0.4782	–	–
$w_{\bullet,5}$	–	–	–	–	–	–	–	–	–	–	4.1833	4.7548
$w_{\bullet,7}$	–	–	1.4156	0.6726	2.2122	0.4253	–	–	–	–	–	–
	$\mu = 1.8159$ kPa		$\mu = 1.9043$ kPa		$\mu_1 = 0.0021$ kPa $\mu_2 = 1.8817$ kPa		$a = 1.8447$ kPa $b = 1.0529$		$\alpha = 1.7597$ kPa $\beta = 1.8399$		$a = 39.7815$ kPa $b = 4.1833$	
	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}
	0.2817 0.6066	0.9394 0.8195	0.3627 0.7588	0.9457 0.8809	0.4021 0.7477	0.9446 0.8784	0.1360 0.6544	0.9499 0.8314	0.2875 0.6239	0.9502 0.8264	0.4845 0.5325	0.9560 0.8001
	basal ganglia		**basal ganglia**		**basal ganglia**		**basal ganglia**		**basal ganglia**		**basal ganglia**	
	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]
$w_{\bullet,1}$	0.4138	1.0624	–	–	0.0000	0.0000	–	–	–	–	–	–
$w_{\bullet,2}$	–	–	–	–	–	–	0.5829	0.7362	–	–	–	–
$w_{\bullet,3}$	–	–	–	–	–	–	–	–	1.3991	0.2960	–	–
$w_{\bullet,5}$	–	–	–	–	–	–	–	–	–	–	1.9013	4.7958
$w_{\bullet,7}$	–	–	0.2628	1.6856	1.0851	0.4132	–	–	–	–	–	–
	$\mu = 0.8792$ kPa		$\mu = 0.8860$ kPa		$\mu_1 = 0.0000$ kPa $\mu_2 = 0.8967$ kPa		$a = 0.8385$ kPa $b = 0.5829$		$\alpha = 0.8283$ kPa $\beta = 1.3991$		$a = 18.2365$ kPa $b = 1.9013$	
	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}
	0.0425 0.5812	0.9684 0.8112	0.3557 0.6969	0.9687 0.8649	0.3033 0.7104	0.9689 0.8683	0.1191 0.5688	0.9700 0.8091	0.2267 0.5494	0.9719 0.8054	0.2195 0.4162	0.9269 0.7553
	corona radiata		**corona radiata**		**corona radiata**		**corona radiata**		**corona radiata**		**corona radiata**	
	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]
$w_{\bullet,1}$	0.5157	0.91250	–	–	0.0200	0.4208	–	–	–	–	–	–
$w_{\bullet,2}$	–	–	–	–	–	–	0.5924	0.7693	–	–	–	–
$w_{\bullet,3}$	–	–	–	–	–	–	–	–	1.0545	0.4548	–	–
$w_{\bullet,5}$	–	–	–	–	–	–	–	–	–	–	3.2397	3.1842
$w_{\bullet,7}$	–	–	0.6742	0.7087	0.5359	0.9047	–	–	–	–	–	–
	$\mu = 0.9412$ kPa		$\mu = 0.9556$ kPa		$\mu_1 = 0.0168$ kPa $\mu_2 = 0.9697$ kPa		$a = 0.9115$ kPa $b = 0.5924$		$\alpha = 0.9592$ kPa $\beta = 1.0545$		$a = 20.6317$ kPa $b = 3.2397$	
	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}
	0.0000 0.3770	0.9509 0.6699	0.0000 0.4977	0.9520 0.7355	0.0000 0.5158	0.9528 0.7414	0.0000 0.3567	0.9517 0.6643	0.0000 0.3840	0.9573 0.6737	0.0000 0.3604	0.9373 0.6702
	corpus callosum		**corpus callosum**		**corpus callosum**		**corpus callosum**		**corpus callosum**		**corpus callosum**	
	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]	$w_{1,\bullet}$ [-]	$w_{2,\bullet}$ [kPa]
$w_{\bullet,1}$	0.6521	0.4131	–	–	0.0053	0.4231	–	–	–	–	–	–
$w_{\bullet,2}$	–	–	–	–	–	–	0.3625	0.7162	–	–	–	–
$w_{\bullet,3}$	–	–	–	–	–	–	–	–	0.3682	0.6894	–	–
$w_{\bullet,5}$	–	–	–	–	–	–	–	–	–	–	1.6928	3.4100
$w_{\bullet,7}$	–	–	0.4124	0.6932	0.7495	0.3731	–	–	–	–	–	–
	$\mu = 0.5388$ kPa		$\mu = 0.5718$ kPa		$\mu_1 = 0.0045$ kPa $\mu_2 = 0.5593$ kPa		$a = 0.5192$ kPa $b = 0.3625$		$\alpha = 0.5077$ kPa $\beta = 0.3682$		$a = 11.5449$ kPa $b = 1.6928$	
	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}	R^2_t R^2_c	R^2_s R^2_{tc}
	0.0000 0.3871	0.9603 0.6579	0.0000 0.5485	0.9577 0.7391	0.0000 0.5328	0.9590 0.7339	0.0000 0.3570	0.9606 0.6487	0.0000 0.3331	0.9587 0.6407	0.0000 0.3739	0.9199 0.6618

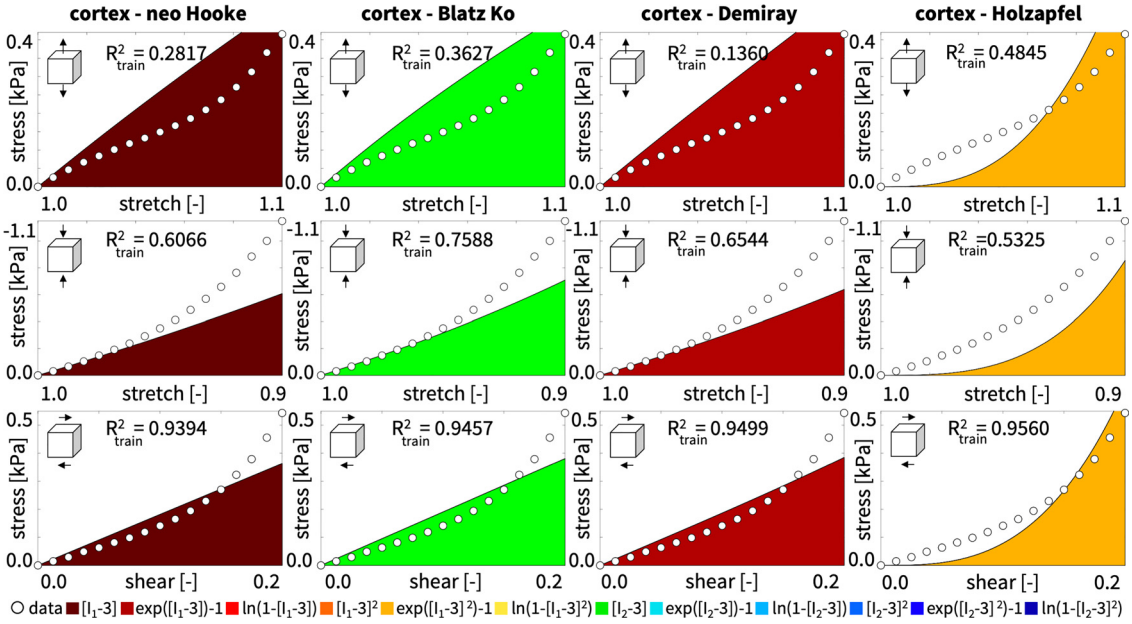

Fig. 10. Special cases of neo Hookean, Blatz Ko, Mooney Rivlin, Demiray, Gent, and Holzapfel models. Nominal stress as a function of stretch and shear for special cases of the isotropic, perfectly incompressible Constitutive Artificial Neural Network from Fig. 3. Dots illustrate the tension, compression, and shear data of the human cortex [7] from Table 1; color-coded areas highlight the terms of the stress function according to Fig. 3 from Table 7.

eter identification and comparison of special cases of the generalized network according to Eqs. (16) to (21). The first set of models, the neo Hookean, Blatz Ko, and Mooney Rivlin models, are all linear in terms of the first invariant, second invariant, or both; the second set, the Demiray, Gent, and Holzapfel models, contain linear exponential, linear logarithmic, or quadratic exponential terms. Table 7 shows that each model, except for the Mooney Rivlin model, activates only one term of our network, either the first, second, third, fifth, or seventh. For all six models, we can convert the weights into a stiffness-like parameter with units [kPa]; the linear Mooney Rivlin model has an additional stiffness-like parameter, and the three nonlinear models have an additional coefficient of nonlinearity. Fig. 10 shows the behavior of the neo Hookean, Blatz Ko, Demiray, and Holzapfel models when simultaneously trained for the tension, compression, and shear data of the human cortex. Notably, for the small stretch and shear strain ranges of $0.9 \leq \lambda \leq 1.1$ and $0.0 \leq \gamma \leq 0.2$, only the Holzapfel model displays a marked strain stiffening, while the neo Hookean, Blatz Ko, and Demiray models remain is their predominantly linear regimes. This allows the Holzapfel model to perform best not only in shear, with an R^2 values of 0.96, but also in tension with a value of 0.48, where most other models fail.

Fig. 11 summarizes and compares the performance of all models, the Constitutive Artificial Neural Network without and with L2 regularization and its special cases, the neo Hookean, Blatz Ko, Demiray, Gent, and Holzapfel models. The graphs in the first three columns result from single-mode training with the individual tension, compression, and shear data [7] from Tables 1 and 2; the last column results from multi-mode training with all three data sets combined. The three rows highlight the coefficients of determination for tension R^2_t, compression R^2_c, and shear R^2_s. The color-coded blocks and error bars represent the means and standard deviations of the R^2 value across all four brain regions. First, in the three graphs on the diagonal that reflect the *training* of the models, the R^2_{train} values of all seven models are close to one, with only three models training poorly, the neo Hookean and Holzapfel models in tension and the Blatz Ko model in compression. Notably, our non-regularized Constitutive Artificial Neural Network outperforms all other models and has the largest R^2_{train} values when trained individually for tension, compression, and shear. Second, from the six off-diagonal graphs that reflect the *testing* of the models, we conclude that the model and parameters trained for tension are generally incapable of predicting the compression behavior and vice versa. However, the tension parameters are reasonably well suited to characterize the shear behavior, with our Constitutive Artificial Neural Network and the Holzapfel model performing best; vice versa, the shear parameters are moderately suited to characterize the tensile behavior, with our Constitutive Artificial Neural Network and the Blatz Ko model performing best. Finally, from the right column that reflects *training* with all three data sets combined, we conclude that our Constitutive Artificial Neural Network performs best for all three modes, followed by its L2 regularized counterpart, and the Blatz Ko model. Interestingly, we observe large R^2 values across the entire bottom row, indicating that, of all three tests, shear tests are generally the easiest to fit and predict for all seven models. Taken together, our non-regularized Constitutive Artificial Neural Network performs best in eight of all twelve cases, second best in two, and fifths in one suggesting that our proposed neural network *successfully discovers both model and parameters* that best describe the data.

4. Discussion

Characterizing human brain tissue is a challenging but important task. Throughout the past decade, driven by the need to improve diagnostic and predictive clinical tools, neuroscience has seen an enormous, growing interest in accurately characterizing and modeling the human brain [4]. Numerous research groups have proposed competing constitutive models to best characterize the behavior of gray and white matter tissue and calibrate the model parameters in response to mechanical loading [14]. Amongst the wide variety of possible models, the neo Hookean [41], Blatz Ko [42], Mooney Rivlin [43,44], Demiray [45], Gent [46], and Holzapfel [47] models have emerged as the most successful candidates to approximate the stress-stretch relations in the human brain. The gold standard strategy of all these approaches is to *first* select a constitutive model, either from the above list or beyond, and *then* tune its parameters by fitting the model to data.

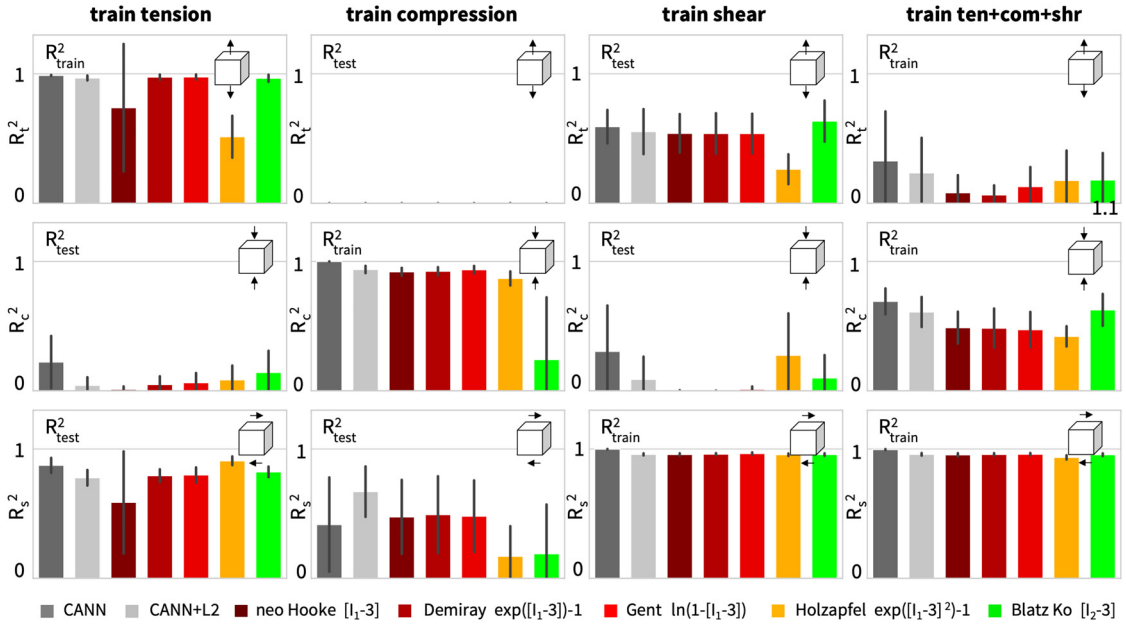

Fig. 11. Goodness of fit for all models. Coefficients of determination R^2 of our Constitutive Artificial Neural Network, without and with L2 regularization, and its special cases, the neo Hookean, Blatz Ko, Demiray, Gent, and Holzapfel models, trained with the tension, compression, and shear data [7] from Tables 1 and 2; color-coded blocks and error bars highlight the means and standard deviations of the coefficient of determination R^2 across all four brain regions, with colors and terms according to Fig. 3.

Often, these data are collected for a single loading mode–tension [62], compression [11], or shear [63]–and the parameters that fit one type of loading fail to predict the behavior for the other modes [7,17]. This simplification can have fatal consequences; for example, it could overestimate the stiffness of the brain in injury simulations. To address these limitations, our group has recently performed a comprehensive set of human brain tissue experiments in tension, compression, and shear and calibrated the neo Hookean, Mooney Rivlin, Demiray, and Gent models for four different brain regions, the cortex, basal ganglia, corona radiata, and corpus callosum [7,8,14]. While this approach is valuable to generate the best sets of parameters for existing models, some natural follow-up questions to ask are: How good are these models in the first place? Which one of them performs best? Are there other models that perform equally well, or even better? And, if so, how can we find them?

Constitutive Artificial Neural Networks are a family of neural networks that a priori satisfy thermodynamic constraints. When searching for generic models that could outperform traditional constitutive models, neural networks are a natural first choice [24]. Neural networks have advanced as a powerful strategy to approximate data by cleverly combining nested weighted activation functions with several thousand unknowns [64]. They have become the go-to strategy to interpolate data within a well-defined domain when the underlying physics are completely unknown [25]. At the same time, as Fig. 5 indicates, classical neural networks typically fail to predict the behavior outside the training domain [28], they violate common physical constraints, and their parameters have no real physical interpretation [32]. This has sparked the recent trend to integrate physical information into classical neural networks [31]. In the spirit of this idea, we propose a new family of neural networks that a priori satisfy common kinematic, thermodynamic, and physical constraints. Towards this goal we consult the non-linear field theories of mechanics [48,50,65] and constrain the network *output* to enforce thermodynamic consistency; the network *input* to enforce material objectivity, and, if desired, material symmetry and incompressibility; the *activation functions* to implement physically reasonable constitutive restrictions; and the network *architecture* to ensure polyconvexity. These ideas are not entirely new. Several recent network models are designed around enforcing thermodynamic constraints [27,35,38], for example through additional terms in their loss function [66]. However, the problem of overfitting sparse data with a large set of physically meaningless network parameters remains [36]. This raises the questions: How do we harness decades of knowledge in constitutive modeling to create a neural network, from easy-to-understand modular building blocks, with well-defined physical parameters, that we can constrain with our domain knowledge?

Constitutive Artificial Neural Networks can be made up of building blocks that feature prominent constitutive models. At a closer look, most popular constitutive models for human brain tissue have a similar functional structure. Here we propose to hardwire this structure into our neural network architecture [20]. Our underlying design paradigm is to *reverse-engineer* a Constitutive Artificial Neural Network that is, by design, a generalization of widely used and commonly accepted constitutive models including the neo Hookean, Blatz Ko, Mooney Rivlin, Demiray, Gent, and Holzapfel models, and a combination of their individual terms. In Fig. 3, we prototype this idea for an isotropic perfectly incompressible feed forward network with two hidden layers and four and twelve nodes. This network takes the scalar-valued first and second invariants of the deformation gradient, $[I_1 - 3]$ and $[I_2 - 3]$, as input and approximates the scalar-valued free energy function, $\psi(I_1, I_2)$, as output. The first layer generates the first and second powers, $(\circ)^1$ and $(\circ)^2$, of the input, and the second layer applies the identity $(\circ)$, the exponential, $(\exp((\circ)) - 1)$, and the natural logarithm $(-\ln(1 - (\circ)))$ to these powers. This results in twelve building blocks, and a total possible combination of $2^{12} - 1 = 4095$ possible models, that additively feed into the final free energy function ψ from which we derive the Piola stress, $\boldsymbol{P} = \partial\psi/\partial\boldsymbol{F}$, following standard arguments of thermodynamics. This strategy is conceptually similar to a recent approach that uses sparse-regression, instead of neural networks, to discover the relevant terms from a library of physics-inspired constitutive building blocks [19]. It is easy to show that

our network is a *generalization* of popular constitutive models with the neo Hookean [41], Blatz Ko [42], Mooney Rivlin [43,44], Demiray [45], Gent [46], and Holzapfel [47] models as special cases. More importantly, through a direct comparison with these models in Eqs. (16) to (21), the weights of our network gain a clear physical interpretation. Table 7 and Fig. 10 show, for example, that we recover the classical neo Hookean model with shear moduli of $\mu =$ 1.82kPa, 0.88kPa, 0.94kPa, 0.54kPa for simultaneous training with the tension, compression, and shear data of the cortex, basal ganglia, corona radiata, and corpus callosum, which agree well with the reported values of $\mu =$ 2.07kPa, 0.99kPa, 1.15kPa, 0.65kPa [7]. Interestingly, both our network and the parameter fit in the literature find that one of the two shear moduli of the Mooney Rivlin model is consistently zero in all four regions, while the other is $\mu_2 =$ 1.88kPa, 0.90kPa, 0.97kPa, 0.56kPa for our approach compared to $\mu_1 =$2.08kPa, 1.00kPa, 1.16kPa, 0.65kPa in the literature [7]. This agrees well with other studies in which one of the Mooney Rivlin shear moduli was also significantly smaller than the other across all brain regions [17]. Fig. 10 reveals several additional universal trends for human brain tissue: First, tension is not only the most challenging test to perform [62], but also the most difficult test to fit, with R^2 values ranging from 0.14 to 0.48, followed by compression with 0.53 to 0.76, and shear with 0.94 to 0.96. Second, when trained simultaneously for tension, compression, and shear, all models consistently overestimate the tensile stiffness and underestimate the compression stiffness, highlighting the pronounced tension-compression asymmetry in all four regions of the human brain [17]. Third, of all existing models, only the Holzapfel model captures the nonlinear stress response [47], suggesting that the classical invariant-based models struggle to reproduce the nonlinear behavior of human brain tissue for small deformations with $0.9 \leq \lambda \leq 1.1$ and $0.0 \leq \gamma \leq 0.2$. This raises the question: Can we design a Constitutive Artificial Neural Network that not only learns the best set of parameters for a given constitutive model, but also learns the model itself?

Constitutive Artificial Neural Networks simultaneously discover both model and parameters. In essence, we propose a radically different approach towards soft tissue modeling and abandon the common strategy to *first* select a constitutive model and *then* tune its parameters by fitting the model to data [20]. Instead, we propose a family of Constitutive Artificial Neural Networks, with the general architecture in Fig. 1, specified for soft tissues in Fig. 3, to *simultaneously discover both, model and parameters* that best describe the data. Probing our network with the tension, compression, and shear experiments from the gray matter cortex in Table 3 and Fig. 6 and from the white matter corona radiata in Table 4 and Fig. 7, reveals several interesting trends: When trained with all three experiments *individually*, the network activates all its twelve terms, and fails to discover a single best model. Nonetheless, with these twelve terms, it succeeds in *interpolating* or *fitting* the training data from one experiment; however, it only performs moderately at *extrapolating* or *predicting* the test data from the other two experiments. This suggests that the data from a single loading mode are not sufficient to characterize the entire breadth of the mechanical response of human brain which agrees well with observations in the literature [7,17]. Notably, when trained with all three experiments *simultaneously*, the Constitutive Artificial Neural Network robustly discovers a single model that best approximates the data: For the cortex, in the last columns of Table 3 and Fig. 6, the network discovers four relevant terms, while the weights of the other eight terms train to zero,

$$\begin{aligned}\psi(I_2) = &\frac{1}{2}\,\mu_2[\,I_2-3\,]^2 + \frac{1}{2}\,\frac{a_2}{b_2}[\,\exp(\,b_2[\,I_2-3\,]^2)-1\,] \\ &-\frac{1}{2}\,\frac{\alpha_1}{\beta_1}\ln(\,1-\beta_1[\,I_2-3\,])-\frac{1}{2}\,\frac{\alpha_2}{\beta_2}\ln(\,1-\beta_2[\,I_2-3\,]^2)\,. \qquad (28)\end{aligned}$$

The non-zero weights translate into physically meaningful cortex parameters with well-defined physical units, the four stiffness-like parameters, $\mu_2 = 7.60$kPa, $a_2 = 6.23$kPa, $\alpha_1 = 1.25$kPa, $\alpha_2 =$ 4.67kPa, and the three nonlinearity parameters, $b_2 = 1.65$, $\beta_1 =$ 0.99, $\beta_2 = 1.40$. For the corona radiata, in the last columns of Table 4 and Fig. 7, the network discovers four relevant terms, while the weights of the other eight terms train to zero,

$$\begin{aligned}\psi(I_2) = &\frac{1}{2}\,\mu_1[\,I_2-3\,] + \frac{1}{2}\,\frac{a_1}{b_1}[\,\exp(\,b_1[\,I_2-3\,])-1\,] \\ &+\frac{1}{2}\,\frac{a_2}{b_2}\,[\exp(b_2[\,I_2-3]^2)-1]-\frac{1}{2}\,\frac{\alpha_2}{\beta_2}\ln(1-\beta_2[\,I_2-3]^2)\,. \qquad (29)\end{aligned}$$

The non-zero weights translate into physically meaningful parameters with well-defined physical units, the four stiffness-like parameters, $\mu_1 = 0.44$kPa, $a_1 = 0.24$kPa, $a_2 = 6.37$kPa, $\alpha_2 = 4.51$kPa, and the three nonlinearity parameters, $b_1 = 0.24$, $b_2 = 1.89$, $\beta_2 = 1.18$. Notably, of all seven models in Fig. 11, the Constitutive Artificial Neural Network performs best in eight of all twelve cases, second best in two, and fifths in one, suggesting that it successfully discovers the model and parameters that best describe the data. Since the network *autonomously self-selects both model and parameters*, the human user no longer needs to decide which model to choose. This could have enormous practical implications, for example, in finite element simulations: Instead of selecting a specific material model from a library of available models, finite element solvers could be designed around a single generalized model, the Constitutive Artificial Neural Network, which would *autonomously discover the model* from experimental data, populate the model parameters, and activate the relevant terms. This brings up the final and probably most interesting question: Can we learn anything from the discovery process itself?

For human brain tissue, the Constitutive Artificial Neural Network robustly discovers I_2based models. Our Constitutive Artificial Neural Network combines the advantages of both, our knowledge of constitutive modeling [48–50,53–55] and the efficiency of neural network algorithms [29,30,64]. For insufficient training data that only probe individual modes, in the three left columns of Figs. 6 and 7, our network approximates the overall function $\psi(I_1, I_2)$ robustly with R^2 values well above of 0.99. Yet, similar to the classical neural network in Fig. 5, the contributions of the individual activation functions are non-unique. Enriching the training data by multi-mode training for tension, compression, and shear in Table 5 and Fig. 8 eliminates this non-uniqueness. For sufficiently rich data that probe all three modes combined, in the right columns of Figs. 6 and 7, our network successfully captures the behavior of both gray and white matter, and consistently identifies the same unique subset of activation functions, *without overfitting* the data. The reduced color spectra in Fig. 8 confirm that the network self-selects only a subset of activation functions, while most of its weights train to zero. For classical neural networks, a common approach to prevent overfitting is to enrich the loss function by L1 or L2 regularization as we suggest in Eq. (23). For L1 regularization, the discovered model and parameters are virtually identical to the plain model in Table 5 and Fig. 8 and we do not report them separately here. For L2 regularization, the network robustly discovers a reduced model with only two terms, a subset of the non-regularized models in Eqs. (28) and (29), while the weights of the other terms train to zero,

$$\psi(I_2) = \frac{1}{2}\,\frac{a}{b}[\,\exp(\,b[\,I_2-3\,])-\frac{1}{2}\,\frac{\alpha}{\beta}\ln(\,1-\beta[\,I_2-3\,])\,. \qquad (30)$$

Table 6 and Fig. 9 summarize the model and parameters for the regularized network with two stiffness-like parameters, a and α, and two nonlinearity parameters b and β. Strikingly, in multi-mode

training, both the standard and L2 regularized Constitutive Artificial Neural Networks consistently discover models in terms of *the second invariant only*, while all terms of the first invariant train to zero. We can easily see this selective activation in the color-coded stress terms in Figs. 8 and 9, which only display cold blue-type colors associated with the second invariant $[I_2 - 3]$. The dominance of the second invariant is in stark contrast with the popularity of models that only feature the first invariant, but consistent with observations in the literature [17]. We hypothesize that the second invariant term $[I_2 - 3]$, that ranges from 0.0264 in tension to 0.0346 in compression is better suited to model the characteristic *tension-compression asymmetry* of human brain tissue than the first invariant term $[I_1 - 3]$, that only ranges only from 0.0282 to 0.0322 for our experimental range. Last but not least, in addition to discovering *the best model and parameters*, the goodness of fit in Fig. 11 also teaches us something about *the best experiment* [21]. If we had to select a single one experiment, tension, compression, or shear, Fig. 11 suggests that the tension experiment, with the largest R^2 values overall, would provide the richest data and the best insight into the complex behavior of human brain tissue.

Current limitations and future applications. In the present work, we demonstrate the use of Constitutive Artificial Neural Networks for human brain under the assumption of perfect incompressibility and isotropy. The general concept extends naturally to *compressibily* or *near incompressibility* and to materials with other symmetry classes, *transverse isotropy* or *orthotropy*, by expanding the network input to other sets of strain invariants [21]. A more extensive extension would be to incorporate viscous effects [8], or rather *history-dependence* or *inelasticity* in general, for example, by replacing the feed forward architecture through a *long short-term memory network* with feedback connections [67], while still keeping the same overall network input, output, activation functions, and selectively connected architecture. Another limitation, which involves more complex changes, is the additive architecture of our network, which facilitates incorporating polyconvexity. Especially for human brain tissues that display a pronounced Poynting effect with shear softening in tension and shear stiffening in compression [5], it could be beneficial to introduce a *multiplicative coupling* between the individual invariants. Expressing the free energy as a truncated infinite series of products of powers of the invariants, instead of a sum of individual invariant terms, would result in a *fully connected feed forward* network architecture for which polyconvexity is cumbersome to include a priori [57]. Another technical limitation we foresee for these more complex networks, is that the majority of weights might no longer train to zero and that a more involved *L1* or *L2 regularization* could become necessary. This could artificially bias the training towards a subset of physical parameters. One interesting future direction along these lines, especially in view of human brain tissue, would be to compare *invariant-based* and *principal-stretch-based* Constitutive Artificial Neural Networks [68,69]. Several recent studies suggest that principal-stretch-based models outperform invariant-based models, especially in the context of combined loading and strain-stiffening [7,15,17,28]. We are currently investigating the performance of principal-stretch-based neural networks, both as stand alone networks, and with invariant-based neural networks combined. Finally, an important extension would be to embed the network in a Bayesian inference to supplement the analysis with *uncertainty quantification* [32]. Rather than training the network on *a single mean data set* as we have done in the present study, we would then train the network on *multiple raw data sets* to account for variations across individual brains and across the study population. Instead of simple *point estimates* for the network parameters, a Bayesian Constitutive Artificial Neural Network would then learn parameter *distributions with means and credible intervals*. In contrast to classical Bayesian Neural Networks, here, these distributions would have a clear physical interpretation, since our network weights have a well-defined physical meaning.

5. Conclusion

Human brain is an ultrasoft material that is difficult to test and challenging to model. Numerous competing constitutive models for human brain tissue exist in the literature, but selecting the most appropriate model remains a matter of user experience and personal preference. The underlying idea of this manuscript is to automate the process of model selection. Towards this goal, we formulate the problem of autonomous model discovery as a neural network and harness the power of gradient-based adaptive optimizers for deep learning to train the network on human brain data. However, rather than using conventional fully-connected feed-forward networks, we reverse engineer a family of Constitutive Artificial Neural Networks with a sparsely-connected architecture from a set of modular building blocks. We rationalize these building blocks from commonly accepted and widely used constitutive models for soft biological tissues, including the neo Hookean, Mooney Rivlin, Demiray, Gent, and Holzapfel models. This strategy guarantees thermodynamic consistency, material objectivity, material symmetry, physical restrictions, and polyconvexity by design. Probably even more importantly, the weights of our Constitutive Artificial Neural Networks gain a clear physical interpretation and translate naturally into common mechanical parameters. When trained with tension, compression, and shear experiments of gray and white matter tissue, the network simultaneously discovers –out of more than 4,000 possible combinations of models– one unique model and set of parameters, that describe each data set better than any of the commonly used invariant-based models. When constrained to its individual building blocks, the network learns weights that translate into shear moduli of 1.82kPa, 0.88kPa, 0.94kPa, and 0.54kPa for the human cortex, basal ganglia, corona radiata, and corpus callosum which agree well with the reported shear moduli in these four brain regions. Taken together, Constitutive Artificial Neural Networks have the potential to enable automated model discovery and could induce a paradigm shift in soft tissue modeling, from user-defined to automated model selection and parameterization.

Data availability

Our source code, data, and examples are available at https://github.com/LivingMatterLab/CANN.

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Acknowledgments

This work was supported by a DAAD Fellowship to Kevin Linka, by a National Science Foundation Graduate Research Fellowship to Sarah St. Pierre, by the Stanford School of Engineering Covid-19 Research and Assistance Fund, and by Stanford Bio-X IIP seed grant to Ellen Kuhl.

References

[1] GBD, 2016 traumatic brain injury and spinal cord injury collaborators (2019) global, regional, and national burden of traumatic brain injury and spinal cord injury, 1990-2016: A systematic analysis for the global burden of disease study, Lancet Neurology 18 (1) (2016) 56–87.

[2] M.C. Dewan, A. Rattani, S. Gupta, R.E. Baticulon, Y.C. Hung, M. Punchak, A. Agarwal, A.O. Adeley, M.G. Shrime, A.M. Rubiano, J.V. Rosenfeld, K.B. Park, Estimating the global incidence of traumatic brain injury, Journal of Neurosurgery 130 (4) (2018) 1080–1097.

[3] Center for Neurological Studies, Facts about brain injury, 2019,. https://www.neurologicstudies.com/facts-about-brain-injury.
[4] A. Goriely, M.G.D. Geers, G.A. Holzapfel, J. Jayamohan, A. Jerusalem, S. Sivaloganathan, W. Squier, J.A.W. van Dommelen, S. Waters, E. Kuhl, Mechanics of the brain: Perspectives, challenges, and opportunities, Biomechanics Modeling and Mechanobiology 14 (2015) 931–965.
[5] V. Balbi, A. Trotta, M. Destrade, A.N. Annaidh, Poynting effect of brain matter in torsion, Soft Matter 15 (2019) 5147.
[6] S. Budday, R. Nay, R. de Rooij, P. Steinmann, T. Wyrobek, T.C. Ovaert, E. Kuhl, Mechanical properties of gray and white matter brain tissue by indentation, Journal of the Mechanical Behavior of Biomedical Materials 46 (2015) 318–330.
[7] S. Budday, G. Sommer, C. Birkl, C. Langkammer, J. Jaybaeck, B.M. Kohnert, F. Paulsen, P. Steinmann, E. Kuhl, G.A. Holzapfel, Mechanical characterization of human brain tissue, Acta Biomaterialia 48 (2017a) 319–340.
[8] S. Budday, G. Sommer, J. Hayback, P. Steinmann, G.A. Holzapfel, E. Kuhl, Rheological characterization of human brain tissue, Acta Biomaterialia 60 (2017b) 315–329.
[9] M. Hoppstadter, D. Pullmann, R. Seydewitz, E. Kuhl, M. Bol, Correlating the microstructural architecture and macrostructural behaviour of the brain, Acta Biomaterialia 151 (2022) 379–395.
[10] T.P. Prevost, A. Balakrishnan, S. Suresh, S. Socrate, Biomechanics of brain tissue, Acta Biomaterialia 7 (2011) 83–95.
[11] B. Rashid, M. Destrade, M.D. Gilchrist, Mechanical characterization of brain tissue in compression at dynamic strain rates, Journal of the Mechanical Behavior of Biomedical Materials 10 (2012) 23–38.
[12] J. Weickenmeier, R. de Rooij, S. Budday, P. Steinmann, T.C. Ovaert, E. Kuhl, Brain stiffness increases with myelin content, Acta Biomaterialia 42 (2016) 265–272.
[13] J. Weickenmeier, M. Kurt, E. Ozkaya, M. Wintermark, K. Butts Pauly, E. Kuhl, Magnetic resonance elastography of the brain: A comparison between pigs and humans, Journal of the Mechanical Behavior of Biomedical Materials 77 (2018) 702–710.
[14] S. Budday, T.C. Ovaert, G.A. Holzapfel, P. Steinmann, E. Kuhl, Fifty shades of brain: A review on the material testing and modeling of brain tissue, Archives of Computational Methods in Engineering 27 (2020) 1187–1230.
[15] L.A. Mihai, L. Chin, P.A. Janmey, A. Goriely, A comparison of hyperelastic constitutive models applicable to brain and fat tissues, Journal of the Royal Society Interface 12 (2015) 20150486.
[16] L.A. Mihai, S. Budday, G.A. Holzapfel, E. Kuhl, A. Goriely, A family of hyperelastic models for human brain tissue, Journal of the Mechanics and Physics of Solids 106 (2017) 60–79.
[17] R. Moran, J.H. Smith, J.J. Garcia, Fitted hyperelastic parameters for human brain tissue from reported tension, compression, and shear tests, Journal of Biomechanics 47 (2014) 3762–3766.
[18] N. Atanasova, L. Todorovski, S. Dzeroski, B. Kompare, Application of automated model discovery from data and expert knowledge to a real-world domain: Lake glumso, Ecological Modeling 212 (2008) 92–98.
[19] M. Flaschel, S. Kumar, L. De Lorenzis, Unsupervised discovery of interpretable hyperelastic constitutive laws, Computer Methods in Applied Mechanics and Engineering 381 (2021) 113852.
[20] K. Linka, E. Kuhl, A new family of constitutive artificial neural networks towards automated model discovery, Computer Methods in Applied Mechanics and Engineering 403 (2023) 115731.
[21] K. Linka, A. Buganza Tepole, G.A. Holzapfel, E. Kuhl, Automated model discovery for skin: Discovering the best model, data, and experiment (2023), doi:10.1101/2022.12.19.520979.
[22] J. Bongard, H. Lipson, Automated reverse engineering of nonlinear dynamical systems, Proceedings of the National Academy of Sciences 104 (2007) 9943–9948.
[23] M. Schmidt, H. Lipson, Distilling free-form natural laws from experimental data, Science 324 (2009) 81–85.
[24] J.J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, Proceedings of the National Academy of Science 79 (1982) 2554–2558.
[25] M. Alber, A. Buganza Tepole, W. Cannon, S. De, S. Dura-Bernal, K. Garikipati, G.E. Karniadakis, W.W. Lytton, P. Perdikaris, L. Petzold, E. Kuhl, Integrating machine learning and multiscale modeling: Perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences, npj Digital Medicine 2 (2019) 115.
[26] F. Masi, I. Stefanou, P. Vannucci, V. Maffi-Berthier, Thermodynamics-based artificial neural networks for constitutive modeling, Journal of the Mechanics and Physics of Solids 147 (2021). 04277
[27] F. As'ad, P. Avery, C. Farhat, A mechanicsnformed artificial neural network approach in constitutive modeling, International Journal for Numerical Methods in Engineering 123 (2022) 2738–2759.
[28] A. Granados, F. Perez-Garcia, M. Schweiger, V. Vakharia, S.B. Vos, A. Miserocchi, A. McEvoy, S. Duncan, R. Sparks, S. Ourselin, A generative model of hyperelastic strain energy density functions for multiple tissue brain deformation, International Journal of Computer Assisted Radiology and Surgery. 16 (2021) 141–150.
[29] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, Journal of Computational Physics 378 (2019) 686–707.
[30] G.E. Karniadakis, I.G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, L. Yang, Physics-informed machine learning, Nature Reviews Physics 3 (2021) 422–440.
[31] K. Linka, M. Hillgartner, K.P. Abdolazizi, R.C. Aydin, M. Itskov, C.J. Cyron, Constitutive artificial neural networks: A fast and general approach to predictive data-driven constitutive modeling by deep learning, Journal of Computational Physics 429 (2021) 110010.
[32] K. Linka, A. Schafer, X. Meng, Z. Zou, G.E. Karniadakis, E. Kuhl, Bayesian physics-informed neural networks for real-world nonlinear dynamical systems, Computer Methods in Applied Mechanics and Engineering 402 (2022a) 115346.
[33] K. Linka, C. Cavinato, J.D. Humphrey, C.J. Cyron, Predicting and understanding arterial elasticity from key microstructural features by bidirectional deep learning by deep learning, Acta Biomaterialia 147 (2022b) 63–72.
[34] Y. Shen, K. Chandrashekhara, W.F. Breig, L.R. Oliver, Neural network based constitutive model for rubber material, Rubber Chemistry and Technology 77 (2004) 257–277.
[35] A. Ghaderi, V. Morovati, R. Dargazany, A physics-informed assembly for feed–forward neural network engines to predict inelasticity in cross-linked polymers, Polymers 12 (2020) 2628.
[36] D.K. Klein, M. Fernandez, R.J. Martin, P. Neff, O. Weeger, Polyconvex anisotropic hyperelasticity with neural networks, Journal of the Mechanics and Physics of Solics 159 (2022) 105703.
[37] C. Zopf, M. Kaliske, Numerical characterisation of uncured elastomers by a neural network based approach, Computers and Structures 182 (2017) 504–525.
[38] V. Tac, F. Sahli Costabal, A. Buganza Tepole, Data-driven tissue mechanics with polyconvex neural ordinary differential equations, Computer Methods in Applied Mechanics and Engineering 398 (2022) 115248.
[39] S. Kakaletsis, E. Lejeune, M.K. Rausch, Can machine learning accelerate soft material parameter identification from complex mechanical test data? Biomechanics and Modeling in Mechanobiology (2022), doi:10.1007/s10237-022-01631-z.
[40] G.A. Holzapfel, K. Linka, S. Sherifova, C. Cyron, Predictive constitutive modelling of arteries by deep learning, Journal of the Royal Socienty Interface 18 (2021) 20210411.
[41] L.R.G. Treloar, Stresses and birefringence in rubber subjected to general homogeneous strain, Proceedings of the Physical Society 60 (1948) 135–144.
[42] P.J. Blatz, W.L. Ko, Application of finite elastic theory to the deformation of rubbery materials, Transactions of the Society of Rheology 6 (1962) 223–251.
[43] M. Mooney, A theory of large elastic deformations, Journal of Applied Physics 11 (1940) 582–590.
[44] R.S. Rivlin, Large elastic deformations of isotropic materials. IV. further developments of the general theory, Philosophical Transactions of the Royal Society of London Series A 241 (1948) 379–397.
[45] H. Demiray, A note on the elasticity of soft biological tissues, Journal of Biomechanics 5 (1972) 309–311.
[46] A. Gent, A new constitutive relation for rubber, Rubber Chemistry and Technology 69 (1996) 59–61.
[47] G.A. Holzapfel, T.C. Gasser, R.W. Ogden, A new constitutive framework for arterial wall mechanics and comparative study of material models, Journal of Elasticity 61 (2000) 1–48.
[48] S.S. Antman, Nonlinear Problems of Elasticity. Second edition, Springer-Verlag, New York, 2005.
[49] G.A. Holzapfel, Nonlinear Solid Mechanics: A Continuum Approach to Engineering, John Wiley & Sons, Chichester, 2000.
[50] C. Truesdell, W. Noll, Non-linear field theories of mechanics, in: S. Flügge (Ed.), Encyclopedia of Physics, Vol. III/3, Spinger, Berlin, 1965.
[51] J. Ghaboussi, J.H. Garrett, X. Wu, Knowledge-based modeling of material behavior with neural networks, Journal of Engineering Mechanics 117 (1991) 132–153.
[52] R. Schulte, C. Karca, R. Ostwald, A. Menzel, Machine learning-assisted parameter identification for constitutive models based on concatenated normalised modes, European Journal of Mechanics A/Solids (2022).
[53] M. Planck, Vorlesungen über Thermodynamik, Verlag von Veit & Comp, Leipzig, 1897.
[54] W. Noll, A mathematical theory of the mechanical behavior of continuous media, Archive of Rational Mechanics Analysis 2 (1958) 197–226.
[55] J.M. Ball, Convexity conditions and existence theorems in nonlinear elasticity, Archive for Rational Mechanics and Analysis 63 (1977) 337–403.
[56] R.S. Rivlin, D.W. Saunders, Large elastic deformations of isotropic materials. VII. experiments on the deformation of rubber, Philosophical Transactions of the Royal Society of London Series A 243 (1951) 251–288.
[57] S. Hartmann, P. Neff, Polyconvexity of generalized polynomial-type hyperelastic strain energy functions for near-incompressibility, International Journal of Solids and Structures 40 (2003) 2767–2791.
[58] J.N. Fuhg, N. Bouklas, On physics-informed data-driven isotropic and anisotropic constitutive models through probabilistic machine learning and space-filling sampling, Computer Methods in Applied Mechanics and Engineering 394 (2022) 114915.
[59] J.N. Fuhg, N. Bouklas, R.E. Jones, Learning hyperelastic anisotropy from data via a tensor basis neural network, Journal of the Mechanics and Physics of Solids 168 (2022) 105022.
[60] P. Chen, J. Guilleminot, Polyconvex neural networks for hyperelastic constitutive models: A rectification approach, Mechanics Research Communications 125 (2022) 103993.
[61] A. Delfino, N. Stergiopulos, J.E. Moore, J.J. Meister, Residual strain effects on the stress field in a thick wall finite element model of the human carotid bifurcation, Journal of Biomechanics 30 (1997) 777–786.

[62] K. Miller, K. Chinzei, Mechanical properties of brain tissue in tension, Journal of Biomechanics 35 (2002) 483–490.
[63] B.R. Donnelly, J. Medige, Shear properties of human brain tissue, Journal of Biomechanical Engineering 119 (1997) 423–432.
[64] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444.
[65] C. Truesdell, Rational Thermodynamics, Lecture 5, McGraw-Hill, New York, 1969.
[66] A. Daw, A. Karpatne, W. Watkins, J. Read, V. Kumar, Physics-guided neural networks (PGNN): An application to lake temperature modeling, 2017. Arxiv:1710.11431.
[67] M.A. Bhouri, F. Sahli Costabal, H. Wang, K. Linka, M. Peirlinck, E. Kuhl, P. Perdikaris, COVID-19 dynamics across the US: A deep learning study of human mobility and social behavior, Computer Methods in Applied Mechanics and Engineering 382 (2021) 113891.
[68] R.W. Ogden, Large deformation isotropic elasticity – on the correlation of theory and experiment for incompressible rubberlike solids, Proceedings of the Royal Society London Series A 326 (1972) 565–584.
[69] S.R. St. Pierre, K. Linka, E. Kuhl, Principal-stretch-based constitutive neural networks autonomously discover a subclass of ogden models for human brain tissue, bioRxiv (2023), doi:10.1101/2023.01.14.524079.