

Full length article

Automated data-driven discovery of material models based on symbolic regression: A case study on the human brain cortex

Jixin Hou^a, Xianyan Chen^b, Taotao Wu^c, Ellen Kuhl^d, Xianqiao Wang^{a,*}^a School of Environmental, Civil, Agricultural and Mechanical Engineering, College of Engineering, University of Georgia, Athens, GA 30602, USA^b Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia, Athens, GA 30602, USA^c School of Chemical, Materials, and Biomedical Engineering, College of Engineering, University of Georgia, Athens, GA 30602, USA^d Department of Mechanical Engineering, Stanford University, Stanford, CA 94305, USA

ARTICLE INFO

Article history:

Received 26 March 2024

Revised 21 August 2024

Accepted 5 September 2024

Available online 17 September 2024

Keywords:

Model discovery

Symbolic regression

Constitutive modeling

Hyperelasticity

Polyconvexity

ABSTRACT

We introduce a data-driven framework to automatically identify interpretable and physically meaningful hyperelastic constitutive models from sparse data. Leveraging symbolic regression, our approach generates elegant hyperelastic models that achieve accurate data fitting with parsimonious mathematic formulas, while strictly adhering to hyperelasticity constraints such as polyconvexity/ellipticity. Our investigation spans three distinct hyperelastic models—invariant-based, principal stretch-based, and normal strain-based—and highlights the versatility of symbolic regression. We validate our new approach using synthetic data from five classic hyperelastic models and experimental data from the human brain cortex to demonstrate algorithmic efficacy. Our results suggest that our symbolic regression algorithms robustly discover accurate models with succinct mathematic expressions in invariant-based, stretch-based, and strain-based scenarios. Strikingly, the strain-based model exhibits superior accuracy, while both stretch-based and strain-based models effectively capture the nonlinearity and tension-compression asymmetry inherent to the human brain tissue. Polyconvexity/ellipticity assessment affirm the rigorous adherence to convexity requirements both within and beyond the training regime. However, the stretch-based models raise concerns regarding potential convexity loss under large deformations. The evaluation of predictive capabilities demonstrates remarkable interpolation capabilities for all three models and acceptable extrapolation performance for stretch-based and strain-based models. Finally, robustness tests on noise-embedded data underscore the reliability of our symbolic regression algorithms. Our study confirms the applicability and accuracy of symbolic regression in the automated discovery of isotropic hyperelastic models for the human brain and gives rise to a wide variety of applications in other soft matter systems.

Statement of significance

Our research introduces a pioneering data-driven framework that revolutionizes the automated identification of hyperelastic constitutive models, particularly in the context of soft matter systems such as the human brain. By harnessing the power of symbolic regression, we have unlocked the ability to distill intricate physical phenomena into elegant and interpretable mathematical expressions. Our approach not only ensures accurate fitting to sparse data but also upholds crucial hyperelasticity constraints, including polyconvexity, essential for maintaining physical relevance.

© 2024 Acta Materialia Inc. Published by Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

1. Introduction

Constitutive relationship, articulating mechanical behaviors within specific configurations, plays indispensable roles in engineering analyses and in silico simulations [1–5]. In the general

context, we formulate constitutive relationships deploying various formats to capture material responses, including elasticity, hyperelasticity, viscoelasticity, or plasticity. A more concise approach involves employing Onsager's thermodynamic framework [6,7], where a scalar thermodynamic potential proves sufficient for fully characterizing the mechanical behaviors of diverse materials. Specifically, this individual potential comprises two independent counterparts: the reversible Helmholtz free energy potential and

* Corresponding author.

E-mail address: xqwang@uga.edu (X. Wang).

the irreversible dissipation potential [8]. In cases with no dissipative effects like plasticity or viscosity, the Helmholtz free energy alone is capable of characterizing the mechanical response, as in our current study.

Constitutive relations are commonly characterized through experiments that typically record data in the form of displacement-force pairs or strain-stress pairs as their derivatives. The conventional strategy involves directly calibrating this data using regression algorithms, such as the least square method. However, in these methods, a pre-established material model must be defined prior to calibration. Therefore, the efficacy of calibration is heavily reliant on the initial model selection, which is significantly influenced and biased by individual experiences. Typically, massive efforts are invested in iteratively seeking an appropriate material model, resulting in a tedious and laborious calibration procedure [8]. More promising and automated approaches are the data-driven techniques, especially neural networks, which are primarily built on machine learning backbones and have emerged as versatile tools for facilitating the discovery of constitutive models. Classic neural networks, such as Feed-Forward Neural Network (FFNN), adopt fully connected structures and are commonly trained with strain input and stress output. While these networks can accurately interpolate experimental data, they may struggle with overfitting and fail to extrapolate well outside the training regime [9,10]. To address these limitations, Physics Informed Neural Networks (PINN) are introduced, with particular considerations of physical laws [11]. This approach mainly operates in two ways: i) customizing loss functions by introducing additional terms to penalize the violation of physical laws [12–14]; and ii) crafting network architectures in accordance with the physical validity constraints [15,16]. Representative examples are the Constitutive Artificial Neural Network (CANN) [17], Input Convex Neural Network (ICNN) [18], and the Neural Ordinary Differential Equations (NODE) [19]. Comprehensive explanations and benchmark tests of the CANN, ICNN, and NODE models are available in a recent review [20]. Additional efforts include the Gaussian process [21], spline approximation [22], and probability inference [23]. Moreover, unsupervised investigations into constitutive models, such as the Efficient Unsupervised Constitutive Law Identification and Discovery (EUCLID) model, have also been explored [8].

Data-driven methods based on machine learning, though promising in automated discovery of constitutive models, usually share the following two weaknesses: i) *Black Box Nature*: The model discovery process often acts as a “black box”, making it impossible to express predictions in explicit mathematic formulas. This lack of transparency largely restricts the interpretability and portability of the predicted model. ii) *Limited Functional Space*: The available functional space for model selection is often constrained. For instance, the CANN are confined to a functional set comprising only 12 exquisitely designed terms [24], while In EUCLID, the identification of the material type and calibration of model parameters are simultaneously achieved by determining finite material parameters within a generalized material library [8]. This limitation could potentially lead to challenges in navigating multiple local optima, thereby missing the global optimal constitutive model.

Symbolic regression, an alternative data-driven approach widely used in scientific research [25–29], differs from the machine learning-based methods by employing genetic programming (GP) algorithms. This approach can automatically decipher mathematic information from pure data without specific need of a priori knowledge about the investigated systems, thereby significantly enhancing interpretability [30]. Operating on tree structures, symbolic regression iteratively searches for candidate algebraic models that gradually match the provided data in an evolutionary manner. Theoretically, the functional space in symbolic regression can be considered infinite.

The utilization of symbolic regression in constitutive modeling has gained popularity in recent years. For example, sparse symbolic regression has been used to identify algebraic stress models from high-fidelity simulation data, with the predicted models exhibiting significant superiority over traditional turbulent models [31]. Additionally, symbolic regression has also been applied in model characterization and parameter calibration in the plasticity regime [32]. A recent study integrated PINNs with symbolic regression and successfully discovered several novel reaction-diffusion models, capable of describing the spatio-temporal diffusion patterns of misfolded tau proteins in Alzheimer's disease [13]. However, few endeavors have been explored in the realm of hyperelasticity. Limited studies involve the first attempt in characterizing the multi-axial loading behaviors of vulcanized rubber [33,34] and a recent work in identifying hyperelastic models for particle-reinforced composites through the cooperation of symbolic regression and FFNN models [35]. In the latter work, the neural network was employed to facilitate the differentiation operation and enforce physical admissibility laws, while the symbolic regression served as a mathematical toolbox to generate algebraic formulas. Though achieving satisfactory accuracy, massive concurrent interactions between symbolic regression and neural networks inevitably increase the computational cost, especially when the anticipated model incorporates a complicated format. Therefore, it remains an open question whether it is possible to discover hyperelastic models solely within the framework of symbolic regression, while rigorously adhering to physical constraints.

In this study, we aim to explore the capabilities of symbolic regression in automatically discovering hyperelastic models that rigorously comply with physical requirements. To achieve this objective, we investigated three distinct hyperelastic scenarios, invariant-based, principal stretch-based, and normal strain-based hyperelastic models. These investigations are based on multi-mode experimental data from the human brain cortex [1]. In accordance with physical constraints, we meticulously design the modeling structures, especially focusing on the objective functions. The structure of this paper is organized as follows: First, we introduce the theory of constitutive modeling, with particular emphasis on physical constraints, and the symbolic regression algorithms, along with implementation details in Section 2. Following validation against multiple synthetic datasets, the approaches are implemented on experimental data to discover hyperelastic models for the human brain cortex, with results and discussion presented in Section 3. Finally, we conclude our findings and outline potential directions for future explorations in Section 4.

2. Theoretical method and symbolic regression algorithm

In this section, we revisit the fundamental theorem of continuum mechanics and delineate crucial constraints essential for ensuring the physical admissibility of strain energy function derived through symbolic regression. First, we briefly review the descriptions pertaining to the kinematic equations and constitutive equations within the framework of continuum mechanics. Then, we delve into the critical conditions necessary to acknowledge the physical constraints of strain energy function, especially the requirement of convexity. Once these foundational aspects are established, we proceed to predict the general form of the strain energy function for human brain tissue based on experimental data using symbolic regression, the algorithm and implementation details of which will be introduced subsequently.

2.1. Constitutive modeling

In the context of continuum mechanics, the kinematics of a continuum body can be described by a one-to-one mapping de-

noted as $\mathbf{x} = \boldsymbol{\varphi}(\mathbf{X})$, where a material particle initially positioned at $\mathbf{X}$ in the reference configuration $\mathcal{B}_0$ is carried to its new position $\mathbf{x}$ in the current configuration $\mathcal{B}_t$. Quantitatively, we employ the deformation gradient $\mathbf{F} = \nabla_{\mathbf{x}}\boldsymbol{\varphi}$ to quantify the mapping of the line element from reference to current configuration, and the Jacobian $J = \det \mathbf{F}$ to describe the associated volume alternation. Left multiplying $\mathbf{F}$ by its transpose $\mathbf{F}^T$ yields the right Cauchy-Green deformation tensor $\mathbf{C} = \mathbf{F}^T \mathbf{F}$, which possesses three complete and irreducible principal scalar invariants,

$$\begin{aligned} I_1 &= \text{tr} \mathbf{C} = \lambda_1^2 + \lambda_2^2 + \lambda_3^2, \\ I_2 &= \text{tr}(\text{cof} \mathbf{C}) = \frac{1}{2}(I_1^2 - \text{tr}(\mathbf{C}^2)) = \lambda_1^2 \lambda_2^2 + \lambda_2^2 \lambda_3^2 + \lambda_1^2 \lambda_3^2, \\ I_3 &= \det \mathbf{C} = \lambda_1^2 \lambda_2^2 \lambda_3^2, \end{aligned} \quad (1)$$

where $\text{cof} \mathbf{C} = \det(\mathbf{C})\mathbf{C}^{-1}$ denotes the cofactor of $\mathbf{C}$; $\text{tr}(\cdot)$ and $\det(\cdot)$ are trace and determinant operators, respectively. The deformation gradient tensor $\mathbf{F}$ can be expressed as $\mathbf{F} = \nabla_{\mathbf{x}}\boldsymbol{\varphi} = \sum_{i=1}^3 \lambda_i \mathbf{n}_i \otimes \mathbf{N}_i$, where λ_i , known as the principal stretches, are the square roots of the eigenvalues of the right Cauchy-Green deformation tensor $\mathbf{C}$. The vectors $\mathbf{n}_i$ and $\mathbf{N}_i$ represent the corresponding principal directions in the current and reference configurations, respectively. In the undeformed state, both the deformation gradient and the Cauchy-Green deformation tensor are identical to the unit tensor: $\mathbf{F} = \mathbf{I}$, $\mathbf{C} = \mathbf{I}$, and the Jacobian equals one, $J = 1$.

Furthermore, we introduce two types of stresses: the symmetric Cauchy stress $\boldsymbol{\sigma}$, denoting the force per deformed area along the outward normal direction $\mathbf{n}_s$, and the asymmetric first Piola-Kirchhoff stress $\mathbf{P}$, defined as the force per undeformed area along the outward normal direction $\mathbf{N}_s$. The transpose of the latter is also known as nominal stress, which is commonly employed as the stress measure in experiments. The relation between these two stresses is characterized by Piola transformation:

$$\mathbf{P} = J\boldsymbol{\sigma}\mathbf{F}^{-T} \text{ or } \boldsymbol{\sigma} = J^{-1}\mathbf{P}\mathbf{F}^T. \quad (2)$$

In general, the second-order stress tensor $\mathbf{P}$ is not symmetric and has nine independent components.

The constitutive relationship establishes the connection between strain and stress in a material, reflecting the material response under external stimuli like applied forces or temperature variation. This relationship is a fundamental aspect of the material behavior and is commonly expressed in mathematical or tensorial form, as exemplified by the first Piola-Kirchhoff stress and the deformation gradient, $\mathbf{P} = \mathbf{P}(\mathbf{F})$. For hyperelastic materials, constitutive relations can be reformulated by positing the existence of the strain energy function (Ψ), i.e., the Helmholtz free energy. The strain energy function, under isothermal conditions, provides an implicit mathematical combination of the strain and stress tensors, $\mathbf{P} = \partial \Psi(\mathbf{F}) / \partial \mathbf{F}$. An elegant hyperelastic model should not only exhibit prominent accuracy in characterizing material behaviors but also preserve precision in dealing with perturbations or noise from experimental measurements [36,37]. To uphold these qualities, hyperelastic models must rigorously adhere to physical laws such as thermodynamic consistency and possess well-posed mathematical representations [38]. Furthermore, these models should be exquisitely crafted to meet the criteria of objectivity, stability, and, if desired, material symmetry and incompressibility [18,39]. For a comprehensive explanation of these criteria, please refer to the supplementary material.

Material stability is crucial to ensure the existence and uniqueness of solutions for boundary-value problems [40,41]. Its mathematical representation equates to the convexity, which plays a pivotal role as it implies ellipticity, thereby assuring material stability in a constitutive model by prescribing convex shapes [9]. Moreover, convexity ensures that the energy function exclusively attains its global minimum at thermodynamic equilibrium within the reference configuration. However, general convexity is often considered

too restrictive from physical perspectives and proves challenging to impose in practical mathematical applications [42]. Instead, a less restrictive requirement is the polyconvexity of the strain energy function [19,43]. Polyconvexity of $\Psi(\mathbf{F})$ demands sufficient convexity concerning the extended domain formed by deformation gradient $\mathbf{F}$, its cofactor $\text{cof} \mathbf{F}$, and determinant $\det \mathbf{F}$ [44]. Therefore, there exists a representative strain energy function $\hat{\Psi}(\mathbf{F})$ such that:

$$\Psi(\mathbf{F}) = \hat{\Psi}(\mathbf{F}, \text{cof} \mathbf{F}, \det \mathbf{F}). \quad (3)$$

Constructing a general function that precisely fulfills this requirement can be challenging, and a more flexible and pragmatic approach is to find subsets through the additive decomposition [45],

$$\hat{\Psi}(\mathbf{F}, \text{cof} \mathbf{F}, \det \mathbf{F}) = \hat{\Psi}_F(\mathbf{F}) + \hat{\Psi}_{\text{cof}}(\text{cof} \mathbf{F}) + \hat{\Psi}_{\det}(\det \mathbf{F}), \quad (4)$$

where $\hat{\Psi}_F$, $\hat{\Psi}_{\text{cof}}$, and $\hat{\Psi}_{\det}$ are convex function with respect to $\mathbf{F}$, $\text{cof} \mathbf{F}$, and $\det \mathbf{F}$, respectively. Above considerations of polyconvexity all pertain to the deformation gradient $\mathbf{F}$, however, the convexity with respect to $\mathbf{F}$ encounters incompatible with the principle of objectivity and is not suitable for finite elasticity. To address this limitation, we reformulate the polyconvexity condition by involving the invariants of the right Cauchy Green deformation tensor $\mathbf{C}$. Moreover, it is noteworthy that non-decreasing substitutions of invariants, $I_1 = \text{tr} \mathbf{C}$, $I_2 = \text{tr}(\text{cof} \mathbf{C})$, $I_3 = \det \mathbf{C}$, preserve convexity [35]. Therefore, the strain energy function can be further simplified as the summation of invariant-based functions,

$$\hat{\Psi}_F(\mathbf{F}) + \hat{\Psi}_{\text{cof}}(\text{cof} \mathbf{F}) + \hat{\Psi}_{\det}(\det \mathbf{F}) = \hat{\Psi}_{I_1}(I_1) + \hat{\Psi}_{I_2}(I_2) + \hat{\Psi}_{I_3}(I_3), \quad (5)$$

where $\hat{\Psi}_{I_1}$, $\hat{\Psi}_{I_2}$, $\hat{\Psi}_{I_3}$ are convex functions with respect to the three invariants. Accounting for incompressibility, the contribution of I_3 can be neglected due to its constant value. Consequently, polyconvexity enforces the forms of the strain energy function as $\Psi(\mathbf{F}) = \hat{\Psi}_{I_1}(I_1) + \hat{\Psi}_{I_2}(I_2)$.

An alternative and less restrictive convexity condition is rank-one convexity, commonly referred to as the strong ellipticity condition. The strain energy function remains elliptic when the Legendre-Hadamard condition is satisfied [42]:

$$(\mathbf{M} \otimes \mathbf{m}) : \frac{\partial^2 \Psi}{\partial \mathbf{F}^2} : (\mathbf{M} \otimes \mathbf{m}) \geq 0 \quad (6)$$

where $\mathbf{M}$ and $\mathbf{m}$ denote arbitrary vectors in the reference or material and current or spatial configurations. The ellipticity condition implies positive-semi-definiteness of the tangent tensor $\partial^2 \Psi / \partial \mathbf{F}^2$, which is critical for maintaining material stability [38,46]. When the vectors are coaxial with the principal direction of the right stretch tensor $\mathbf{U} = \sqrt{\mathbf{C}}$, i.e., $\mathbf{N}$ and $\mathbf{n}$, the tangent tensor reduces to the following Hessian matrix,

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 \Psi}{\partial \lambda_1^2} & \frac{\partial^2 \Psi}{\partial \lambda_1 \partial \lambda_2} & \frac{\partial^2 \Psi}{\partial \lambda_1 \partial \lambda_3} \\ \frac{\partial^2 \Psi}{\partial \lambda_2 \partial \lambda_1} & \frac{\partial^2 \Psi}{\partial \lambda_2^2} & \frac{\partial^2 \Psi}{\partial \lambda_2 \partial \lambda_3} \\ \frac{\partial^2 \Psi}{\partial \lambda_3 \partial \lambda_1} & \frac{\partial^2 \Psi}{\partial \lambda_3 \partial \lambda_2} & \frac{\partial^2 \Psi}{\partial \lambda_3^2} \end{bmatrix}, \quad (7)$$

here, λ_1 , λ_2 , λ_3 are the three eigenvalues of $\mathbf{U}$, corresponding to the principal stretches along three principal directions. If all three eigenvalues are positive and the determinant of the Hessian matrix $\mathbf{H}$ is positive, we can confirm the positive definiteness of $\mathbf{H}$. However, it is important to note that satisfying Eq. (7) only ensures ellipticity along the principal directions. To verify the ellipticity of strain energy function $\Psi(\mathbf{F})$ across the entire space, we further examined the contours of the strain energy function in the principal

stretch space [47]. Analogously, the additive decomposition can be applied to refine the available function subsets:

$$\Psi(\lambda_1, \lambda_2, \lambda_3) = \Psi_1(\lambda_1) + \Psi_2(\lambda_2) + \Psi_3(\lambda_3). \quad (8)$$

This decomposition ensures that all off-diagonal components of the Hessian matrix are zero, and its three eigenvalues directly correspond to the diagonal components: $\partial^2 \Psi / \partial \lambda_1^2$, $\partial^2 \Psi / \partial \lambda_2^2$, $\partial^2 \Psi / \partial \lambda_3^2$. Notably, the polyconvexity condition is slightly more restrictive than the rank-one convexity condition. However, both conditions are sufficient conditions for the existence of minimizers, indicating the potential existence of multiple local minimizers under each condition. To address this issue, a coercivity condition can be introduced. The coercivity condition, also known as the growth condition, demands that the stresses should grow unboundedly with infinite strains [35,48]. For simplicity, the strain energy Ψ is considered infinite for infinite compression, $J \rightarrow 0$, and infinite expansion, $J \rightarrow \infty$.

In the current study, the polyconvexity condition, along with the coercivity condition, is employed to determine the existence and uniqueness of the invariant-based strain energy function $\Psi(I_1, I_2)$, while the rank-one convexity condition, along with the coercivity condition, is utilized to verify the existence of global minimizers for the principal stretch-based or strain-based strain energy function $\Psi(\lambda_1, \lambda_2, \lambda_3)$.

2.2. Symbolic regression

Symbolic regression stands out as a distinctive form of regression, wherein a mathematic expression is autonomously identified to best fit the provided dataset. Unlike conventional regression or data-driven methods that require predefined model structures, symbolic regression is capable of generating analytical expressions purely from data without the specific need of prior knowledge, thereby significantly enhancing the interpretability, generalizability, and flexibility of the model discovery process [30].

The algorithm for symbolic regression unfolds in an evolutionary manner, known as genetic programming, which draws inspirations from the Darwinian principles of natural selection. Within GP, functional expressions are efficiently represented using a binary-tree structure, comprising nodes and branches, as illustrated in Fig. 1a. A complete tree structure involves variables, mathematic operators (either unary or binary), and constants. In the initial stages, the algorithm randomly generates a population of symbolic tree expressions based on user-defined variables and operators, serving as candidate functionals. For each candidate expression, the fitness is evaluated through the calculation of the mean square error (MSE) between predicted outputs and target values. Expressions with higher fitness values are more likely to be selected as baselines for subsequent optimization, where the expression trees

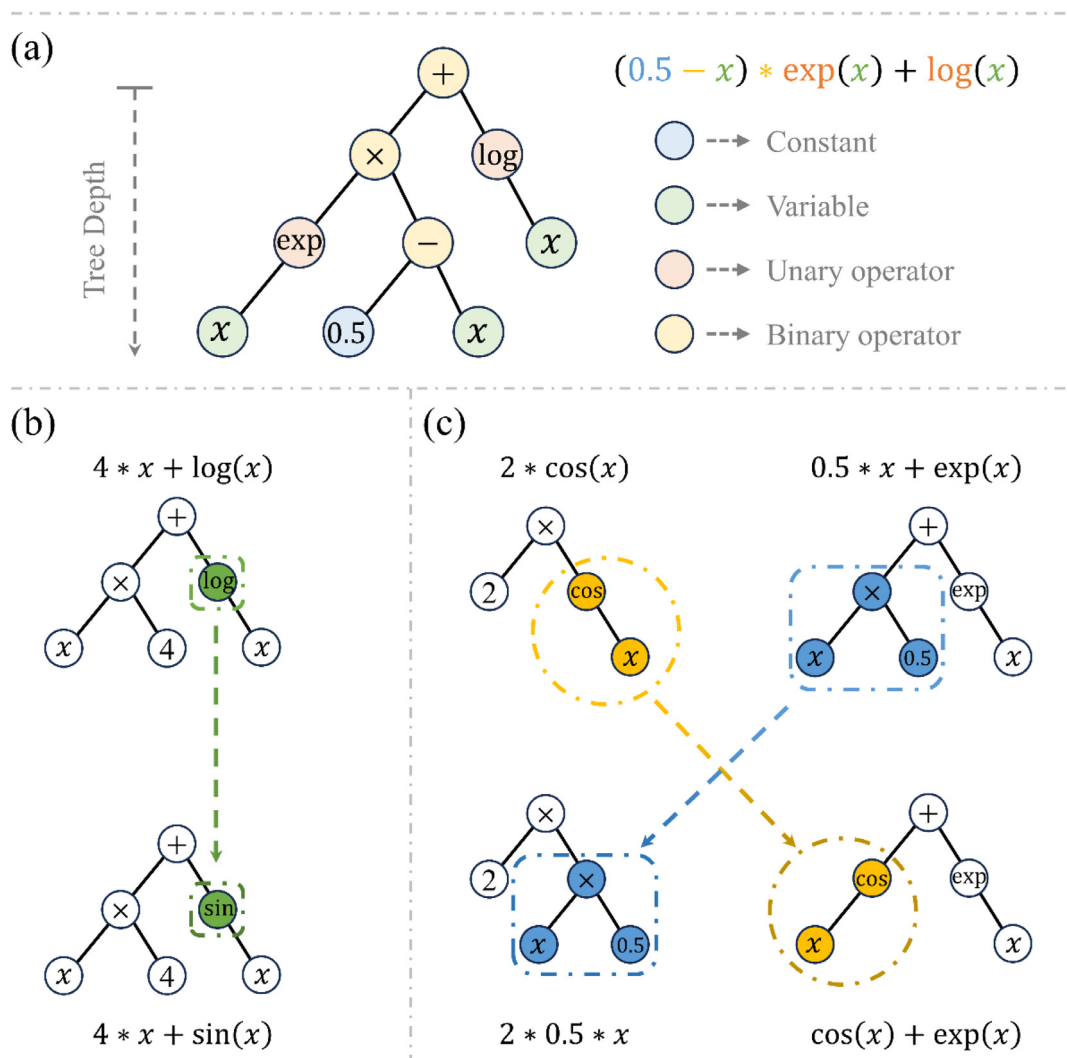


Fig. 1. Structure and operations of expression tree. (a). Representation of expression tree for an algebraic expression $((0.5 - x) * \exp(x) + \log(x))$; (b). An example of the mutation operation; (c). An example of the crossover operation.

are modified primarily through two genetic operations: mutation and crossover. The mutation operation entails randomly altering some nodes in an expression tree, introducing genetic diversity in the population. An example is shown in Fig. 1b, where a new offspring is generated by replacing the unary operator “log” with “sin”. On the other hand, the crossover operation permits the algorithms to create new offspring by combining building blocks from different parent individuals, as demonstrated in Fig. 1c. These newly generated expressions become candidates for the next iteration. The iterative process of evaluation, selection, mutation, and crossover constitutes one evolution of the regression analysis. This cycle is repeated until the optimal expression is obtained or the maximum number of generations is reached [49,50].

To enhance interpretability and mitigate potential overfitting, achieving a balance between model accuracy and complexity is crucial. However, there is still a lack of consensus on the precise definition of these two terms. Herein, we adopt a proposed measure [51] that defines complexity as the number of nodes in an expression tree. The loss of expression $\mathcal{L}(Expr)$ is then evaluated as a combination of the predictive loss $\mathcal{L}_{pred}(Expr)$ and the complexity measure $C(Expr)$,

$$\mathcal{L}(Expr) = \mathcal{L}_{pred}(Expr) \cdot \exp(\text{freq}[C(Expr)]), \quad (9)$$

where $\text{freq}[C(Expr)]$ defines a combined measure of the frequency and recency of expressions occurring at the current complexity in the population [51]. The optimal expression is determined based on a score metric, as the negated derivative of the log-loss with respect to the complexity, $-d[\log(\text{MAE})]/dC$, here MAE denotes the mean absolute error between the prediction and the data. Figure S1 presents an example illustrating the evolutionary process in searching for the target strain energy function.

In this manuscript, we employ the symbolic regression algorithms introduced above to predict the optimal strain energy function for human brain tissue. To explore the capabilities of symbolic regression, we incorporate three distinct sets of variables as model inputs: the invariants (I_1, I_2), principal stretches ($\lambda_1, \lambda_2, \lambda_3$), and principal strains ($\epsilon_1, \epsilon_2, \epsilon_3$). The corresponding strain energy functions we seek to discover are $\Psi(I_1, I_2)$, $\Psi(\lambda_1, \lambda_2, \lambda_3)$, $\Psi(\epsilon_1, \epsilon_2, \epsilon_3)$, respectively. For clarity, we term the related algorithms as “Invariant-based Symbolic Regression”, “Stretch-based Symbolic Regression”, and “Strain-based Symbolic Regression”. In each algorithm, we meticulously craft the model structure and objective functions to ensure the physical admissibility of the strain energy functions by adhering to the physical constraints outlined in Section 2.1.

2.2.1. Invariant-based symbolic regression

Invariant-based Symbolic Regression takes the invariants (I_1, I_2, I_3) of the right Cauchy-Green tensor $\mathbf{C}$ as input and the strain energy function Ψ as output. The data used to train these symbolic regression models include the deformation gradient $\mathbf{F}$ and the first Piola-Kirchhoff stresses $\mathbf{P}$. Below, we briefly introduce the mathematical representations of these quantities based on the experimental measures for each loading mode. For a detailed derivation process, please refer to the Supplementary Material.

In the case of unconfined uniaxial tension and compression tests, the specimen undergoes uniform deformation along the stretching direction, denoting as λ . Under the assumptions of material isotropy and perfect incompressibility, the uniaxial stress can be calculated as:

$$P_{11} = 2 \left(\frac{\partial \Psi}{\partial I_1} + \frac{1}{\lambda} \frac{\partial \Psi}{\partial I_2} \right) \left(\lambda - \frac{1}{\lambda^2} \right) \quad P_{22} = P_{33} = 0. \quad (10)$$

During the simple shear test, assuming a specified amount of shear (γ) is applied in the x-y plane of an isotropic and perfectly in-

compressible specimen, the shear stress has the following explicit form:

$$P_{12} = 2 \left(\frac{\partial \Psi}{\partial I_1} + \frac{\partial \Psi}{\partial I_2} \right) \gamma. \quad (11)$$

In this manuscript, we employ the normalized mean square error to evaluate the loss between the predicted stress $\mathbf{P}^*$ and the experimental stresses $\mathbf{P}$,

$$L_{pred} = \frac{1}{N_{ut}} \sum_{i=1}^{N_{ut}} \left\| \frac{P_{ut,i} - P_{ut,i}^*}{P_{ut,i}^{max}} \right\|^2 + \frac{1}{N_{uc}} \sum_{i=1}^{N_{uc}} \left\| \frac{P_{uc,i} - P_{uc,i}^*}{P_{uc,i}^{max}} \right\|^2 + \frac{1}{N_{ss}} \sum_{i=1}^{N_{ss}} \left\| \frac{P_{ss,i} - P_{ss,i}^*}{P_{ss,i}^{max}} \right\|^2. \quad (12)$$

Here, each loss term is normalized by the maximum experimental stress ($P_{ut,i}^{max}, P_{uc,i}^{max}, P_{ss,i}^{max}$) to mitigate the impact introduced by the choice of stress measure [52,53]. $P_{ut} = P_{11}$ for tension ($\lambda > 1$), $P_{uc} = P_{11}$ for tension ($\lambda < 1$), and $P_{ss} = P_{12}$ for simple shear ($\gamma > 0$).

The data fed into *Invariant-based Symbolic Regression* involve the invariants (I_1, I_2) and stresses (P_{ut}, P_{uc}, P_{ss}). Performing symbolic regression on these data directly yields a relation between stresses and invariants, such as $P_{ut}(I_1, I_2)$. However, thermodynamic consistency dictates an indirect relation, $\mathbf{P} = \partial \Psi / \partial \mathbf{F}$, implying that the target output of symbolic regression should be the strain energy $\Psi(I_1, I_2)$. To address this, we customize the objective function to enable auto-differentiation inside the loss function, as shown in Algorithm 1. At each step, the derivatives of Ψ with respect to I_1 and I_2 are calculated and stored as $\partial \Psi / \partial I_1$ and $\partial \Psi / \partial I_2$. These two derivatives are essential for determining the first Piola-Kirchhoff stresses, as indicated in Eqs. (10) and (11). Furthermore, to ensure a stress-free reference configuration, both invariants are shifted by 3 at the initial stage. Regarding the polyconvexity condition, we enforce non-negativities for all constants by imposing an extremely large values into the loss function whenever any node constant becomes negative, ensuring that the base models used to construct the hyperelastic model remain convex and non-decreasing (Supplementary Section 3) [17,35].

In symbolic regression, the expression tree theoretically can take an arbitrary functional shape. However, in consideration of the computational costs, we constrain the evolving expressions to be within the domain constructed by polynomial, exponential, and logarithmic functions. These forms are commonly utilized in the classical hyperelastic models, such as Mooney Rivlin model [54], Gent model [55], and Holzapfel model [56]. The detailed model setups are summarized in Table S1. Note that, the complexities and constraints can be flexibly tuned for specific problems. For exam-

Algorithm 1

Framework of the Customized Loss Function for the *Invariant-based Symbolic Regression*.

Input: First and second invariants calculated from experimental stretches of uniaxial tension, uniaxial compression, and simple shear, I_1, I_2 ; First Piola-Kirchhoff stress from tension P_{ut} , compression P_{uc} , and shear P_{ss} ;

Output: Normalized mean square error, L_{pred} ;

- 1: Shift I_1 and I_2 with 3 to ensure stress-free state at initial configuration;
- 2: Concatenate the inputs of three loading modes data along the raw direction;
- 3: Constrain all the constants to be non-negative;
- 4: Calculate the derivatives of strain energy density w.r.t invariants, $\frac{\partial \Psi}{\partial I_1}, \frac{\partial \Psi}{\partial I_2}$;

- 5: Calculate the stretches or shear for each loading mode, $\lambda_{ut}, \lambda_{uc}, \lambda_{ss}$;
- 6: Determine the predicted first Piola-Kirchhoff stress:

$$P_{ut}^* = 2 \left(\lambda_{ut} - \frac{1}{\lambda_{ut}} \right) \left(\frac{\partial \Psi}{\partial I_1} + \frac{1}{\lambda_{ut}} \frac{\partial \Psi}{\partial I_2} \right), \quad P_{uc}^* = 2 \left(\lambda_{uc} - \frac{1}{\lambda_{uc}} \right) \left(\frac{\partial \Psi}{\partial I_1} + \frac{1}{\lambda_{uc}} \frac{\partial \Psi}{\partial I_2} \right),$$

$$P_{ss}^* = 2 \left(\frac{\partial \Psi}{\partial I_1} + \frac{\partial \Psi}{\partial I_2} \right) \gamma_{ss};$$

- 7: Evaluate the loss:

$$L_{pred} = \frac{1}{N_{ut}} \sum_{i=1}^{N_{ut}} \left\| \frac{P_{ut,i} - P_{ut,i}^*}{P_{ut,i}^{max}} \right\|^2 + \frac{1}{N_{uc}} \sum_{i=1}^{N_{uc}} \left\| \frac{P_{uc,i} - P_{uc,i}^*}{P_{uc,i}^{max}} \right\|^2 + \frac{1}{N_{ss}} \sum_{i=1}^{N_{ss}} \left\| \frac{P_{ss,i} - P_{ss,i}^*}{P_{ss,i}^{max}} \right\|^2;$$

- 8: **return** L_{pred} .

ple, we can adjust the complexity of the multiplication operator (“*”) to a much larger value than that of addition operator (“+”) if we need to restrict the use of “*” during functional evolution.

2.2.2. Stretch-based symbolic regression

Stretch-based Symbolic Regression takes the principal stretches $(\lambda_1, \lambda_2, \lambda_3)$ of the right stretch tensor $\mathbf{U}$ as input and the strain energy function $\Psi(\lambda_1, \lambda_2, \lambda_3)$ as output. For this particular regression, we confined the functional format to rigorously follow the generalized Ogden model due to its demonstrated efficiency and suitability in describing the nonlinear and asymmetric mechanical behaviors of isotropic soft tissue, such as the human brain [57],

$$\Psi(\lambda_1, \lambda_2, \lambda_3) = \sum_{k=1}^n \frac{\mu_k}{\alpha_k^2} [\lambda_1^{\alpha_k} + \lambda_2^{\alpha_k} + \lambda_3^{\alpha_k} - 3], \quad (13)$$

where μ_k represents the shear stiffness, α_k the nonlinearity parameter. Prior to application, the first step is to convert the experimental measure of the right stretch tensor into its spectral representation, $[\mathbf{U}] = \text{diag}(\lambda_1, \lambda_2, \lambda_3)$. For the case of unconfined uniaxial tension and compression, the nominal uniaxial stress can be calculated as:

$$P_{11} = \frac{\partial \Psi}{\partial \lambda_1} - \frac{1}{\lambda_1 \sqrt{\lambda_1}} \frac{\partial \Psi}{\partial \lambda_2}. \quad (14)$$

In the simple shear test prescribed in the x-y plane, the shear stress takes the following succinct expression:

$$P_{12} = \frac{\lambda_1^2}{\lambda_1^2 + 1} \frac{\partial \Psi}{\partial \lambda_1} - \frac{\lambda_2^2}{\lambda_2^2 + 1} \frac{\partial \Psi}{\partial \lambda_2}. \quad (15)$$

For a detailed mathematic derivation of Eq. (15), please refer to the Appendix B in our recently published paper [53].

Analogously, for *Stretch-based Symbolic Regression*, we employ the normalized mean square error to evaluate the loss between the predicted stress $\mathbf{P}^*$ and experimental stresses $\mathbf{P}$, with the expression identical to Eq. (12). Again, we customize the objective function to ensure the thermodynamic consistency, with its framework depicted in Algorithm 2.

At each step, the derivatives of Ψ with respect to λ_1, λ_2 , and λ_3 are calculated and stored in $\partial \Psi / \partial \lambda_1, \partial \Psi / \partial \lambda_2$, and $\partial \Psi / \partial \lambda_3$, respectively. These derivatives further contribute to determining the first Piola-Kirchhoff stresses P_{ut}^* , P_{uc}^* , and P_{ss}^* , as indicated in Eqs. (14) and (15). In the algorithm, the first step is to concatenate all the principal stretches and stresses into one column, respectively. This tricky recombination is crucial for code execution because the expression tree is constrained to follow the format of

the Ogden model. Thus, differentiation operations of Ψ with respect to each principal stretch share equal weight. This allows us to simplify the derivatives of Ψ with respect to a single variable, e.g., $\frac{\partial \Psi}{\partial \lambda}|_{\lambda=\lambda_1, \lambda_2, \lambda_3}$. Notably, the convexity requirement is not enforced as described for the *Invariant-based Symbolic Regression* because the nonlinearity parameter α_k is allowed to be negative. However, the coefficient μ_k / α_k^2 must be strictly positive to ensure the positivity of shear stiffness [58]. Hence, the rank-one convexity condition will be validated post hoc by determining the positive definiteness of the Hessian matrix, as described in Section 2.1. The evolving expressions are restricted to polynomial functions.

2.2.3. Strain-based symbolic regression

Strain-based Symbolic Regression takes the principal strains $(\epsilon_1, \epsilon_2, \epsilon_3)$ of the right stretch tensor $\mathbf{U}$ as input and the strain energy function $\Psi(\epsilon_1, \epsilon_2, \epsilon_3)$ as output. The principal strains are also referred to as Biot strains, representing the strain measure in the normal direction. The relation between Biot strain and principal stretch is described as:

$$\epsilon_i = \lambda_i - 1, \quad \text{with } i \text{ in } 1, 2, 3 \quad (16)$$

In contrast to *Stretch-based Symbolic Regression*, where an Ogden functional format is specified for the expression tree, *Strain-based Symbolic Regression* considers the strain energy function as a polynomial series of the Biot strain measure [53],

$$\Psi(\epsilon_1, \epsilon_2, \epsilon_3) = \sum_{k=1}^n \beta_k (\epsilon_1^k + \epsilon_2^k + \epsilon_3^k). \quad (17)$$

Using the relation defined in Eq. (16), we can reformulate the mathematic representation of stresses in terms of the principal strains $(\epsilon_1, \epsilon_2, \epsilon_3)$. For the case of unconfined uniaxial tension and compression, the nominal uniaxial stress can be determined as:

$$P_{11} = \frac{\partial \Psi}{\partial \epsilon_1} - \frac{1}{(\epsilon_1 + 1)\sqrt{(\epsilon_1 + 1)}} \frac{\partial \Psi}{\partial \epsilon_2}. \quad (18)$$

In the simple shear test prescribed in the x-y plane, the shear stress has the following expression:

$$P_{12} = \frac{(\epsilon_1 + 1)^2}{(\epsilon_1 + 1)^2 + 1} \frac{\partial \Psi}{\partial \epsilon_1} - \frac{(\epsilon_2 + 1)^2}{(\epsilon_2 + 1)^2 + 1} \frac{\partial \Psi}{\partial \epsilon_2}. \quad (19)$$

For *Strain-based Symbolic Regression*, we also use the normalized mean square error to assess the difference between the predicted stress $\mathbf{P}^*$ and experimental stresses $\mathbf{P}$. Additionally, we customize the objective function to ensure thermodynamic consistency, as outlined in Algorithm 3. In each iteration, the derivatives of Ψ

Algorithm 3

Framework of the customized loss function for the *strain-based symbolic regression*.

Input: Principal strains of uniaxial tension $\epsilon_{k,ut}$, uniaxial compression $\epsilon_{k,uc}$, and simple shear $\epsilon_{k,ss}$, with k in 1, 2, 3; First Piola-Kirchhoff stress from tension P_{ut} , compression P_{uc} , and shear P_{ss} ;

Output: Normalized mean square error, L_{pred} ;

1. Concatenate all the principal strains and stresses into one column, respectively;
2. Calculate the derivatives of strain energy density w.r.t principal strains, $\frac{\partial \Psi}{\partial \epsilon_1}, \frac{\partial \Psi}{\partial \epsilon_2}, \frac{\partial \Psi}{\partial \epsilon_3}$;
3. Derive the principal stretches based on given principal strains $\lambda_{k,ut} = \epsilon_{k,ut} + 1, \lambda_{k,uc} = \epsilon_{k,uc} + 1, \lambda_{k,ss} = \epsilon_{k,ss} + 1$, with k in 1, 2, 3;
4. Determine the predicted first Piola-Kirchhoff stress: $P_{ut}^* = \frac{\partial \Psi}{\partial \epsilon_1} - \frac{\lambda_{2,ut}}{\lambda_{1,ut}} \frac{\partial \Psi}{\partial \epsilon_2}$, $P_{uc}^* = \frac{\partial \Psi}{\partial \epsilon_1} - \frac{\lambda_{2,uc}}{\lambda_{1,uc}} \frac{\partial \Psi}{\partial \epsilon_2}$, $P_{ss}^* = \frac{\lambda_{1,ss}^2}{\lambda_{1,ss}^2 + 1} \frac{\partial \Psi}{\partial \epsilon_1} - \frac{\lambda_{2,ss}^2}{\lambda_{2,ss}^2 + 1} \frac{\partial \Psi}{\partial \epsilon_2}$;
5. Evaluate the loss: $L_{pred} = \frac{1}{N_{ut}} \sum_{i=1}^{N_{ut}} \left\| \frac{P_{ut,i} - P_{ut,i}^*}{P_{ut,i}^{max}} \right\|^2 + \frac{1}{N_{uc}} \sum_{i=1}^{N_{uc}} \left\| \frac{P_{uc,i} - P_{uc,i}^*}{P_{uc,i}^{max}} \right\|^2 + \frac{1}{N_{ss}} \sum_{i=1}^{N_{ss}} \left\| \frac{P_{ss,i} - P_{ss,i}^*}{P_{ss,i}^{max}} \right\|^2$;
6. **return** L_{pred} .

Algorithm 2

Framework of the Customized Loss Function for the *Stretch-based Symbolic Regression*.

Input: Principal stretches of uniaxial tension $\lambda_{k,ut}$, uniaxial compression $\lambda_{k,uc}$, and simple shear $\lambda_{k,ss}$, with k in 1, 2, 3; First Piola-Kirchhoff stress from tension P_{ut} , compression P_{uc} , and shear P_{ss} ;

Output: Normalized mean square error, L_{pred} ;

1. Concatenate all the principal stretches and stresses into one column, respectively;
2. Calculate the derivatives of strain energy density w.r.t principal stretches, $\frac{\partial \Psi}{\partial \lambda_1}, \frac{\partial \Psi}{\partial \lambda_2}, \frac{\partial \Psi}{\partial \lambda_3}$;
3. Determine the predicted first Piola-Kirchhoff stress: $P_{ut}^* = \frac{\partial \Psi}{\partial \lambda_1} - \frac{\lambda_{2,ut}}{\lambda_{1,ut}} \frac{\partial \Psi}{\partial \lambda_2}$, $P_{uc}^* = \frac{\partial \Psi}{\partial \lambda_1} - \frac{\lambda_{2,uc}}{\lambda_{1,uc}} \frac{\partial \Psi}{\partial \lambda_2}$, $P_{ss}^* = \frac{\lambda_{1,ss}^2}{\lambda_{1,ss}^2 + 1} \frac{\partial \Psi}{\partial \lambda_1} - \frac{\lambda_{2,ss}^2}{\lambda_{2,ss}^2 + 1} \frac{\partial \Psi}{\partial \lambda_2}$;
4. Evaluate the loss: $L_{pred} = \frac{1}{N_{ut}} \sum_{i=1}^{N_{ut}} \left\| \frac{P_{ut,i} - P_{ut,i}^*}{P_{ut,i}^{max}} \right\|^2 + \frac{1}{N_{uc}} \sum_{i=1}^{N_{uc}} \left\| \frac{P_{uc,i} - P_{uc,i}^*}{P_{uc,i}^{max}} \right\|^2 + \frac{1}{N_{ss}} \sum_{i=1}^{N_{ss}} \left\| \frac{P_{ss,i} - P_{ss,i}^*}{P_{ss,i}^{max}} \right\|^2$;
5. **return** L_{pred} .

with respect to ϵ_1 , ϵ_2 , and ϵ_3 are calculated and stored in $\partial\Psi/\partial\epsilon_1$, $\partial\Psi/\partial\epsilon_2$, and $\partial\Psi/\partial\epsilon_3$, respectively. Furthermore, concatenation is also performed to facilitate the differentiation operation. Herein, we impose no restriction on the sign of β_k , however, the positivity of shear modulus is strictly enforced by considering the consistency condition, namely, the isotropic hyperelastic model should be consistent with linear elasticity theory for small strains [59],

$$\mu = \frac{1}{2} \left[\frac{\partial^2 \Psi(0, 0, 0)}{\partial \epsilon_i^2} - \frac{\partial^2 \Psi(0, 0, 0)}{\partial \epsilon_i \partial \epsilon_j} + \frac{\partial^2 \Psi(0, 0, 0)}{\partial \epsilon_i} \right] > 0. \quad (20)$$

Here, $\partial\Psi(0, 0, 0)/\partial\epsilon_i$ indicates the derivatives of $\Psi(\epsilon_1, \epsilon_2, \epsilon_3)$ in the reference configuration where all principal strains have a constant value of 0. The polyconvexity condition will be post-checked as described in Section 2.1 and the evolving expressions are confined to polynomial functions.

2.2.4. Training data and implementation details

In our study, we trained our models simultaneously using data from three loading modes—uniaxial tension, uniaxial compression, and simple shear—and evaluated their fitting accuracy within this same data regime. This approach was selected based on insights gained from testing different training scenarios. As illustrated in Figures S4–S6, employing a multi-mode training approach results in more accurate overall fittings across the three loading modes compared to the single-mode training, particularly excelling in capturing the inherent nonlinearity of soft tissues under large deformations. Both the synthetic and experimental datasets comprise 41 data points for each loading scenarios, resulting in a total of 123 data points included in the training dataset for multi-mode training. The synthetic training data was generated based on presumed function formats with the inputs of stretch (λ , γ) and outputs of the stresses (P_{11} , P_{12}). Experimental data was extracted from the literature using an open-source digitizer tool [4].

All symbolic regression analyses were performed using PySR [51], a powerful open-source package developed alongside the Julia library *SymbolicRegression.jl*. During the training of *Invariant-based Symbolic Regression*, the binary operators are restricted to addition (+) and multiplication (*), while the unary operators are limited to the exponential (exp), square (x^2), cube (x^3), and user-defined logarithmic functions ($\ln(1/(1-x))$). The training time is set to 30 min or a maximum of 1000 iterations or an early stopping criterion of MSE lower than $1e-3$, whichever is reached first. The maximum depth of the expression tree is set to 10, and the maximum complexity is constrained to 100. Notably, we customize the complexity of each operator to penalize their occurrence preference (the default is 1). Additionally, the nested behavior of the exponential and logarithmic functions is forbidden, while the square and cube operators are restricted to occur, if desired, only once inside the exponential and logarithmic functions. These nested constraints are defined for sufficient simplification of the enforcement regarding the convex and non-decreasing requirements.

During the training of *Stretch-based Symbolic Regression*, the binary operators are restricted to addition (+), multiplication (*), and polynomial functions. Here, the exponents of polynomial functions are limited to normal values within the range of $(-30, 30)$. The training time, maximum depth, and maximum complexity are the same as those of the *Invariant-based Symbolic Regression*. However, we introduce functional constraints and penalize variables more than constant to avoid the occurrence of variable exponents. Again, the nested behavior of polynomial function is forbidden. A consistent training setup is employed for the *Strain-based Symbolic Regression*. We adopt the default criterion (“best”) to guide the model selection process. Detailed training setups are summarized in Table S1. For each algorithm, the training is repeated at least three times, and favorable candidates, such as models with the top four scores, are selected as the target models. All trainings were performed on

a Legion PC equipped with a six-core Intel Core i7–8750H 2.2 GHz CPU, 4 GB NVIDIA GTX 1050Ti GPU, and 24GB of memory.

3. Numerical results and discussion

In this study, our primary objective is to investigate the capability of symbolic regression algorithms in autonomously identifying suitable hyperelastic models for soft tissues, specifically focusing on the human brain cortex. The suitability of a hyperelastic model is characterized by its accuracy, generalizability, and physical admissibility, namely, adherence to the physical constraints, as outlined in Section 2.1. Prior to application, we initially conducted equation search on synthetic dataset to validate the applicability of the algorithms. Subsequently, three distinct endeavors were undertaken for models’ discovery on the human brain cortex based on multi-mode experimental data, wherein invariants, principal stretches, principal strains were employed as the model inputs, respectively. Furthermore, the predictive capabilities of the three algorithms were evaluated regarding the interpolation and extrapolation performance. Finally, we assessed the robustness of symbolic regression by testing the model discovery performance on synthetic data embedded with varying levels of noise.

3.1. Model verification with synthetic data

The synthetic data was generated based on five classical hyperelastic models that are commonly utilized to characterize the material behavior of soft tissues: Mooney Rivlin model, Gent model, Demiray model, Holzapfel model, and Ogden model. The detailed mathematic expressions for each model are as follows:

$$\begin{aligned} \Psi_{\text{Gent}} &= -1.9 \ln(1 - 1.2[I_1 - 3]), \\ \Psi_{\text{Ogden}} &= 0.01(\lambda_1^{-18} + \lambda_2^{-18} + \lambda_3^{-18} - 3), \\ \Psi_{\text{Demiray}} &= 1.66(\exp(0.88[I_1 - 3]) - 1), \\ \Psi_{\text{Holzapfel}} &= 5.6 \left(\exp(3[I_1 - 3]^2) - 1 \right), \\ \Psi_{\text{Mooney Rivlin}} &= 0.87[I_1 - 3] + 0.86[I_1 - 3]^2 + 0.98[I_2 - 3] + 0.43[I_2 - 3]^2. \end{aligned} \quad (21)$$

Three loading modes (uniaxial tension, uniaxial compression, and simple shear) were considered, with stress data generated in terms of predefined stretches using mechanical expressions defined in Section 2.2. The *Invariant-based Symbolic Regression* algorithm was employed for the first four invariant-based models, while the Ogden model data was trained using the *Stretch-based Symbolic Regression* algorithm. The predicted models are visually presented in Fig. 2. As shown, all models were successfully predicted using our algorithms, precisely aligning with the synthetic data.

Although the model setups primarily follow the patterns outlined in Table S1, special considerations were made for each algorithm to ensure the accuracy and efficiency in discovering the target functions. First, a constant shift was manually added for Ogden model, Demiray model, and Holzapfel model to enforce the normalization or stress-free condition at the reference state. For example, in the case of the Demiray model with a direct output of $\Psi = 1.661 \exp(0.8802[I_1 - 3])$, a constant shift value of -1.661 was prescribed to ensure zero strain energy ($\Psi = 0$) in the reference configuration ($I_1 = 3$). This constant shift would not affect the stress measure as it may vanish under the differentiation operation. An alternative but automated approach involves customizing the exponential operator as $(\exp(x) - 1)$, like the approach used for logarithmic function $(-\ln(1-x))$, as seen in Figure S2. However, it should be noted that integrating customized operators would significantly increase the training time: the algorithm with the embedded operator can discover the target model within merely 100 iterations, whereas the one with customized operator requires nearly 350 iterations to find the same model. Second, we

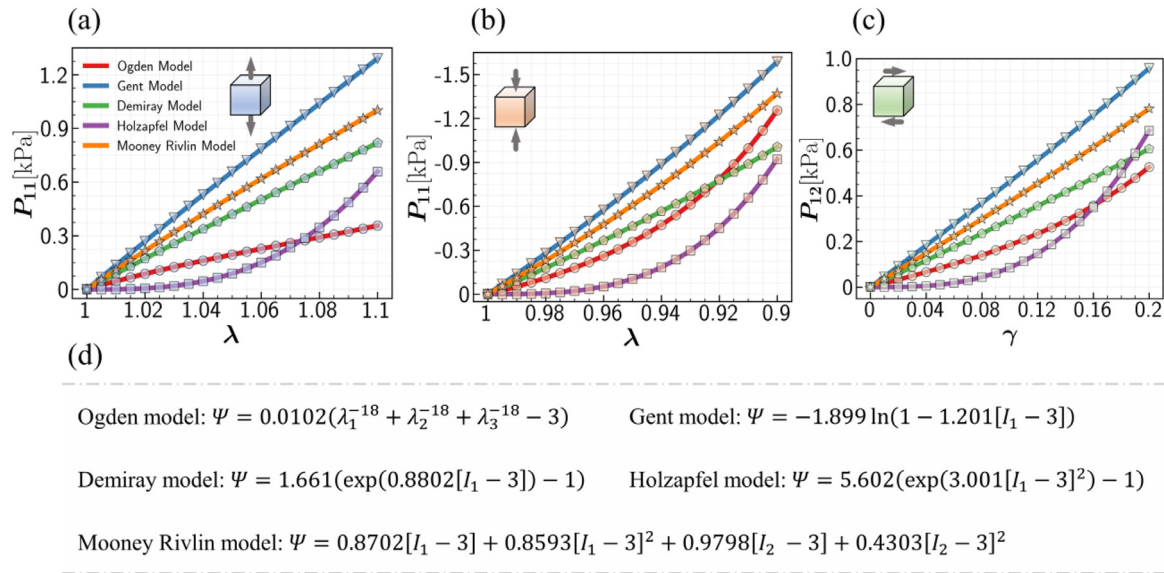


Fig. 2. Validation of symbolic regression algorithm. Models were trained simultaneously with data from three loading modes, and tested with tension (a), compression (b) and shear (c) data individually. Dots illustrate the synthetic data generated from five classical hyperelastic models that widely used for human brain characterization. Specific mathematic expressions of predicted hyperelastic models are provided in (d).

constrained the order of polynomial operator (x^y) to be strictly positive by introducing a ceiling function. To further improve efficiency in the equation search, the value of y was restricted to the range of (0, 5) and (−30, 30) for the Mooney Rivlin model and Ogden model, respectively. The successful reproductions of these models confirm the applicability and accuracy of our algorithms in the automated discovery of strain energy functions within hyperelastic regimes. Upon validation, we started to perform the symbolic regression based on experimental data, aiming to discover a suitable hyperelastic model for the human brain cortex.

3.2. Hyperelastic models discovered by the invariant-based symbolic regression algorithm

Fig. 3 illustrates four superior hyperelastic models discovered for the human brain cortex using the *Invariant-based Symbolic Regression* algorithm. These models were trained simultaneously with data from three loading modes, thus the combined loss function defined in Eq. (12) was employed for regression optimization and tested individually for each loading mode. The fitting performance of each model is evaluated by the R^2 value, defined as $R^2 = 1 - \sum_{i=1}^N (\mathbf{P}_i - \mathbf{P}_i^*)^2 / \sum_{i=1}^N (\mathbf{P}_i - \bar{\mathbf{P}})^2$, where $\bar{\mathbf{P}}$ is the mean of experimental stress. Detailed mathematical expressions of each model are provided at the bottom of the figure. As shown, these four predicted models, though employing distinct formulas, exhibit satisfactory performances in characterizing the material behaviors in uniaxial tension, uniaxial compression, and simple shear scenarios. Notably, all models are exclusively dependent on the second invariant. This observation aligns with recent findings [24,60]. Examining at curves, a consistent trend is observed across all models: overestimation in tension occurs once the stretch is greater than 5 %, underestimation persists throughout compression, and accurate fitting is observed with simple shear data. The consistent underperformance in tension and compression data indicates potential data inconsistency with the hyperelasticity assumption [20]. In essence, the experimental data may not be equally reliable for different loading modes. To address this issue, we propose introducing weighting factors in Eq. (12) to demonstrate the contribution of each loading mode to the combined loss function.

Interestingly, the predicted models Ψ_a and Ψ_b closely resemble the models discovered by the CANN regularized with subset selection (L_0 regularization) [52]. However, our identified models Ψ_c and Ψ_d , which exhibit comparable fitting performance, are not covered in their research findings. Conversely, two hyperelastic models presented in their work with satisfactory fitting accuracy are not discovered by our current algorithm. One is constructed by terms $([I_2 - 3]^2)$ and $(\exp([I_2 - 3]) - 1)$, and another is formed by $(\exp([I_2 - 3]) - 1)$ and $(\exp([I_2 - 3]^2) - 1)$. This discrepancy suggests the potential existence of multiple optima for the current optimization problem. For further exploration, it is crucial to consider either an enriched function space or more diverse loading modes when performing symbolic regression or CANN. Among the four models we discovered, the third model $\Psi_c = 0.017(\exp(27.91[I_2 - 3]) - 1)$ exhibits the highest fitting accuracy with the simplest form. Therefore, it serves as the optimal model discovered by the *Invariant-based Symbolic Regression* algorithm.

Within the framework of invariants, Fig. 4 provides a comparison on the fitting performance of the optimal hyperelastic model discovered by symbolic regression, multiple regression, and artificial neural networks. The results for the latter two models are derived from our recent paper [60], where artificial neural networks followed the idea of CANNs [17], but utilized a different loss function, the mean absolute percentage error (MAPE). The comparison reveals no significant differences among the three models, except that the symbolic regression model exhibits a smaller underestimation in tension, leading to a marginally higher R^2 value of 0.908, as shown in Fig. 4a. Though achieving comparable performance, the symbolic regression algorithm demonstrates significant advantages in computational cost. On our platform, a complete execution of symbolic regression took nearly 20 min. In comparison, CANN cost 30 min for training, while multiple regression requires almost 1 hour due to its exhaustive traversal over the entire candidate space.

For invariant-based hyperelastic models, critical physical admissible conditions elucidated in Section 2.1, such as the polyconvexity requirement, have been predefined and embedded into the symbolic regression framework. Therefore, the four models unveiled in Fig. 3 are inherently designed to satisfy the convexity requirement.

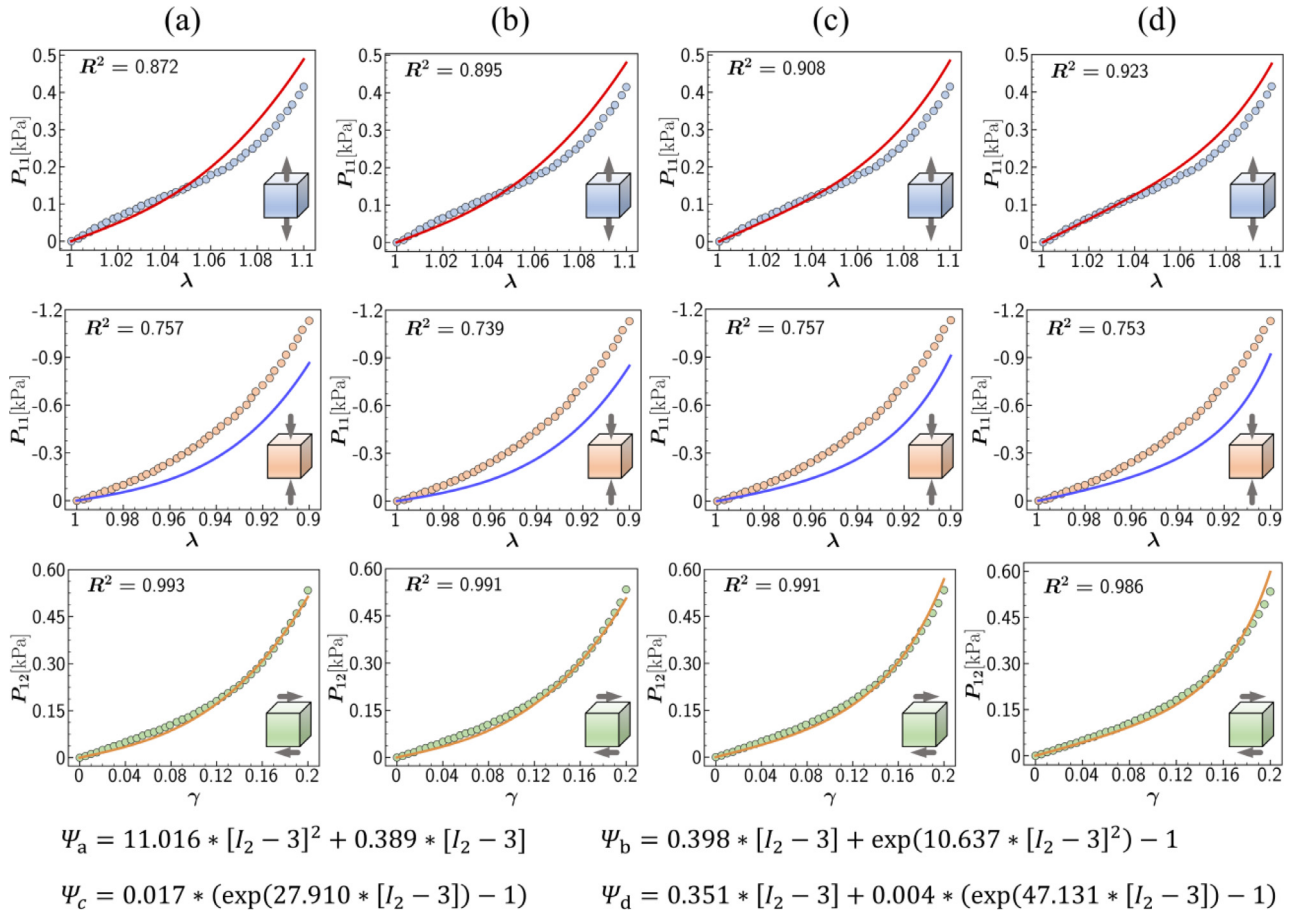


Fig. 3. Four distinct hyperelastic models discovered with invariant-based algorithm. Models are trained simultaneously with data from three loading modes, and tested with tension, compression and shear data individually. Dots illustrate the experimental data of the human brain cortex. R^2 indicates the goodness of fit. Mathematical expressions for each strain energy function are provided at the bottom of the figure.

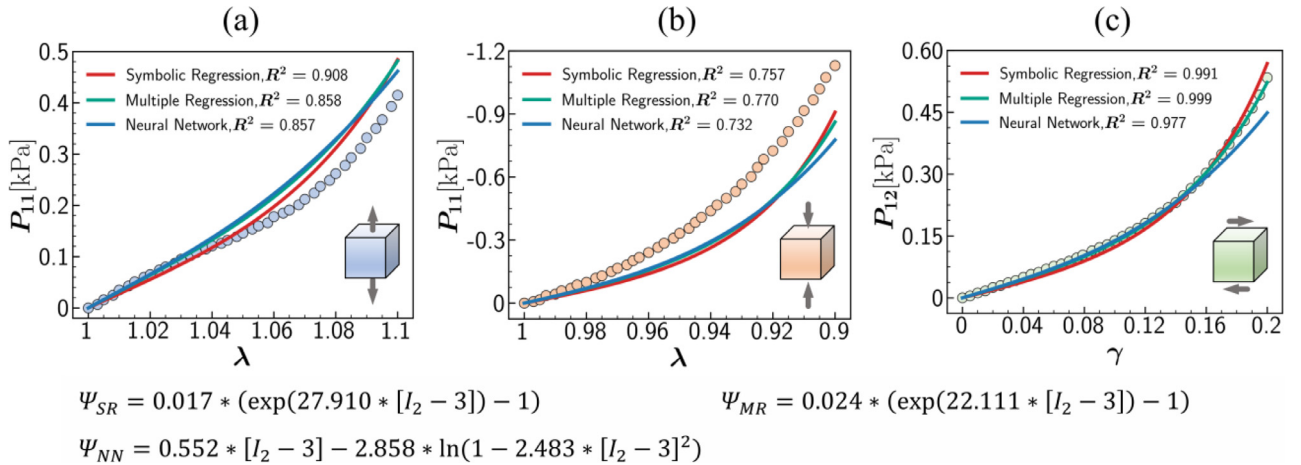


Fig. 4. Symbolic regression vs Multiple regression and artificial neural network. Comparison on fitting performance of invariant-based hyperelastic models derived from symbolic regression (SR), multiple regression (MR), and neural network (NN). Models are trained simultaneously with data from three loading modes, and tested with tension (a), compression (b), and shear data (c), individually. Dots illustrate the experimental data of the human brain cortex. R^2 indicates the goodness of fit. Mathematical expressions for each strain energy function are provided at the bottom of the figure.

To substantiate this claim, we illustrate the contours of the identified strain energy functions with respect to principal stretches λ_1 and λ_2 , as depicted in Figs. 5 and 6. Despite incorporating distinct functional operators, the contour lines of these strain energy functions consistently display elliptic shapes encircling the center point, representing the stress-free reference state ($\lambda_1 = 1$, $\lambda_2 = 1$), particularly evident in Fig. 5. This observation signifies the rigor-

ous fulfillment of convexity for the strain energy function within the training data regime. Furthermore, the convexity is preserved beyond the training range, as shown in Fig. 6. It is noteworthy that the different deformation ranges illustrated in Fig. 6 are chosen for enhanced visualization. The fulfillment of convexity is more straightforwardly illustrated in the 3D surface plot of the strain energy function, as shown in Figure S3. Ultimately, the physical

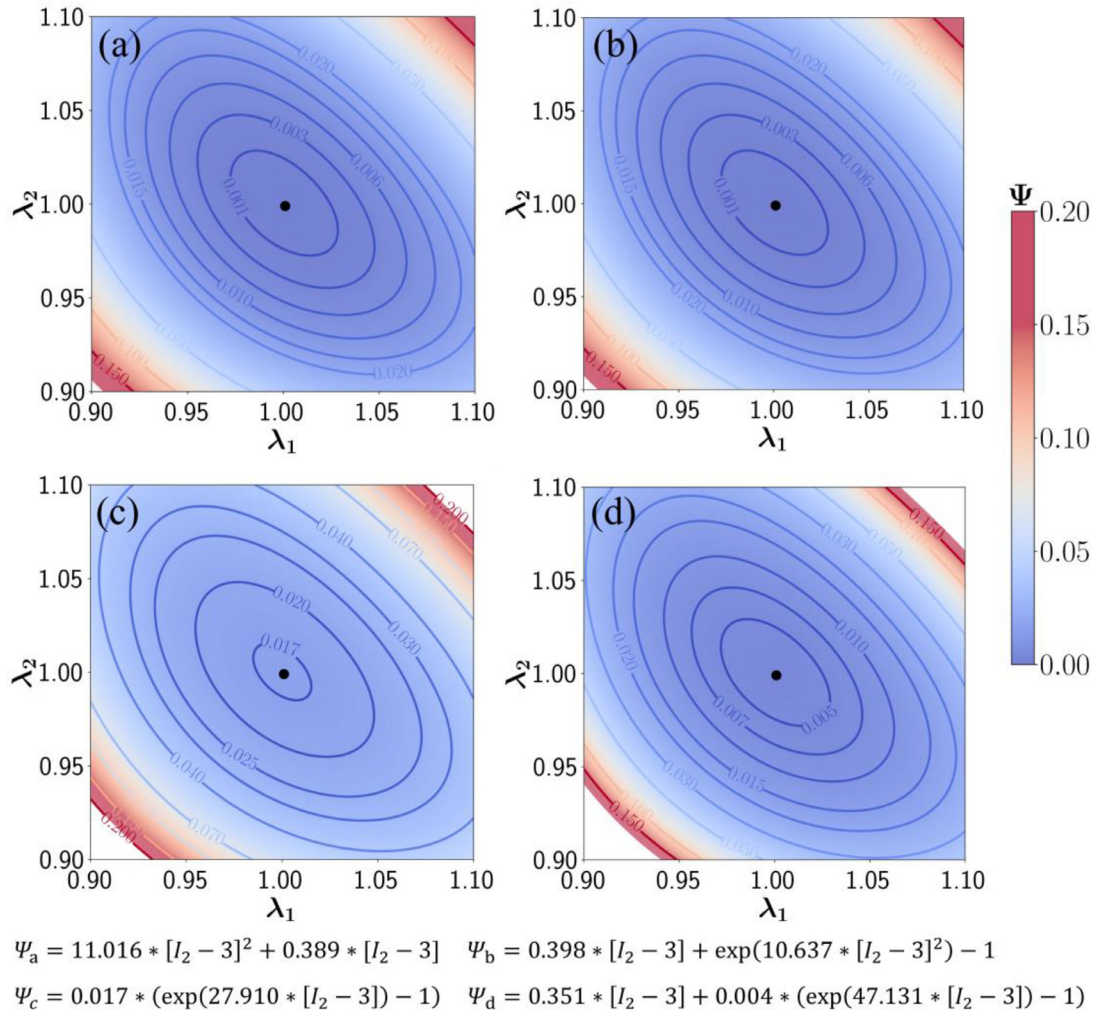


Fig. 5. Convexity of four invariant-based hyperelastic models within the training data regime. Contours of the four strain energy functions discovered by *Invariant-based Symbolic Regression* algorithm within the training data regime, (a) Ψ_a , (b) Ψ_b , (c) Ψ_c , (d) Ψ_d . Black dots represent the location of global minimum. Mathematical expressions for each strain energy function are provided at the bottom of the figure.

validity regarding the polyconvexity condition have been clearly demonstrated for the four hyperelastic models discovered by the *Invariant-based Symbolic Regression* algorithm.

3.3. Hyperelastic models identified by the stretch-based symbolic regression algorithm

Fig. 7 shows the four superior hyperelastic models discovered for the human brain cortex, using the *Stretch-based Symbolic Regression* algorithm. All four predicted models exhibit promising fitting accuracy in describing material behaviors within uniaxial tension, uniaxial compression, and simple shear scenarios, each achieving an R^2 value greater than 0.9 for all loading modes. In contrast to the performance of invariant-based models that show limitations in uniaxial tension and compression, the stretch-based hyperelastic models significantly outperform them in describing uniaxial deformations, particularly in compressions. All stretch-based models demonstrate perfect alignment with the compressive data. Conversely, the invariant-based models consistently underestimate compressive forces throughout deformation, leading to R^2 values all below 0.8, as evident in Fig. 3. This observation is consistent with recent findings [58]. The superior performance of stretch-based hyperelastic models is particularly pronounced in nonlinear stages, such as in regions where tension or compression

exceeds 5 %. This demonstrates their effectiveness in capturing the inherent nonlinearity of the experimental data [61,62].

Similar efforts have been made using stretch-based approaches, particularly the generalized Ogden model, to characterize the material behavior of the human brain cortex [1,58]. The one-term Ogden model in Ref. [1] shares an identical expression with our first model (Ψ_a in Fig. 7). The generalized Ogden models [58] identified by the principal stretch-based CANN exhibit lower accuracy compared to our predicted models (Ψ_b , Ψ_c , Ψ_d), achieving R^2 values of 0.938, 0.985, and 0.987 for tension, compression, and shear, respectively. Notably, their model includes a significantly greater number of terms compared to ours [58], indicating that we achieved higher accuracy with a more succinct model form. This disparity is primarily attributed to the smaller function space they employed, where the polynomial order is restricted to the range of $(-30, 10)$. In contrast, this value is searched from a wider range of $(-30, 30)$ in our study.

Among the four models presented in Fig. 7, the last two models, $\Psi_c = 0.0035 \sum_i (\lambda_i^{-24} - 1) + 0.0003 \sum_i (\lambda_i^{30} - 1)$ and $\Psi_d = 0.004 \sum_i (\lambda_i^{-23} - 1) + 0.0003 \sum_i (\lambda_i^{29} - 1)$, demonstrate comparably high accuracy in fitting the multi-mode data. As such, these models serve as the optimal outcomes derived from the *Stretch-based Symbolic Regression* algorithm. Unlike the invariant-based

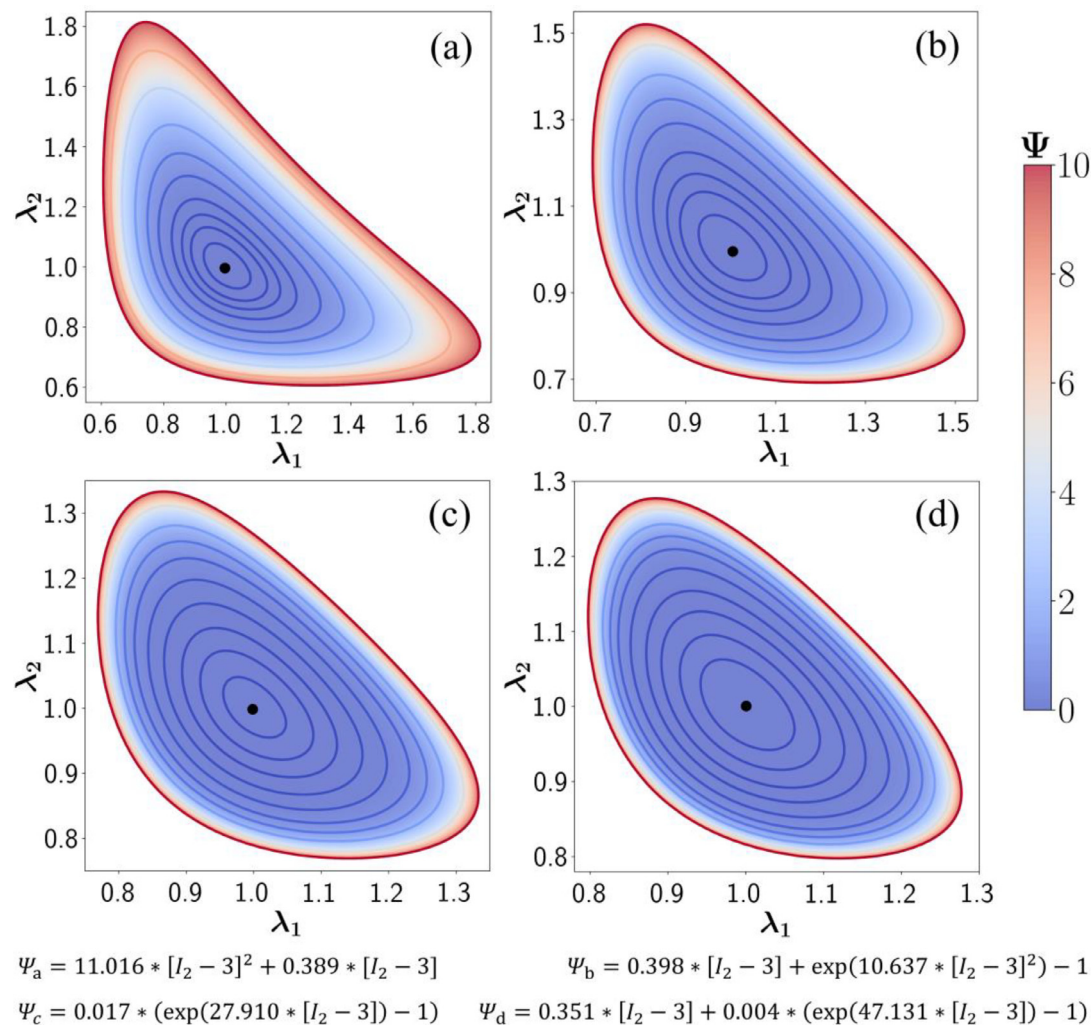


Fig. 6. Convexity of four invariant-based hyperelastic models beyond the training data regime. Contours of the four strain energy functions discovered by *Invariant-based Symbolic Regression* algorithm beyond the training data regime, (a) Ψ_a , (b) Ψ_b , (c) Ψ_c , (d) Ψ_d . Black dots represent the location of global minimum. Mathematical expressions for each strain energy function are provided at the bottom of the figure.

hyperelastic models, which enforce strict polyconvexity on the output function, stretch-based may exhibit non-convexity due to the random combinations of polynomial series with arbitrary orders [47]. Hence, we checked the ellipticity along the principal directions for each model derived from the *Stretch-based Symbolic Regression* algorithm by assessing the positive definiteness of the Hessian matrix, as defined in Eq. (7). The results are pre-

sented in Table 1. The positive determinants and real eigenvalues of the Hessian matrix confirm the positive definiteness of tangent ($\partial^2 \Psi / \partial \lambda_i^2$), thereby validating the local ellipticity of the strain energy functions within the training regime.

Representations of the convexity of all four stretch-based hyperelastic models are present in Figs. 8 and 9. In Fig. 8, the contour lines all exhibit near-elliptic shapes encircling the center points

Table 1					
Ellipticity checks for stretch-based hyperelastic models. $\partial^2 \Psi / \partial \lambda_i^2$ are the second derivatives of Ψ with respect to principal stretch λ_i , with i in 1, 2, 3; $\min(\det[H])$ is the minimal determinant of the Hessian matrix $[H]$; $\min(\partial^2 \Psi / \partial \lambda_i^2)$ denotes the minimal value of the i th eigenvalue of $[H]$. Ψ_a , Ψ_b , Ψ_c , Ψ_d correspond to the four stretch-based hyperelastic models shown in Fig. 7. Experimental data of the human brain cortex are employed for calculations. Here, ellipticity along the principal directions was assessed.					
Ellipticity Checks	Ψ_a	Ψ_b	Ψ_c	Ψ_d	
$\partial^2 \Psi / \partial \lambda_i^2$	$2.998 \lambda_i^{-21}$	$1.751 \lambda_i^{-27}$	$0.274 \lambda_i^{28} + 2.081 \lambda_i^{-26}$	$0.260 \lambda_i^{27} + 2.2051 \lambda_i^{-25}$	
$\min(\det[H])$	26.951	5.369	13.074	14.983	
$\min(\partial^2 \Psi / \partial \lambda_1^2)$	0.405	0.134	1.569	1.577	
$\min(\partial^2 \Psi / \partial \lambda_2^2)$	0.405	0.134	1.569	1.577	
$\min(\partial^2 \Psi / \partial \lambda_3^2)$	0.036	0.006	1.568	1.576	
? Positive definite	Yes	Yes	Yes	Yes	
? Ellipticity of Ψ	Yes	Yes	Yes	Yes	

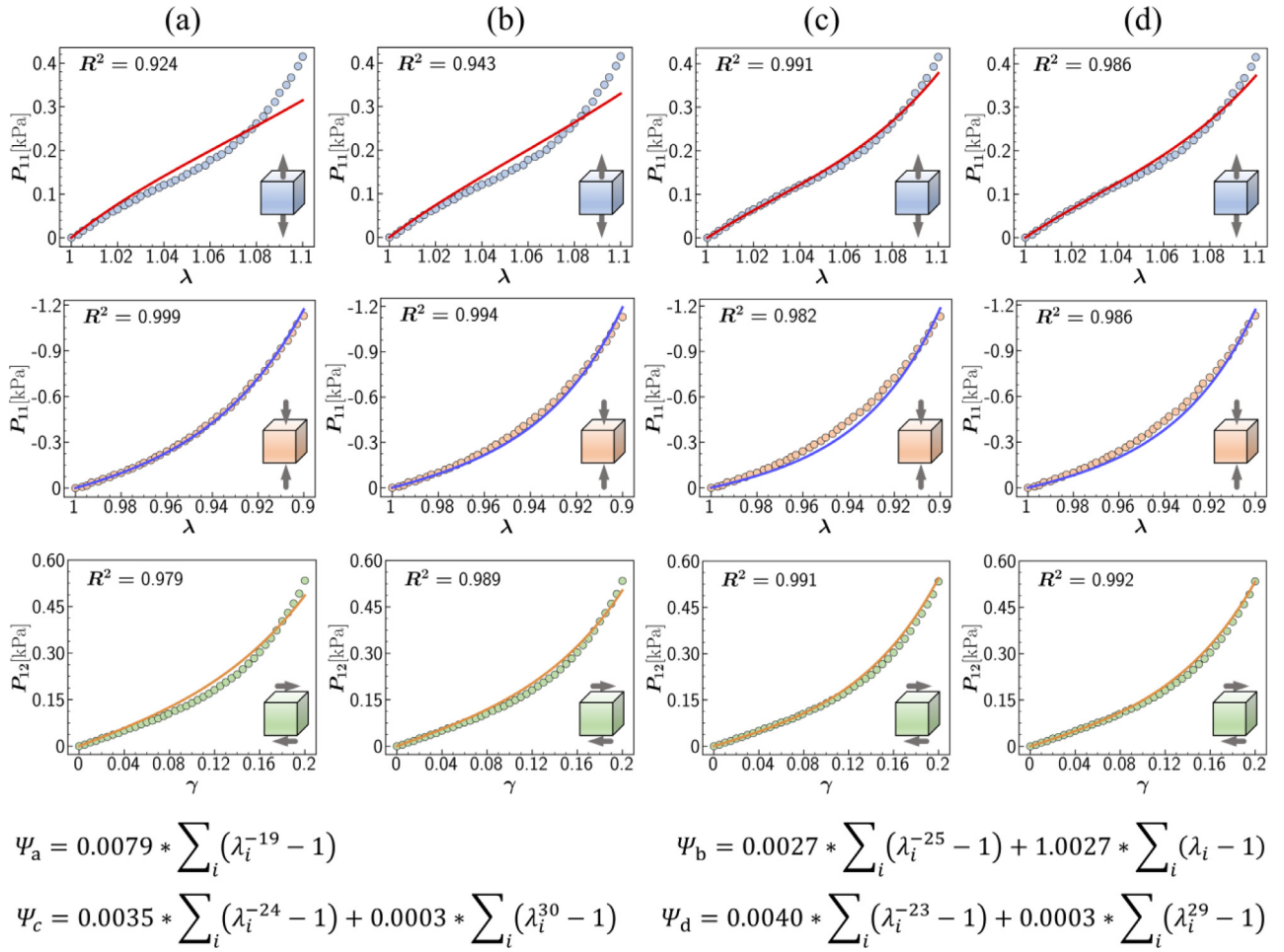


Fig. 7. Four distinct Ogden-form hyperelastic models discovered with stretch-based algorithm. Models are trained simultaneously with data from three loading modes, and tested with tension, compression, and shear data individually. Dots illustrate the experimental data of the human brain cortex. R^2 indicates the goodness of fit. Mathematical expressions for each strain energy function are provided at the bottom of the figure.

that correspond to the stress-free reference state ($\lambda_1 = 1$, $\lambda_2 = 1$). This observation confirms that the predicted strain energy functions satisfy the convexity requirement within the training data. Furthermore, we expanded the synthetic stretches ranges to examine whether the convexity is preserved beyond the training range. Results are shown in Fig. 9, where different deformation ranges are selected for enhanced visualization. Though all models present concave shapes with global minimum locating at the center and higher value at the outer boundaries, the strict adherence to convexity requirements is not maintained, particularly for Ψ_a and Ψ_b . In Fig. 9a, the contour lines resemble a right triangle with the hypotenuse posing a significant non-convex shape. This issue is also evident for Ψ_b in Fig. 9b, suggesting a potential loss of convexity for Ψ_a and Ψ_b under large deformations. Similar observations were also reported in investigations of the one-term Ogden model [47]. Nonetheless, the remaining two models (Ψ_c and Ψ_d) still preserve their convexity beyond the training data regime, as illustrated in Fig. 9c and d. A clearer depiction of the convexity fulfillment is shown in Figure S4, where we generated 3D surface plots to illustrate the strain energy function with respect to principal stretches λ_1 and λ_2 . These plots effectively illustrate the convexity behaviors of hyperelastic models discovered by the *Stretch-based Symbolic Regression* algorithm. However, it is important to note that there may be potential convexity loss in certain scenarios, as observed in Ψ_a and Ψ_b under large deformations.

3.4. Hyperelastic models predicted by the strain-based symbolic regression algorithm

Unlike the *Invariant-based Symbolic Regression* or principal *Stretch-based Symbolic Regression* algorithms, the *Strain-based Symbolic Regression* algorithm identified a single, unique model,

$$\Psi = \sum_i (2820.76\epsilon_i^6 + 43.27\epsilon_i^4 - 13.72\epsilon_i^3 + 1.37\epsilon_i^2). \quad (22)$$

Here, ϵ_i represents the normal strain and is related to principal stretch by $\epsilon_i = \lambda_i - 1$. The fitting performance of this strain-based model is shown in Fig. 10. As illustrated, the strain-based model exhibits remarkable fitting accuracy in capturing material behaviors in uniaxial tension, uniaxial compression, and simple shear scenarios, achieving R^2 values greater than 0.99.

Fig. 11 depicts a comparison among the optimal models predicted by *Invariant-based Symbolic Regression*, *Stretch-based Symbolic Regression*, and *Strain-based Symbolic Regression* algorithms, respectively. As seen, the strain-based model exhibits the highest fitting accuracy with experimental data across all three loading modes. This promising fitting behavior pertaining to the strain-based model was also reported in our recently published paper [53], where the multiple regression method, combined with the Akaike's information criteria (AIC), was used to identify the optimal hyperelastic model within a confined polynomial space, with an order range of (0,10). Interestingly, the polynomial series

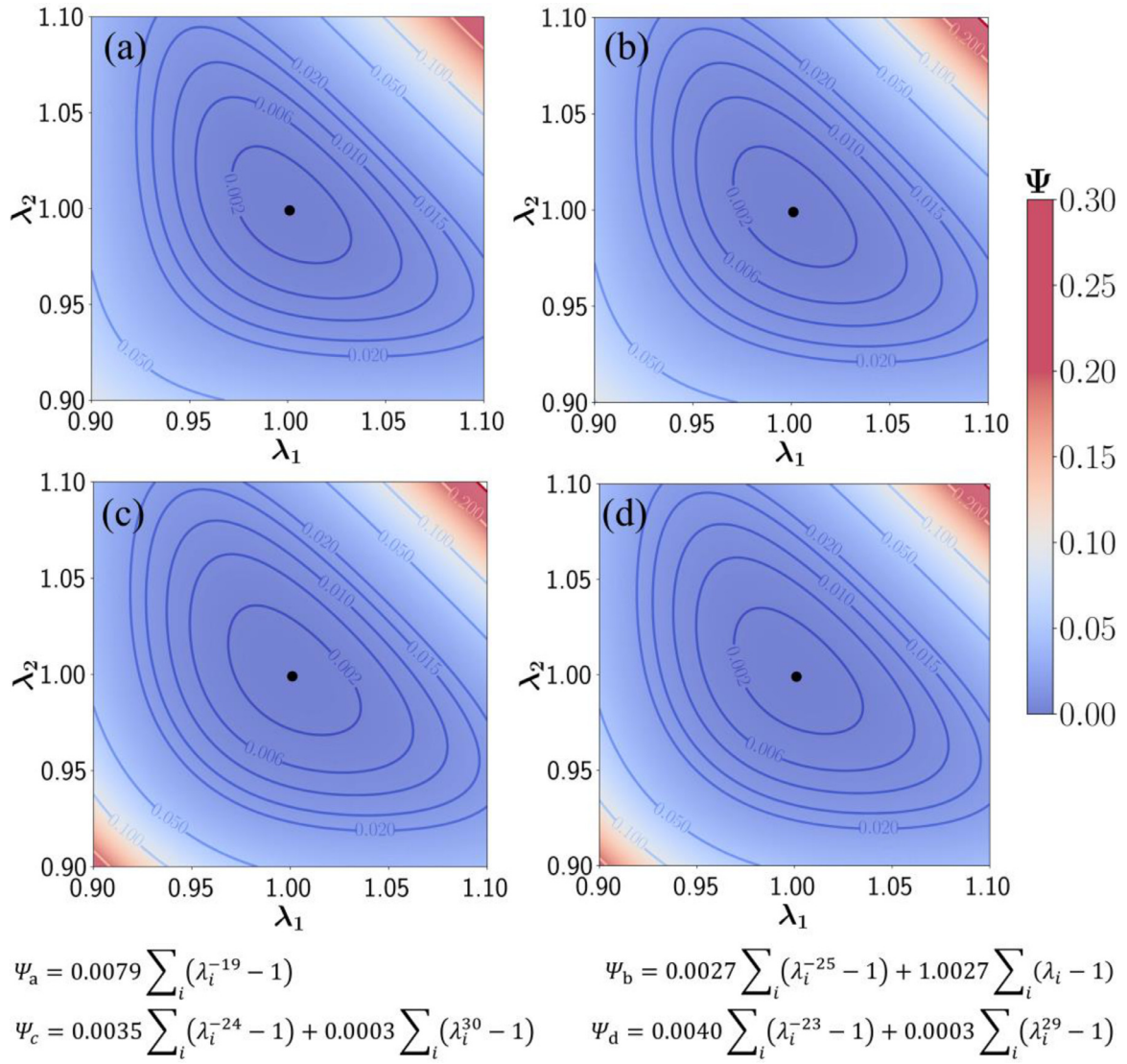


Fig. 8. Convexity of the four stretch-based hyperelastic models within the training data regime. Contours of the four strain energy functions discovered by *Stretch-based Symbolic Regression* algorithm within the training data regime, (a) Ψ_a , (b) Ψ_b , (c) Ψ_c , (d) Ψ_d . Black dots represent the location of global minimum. Mathematical expressions for each strain energy function are provided at the bottom of the figure.

selected by multiple regression model is identical to those identified by our symbolic regression model (ϵ_i^6 , ϵ_i^4 , ϵ_i^3 , ϵ_i^2), though the coefficients differ. Notably, the polynomial basis available for symbolic regression algorithm are confined within $(-30, 30)$, indicating that the polynomial set discovered by multiple regression is indeed optimal for the given range. It is noteworthy that though multiple regression demonstrates satisfactory capability in model discovery, substantial human efforts and computational costs are required, especially when dealing with a vast functional space. For example, when the polynomial order is confined within $(-30, 30)$, the potential equation combinations amount to 2^{60} , making it impractical to traverse completely using multiple regression. This underscores the advantages of evolution algorithms over traditional regression methods like multiple regression in functions searching and parameter optimization [63,64].

Analogously, the ellipticity of the strain-based model along principal directions was evaluated by examining the positive-definiteness of the Hessian matrix. Within the training data regime, the minimal determinant of the Hessian matrix is 3.158, while the minimal values for the three eigenvalues ($\partial^2 \Psi / \partial \lambda_1^2$,

$\partial^2 \Psi / \partial \lambda_2^2$, $\partial^2 \Psi / \partial \lambda_3^2$) are 0.422, 0.422, and 0.420, respectively. The positive determinant and real eigenvalues guarantee the positive definiteness of the Hessian matrix, which further confirms the local ellipticity of the identified strain energy function within the training regime. Furthermore, the consistency condition (positivity of shear modulus), as defined in Eq. (20), is also satisfied, as the coefficient before the second order term (ϵ_i^2) is positive (1.37).

A more straightforward illustration of convexity is provided by the 3D surface plots and contour plots, as illustrated in Fig. 12. Within the incompressibility framework, the strain energy function exhibits a distinct concave shape, with its minimum occurring at the reference state ($\lambda_1 = \lambda_2 = 1$), as seen in Fig. 12a and b. Additionally, the coercivity condition is also satisfied, as the strain energy function continues to achieve its local maximum with increasing tensile or compressive deformations. However, the peak strain energy function under compression is significantly higher than its tensile counterpart under the same level of deformation. The dissimilar behaviors under tension and compression potentially explains the prominence of strain-based models in capturing the data nonlinearity, as shown in Fig. 11. If

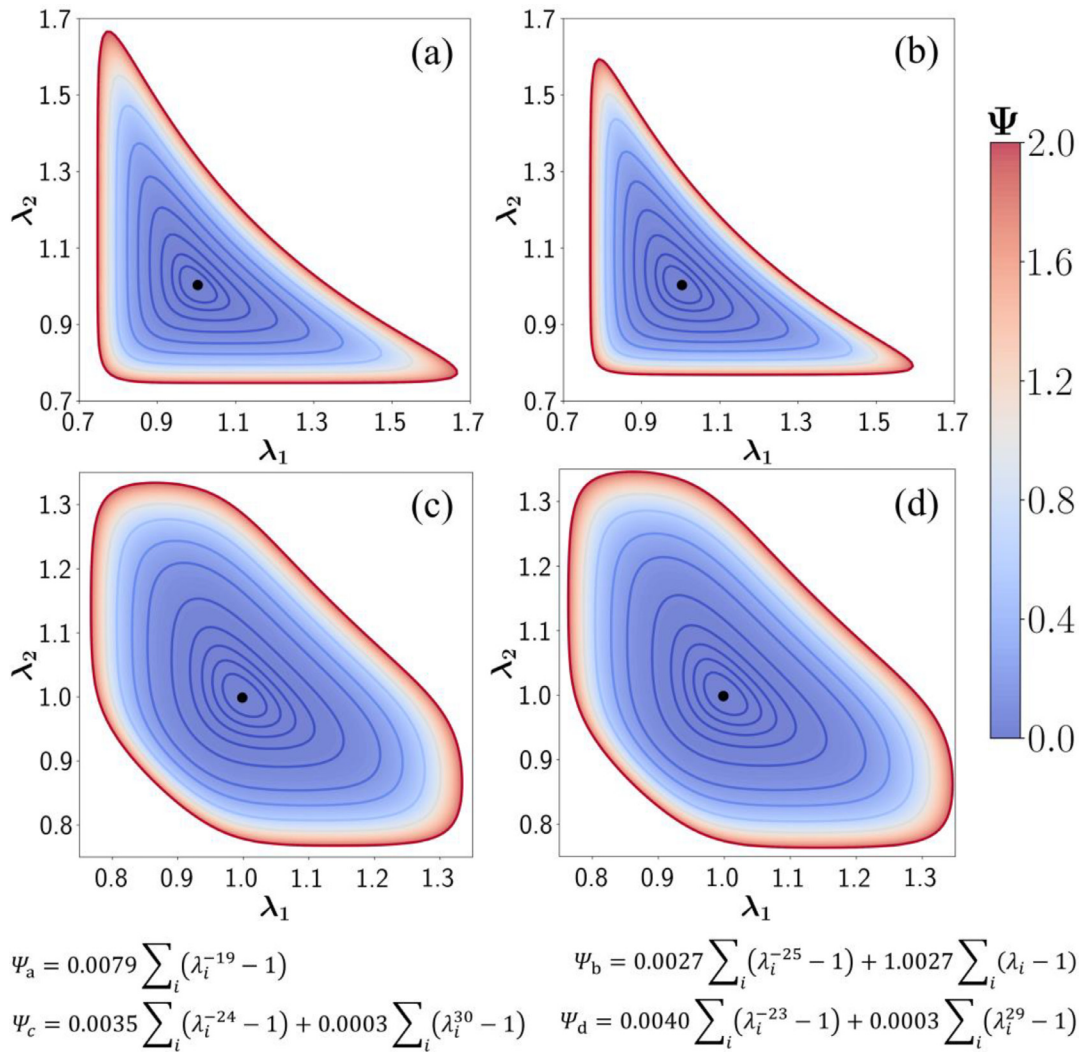


Fig. 9. Convexity of the four stretch-based hyperelastic models beyond the training data regime. Contours of the four strain energy functions discovered by *Stretch-based Symbolic Regression* algorithm beyond the training data regime, (a) Ψ_a , (b) Ψ_b , (c) Ψ_c , (d) Ψ_d . Black dots represent the location of global minimum. Mathematical expressions for each strain energy function are provided at the bottom of the figure.

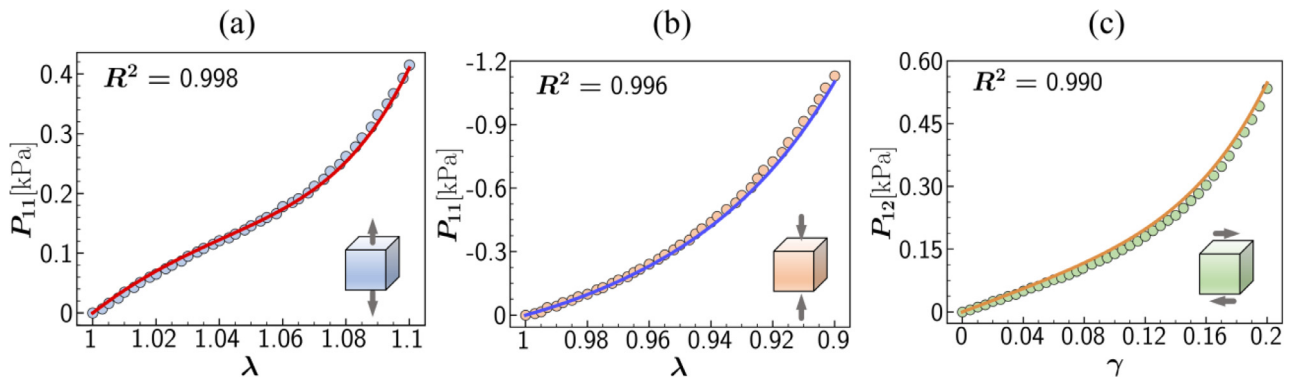


Fig. 10. Hyperelastic model discovered with strain-based algorithm. Models are trained simultaneously with data from three loading modes, and tested with tension (a), compression (b) and, shear (c) data individually. Dots illustrate the experimental data of the human brain cortex. R^2 indicates the goodness of fit.

we relax the compressibility constraint, the strain energy demonstrates perfect symmetry with respect to λ_1 and λ_2 , as shown in Fig. 12c and d.

To assess the convexity beyond the training data regime, we expanded the synthetic stretches ranges to 60 % tension and compression and generated a contour plot, as presented in Fig. 13.

Noted, the incompressibility constraint is imposed in this situation. From the figure, it is evident that the convexity observed within the training regime (Fig. 13a) is maintained beyond the training dataset (Fig. 13b), even in the case with 60 % deformation. This result is intriguing because, in the Strain-based Symbolic Regression algorithm, we applied more relaxed restrictions on the func-

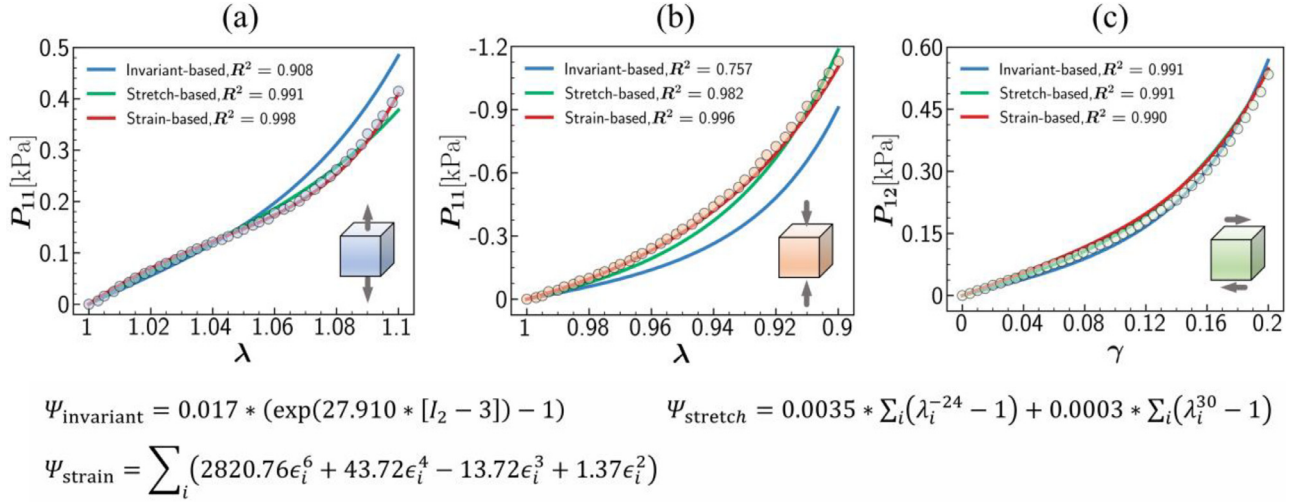


Fig. 11. Symbolic regression with invariant-based algorithm vs stretch-based algorithm and strain-based algorithm. Comparison on fitting performance of optimal hyperelastic models derived from symbolic regression using three distinct algorithms. Models are trained simultaneously with data from three loading modes, and tested with tension (a), compression (b), and shear (c) data individually. Dots illustrate the experimental data of the human brain cortex. R^2 indicates the goodness of fit. Mathematical expressions for each strain energy function are provided at the bottom of the figure.

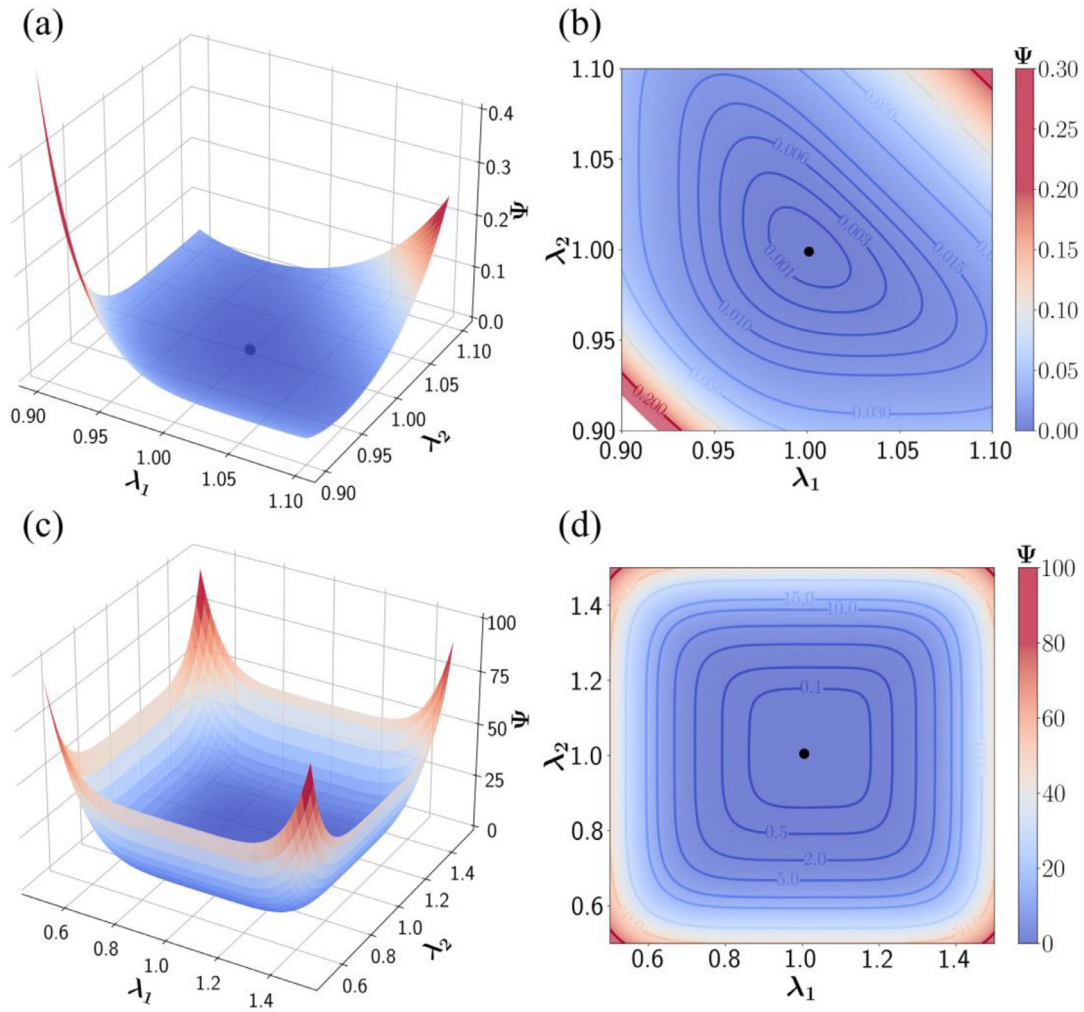


Fig. 12. Convexity of the hyperelastic model derived from strain-based algorithm. (a), 3D visualization of the strain energy function Ψ with respect to principal stretch λ_1 and λ_2 , ($\lambda_3 = 1/(\lambda_1\lambda_2)$) within the training dataset of the human brain cortex; (b), Contour of the strain energy function in (a); (c), 3D visualization of the strain energy function Ψ with respect to principal stretch λ_1 and λ_2 , ($\lambda_3 = 1$) from synthetic data; (d), Contour of the strain energy function in (c). Black dots represent the location of global minimum.

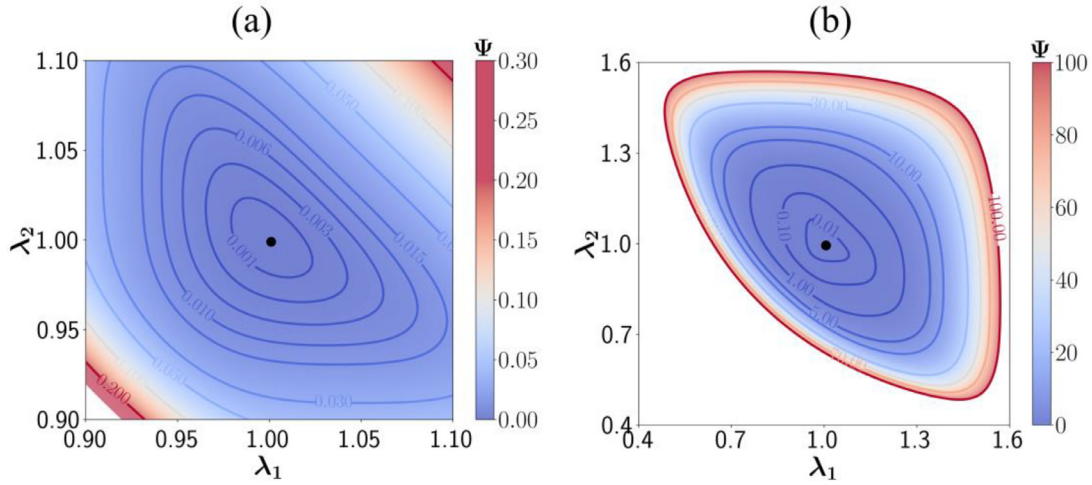


Fig. 13. Convexity of the strain-based hyperelastic model within and beyond the training data regime. Contours of the strain energy function discovered by *Strain-based Symbolic Regression* algorithm within (a) and beyond (b) the training data regime. Black dots represent the location of global minimum.

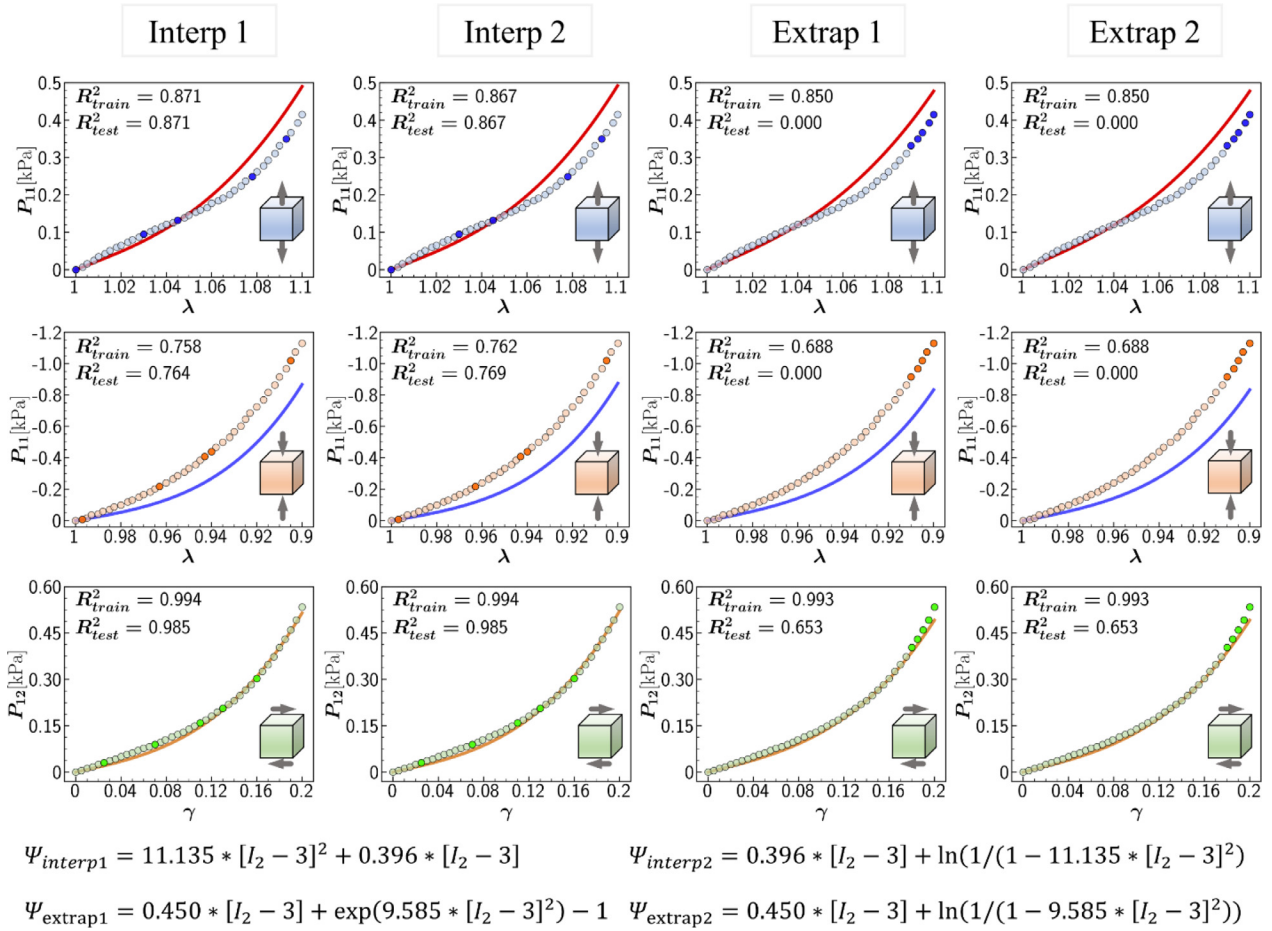


Fig. 14. Interpolation and extrapolation capabilities of invariant-based model. Four distinct hyperelastic models discovered with invariant-based symbolic regression algorithm. From left to right, the first two models are present to evaluate models' predictive capability regarding interpolation (interp), while the last two models are for extrapolation (extrap). All models are trained with 90 % of the multi-mode data, as indicated by dots with light color, and tested with remaining 10 % dataset, as indicated by dots with dark color. R^2 indicates the goodness of fitting. Mathematical expressions for each strain energy function are provided at the bottom of the figure.

tion format, allowing both coefficients and polynomial orders to be unrestricted, which led to negative coefficients as seen in Eq. (22). Despite this, the strain-based energy function maintains rigorous ellipticity and preserves convexity even under large deformations, whereas the stretch-based energy functions may lose convexity, as illustrated in Fig. 9. Moreover, the strain-based functions $f(\epsilon_i)$ can

be expressed as stretch-based polynomials $g(\lambda_i)$ using the relations $\epsilon_i = \lambda_i - 1$. This suggests that the constraints enforcing positive coefficients in the stretch-based symbolic regression algorithm may be overly restrictive. A less stringent prior confinement, combined with rigorous posterior checks, could potentially yield more satisfactory results.

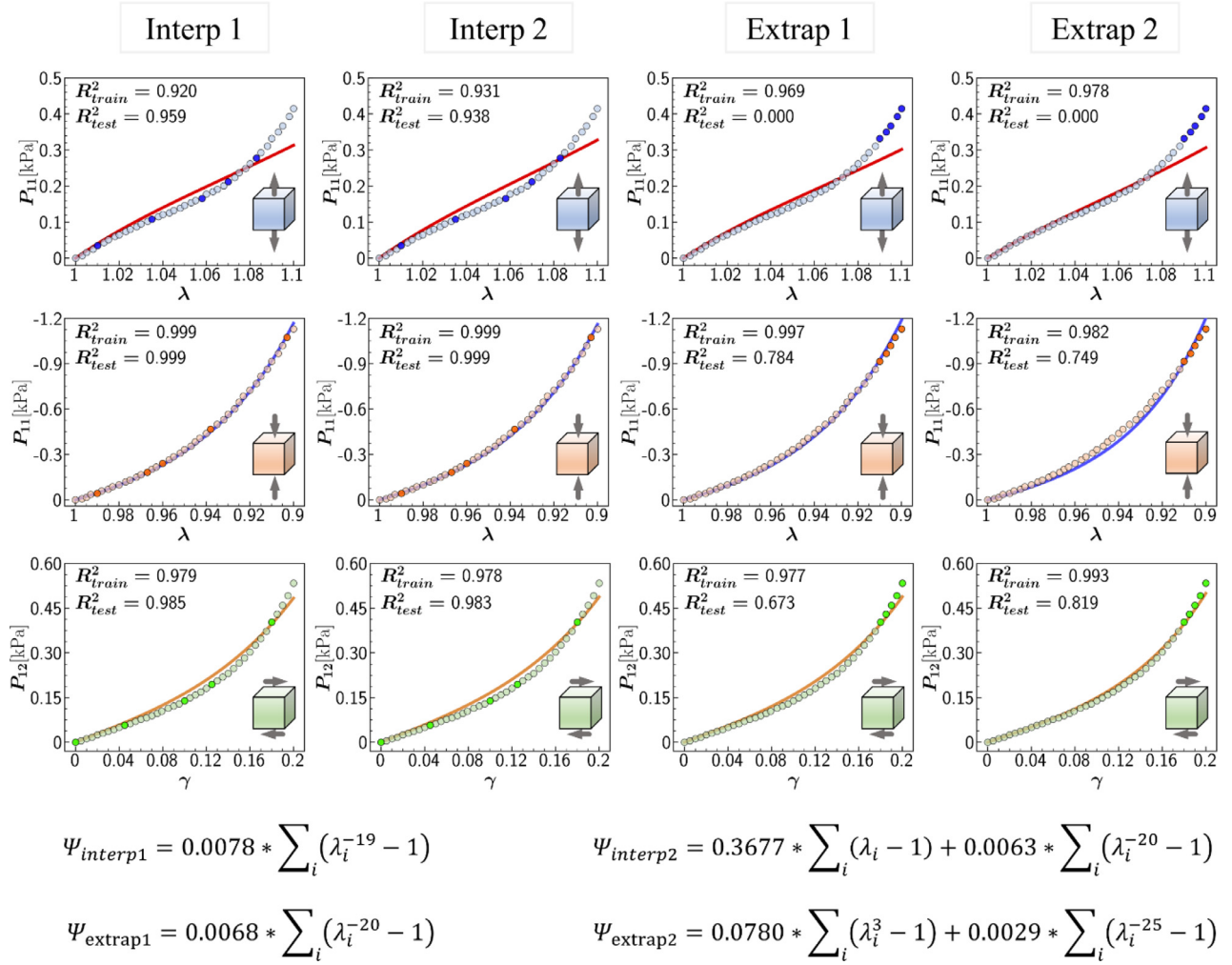


Fig. 15. Interpolation and extrapolation capabilities of stretch-based model. Four distinct hyperelastic models discovered with stretch-based symbolic regression algorithm. From left to right, the first two models are present to evaluate models' predictive capability regarding interpolation (interp), while the last two models are for extrapolation (extrap). All models are trained with 90 % of the multi-mode data, as indicated by dots with light color, and tested with remaining 10 % dataset, as indicated by dots with dark color. R^2 indicates the goodness of fitting. Mathematical expressions for each strain energy function are provided at the bottom of the figure.

3.5. Interpolation and extrapolation capabilities of symbolic regression algorithms

In this section, we evaluated the interpolation and extrapolation capabilities of our symbolic regression algorithms for hyperelastic model discovery. Given the critical role of data in model training and the typical scarcity of experimental data, it is imperative to address various challenges such as the absence of diverse loading modes due to limited experimental apparatus. As depicted in Figures S5–S7, insights from single-mode training endeavors provide guidance on optimal loading modes for model characterization. For example, simple shear loading is favorable for characterizing models using the invariant-based approach, whereas uniaxial compression is suggested for stretch-based or strain-based methods. Another type of data scarcity involves oversized intervals in data sampling or a limited range in testing apparatus. The former may result in potential oversights regarding intermediate data information [65], while the latter may could miss capturing data features during large deformations [66]. Therefore, we also investigated the interpolation and extrapolation capabilities of our algorithms. In both cases, 90 % of the data were allocated to the training dataset, with the remaining 10 % designated for testing. Specifically, to examine the interpolation capabilities, we divided the full-field training data into five subintervals for each training mode and randomly

selected one data point from each interval to construct the testing dataset, using the remaining data for training. For assessing extrapolation capabilities, we constructed the training dataset from the bottom 90 % of stretch-stress paired data (sorted by increasing stretch), with the remaining 10 % serving as the testing dataset.

Fig. 14 illustrates the fitting accuracy of invariant-based hyperelastic models trained with 90 % of the entire dataset, indicated by light colors. Two candidate models are presented to evaluate the predictive capabilities regarding interpolation and extrapolation (referred to as “interpolated models” and “extrapolated models” in the following context unless otherwise noted). As shown in the figure, despite being trained on different dataset, both interpolated and extrapolated models demonstrate consistent predictive trends: overestimating tension data, underestimating compression data, and aligning well with shear data. These trends are also observed in models trained with full-field data, as seen in Figs. 3 and 4. This consistency may indicate the constrained capability of invariant-based models in capturing the significant non-linearity occurring within the large deformation range. Additionally, the interpolated models demonstrate comparable training and testing accuracy for each loading mode, while the extrapolated models tend to fairly predict the testing data, especially for tension and compression. This suggests that our algorithms possess superior interpolation capabilities compared to extrapolation capabilities. In-

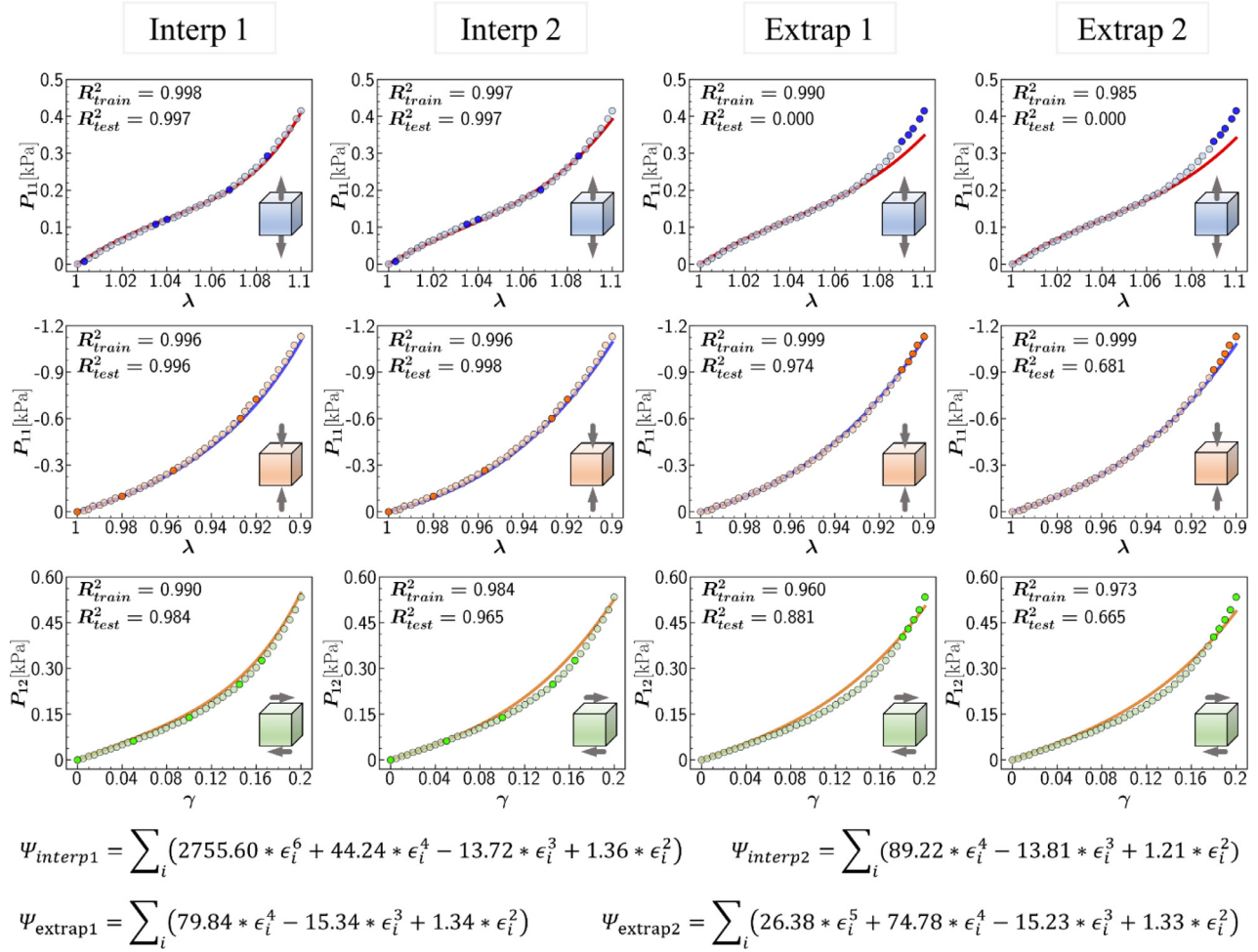


Fig. 16. Interpolation and extrapolation capabilities of strain-based model. Four distinct hyperelastic models discovered with strain-based symbolic regression algorithm. From left to right, the first two models are present to evaluate models' predictive capability regarding interpolation (interp), while the last two models are for extrapolation (extrap). All models are trained with 90 % of the multi-mode data, as indicated by dots with light color, and tested with remaining 10 % dataset, as indicated by dots with dark color. R^2 indicates the goodness of fitting. Mathematical expressions for each strain energy function are provided at the bottom of the figure.

triguously, one of the interpolating models ($\Psi_{interp1}$) and one of the extrapolating models ($\Psi_{extrap1}$) share the exact same format as the models trained with full-field data (Ψ_a and Ψ_b in Fig. 3), differing only in the constants. This potentially validates the robustness of our algorithms.

Fig. 15 illustrates the fitting accuracy of stretch-based hyperelastic models trained with 90 % of the entire dataset, indicated by light colors. Analogously, both interpolated and extrapolated models exhibit consistent predictive trends. Unlike the invariant-based models, all stretch-based models achieve promising fitting accuracy in predicting compression and shear data. Even for tension, the models still maintain satisfying performances until stretches exceed 8 %. Furthermore, the interpolated models show comparable training and testing accuracy for each loading mode, while the extrapolated models moderately predict the testing dataset for compression and shear. However, extrapolated stretch-based models still fail to predict the tension within the testing data regime. Similarly, the strain-based models demonstrate satisfying performance comparable to the stretch-based models, as shown in Fig. 16. The strain-based models also accurately predict tension data, especially for the two interpolated models. Despite this enhanced performance, the extrapolated strain-based models still fail to predict tension within the testing data regime. Interestingly, the second interpolated model ($\Psi_{interp2}$) exhibits comparable performance as the model trained with full-field data (see Eq. (22) and

Fig. 10), using only three terms. This suggests a more favorable model selection, considering the tradeoff between model's accuracy and complexity.

3.6. Robustness evaluation for symbolic regression algorithms

In this section, we investigated the robustness of our symbolic regression algorithms for hyperelastic model discovery. All evaluations were performed on synthetic data generated from predefined hyperelastic models, similar in form to our identified models: $\Psi = 0.0170 * (\exp(27.9100 * [I_2 - 3]) - 1)$ for the *Invariant-based Symbolic Regression* algorithm, $\Psi = 0.0079 * \sum_i (\lambda_i^{-19} - 1)$ for the *Stretch-based Symbolic Regression* algorithm, and $\Psi = \sum_i (43.2700 * \epsilon_i^4 + 1.3700 * \epsilon_i^2)$ for the *Strain-based Symbolic Regression* algorithm. Artificial Gaussian noises were incorporated into the synthetic data to mimic the perturbations encountered in real experiments,

$$P_{i,k}^{test} = P_{i,k}^{synthetic} + P_{i,k}^{noise},$$

$$P_{i,k}^{noise} \sim \mathcal{N}(0, \sigma_k) \quad \forall i \in \{1, \dots, n_{data}\}, k \in \{ut, uc, ss\}, \quad (23)$$

where k represents loading modes: uniaxial tension (ut), uniaxial compression (uc), and simple shear (ss); $P_{i,k}^{synthetic}$ means the ith

Table 2

Robustness test for the invariant-based algorithm. Effects of prescribed noise on symbolic regression predictions. The target strain energy function is represented as $\Psi = 0.0170 * (\exp(27.9100 * [I_2 - 3]) - 1)$. A prediction is deemed right when the mathematic format of the prediction model coincides with the target model.

Noise, σ	MSE	Predicted Ψ	? Right Prediction
0	1.50×10^{-11}	$0.0170 * (\exp(27.9100 * [I_2 - 3]) - 1)$	Yes
0.0001	2.20×10^{-10}	$0.0170 * (\exp(27.9085 * [I_2 - 3]) - 1)$	Yes
0.001	1.02×10^{-8}	$0.0170 * (\exp(27.9063 * [I_2 - 3]) - 1)$	Yes
0.002	1.51×10^{-6}	$0.0171 * (\exp(27.8814 * [I_2 - 3]) - 1)$	Yes
0.005	1.15×10^{-5}	$0.0170 * (\exp(27.8227 * [I_2 - 3]) - 1)$	Yes
0.01	3.99×10^{-5}	$0.0170 * (\exp(27.8616 * [I_2 - 3]) - 1)$	Yes
0.02	2.04×10^{-4}	$0.0168 * (\exp(28.0941 * [I_2 - 3]) - 1)$	Yes
0.05	1.05×10^{-3}	$0.0168 * (\exp(28.1500 * [I_2 - 3]) - 1)$	Yes
0.1	4.79×10^{-3}	$0.0227 * (\exp(23.4639 * [I_2 - 3]) - 1)$	Yes
0.2	2.10×10^{-2}	$[I_2 - 3]$	No

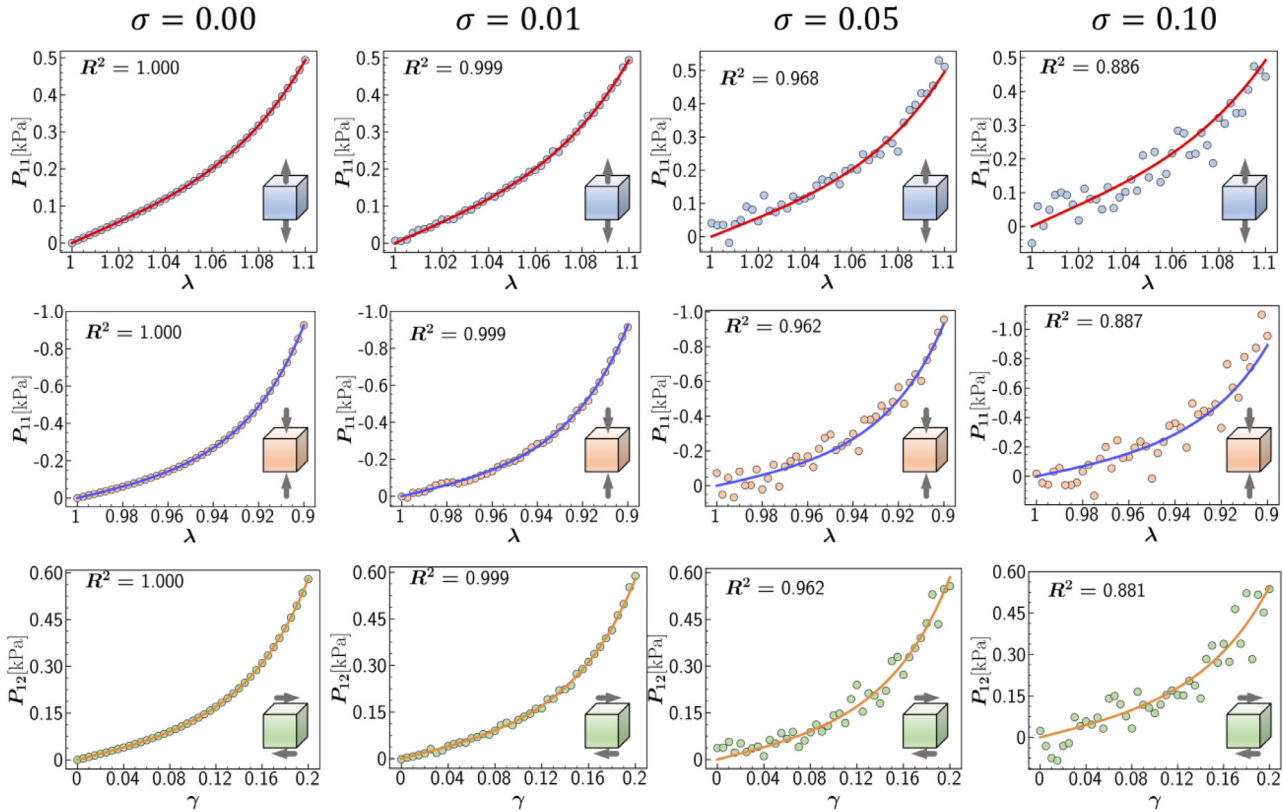


Fig. 17. Robustness test for invariant-based algorithm. Effects of prescribed noise on the predictions of symbolic regression. Models are trained simultaneously with synthetic data from three loading modes, and tested with tension, compression, and shear data individually. Dots illustrate the generated synthetic data. R^2 indicates the goodness of fit.

synthetic stress data under k th loading mode; $p_{i,k}^{\text{noise}}$ denotes the prescribed noise, sampled from a Gaussian distribution with zero mean and standard deviation σ_k . Here, we applied a consistent relative deviation to each loading mode, thus the actual standard deviation was scaled based on the maximum stress of each mode, $\sigma_k = \sigma * P_{\max,k}$. Relative deviations ranging from 0 to 20 % were utilized to assess the robustness of our symbolic regression algorithms. The training setups for each noise case remained consistent as outlined in Table S1.

For invariant-based models, prediction results along with the corresponding MSE are shown in Table 2, while the fitting performances under four typical noise scenarios are illustrated in Fig. 17. As seen, our algorithm is capable of precisely discovering the pre-defined mathematic format of the target strain energy function even with a 10 % noise prescription. However, the accuracy of data fitting continues to diminish as noise amplifies. With a 20 % noise imposition, the data distributions become too random to reveal

a discernible mathematic trend, resulting in the failure of precise predictions. An analogous effect is observed for the stretch-based and strain-based symbolic regression algorithms, as illustrated in Figures S8 and S9, with their detailed functions presented in Table S2 and S3, respectively. Consequently, it is evident that our algorithms demonstrate satisfactory robustness in the current model discovery scenarios.

4. Conclusion and future endeavors

We proposed a symbolic regression framework capable of autonomously identifying interpretable hyperelastic models from sparse experimental data while ensuring adherence to physical laws. Our study explored three distinct approaches to hyperelastic models—invariant-based, principal stretch-based, and normal strain-based—to unveil the capabilities of our symbolic regression algorithms. To ensure the physical validity of the predicted con-

stitutive models, we customized the symbolic regression algorithm by integrating the polyconvexity condition into the objective functions for the invariant-based algorithm and validating rank-one convexity post hoc for model predictions from stretch-based and strain-based algorithms. After validating our algorithms on synthetic data, we extended our study to the human brain cortex using experimental data across three loading modes. We demonstrated the convexity requirements both within and beyond the experimental data regime for each discovered model. Additionally, we examined the interpolation and extrapolation capabilities of our algorithms based on partially selected training dataset. Finally, we assessed the robustness of our algorithms using synthetic data embedded with artificial Gaussian noises.

Our results reveal that symbolic regression can discover accurate hyperelastic models with parsimonious mathematic expressions in invariant-based, stretch-based, and strain-based scenarios. Among all discovered models, the strain-based model exhibits superior performance in fitting the experimental data, with an R^2 value exceeding 0.99 for all loading modes. Additionally, both principal stretch-based and strain-based models effectively capture the nonlinearity and tension-compression asymmetry inherent in the human brain cortex. Convexity checks validate the rigorous fulfillment of polyconvexity/ellipticity within and beyond the training data regime, except for certain stretch-based hyperelastic models that may lose convexity under large deformations. The evaluation of predictive capabilities indicates promising interpolation capabilities for all three models and acceptable extrapolation performance for stretch-based and strain-based models. Robustness tests underscore the accuracy and precision of our proposed symbolic regression algorithms.

In present study, we leveraged symbolic regression for identifying constitutive material models for the human brain cortex within the hyperelasticity context. Naturally, our approach is readily applicable to model discovery for other brain regions, including the corona radiata and corpus callosum [1], or other soft tissues like skin [67,68] and muscles [69]. In addition to hyperelasticity, the exploration of other constitutive behaviors in soft tissues such as viscosity and plasticity presents intriguing future avenues, akin to similar investigations in alloy composites [32] and concrete beams [49]. Furthermore, our current model discovery utilized data from three loading modes, uniaxial tension, uniaxial compression, and simple shear. Incorporating a more diverse range of loading scenarios, such as biaxial experiments [70], will significantly contribute to the comprehensive characterization of material behaviors, particularly in complex loading cases. While the current study focuses on identifying hyperelastic models under incompressibility, future research will investigate the effects of relaxing incompressibility. Moreover, the framework can also incorporate with Finite Element models to perform inverse parameter identification [71].

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRedit authorship contribution statement

Jixin Hou: Writing – original draft, Validation, Software, Methodology, Investigation. **Xianyan Chen:** Writing – review & editing, Formal analysis. **Taotao Wu:** Writing – review & editing, Validation. **Ellen Kuhl:** Writing – review & editing, Validation. **Xianqiao Wang:** Writing – review & editing, Writing – original draft, Validation, Supervision, Funding acquisition, Conceptualization.

Acknowledgement

JH and XW acknowledges the support from [National Science Foundation \(IIS-2011369\)](#) and National Institutes of Health ([1R01NS135574-01](#)). EK acknowledges the partial support from [National Science Foundation \(CMMI-2320933\)](#).

Data availability statement

The original contributions presented in the study are included in the article/supplemental material. Supplementary material associated with this article can be found in the online version. Further inquiries can be directed to the corresponding authors.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.actbio.2024.09.005](https://doi.org/10.1016/j.actbio.2024.09.005).

References

- [1] S. Budday, G. Sommer, C. Birkl, C. Langkammer, J. Haybaeck, J. Kohnert, M. Bauer, F. Paulsen, P. Steinmann, E. Kuhl, Mechanical characterization of human brain tissue, *Acta Biomater.* 48 (2017) 319–340.
- [2] J. Hou, B. Park, C. Li, X. Wang, A multiscale computation study on bruise susceptibility of blueberries from mechanical impact, *Postharvest Biol. Technol.* 208 (2024) 112660.
- [3] K. Mendis, R. Stalnaker, S. Advani, A constitutive relationship for large deformation finite element modeling of brain tissue, (1995).
- [4] P. Thamburaja, N. Nikabdullah, A macroscopic constitutive model for shape-memory alloys: theory and finite-element simulations, *Comput. Methods Appl. Mech. Eng.* 198 (9) (2009) 1074–1086.
- [5] S. Budday, G. Sommer, J. Haybaeck, P. Steinmann, G.A. Holzapfel, E. Kuhl, Rheological characterization of human brain tissue, *Acta Biomater.* 60 (2017) 315–329.
- [6] L. Onsager, Reciprocal relations in irreversible processes, I, *Physical review* 37 (4) (1931) 405.
- [7] L. Onsager, Reciprocal relations in irreversible processes, II, *Physical review* 38 (12) (1931) 2265.
- [8] M. Flaschel, S. Kumar, L. De Lorenzis, Automated discovery of generalized standard material models with EUCLID, *Comput. Methods Appl. Mech. Eng.* 405 (2023) 115867.
- [9] S. Hartmann, P. Neff, Polyconvexity of generalized polynomial-type hyperelastic strain energy functions for near-incompressibility, *Int. J. Solids. Struct.* 40 (11) (2003) 2767–2791.
- [10] L. Qingbin, J. Zhong, L. Mabao, W. Shichun, Acquiring the constitutive relationship for a thermal viscoplastic material using an artificial neural network, *J. Mater. Process. Technol.* 62 (1–3) (1996) 206–210.
- [11] G.E. Karniadakis, I.G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, L. Yang, Physics-informed machine learning, *Nature Reviews Physics* 3 (6) (2021) 422–440.
- [12] Y. Leng, V. Tac, S. Calve, A.B. Tepole, Predicting the mechanical properties of biopolymer gels using neural networks trained on discrete fiber network data, *Comput. Methods Appl. Mech. Eng.* 387 (2021) 114160.
- [13] Z. Zhang, Z. Zou, E. Kuhl, G.E. Karniadakis, Discovering a reaction-diffusion model for Alzheimer's disease by combining PINNs with symbolic regression, *Comput. Methods Appl. Mech. Eng.* 419 (2024) 116647.
- [14] I. Jeong, M. Cho, H. Chung, D.-N. Kim, Data-driven nonparametric identification of material behavior based on physics-informed neural network with full-field data, *Comput. Methods Appl. Mech. Eng.* 418 (2024) 116569.
- [15] L. Linden, D.K. Klein, K.A. Kalina, J. Brummund, O. Weeger, M. Kästner, Neural networks meet hyperelasticity: a guide to enforcing physics, *J. Mech. Phys. Solids.* 179 (2023) 105363.
- [16] P. Chen, J. Guilleminot, Polyconvex neural networks for hyperelastic constitutive models: a rectification approach, *Mech. Res. Commun.* 125 (2022) 103993.
- [17] K. Linka, E. Kuhl, A new family of Constitutive Artificial Neural Networks towards automated model discovery, *Comput. Methods Appl. Mech. Eng.* 403 (2023) 115731.
- [18] F. As'ad, P. Avery, C. Farhat, A mechanics-informed artificial neural network approach in data-driven constitutive modeling, *Int. J. Numer. Methods Eng.* 123 (12) (2022) 2738–2759.
- [19] V. Tac, F. Sahli Costabal, A.B. Tepole, Data-driven tissue mechanics with polyconvex neural ordinary differential equations, *Comput. Methods Appl. Mech. Eng.* 398 (2022) 115248.
- [20] V. Taç, K. Linka, F. Sahli-Costabal, E. Kuhl, A.B. Tepole, Benchmarking physics-informed frameworks for data-driven hyperelasticity, *Comput. Mech.* (2023).
- [21] N. Ellmer, R. Ortigosa, J. Martínez-Frutos, A.J. Gil, Gradient enhanced gaussian process regression for constitutive modelling in finite strain hyperelasticity, *Comput. Methods Appl. Mech. Eng.* 418 (2024) 116547.
- [22] A.K. Açan, O.Z. Tikenogullari, H. Dal, A data-driven constitutive model for soft biological tissues, *PAMM* 23 (4) (2023) e202300239.

- [23] A. Aggarwal, L.T. Hudson, D.W. Laurence, C.-H. Lee, S. Pant, A Bayesian constitutive model selection framework for biaxial mechanical testing of planar soft tissues: application to porcine aortic valves, *J. Mech. Behav. Biomed. Mater.* 138 (2023) 105657.
- [24] K. Linka, S.R.S. Pierre, E. Kuhl, Automated model discovery for human brain using Constitutive Artificial Neural Networks, *Acta Biomater.* 160 (2023) 134–151.
- [25] Y. Wang, N. Wagner, J.M. Rondinelli, Symbolic regression in materials science, *MRS. Commun.* 9 (3) (2019) 793–805.
- [26] P. Orzechowski, W. La Cava, J.H. Moore, Where are we now? A large benchmark study of recent symbolic regression methods, in: *Proceedings of the Genetic and Evolutionary Computation Conference*, 2018, pp. 1183–1190.
- [27] S.-M. Udrescu, M. Tegmark, A.I. Feynman, A physics-inspired method for symbolic regression, *Sci. Adv.* 6 (16) (2020) eaay2631.
- [28] M. Ansari, H.A. Gandhi, D.G. Foster, A.D. White, Iterative symbolic regression for learning transport equations, *AIChE Journal* 68 (6) (2022) e17695.
- [29] P. Neumann, L. Cao, D. Russo, V.S. Vassiliadis, A.A. Lapkin, A new formulation for symbolic regression to identify physico-chemical laws from experimental data, *Chem. Eng. J.* 387 (2020) 123412.
- [30] D. Angelis, F. Sofos, T.E. Karakasidis, Artificial intelligence in physical sciences: symbolic regression trends and perspectives, *Arch. Comput. Methods Eng.* 30 (6) (2023) 3845–3865.
- [31] M. Schmelzer, R.P. Dwight, P. Cinnella, Discovery of Algebraic Reynolds-Stress Models Using Sparse Symbolic Regression, *Flow. Turbul. Combust.* 104 (2) (2020) 579–603.
- [32] E. Kablman, A.H. Kolody, J. Kronsteiner, M. Kommenda, G. Kronberger, Application of symbolic regression for constitutive modeling of plastic deformation, *Appl. Eng. Sci.* 6 (2021) 100052.
- [33] R. Abdusalamov, M. Hillgärtner, M. Itskov, Hyperelastic material modelling using symbolic regression, *PAMM* 22 (1) (2023) e202200263.
- [34] R. Abdusalamov, M. Hillgärtner, M. Itskov, Automatic generation of interpretable hyperelastic material models by symbolic regression, *Int. J. Numer. Methods Eng.* 124 (9) (2023) 2093–2104.
- [35] B. Bahmani, W. Sun, Physics-constrained symbolic model discovery for polyconvex incompressible hyperelastic materials, *arXiv preprint arXiv:2310.04286* (2023).
- [36] A. Anssari-Benam, G. Saccomandi, Continuous Softening as a State of Hyperelasticity: examples of Application to the Softening Behavior of the Brain Tissue, *J. Biomech. Eng.* 146 (9) (2024).
- [37] G. He, B. Xia, Y. Feng, Y. Chen, L. Fan, D. Zhang, Modeling the damage-induced softening behavior of brain white matter using a coupled hyperelasticity-damage model, *J. Mech. Behav. Biomed. Mater.* 141 (2023) 105753.
- [38] D.K. Klein, M. Fernández, R.J. Martin, P. Neff, O. Weeger, Polyconvex anisotropic hyperelasticity with neural networks, *J. Mech. Phys. Solids* 159 (2022) 104703.
- [39] K. Willam, Constitutive models for engineering materials, *Encyclopedia. Phys. Sci. Technol.* 3 (2002) 603–633.
- [40] T. Mai, J.R. Walton, On strong ellipticity for implicit and strain-limiting theories of elasticity, *Math. Mech. Solids* 20 (2) (2015) 121–139.
- [41] G. Chagnon, M. Rebouah, D. Favier, Hyperelastic energy densities for soft biological tissues: a review, *J. Elast* 120 (2015) 129–160.
- [42] E. Kuhl, H. Askes, P. Steinmann, An illustration of the equivalence of the loss of ellipticity conditions in spatial and material settings of hyperelasticity, *Eur. J. Mech. - A/Solids* 25 (2) (2006) 199–214.
- [43] D. Yang Gao, P. Neff, I. Roventa, C. Thiel, On the Convexity of Nonlinear Elastic Energies in the Right Cauchy-Green Tensor, *J. Elast* 127 (2) (2017) 303–308.
- [44] J.M. Ball, Convexity conditions and existence theorems in nonlinear elasticity, *Archive for rational mechanics and Analysis* 63 (1976) 337–403.
- [45] J. Schröder, P. Neff, Invariant formulation of hyperelastic transverse isotropy based on polyconvex free energy functions, *Int. J. Solids Struct.* 40 (2) (2003) 401–445.
- [46] G.A. Holzapfel, T.C. Gasser, R.W. Ogden, A new constitutive framework for arterial wall mechanics and a comparative study of material models, *J. Elasticity Phys. Sci. Solids* 61 (1) (2000) 1–48.
- [47] A. Anssari-Benam, M. Destrade, G. Saccomandi, Modelling brain tissue elasticity with the Ogden model and an alternative family of constitutive models, *Philosoph. Trans. R. Soc. A: Math., Phys. Eng. Sci.* 380 (2234) (2022) 20210325.
- [48] W. Liu, M. Röckner, Local and global well-posedness of SPDE with generalized coercivity conditions, *J. Differ. Equ* 254 (2) (2013) 725–755.
- [49] W. Ben Chaabene, M.L. Nehdi, Genetic programming based symbolic regression for shear capacity prediction of SFRC beams, *Constr. Build. Mater.* 280 (2021) 122523.
- [50] D.A. Nembhard, Y. Sun, A symbolic genetic programming approach for identifying models of learning-by-doing, *Comput. Ind. Eng.* 131 (2019) 524–533.
- [51] M. Cranmer, Interpretable machine learning for science with PySR and SymbolicRegression, *jl*, *arXiv preprint arXiv:2305.01582* (2023).
- [52] J.A. McCulloch, S.R.S. Pierre, K. Linka, E. Kuhl, On sparse regression, Lp-regularization, and automated model discovery, *arXiv preprint arXiv:2310.06872* (2023).
- [53] N. Filla, J. Hou, T. Liu, S. Budday, X. Wang, Accuracy meets simplicity: a constitutive model for heterogeneous brain tissue, *J. Mech. Behav. Biomed. Mater.* 150 (2024) 106271.
- [54] M. Mooney, A theory of large elastic deformation, *J. Appl. Phys.* 11 (9) (1940) 582–592.
- [55] A.N. Gent, A new constitutive relation for rubber, *Rubber Chem. Technol.* 69 (1) (1996) 59–61.
- [56] G.A. Holzapfel, T.C. Gasser, R.W. Ogden, A new constitutive framework for arterial wall mechanics and a comparative study of material models, *J. Elast. Phys. Sci. Solids* 61 (2000).
- [57] R.W. Ogden, Large deformation isotropic elasticity—on the correlation of theory and experiment for incompressible rubberlike solids, *Proceedings of the Royal Society of London, A. Math. Phys. Sci.* 326 (1567) (1972) 565–584.
- [58] S.R. St. Pierre, K. Linka, E. Kuhl, Principal-stretch-based constitutive neural networks autonomously discover a subclass of Ogden models for human brain tissue, *Brain Multiphys.* 4 (2023) 100066.
- [59] R.W. Ogden, *Non-linear Elastic Deformations*, Courier Corporation, 1997.
- [60] J. Hou, N. Filla, X. Chen, M.J. Razavi, T. Liu, X. Wang, Exploring hyperelastic material model discovery for human brain cortex: multivariate analysis vs. artificial neural network approaches, *arXiv preprint arXiv:2310.10762* (2023).
- [61] N. Firouzi, A. Misra, New insight into large deformation analysis of stretch-based and invariant-based rubber-like hyperelastic elastomers, *Thin-Walled Structures* 192 (2023) 111162.
- [62] H. Dal, K. Açıkgöz, Y. Badienia, On the Performance of Isotropic Hyperelastic Constitutive Models for Rubber-Like Materials: a State of the Art Review, *Appl. Mech. Rev.* 73 (2) (2021).
- [63] P.A. Vikhar, in: *2016 International conference on global trends in signal processing, information computing and communication (ICGTSPICC)*, IEEE, 2016, pp. 261–265.
- [64] H. Mühlenbein, M. Gorges-Schleuter, O. Krämer, Evolution algorithms in combinatorial optimization, *Parallel. Comput.* 7 (1) (1988) 65–85.
- [65] M. Nagle, H.C. Broderick, C. Vedel, M. Destrade, M. Fop, A.N. Annaidh, A Gaussian process approach for rapid evaluation of skin tension, *Acta Biomater.* (2024).
- [66] B. Bahmani, W. Sun, Physics-constrained symbolic model discovery for polyconvex incompressible hyperelastic materials, *Int. J. Numer. Methods Eng.* (2024) e7473 n/a(n/a).
- [67] V. Tac, V.D. Sree, M.K. Rausch, A.B. Tepole, Data-driven modeling of the mechanical behavior of anisotropic soft biological tissue, *Eng. Comput.* (2022).
- [68] K. Linka, A.B. Tepole, G.A. Holzapfel, E. Kuhl, Automated model discovery for skin: discovering the best model, data, and experiment, *Comput. Methods Appl. Mech. Eng.* 410 (2023) 116007.
- [69] L.M. Wang, K. Linka, E. Kuhl, Automated model discovery for muscle using constitutive recurrent neural networks, *J. Mech. Behav. Biomed. Mater.* 145 (2023) 106021.
- [70] K.M. Labus, C.M. Puttlitz, An anisotropic hyperelastic constitutive model of brain white matter in biaxial tension and structural-mechanical relationships, *J. Mech. Behav. Biomed. Mater.* 62 (2016) 195–208.
- [71] J.S. Giudice, A. Alshareef, T. Wu, A.K. Knutsen, L.V. Hiscox, C.L. Johnson, M.B. Panzer, Calibration of a heterogeneous brain model using a subject-specific inverse finite element approach, *Front. Bioeng. Biotechnol.* 9 (2021).