journal of the mechanical behavior of biomedical materials 145 (2023) 106021

Contents lists available at ScienceDirect

Journal of the Mechanical Behavior of Biomedical Materials

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

Research paper

Automated model discovery for muscle using constitutive recurrent neural networks

Lucy M. Wang, Kevin Linka, Ellen Kuhl *

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

ARTICLE INFO

Dataset link: https://github.com/LivingMatterLa

Keywords:
Automated model discovery
Hyperelasticity
Viscoelasticity
Constitutive neural networks
Recurrent neural networks
Skeletal muscle

ABSTRACT

The stiffness of soft biological tissues not only depends on the applied deformation, but also on the deformation rate. To model this type of behavior, traditional approaches select a specific time-dependent constitutive model and fit its parameters to experimental data. Instead, a new trend now suggests a machine-learning based approach that simultaneously discovers both the best model and best parameters to explain given data. Recent studies have shown that feed-forward constitutive neural networks can robustly discover constitutive models and parameters for hyperelastic materials. However, feed-forward architectures fail to capture the history dependence of viscoelastic soft tissues. Here we combine a feed-forward constitutive neural network for the hyperelastic response and a recurrent neural network for the viscous response inspired by the theory of quasi-linear viscoelasticity. Our novel rheologically-informed network architecture discovers the time-independent initial stress using the feed-forward network and the time-dependent relaxation using the recurrent network. We train and test our combined network using unconfined compression relaxation experiments of passive skeletal muscle and compare our discovered model to a neo Hookean standard linear solid, to an advanced mechanics-based model, and to a vanilla recurrent neural network with no mechanics knowledge. We demonstrate that, for limited experimental data, our new constitutive recurrent neural network discovers models and parameters that satisfy basic physical principles and generalize well to unseen data. We discover a Mooney–Rivlin type two-term initial stored energy function that is linear in the first invariant I_1 and quadratic in the second invariant I_2 with stiffness parameters of 0.60 kPa and 0.55 kPa. We also discover a Prony-series type relaxation function with time constants of 0.362s, 2.54s, and 52.0s with coefficients of 0.89, 0.05, and 0.03. Our newly discovered model outperforms both the neo Hookean standard linear solid and the vanilla recurrent neural network in terms of prediction accuracy on unseen data. Our results suggest that constitutive recurrent neural networks can autonomously discover both model and parameters that best explain experimental data of soft viscoelastic tissues. Our source code, data, and examples are available at https://github.com/LivingMatterLab.

1. Introduction

The mechanical behavior of biological tissues exhibits many complexities that need to be taken into consideration in constitutive modeling (Holzapfel and Ogden, 2006). For example, we can observe nonlinearity (Nicolle et al., 2010), heterogeneity (Budday et al., 2017), anisotropy (Liu et al., 2020b), and viscoelasticity (Dehoff, 1978) in a variety of tissues ranging from tendon (Maganaris and Paul, 1999) to liver (Nicolle et al., 2010). As an alternative to traditional constitutive modeling, researchers are now exploring the ability of neural networks to capture the intricate mechanical response of various biological tissues (Liu et al., 2020a) including skin (Linka et al., 2023b; Tac et al., 2023a), arteries (Holzapfel et al., 2021), the brain (St. Pierre et al., 2023; Linka et al., 2023a, 2021b), and artificial meat (St. Pierre et al., 2023). Traditional constitutive modeling assumes a certain functional form, and fits the parameters to the measured data. This may introduce significant modeling errors if the assumed functional form does not represent the material behavior well. In contrast, neural networks have the potential to learn both the functional form and its parameters (Kalina et al., 2022; As'ad et al., 2022; Shen et al., 2004), creating a more accurate representation of the material behavior.

A classical feed-forward neural network can capture traits such as nonlinearity (Linka et al., 2021b, 2023a), heterogeneity (St. Pierre et al., 2023), and anisotropy (Holzapfel et al., 2021; Linka et al., 2023b), but these architectures are not well suited for modeling viscoelasticity. Instead, recurrent neural network architectures can model history-dependent behaviors such as viscoelasticity, where information from previous time steps informs the material response at the

* Corresponding author.
E-mail addresses: lwang8@stanford.edu (L.M. Wang), kevin.linka@me.com (K. Linka), ekuhl@stanford.edu (E. Kuhl).

https://doi.org/10.1016/j.jmbbm.2023.106021
Received 9 May 2023; Received in revised form 18 June 2023; Accepted 6 July 2023
Available online 11 July 2023

current time point (Bonatti and Mohr, 2021). In the classical deep learning realm, recurrent neural networks are successfully applied to sequence problems (Medsker and Jain, 1999) like natural language processing (Goldberg, 2016), signal processing (Übeyli, 2009), and robot control (Zhang and Chu, 2012). In constitutive modeling, researchers are now beginning to evaluate the potential of recurrent neural networks to model plasticity (Borkowski et al., 2022; Tancogne-Dejean et al., 2021), viscoelasticity (Abdolazizi et al., 2023; Chen, 2021; Oeser and Freitag, 2016), and fatigue damage (Yang et al., 2021). These recurrent neural networks also include various approaches such as directly predicting stress from strain input (Chen, 2021) and incorporating the recurrent neural network as a surrogate for micro level response in multiscale simulations (Ghavamian and Simone, 2019).

In the early applications of recurrent neural networks to constitutive modeling, researchers directly adopted recurrent network architectures from the classical deep learning field (Zhu et al., 2011; Chen, 2021; Gorji et al., 2020; Tancogne-Dejean et al., 2021). While these classical recurrent network architectures reproduce history-dependent material behaviors, they typically involve hundreds to thousands of parameters, if not more. As a result, classical recurrent networks require large amounts of training data which may not be practical when building a constitutive model based on limited experimental measurements (Alber et al., 2019). Furthermore, the parameters in these classical networks have no clear physical interpretation, and their predictions may violate physical laws and constraints (Linka and Kuhl, 2023).

To address these concerns, researchers have proposed physics-informed neural networks that incorporate physics knowledge into neural network design (Danoun et al., 2022; Zhang et al., 2020). Integrating our prior physics knowledge reduces the amount of required training data and constrains solutions to a physically admissible subspace. Two conceptually different approaches have emerged to incorporate physics knowledge: The first approach adds physical constraints to the loss function to enforce thermodynamic principles (Raissi et al., 2019; Borkowski et al., 2022; Linka et al., 2022); the second approach hardwires physical constraints into the network input, architecture, and output to learn stored energy functions and evolution laws for internal state variables (Linka et al., 2021a; He and Chen, 2022; Masi et al., 2021; Linka and Kuhl, 2023). The stress then follows from these intermediate functions by direct calculation of mechanics equations.

Here we propose a new physics-informed approach that combines a feed-forward hyperelastic neural network (St. Pierre et al., 2023; Linka and Kuhl, 2023) with a recurrent linear viscous neural network to model the viscoelastic behavior of soft biological tissues. To motivate our new network architecture, we briefly revisit the theory of quasi-linear viscoelasticity (Fung et al., 1970) in Section 2, and illustrate how it decomposes the total stress response into a time-independent initial elastic stress and a time-dependent viscous overstress. In Section 3, we map this theory onto a new network architecture that integrates a feed-forward hyperelastic network and a recurrent viscous network. We illustrate the modular nature of this approach by probing two alternative hyperelastic networks: principal-stretch-based (St. Pierre et al., 2023) and invariant-based (Linka and Kuhl, 2023). In Section 4, we test and train both networks on five unconfined compression relaxation tests of passive skeletal muscle (Van Loocke et al., 2008), and compare both against an overly constrained model, the neo Hookean standard linear solid, a traditional phenomenological model, the Van Loocke Lyons Simms model, and an overly flexible model, a vanilla recurrent neural network. We perform two separate tasks, train on one and test on four versus train on four and test on one, and demonstrate that our new networks can uniquely discover model and parameters that best explain the experimental data. We close with some limitations of our study in Section 5 and with a brief conclusion in Section 6.

2. Theory of quasi-linear viscoelasticity

We begin by revisiting the theory of quasi-linear viscoelasticity, first in three dimensions, and then for the special case of uniaxial tension and compression. 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. 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,

$$\boldsymbol{F} = \nabla_{\boldsymbol{X}} \boldsymbol{\varphi} = \sum_{i=1}^{3} \lambda_i \, \boldsymbol{n}_i \otimes \boldsymbol{N}_i \quad \text{with} \quad J = \det(\boldsymbol{F}) > 0 \,. \tag{1}$$

The spectral representation introduces the principal stretches λ_i and the principal directions $\boldsymbol{N}_i$ and $\boldsymbol{n}_i$ in the undeformed and deformed configurations, where $\boldsymbol{F} \cdot \boldsymbol{N}_i = \lambda_i \boldsymbol{n}_i$. We further introduce the left Cauchy Green deformation tensor,

$$\boldsymbol{b} = \boldsymbol{F} \cdot \boldsymbol{F}^{\mathrm{t}} = \sum_{i=1}^{3} \lambda_i^2 \, \boldsymbol{n}_i \otimes \boldsymbol{n}_i \,. \tag{2}$$

To characterize the isotropic material behavior, we introduce the three principal invariants, I_1, I_2, I_3, either in terms of the principal stretches $\lambda_1, \lambda_2, \lambda_3$,

$$I_1 = \lambda_1^2 + \lambda_2^2 + \lambda_3^2 \quad I_2 = \lambda_1^2\lambda_2^2 + \lambda_2^2\lambda_3^2 + \lambda_1^2\lambda_3^2 \quad I_3 = \lambda_1^2\lambda_2^2\lambda_3^2, \tag{3}$$

or in terms of the deformation gradient $\boldsymbol{F}$, with derivatives, $\partial_{\boldsymbol{F}} I_1$, $\partial_{\boldsymbol{F}} I_2$, $\partial_{\boldsymbol{F}} I_3$,

$$\begin{aligned} I_1 &= \boldsymbol{F} : \boldsymbol{F} & \partial_{\boldsymbol{F}} I_1 &= 2\,\boldsymbol{F} \\ I_2 &= \tfrac{1}{2}\,[I_1^2 - [\boldsymbol{F}^{\mathrm{t}} \cdot \boldsymbol{F}] : [\boldsymbol{F}^{\mathrm{t}} \cdot \boldsymbol{F}]] & \partial_{\boldsymbol{F}} I_2 &= 2\,[I_1 \boldsymbol{F} - \boldsymbol{F} \cdot \boldsymbol{F}^{\mathrm{t}} \cdot \boldsymbol{F}] \\ I_3 &= \det(\boldsymbol{F}^{\mathrm{t}} \cdot \boldsymbol{F}) = J^2 & \partial_{\boldsymbol{F}} I_3 &= 2\,I_3\,\boldsymbol{F}^{-\mathrm{t}}. \end{aligned} \tag{4}$$

For isotropic, perfectly incompressible materials, the third invariant always remains identical to one, $I_3 = \lambda_1^2\lambda_2^2\lambda_3^2 = J^2 = 1$, and the set of invariants reduces to I_1 and I_2. Following standard arguments of thermodynamics, we introduce an instantaneous elastic free energy function $\psi_0(\boldsymbol{F})$ from which we derive the instantaneous elastic Cauchy stress $\boldsymbol{\sigma}_0$,

$$\boldsymbol{\sigma}_0 = \frac{1}{J} \frac{\partial \psi_0}{\partial \boldsymbol{F}} \cdot \boldsymbol{F}^{\mathrm{t}} - p\,\boldsymbol{I} \,, \tag{5}$$

where p is the hydrostatic pressure that we determine from the boundary conditions and $\boldsymbol{I}$ is the second-order unit tensor. For an initial free energy expressed in terms of the principal stretches, $\psi_0(\lambda_1, \lambda_2, \lambda_3)$, the initial Cauchy stress takes the following explicit representation,

$$\boldsymbol{\sigma}_0 = \frac{1}{J} \frac{\partial \psi_0}{\partial \boldsymbol{F}} \cdot \boldsymbol{F}^{\mathrm{t}} - p\,\boldsymbol{I} = \frac{1}{J} \sum_{i=1}^{3} \lambda_i \frac{\partial \psi_0}{\partial \lambda_i} \boldsymbol{n}_i \cdot \boldsymbol{n}_i - p\,\boldsymbol{I} \,, \tag{6}$$

whereas for an energy function in terms of the invariants $\psi_0(I_1, I_2)$, the initial Cauchy stress takes the following form,

$$\boldsymbol{\sigma}_0 = 2 \left[\frac{\partial \psi_0}{\partial I_1} + I_1 \frac{\partial \psi_0}{\partial I_2} \right] \boldsymbol{b} - 2 \frac{\partial \psi_0}{\partial I_2} \boldsymbol{b}^2 - p\,\boldsymbol{I} \,. \tag{7}$$

We pull the Cauchy stress back onto the undeformed reference configuration to obtain the initial Piola Kirchhoff stress,

$$\boldsymbol{S}_0 = J\,\boldsymbol{F}^{-1} \cdot \boldsymbol{\sigma}_0 \cdot \boldsymbol{F}^{-\mathrm{t}} \,. \tag{8}$$

Following the quasi-linear viscoelastic theory (Fung et al., 1970), we introduce the viscoelastic Piola Kirchhoff stress through the following convolution integral (Pascalis et al., 2014),

$$\boldsymbol{S}_t = \int_{-\infty}^{t} \mathbb{G}(t-s) : \frac{\mathrm{d}}{\mathrm{d}s} \boldsymbol{S}_0 \, \mathrm{d}s \tag{9}$$

where $\mathbb{G}(t)$ is the time-dependent fourth-order viscous relaxation tensor and $\mathrm{d}\boldsymbol{S}_0/\mathrm{d}s$ is the material time derivative of the instantaneous elastic Piola Kirchhoff stress according to Eq. (8). We assume that the relaxation is isotropic, and express it in terms of a discrete Prony series,

$$\mathbb{G}(t) = G(t)\,\mathbb{I} \quad \text{with} \quad G(t) = \gamma_\infty + \sum_{i=1}^{n_{\mathrm{prn}}} \gamma_i \, \exp\left(-t/\tau_i\right), \tag{10}$$

where $G(t)$ is the time-dependent viscous relaxation function, $\mathbb{I}$ is the fourth-order unit tensor, γ_∞ and γ_i are the long-term modulus and the viscous relaxation coefficients with $\gamma_\infty + \sum_{i=1} \gamma_i = 1$ and $0 \leq \gamma_\infty, \gamma_i \leq 1$, τ_i are the viscous relaxation times of the $i = 1, \dots, n_{\text{prn}}$ Maxwell elements, with $\tau_i > 0$. We substitute the relaxation function (10) into the total stress expression (9), and push it forward into the deformed current configuration to obtain the total Cauchy stress,

$$\boldsymbol{\sigma}(t) = \gamma_\infty \, \boldsymbol{\sigma}_0(t) + \sum_{i=1}^{n_{\text{prn}}} \gamma_i \boldsymbol{h}_i(t) \,. \tag{11}$$

Here we have introduced a set of stress-like variables, the tensor-valued spatial overstresses $\boldsymbol{h}_i$ for each Maxwell element i, as a push-forward of their material counterparts,

$$\boldsymbol{h}_i(t) = \frac{1}{J} \boldsymbol{F} \cdot \int_0^t \exp\left(-\frac{t-s}{\tau_i}\right) \frac{\mathrm{d}}{\mathrm{d}s} \boldsymbol{S}_0 \, \mathrm{d}s \cdot \boldsymbol{F}^{\mathrm{t}} \,. \tag{12}$$

Since the exponential function gradually decays to zero, the viscous overstresses in Eq. (12) gradually decay in time, $\boldsymbol{h}_i(t \to \infty) = \boldsymbol{0}$. The total stress, $\boldsymbol{\sigma}$, in Eq. (11) converges towards the long-term equilibrium stress, $\boldsymbol{\sigma}(t \to \infty) = \gamma_\infty \, \boldsymbol{\sigma}_0$.

Uniaxial tension and compression. For the special case of uniaxial tension and compression, we deform the specimen in one direction, $F_{11} = \lambda_1 = \lambda$, where the stretch $\lambda = l/L$ denotes the relative change in length. 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 inverse of 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,

$$\boldsymbol{F} = \text{diag}\,\{\lambda, \lambda^{-1/2}, \lambda^{-1/2}\} \quad \text{with} \quad J = 1 \,, \tag{13}$$

and left Cauchy Green deformation tensor,

$$\boldsymbol{b} = \text{diag}\,\{\lambda^2, \lambda^{-1}, \lambda^{-1}\} \,, \tag{14}$$

we calculate the first and second invariants and their derivatives,

$$\begin{aligned} I_1 &= \lambda^2 + 2/\lambda \quad \partial_\lambda I_1 = 2\,[\lambda - 1/\lambda^2] \\ I_2 &= 2\lambda + 1/\lambda^2 \quad \partial_\lambda I_2 = 2\,[1 - 1/\lambda^3] \,. \end{aligned} \tag{15}$$

We derive the instantaneous elastic Cauchy stress $\sigma_0 = \sigma_{11}$ in the loading direction in terms of the instantaneous elastic free energy function ψ_0 for perfectly incompressible materials using standard arguments of thermodynamics,

$$\sigma_{ii} = \frac{\partial \psi}{\partial I_1} \frac{\partial I_1}{\partial \lambda_i} \lambda_i + \frac{\partial \psi}{\partial I_2} \frac{\partial I_2}{\partial \lambda_i} \lambda_i - p \,, \tag{16}$$

for $i = 1, 2, 3$. Here, p denotes the hydrostatic pressure that we determine from the zero stress condition in the transverse directions, $\sigma_{22} = 0$ and $\sigma_{33} = 0$, using Eq. (16) as $p = [2/\lambda] \, \partial\psi/\partial I_1 + [2\lambda + 2/\lambda^2] \, \partial\psi/\partial I_2$. This results in the following instantaneous elastic uniaxial stress–stretch relation for perfectly incompressible, isotropic materials,

$$\sigma_0 = 2 \left[\frac{\partial \psi}{\partial I_1} + \frac{1}{\lambda} \frac{\partial \psi}{\partial I_2} \right] \left[\lambda^2 - \frac{1}{\lambda} \right] . \tag{17}$$

The underlying assumption of the theory of quasi-linear viscoelasticity is that we can multiplicatively decompose the total viscoelastic stress $\sigma(t)$ into the strain-dependent function σ_0 from Eq. (17) and a dimensionless time-dependent function $g(t)$ (Fung et al., 1970). We can then represent the total viscoelastic Cauchy stress through the convolution integral,

$$\sigma(t) = \int_{-\infty}^t g(t-s) \frac{\mathrm{d}}{\mathrm{d}s} \sigma_0 \, \mathrm{d}s \,, \tag{18}$$

where $g(t)$ is the time-dependent viscous relaxation function and $\mathrm{d}\sigma_0/\mathrm{d}s$ is the material time derivative of the instantaneous elastic uniaxial Cauchy stress according to Eq. (17). According to the theory of quasi-linear viscoelasticity and motivated by the generalized Maxwell model, we choose a discrete Prony series for the relaxation function,

$$g(t) = \gamma_\infty + \sum_{i=1}^{n_{\text{prn}}} \gamma_i \, \exp(-t/\tau_i), \tag{19}$$

where γ_∞ and γ_i are the long-term modulus and the viscous relaxation coefficients with $\gamma_\infty + \sum_i \gamma_i = 1$ and $0 \leq \gamma_\infty, \gamma_i \leq 1$, τ_i are the viscous relaxation times of the $i = 1, \dots, n_{\text{prn}}$ Maxwell elements, with $\tau_i > 0$. By substituting Eq. (19) into Eq. (18) we obtain a convenient expression for the total Cauchy stress,

$$\sigma(t) = \gamma_\infty \, \sigma_0(t) + \sum_{i=1}^{n_{\text{prn}}} \gamma_i h_i(t) \,, \tag{20}$$

where we have introduced a set of new stress-like variables, the scalar-valued overstresses, $h_i(t)$, for each Maxwell element i,

$$h_i(t) = \int_0^t \exp\left(-(t-s)/\tau_i\right) \frac{\mathrm{d}}{\mathrm{d}s} \sigma_0(s) \, \mathrm{d}s \,. \tag{21}$$

In general, the above convolution integral does not have a closed form solution, and we solve it numerically using an explicit Euler forward time integration scheme (Taylor et al., 1970). We discretize the time interval of interest and march forward from the previous time point t_n to the current time point t_{n+1} with the discrete time step size, $\Delta t = t_{n+1} - t_n$, using the following update equations for the total stress (Simo and Hughes, 2000; Holzapfel, 2002),

$$\sigma_{n+1} = \gamma_\infty \, \sigma_{0,n+1} + \sum_{i=1}^{n_{\text{prn}}} \gamma_i \, h_{i,n+1}, \tag{22}$$

in terms of the $i = 1, \dots, n_{\text{prn}}$ overstresses,

$$h_{i,n+1} = \exp(-\Delta t/\tau_i) h_{i,n} + \exp(-\Delta t/2\tau_i)[\sigma_{0,n+1} - \sigma_{0,n}]. \tag{23}$$

Before we move on to the next time step, we need to store both the total stress σ_{n+1} and the overstresses $h_{i,n+1}$ of the current time step. Motivated by the considerations in this section, we will now introduce four different approaches to model the viscoelastic behavior of passive skeletal muscle.

3. Constitutive recurrent neural networks

In this section, we propose two constitutive recurrent neural networks inspired and informed by the quasi-linear viscoelastic theory in Section 2: one principal-stretch-based and one invariant-based. For comparison, we also introduce a neo Hookean standard linear solid, a Van Loocke Lyons Simms model, and a vanilla recurrent neural network. Fig. 1 illustrates our constitutive recurrent neural network that combines a feed-forward network, top left and right, to discover the initial elastic stress and its parameters with a recurrent neural network, bottom, to discover the viscous overstress with its viscous parameters. We demonstrate both a principal-stretch-based network, top left, and a feed-forward invariant-based network, top right, for the initial elastic stress.

Principal-stretch-based elastic stress. The principal-stretch-based network in Fig. 1, top left, is parameterized in terms of the principal stretches, $\lambda_1, \lambda_2, \lambda_3$, that act as input to a single dense layer (St. Pierre et al., 2023). Its initial stored energy function is inspired by the classical Ogden model (Ogden, 1972) with n_{ogd} terms with fixed exponents,

$$\psi_0 = \sum_{j=1}^{n_{\text{ogd}}/2} w_{1,j}(\lambda_1^{-j} + \lambda_2^{-j} + \lambda_3^{-j} - 3) + w_{1,2j}(\lambda_1^{j} + \lambda_2^{j} + \lambda_3^{j} - 3), \tag{24}$$

where $w_{1,j}$ and $w_{1,2j}$ are the weights learned by the network during training that we constrain to be positive, and $(\lambda_1^{-j} + \lambda_2^{-j} + \lambda_3^{-j} - 3)$ and $(\lambda_1^{j} + \lambda_2^{j} + \lambda_3^{j} - 3)$, are the activation functions that raise the principal stretches to fixed powers $j = -\frac{1}{2} n_{\text{ogd}}, \dots, +\frac{1}{2} n_{\text{ogd}}$. Unlike classical neural network architectures, the activation functions are applied first and the weights second. We can then derive the initial Cauchy stress as illustrated in Section 2 using Eq. (6) for the full stress tensor,

$$\boldsymbol{\sigma}_0 = \sum_{i=1}^{3} \lambda_i \frac{\partial \psi}{\partial \lambda_i} \boldsymbol{n}_i \otimes \boldsymbol{n}_i - p \, \boldsymbol{I}, \tag{25}$$

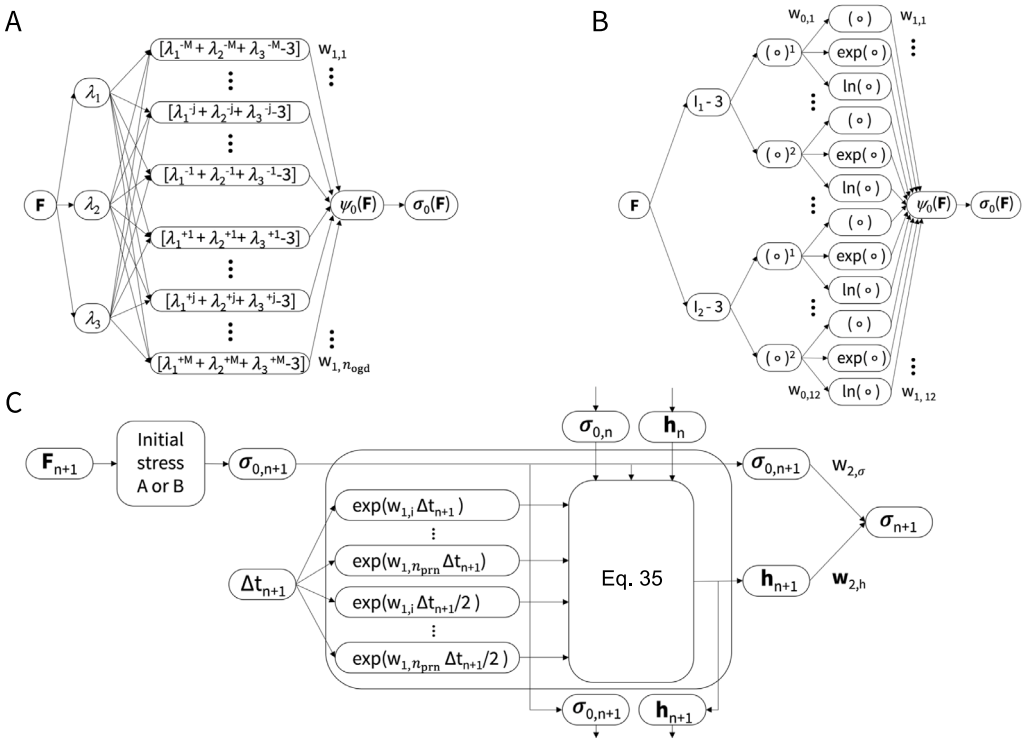

Fig. 1. Principal-stretch-based and invariant-based recurrent neural network. The network takes the deformation gradient $\boldsymbol{F}_{n+1}$ and time step size Δt_{n+1} as input and outputs the total Cauchy stress $\boldsymbol{\sigma}_{n+1}$. The network architecture combines a feed-forward constitutive neural network block to discover the time-independent initial stress function $\boldsymbol{\sigma}_{0,n+1}$ and its elastic parameters, followed by a recurrent neural network block to discover the time-dependent relaxation function and its Prony parameters. The network has a modular architecture and can seamlessly integrate different building blocks, for example, a principal-stretch or invariant-based neural network for the initial stress.

where $\boldsymbol{n}_i$ are the principal directions, or Eq. (16) for the initial uniaxial stress in the loading direction,

$$\sigma_0 = \lambda_1 \frac{\partial \psi}{\partial \lambda_1} - p\,. \tag{26}$$

The hydrostatic pressure p follows from the zero-normal-stress condition as described in Section 2.

Invariant-based elastic stress. In contrast to the principal-stretch-based network, the invariant-based network in Fig. 1, top right, is parameterized in terms of the first and second invariants, I_1, I_2, Linka et al. (2023a). The network calculates both invariants using Eq. (4), and feeds them into two hidden layers with linear $(\circ)$ and quadratic $(\circ)^2$ activation functions in the first layer and linear $(\circ)$, exponential $(\exp((\circ)-1))$, and natural logarithmic $(-\ln(1-(\circ)))$ activation functions in the second layer. The resulting initial stored energy function has a total of $n_{\text{inv}} = 12$ terms,

$$\begin{aligned}\psi_0 \;=\; & w_{1,1}\;\; w_{0,1}[I_1-3] && + w_{1,2}\;\; \exp(w_{0,2}[I_1-3]-1)\\ & - w_{1,3}\;\; \ln(1-w_{0,3}[I_1-3]) && + w_{1,4}\;\; w_{0,4}[I_1-3]^2\\ & + w_{1,5}\;\; \exp(w_{0,5}[I_1-3]^2-1) && - w_{1,6}\;\; \ln(1-w_{0,6}[I_1-3]^2)\\ & + w_{1,7}\;\; w_{0,7}[I_2-3] && + w_{1,8}\;\; \exp(w_{0,8}[I_2-3]-1)\\ & - w_{1,9}\;\; \ln(1-w_{0,9}[I_2-3]) && + w_{1,10}\;\; w_{0,10}[I_2-3]^2\\ & + w_{1,11}\;\; \exp(w_{0,11}[I_2-3]^2-1) && - w_{1,12}\;\; \ln(1-w_{0,12}[I_2-3]^2),\end{aligned} \tag{27}$$

where $w_{0,1-12}$ and $w_{1,1-12}$ are the weights of the first and second layers, and are constrained to be positive. We note that the four logarithmic terms, $(-\ln(1-(\circ)))$, are only physically meaningful for positive arguments, $(1-w\,(\circ))$, which might require special attention in the case of extreme tension with large I_1 and I_2. We can then derive the initial Cauchy stress as illustrated in Section 2 using Eq. (7) for the full stress tensor,

$$\boldsymbol{\sigma}_0 = 2\left[\frac{\partial \psi_0}{\partial I_1} + I_1 \frac{\partial \psi_0}{\partial I_2}\right]\boldsymbol{b} - 2\frac{\partial \psi_0}{\partial I_2}\boldsymbol{b}^2 - p\,\boldsymbol{I}\,, \tag{28}$$

where the hydrostatic pressure p follows from the zero-normal-stress condition, or using Eq. (17) for the initial uniaxial stress in the loading direction,

$$\sigma_0 = 2\left[\frac{\partial \psi}{\partial I_1} + \frac{1}{\lambda}\frac{\partial \psi}{\partial I_2}\right]\left[\lambda^2 - \frac{1}{\lambda}\right]. \tag{29}$$

Time-dependent viscous overstress. Fig. 1, bottom, shows the recurrent neural network inspired by the relaxation function of the Prony series of Eq. (19),

$$g(t) = \gamma_\infty + \sum_{i=1}^{n_{\text{prn}}} \gamma_i \exp(-t/\tau_i)\,, \tag{30}$$

from which we derive the scalar-valued overstresses $h_{i,n+1}$ according to Eq. (21),

$$h_i(t) = \int_0^t \exp\left(-(t-s)/\tau_i\right)\frac{\mathrm{d}}{\mathrm{d}s}\sigma_0(s)\,\mathrm{d}s\,. \tag{31}$$

Time-discrete update equations. Following Section 2, we obtain the time-discrete updates of the total stress according to Eq. (22),

$$\sigma_{n+1} = \gamma_\infty\,\sigma_{0,n+1} + \sum_{i=1}^{n_{\text{prn}}} \gamma_i\, h_{i,n+1}\,, \tag{32}$$

and of the overstresses according to Eq. (23),

$$h_{i,n+1} = \exp(-\Delta t/\tau_i)h_{i,n} + \exp(-\Delta t/2\tau_i)[\sigma_{0,n+1} - \sigma_{0,n}]. \tag{33}$$

The architecture of our custom recurrent neural network in Fig. 1, bottom, is inspired by these two update formulas. Its inputs are the initial stress σ_0, either from Eq. (26) for the principal-stretch-based neural network, or from Eq. (29) for the invariant-based neural network, and the time step size Δt. The recurrent neural network first calculates two sets of intermediate terms, a_i and b_i, from the time step size, Δt_{n+1},

$$a_i = \exp(w_{1,i}\,\Delta t_{n+1}) \;\text{ and }\; b_i = \exp(w_{1,i}\,\Delta t_{n+1}/2), \tag{34}$$

where $w_{1,i}$ are the $i = 1, \dots, n_{\text{prn}}$ weights that the network learns during training. Upon comparison with Eq. (33), we see that the network weights are equal to the negative inverse relaxation times, $w_{1,i} = -1/\tau_i$. To ensure that the time constants are always positive, we calculate τ_i from the network weights $w_{1,i}$ and then apply a ReLU function to the calculated τ_i values. After this initial layer, the network updates the overstresses h_i using Eq. (33),

$$h_{i,n+1} = a_i h_{i,n} + b_i(\sigma_{0,n+1} - \sigma_{0,n}), \tag{35}$$

where a_i and b_i are the intermediate terms from Eqs. (34), $\sigma_{0,n+1}$ is the initial stress output from the principal-stretch-based or invariant-based neural network, and $\sigma_{0,n}$ and $h_{i,n}$ are passed forward from the previous time step t_n. The final layer is a dense layer that updates the total Cauchy stress according to Eq. (32),

$$\sigma_{n+1} = w_{2,0}\, \sigma_{0,n+1} + \sum_{i=1}^{n_{\text{prn}}} w_{2,i}\, h_{i,n+1}, \tag{36}$$

where $w_{2,0}$ and $w_{2,i}$ are the weights learned by the model during training. Upon comparison to Eq. (32), we see that the weights are equal to the long-term modulus, $w_{2,0} = \gamma_\infty$, and the viscous relaxation coefficients, $w_{2,i} = \gamma_i$. We constrain the weights to be positive and to sum to one in accordance with $\gamma_\infty + \sum_i \gamma_i = 1$.

Loss function. We select a principal-stretch-based network with $n_{\text{ogd}} = 20$ Ogden terms in Eq. (24) and $n_{\text{prn}} = 10$ Prony series terms in Eq. (36). For these values, the principal-stretch-based network has a total of 41 trainable parameters or weights, 20 for the principal stretch terms $(\lambda_1^j + \lambda_2^j + \lambda_3^j - 3)$, ten for the negative inverse relaxation times $-1/\tau_i$, one for the initial stress $\sigma_{0,n+1}$, and ten for the overstresses $h_{i,n+1}$. We select an invariant-based network with $n_{\text{inv}} = 12$ invariant terms in Eq. (27) and $n_{\text{prn}} = 10$ Prony series terms in Eq. (30). For these values, the invariant-based network has a total of 45 trainable parameters or weights, two times twelve for the invariant terms, ten for the negative inverse relaxation times $-1/\tau_i$, one for the initial stress $\sigma_{0,n+1}$, and ten for the overstresses $h_{i,n+1}$. We learn these network weights $\boldsymbol{\theta} = \{\boldsymbol{w}\}$ by minimizing a loss function parameterized in terms of the stretch λ and time t. To select the type of loss function, we compare the mean absolute error, the mean squared error, and the logarithmic hyperbolic cosine functions. We choose the loss function of mean squared error type,

$$L(\boldsymbol{\theta}; \lambda, t) = \frac{1}{n_{\text{train}}} \frac{1}{n_{\text{time}}} \sum_{i=1}^{n_{\text{train}}} \omega_i \sum_{j=1}^{n_{\text{time}}} (\sigma(\lambda_{ij}, t_{ij}) - \hat{\sigma}_{ij})^2 \to \min, \tag{37}$$

for which the error is smallest. Here $i = 1, \dots, n_{\text{train}}$ denotes the number of training sets, $j = 1, \dots, n_{\text{time}}$ denotes the number of data points in the time series, and $(\sigma(\lambda_{ij}, t_{ij}) - \hat{\sigma}_{ij})$ is the difference between the model stress $\sigma(\lambda_{ij}, t_{ij})$ at stretch λ_{ij} and time t_{ij} and the experimental stress $\hat{\sigma}_{ij}$. We weight the loss associated with each training set with the coefficient, $\omega_i = \bar{\sigma}_i^{-1} / \sum_{j=1}^{n_{\text{train}}} \bar{\sigma}_j^{-1}$ to normalize the contributions from each of the training sets since they have different mean stress values $\bar{\sigma}_i$. We train both networks by minimizing the loss function for 5000 epochs using the Adam optimizer, with a learning rate $\alpha = 0.001$ and parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$.

Regularization. To investigate the effect of regularization on model discovery, we add a penalty term to the loss function in Eq. (37) to penalize network weights with large magnitudes. After preliminary analysis of both L1 and L2 regularization, we choose to further investigate the L2 regularization, which delivered a better performance. We assess the effect of L2 regularization in both the feed-forward network and the recurrent neural network. With regularization in the feed-forward network, the loss function becomes

$$L(\boldsymbol{\theta}; \lambda, t) = \frac{1}{n_{\text{train}}} \frac{1}{n_{\text{time}}} \sum_{i=1}^{n_{\text{train}}} \omega_i \sum_{j=1}^{n_{\text{time}}} (\sigma(\lambda_{ij}, t_{ij}) - \hat{\sigma}_{ij})^2 + \alpha \sum_{k=1}^{n_{\text{node}}} w_{1,k}^2 . \tag{38}$$

Here, α is the regularization parameter and $w_{1,k}$ are the weights of the last hidden layer of the feed-forward network with either $n_{\text{node}} = n_{\text{ogd}} = 20$ or $n_{\text{node}} = n_{\text{inv}} = 12$. We compare the effect of varying regularization parameters, α, ranging from 10^{-7} to 10^{-1} on the invariant-based network. With regularization in the recurrent network, the loss function becomes

$$L(\boldsymbol{\theta}; \lambda, t) = \frac{1}{n_{\text{train}}} \frac{1}{n_{\text{time}}} \sum_{i=1}^{n_{\text{train}}} \omega_i \sum_{j=1}^{n_{\text{time}}} (\sigma(\lambda_{ij}, t_{ij}) - \hat{\sigma}_{ij})^2 + \beta \left[w_{2,\sigma}^2 + \sum_{k=1}^{n_{\text{prn}}} (w_{1,k}^2 + w_{2,k}^2) \right]. \tag{39}$$

Here, β is the regularization parameter and $w_{2,\sigma}$, $w_{1,k}$, and $w_{2,k}$ are the weights of the recurrent neural network. We compare the effect of varying β from 10^{-4} to 10^{-1}.

Special case: Neo Hookean standard linear solid. A special case of both the principal-stretch and invariant-based recurrent neural networks is the neo Hookean standard linear solid. It is the simplest of all physics-informed models, and also the most frequently used. We use this model as an overly-constrained, low-end comparison and expect that it will have difficulties fitting the nonlinear elastic behavior, but will not display non-physical responses for small data. The initial stored energy function of the neo Hookean standard linear solid is linear in the sum of the three principal stretches, λ_1, λ_2, λ_3, squared, or in other words, linear in the first invariant I_1,

$$\psi_0 = \frac{1}{2}\mu\,[\lambda_1^2 + \lambda_2^2 + \lambda_3^2 - 3] = \frac{1}{2}\mu\,[I_1 - 3]\,. \tag{40}$$

The shear modulus μ follows from the weights of the principal-stretch-based network as $\mu = 2w_{1,12}$ and from the weights of the invariant-based network as $\mu = 2w_{0,1}w_{1,1}$ with all other weights identical to zero. For the special case of uniaxial tension and compression, the instantaneous elastic uniaxial stress–stretch relation for a perfectly incompressible, isotropic material of Eq. (17) becomes

$$\sigma_0 = \mu\,[\lambda^2 - 1/\lambda]\,. \tag{41}$$

For the neo Hookean standard linear solid, the Prony series only has a single one term, $n_{prn} = 1$. To numerically solve for its total stress, we discretize the time interval of interest and adopt an explicit Euler forward scheme to march from time t_n to t_{n+1} in increments of $\Delta t = t_{n+1} - t_n$ using the update rules according to Eq. (22) for the total stress,

$$\sigma_{n+1} = \gamma_\infty\, \sigma_{0,n+1} + \gamma\, h_{n+1}\,, \tag{42}$$

and Eq. (23) for the viscous overstress,

$$h_{n+1} = \exp(-\Delta t/\tau) h_n + \exp(-\Delta t/2\tau)(\sigma_{0,n+1} - \sigma_{0,n})\,. \tag{43}$$

The classical neo Hookean standard linear solid has four parameters that we need to fit to the data: the shear modulus μ to characterize the instantaneous elastic behavior, and the long-term modulus γ_∞, the relaxation coefficient γ, and the relaxation time τ to characterize the time-dependent viscous behavior.

Benchmark case: Van Loocke Lyons Simms model. As a benchmark comparison, we adopt a constitutive model inspired by the initial model for the muscle data of our current study (Van Loocke et al., 2008). This model follows the quasi-linear viscoelasticity theory of Section 2. Its initial stored energy is a cubic polynomial function,

$$\psi_0 = \frac{1}{2}\mu_1\,[I_1 - 3] + \frac{1}{2}\mu_2\,[I_1 - 3]^2 + \frac{1}{2}\mu_3\,[I_1 - 3]^3, \tag{44}$$

where μ_1, μ_2, and μ_3 are stiffness-like material parameters. The instantaneous elastic stress follows from this initial stored energy function using Eq. (17),

$$\sigma_0 = [\mu_1 + 2\,\mu_2\,[I_1 - 3] + 3\,\mu_3\,[I_1 - 3]^2][\lambda^2 - 1/\lambda]\,. \tag{45}$$

The relaxation function is a Prony series similar to Eq. (19), $g(t) = \gamma_\infty + \sum_{i=1}^{n_{\text{prn}}} \gamma_i \exp(-t/\tau_i)$, with five terms, $n_{\text{prn}} = 5$. To numerically solve for the total stress, we discretize the time interval of interest and adopt

Table 1
Unconfined compression relaxation data of passive skeletal muscle. Compression of gluteus muscle samples at fixed stretch rate of $\dot{\lambda}_{med} = 0.01/s$ and varying stretches of $\lambda_{low} = 0.9$, $\lambda_{med} = 0.8$, and $\lambda_{high} = 0.7$, and at fixed stretch $\lambda_{high} = 0.7$ and varying stretch rates of $\dot{\lambda}_{slow} = 0.005/s$, $\dot{\lambda}_{med} = 0.010/s$, and $\dot{\lambda}_{fast} = 0.050/s$. Cauchy stresses are reported as means from $n = 6$ samples (Van Loocke et al., 2008).

$\lambda_{low} = 0.9$, $\dot{\lambda}_{med} = 0.010/s$			$\lambda_{med} = 0.8$, $\dot{\lambda}_{med} = 0.010/s$			$\lambda_{high} = 0.7$, $\dot{\lambda}_{med} = 0.010/s$			$\lambda_{high} = 0.7$, $\dot{\lambda}_{slow} = 0.005/s$			$\lambda_{high} = 0.7$, $\dot{\lambda}_{fast} = 0.050/s$		
t [s]	λ [–]	σ [kPa]	t [s]	λ [–]	σ [kPa]	t [s]	λ [–]	σ [kPa]	t [s]	λ [–]	σ [kPa]	t [s]	λ [–]	σ [kPa]
0.0	1.00	0.000	0.0	1.00	0.000	0.0	1.00	0.000	0.0	1.00	0.000	0.0	1.00	0.000
2.0	0.98	−0.060	4.0	0.96	−0.119	6.0	0.94	−0.167	12.0	0.94	−0.077	1.2	0.94	−0.223
4.0	0.96	−0.101	8.0	0.92	−0.209	12.0	0.88	−0.343	24.0	0.88	−0.193	2.4	0.88	−0.427
6.0	0.94	−0.148	12.0	0.88	−0.342	18.0	0.82	−0.602	36.0	0.82	−0.355	3.6	0.82	−0.664
8.0	0.92	−0.205	16.0	0.84	−0.511	24.0	0.76	−1.058	48.0	0.76	−0.702	4.8	0.76	−1.518
10.0	0.90	−0.266	20.0	0.80	−0.711	30.0	0.70	−1.844	60.0	0.70	−1.424	6.0	0.70	−2.315
12.0	0.90	−0.240	22.0	0.80	−0.616	32.0	0.70	−1.593	62.0	0.70	−1.235	8.0	0.70	−1.694
15.0	0.90	−0.212	25.0	0.80	−0.569	35.0	0.70	−1.473	65.0	0.70	−1.120	10.0	0.70	−1.454
20.0	0.90	−0.192	30.0	0.80	−0.529	40.0	0.70	−1.376	70.0	0.70	−1.025	15.0	0.70	−1.267
30.0	0.90	−0.174	40.0	0.80	−0.487	50.0	0.70	−1.273	80.0	0.70	−0.917	20.0	0.70	−1.161
40.0	0.90	−0.160	50.0	0.80	−0.465	60.0	0.70	−1.212	90.0	0.70	−0.861	30.0	0.70	−1.058
60.0	0.90	−0.147	70.0	0.80	−0.433	80.0	0.70	−1.139	110.0	0.70	−0.782	40.0	0.70	−0.992
80.0	0.90	−0.138	90.0	0.80	−0.411	100.0	0.70	−1.090	130.0	0.70	−0.734	60.0	0.70	−0.904
100.0	0.90	−0.130	110.0	0.80	−0.395	120.0	0.70	−1.057	150.0	0.70	−0.695	80.0	0.70	−0.844
130.0	0.90	−0.125	140.0	0.80	−0.377	150.0	0.70	−1.014	180.0	0.70	−0.654	120.0	0.70	−0.774
170.0	0.90	−0.113	180.0	0.80	−0.357	190.0	0.70	−0.970	220.0	0.70	−0.610	180.0	0.70	−0.708
220.0	0.90	−0.106	230.0	0.80	−0.339	240.0	0.70	−0.932	270.0	0.70	−0.578	240.0	0.70	−0.658
300.0	0.90	−0.097	310.0	0.80	−0.322	320.0	0.70	−0.881	350.0	0.70	−0.531	300.0	0.70	−0.625

an explicit Euler forward scheme to march from time t_n to t_{n+1} in increments of $\Delta t = t_{n+1} - t_n$ using the update rules according to Eq. (22) for the total stress,

$$\sigma_{n+1} = \gamma_\infty \, \sigma_{0,n+1} + \sum_{i=1}^{5} \gamma_i \, h_{i,n+1}, \tag{46}$$

and Eq. (23) for the viscous overstress,

$$h_{i,n+1} = \exp(-\Delta t/\tau_i) h_{i,n} + \exp(-\Delta t/2\tau_i)[\sigma_{0,n+1} - \sigma_{0,n}]. \tag{47}$$

This benchmark model has a total of 14 parameters that we fit to the data, eleven for the relaxation function and three for the initial stored energy function.

General case: Vanilla recurrent neural network. Recurrent neural networks have a built-in memory structure and learn functions that, at any given time point, depend on all past inputs. Here we adopt a vanilla recurrent neural network as an overly-flexible, high-end comparison and expect that it will not have difficulties fitting the nonlinear elastic behavior, but might display overfitting and non-physical responses for small data. For this vanilla type recurrent neural network, the new stress state σ_{n+1} at time t_{n+1} not only depends on the stretch and the new time point $\{\lambda_{n+1}, t_{n+1}\}$, but also on a history vector h_n from the previous time point t_n (Goodfellow et al., 2016). At each new time point, the network updates the history vector,

$$h_{n+1} = g_h(w_{hh} \, h_n + w_{h\lambda} \, \{\lambda_{n+1}, t_{n+1}\} + b_h). \tag{48}$$

and calculates the new stress state using an additional feed-forward layer,

$$\sigma_{n+1} = g_\sigma(w_{\sigma h} \, h_{n+1} + b_\sigma). \tag{49}$$

Importantly, a recurrent neural network learns the same weights w_{hh}, $w_{h\lambda}, w_{\sigma h}$ and biases b_h, b_σ for all time steps, so Eqs. (48) and (49) closely mirror the update rules for the viscous overstress (23) and the elastic stress (22) of the theory of quasi-linear viscoelasticity in Section 2. Here we choose a hyperbolic tangent activation function $g_h(\circ)$ for the history vector (48), a linear activation function $g_\sigma(\circ)$ for the stress (49). Our network input $\{\lambda, t\}$ is a two-unit vector of stretch and time and our network output σ is a scalar. For the history vector h, we select an 8-unit vector, such that $w_{h\lambda}$ is an 8×2 matrix, w_{hh} is an 8×8 matrix, $w_{\sigma h}$ is a 1×8 vector, b_h is an 8×1 vector, and b_σ is a scalar. This results in $16 + 64 + 8 + 8 + 1 = 97$ total trainable parameters, $\theta = \{\boldsymbol{w}, \boldsymbol{b}\}$. We choose this architecture for the vanilla recurrent neural network because it is the smallest and simplest architecture that could fit our training data almost perfectly. The network learns its parameters by minimizing the loss function that penalizes the error between model and data. Similar to the constitutive recurrent neural network, we select a loss function by comparing the mean absolute error, mean squared error, and logarithmic hyperbolic cosine loss functions. For the vanilla recurrent neural network, we find the lowest errors using a loss function of logarithmic hyperbolic cosine type. The loss function is parameterized in terms of the stretch λ and time t,

$$L(\boldsymbol{\theta}; \lambda, t) = \sum_{i=1}^{n_{\text{train}}} \omega_i \sum_{j=1}^{n_{\text{time}}} \log(\cosh(\sigma(\lambda_{ij}, t_{ij}) - \hat{\sigma}_{ij})) \to \min \tag{50}$$

where $i = 1, \ldots, n_{\text{train}}$ denotes the number of training sets, $j = 1, \ldots, n_{\text{time}}$ denotes the number of data points in the time series, and $(\sigma(\lambda_{ij}, t_{ij}) - \hat{\sigma}_{ij})$ is the difference between the model stress $\sigma(\lambda_{ij}, t_{ij})$ at time t_{ij} and the experimental stress $\hat{\sigma}_{ij}$. We normalize the loss associated with each training set with the weight, $\omega_i = \bar{\sigma}_i^{-1} / \sum_{j=1}^{n_{\text{train}}} \bar{\sigma}_j^{-1}$ since each training set has a different mean stress values $\bar{\sigma}_i$. We train the model for 5000 epochs using the Adam optimizer, with a learning rate $\alpha = 0.1$ and parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$.

3.1. Data

We train and test all models on unconfined compression relaxation tests of passive skeletal muscle collected from porcine gluteus muscle tissue (Van Loocke et al., 2008). The data consist of recordings from five different experiments in which the muscle is compressed to three different stretch levels at three different stretch rates. After an initial ramp loading, the final stretch level is held constant for 300 s. For three experiments, the stretch rate is fixed at $\dot{\lambda}_{med} = 0.01/s$, and the minimum compressive stretch is varied as $\lambda_{low} = 0.9$, $\lambda_{med} = 0.8$, and $\lambda_{high} = 0.7$. For three experiments, the minimum compressive stretch is fixed at $\lambda_{high} = 0.7$, and the stretch rate is varied as $\dot{\lambda}_{slow} = 0.005/s$, $\dot{\lambda}_{med} = 0.010/s$, and $\dot{\lambda}_{fast} = 0.050/s$. As the $\lambda_{high} = 0.7$, $\dot{\lambda}_{med} = 0.010/s$ case is repeated, this results in a total of five data sets. To minimize the effects of tissue anisotropy, all experiment are performed with compression in the muscle fiber direction under the assumption that the compressed fibers do not carry any load. Table 1 summarizes the five data sets, reported as means over $n = 6$ tests per data set.

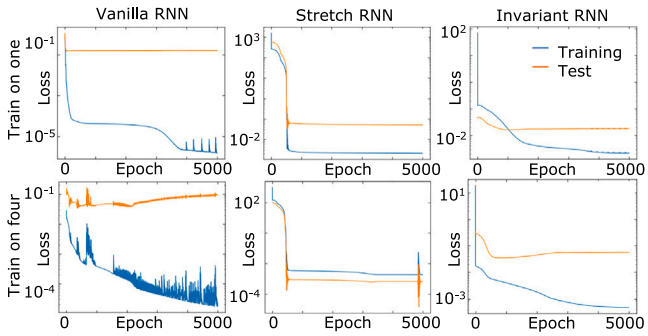

Fig. 2. Representative evolution of training and test losses during model training. Columns display the three networks, the vanilla, principal-stretch-based, and invariant-based recurrent neural networks; rows represent the two tasks, train on one and test on the remaining four, and train on four and test on the remaining one.

3.2. Statistical analysis

We use two error measures to quantify the model performance during testing and training. The first error measure is the normalized root mean squared error, $NRMSE$,

$$NRMSE = \frac{1}{|\bar{\sigma}|}\sqrt{\frac{1}{n_{\text{time}}}\sum_{j=1}^{n_{\text{time}}}(\sigma(\lambda_j,t_j)-\hat{\sigma}_j)^2}\,, \tag{51}$$

where $(\sigma(\lambda_j,t_j)-\hat{\sigma}_j)$ is the difference between the model stress $\sigma(\lambda_j,t_j)$ and the experimental stress $\hat{\sigma}_j$, and $j=1,\dots,n_{\text{time}}$ denotes the number of data points in the time series. We normalize the error by the mean stress, $|\bar{\sigma}| = \frac{1}{n_{\text{time}}}\sum_{j=1}^{n_{\text{time}}}|\hat{\sigma}_j|$ of each time series since the mean stress values of the five data curves in Section 3.1 vary by over an order of magnitude. The second error measure is the coefficient of determination, R^2,

$$R^2 = 1-\sum_{j=1}^{n_{\text{time}}}(\sigma(\lambda_j,t_j)-\hat{\sigma}_j)^2 \Big/ \sum_{j=1}^{n_{\text{time}}}(\bar{\sigma}-\hat{\sigma}_j)^2\,, \tag{52}$$

where $(\sigma(\lambda_j,t_j)-\hat{\sigma}_j)$ is the difference between the model stress $\sigma(\lambda_j,t_j)$ and the experimental stress $\hat{\sigma}_j$, and $\bar{\sigma} = \sum_{j=1}^{n_{\text{time}}}\hat{\sigma}_j/n_{\text{time}}$ is the mean stress of the time series. In the discussion, we use the normalized root mean squared error to compare the different models because it provides more insight into the magnitude of the prediction errors compared to the magnitude of the data themselves. However, we also include the coefficient of determination to aid comparison to other studies. We performed all statistical analyses using the SciPy 1.7.3 Python library (Virtanen et al., 2020).

4. Results and discussion

4.1. Recurrent neural network training

Fig. 2 shows representative plots of the training and test losses for all three recurrent neural networks. The training loss of all three recurrent neural networks, vanilla, principal-stretch-based, and invariant based, converged within 5000 epochs. After confirming that 5000 epochs were generally sufficient for the training loss to plateau, we performed all following network training trials using 5000 epochs. The test loss of the principal-stretch and invariant-based networks displayed a similar tendency and dropped visibly within the first 2000 epochs. Notably, the test loss of the vanilla network remained unchanged across all epochs for the train on one case, and even increased after about 2000 epochs for the train on four case. This generalization gap indicates the poor predictive performance of the vanilla network when trained on limited data.

4.2. Model performance

We evaluate all five models on two separate tasks, train on one data set and test on the remaining four; and train on four data sets and test on the remaining one. Figs. 3 through 12 illustrate the models' performance for both tasks. In these figures, each column represents one of five experiments with the corresponding column label signifying the training curve for the train on one task or the test curve for the train on four task. Each row represents a data curve, and the plots in that row display the model's performance in fitting that data curve. If the curve was included in the training set, the annotation for the coefficient of determination is labeled R^2_{train}, and if it was not included in the training set, the annotation is labeled R^2_{test}.

Train on one, test on four. In the first evaluation task, we trained the models on one of the five compression relaxation data sets from Table 1 and then tested the models' predictive abilities on the remaining four curves. Fig. 3 shows the results for the neo Hookean standard linear solid. As expected, since the model parameters are fit to the training curve, the errors are lowest for the training set (average NRMSE: 0.136) and higher for the test curves (average NRMSE: 0.526). The simplicity of the model makes it impossible to exactly match the shape of the experimental curves. For example, the neo Hookean model assumes the stress will change linearly during the ramp portion of the experiment, but this is not how the tissue behaves. Therefore, the model will not fit the experimental data well for any combination of parameter values. This extreme case illustrates the need to select a model whose shape can represent the data well. In contrast to the neo Hookean standard linear solid model, the more advanced van Loocke Lyons Simms model more closely matches the shape of the experimental data as seen in Fig. 4. Now, with the more sophisticated model, the training (average NRMSE: 0.034) and test (average NRMSE: 0.206) errors are both reduced.

As an alternative to first assuming a functional form, vanilla recurrent neural networks learn both the functional form and its parameters from the data themselves. Fig. 5 shows the performance of the vanilla recurrent neural network. With 97 trainable parameters, the vanilla recurrent neural network fits the training data sets almost perfectly (average NRMSE: 0.012), but is at risk for overfitting to the training data. Overfitting can be observed clearly in the middle column of Fig. 5 where the neural network predicts almost identical curves in the first three rows. The average errors on the test set (average NRMSE: 1.622) are much higher than those of the training set. Besides overfitting, the vanilla recurrent neural network sometimes produces predictions that violate physical principles. For example, the predicted stresses sometimes switch between increasing and decreasing during the ramp portion of the time history.

Figs. 6 and 7 show the results for the two constitutive recurrent neural network models. The performance of both the principal stretch version in Fig. 6 and the invariant version in Fig. 7 falls between that of the neo Hookean standard linear solid and the vanilla recurrent neural network when looking at the training sets (average NRMSE: 0.048 and 0.022). Both constitutive recurrent neural networks discover a functional form from the data and can approximate the data more closely compared to the restrictive neo Hookean standard linear solid model. However, the functional forms of both constitutive networks are more limited in scope than the vanilla network. Therefore, the constitutive networks cannot fit the training data as perfectly as the vanilla network. This limitation on the discovered functional forms provides a benefit, however, when looking at the test data sets. Compared to the vanilla network, both constitutive networks exhibit less overfitting as evidenced by the lower test errors (average NRMSE: 0.289 and 0.243). Importantly, unlike the vanilla network, the constitutive recurrent neural networks do not produce unphysical solutions with spurious oscillatory stress responses. Both training and test errors for the constitutive recurrent neural networks are comparable to those of the van Loocke Lyons Simms model, suggesting that the models

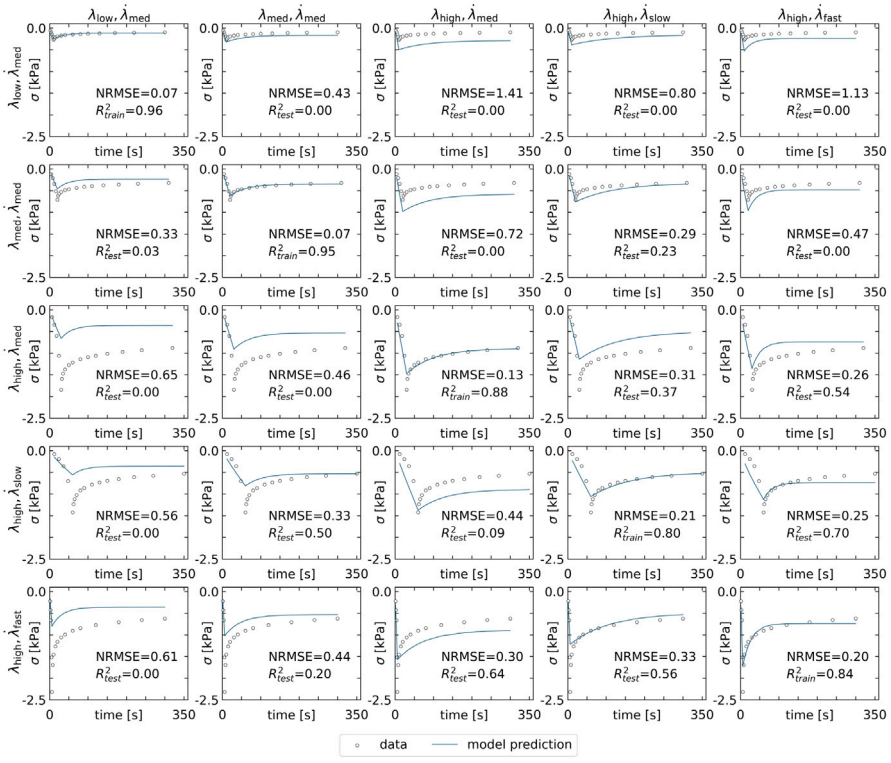

Fig. 3. Performance of neo Hookean standard linear solid trained on one curve. Experimental data and discovered model for the constitutive behavior of muscle tissue after discovering the model parameters for just one of the data sets. Columns represent all five training runs with each column label signifying the training set. Rows display the model performance for all five data sets with the training cases on the diagonal and the test cases on the off-diagonal.

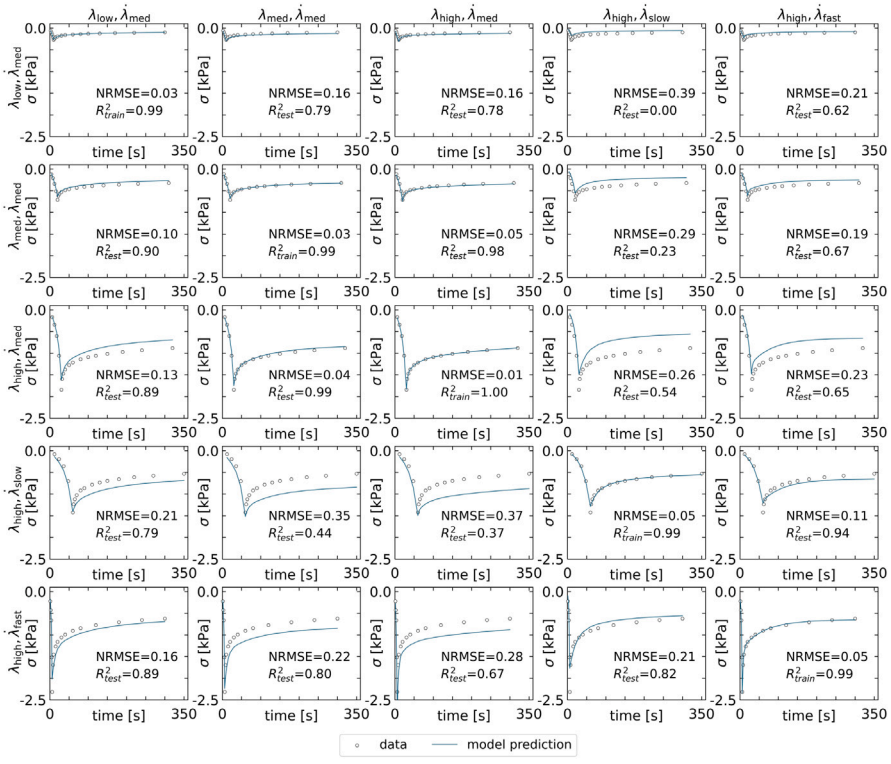

Fig. 4. Performance of Van Loocke Lyons Simms model trained on one curve. Experimental data and fitted model for the constitutive behavior of muscle tissue after fitting the model parameters to just one of the data sets. Columns represent all five training runs with each column label signifying the training set. Rows display the model performance for all five data sets with the training cases on the diagonal and the test cases on the off-diagonal.

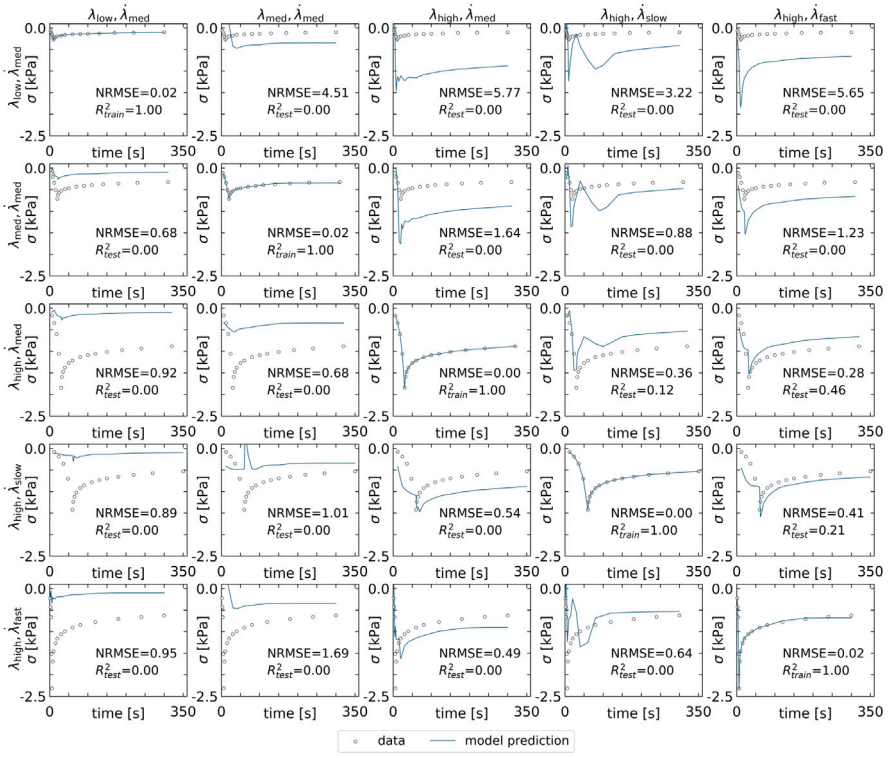

Fig. 5. Performance of vanilla recurrent neural network trained on one curve. Experimental data and discovered model for the constitutive behavior of muscle tissue after discovering the model parameters for just one of the data sets. Columns represent all five training runs with each column label signifying the training set. Rows display the model performance for all five data sets with the training cases on the diagonal and the test cases on the off-diagonal.

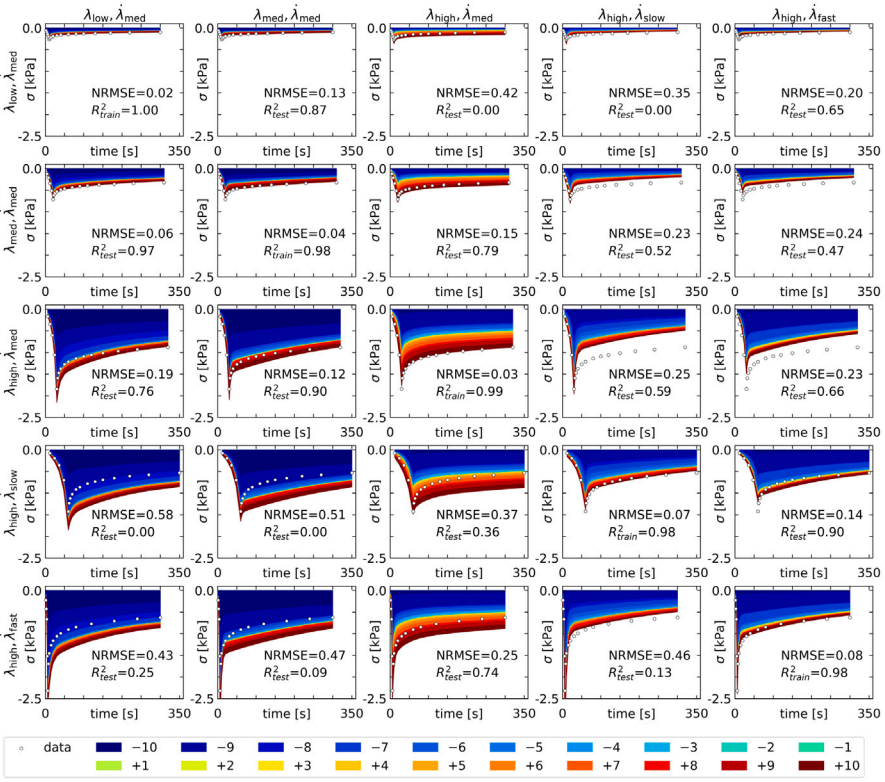

Fig. 6. Performance of principal-stretch-based recurrent neural network trained on one curve. Experimental data and discovered model for the constitutive behavior of muscle tissue after discovering the model parameters for just one of the data sets. Columns represent all five training runs with each column label signifying the training set. Rows display the model performance for all five data sets with the training cases on the diagonal and the test cases on the off-diagonal. The contributions of the principal stretch terms in the initial stored energy function are illustrated in colors.

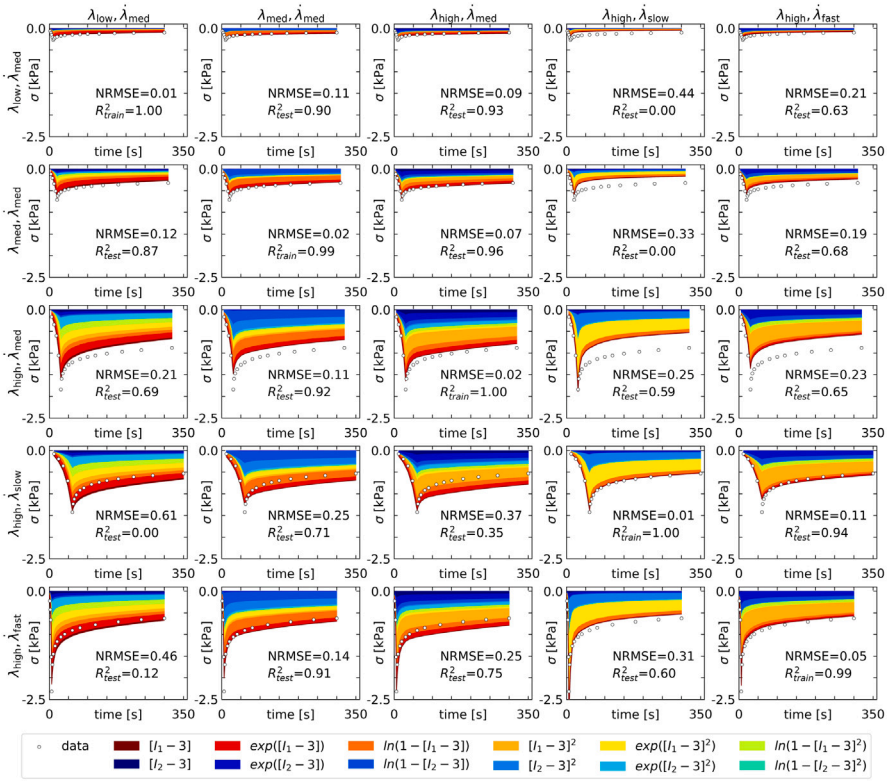

Fig. 7. Performance of invariant-based recurrent neural network trained on one curve. Experimental data and discovered model for the constitutive behavior of muscle tissue after discovering the model parameters for just one of the data sets. Columns represent all five training runs with each column label signifying the training set. Rows display the model performance for all five data sets with the training cases on the diagonal and the test cases on the off-diagonal. The contributions of the principal stretch terms in the initial stored energy function are illustrated in colors.

discovered by both constitutive recurrent neural networks represent the data as well as this hand-tailored benchmark model.

Train on four, test on one. In the second evaluation task, we trained the five models on four out of five data curves and tested the models' predictive abilities on the final remaining curve. Fig. 8 shows the results for the neo Hookean standard linear solid model. Compared to the train on one task, the errors on the training set are higher for the train on four task, (increasing average NRMSE: from 0.136 to 0.370). The model struggles to find a single set of parameters that can describe four different curves simultaneously. With the increase in training data, however, the error on the test set becomes smaller (decreasing average NRMSE: from 0.526 to 0.462). Similar to the train on one task, the neo Hookean standard linear solid continues to be restricted by its simple functional form. It simply cannot fit the data well because the tissue does not behave in accordance with the assumed functional form. As in the train on one case, the van Loocke Lyons Simms model outperforms the simple neo Hookean standard linear solid model because its more complex form has been tailored to fit the shape of the data more closely. With the more difficult task of simultaneously fitting four curves, errors on the training sets are higher (increasing average NRMSE: from 0.034 to 0.123), but errors on the test set decrease (decreasing average NRMSE: from 0.206 to 0.188).

Fig. 10 shows that the vanilla recurrent neural network almost perfectly fits the training data set, similar to the train on one task but with a slightly larger error (increasing average NRMSE: from 0.012 to 0.040). In contrast to the traditional mechanics-based models, the vanilla recurrent neural network has the ability to learn a complex enough functional form to simultaneously fit four different curves using the same set of parameters. With the additional training data, the average test error goes down (decreasing average NRMSE: from 1.622 to 0.838). However, the issue of overfitting remains apparent in comparing the train and test errors (average NRMSE: 0.040 vs. 0.838). Additionally, the issue of unphysical predicted solutions remains as seen clearly in the bottom left subplot of Fig. 10. The vanilla network prediction in this test case shows the stress decreasing and then increasing all during the hold portion of the experiment where the stress is expected to monotonically decrease. These observations suggest that although vanilla recurrent neural network architectures can successfully learn history-dependent constitutive laws (Zhu et al., 2011; Chen, 2021; Gorji et al., 2020; Tancogne-Dejean et al., 2021), the amount of required training data may not be practical to collect in a typical experimental setup. Importantly, in this case, increasing the amount of data does not imply increasing the number of data points on a curve but rather increasing the number of different curves from varying experiments.

Figs. 11 and 12 suggest that the performance of both constitutive recurrent neural network models on the training set lies, as in the train on one task, between the performance of the neo Hookean standard linear solid model and that of the vanilla recurrent neural network. Similar to these two, the training error for the principal-stretch and invariant-based networks increases for the train on four task compared to the train on one task (increasing average NRMSE: from 0.048 to 0.208 and from 0.022 to 0.108). Looking at the test cases, the constitutive network models exhibit less overfitting compared to the vanilla network (decreasing average NRMSE: from 0.289 to 0.196 and from 0.243 to 0.159). Comparison of the train and test errors for the constitutive network models reveals that the errors are similar between train and test sets. This provides evidence that the constitutive network models experience less overfitting compared to the vanilla network for which the train and test errors show a larger difference. As in the train on one case, the performance of both constitutive recurrent neural network models is comparable to the van Loocke Lyons Simms benchmark model with similar average NRMSE values.

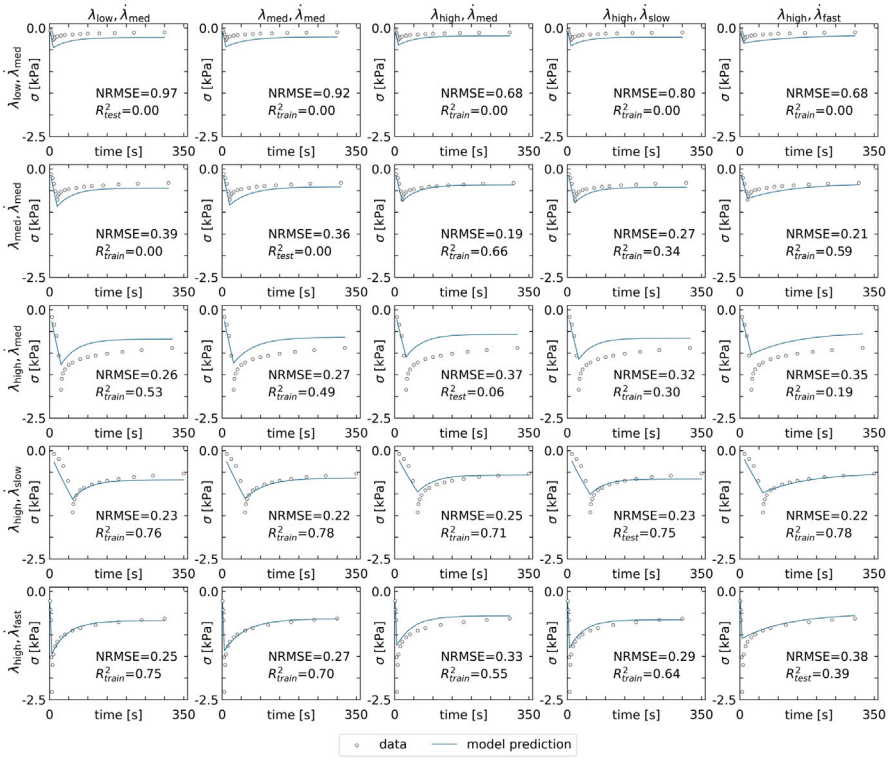

Fig. 8. Performance of neo Hookean standard linear solid trained on four curves. Experimental data and discovered model for the constitutive behavior of muscle tissue after discovering the model parameters for four of the data sets combined. Columns represent all five training runs with each column label signifying the test set. Rows display the model performance for all five data sets with the training cases on the off-diagonal and the test cases on the diagonal.

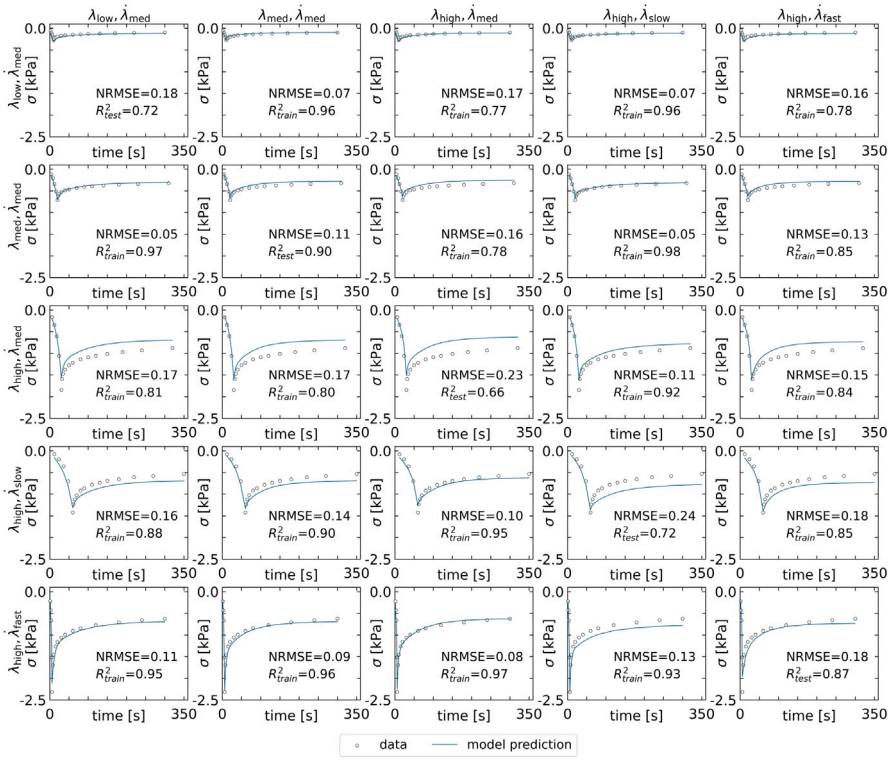

Fig. 9. Performance of Van Loocke Lyons Simms model trained on four curves. Experimental data and fitted model for the constitutive behavior of muscle tissue after fitting the model parameters to four of the data sets combined. Columns represent all five training runs with each column label signifying the test set. Rows display the model performance for all five data sets with the training cases on the off-diagonal and the test cases on the diagonal.

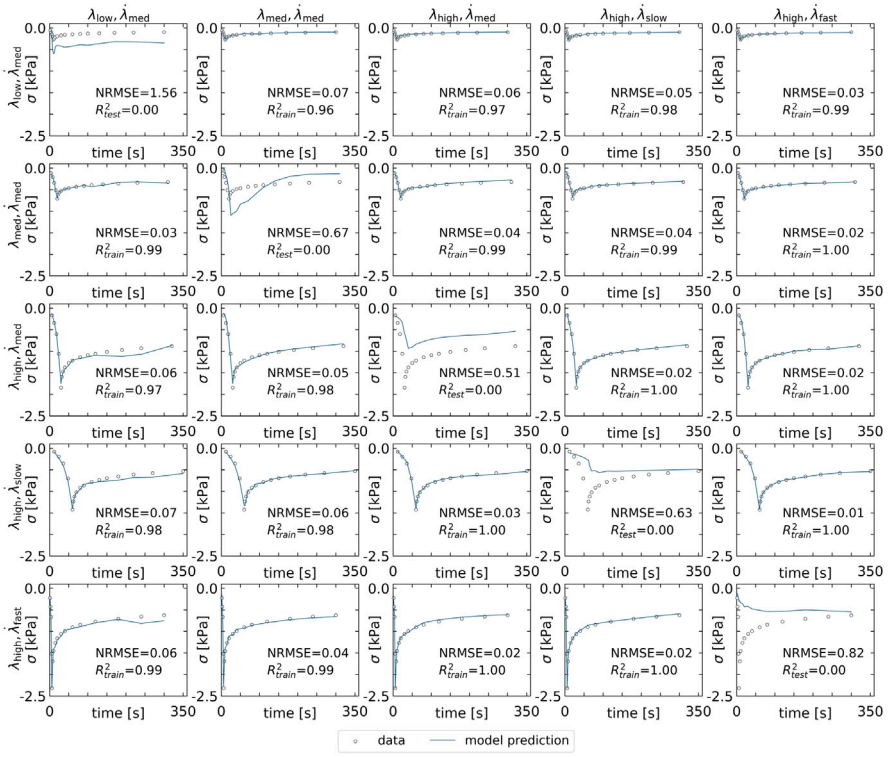

Fig. 10. Performance of vanilla recurrent neural network trained on four curves. Experimental data and discovered model for the constitutive behavior of muscle tissue after discovering the model parameters for four of the data sets combined. Columns represent all five training runs with each column label signifying the test set. Rows display the model performance for all five data sets with the training cases on the off-diagonal and the test cases on the diagonal.

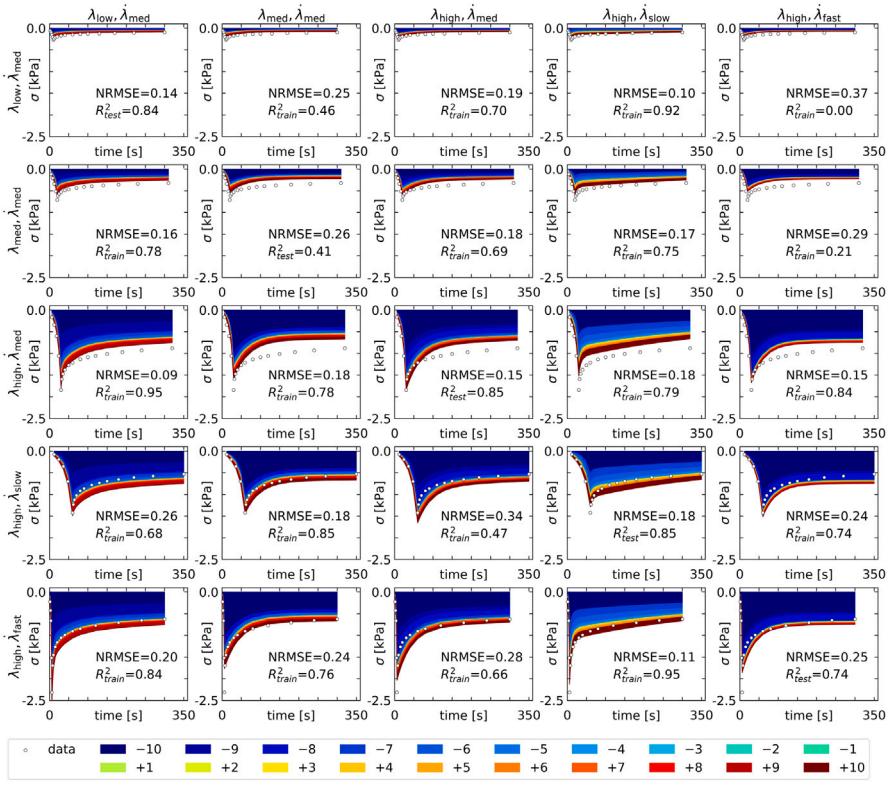

Fig. 11. Performance of principal-stretch-based recurrent neural network trained on four curves. Experimental data and discovered model for the constitutive behavior of muscle tissue after discovering the model parameters for four of the data sets combined. Columns represent all five training runs with each column label signifying the test set. Rows display the model performance for all five data sets with the training cases on the off-diagonal and the test cases on the diagonal. The contributions of the principal stretch terms in the initial stored energy function are illustrated in colors.

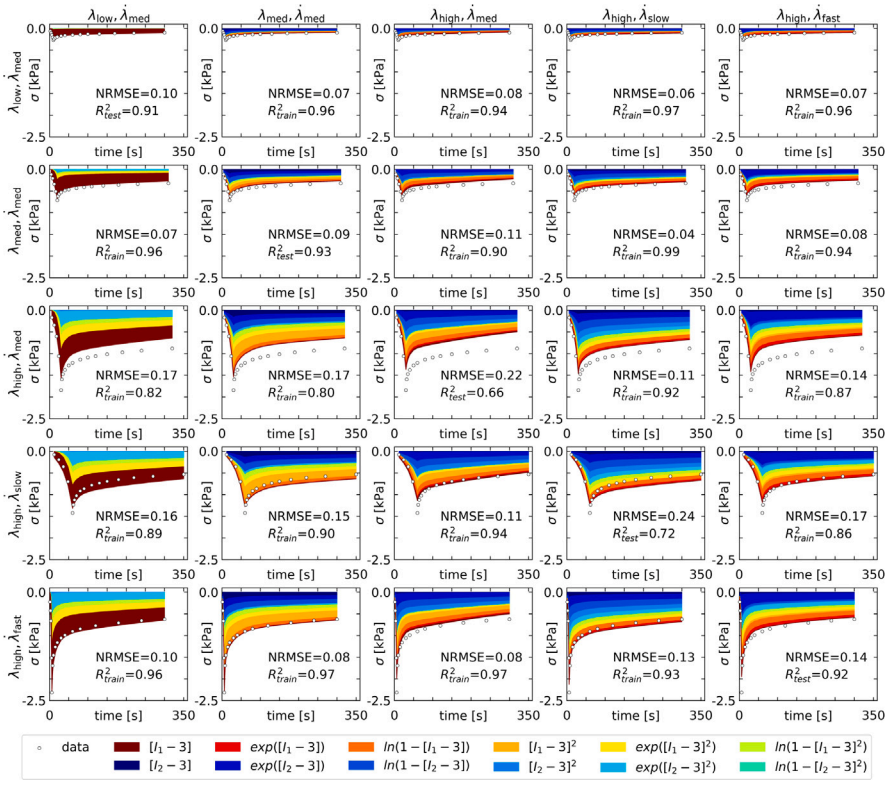

Fig. 12. Performance of invariant-based recurrent neural network trained on four curves. Experimental data and discovered model for the constitutive behavior of muscle tissue after discovering the model parameters for four of the data sets combined. Columns represent all five training runs with each column label signifying the test set. Rows display the model performance for all five data sets with the training cases on the off-diagonal and the test cases on the diagonal. The contributions of the principal stretch terms in the initial stored energy function are illustrated in colors.

4.3. Model discovery

Fig. 13 summarizes the performance of all five models on both evaluation tasks. The left column corresponds to the first evaluation task with training on just one curve, and the right column corresponds to the second task with training on four of the five curves. The first row displays the training errors and the second row the test errors. All displayed NRMSE values are the averages per curve in the training or test set.

The same general trend appears in all five models where the training errors are lower than the test errors. This is expected as the models are fit to the training data and have no prior knowledge of the test data. Moving from the train on one task to the train on four task, the training errors increase as the models face the more challenging task of representing four curves using just one set of parameters. With the additional training data, however, the test errors are generally lower for the train on four case compared to the train on one case.

Focusing in on the performance of individual models, the vanilla recurrent neural network clearly fits the training data the best with errors equal to or close to zero. However, it also exhibits the largest test errors, highlighting an issue with overfitting. Both the traditional mechanics-based models and the constitutive recurrent neural network models exhibit less overfitting, as their train and test errors are closer in magnitude. Both constitutive recurrent neural networks are comparable to the performance of the more advanced benchmark model with the invariant-based version seemingly fitting the data slightly better than the principal-stretch-based model.

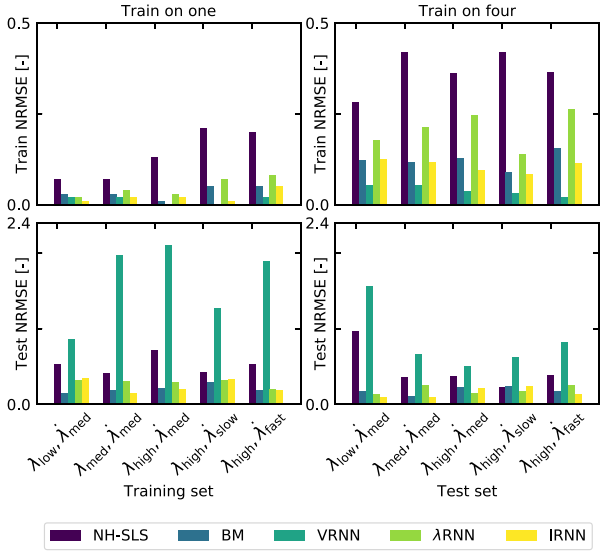

Fig. 13. Comparison of all five models for training and testing on one and four curves. Normalized root mean squared error (NRMSE) for the train on one and train on four tasks for all five models, neo Hookean standard linear solid (NH-SLS), van Loocke Lyons Simms benchmark model (BM), vanilla recurrent neural network (VRNN), principal-stretch-based (λRNN) and invariant-based (IRNN) recurrent neural networks. The plotted values are the average NRMSE per curve in a given test or training set. Some bars are invisible in the top row because the errors are equal or very close to zero.

4.4. Parameter discovery

The constitutive recurrent neural network models discover two sets of weights: one corresponding to the parameters of the initial stored energy function and one corresponding to the parameters of the Prony series relaxation function. The distribution of the discovered parameters for the initial stored energy functions is displayed using

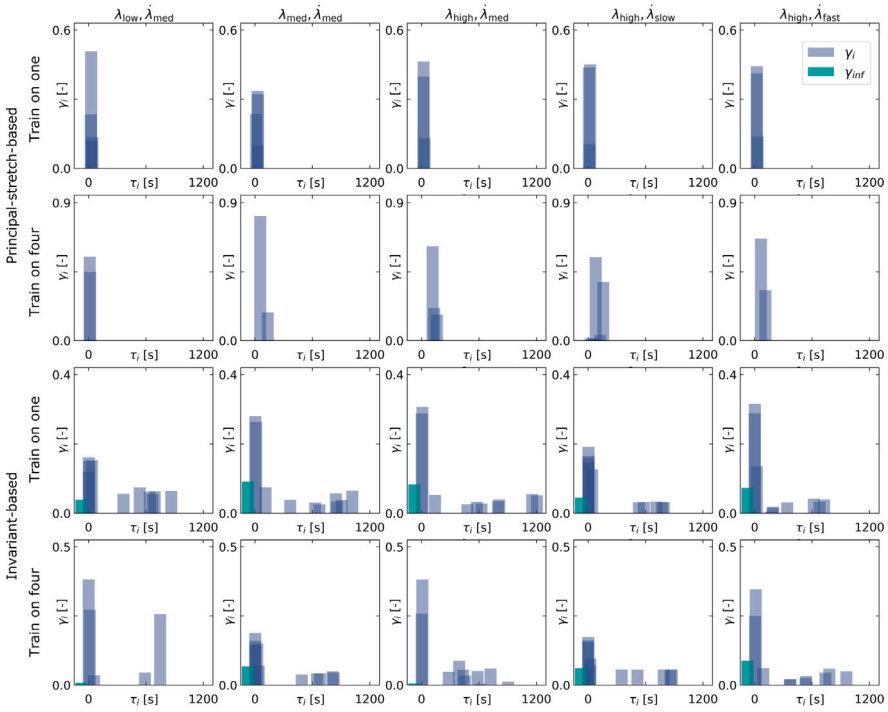

Fig. 14. Comparison of Prony parameters for principal-stretch and invariant-based recurrent neural networks. Each bar corresponds to a time constant, τ_i, and its height represents the magnitude of the corresponding weight, γ_i. The bars are not fully opaque to aid the visualization of multiple terms located at similar horizontal axis locations. The long-term modulus, γ_∞, is represented by a single green bar to the far left; however, it is invisible in some plots where its discovered value is close to zero.

color coding in Figs. 6 and 11 for the principal-stretch-based model and Figs. 7 and 12 for the invariant-based model. In these figures, we assign each term in the initial stored energy function a color. The thickness of each color band indicates the relative contribution of that term to the final initial stored energy function.

Looking at the principal-stretch-based model in Figs. 6 and 11, in each training scenario, the model learns different combinations of terms. However, the parameter distributions all share some general trends. For both the train on one and train on four tasks, the discovered models show a preference for negative exponent terms in the initial stored energy function. Similarly, all of the results show a preference for terms at either extreme, large negative and large positive exponents. In a study fitting a single-term Ogden model to muscle data, the resulting exponent was 14.00 for bovine and porcine tissue and 8.97 for human tissue (Mo et al., 2020). Similarly, a study fitting a two-term compressible Ogden model found exponents of 11.77 and 14.34 for tensile data measured on guinea pig ventricular papillary muscle (Hassan et al., 2012). While our discovered models do not exactly match these fitted functional forms, our discovered exponents agree well with these large exponents reported in the literature. This suggests that the trend of our network to discover terms at either extreme agrees well with reported observations for muscle tissue.

The invariant-based model in Figs. 7 and 12 behaves similarly, learning different combinations of terms for different training scenarios. However, the invariant-based model does not show any strong preference towards terms involving the first or second invariant. The model similarly does not show any strong preference towards the linear, quadratic, exponential, or logarithmic terms. Section 4.5 discusses the terms learned by the invariant-based model in more detail.

Fig. 14 illustrates the distributions of the Prony parameters discovered by both constitutive recurrent neural networks. The constitutive recurrent neural network with $n_{\text{prn}} = 10$ in Eq. (32) discovers ten Prony terms, each with a corresponding time constant, τ_i, and weight, γ_i. Each bar in Fig. 14 represents one of these terms with its horizontal location corresponding to τ_i and its height corresponding to γ_i. The presence of multiple bars in the same location is indicated by a darker color since the bars are not fully opaque. The long-term modulus, γ_∞, appears as a single green bar to the very left of each graph.

For the principal-stretch-based model, the discovered time constants, τ_i, in the train on one task range from 4.3 s to 41 s with the majority of terms concentrated around 21 s to 30 s. In the train on four task, the learned Prony parameters show a slightly wider spread of τ_i values ranging from 13 s to 161 s. In both the train on one and train on four cases, the principal-stretch-based model shows little contribution from the long term modulus, γ_∞.

For the invariant-based model, the network discovered a greater range of τ_i values compared to the principal-stretch-based model. The results for both the train on one and train on four tasks look similar with τ_i values ranging from 1.16 s to 1205 s and 1.77 s to 959 s, respectfully. The discovered parameters for both tasks display large contributions from time constants clustered around 1 s to 4 s, 8 s to 13 s, and 22 s to 40 s. These main clusters are accompanied by scattered contributions from larger time constants in the range of 100 s to 1200 s. In contrast to the principal-stretch-based model, the invariant-based model also predicts a more significant contribution of the long term modulus, γ_∞.

In a study fitting a five-term Prony series to the same muscle data (Van Loocke et al., 2008), the resulting τ_i values were 0.6 s, 6 s, 30 s, 60 s, and 300 s with the two smaller time constants weighted more heavily than the larger three, with corresponding weights of 0.465, 0.200, 0.057, 0.066, and 0.089. Another study fitting a two-term Prony series to guinea pig ventricular muscle data found time constants of 1.74 s and 52.16 s (Hassan et al., 2012). The general trends of these findings match the distributions discovered in this study with heavily-weighted contributions from time constants on the order of 1 s to 10 s accompanied by lesser-weighted contributions from time constants on the order of 100 s.

4.5. Regularization

Looking at the color distributions in Figs. 6, 7, 11, and 12, we recognize that both constitutive recurrent neural networks discover a broad spectrum of terms for the initial stored energy function, rather

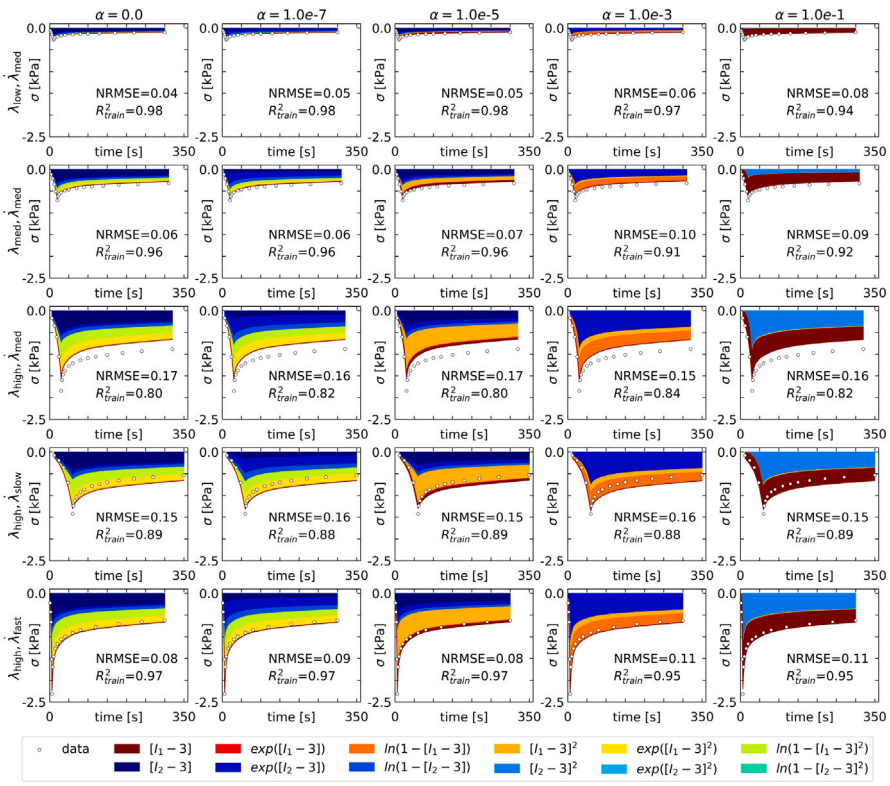

Fig. 15. Effect of L2 regularization on initial stored energy of invariant-based neural network. Displayed results from simultaneously fitting the invariant-based recurrent neural network to all five data sets for varying regularization parameters α. Increasing the regularization parameter from zero via 10^{-7} to 10^{-1} decreases the number of discovered terms of the initial stored energy function from six to two. At $\alpha = 10^{-1}$, the network discovers a two-term model that is linear in the first invariant I_1 and quadratic in the second invariant I_2.

than focusing on a select few. Towards the goal of preventing potential overfitting, we investigated the ability of L2 regularization to reduce the number of discovered terms for the invariant-based model, both for the elastic and viscous parts.

First, we varied the regularization parameter α in Eq. (38) from 10^{-7} to 10^{-1}. For each α value, we simultaneously trained the invariant-based model on all five data sets. Fig. 15 displays the results for the case with no regularization in the leftmost column and cases with a gradually increasing regularization parameter α from left to right. From the color spectrum in Fig. 15, we conclude that, as the regularization parameter α increases from zero to 10^{-1}, the number of terms in the initial stored energy function decreases from six to two. This controlled reduction of terms agrees with the results of previous studies on the effects of regularization in constitutive neural networks for time-independent hyperelasticity (St. Pierre et al., 2023; Linka and Kuhl, 2023). With no regularization, our invariant-based network discovers a six-term initial stored energy function of the following form,

$$\begin{aligned}\psi_0 \;=\; & w_{1,1}\; w_{0,1}[I_1-3] && +\; w_{1,5}\; \exp(w_{0,5}[I_1-3]^2-1)\\ & -\; w_{1,6}\; \ln(1-w_{0,6}[I_1-3]^2) && +\; w_{1,7}\; w_{0,7}[I_2-3]\\ & +\; w_{1,8}\; \exp(w_{0,8}[I_2-3]-1) && -\; w_{1,9}\; \ln(1-w_{0,9}[I_2-3])\end{aligned} \tag{53}$$

with the following weights, $w_{0,4} = 0.19$, $w_{1,4} = 0.30$ kPa, $w_{0,5} = 0.12$, $w_{1,5} = 3.69$ kPa, $w_{0,6} = 0.13$, $w_{1,6} = 4.41$ kPa, $w_{0,7} = 1.71$, $w_{1,7} = 0.23$ kPa, $w_{0,8} = 0.11$, $w_{1,8} = 0.68$ kPa, $w_{0,9} = 0.10$, and $w_{1,9} = 1.07$ kPa. Notably, all other weights in Eq. (27) train to zero.

With the maximal regularization parameter of $\alpha = 10^{-1}$, our invariant-based network discovers a two-term initial stored energy function of the following form,

$$\psi_0 = w_{1,1}\, w_{0,1}\, [I_1-3] + w_{1,10}\, w_{0,10}\, [I_2-3]^2, \tag{54}$$

where $w_{0,1} = 19.97$ kPa, $w_{1,1} = 0.03$, $w_{0,10} = 18.37$ kPa, and $w_{1,10} = 0.03$, while all remaining terms train to zero. Strikingly, this final reduced version of the discovered initial stored energy function in Eq. (54), with a linear term in the first invariant I_1 and a quadratic term in the second invariant I_2, belongs to the family of generalized Mooney–Rivlin models (Mooney, 1940; Rivlin, 1948), $\psi_0 = \sum_{i=0}^{n_i} \sum_{j=0}^{n_j} C_{ij}\, [I_1-3]^i [I_2-3]^j$, and its discovered weights translate into the Mooney–Rivlin coefficients $C_{10} = w_{1,1}\, w_{0,1} = 0.60$ kPa and $C_{02} = w_{1,10}\, w_{0,10} = 0.55$ kPa, with all other coefficients, C_{ij}, equal to zero. In a study using the Mooney–Rivlin model with considerations for intramuscular pressure (Wheatley et al., 2017), the resulting coefficients were $C_{10} = 0.05$ kPa and $C_{01} = 0.50$ kPa for the rabbit tibialis anterior muscle. These values were obtained for modeling hyperelasticity rather than for use in modeling the inital stress response of a viscoelastic model, so direct comparison to our values is not possible. However, these values from the literature suggest that our results are of a reasonable order of magnitude.

Second, we investigated the ability of L2 regularization to reduce the number of discovered terms in the Prony series relaxation function of the invariant-based recurrent neural network. We varied the regularization parameter β in Eq. (39) from 10^{-4} to 10^{-1}. For each β value, we simultaneously trained the invariant-based model on all five data curves. Fig. 16 displays the results for the case with no regularization in the leftmost column and cases with a gradually increasing regularization parameter β from left to right. The top row displays the distribution of learned Prony parameters in the same format as Fig. 14, and the remaining rows show the model predictions on the five curves. Looking at the top row of Fig. 16, as we increase the regularization parameter β from zero to 10^{-1}, the number of learned Prony terms decreases from seven to two. However, for the largest value of $\beta = 10^{-1}$, the errors between the model predictions and the experimental data increase significantly. This suggests that the optimal regularization parameter lies below this value, $\beta < 10^{-1}$. With no regularization, the invariant-based network discovers a seven-term Prony series for the relaxation

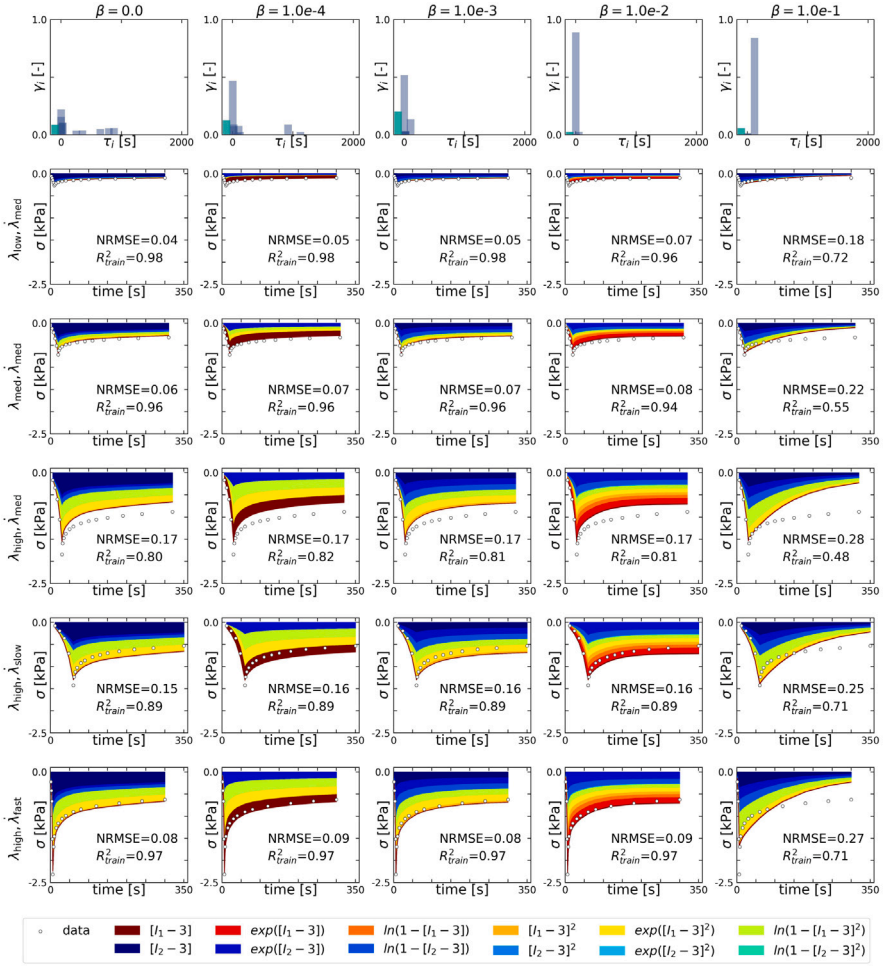

Fig. 16. Effect of L2 regularization on Prony parameters of recurrent neural network. Displayed results from simultaneously fitting the invariant-based recurrent neural network to all five data sets for varying regularization parameters β. Increasing the regularization parameter from zero via 10^{-4} to 10^{-1} decreases the number of discovered terms of the Prony series from six to two. At $\beta = 10^{-2}$, the network discovers a three-term Prony series with time constants of 0.362 s, 2.54 s, and 52.0 s.

function:

$$\begin{aligned} g(t) \;=\; & 0.08 \;+\; 0.48\,\exp(-t/2.59\text{ s}) \\ & +\; 0.21\,\exp(-t/25.9\text{ s}) \;+\; 0.04\,\exp(-t/254\text{ s}) \\ & +\; 0.04\,\exp(-t/356\text{ s}) \;+\; 0.05\,\exp(-t/651\text{ s}) \\ & +\; 0.05\,\exp(-t/793\text{ s}) \;+\; 0.05\,\exp(-t/877\text{ s})\,. \end{aligned} \tag{55}$$

As we increase the regularization parameter β, the model discovers fewer terms; yet, at the same time, the prediction error increases significantly. A value of $\beta = 10^{-2}$ for which the network discovers a three-term relaxation function,

$$\begin{aligned} g(t) \;=\; & 0.03 \;+\; 0.89\,\exp(-t/0.362\text{ s}) \\ & +\; 0.05\,\exp(-t/2.54\text{ s}) \;+\; 0.03\,\exp(-t/52.0\text{ s})\,, \end{aligned} \tag{56}$$

seems to be a reasonable trade off between model terms and model error. The magnitudes of the three time constants discovered by the regularized model closely match the time constants of 1.74 s and 52.16 s reported for a two-term Prony series fitted to guinea pig ventricular muscle data (Hassan et al., 2012). Our three discovered time constants are also comparable to the range of time constants, from 0.6 s–300 s, used to fit the same muscle data to a five-term Prony series (Van Loocke et al., 2008). Our discovered time constants lie in the lower end of the 0.6–300 s range, and we note that the L2 regularization trends towards smaller time constants for increasing β as seen in Fig. 16.

5. Limitations and future outlook

While our two constitutive recurrent neural networks in their principal-stretch-based and invariant-based versions show promise in the discovery of constitutive models for muscle tissue, our current model also has some limitations to address: First, as a proof of concept, we have focused on a very specific case of uniaxial unconfined compression and have developed a one-dimensional model in accordance with this loading configuration. Second, for simplicity, we only consider incompressible and isotropic materials and assume that both, the elastic and viscous response, are incompressible and isotropic. We plan to investigate other loading modes such as tension, and we will expand our general framework to accommodate additional loading scenarios by extending it to three dimensions (Calvo et al., 2014; Linka and Kuhl, 2023). We will also incorporate considerations for compressibility (Wheatley et al., 2017; Hassan et al., 2012) and anisotropy (Van Loocke et al., 2006; Böl et al., 2014; Kuthe and Uddanwadiker, 2016; Takaza et al., 2013; Kohn et al., 2021; Linka et al., 2023b; Tac et al., 2023a). Measurements of muscle fiber volume under tension and compression have shown evidence of muscle volume change under compression (Böl et al., 2020). Future incorporation of compressibility into our model will facilitate an examination of the assumption of incompressibility applied here. A three-dimensional formulation would allow the investigation of different loading modes such tension, compression, and shear (Böl et al., 2020) or biaxial stretch (Wheatley, 2020) and the investigation of tension–compression asymmetry (Latorre et al., 2018; St. Pierre et al., 2023). For hyperelastic materials, we have shown that including different loading modes in the training process is critical to reproducibly discover a single unique model (Linka and Kuhl, 2023), and we are confident that this stabilizing effect will translate to viscoelastic model discovery (Marino et al.,

2023). Third, our model is limited by the assumptions of the theory of quasi-linear viscoelasticity (Fung et al., 1970). Following this theory, our model assumes that the material response can be decomposed into a time-independent initial stress function that depends only on the current stretch and a time-dependent relaxation function that depends only on time. The theory of quasi-linear viscoelasticity is widely used to model single cells (Wang and Kuhl, 2020) and tissues, such as tendons (Provenzano et al., 2001; Woo, 1982), skeletal muscle (Van Loocke et al., 2008), or the heart (Tikenogullari et al., 2022). Materials for which this assumption does not hold will require a different, fully-nonlinear modeling approach (Latorre and Montáns, 2017; Wheatley et al., 2015, 2016; Holthusen et al., 2023; Tac et al., 2023b; Abdolazizi et al., 2023). However, as a reasonable first approximation, the theory of quasi-linear viscoelasticity characterizes the behavior of muscle tissue well (Wheatley et al., 2017; Rehorn et al., 2014; Then et al., 2012; Hassan et al., 2012; Mo et al., 2020; Van Loocke et al., 2008) and produces comparable results in this study. Finally, here we focus only on the effects of L2 regularization on down-selection of terms in our discovered models. Other regularization methods will be the focus of future studies.

6. Conclusions

Constitutive artificial neural networks are pioneering a new approach to constitutive modeling where both model and parameters are fit to the data themselves. In this study, we illustrate the potential of neural networks as powerful function approximators in biomechanics. However, their application to material modeling requires special attention: The amount of data available from mechanical testing is typically much smaller than the data used in traditional deep learning. When trained on these limited datasets, a vanilla recurrent neural network with no mechanics knowledge struggles to learn a constitutive law that generalizes well to previously unseen load cases. Worse yet, vanilla recurrent neural networks make predictions that may violate physical laws. This emphasizes the critical need to integrate mechanics-based knowledge into the network design. Inspired by the theory of quasi-linear viscoelasticity, we design a new class of constitutive recurrent neural networks that incorporate our domain knowledge and extend recent feed-forward networks to model the history-dependent behavior of viscoelastic materials. Even for limited training data, our constitutive recurrent neural networks robustly discover constitutive models that obey physical principles, generalize well to unseen data, and match the performance of hand-tailored models. The modular structure of our network architecture takes advantage of past and ongoing research on hyperelastic feed-forward networks by extending these architectures with an additional recurrent layer. This modular design encourages the plug-and-play of different families of hyperelastic models – principal-stretch or invariant-based, isotropic or anisotropic, incompressible or compressible – and provides a clear road map for automated model discovery in computational inelasticity.

CRediT authorship contribution statement

Lucy M. Wang: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Conceptualization. **Kevin Linka:** Writing – review & editing, Methodology. **Ellen Kuhl:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Lucy M. Wang reports financial support was provided by the National Science Foundation Graduate Research Fellowship DGE 1656518, the Bio-X Bowes Fellowship, and the Stanford School of Engineering Fellowship. Ellen Kuhl reports financial support was provided by the Wu Tsai Human Performance Alliance and the National Science Foundation Proposal CMMI 2320933 Automated Model Discovery for Soft Matter.

Data availability

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

Funding

This work was supported by the National Science Foundation Graduate Research Fellowship DGE 1656518, the Bio-X Bowes Fellowship, and the Stanford School of Engineering Fellowship to Lucy M. Wang and by the Wu Tsai Human Performance Alliance and the National Science Foundation Proposal CMMI 2320933 Automated Model Discovery for Soft Matter to Ellen Kuhl. The sponsors had no role in study design; collection, analysis and interpretation of data; in writing the report; or in the decision to submit the article for publication.

References

Abdolazizi, K.P., Linka, K., Cyron, C.J., 2023. Viscoelastic constitutive artificial neural Networks (vCANNs)–a framework for data-driven anisotropic nonlinear finite viscoelasticity. arXiv.

Alber, M., Buganza Tepole, A., Cannon, W., De, S., Dura-Bernal, S., Garikipati, K., Karniadakis, G.E., Lytton, W.W., Perdikaris, P., Petzold, L., Kuhl, E., 2019. Integrating machine learning and multiscale modeling: Perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. Npj Digit. Med. 2, 115.

As'ad, F., Avery, P., Farhat, C., 2022. A mechanics-informed artificial neural network approach in data-driven constitutive modeling. Internat. J. Numer. Methods Engrg. 123, 2738–2759.

Böl, M., Ehret, A., Leichsenring, K., Weichert, C., Kruse, R., 2014. On the anisotropy of skeletal muscle tissue under compression. Acta Biomater. 10, 3225–3234.

Böl, M., Iyer, R., Garces-Schröder, M., Dietzel, A., 2020. Mechano-geometrical skeletal muscle fibre characterisation under cyclic and relaxation loading. J. Mech. Behav. Biomed. Mater. 110, 104001.

Bonatti, C., Mohr, D., 2021. One for all: Universal material model based on minimal state-space neural networks. Sci. Adv. 7, 3658–3681.

Borkowski, L., Sorini, C., Chattopadhyay, A., 2022. Recurrent neural network-based multiaxial plasticity model with regularization for physics-informed constraints. Comput. Struct. 258, 106678.

Budday, S., Sommer, G., Birkl, C., Langkammer, C., Haybaeck, J., Kohnert, J., Bauer, M., Paulsen, F., Steinmann, P., Kuhl, E., Holzapfel, G.A., 2017. Mechanical characterization of human brain tissue. Acta Biomater. 48, 319–340.

Calvo, B., Sierra, M., Grasa, J., Muñoz, M.J., Peña, E., 2014. Determination of passive viscoelastic response of the abdominal muscle and related constitutive modeling: Stress-relaxation behavior. J. Mech. Behav. Biomed. Mater. 36, 47–58.

Chen, G., 2021. Recurrent neural networks (rnns) learn the constitutive law of viscoelasticity. Comput. Mech. 67, 1009–1019.

Danoun, A., Prulière, E., Chemisky, Y., 2022. Thermodynamically consistent recurrent neural networks to predict non linear behaviors of dissipative materials subjected to non-proportional loading paths. Mech. Mater. 173, 104436.

Dehoff, P.H., 1978. On the nonlinear viscoelastic behavior of soft biological tissues. J. Biomech. 11, 35–40.

Fung, Y., Perrone, N., Anliker, M., 1970. Biomechanics: Its Foundation and Objectives. Prentice-Hall.

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

Goldberg, Y., 2016. A primer on neural network models for natural language processing. J. Artificial Intelligence Res. 57, 345–420.

Goodfellow, I.J., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press, Cambridge, MA, USA, http://www.deeplearningbook.org.

Gorji, M.B., Mozaffar, M., Heidenreich, J.N., Cao, J., Mohr, D., 2020. On the potential of recurrent neural networks for modeling path dependent plasticity. J. Mech. Phys. Solids 143, 103972.

Hassan, M., Hamdi, M., Noma, A., 2012. The nonlinear elastic and viscoelastic passive properties of left ventricular papillary muscle of a guinea pig heart. J. Mech. Behav. Biomed. Mater. 5, 99–109.

He, X., Chen, J.S., 2022. Thermodynamically consistent machine-learned internal state variable approach for data-driven modeling of path-dependent materials. Comput. Methods Appl. Mech. Engrg. 402, 115348.

Holthusen, H., Rothkranz, C., Lamm, L., Brepols, T., Reese, S., 2023. Inelastic material formulations based on a co-rotated intermediate configuration–Applications to bioengineered tissues. J. Mech. Phys. Solids 172, 105174.

Holzapfel, G.A., 2002. Nonlinear solid mechanics: A continuum approach for engineering science. Meccanica 37, 489–490.

Holzapfel, G.A., Linka, K., Sherifova, S., Cyron, C.J., 2021. Predictive constitutive modelling of arteries by deep learning. J. R. Soc. Interface 18.
Holzapfel, G., Ogden, R. (Eds.), 2006. Mechanics of Biological Tissue, first ed. Springer Verlag, Germany.
Kalina, K.A., Linden, L., Brummund, J., Metsch, P., Kästner, M., 2022. Automated constitutive modeling of isotropic hyperelasticity based on artificial neural networks. Comput. Mech. 69, 213–232.
Kohn, S., Leichsenring, K., Kuravi, R., Ehret, A.E., Böl, M., 2021. Direct measurement of the direction-dependent mechanical behavior of skeletal muscle extracellular matrix. Acta Biomater. 122, 249–262.
Kuthe, C.D., Uddanwadiker, R.V., 2016. Investigation of effect of fiber orientation on mechanical behavior of skeletal muscle. J. Appl. Biomater. Funct. Mater. 14, e154–e162.
Latorre, M., Mohammadkhah, M., Simms, C.K., Montans, F.J., 2018. A continuum model for tension-compression asymmetry in skeletal muscle. J. Mech. Behav. Biomed. Mater. 77, 455–460.
Latorre, M., Montáns, F.J., 2017. Strain-level dependent nonequilibrium anisotropic viscoelasticity: Application to the abdominal muscle. J. Biomech. Eng. 139.
Linka, K., Hillgärtner, M., Abdolazizi, K.P., Aydin, R.C., Itskov, M., Cyron, C.J., 2021a. Constitutive artificial neural networks: A fast and general approach to predictive data-driven constitutive modeling by deep learning. J. Comput. Phys. 429, 110010.
Linka, K., Kuhl, E., 2023. A new family of Constitutive Artificial Neural Networks towards automated model discovery. Comput. Methods Appl. Mech. Engrg. 403, 115731.
Linka, K., Reiter, N., Würges, J., Schicht, M., Bräuer, L., Cyron, C.J., Paulsen, F., Budday, S., 2021b. Unraveling the local relation between tissue composition and human brain mechanics through machine learning. Front. Bioeng. Biotechnol. 9, 712.
Linka, K., Schafer, A., Meng, X., Zou, Z., Karniadakis, G.E., Kuhl, E., 2022. Bayesian Physics-Informed Neural Networks for real-world nonlinear dynamical systems. Comput. Methods Appl. Mech. Engrg. 402, 115346.
Linka, K., St. Pierre, .S., Kuhl, E., 2023a. Automated model discovery for human brain using constitutive artificial neural networks. Acta Biomater. 160, 134–151.
Linka, K., Tepole, A.B., Holzapfel, G.A., Kuhl, E., 2023b. Automated model discovery for skin: Discovering the best model, data, and experiment. Comput. Methods Appl. Mech. Engrg. 410, 116007.
Liu, M., Liang, L., Sun, W., 2020a. A generic physics-informed neural network-based constitutive model for soft biological tissues. Comput. Methods Appl. Mech. Engrg. 372, 113402.
Liu, Z., Zhang, Z., Ritchie, R.O., Liu, Z.Q., Zhang, Z.F., Ritchie, R.O., 2020b. Structural orientation and anisotropy in biological materials: Functional designs and mechanics. Adv. Funct. Mater. 30, 1908121.
Maganaris, C.N., Paul, J.P., 1999. In vivo human tendon mechanical properties. J. Physiol. 521, 307–313.
Marino, E., Flaschl, M., Kumar, S., Laurenzis, L.D., 2023. Automated identification of linear viscoelastic constitutive laws with EUCLID. Mech. Mater. 181, 104643.
Masi, F., Stefanou, I., Vannucci, P., Maffi-Berthier, V., 2021. Thermodynamics-based artificial neural networks for constitutive modeling. J. Mech. Phys. Solids 147, 104277.
Medsker, L., Jain, L.C. (Eds.), 1999. Recurrent Neural Networks: Design and Applications. CRC Press.
Mo, F., Zheng, Z., Zhang, H., Li, G., Yang, Z., Sun, D., 2020. In vitro compressive properties of skeletal muscles and inverse finite element analysis: Comparison of human versus animals. J. Biomech. 109, 109916.
Mooney, M., 1940. A theory of large elastic deformations. J. Appl. Phys. 11, 582–590.
Nicolle, S., Vezin, P., Palierne, J.F., 2010. A strain-hardening bi-power law for the nonlinear behaviour of biological soft tissues. J. Biomech. 43, 927–932.
Oeser, M., Freitag, S., 2016. Fractional derivatives and recurrent neural networks in rheological modelling – part i: theory. Int. J. Pavem. Eng. 17, 87–102.
Ogden, R.W., 1972. Large deformation isotropic elasticity–on the correlation of theory and experiment for incompressible rubberlike solids. Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 326, 565–584.
Pascalis, R.D., Abrahams, D., Parnell, W.J., 2014. On nonlinear viscoelastic deformations: a reappraisal of Fung's quasi-lineear viscoelastic model. Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 470, 20140058.
Provenzano, P., Lakes, R., Keenan, T., Vanderby, R., 2001. Nonlinear ligament viscoelasticity. Ann. Biomed. Eng. 29, 908–914.
Raissi, M., Perdikaris, P., Karniadakis, G.E., 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J. Comput. Phys. 378, 686–707.
Rehorn, M.R., Schroer, A.K., Blemker, S.S., 2014. The passive properties of muscle fibers are velocity dependent. J. Biomech. 47, 687–693.
Rivlin, R.S., 1948. Large elastic deformations of isotropic materials. Philos. Trans. R. Soc. Lond. Ser. A 241, 79–397.
Shen, Y., Chandrashekhara, K., Breig, W.F., Oliver, L.R., 2004. Neural network based constitutive model for rubber material. Rubber Chem. Technol. 77, 257–277.
Simo, J., Hughes, T., 2000. Computational Inelasticity. In: Interdisciplinary Applied Mathematics, Springer New York.
St. Pierre, .S., Linka, K., Kuhl, E., 2023. Principal-stretch-based constitutive neural networks autonomously discover a subclass of Ogden models for human brain tissue. Brain Multiphys. 4, 100066.
St. Pierre, S.R., Rajasekharan, D., Darwin, E.C., Linka, K., Levenston, M.E., Kuhl, E., 2023. Discovering the mechanics of artificial and real meat. Comput. Methods Appl. Mech. Engrg. http://dx.doi.org/10.1016/j.cma.2023.116236.
Tac, V., Linka, K., Sahli Costabal, F., Kuhl, E., Buganza Tepole, A., 2023a. Benchmarking physics-informed frameworks for data-driven hyperelasticity. Comput. Mech..
Tac, V., Rausch, M., Costabal, F.Sahli., Tepole, A.Buganza., 2023b. Data-driven anisotropic finite viscoelasticity using neural ordinary differential equations. arXiv.
Takaza, M., Moerman, K.M., Gindre, J., Lyons, G., Simms, C.K., 2013. The anisotropic mechanical behaviour of passive skeletal muscle tissue subjected to large tensile strain. J. Mech. Behav. Biomed. Mater. 17, 209–220.
Tancogne-Dejean, T., Gorji, M.B., Zhu, J., Mohr, D., 2021. Recurrent neural network modeling of the large deformation of lithium-ion battery cells. Int. J. Plast. 146, 103072.
Taylor, R.L., Pister, K.S., Goudreau, G.L., 1970. Thermomechanical analysis of viscoelastic solids. Internat. J. Numer. Methods Engrg. 2, 45–59.
Then, C., Vogl, T., Silber, G., 2012. Method for characterizing viscoelasticity of human gluteal tissue. J. Biomech. 45, 1252–1258.
Tikenogullari, O.Z., Costabal, F.S., Yao, J., Marsden, A., Kuhl, E., 2022. How viscous is the beating heart? Insights from a computational study. Comput. Mech. 70, 565–579.
Übeyli, E.D., 2009. Analysis of eeg signals by implementing eigenvector methods/recurrent neural networks. Digit. Signal Process. 19, 134–143.
Van Loocke, M., Lyons, C.G., Simms, C.K., 2006. A validated model of passive muscle in compression. J. Biomech. 39, 2999–3009.
Van Loocke, M. ., Lyons, C.G., Simms, C.K., 2008. Viscoelastic properties of passive skeletal muscle in compression: Stress-relaxation behaviour and constitutive modelling. J. Biomech. 41, 1555–1566.
Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İlhan, Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., Mulbregt, P.van., Vijaykumar, A., Bardelli, A.P., Rothberg, A., Hilboll, A., Kloeckner, A., Scopatz, A., Lee, A., Rokem, A., Woods, C.N., Fulton, C., Masson, C., Häggström, C., Fitzgerald, C., Nicholson, D.A., Hagen, D.R., Pasechnik, D.V., Olivetti, E., Martin, E., Wieser, E., Silva, F., Lenders, F., Wilhelm, F., Young, G., Price, G.A., Ingold, G.L., Allen, G.E., Lee, G.R., Audren, H., Probst, I., Dietrich, J.P., Silterra, J., Webber, J.T., Slavič, J., Nothman, J., Buchner, J., Kulick, J., Schönberger, J.L., de Miranda Cardoso, J.V., Reimer, J., Harrington, J., Rodríguez, J.L.C., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M., Kümmerer, M., Bolingbroke, M., Tartre, M., Pak, M., Smith, N.J., Nowaczyk, N., Shebanov, N., Pavlyk, O., Brodtkorb, P.A., Lee, P., McGibbon, R.T., Feldbauer, R., Lewis, S., Tygier, S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T., Oshima, T., Pingel, T.J., Robitaille, T.P., Spura, T., Jones, T.R., Cera, T., Leslie, T., Zito, T., Krauss, T., Upadhyay, U., Halchenko, Y.O., Vázquez-Baeza, Y., 2020. Scipy 1.0. fundamental algorithms for scientific computing in Python. Nat. Methods 17, 261–272, 2020 17:3.
Wang, L., Kuhl, E., 2020. Viscoelasticity of the axon limits stretch-mediated growth. Comput. Mech. 65, 587–595.
Wheatley, B.B., 2020. Investigating passive muscle mechanics with biaxial stretch. Front. Physiol. 11, 1021.
Wheatley, B.B., Morrow, D.A., Odegard, G.M., Kaufman, K.R., Donahue, T.L.H., 2016. Skeletal muscle tensile strain dependence: Hyperviscoelastic nonlinearity. J. Mech. Behav. Biomed. Mater. 53, 445–454.
Wheatley, B.B., Odegard, G.M., Kaufman, K.R., Donahue, T.L.H., 2017. A validated model of passive skeletal muscle to predict force and intramuscular pressure. Biomech. Model. Mechanobiol. 16, 1011–1022.
Wheatley, B.B., Pietsch, R.B., Donahue, T.L.H., Williams, L.N., 2015. Fully non-linear hyper-viscoelastic modeling of skeletal muscle in compression. Comput. Methods Biomech. Biomed. Eng. 19, 1181–1189.
Woo, S.L., 1982. Mechanical properties of tendons and ligaments, i. quasi-static and nonlinear viscoelastic properties. Biorheology 19, 385–396.
Yang, E., Tang, Y., Li, L., Yan, W., Huang, B., Qiu, Y., 2021. Research on the recurrent neural network-based fatigue damage model of asphalt binder and the finite element analysis development. Constr. Build. Mater. 267, 121761.
Zhang, M.J., Chu, Z.Z., 2012. Adaptive sliding mode control based on local recurrent neural networks for underwater robot. Ocean Eng. 45, 56–62.
Zhang, R., Liu, Y., Sun, H., 2020. Physics-informed multi-lstm networks for metamodeling of nonlinear structures. Comput. Methods Appl. Mech. Engrg. 369, 113226.
Zhu, J.-H., Zaman, M., Anderson, S., 2011. Modeling of soil behavior with a recurrent neural network. Can. Geotech. J. 35, 858–872.