International Journal of Mechanical Sciences 321 (2026) 111696

Contents lists available at ScienceDirect

International Journal of Mechanical Sciences

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

A LLM-inspired experimental-data-driven framework for viscoelastic soft structures

Yicheng Lu [a,b], Xinjie Hu [c], Yichen Pu [c], Zefeng Yu [a,b], Ning An [c], Shan Tang [a,b,*], Xu Guo [a,b,*]

[a] State Key Laboratory of Structural Analysis, Optimization and CAE Software for Industrial Equipment, Department of Engineering Mechanics, Dalian University of Technology, Dalian, 116023, PR China
[b] International Research Center for Computational Mechanics, Dalian University of Technology, Dalian, 116023, PR China
[c] Key Laboratory of Advanced Spatial Mechanism and Intelligent Spacecraft, Ministry of Education, School of Aeronautics and Astronautics, Sichuan University, Chengdu 610065, PR China

ARTICLE INFO

Keywords:
Soft structures
Neural networks
Data-driven
Finite-strain
Viscoelastic constitutive model
Mechanics-informed

ABSTRACT

Soft structures such as soft fluidic actuators (SFAs) are key components in soft robotics, yet accurate prediction of their viscoelastic response under finite deformation remains challenging, as most existing approaches focus on time-independent hyperelastic behavior. Inspired by large language models (LLMs), this study proposes a mechanics-informed, experimental-data-driven framework for predicting the viscoelastic responses of Sil 950-based SFAs. A tailored finite-strain memory decay network (fMDN) is trained through pre-training, fine-tuning, and inference stages using a small amount of experimental data, including material-level uniaxial tests and SFA experiments, supplemented by limited numerical data. During training, viscoelastic memory decay and thermodynamic consistency are enforced, and a UMAT subroutine is implemented for finite element analysis. At the inference stage, the fMDN achieves more accurate predictions than conventional viscoelastic constitutive models for both short- and long-term responses, enables extrapolation beyond the training loading durations, and does not rely on postulated forms of hyperelastic strain energy or viscous relaxation functions. Moreover, the proposed approach exhibits improved computational efficiency for composite structures. These results demonstrate the strong generalization capability of the proposed framework and its potential applicability to other soft viscoelastic structures, in line with the design philosophy of LLM-inspired learning strategies.

1. Introduction

Soft structures, owing to their high flexibility, extensibility, and biocompatibility, are widely used in diverse applications, including surgical interventions [1,2], human assistive devices [3,4], rehabilitation nursing [5,6], human–robot interaction [7], and fragile object manipulation [8,9]. Typically, soft structures are composed of rubber-like materials that exhibit time-dependent viscoelastic mechanical response. The time-dependent characteristics of these materials play a critical role in the performance of soft structures, particularly when subjected to sustained mechanical loading under finite strain conditions. In addition, soft materials attract attention in applications of tunable metamaterials and phononic crystals [10–16]. SFAs, as a typical class of soft structures, are the essential building blocks responsible for producing motion in soft robotics, which are commonly fabricated from the viscoelastic silicone rubber material via 3D printing techniques [17,18]. However, most current researches on SFA mechanical properties predominantly focus on hyperelastic behavior while neglecting the complex time-dependent effects arising from material viscosity [19–25], which highlights the importance of incorporating a comprehensive viscoelastic characterization in SFA design and analysis to account for both instantaneous and time-dependent mechanical responses. Accurately capturing the complex geometry and time-dependent viscoelastic behavior under finite deformation substantially increases the modeling complexity for SFAs [26].

Conventional viscoelastic constitutive models have been extensively developed over the past six decades. Schapery [27] proposed a nonlinear single-integral constitutive model that extends the Boltzmann superposition principle to finite-strain through incorporating nonlinear factors. Zapas and Craft [28] established a hereditary integral formulated viscoelastic model based on strain energy functions. In addition, some studies advance the conventional differential viscoelastic models (such as the Maxwell model and the Kelvin–Voigt model) to account for finite-strain conditions by incorporating time-dependent or rate-dependent nonlinear terms [29]. Bagley and Torvik [30] introduced

---

* Corresponding authors.
E-mail addresses: shantang@dlut.edu.cn (S. Tang), guoxu@dlut.edu.cn (X. Guo).

https://doi.org/10.1016/j.ijmecsci.2026.111696
Received 17 January 2026; Received in revised form 1 May 2026; Accepted 3 May 2026

fractional calculus formulations to capture the memory decay propertys and frequency-dependent behavior of viscoelasticity. Simo [31] and Bergström and Boyce [32] described the viscoelastic behavior by multiplicative decomposing the deformation gradient into elastic and viscous components. However, the mathematical formulations of these constitutive models inevitably exhibit significant complexity, presenting challenges in both model selection and parameter identification — particularly for empirically derived parameters that require extensive multi-condition experimental calibration [33]. This limitation is especially pronounced in SFA applications due to the coupled multi-physics loading environment involving pressurized actuation and gravitational effects [34–37], which increases the cost and difficulty of obtaining experimental data.

Fortunately, the rapid development of artificial intelligence (AI) technologies presents transformative opportunities to overcome the above challenges posed by conventional constitutive models [38–49]. Advanced machine learning techniques, particularly deep neural networks, enable effective modeling of complex viscoelastic behaviors through comprehensive analysis of experimental data. As a result, a substantial body of experimental and theoretical research has emerged, focusing on soft structures and facilitating their applications across a wide range of fields [50–56]. Currently, data-driven constitutive modeling approaches are often developed based on relatively extensive experimental datasets, including multiaxial loading tests, full-field deformation measurements, and densely sampled stress–strain databases that cover a wide range of deformation paths. For example, artificial neural network (ANN) models require comprehensive datasets spanning uniaxial, biaxial, shear, and volumetric loading conditions [57], while full-field approaches rely on high-dimensional displacement and strain field data [58,59]. Similarly, model-free data-driven computing frameworks necessitate densely populated material databases to ensure accuracy [60]. Specifically, Al-Haik et al. [61] developed an ANN model to predict the one-dimensional stress relaxation behavior of polymer matrix composites. This ANN model was trained and validated using 9000 experimental datasets obtained from stress relaxation tests under different temperatures. Jordan et al. [62] utilized a simple ANN to characterize the one-dimensional rate-dependent behavior of polypropylene under finite deformation. This model takes strain and temperature as the input, and loading strain rate as the output. These works, the same as the other pure data-driven approaches [63–67], struggle to directly extract physics-related features embedded within extensive datasets and often fail to enforce fundamental mechanical principles, including viscoelastic memory decay and thermodynamic consistency. Consequently, these works typically require constructing large-scale deep learning architectures that nevertheless exhibit limited predictive capabilities, which perform adequately only within specific experimental conditions (e.g., one-dimensional setting) on which they were trained. When applied to more complex loading conditions, a poor generalization capacity may be observed [68]. In recent years, significant research efforts have been devoted to developing mechanics-informed data-driven approaches that simultaneously enhance predictive capability while ensuring strict adherence to fundamental physical principles. These approaches strategically integrate well-established governing laws with deep learning frameworks to address the limitations of purely data-driven models [69–80]. For instance, Bozkurt and Tagarielli [75] developed a knowledge-based ANN to characterize the homogenized mechanical response of viscoelastic porous elastomers at the unit cell level, utilizing numerical datasets generated through explicit finite element simulations. Compared with pure regression data-driven model, this knowledge-based model requires a simpler ANN and incurs a smaller computational cost. Additionally, based on a general mechanistic data science framework, Mojumder et al. [76] employed low-dimensional surrogate representations together with physics-guided learning based on regression and classification to characterize polymer matrix composites. To mitigate the substantial dependence on training datasets, Dekhovich et al. [77] developed an innovative continual learning framework for inelastic material modeling that integrates cooperative data-driven strategies. This framework achieves reducing the number of training data required to learn a new task, while maintaining precise prediction accuracy for previously learned tasks. Despite significant advancements in data-driven constitutive modeling, critical challenges remain in accurately characterizing the memory decay property of viscoelastic materials, while simultaneously satisfying thermodynamic consistency and requiring a small amount of experimental data. These limitations persist as challenges in the relevant literature, particularly in accurately predicting the viscoelastic mechanical behavior of soft structure applications. Although these effects are widely observed in many experiments involving rubber-like materials, they remain difficult to model effectively [81, 82].

Using SFAs as a representative case of soft structures, this paper develops a mechanics-informed experimental-data-driven framework that achieves more accurate and comprehensive characterization of both short- and long-term viscoelastic behavior than conventional viscoelastic constitutive models, for Sil 950-based SFAs subjected to hydraulic or pneumatic actuation. Inspired by the training strategy of LLMs, this approach employs the standard three-stage process of pre-training, fine-tuning, and inference to train our tailored fMDN, as detailed in Fig. 1. The proposed fMDN employs an ANN and two specialized convolutional neural networks with trainable kernel lengths (TKL-CNNs) to model the hyperelastic parameters and the relaxation function related parameters — key physical quantities governing the viscoelastic memory decay property and thermodynamic consistency. First, in the pre-training stage, a representative volume element (RVE) incorporating material parameters preliminarily characterized through material-level uniaxial tests. This RVE generates numerical pre-training datasets while extracting feature variables that control the memory decay property and thermodynamic consistency. The fMDN then undergoes mechanics-constrained pre-training using these datasets, which helps reduce optimization costs and enhance the initial accuracy for subsequent fine-tuning. Subsequently, in the fine-tuning stage, this pre-trained data-driven constitutive model is implemented within the finite element method (FEM) framework and further fine-tuned using SFA experimental data. Finally, the trained model enables the precise description of the viscoelastic behavior of SFAs in the inference stage. Notably, the entire training protocol utilizes a small amount of experimental data, including material-level uniaxial test results and SFAs test results, supplemented with a small amount of augmented numerical data. Meanwhile, mechanics-informed constraints are rigorously applied throughout the training process.

The structure of this paper is as follows. Section 2 presents the extraction of viscoelastic mechanical features. Section 3 delineates the pre-training methodology by utilizing the material-level test data combined with an augmented numerical dataset. The architecture of fMDN and the implementation of mechanics-informed constraints are also detailed to enforce viscoelastic fading memory and thermodynamic consistency. In Section 4, the methodology of integrating the fMDN model into the FEM framework is first introduced. Then, the fine-tuning approach by utilizing the SFA experimental results is presented. Section 5 demonstrates the predictive capability of our data-driven approach in the inference stage for describing the viscoelastic behavior of SFAs. Finally, concluding remarks are provided in Section 6.

2. Mechanics-related characteristics of viscoelasticity

In this section, the physical meaning and constitutive roles of the characteristic variables used for fMDN training are first clarified, with particular emphasis on their connections to viscoelastic fading memory and thermodynamic consistency. On this basis, a detailed review of the mechanics theory of viscoelasticity is then presented to establish the theoretical foundation for the subsequent model development.

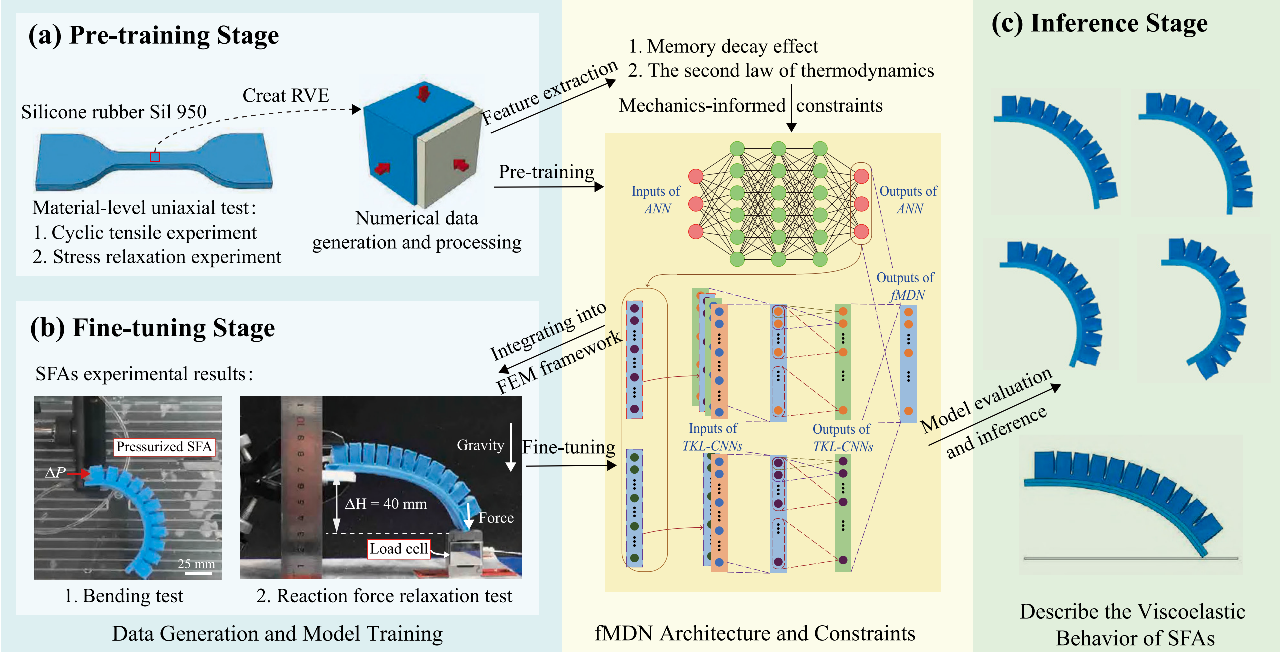

Fig. 1. Flowchart of the proposed experimental-data-driven framework for viscoelastic modeling of SFAs: (a) Pre-training stage; (b) Fine-tuning stage; (c) Inference stage.

From this point onward, to clearly distinguish stress quantities defined under different configurations, the stress symbol $\boldsymbol{\tau}$, regardless of the subscripts, consistently denotes the corresponding Kirchhoff stress. First, we decompose the Kirchhoff stress of viscoelasticity $\boldsymbol{\tau}(t)$ into the instantaneous hyperelastic part $\boldsymbol{\tau}_h(t)$ and the nonequilibrium viscous part $\boldsymbol{\tau}_v(t)$, and then decouple these two parts into the deviatoric and volumetric responses, respectively, as follows:

$$\boldsymbol{\tau} = \boldsymbol{\tau}_h(t) + \boldsymbol{\tau}_v(t) = \boldsymbol{\tau}_{h,dev}(t) + \boldsymbol{\tau}_{h,vol}(t) + \boldsymbol{\tau}_{v,dev}(t) + \boldsymbol{\tau}_{v,vol}(t), \tag{1}$$

where $\boldsymbol{\tau}_{v,dev}(t)$ and $\boldsymbol{\tau}_{v,vol}(t)$ are the deviatoric and volumetric parts of viscous stress respectively, and the deviatoric and volumetric parts of instantaneous hyperelastic stress, $\boldsymbol{\tau}_{h,dev}(t)$, $\boldsymbol{\tau}_{h,vol}(t)$, are defined as:

$$\begin{cases} \boldsymbol{\tau}_{h,dev}(t) = \mathrm{dev}[\boldsymbol{\tau}_h(t)] \\ \boldsymbol{\tau}_{h,vol}(t) = \dfrac{1}{3}\mathrm{tr}[\boldsymbol{\tau}_h(t)]\mathbf{1}. \end{cases} \tag{2}$$

Here, $\mathrm{dev}(\bullet) = (\bullet) - \frac{1}{3}\mathrm{tr}(\bullet)\mathbf{1}$, $\mathrm{tr}(\bullet) = (\bullet)_{ii}$, and $\mathbf{1}$ is the second-order unit tensor.

Unlike the elastic materials, the mechanical behavior of viscoelasticity shows strong time-dependent characteristics, which means the stress applied at historic time $t - t'$ influences the stress at current time t. Therefore, in the finite-strain regime, it is essential to map the stress that existed in the configuration at time $t - t'$ into the configuration at time t, which is done by a transformation with the relative deformation gradient $\boldsymbol{F}_{t-t'}(t)$ from the state at $t-t'$ to the state at t. This process is called standard push-forward transformation [31]. It is worth emphasizing that the volume strains of viscoelasticity are generally small and the mapping of viscous volumetric stress from time $t - t'$ to time t is not considered here [83]. Using the hereditary integral in the reference configuration for finite-strain that captures the deformation history, the nonequilibrium viscous responses $\boldsymbol{\tau}_{v,dev}(t)$ and $\boldsymbol{\tau}_{v,vol}(t)$ are expressed as follows:

$$\begin{cases} \boldsymbol{\tau}_{v,dev}(t) = \mathrm{dev}[\int_0^t -\alpha_G \dot{g}_R(t')\overline{\boldsymbol{F}}_t^{-1}(t-t')\cdot\boldsymbol{\tau}_{h,dev}(t-t')\cdot\overline{\boldsymbol{F}}_t^{-T}(t-t')dt'] & \text{(3a)} \\ \boldsymbol{\tau}_{v,vol}(t) = \int_0^t -\alpha_K \dot{k}_R(t')\boldsymbol{\tau}_{h,vol}(t-t')dt', & \text{(3b)} \end{cases}$$

where $\overline{\boldsymbol{F}} = J^{-\frac{1}{3}}\boldsymbol{F}$ is the distortional deformation gradient related to the deformation gradient $\boldsymbol{F}$, $J = \det(\boldsymbol{F})$ is the determinant of deformation gradient, $\overline{\boldsymbol{F}}_t(t-t')$ is the distortional deformation gradient of the state at $t - t'$ relative to the state at t, and $\alpha_G = G_R/G_0$ and $\alpha_K = K_R/K_0$ are the dimensionless initial relaxation shear and bulk moduli, with G_0 and K_0 are the initial shear and bulk moduli. Also, $\dot{g}_R(t')$ and $\dot{k}_R(t')$ are respectively the derivative of deviatoric and volumetric relaxation functions with respect to time, varying according to the time interval between the historical time t' and present time t. The time dependence in viscoelasticity arises directly from this characteristic of relaxation functions. The negative sign preceding the dimensionless initial relaxation moduli is introduced to facilitate the subsequent derivation of viscous characteristics.

To capture the viscoelastic time-dependent response, a memory decay formulation is employed, where the relaxation functions $g_R(t')$ and $k_R(t')$ are modeled as monotonically decreasing functions with progressively diminishing decay rates that asymptotically stabilize. As a result, the contribution of past loading conditions to the current stress state decays progressively with time [84]. The extension of this constraint to the derivative of relaxation functions, $\dot{g}_R(t')$ and $\dot{k}_R(t')$, ensures that they also exhibit a monotonically decreasing trend, as influenced by the negative sign in Eq. (3). Specifically, it requires that:

$$\begin{cases} \dot{g}_R(t_1) < \dot{g}_R(t_2) \\ \dot{k}_R(t_1) < \dot{k}_R(t_2) \end{cases} \text{if } t_1 > t_2. \tag{4}$$

These viscoelastic memory decay constraints will be incorporated into the fMDN in Section 3.3.1.

Exploiting the classical Clausius-Duhem form of the second law of thermodynamics, the dissipation Ξ in a material must be non-negative. When thermal effects are neglected (adiabatic assumption), the viscous dissipation function Ξ takes the following form:

$$\Xi(\boldsymbol{\tau}_{v,dev}, \boldsymbol{\tau}_{v,vol}, \boldsymbol{Q}_{dev}, \boldsymbol{Q}_{vol}) = \frac{\boldsymbol{\tau}_{v,dev}}{G_R} : \dot{\boldsymbol{Q}}_{dev} + \frac{\boldsymbol{\tau}_{v,vol}}{K_R} : \dot{\boldsymbol{Q}}_{vol} \geq 0, \tag{5}$$

where a superscript dot denotes differentiation by time, and the $\boldsymbol{Q}_{dev}$ and $\boldsymbol{Q}_{vol}$ are the time-evolving internal variables. The following set of rate equations governs the evolution of these internal variables:

$$\begin{cases} \dot{\boldsymbol{Q}}_{dev}(t) + \dfrac{1}{\tau^G}\boldsymbol{Q}_{dev}(t) = \dfrac{\alpha_G}{\tau^G}\boldsymbol{\tau}_{h,dev}(t) \\ \lim\limits_{t\to 0} \boldsymbol{Q}_{dev}(t) = 0, \end{cases} \tag{6}$$

and

$$\begin{cases} \dot{\boldsymbol{Q}}_{vol}(t) + \dfrac{1}{\tau^K}\boldsymbol{Q}_{vol}(t) = \dfrac{\alpha_K}{\tau^K}\boldsymbol{\tau}_{h,vol}(t) \\ \lim\limits_{t\to 0} \boldsymbol{Q}_{vol}(t) = 0, \end{cases} \tag{7}$$

where τ^G and τ^K are the deviatoric and volumetric relaxation time, respectively. The inequality condition in Eq. (5) imposes thermodynamic consistency constraints on the relaxation function derivatives $\dot{g}_R$ and $\dot{k}_R$. These constraints require that the derivatives must monotonically decrease over time such that their current values at time t are always less than their historical values at any prior time t', while asymptotically approaching zero as $t \to \infty$ to satisfy thermodynamic consistency according to [85]. These conditions can be formally expressed as:

$$\min_{t'} \dot{g}_R(t) \le \dot{g}_R(t - t') \tag{8a}$$

$$\lim_{t\to\infty} \dot{g}_R(t) = 0, \tag{8b}$$

and

$$\min_{t'} \dot{k}_R(t) \le \dot{k}_R(t - t') \tag{9a}$$

$$\lim_{t\to\infty} \dot{k}_R(t) = 0. \tag{9b}$$

When these constraints are coupled with the viscoelastic fading memory constraint (Eq. (4)), Eqs. (8a) and (9a) are inherently fulfilled. The remaining conditions, Eqs. (8b),(9b), stem exclusively from the second law of thermodynamics and are subsequently embedded into the fMDN in Section 3.3.2.

In summary, the dimensionless initial relaxation shear and bulk moduli α_G and α_K, and the time derivatives of relaxation functions $\dot{g}_R$ and $\dot{k}_R$, serve as fundamental parameters that critically govern the time-dependent memory decay characteristics and the thermodynamic consistency of viscoelasticity. We note, however, that these fundamental parameters present significant challenges for direct experimental measurement.

3. Pre-training stage

In this section, the implementation of the pre-training stage is presented. Specifically, Section 3.1 describes the generation of the pre-training dataset. Section 3.2 then introduces the architecture of the proposed fMDN model. Finally, Section 3.3 presents the associated mechanics-informed constraints.

3.1. Pre-training dataset preparation

This section details the data generation methodology for pre-training and elucidates the feature extraction process for developing a data-driven viscoelastic model. The proposed framework begins with material-level experimental characterization to preliminarily evaluate the material parameters of Sil 950 silicone rubber. These material parameters are then used to construct a homogeneous RVE for generating an augmented numerical pre-training dataset. Feature quantities that adhere to the memory decay property and thermodynamic consistency are also extracted from the RVE simulations.

3.1.1. Material-level uniaxial experiments

Here, the stress relaxation experiment and the uniaxial cyclic tensile experiment were conducted on Sil 950 specimens with identical geometry as shown in Fig. 2(a).

The first experiment was designed to characterize the long-term stress relaxation behavior of Sil 950. Specimens were extended to 66 mm (200% engineering strain) at a rate of 1000 mm/min and held at this extension for 1.5 h. The resulting stress relaxation curve is shown in Fig. 2(b). The shaded region represents the 95% confidence interval obtained from repeated experimental measurements. As observed in the figure, Sil 950 exhibits a pronounced stress relaxation behavior, with approximately 30% stress decay during the relaxation process and a characteristic relaxation time slightly exceeding 1.5 h.

In the second experiment, an uniaxial cyclic tensile test was conducted to characterize the instantaneous hyperelastic response of Sil 950. Specimens were subjected to 10 complete loading–unloading cycles between 0 mm and 77 mm extension (corresponding to approximately 230% engineering strain) at a constant rate of 100 mm/min. Under such large deformation conditions, true stress was adopted to account for the variation of the specimen cross-sectional area. The true stress was calculated assuming the material to be nearly incompressible, which is a common approximation for silicone elastomers [81, 86]. At the specified loading rate, each loading cycle was completed within approximately 1.5 min, which is significantly shorter than the characteristic relaxation time identified in the stress relaxation experiment. Consequently, the contribution of time-dependent viscous effects during each loading cycle is expected to be limited. Therefore, the cyclic stress–strain response shown in Fig. 2(c) can be primarily attributed to the instantaneous hyperelastic behavior of the material. Additionally, Fig. 2(c) clearly demonstrates irreversible changes in the material response after the first cycle. The observed reduction in stiffness provides clear evidence of the Mullins effect [87,88] in this material. This observation is consistent with Case et al. [81].

Based on the obtained uniaxial experimental data, we can preliminarily determine the material parameters. This step was accomplished using the parameter evaluation tool in ABAQUS, where the uniaxial experimental data were directly input into a specified constitutive model to evaluate material parameters. Consistent with our experimental design considerations, the instantaneous hyperelastic parameters were derived from uniaxial cyclic tensile data. To eliminate the influence of the Mullins effect, the second loading cycle was used for instantaneous parameter identification using a conventional hyperelastic model (Yeoh or Arruda-Boyce), both of which are widely adopted for soft materials [19,86]. It should be noted that this treatment effectively limits the analysis to stabilized material behavior following pre-cycling. Additionally, the stress relaxation data were utilized to characterize the viscous properties of the material. Note that our methodology maintains complete generality in the mathematical formulation of the dimensionless initial relaxation moduli or the time derivatives of relaxation functions, requiring only that they compliance with the material's memory decay effect. Therefore, to facilitate the preliminary evaluation of material parameters and the generation of the augmented numerical datasets, the classical time-domain Prony series model, an exponential form representation, was employed to evaluate the viscous parameters. The explicit formulations of the Yeoh, Arruda-Boyce, and Prony series models are provided in Appendix A. Notably, these conventional constitutive models are first employed to efficiently provide an initial, coarse representation of the material behavior, which serves as a high-quality initialization for the subsequent data-driven constitutive modeling, thereby accelerating the overall training process.

3.1.2. Generation and processing of augmented numerical data

Upon determining the Yeoh hyperelastic and Prony Series viscoelastic parameters, a homogeneous RVE was constructed to generate pre-training datasets, as illustrated in Fig. 3. The RVE can be considered as being connected to a macro-scale material point (X^M, Y^M, Z^M) within the specimen. Here, and throughout the subsequent discussion, superscript M indicates parent-domain macro-scale quantities while superscript m denotes RVE-level meso-scale quantities. We select a cuboidal RVE in a Cartesian frame with principal axes x^m, y^m, z^m and

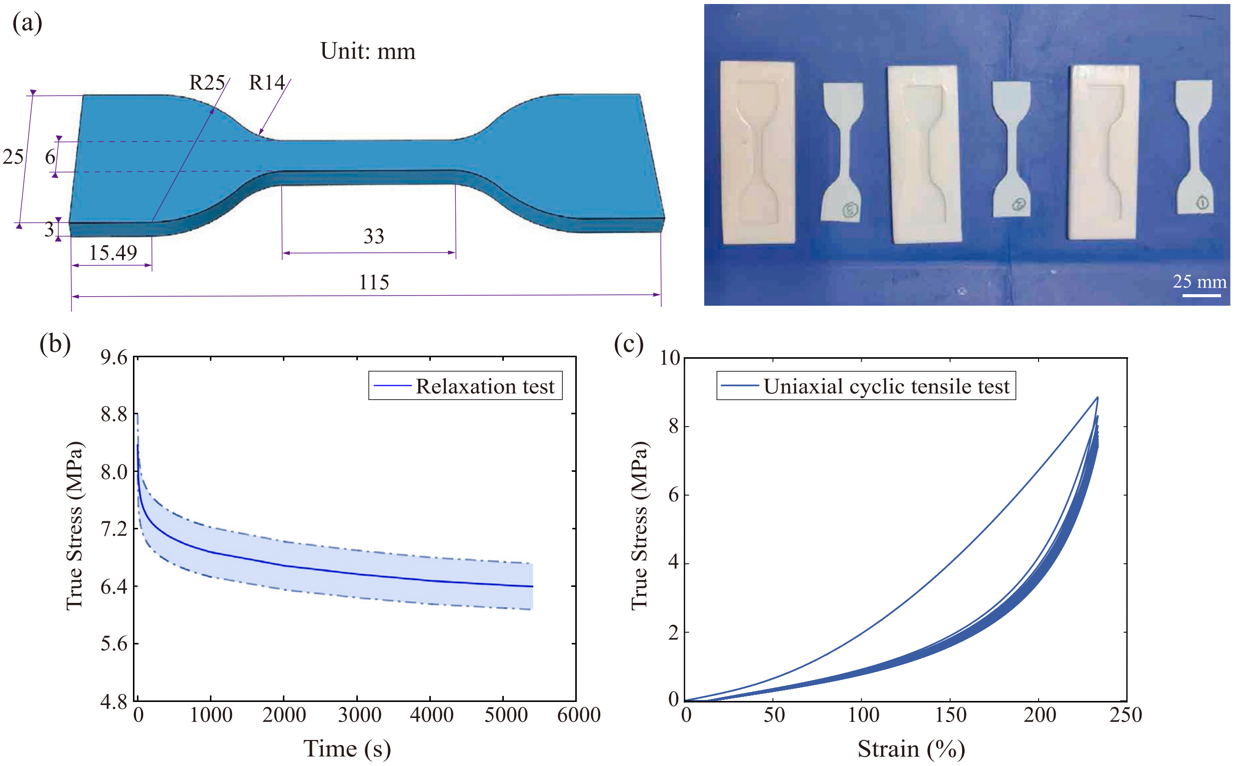

Fig. 2. The uniaxial experimental specimens and results for Sil 950: (a) Geometry of the Sil 950 specimens used in the uniaxial experiments, along with the specimen molds fabricated from Resin 8200Pro and the corresponding fabricated specimens; (b) Relaxation stress as a function of time over 1.5 h at a fixed strain of 200%, with the shaded area indicating the 95% confidence interval from repeated experiments; (c) Cyclic stress–strain curves over 10 cycles at a constant loading rate of 100 mm/min.

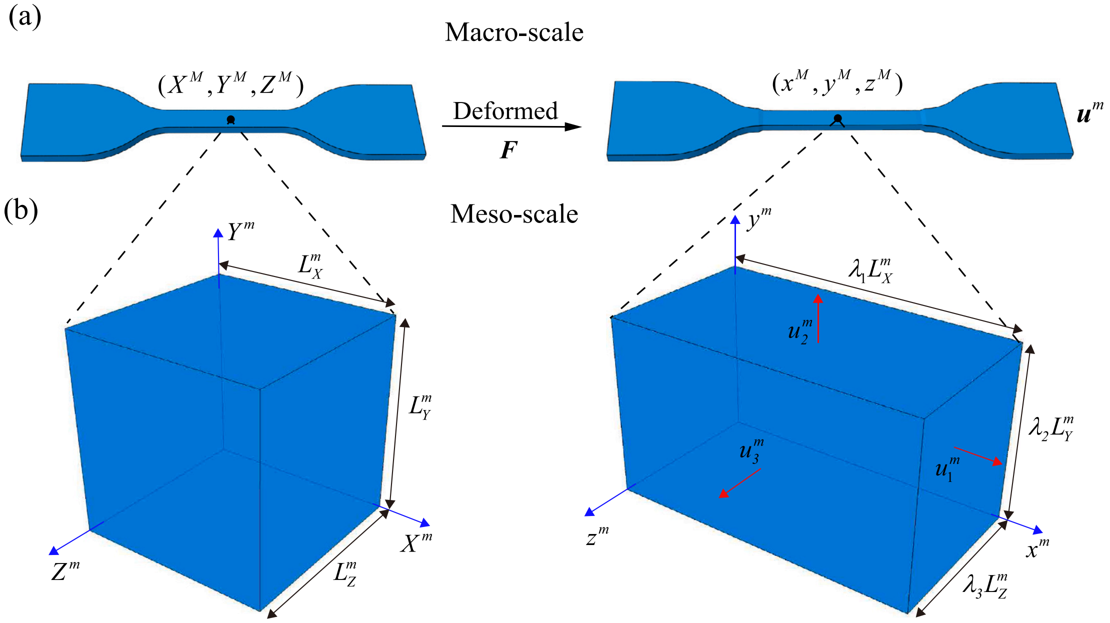

Fig. 3. The construction of RVE: (a) (X^M, Y^M, Z^M) and (x^M, y^M, z^M) represent the material points of the specimen at the macro-scale in the initial and deformed configurations, respectively. $\boldsymbol{u}^M$ denotes the imposed displacement field on the boundary of the continuum body; (b) The initial and deformed configurations of the RVE at the meso-scale.

corresponding dimensions L_x^m, L_y^m, L_z^m. The RVE boundary conditions are imposed as follows:

$$\begin{cases} u_1^m = 0 \ , \forall \ x^m = 0 \\ u_2^m = 0 \ , \forall \ y^m = 0 \\ u_3^m = 0 \ , \forall \ z^m = 0 \\ u_1^m = (\lambda_1 - 1)L_x^m \ , \forall \ x^m = L_x^m \\ u_2^m = (\lambda_2 - 1)L_y^m \ , \forall \ y^m = L_y^m \\ u_3^m = (\lambda_3 - 1)L_z^m \ , \forall \ z^m = L_z^m \end{cases} , \tag{10}$$

where u_i^m $(i = 1, 2, 3)$ are the principal displacement components. The principal stretches λ_i $(i = 1, 2, 3)$ and the corresponding principal deformation gradient components F_i $(i = 1, 2, 3)$ are selected as the loading parameters, namely:

$$\{\lambda_1, \lambda_2, \lambda_3\}^{[\alpha,\beta]} = \{F_1, F_2, F_3\}^{[\alpha,\beta]} = \boldsymbol{F}_\alpha^{pre} \frac{\beta}{N_S} \ (\alpha = 1, \dots, N_L; \beta = 1, \dots, N_S) \ , \tag{11}$$

where α represents the loading path index (with maximum value N_L), and β indexes the steps in each loading history (with maximum value N_S). The pre-training deformation gradient $\boldsymbol{F}_\alpha^{pre}$ defines the loading direction corresponding to path α in the principal strain space. To satisfy the i.i.d. (independent and identically distributed) condition for comprehensive material characterization, these loading directions are generated to achieve a uniform distribution over the principal stress space. In our present work, as shown in Table 1, all 14 loading paths are required to obtain the principal deformation gradient $\{F_1, F_2, F_3\}^{[\alpha,\beta]}$ for generating instantaneous hyperelastic data, and the 7 paths marked with * are used to obtain the principal deformation gradient $\{F_1^*, F_2^*, F_3^*\}^{[\alpha,\beta]}$ for generating time-dependent viscous data. To account for viscous dissipation in both the deviatoric and volumetric responses, the pre-training loading paths listed in Table 1 are not volume-conserving. When nearly incompressible materials such as Sil 950 are considered, the initial bulk modulus is assumed to be approximately 100 times the initial shear modulus (corresponding to a Poisson's ratio of about 0.495). This assumption ensures that the material response can still be evaluated under the predefined pre-training loading paths using hybrid elements. Note that for very fast or very slow loading processes, the material response reduces to general finite elastic behavior, which is typically captured by classical hyperelasticity models [31]. Therefore, for a viscoelasticity subjected to any nonlinear loading, when the loading time is relatively short compared to the relaxation time (the loading time requirement in 14 paths), the viscoelastic mechanical behavior within each loading step can be approximately considered as an instantaneous hyperelastic response (with $N_L = 14$ and $N_S = 500$). Theoretically, reducing the loading duration helps isolate the instantaneous response by minimizing time-dependent effects. Taking advantage of computational simulation capabilities, we implemented a very short loading time of 0.001s for generating instantaneous hyperelastic data — a setting that is generally applicable for most viscoelastic materials. Conversely, when the loading time is relatively long compared to the relaxation time (the loading time requirement in 7 paths with * designated), the time-dependence of viscoelasticity must be considered in its mechanical response within each loading step. To properly train a data-driven model on the relaxation phenomenon of viscoelasticity [84], the training dataset is desired to include the complete temporal evolution of material responses for each loading path α, from initial loading through to steady-state equilibrium. While theoretical considerations might suggest infinite loading time are necessary, our mechanics-informed data-driven framework demonstrates that extending the loading duration beyond the characteristic relaxation time in the 7 paths (marked *) suffices to capture the essential viscoelastic responses, as detailed in Section 3.3.2. For these 7 loading paths, two distinct loading protocols are implemented. The first protocol involves uniformly distributing the total loading duration across 500 steps (with $N_L = 7$ and $N_S = 500$), analogous to the method used for acquiring instantaneous hyperelastic data but with an extended loading time. The second protocol is the relaxation test, which employs a transient loading within the initial 1% of loading steps (5 steps), followed by maintaining the constant load for the remaining 495 steps. This approach specifically captures stress relaxation behavior, thereby enhancing the network's capability to learn the associated viscous response characteristics. It is worth noting that when the loading time used in the training dataset to characterize the viscous behavior is sufficiently long, the initially prescribed value of N_S can be moderately increased to ensure that the trained convolution kernels remain sufficiently smooth and accurate. A more detailed discussion of this aspect is provided in Section 4.1.

Table 1
Pre-training loading paths α and associated direction vectors $\boldsymbol{F}_\alpha^{pre}$ for numerical data generation. All 14 loading paths are utilized to generate hyperelastic data (the loading time is relatively short compared to the relaxation time), with the 7 paths with * designated for generating viscous data (the loading time is relatively long compared to the relaxation time).

Path ID (α)	Loading direction ($\boldsymbol{F}_\alpha^{pre}$)	Path ID (α)	Loading direction ($\boldsymbol{F}_\alpha^{pre}$)
1	{ 0.785, 0.750, 0.625}	8*	{ 1.375, 0.785, 0.750}
2*	{ 1.000, 0.750, 0.565}	9	{ 1.375, 1.215, 0.750}
3	{ 1.000, 1.000, 0.500}	10*	{ 1.375, 1.250, 0.785}
4*	{ 1.215, 0.750, 0.625}	11	{ 1.375, 1.250, 1.215}
5	{ 1.250, 0.785, 0.625}	12*	{ 1.435, 1.000, 0.750}
6*	{ 1.250, 1.000, 0.565}	13	{ 1.435, 1.250, 1.000}
7	{ 1.250, 1.215, 0.625}	14*	{ 1.500, 1.000, 1.000}

First, the critical mechanical features required for representing instantaneous hyperelastic behaviors can be generated through the above 14 loading paths. Based on homogenization theory [89–91], the homogenized principal left Cauchy–Green tensor $\{B_1^M, B_2^M, B_3^M\}^{[\alpha,\beta]}$ can be obtained as:

$$\{B_1^M, B_2^M, B_3^M\}^{[\alpha,\beta]} = \{F_1^2, F_2^2, F_3^2\}^{[\alpha,\beta]}, \tag{12}$$

and the energy-conjugate instantaneous Kirchhoff stress $\{\tau_{h,1}^M, \tau_{h,2}^M, \tau_{h,3}^M\}^{[\alpha,\beta]}$can be expressed as:

$$\begin{cases} \tau_{h,1}^M = L_x^m \int_0^{L_y^m} \int_0^{L_z^m} t_1^m dy^m dz^m \\ \tau_{h,2}^M = L_y^m \int_0^{L_x^m} \int_0^{L_z^m} t_2^m dx^m dz^m \\ \tau_{h,3}^M = L_z^m \int_0^{L_x^m} \int_0^{L_y^m} t_3^m dx^m dy^m \end{cases} , \tag{13}$$

where t_1^m, t_2^m and t_3^m represent the surface traction force components imposed on the RVE's boundary. Therefore, the deviatoric part and volumetric part of homogenized principal instantaneous Kirchhoff stress, $\boldsymbol{\tau}_{h,dev}^M$ and $\boldsymbol{\tau}_{h,vol}^M$, can be obtained from Eq. (2).

Similarly, the critical features required for representing time -dependent viscous behaviors can be generated through the 7 loading paths with * designated. The time-dependent viscous Kirchhoff stress $\{\tau_{v,1}^M, \tau_{v,2}^M, \tau_{v,3}^M\}^{[\alpha,\beta]}$ can be obtained through an expression similar to Eq. (13), simply by replacing traction force components t_i^m calculated by the 14 loading paths to the t_i^{*m} calculated by the 7 loading paths with * designated. A critical preprocessing requirement is the homogenization of all RVE-derived physical quantities prior to fMDN training. For notational simplicity, we hereafter omit the superscript "M" denoting homogenized quantities unless explicitly required for disambiguation. To comprehensively describe the viscous responses as expressed in Eq. (3), $J^{[\alpha,\beta]}$ and the homogenized principal variables $\boldsymbol{F}^{[\alpha,\beta]}$, $\overline{\boldsymbol{F}}^{[\alpha,\beta]}$

(including their inverse and transpose) can be computed as:

$$\begin{cases} J^{[\alpha,\beta]} = \det(\boldsymbol{F}^{[\alpha,\beta]}) = F_1^{[\alpha,\beta]} F_2^{[\alpha,\beta]} F_3^{[\alpha,\beta]} \\ (\boldsymbol{F}^T)^{[\alpha,\beta]} = \boldsymbol{F}^{[\alpha,\beta]},\ (\boldsymbol{F}^{-1})^{[\alpha,\beta]} = (\boldsymbol{F}^{[\alpha,\beta]})^{-1}\ , \\ \overline{\boldsymbol{F}}^{[\alpha,\beta]} = (\overline{\boldsymbol{F}}^T)^{[\alpha,\beta]} = (J^{[\alpha,\beta]})^{-\frac{1}{3}} \boldsymbol{F}^{[\alpha,\beta]},\ (\overline{\boldsymbol{F}}^{-1})^{[\alpha,\beta]} = (J^{[\alpha,\beta]})^{\frac{1}{3}} (\boldsymbol{F}^{[\alpha,\beta]})^{-1} \end{cases} \tag{14}$$

Hence, the homogeneous principal values of relative distortional deformation gradient $\overline{\boldsymbol{F}}_t(t-t')$ and its inverse and transpose can be computed by the following form:

$$\begin{cases} \overline{\boldsymbol{F}}_{[\alpha,\beta]}([\alpha,\beta-\gamma]) = \overline{\boldsymbol{F}}^{[\alpha,\beta-\gamma]} \cdot (\overline{\boldsymbol{F}}^{-1})^{[\alpha,\beta]} \\ \overline{\boldsymbol{F}}^{-1}_{[\alpha,\beta]}([\alpha,\beta-\gamma]) = \overline{\boldsymbol{F}}^{[\alpha,\beta]} \cdot (\overline{\boldsymbol{F}}^{-1})^{[\alpha,\beta-\gamma]} (\gamma = 0,\dots,\beta-1)\ , \\ \overline{\boldsymbol{F}}^{-T}_{[\alpha,\beta]}([\alpha,\beta-\gamma]) = (\overline{\boldsymbol{F}}^{-1})^{[\alpha,\beta-\gamma]} \cdot \overline{\boldsymbol{F}}^{[\alpha,\beta]} \end{cases} \tag{15}$$

At this point, all the numerical dataset required for pre-training our data-driven viscoelastic model has been collected. It is important to emphasize that the original dataset requires only the principal deformation gradient $\boldsymbol{F}^{[\alpha,\beta]}$, instantaneous Kirchhoff stress $\boldsymbol{\tau}_h^{[\alpha,\beta]}$ (generated by 14 paths) and total Kirchhoff stress $\boldsymbol{\tau}^{[\alpha,\beta]}$ (generated by 7 paths with * designated). All other homogenized quantities can be derived from this original dataset through post-processing. The pre-training dataset is partitioned into training and validation sets through random sampling, with 85% allocated for model training and the remaining 15% reserved for validation across all loading paths.

3.2. The architecture of fMDN

The proposed fMDN combines different neural network architectures by coupling the nonlinear approximation strengths of ANNs with the feature learning capability of CNNs. The fMDN is tailored designed to model viscoelastic material behavior through carefully structured input–output mappings and mechanics-informed constraints, ensuring adherence to the memory decay property and thermodynamic consistency. As illustrated in Fig. 4, the fMDN architecture integrates an ANN to capture the instantaneous hyperelastic response of the material (light yellow region) along with two distinct TKL-CNNs that characterize the time-dependent viscous behavior (light blue and light green regions). First, the instantaneous dataset $\mathbb{T}$ obtained from 14 loading paths are used to train the ANN, with left Cauchy–Green tensor $\boldsymbol{B}^{[\alpha,\beta]}$ as the input and instantaneous Kirchhoff stress $\boldsymbol{\tau}_h^{[\alpha,\beta]}$ as the output. All outputs of ANN preceding step β are then post-processed into the deviatoric parts $\boldsymbol{\tau}_{h,dev}^{[\alpha,\beta-\gamma]}$ and volumetric parts $\boldsymbol{\tau}_{h,vol}^{[\alpha,\beta-\gamma]}$ $(\gamma = 0,\dots,\beta-1)$, which serve as partial inputs for two TKL-CNNs. These partial inputs are combined with the long-term dataset $\mathbb{T}^*$ acquired from 7 loading paths with * designated to train the two TKL-CNNs to represent the viscous responses, $\boldsymbol{\tau}_{v,dev}^{[\alpha,\beta]}$ and $\boldsymbol{\tau}_{v,vol}^{[\alpha,\beta]}$. Finally, the sum of the outputs of two TKL-CNNs and ANN, which constitutes the fMDN's outputs, represents the total Kirchhoff stress $\boldsymbol{\tau}^{[\alpha,\beta]}$. This section focuses exclusively on the forward propagation process of the fMDN model. The optimization methodology for trainable parameters and implementation of mechanics-informed constraints will be presented in Section 3.3. From this point onward, all the quantities with a superscript "∧" represent output values predicted by neural networks.

3.2.1. ANN-based instantaneous hyperelastic branch

ANNs have been widely used to capture the mechanical behavior of hyperelastic materials by establishing nonlinear mappings between strain measures and stress responses [92]. Motivated by this general approach, the present work employs an ANN-based constitutive model formulated in terms of an alternative pair of energy-conjugate variables. In particular, the left Cauchy–Green tensor and the Kirchhoff stress are adopted as the input and output variables, respectively, which are introduced in detail below.

As shown in the light yellow region in Fig. 4, the ANN architecture consists of 5 layers (one input layer, three hidden layers, and one output layer), with the number of neurons being 3,6,6,6 and 3, respectively. These hyperparameters are selected based on multiple rounds of empirical evaluation, aiming to achieve a balance between training efficiency, predictive accuracy, and generalization capability. Similar compact architectures have been shown in prior studies to be sufficient for accurately and efficiently capturing the nonlinear mappings involved in constitutive modeling with low-dimensional input and output variables [92–95]. The homogenized principal instantaneous Kirchhoff stress $\hat{\boldsymbol{\tau}}_h^{[\alpha,\beta]}$ as represented by the trainable function $\boldsymbol{f}_{\text{ANN}}$, with inputs $\boldsymbol{B}^{[\alpha,\beta]}$, weights $\boldsymbol{w}_{\text{ANN}}$ and biases $\boldsymbol{b}_{\text{ANN}}$, can be expressed as:

$$\begin{aligned} \hat{\boldsymbol{\tau}}_h^{[\alpha,\beta]} &= \boldsymbol{f}_{\text{ANN}}(\boldsymbol{B}^{[\alpha,\beta]}; \boldsymbol{w}_{\text{ANN}}, \boldsymbol{b}_{\text{ANN}}) \\ &= \tanh(\tanh(\tanh(\boldsymbol{B}^{[\alpha,\beta]} \boldsymbol{w}^2_{\text{ANN}} + \boldsymbol{b}^2_{\text{ANN}})\boldsymbol{w}^3_{\text{ANN}} \\ &\quad + \boldsymbol{b}^3_{\text{ANN}})\boldsymbol{w}^4_{\text{ANN}} + \boldsymbol{b}^4_{\text{ANN}})\boldsymbol{w}^5_{\text{ANN}} + \boldsymbol{b}^5_{\text{ANN}}\ , \end{aligned} \tag{16}$$

where $\boldsymbol{w}^i_{\text{ANN}}$ and $\boldsymbol{b}^i_{\text{ANN}}$ $(i = 2,3,4,5)$ are the weights and biases for the link between $(i-1)$th layer and ith layer, respectively. The activation function between each layer is the hyperbolic tangent function $\tanh(\cdot)$. The weights and biases of ANN are the trainable variables.

3.2.2. TKL-CNNs-based time-dependent viscous branch

After completing the training of the ANN, we obtain the output instantaneous Kirchhoff stress $\hat{\boldsymbol{\tau}}_h^{[\alpha,\beta]}$ represented by $\boldsymbol{f}_{\text{ANN}}$. By applying Eq. (2), $\hat{\boldsymbol{\tau}}_h^{[\alpha,\beta]}$ can be decomposed into the deviatoric part $\hat{\boldsymbol{\tau}}_{h,dev}^{[\alpha,\beta]}$ and the volumetric part $\hat{\boldsymbol{\tau}}_{h,vol}^{[\alpha,\beta]}$. Similarly, all values of $\hat{\boldsymbol{\tau}}_{h,dev}^{[\alpha,\beta-\gamma]}$ and $\hat{\boldsymbol{\tau}}_{h,vol}^{[\alpha,\beta-\gamma]}$ $(\gamma = 0,\dots,\beta-1)$ preceding step β can be obtained, which will serve as part of the inputs for TKL-CNNs. While TKL-CNN1 captures the deviatoric viscous Kirchhoff stress $\boldsymbol{\tau}_{v,dev}(t)$ and TKL-CNN2 represents its volumetric counterpart $\boldsymbol{\tau}_{v,vol}(t)$, we primarily details the architecture of TKL-CNN1, as TKL-CNN2 follows an analogous implementation.

As defined in Eq. (3a), the deviatoric viscous Kirchhoff stress $\boldsymbol{\tau}_{v,dev}(t)$ is formulated through a convolution operation between the push-forward transformed deviatoric instantaneous stress, $\overline{\boldsymbol{F}}_t^{-1}(t-t') \cdot \boldsymbol{\tau}_{h,dev}(t-t') \cdot \overline{\boldsymbol{F}}_t^{-T}(t-t')$, and a convolution function $\alpha_G \dot{g}_R(t')$. To model this relationship, TKL-CNN1 employs a convolutional layer (with a maximum kernel length N_S), which inherently approximates $\alpha_G \dot{g}_R(t')$ while learning the mapping between the transformed stress history and $\boldsymbol{\tau}_{v,dev}(t)$. Since the data generation process yields only discrete variable values at each time increment, we discretize Eq. (3a) as follows:

$$\hat{\boldsymbol{\tau}}_{v,dev}^{[\alpha,\beta]} = \operatorname{dev}\Big[\sum_{\gamma=0}^{\beta-1} -\alpha_G \dot{g}_R^{[\alpha,\gamma]} \overline{\boldsymbol{F}}^{-1}_{[\alpha,\beta]}([\alpha,\beta-\gamma]) \cdot \hat{\boldsymbol{\tau}}_{h,dev}^{[\alpha,\beta-\gamma]} \cdot \overline{\boldsymbol{F}}^{-T}_{[\alpha,\beta]}([\alpha,\beta-\gamma])\Delta t\Big], \tag{17}$$

where Δt is the time increment, and $\hat{\boldsymbol{\tau}}_{h,dev}^{[\alpha,\beta-\gamma]}$ can be obtained from the outputs of trained ANN. Hence, the TKL-CNN1 inputs, $\overline{\boldsymbol{F}}^{-1}_{[\alpha,\beta]}([\alpha,\beta-\gamma]) \cdot \hat{\boldsymbol{\tau}}_{h,dev}^{[\alpha,\beta-\gamma]} \cdot \overline{\boldsymbol{F}}^{-T}_{[\alpha,\beta]}([\alpha,\beta-\gamma])\Delta t$ $(\gamma = 0,\dots,\beta-1)$, are the push-forward transformed deviatoric instantaneous stress within the time increment Δt. The outputs, $\hat{\boldsymbol{\tau}}_{v,dev}^{[\alpha,\beta]}$, can be expressed by TKL-CNN1 as:

$$\begin{aligned} \hat{\boldsymbol{\tau}}_{v,dev}^{[\alpha,\beta]} = \boldsymbol{f}_{\text{TKL-CNN1}}\Big(&\overline{\boldsymbol{F}}^{-1}_{[\alpha,\beta]}([\alpha,\beta-\gamma]) \cdot \hat{\boldsymbol{\tau}}_{h,dev}^{[\alpha,\beta-\gamma]} \cdot \overline{\boldsymbol{F}}^{-T}_{[\alpha,\beta]}([\alpha,\beta-\gamma])\Delta t \\ &; k_{\text{TKL-CNN1}}(1,\dots,N_G)\Big), \end{aligned} \tag{18}$$

where $k_{\text{TKL-CNN1}}(1,\dots,N_G)$ denote the components of the trained convolution kernel $\boldsymbol{k}_{\text{TKL-CNN1}}$ of TKL-CNN1, and N_G represents the trained length (with $N_G \in [1, N_S]$) of $\boldsymbol{k}_{\text{TKL-CNN1}}$ that undergoes updating during each training iteration. Here, the trained mapping function

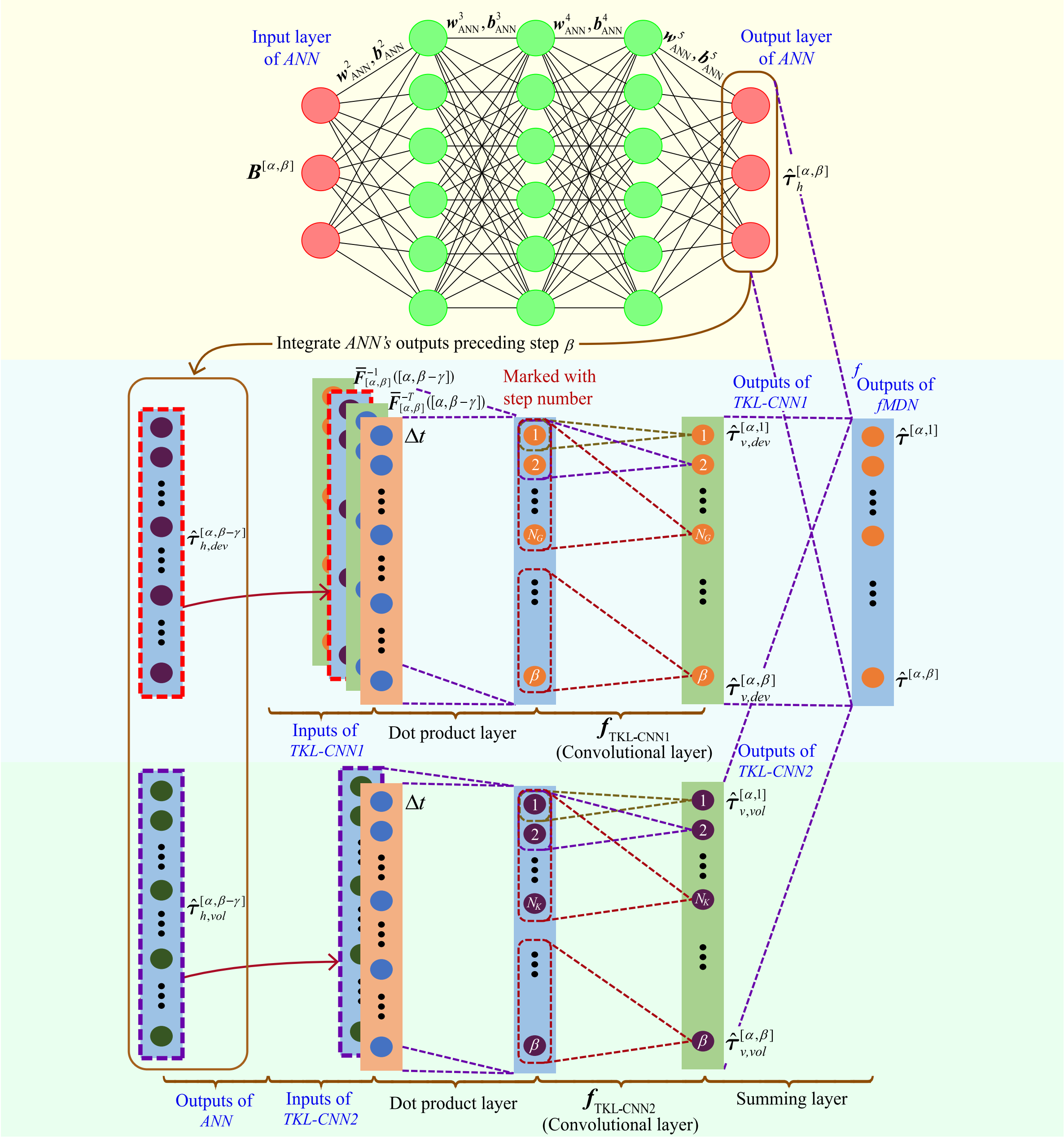

Fig. 4. The architecture of fMDN. First, in the yellow region, the instantaneous hyperelastic behavior is trained by ANN, with the left Cauchy–Green tensor, $\boldsymbol{B}^{[\alpha,\beta]}$, as input and the instantaneous Kirchhoff stress, $\boldsymbol{\tau}_h^{[\alpha,\beta]}$, as output. Next, all the outputs of ANN preceding step β serve as part of the inputs of the TKL-CNNs. In the blue and green regions, TKL-CNN1 and TKL-CNN2 are used to characterize the deviatoric and volumetric viscous responses, $\boldsymbol{\tau}_{v,dev}^{[\alpha,\beta]}$ and $\boldsymbol{\tau}_{v,vol}^{[\alpha,\beta]}$, respectively. The total Kirchhoff stress $\boldsymbol{\tau}^{[\alpha,\beta]}$ is then obtained by integrating the contributions from the ANN and the two TKL-CNN branches.

$\boldsymbol{f}_{\text{TKL}-\text{CNN1}}$ is expressed as:

$$\boldsymbol{f}_{\text{TKL}-\text{CNN1}} = \begin{cases} \displaystyle\sum_{\gamma=0}^{\beta-1} -k_{\text{TKL}-\text{CNN1}}(\beta-\gamma)\overline{\boldsymbol{F}}^{-1}_{[\alpha,\beta]}([\alpha,\gamma+1]) \cdot \hat{\boldsymbol{\tau}}_{h,dev}^{[\alpha,\gamma+1]} \cdot \overline{\boldsymbol{F}}^{-T}_{[\alpha,\beta]}([\alpha,\gamma+1])\Delta t, \\ \quad \text{if } \beta < N_G \\ \displaystyle\sum_{\gamma=0}^{N_G-1} -k_{\text{TKL}-\text{CNN1}}(\gamma+1)\overline{\boldsymbol{F}}^{-1}_{[\alpha,\beta]}([\alpha,\beta-\gamma]) \cdot \hat{\boldsymbol{\tau}}_{h,dev}^{[\alpha,\beta-\gamma]} \cdot \overline{\boldsymbol{F}}^{-T}_{[\alpha,\beta]}([\alpha,\beta-\gamma])\Delta t, \\ \quad \text{otherwise.} \end{cases} \tag{19}$$

When evaluating the output $\boldsymbol{\tau}_{v,dev}$ at step β, if $\beta < N_G$, only the first β values of $\boldsymbol{k}_{\text{TKL}-\text{CNN1}}$ and all inputs prior to step β are used; if $\beta \geq N_G$, the full kernel $\boldsymbol{k}_{\text{TKL}-\text{CNN1}}$ is applied to only the most recent N_G steps of inputs, reflecting that earlier inputs have negligible influence. Meanwhile, this approach ensures the dimensional consistency between inputs and outputs, analogous to zero-padding technique [73,96]. Physically, The trained convolution kernel $\boldsymbol{k}_{\text{TKL}-\text{CNN1}}$ corresponds to the product of the dimensionless initial relaxation shear modulus α_G and the time derivative of deviatoric relaxation function $\dot{g}_R$. Its optimized kernel length N_G corresponds to the characteristic number of incremental steps over which relaxation effects influence the deviatoric response. As shown in the light blue region of Fig. 4, the outputs of TKL-CNN1 directly correspond to the deviatoric component of viscous Kirchhoff stress $\boldsymbol{\tau}_{v,dev}$.

Similar to TKL-CNN1's architecture, TKL-CNN2 implements Eq. (3b) to map volumetric Kirchhoff stress history (can be obtained from the outputs of trained ANN) $\hat{\boldsymbol{\tau}}_{h,vol}^{[\alpha,\gamma]}\Delta t$ ($\gamma = 0,\ldots,\beta-1$) to $\hat{\boldsymbol{\tau}}_{v,vol}^{[\alpha,\beta]}$ via convolution with the trained kernel $\boldsymbol{k}_{\text{TKL-CNN2}}$, whose components $k_{\text{TKL-CNN2}}(1,\ldots,N_K)$ denote the trained convolution weights of TKL-CNN2, with a trainable length $N_K \in [1, N_S]$. Formally,

$$\hat{\boldsymbol{\tau}}_{v,vol}^{[\alpha,\beta]} = \boldsymbol{f}_{\text{TKL-CNN2}}\Big(\hat{\boldsymbol{\tau}}_{h,vol}^{[\alpha,\beta-\gamma]}\Delta t\ ; k_{\text{TKL-CNN2}}(1,\ldots,N_K)\Big), \tag{20}$$

where the trained mapping function $\boldsymbol{f}_{\text{TKL-CNN2}}$ is expressed as:

$$\boldsymbol{f}_{\text{TKL-CNN2}} = \begin{cases} \displaystyle\sum_{\gamma=0}^{\beta-1} -k_{\text{TKL-CNN2}}(\beta-\gamma)\hat{\boldsymbol{\tau}}_{h,vol}^{[\alpha,\gamma+1]}\Delta t, & \text{if } \beta < N_K \\ \displaystyle\sum_{\gamma=0}^{N_K-1} -k_{\text{TKL-CNN2}}(\gamma+1)\hat{\boldsymbol{\tau}}_{h,vol}^{[\alpha,\beta-\gamma]}\Delta t, & \text{otherwise.} \end{cases} \tag{21}$$

Notably, the trained convolution kernels $\boldsymbol{k}_{\text{TKL-CNN2}}$ corresponds to the product of the dimensionless initial relaxation bulk modulus α_K and the time derivative of volumetric relaxation function $\dot{k}_R$. Its optimized kernel length N_K corresponds to the characteristic number of incremental steps over which relaxation effects influence the volumetric response. Consequently, as shown in the light green region of Fig. 4, the trained output of TKL-CNN2 represents the volumetric component of viscous Kirchhoff stress $\boldsymbol{\tau}_{v,vol}$.

As such, the viscous Kirchhoff stress $\hat{\boldsymbol{\tau}}_v^{[\alpha,\beta]}$ is computed as a linear superposition of the outputs from TKL-CNN1 and TKL-CNN2, expressed as:

$$\begin{aligned}\hat{\boldsymbol{\tau}}_v^{[\alpha,\beta]} =& \boldsymbol{f}_{\text{TKL-CNN1}}\ \Big(\ \overline{\boldsymbol{F}}_{[\alpha,\beta]}^{-1}([\alpha,\beta-\gamma])\cdot\hat{\boldsymbol{\tau}}_{h,dev}^{[\alpha,\beta-\gamma]}\cdot\overline{\boldsymbol{F}}_{[\alpha,\beta]}^{-T}([\alpha,\beta-\gamma])\Delta t\ ; \\ & k_{\text{TKL-CNN1}}(1,\ldots,N_G)\ \Big)\ + \\ & \boldsymbol{f}_{\text{TKL-CNN2}}\Big(\hat{\boldsymbol{\tau}}_{h,vol}^{[\alpha,\beta-\gamma]}\Delta t\ ; k_{\text{TKL-CNN2}}(1,\ldots,N_K)\Big).\end{aligned} \tag{22}$$

Finally, by summing Eqs. (16) and (22), the total Kirchhoff stress $\hat{\boldsymbol{\tau}}^{[\alpha,\beta]}$ can be expressed as:

$$\begin{aligned}\hat{\boldsymbol{\tau}}^{[\alpha,\beta]} = \boldsymbol{f}_{\text{fMDN}}\ \Big(&\overline{\boldsymbol{F}}_{[\alpha,\beta]}^{-1}([\alpha,\beta-\gamma])\cdot\hat{\boldsymbol{\tau}}_{h,dev}^{[\alpha,\beta-\gamma]}\cdot\overline{\boldsymbol{F}}_{[\alpha,\beta]}^{-T}([\alpha,\beta-\gamma])\Delta t, \hat{\boldsymbol{\tau}}_{h,vol}^{[\alpha,\beta-\gamma]}\Delta t; \\ & \boldsymbol{w}_{\text{ANN}}, \boldsymbol{b}_{\text{ANN}}, N_G, N_K, \boldsymbol{k}_{\text{TKL-CNN1}}, \boldsymbol{k}_{\text{TKL-CNN2}}\ \Big),\end{aligned} \tag{23}$$

where all parameters $\{\boldsymbol{w}_{\text{ANN}}, \boldsymbol{b}_{\text{ANN}}, N_G, N_K, \boldsymbol{k}_{\text{TKL-CNN1}}, \boldsymbol{k}_{\text{TKL-CNN2}}\}$ are trainable within the fMDN architecture. The global network function $\boldsymbol{f}_{\text{fMDN}}$, encapsulating the full forward propagation operation, is formulated as:

$$\boldsymbol{f}_{\text{fMDN}} = \boldsymbol{f}_{\text{ANN}} + \boldsymbol{f}_{\text{TKL-CNN1}} + \boldsymbol{f}_{\text{TKL-CNN2}}. \tag{24}$$

3.3. Mechanics-informed pre-training of the fMDN

This section begins by outlining the training configuration, then develops mechanics-informed constraints for the trained viscoelastic constitutive model to satisfy two fundamental physical principles: viscoelastic memory decay property and thermodynamic consistency.

The preliminary parameter update scheme in fMDN follows standard deep learning backpropagation schemes [97]. As discussed in Section 3.2.1, the first step in training fMDN is to optimize the ANN function Eq. (16), which allows us to obtain accurate outputs $\hat{\boldsymbol{\tau}}_h^{[\alpha,\beta]}$. These outputs will then be post-processed and used as partial inputs for two TKL-CNNs. Here, the trained variables of ANN, weights $\boldsymbol{w}_{\text{ANN}}$ and biases $\boldsymbol{b}_{\text{ANN}}$, are optimized by solving the following problem of minimizing the mean squared error (MSE) loss function:

$$\min_{\boldsymbol{w}_{\text{ANN}},\boldsymbol{b}_{\text{ANN}}} \sum_{[\alpha,\beta]\in\mathbb{T}}\sum_{i=1}^{3}(\tau_{h,i}^{[\alpha,\beta]} - \hat{\tau}_{h,i}^{[\alpha,\beta]})^2\ , \tag{25}$$

where $\tau_{h,i}^{[\alpha,\beta]}$ and $\hat{\tau}_{h,i}^{[\alpha,\beta]}$ respectively denote the computed principal component of the instantaneous Kirchhoff stress and their corresponding ANN predictions, evaluated across 14 distinct loading paths in $\mathbb{T}$.

Once accurate instantaneous Kirchhoff stress $\hat{\boldsymbol{\tau}}_h^{[\alpha,\beta]}$ (the outputs of ANN) are obtained, the training of the TKL-CNNs can then begin. The trainable variables in TKL-CNNs are first optimized via backpropagation using gradient descent methods [98], which minimizes the MSE loss function through the following optimization problem:

$$\min_{\substack{N_G,\boldsymbol{k}_{\text{TKL-CNN1}},\\ N_K,\boldsymbol{k}_{\text{TKL-CNN2}}}} \sum_{[\alpha,\beta]\in\mathbb{T}^*}\sum_{i=1}^{3}(\tau_i^{[\alpha,\beta]} - \hat{\tau}_i^{[\alpha,\beta]})^2\ , \tag{26}$$

where $\tau_i^{[\alpha,\beta]}$ and $\hat{\tau}_i^{[\alpha,\beta]}$ respectively denote the computed principal component of the total Kirchhoff stress and their corresponding fMDN predictions, evaluated across 7 distinct loading paths in $\mathbb{T}^*$.

Building upon the above conventional deep learning parameter update schemes, our methodology incorporates a subsequent mechanics-informed constrained optimization phase to ensure compliance with viscoelastic memory decay property and thermodynamic consistency.

3.3.1. Mechanics-informed constraints: Viscoelastic memory decay property

As prescribed by Eq. (4), the viscoelastic memory decay property requires strict constraints on the relaxation function derivatives $\dot{g}_R$ and $\dot{k}_R$. Physically, the dimensionless initial moduli α_G and α_K are positive constants, and their multiplicative relationship with $\dot{g}_R$ and $\dot{k}_R$ preserves the same constraint structure of Eq. (4). This permits direct application of identical constraints to the convolution kernels $\boldsymbol{k}_{\text{TKL-CNN1}}$ and $\boldsymbol{k}_{\text{TKL-CNN2}}$, which denote the $\alpha_G\dot{g}_R$ and $\alpha_K\dot{k}_R$ respectively, as the following formulations:

$$k_{\text{TKL-CNN1}}(\eta{+}1) = \begin{cases} k_{\text{TKL-CNN1}}(\eta),\ \text{if } k_{\text{TKL-CNN1}}(\eta+1) > k_{\text{TKL-CNN1}}(\eta) \\ \text{Unchanged},\ \text{otherwise} \end{cases} \eta = 1,\ldots,N_S-1 \tag{27}$$

$$k_{\text{TKL-CNN2}}(\eta{+}1) = \begin{cases} k_{\text{TKL-CNN2}}(\eta),\ \text{if } k_{\text{TKL-CNN2}}(\eta+1) > k_{\text{TKL-CNN2}}(\eta) \\ \text{Unchanged},\ \text{otherwise} \end{cases} \eta = 1,\ldots,N_S-1. \tag{28}$$

The mechanics-informed evaluation criteria applied during the secondary updating procedure enforce the monotonicity constraint specified in Eq. (4) for each convolution kernel following preliminary optimization. This constrained updating procedure ensures that the trained fMDN correctly exhibits viscoelastic memory decay property.

3.3.2. Mechanics-informed constraints: Thermodynamic consistency

Consistent with the second law of thermodynamics (Eqs. (8)–(9)), the time derivatives of relaxation functions must obey thermodynamic consistency constraints. While Eqs. (27)–(28) ensure compliance with Eqs. (8a) and (9a), the remaining conditions ((8b),(9b)) are enforced through direct constraints on the convolutional kernels as follows:

$$k_{\text{TKL-CNN1}}(\eta) = \begin{cases} 0,\ \text{if } k_{\text{TKL-CNN1}}(\eta) < 0 \\ \text{Unchanged},\ \ \text{otherwise} \end{cases} \eta = 1,\ldots,N_S \tag{29}$$

$$k_{\text{TKL-CNN2}}(\eta) = \begin{cases} 0,\ \text{if } k_{\text{TKL-CNN2}}(\eta) < 0 \\ \text{Unchanged},\ \ \text{otherwise} \end{cases} \eta = 1,\ldots,N_S. \tag{30}$$

The above mechanics-informed evaluation criterion guarantees non-negativity of all convolutional kernels. Furthermore, our data-driven framework is designed to fully characterize relaxation behavior, which requires asymptotic convergence of terminal kernel values, $k_{\text{TKL-CNN1}}(N_G)$ and $k_{\text{TKL-CNN2}}(N_K)$, to zero. Consequently, two mechanics-informed evaluation criteria are additionally enforced during the secondary updating procedure — the non-negativity of all convolutional kernels and the asymptotic decay of terminal kernel values to zero. In an idealized implementation, the terminal kernel values must decay below a prescribed tolerance threshold $\mathfrak{R}$ (defined as 0.01%

of their initial values, $k_{\mathrm{TKL-CNN1}}(1)$ and $k_{\mathrm{TKL-CNN2}}(1)$, respectively). The training procedure iteratively updates fMDN's trainable parameters until all convergence criteria are satisfied. Clearly, the adaptive kernel lengths N_G and N_K (bounded by N_S) are learned during training as the number of effective (nonzero) kernel components induced by the mechanics-informed constraints, thereby embodying the key feature of adaptive convolutional kernel lengths in the TKL-CNN architecture. Physically, this feature encodes relaxation effects that are confined to N_G steps for deviatoric responses and N_K steps for volumetric responses.

The resulting trained fMDN thus achieves thermodynamic consistency and demonstrates predictive capability extending beyond the temporal range of its training data, as justified by neglecting kernel contributions below the $\mathfrak{R}$ threshold. Moreover, it is particularly noteworthy that when either long-term viscoelastic behavior need not be considered (i.e., when the structural operational time does not exceed the material-specific relaxation time), or practical constraints prevent acquisition of long-term data, we can substantially reduce both data acquisition and computational costs by aligning the RVE loading time for the 7 paths marked with * (as shown in Table 1) with the actual structural service time. Remarkably, this numerical loading time setting achieves reliable predictive accuracy within the operational time domain without necessitating the strict convergence criterion, which typically requires $k_{\mathrm{TKL-CNN1}}(N_G)$ and $k_{\mathrm{TKL-CNN2}}(N_K)$ to be smaller than $\mathfrak{R}$, as is typically needed for full-time viscoelastic characterization. This adaptable numerical loading time configuration further demonstrates the inherent flexibility of our data-driven methodology, which enables customized generation of a numerical training dataset tailored to specific operational requirements.

Given the Sil 950 material's characteristic relaxation time (approximately 1.5 h), complete characterization of this relaxation behavior through fMDN would theoretically require equivalent RVE loading time for the 7 paths with * designated. However, to optimize computational efficiency while demonstrating methodological flexibility, we aligned the RVE loading time of these 7 paths with the maximum experimental duration (2500 s) from subsequent SFA tests. The numerical dataset generated by the RVE is used for fMDN pre-training. It is noteworthy that Sil 950 exhibits nearly incompressible behavior, which means the viscous effects are negligible in the volumetric response. Accordingly, the fMDN training incorporates only the deviatoric part of viscous response by training $\boldsymbol{k}_{\mathrm{TKL-CNN1}}$ with $N_G = 500$, while explicitly neglecting the volumetric viscous dissipation by constraining $\boldsymbol{k}_{\mathrm{TKL-CNN2}} = 0$ with $N_K = 1$. The fMDN pre-training phase reaches completion upon simultaneous satisfaction of all mechanics-derived constraints and optimization objectives.

4. Fine-tuning stage

In this section, the implementation of the fine-tuning stage is presented. Specifically, Section 4.1 describes how the trained fMDN model is integrated into the FEM framework to enable the data-driven model to solve complex boundary value problems. Section 4.2 then presents a series of SFA experiments designed to evaluate and quantify the discrepancy between the experimental measurements and the predictions of the pre-trained fMDN model. Finally, Section 4.3 introduces an intelligent algorithm for further fine-tuning the fMDN, thereby enabling high-accuracy prediction of the viscoelastic behavior of the SFA under various operating conditions.

From this point onward, the pre-trained fMDN framework is no longer constrained by the pretraining loading paths α or the loading steps β, and can directly operate on experimental data to establish the input–output mapping. Accordingly, the superscript notation $[\alpha, \beta]$ is no longer used, allowing for a more general formulation of the fMDN framework.

4.1. Computational mechanics with the trained fMDN model

With the trained data-driven model $\boldsymbol{f}_{\mathrm{fMDN}}$ established, stress state updates can be computed via standard return mapping algorithm. Provided that all the historic responses (such as Kirchhoff stress $\boldsymbol{\tau}$, deformation gradient $\boldsymbol{F}$, and incremental time Δt) preceding step n, are known, the following procedure is employed. In the trained fMDN model, the relaxation kernels are uniformly discretized with respect to fixed time increments. However, variable time stepping is typically adopted in numerical simulations. Therefore, the non-uniform time-step history needs to be mapped onto a fixed kernel time grid Δt_f. Specifically, if the elapsed time $t^n - t^i$ between the current loading step n and the ith historical loading increment falls between $j\Delta t_f$ and $(j+1)\Delta t_f$, the corresponding viscoelastic kernel value is obtained through linear interpolation between the trained kernel values at the jth and $(j+1)$th grid points. To mitigate interpolation-induced numerical errors, the initial convolution kernel length in the fMDN training process can be intentionally chosen to be sufficiently large. This leads to a smaller fixed loading interval Δt_f in the training dataset. In addition, to preserve and recover the original trained network parameters as accurately as possible, the maximum allowable increment size in the numerical simulation is chosen to be no longer than Δt_f. As a result, the kernel values obtained through interpolation in numerical simulations become smoother, ensuring numerical consistency between the adaptive time stepping adopted in the simulations and the uniformly discretized relaxation kernel used in the constitutive model.

From this point onward, to maintain the simplicity in our notation, all variables that are not explicitly labeled with a step index in the superscript will represent the variables at step n, except when necessary for clarity. The preceding algorithm defines the predicted Kirchhoff stress $\hat{\boldsymbol{\tau}}$ by Eq. (23), the associated tangent modulus $\boldsymbol{D}^M$ can be calculated by the following:

$$
\begin{aligned}
\boldsymbol{D}^M &= 2\frac{\partial\left(\boldsymbol{F}^{-1}\right)^n \cdot \hat{\boldsymbol{\tau}}^n \cdot \left(\boldsymbol{F}^{-T}\right)^n}{\partial \boldsymbol{C}^n} \\
&= 2\left(\frac{\partial \boldsymbol{F}^{-1}}{\partial \boldsymbol{F}} : \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{C}} \cdot \hat{\boldsymbol{\tau}}^n \cdot \boldsymbol{F}^{-T} + \boldsymbol{F}^{-1} \cdot \frac{\partial \hat{\boldsymbol{\tau}}^n}{\partial \boldsymbol{F}^n} : \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{C}} \cdot \boldsymbol{F}^{-T}\right. \\
&\quad \left. + \boldsymbol{F}^{-1} \cdot \hat{\boldsymbol{\tau}}^n \cdot \frac{\partial \boldsymbol{F}^{-T}}{\partial \boldsymbol{F}} : \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{C}}\right),
\end{aligned}
\tag{31}
$$

where $\boldsymbol{C} = \boldsymbol{F}^T \cdot \boldsymbol{F}$ is the right Cauchy–Green tensor at step n, $\hat{\boldsymbol{\tau}}^n$ is the predicted Kirchhoff stress calculated by $\boldsymbol{f}_{\mathrm{fMDN}}$ at step n. Here, $\frac{\partial \boldsymbol{F}^{-1}}{\partial \boldsymbol{F}}$ and $\frac{\partial \boldsymbol{F}^{-T}}{\partial \boldsymbol{F}}$ can be derived as follows:

$$
\begin{cases}
\dfrac{\partial F^{-1}_{ij}}{\partial F_{kl}} = -F^{-1}_{ik} F^{-1}_{lj} \\[2ex]
\dfrac{\partial F^{-T}_{ij}}{\partial F_{kl}} = -F^{-1}_{jk} F^{-1}_{li}.
\end{cases}
\tag{32}
$$

Additionally, the detailed derivation of $\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{C}}$ in Eq. (31) is given in Appendix B. $\frac{\partial \hat{\boldsymbol{\tau}}^n}{\partial \boldsymbol{F}^n}$ in Eq. (31) can be calculated with the partial derivative of fMDN, as follows:

$$
\frac{\partial \hat{\boldsymbol{\tau}}^n}{\partial \boldsymbol{F}^n} = \frac{\partial \boldsymbol{f}_{\mathrm{fMDN}}}{\partial \boldsymbol{F}^n} = \frac{\partial \boldsymbol{f}_{\mathrm{ANN}}}{\partial \boldsymbol{F}} + \frac{\partial \boldsymbol{f}_{\mathrm{TKL-CNN1}}}{\partial \boldsymbol{F}^n} + \frac{\partial \boldsymbol{f}_{\mathrm{TKL-CNN2}}}{\partial \boldsymbol{F}^n},
\tag{33}
$$

where the three terms on the far right can be obtained using the backpropagation algorithm, with the derivation provided in Appendix C.

Consequently, the closed-form expression for the tangent modulus $\boldsymbol{D}^M$ can be derived from the aforementioned computational results. According to the Holzapfel [99], tangent modulus based on Truesdell rate $\boldsymbol{D}^T$ can be obtained by pushing forward of $\boldsymbol{D}^M$ as follows:

$$
D^T_{ijkl} = \frac{1}{J} F_{im} F_{jn} F_{kp} F_{lq} D^M_{mnpq}.
\tag{34}
$$

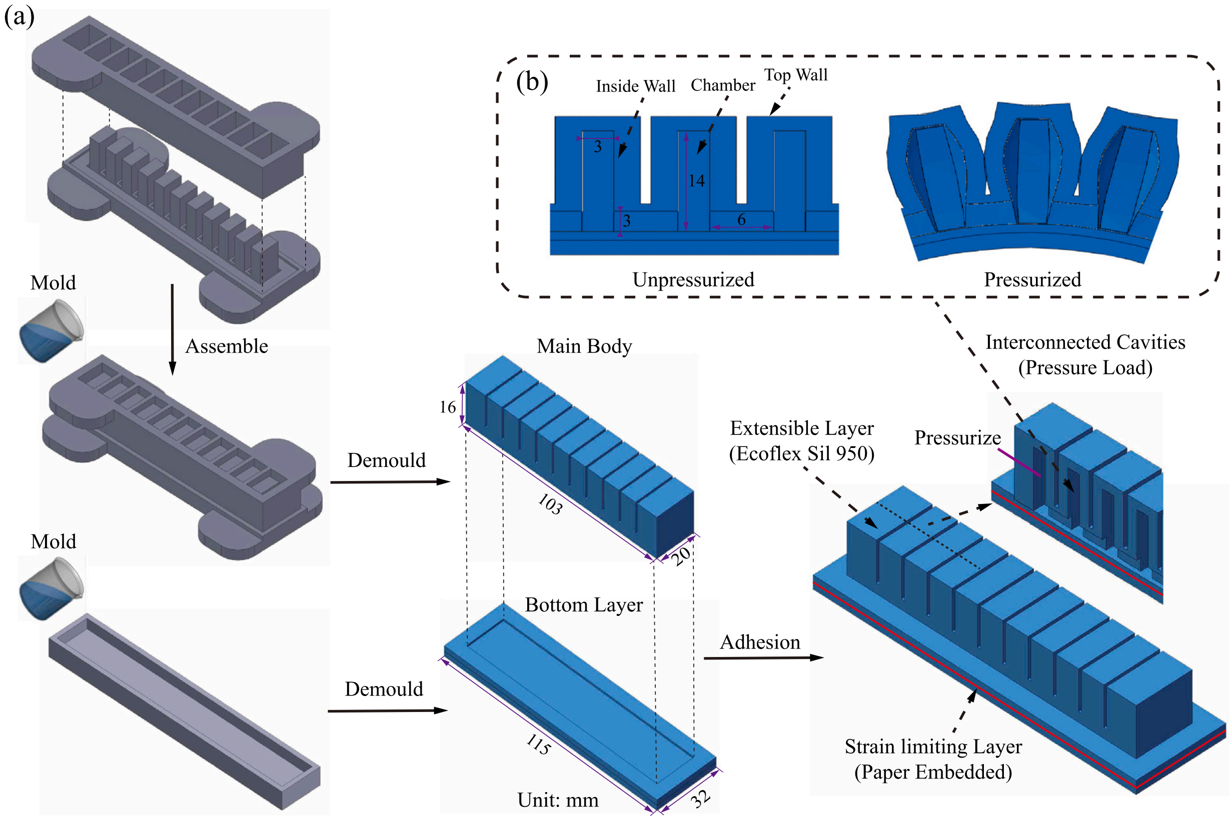

Fig. 5. (a) The fabrication process and geometry of SFA. The extensible body (Ecoflex Sil 950) and the strain limiting layer (Paper Embedded) are also identified. (b) The interconnected unpressurized/pressurized cavities of SFA.

The Jaumann-rate tangent modulus C^J can be derived through the following transformation:

$$D^J_{ijkl} = D^T_{ijkl} + \frac{1}{2J}(\tau_{il}\delta_{jk} + \tau_{jl}\delta_{ik} + \tau_{ik}\delta_{jl} + \tau_{jk}\delta_{il}) - \frac{1}{J}\tau_{ij}\delta_{kl}. \tag{35}$$

The equilibrium equation can be solved with these closed-form tangent moduli. For practical engineering applications, we have implemented this data-driven viscoelastic model as a UMAT subroutine in the ABAQUS finite element environment, without using additional numerical stabilization techniques.

4.2. Fine-tuning dataset preparation: Soft fluidic actuator experiments

A representative silicone rubber (Sil 950) and a strain limiting layer (paper embedded) were employed in the SFA, as illustrated in Fig. 5(a). The internally interconnected cavities of the actuator ensure that pressurization induces corresponding bending deformation, as shown in Fig. 5(b). The structure and fabrication process of the SFA were implemented using the Fast Penu-net (fPN) model proposed by Mosadegh et al. [23]. The time-dependent viscoelastic properties — a critically important yet frequently overlooked characteristic of SFAs — are taken into consideration to design the SFA experiments. Commonly, SFAs are pressurized by hydraulic or pneumatic actuator [100,101]. Here, we designed two SFA experiments for these conditions: a bending test (with a hydraulic actuator) and a reaction force relaxation test (with a pneumatic actuator). Notably, all the SFAs were subjected to preloaded cycles of loading-unloading prior to formal testing to eliminate the influence of the Mullins effect.

4.2.1. Bending test of the SFA

To access the bending deformation, as depicted in Fig. 6(a), the SFA was hydraulically pressurized through a syringe pump, with the internal pressure monitored in real-time via a digital pressure gauge connected to the SFA. In this test, the internal pressure was increased from 0 kPa to 100 kPa within 55 s, as shown in Fig. 7(a). Notably, to eliminate the effects of gravity and ensure that only hydraulic pressure remained inside the SFA, the entire SFA was submerged in a water tank. Additionally, before each measurement cycle, air was removed from all supply tubes and the SFA, and the pressure was calibrated to atmospheric pressure. Moreover, a camera was positioned at the SFA's distal end to capture its deformation throughout the actuation cycle. As illustrated in Fig. 6(b), both the unpressurized and pressurized configurations of the SFA are presented, demonstrating the significant deformation characteristics induced by hydraulic actuation. To quantitatively characterize the temporal evolution of SFA bending deformation, we extracted some discrete time-point configurations by tracking the arc-shaped variation along the lower surface of the strain-limiting layer. A coordinate system was established with its origin fixed at the left end. The dashed curves with triangular markers in Fig. 8(a)–(e) respectively represent the derived arc-shaped results at 15, 30, 35, 40, and 55 s of the bending test.

4.2.2. Reaction force relaxation test of the SFA

A prevalent approach in soft robotics involves utilizing SFAs as the fundamental building blocks for bio-inspired grippers capable of adaptive object manipulation. Given this widespread application, it is imperative to account for the time-dependent relaxation behavior of gripping forces attributable to the intrinsic viscoelastic properties of SFAs, which is critical for preventing accidental object release during prolonged grasping. To characterize this phenomenon, we developed a reaction force relaxation test that systematically characterizes the decay trend of reaction force under a constant input air volume condition. A load cell was employed to continuously monitor the reaction force of the SFA (with a separation distance of $\Delta H = 40$ mm) under coupled gravitational and pneumatic loading conditions. As shown in Fig. 6(c),

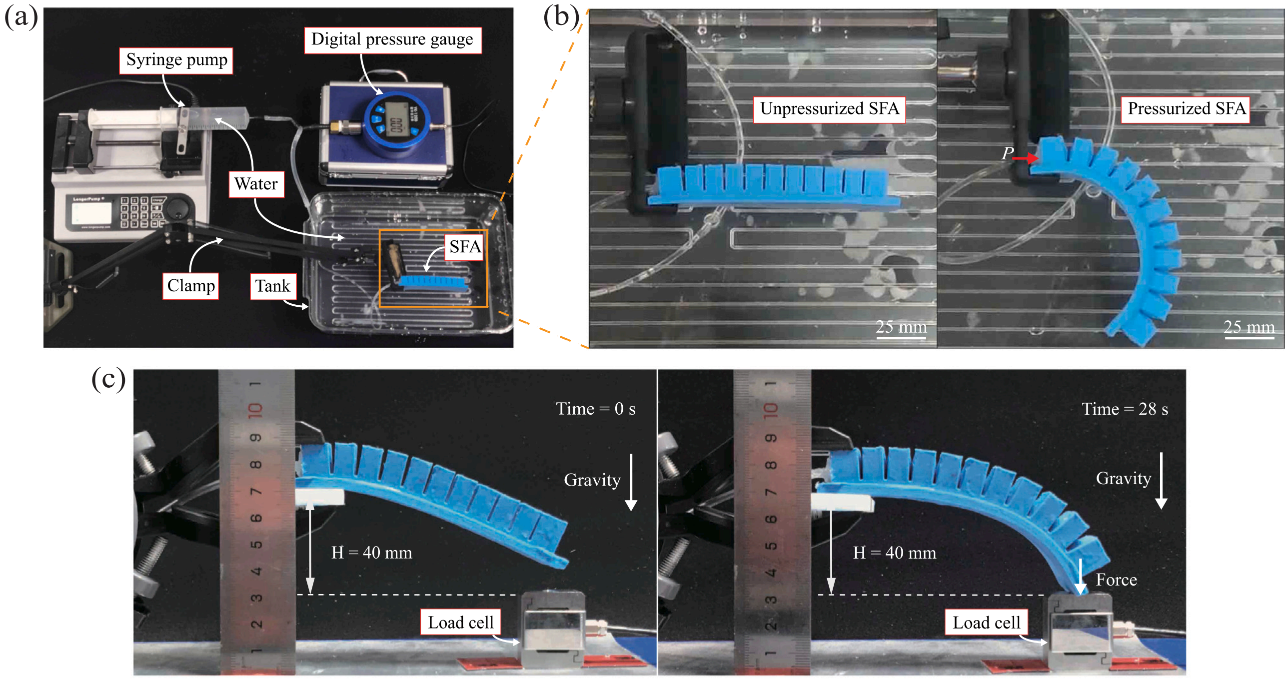

Fig. 6. Bending and relaxation tests of the SFA: (a) Experimental setup for the bending test; (b) SFA configurations in the bending test, showing the unpressurized (initial) and pressurized states; (c) SFA configurations in the relaxation test under the coupled effects of gravity and rapid inflation loading at $T = 0$ s (initial state) and $T = 28$ s.

at 0 s (initial state), the SFA hangs naturally under gravity alone, with its left end fixed and connected to the digital pressure gauge and syringe pump, while its right end remains out of contact with the load cell. Under the coupled effects of gravity and rapid inflation loading, the SFA gradually deforms, and at 28 s, its right end comes into contact with the load cell. The left end is fixed and connected to a syringe pump, imposing a rapid airflow rate of $\dot{V}/V_0 = 0.9$/min, with $\dot{V}$ denoting the cavity volume change rate and $\dot{V}$ the initial cavity volume. The entire inflation process was maintained for 28 s until $\Delta V/V_0 = 0.42$. Subsequently, the input air volume was maintained constant for 2400 s, and the decay trend of the reaction force returned by the load cell during this period was monitored, as represented by the gray-shaded region (95% confidence interval) in Fig. 8(f). Throughout the entire 2400-second relaxation period, reaction force measurements were systematically recorded at consistent 5-second intervals. It is worth emphasizing that the pre-training loading time (2500 s) set in Section 3.1 was based on an approximate value derived from the overall test duration of this experiment, which includes a rapid pressurization time of 28 s and a relaxation time of 2400 s.

4.3. fMDN fine-tuning with SFA experimental results

To systematically evaluate the predictive performance of the fMDN model in the aforementioned SFA experimental scenarios, we first establish corresponding finite element models that accurately replicate the experimental conditions. For the bending test, the main body and the bottom layer were meshed with three-dimensional four-node tetrahedral hybrid elements (C3D4H). The paper embedded within the strain limiting layer was modeled with a Young's modulus of 6500 MPa and a Poisson ratio of 0.2, and discretized using three-node triangular shell elements (S3). The upper and lower surfaces of the SFA, spanning from the leftmost end to the first fluid cavity, were fully fixed. A self-contact was employed between the outer surfaces of adjacent chambers. Hydraulic pressurization of the model was implemented through a fluid cavity interaction, with the fluid represented as a hydraulic medium of density 1000 kg/m^3 and bulk modulus of 2000 MPa. The same as the experimental conditions, the pressurization process was simulated by applying a pressure of 100 kPa over 55 s to the fluid cavity. Through post-processing, the arc-shaped profiles along the lower surface of the strain limiting layer can be extracted.

Similarly, a corresponding finite element model was established for the reaction force relaxation experiment, as illustrated in Fig. 7(b). The identical element types were employed for the SFA as those implemented in the bending simulation model. Additionally, a rigid body plate was positioned $\Delta H = 40$ beneath the SFA, which is meshed using three-dimensional 8-node elements (C3D8), to simulate the load cell. This plate was kinematically coupled to a fully fixed reference point, enabling the extraction of reaction forces at the reference point to characterize the time-dependent relaxation behavior. The upper and lower surfaces of the SFA, spanning from the leftmost end to the first fluid cavity, were fully constrained (consistent in both bending and force relaxation simulations). Two distinct contact interactions were incorporated into this model: a self-contact governing the interfacial interactions between the outer surfaces of adjacent chambers, and a surface-to-surface contact regulating the mechanical interaction between the SFA and the rigid plate. Moreover, a gravitational load of 9810 N/ton was uniformly applied to the entire SFA structure. In contrast to the bending case, incompressible air with a density of 1.225 kg/m^3 was considered. Inflation was achieved by controlling the cavity air volume via a fictitious thermal expansion induced by an imposed temperature change as follows:

$$\Delta V/V_0 = 3\alpha\Delta T, \tag{36}$$

where α is the fluid thermal expansion coefficient, and ΔT is the temperature change. Here, $\alpha = 1$ and $\Delta T = 0.14$.

With the finite element model of the SFA experiments now established and integrated with the stress update algorithm presented in Section 4.1, the predictive capability of the fMDN model can be evaluated through direct experimental-numerical comparison. The performance

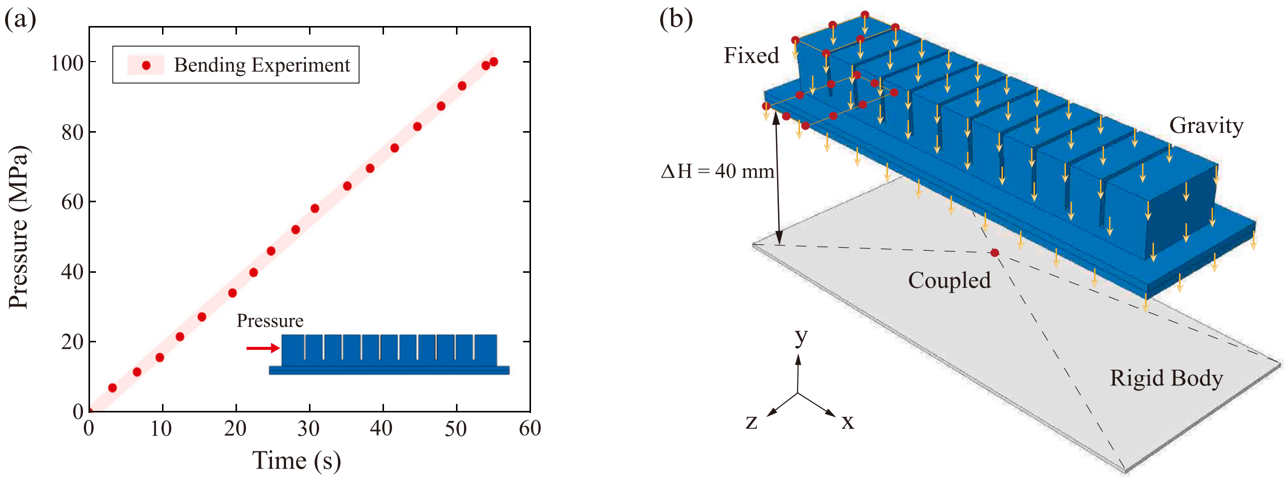

Fig. 7. The boundary conditions of the SFA simulation models: (a) The hydraulic pressure employed in the SFA bending test. The gray-shaded area presents the experimental results with 95% confidence intervals; (b) The boundary conditions of the reaction force relaxation test.

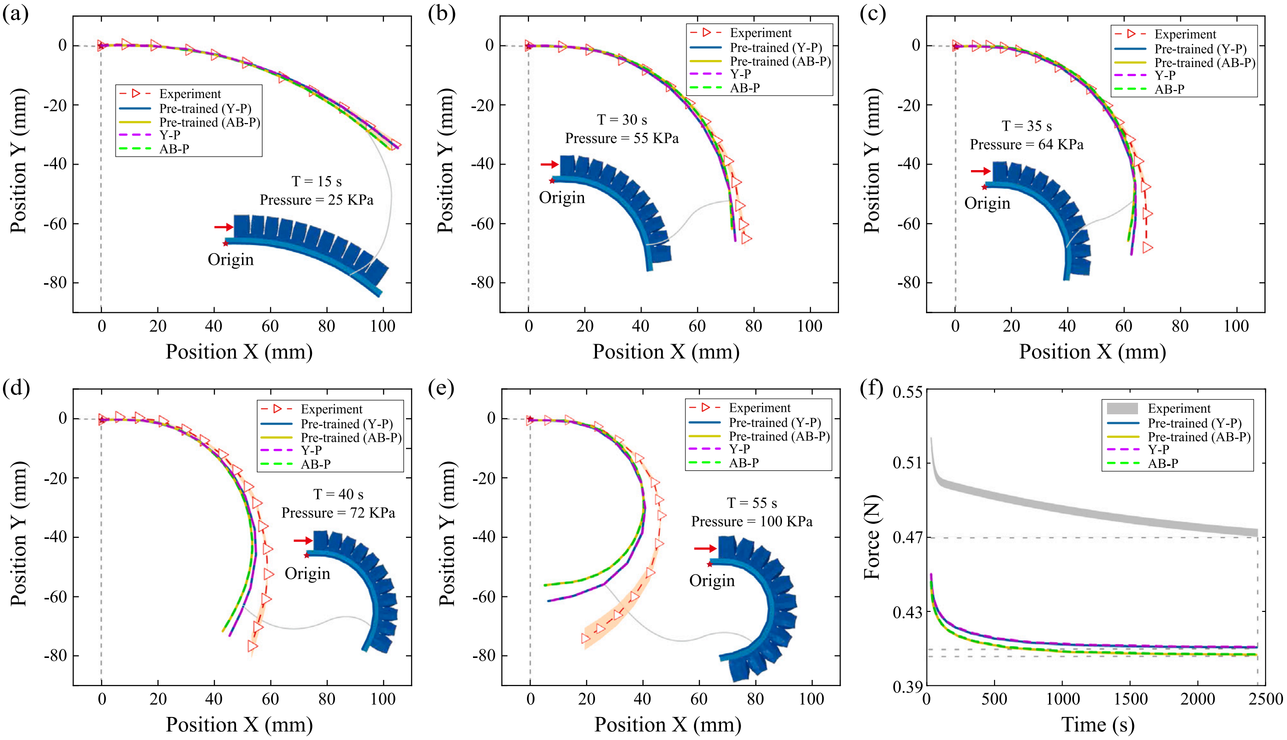

Fig. 8. Performance of the conventional viscoelastic constitutive models and the corresponding pre-trained fMDN models in SFA tests. Arc-shaped bending deformation at different pressures, with the corresponding deformation configurations of the pre-trained (Y-P) model also provided for each pressure level: (a) 25 kPa at 15 s; (b) 55 kPa at 30 s; (c) 64 kPa at 35 s; (d) 72 kPa at 40 s; (e) 100 kPa at 55 s. (f) The reaction force relaxation responses versus the loading time. The dashed red lines with triangular markers denote the experimental measurements, and all the shaded areas indicate the corresponding 95% confidence intervals. The dashed lines show the predictions of the conventional viscoelastic models preliminarily evaluated from the material-level experiments, with purple and green indicating the Yeoh + Prony (Y-P) and Arruda-Boyce + Prony (AB-P) models, respectively, while the solid blue and yellow lines represent the predictions of the corresponding pre-trained fMDN models.

of the conventional viscoelastic constitutive models preliminarily evaluated from the material-level experiments in Section 3.1.1 and the corresponding pre-trained fMDN models in the bending test of SFA is presented in Fig. 8(a)–(e). It can be observed that the pre-trained fMDN models exhibits high predictive accuracy at the early stage (25 kPa at 15 s). As the pressure increases, the discrepancy between the experimental measurements and the pre-trained predictions grows, with the predicted bending amplitude becoming increasingly larger than the experimental measurements. This indicates that the material properties represented by the pre-trained fMDN models are softer. Here, we select three discrete configurations of the bending experiment (30 s, 35 s, and 40 s) from the intermediate phase where the pre-trained fMDN models exhibit significant prediction errors to constitute the fine-tuning dataset $\mathbb{T}^{bend}$, while both preceding (15 s) and subsequent

(55 s) configurations are used as the validation set in the inference stage to evaluate the prediction accuracy of the corresponding fine-tuned fMDN models. This methodology establishes a robust validation framework at the inference stage, enabling the assessment of both the fine-tuned models' ability to retain predictive accuracy for early-stage configurations (15 s), which are well characterized by the pre-trained fMDN models, and their enhanced capability to predict configurations beyond the fine-tuning dataset (55 s), corresponding to the maximum error condition of the pre-trained fMDN models.

Additionally, the performance of the conventional viscoelastic constitutive models and the corresponding pre-trained fMDN models in the reaction force relaxation test of SFA is presented in Fig. 8(f). It can be seen that at the beginning of force relaxation stage, the pre-trained fMDN models' initial predictions are approximately 14% lower than the experimental measurements. This discrepancy is attributed to the model's inaccuracy in predicting short-term viscoelastic responses, which is also reflected in its performance during the bending experiments. Furthermore, the lower initial prediction also demonstrates that the material properties predicted by the pre-trained model are softer, which is consistent with the conclusions drawn from the bending prediction. During the subsequent 2400-second relaxation phase, the pre-trained model predictions and experimental results exhibit a similar degree of decay in the reaction force, both falling within the 9%–10% range. However, the pre-trained model predictions exhibit a faster initial decay rate compared to experimental measurements, resulting in a smoother decay trend in later durations. Here, the force values obtained from the reaction force relaxation experiment of SFA are uniformly partitioned, with 80% of the measurements selected to constitute the fine-tuning dataset $\mathbb{T}^{force}$ and the remaining 20% reserved as an independent validation set to assess model prediction accuracy in the inference stage.

As anticipated, the fMDN pre-trained solely with material parameters calibrated from uniaxial test data proved insufficient to reliably capture the mechanical response of SFAs under complex operating conditions. Consequently, leveraging the neural network's high generalization capability and tunability, the pre-trained fMDN parameters were further fine-tuned based on the above SFA experimental results through the genetic algorithm (GA). The hyperelastic response parameters in the ANN part of fMDN will be primarily optimized using the bending test data, as the relatively short pressurization duration (compared to the Sil 950 material's characteristic relaxation time) renders minimal influence on the TKL-CNNs branch responsible for viscous response characterization. Meanwhile, the viscous response parameters in the TKL-CNNs branch will be independently optimized using the reaction force relaxation test data to capture the long-term material behavior. Two objective functions, f_{min}^{bend} and f_{min}^{force}, were formulated to quantitatively evaluate the accuracy of the fMDN parameters, serving as a metric to minimize the discrepancy between the model predictions and experimental observations through optimization. f_{min}^{bend} was defined as follows:

$$f_{min}^{bend} = \sum_{T \in \mathbb{T}^{bend}} \sum_{i=1}^{n} \sqrt{(X_{i,T} - \hat{X}_{i,T})^2 + (Y_{i,T} - \hat{Y}_{i,T})^2} \ , \tag{37}$$

where $\mathbb{T}^{bend}$ denotes three discretely selected time points (30 s, 35 s, and 40 s) in the fine-tuning dataset of bending test, n represents 24 uniformly distributed fixed nodes extracted along the lower surface of the SFA to provide spatial measurements of deformation characteristics, $X_{i,T}$ and $Y_{i,T}$ denote the experimentally measured coordinate values of the ith node at time point T, and $\hat{X}_{i,T}$ and $\hat{Y}_{i,T}$ denote the fMDN predicted coordinate values. Additionally, f_{min}^{force} was defined as follows:

$$f_{min}^{force} = \sum_{T \in \mathbb{T}^{force}} (RF_T - \hat{RF}_T)^2 \ , \tag{38}$$

where RF_T represents the experimentally measured reaction force at time T sampled at the fine-tuning dataset $\mathbb{T}^{force}$ of reaction force relaxation test, and $\hat{RF}_T$ represents the fMDN predicted reaction force. Notably, consistent with the pre-training stage, the mechanics-informed constraints outlined in Sections 3.3.1 and 3.3.2 are verified after each parameter fine-tuning step to ensure rigorous adherence to the memory decay property and thermodynamic consistency throughout the fine-tuning process. Moreover, the closed-form tangent modulus of the fine-tuned fMDN model can also be obtained using the algorithm presented in Section 4.1.

As shown in Fig. 8, the conventional viscoelastic models exhibit distinct predictive performances, with the Y-P model slightly outperforming the AB-P model in both experiments. For further comparison, same genetic algorithm are applied to fine-tune the parameters of the Y-P and AB-P models, enabling a direct comparison with the fine-tuned fMDN results.

Finally, the capability of the fMDN model to learn conventional viscoelastic constitutive behavior is specifically examined. As shown in Fig. 8, for both short-term bending tests and long-term reaction force relaxation tests, the fMDN pre-trained using the Y-P or AB-P models exhibit nearly identical predictive performance to their corresponding conventional viscoelastic models. This result indicates that, when the objective is to reproduce the mechanical response of viscoelastic materials as described by conventional constitutive formulations, the pre-training strategy introduced in Section 3 is sufficient. Moreover, such high-fidelity reproduction of the responses of conventional constitutive models is not restricted by the specific choice of conventional constitutive model, demonstrating strong generalization capability. However, as discussed previously, the primary advantage of the proposed data-driven constitutive modeling framework lies in its ability to bypass complex mathematical derivations and avoid reliance on empirical assumptions during material parameter identification. The framework does not depend on any specific analytical form of the hyperelastic strain energy density function or viscous relaxation function. Moreover, owing to the incorporation of mechanics-informed constraints, accurate predictions can be achieved using only a small and simple training dataset, while maintaining the capability to predict material responses under arbitrary loading durations, including durations exceeding those used in the construction of the training dataset. To further demonstrate these advantages, Appendix D presents a two-phase soft cantilever beam actuator with random particle inclusions. In this example, the matrix and particle phases are modeled using alternative conventional hyperelastic models (Ogden and Neo-Hookean) combined with different-order Prony series models accounting for volumetric relaxation effects, under a highly complex loading condition. The results further confirm that, compared with conventional viscoelastic constitutive models, the proposed method can significantly accelerate finite element simulations by approximately one order of magnitude.

5. Inference stage: Describing the viscoelastic behavior of SFAs

In this section, the fine-tuned fMDN and conventional viscoelastic models are first evaluated at the inference stage to compare their accuracy in capturing the mechanical behavior of SFAs.

First, the performance of the fine-tuned conventional viscoelastic models is examined. For the short-term responses of SFAs under hydraulic loading, as indicated by the purple and green dashed lines in Fig. 9(a)–(e), the predictions of the Y-P(GA) and AB-P(GA) models after GA-based fine-tuning show noticeable improvements over their original versions for the arc-shaped deformation results at 5 different loading times. Nevertheless, the predictions do not entirely fall within the 95% confidence intervals of the experimental observations. In addition, the two models exhibit different prediction accuracies at different time instants. Specifically, at 15 s, 30 s, and 35 s, the Y-P(GA) model provides slightly higher prediction accuracy than the AB-P(GA) model, whereas at the later stages of 40 s and 55 s, the AB-P(GA) model performs marginally better. Similarly, for the long-term responses of SFAs under pneumatic loading with gravitational effects, as shown in Fig. 9(f), the fine-tuned Y-P(GA) model yields predictions

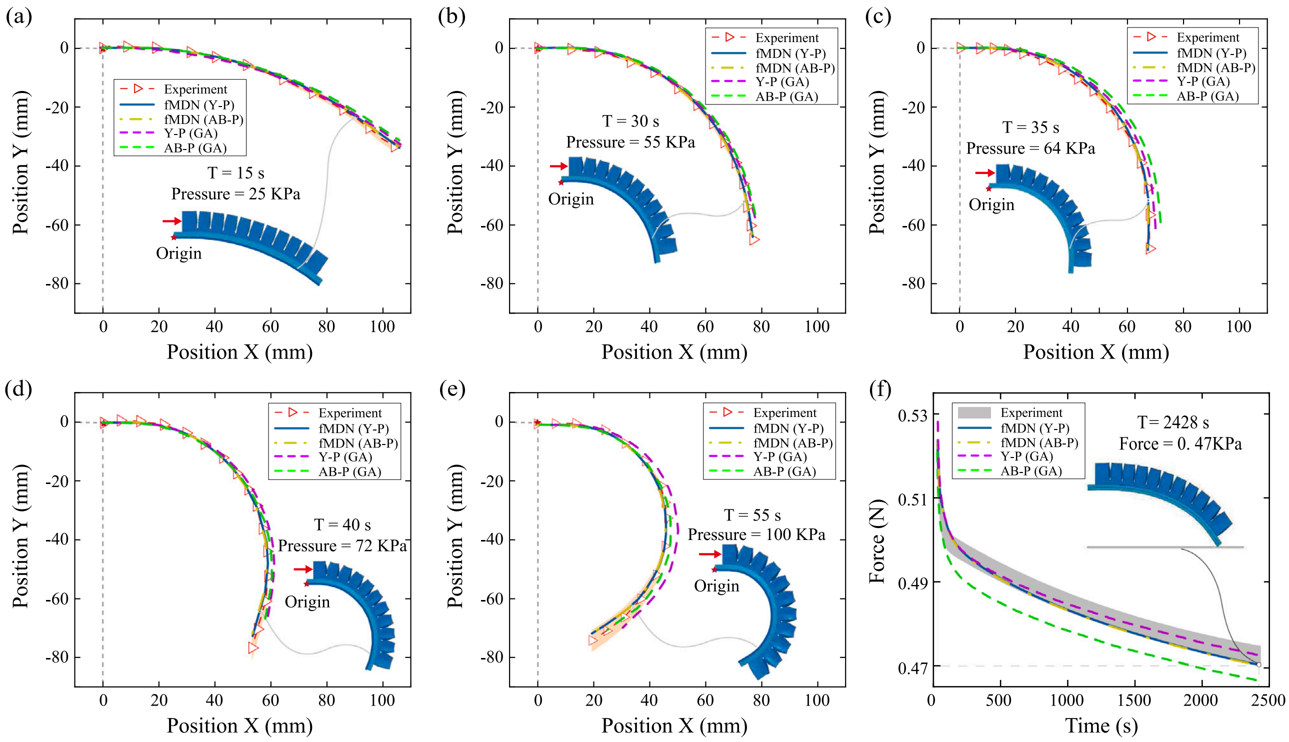

Fig. 9. Inference-stage performance of fine-tuned fMDN models and fine-tuned conventional viscoelastic models in SFAs tests. Arc-shaped bending deformation at different pressures, with the corresponding deformation configurations of the fine-tuned fMDN (Y-P) model also provided for each pressure level: (a) 25 kPa at 15 s; (b) 55 kPa at 30 s; (c) 64 kPa at 35 s; (d) 72 kPa at 40 s; (e) 100 kPa at 55 s. (f) The reaction force relaxation responses versus the loading time of the reference point in the *y* direction, and the display of the fine-tuned fMDN(Y-P) model calculated at 2428 s. The dashed red lines with triangular markers denote the experimental measurements, and all the shaded areas indicate the corresponding 95% confidence intervals. The dashed purple and green lines show the predictions of the fine-tuned Y-P(GA) and AB-P(GA) models using genetic algorithm, respectively. The solid blue and the dot-dashed yellow lines denote the predictions of the fine-tuned fMDN models, which were pre-trained on the Y-P and AB-P models, respectively.

of the reaction force relaxation process that mostly fall within the 95% confidence interval of the experimental data, except for slightly higher values at the initial stage of relaxation. In contrast, although the fine-tuned AB-P(GA) model also exhibits a clear improvement compared with its original version, its predictive performance is inferior to that of the Y-P(GA) model. Except for the initial relaxation stage, the majority of AB-P(GA) model's predictions do not fall within the 95% confidence interval of the experimental measurements.

Next, the performance of the fine-tuned fMDN models is discussed. For the short-term responses of SFAs, as indicated by the solid blue and dot-dashed yellow lines in Fig. 9(a)–(e), the fine-tuned fMDN models, whether pre-trained on the Y-P or AB-P models, exhibit nearly identical predictive results. This indicates that the proposed fMDN framework is not sensitive to the choice of the pre-training conventional constitutive model. Furthermore, nearly all coordinate values predicted by the proposed data-driven approach consistently fall within the 95% confidence intervals of the experimental measurements, demonstrating strong agreement between the predicted and observed responses. In addition, compared with directly fine-tuning conventional constitutive models, the proposed framework demonstrates superior predictive accuracy. This indicates that merely fine-tuning the parameters of conventional constitutive models is insufficient to achieve the desired level of accuracy, as their representation capability is fundamentally constrained by fixed explicit functional forms. It further confirms that the fMDN framework, by not relying on any specific analytical form of the hyperelastic strain energy density function or viscous relaxation function, exhibits enhanced generalization capability. Notably, these enhanced predictive performances are achieved through the fine-tuning stage using only three selected time points (30 s, 35 s, and 40 s) from the bending test data, which themselves show markedly improved prediction accuracy, as illustrated in Fig. 9(b)–(d). More importantly, the fine-tuned models maintain high accuracy for early-stage configurations at 15 s, as shown in Fig. 9(a). Furthermore, the pre-trained model is fine-tuned within a maximum time range of 40 s, while the predictions are extended up to 55 s, corresponding to an extrapolation factor of 1.375. Table 2 provides a clear comparison of the proportion of predicted values from both the pre-trained and fine-tuned models that fall outside the 95% confidence intervals of the experimental observations, along with the maximum relative error, at different time instances. It can be observed that substantial improvement is achieved in predicting extended temporal behavior at 55 s compared with the pre-trained models. This multi-timescale enhancement demonstrates that the proposed data-driven framework possesses a measurable temporal extrapolation capability in predicting the viscoelastic response of SFAs. However, it should be noted that part of this extrapolation capability is influenced by the GA-based fine-tuning procedure, as conventional constitutive models after GA (i.e., Y-P(GA) and AB-P(GA) models) also exhibit a certain degree of extrapolation capability at 55 s. To further isolate and assess the intrinsic extrapolation capability of the proposed fMDN framework, additional analyses have been conducted without the use of the genetic algorithm, as presented in Appendix D. In this case, the model is trained within a time range up to 3600 s, and the predictions are extended to 5850 s, corresponding to an extrapolation factor of 1.625. The results demonstrate that, even over such extended time scales, the proposed fMDN framework is capable of accurately capturing the time-dependent behavior of viscoelastic materials. Moreover, the predictive accuracy remains comparable to that within the training time range, which further confirms that the proposed fMDN framework exhibits a robust and well-demonstrated intrinsic extrapolation capability. In

Table 2
Comparison of the proportion of predicted values from both the pre-trained and fine-tuned models outside the 95% confidence intervals (CI) of the experimental observations, along with the maximum relative error, at different time instances.

Time	Metric	The pre-trained models		The fine-tuned models		
		Y-P pre-trained (Y-P)	AB-P pre-trained (AB-P)	Y-P (GA)	AB-P (GA)	fMDN (Y-P) fMDN (AB-P)
15 s	Outside 95% CI	4%	1%	10%	17%	1%
	Max relative error	**1.5%**	**0.4%**	0.9%	2.1%	**0.4%**
30 s	Outside 95% CI	40%	24%	14%	19%	1%
	Max relative error	**5.4%**	**7.6%**	5.1%	6.9%	**0.5%**
35 s	Outside 95% CI	44%	32%	56%	62%	2%
	Max relative error	**6.2%**	**8.7%**	6.9%	8.5%	**0.9%**
40 s	Outside 95% CI	68%	70%	43%	30%	1%
	Max relative error	**8.9%**	**11.7%**	7.0%	8.8%	**0.4%**
55 s	Outside 95% CI	80%	76%	67%	33%	3%
	Max relative error	**21.5%**	**27.9%**	12.1%	10.3%	**1.3%**

addition, for the long-term responses of SFAs, as shown in Fig. 9(f), the fine-tuned fMDN(Y-P) and fMDN(AB-P) models exhibit significant performance improvements and nearly identical predictive results, with all predictions falling within the 95% confidence intervals of the experimental observations. These results further validate the effectiveness of the proposed data-driven approach in accurately characterizing the long-term viscoelastic behavior of SFAs.

In summary, while the fine-tuned Y-P(GA) and AB-P(GA) models demonstrate substantial improvements in predicting both the short- and long-term viscoelastic behavior of SFAs, they still fail to provide an accurate and comprehensive description of the viscoelastic behavior, with non-negligible discrepancies remaining. Compared with these conventional constitutive models, the proposed framework provides more accurate and comprehensive predictions of both short- and long-term viscoelastic behaviors of SFAs, with predictions remaining within the 95% confidence intervals of the experimental measurements. Meanwhile, the proposed framework exhibits high robustness to the choice of conventional viscoelastic constitutive models adopted in the pre-training stage for preliminary parameter evaluation. Moreover, the framework maintains robust predictive capability across multi-physical loading scenarios, including hydraulic and pneumatic loading, without degradation due to gravitational effects. As a supplement, the LLM-inspired design principles of the fMDN framework are presented in detail in Appendix E, highlighting the distinctions and connections between the proposed framework and standard LLM architectures.

6. Conclusion

This study presents a mechanics-informed experimental-data-driven framework for modeling the viscoelastic behavior of soft structures. The proposed framework adopts a three-stage training strategy inspired by LLMs and embeds mechanics-informed constraints into the learning process. Compared with conventional viscoelastic constitutive models, it achieves more accurate and comprehensive predictions of both short- and long-term viscoelastic responses of Sil 950-based SFAs under hydraulic and pneumatic actuation, while explicitly accounting for gravitational effects. Moreover, this superior predictive performance is insensitive to the choice of conventional viscoelastic constitutive models adopted in the pre-training stage for preliminary parameter evaluation.

The core of the framework is a tailored fractional multi-branch deep network (fMDN) that combines an ANN for instantaneous hyperelastic response with two dedicated TKL-CNN branches for the deviatoric and volumetric components of time-dependent viscous behavior. This hybrid structure allows the equilibrium and non-equilibrium aspects of viscoelasticity to be learned simultaneously. In contrast to purely data-driven black-box models, the proposed approach explicitly enforces viscoelastic memory decay and thermodynamic consistency, thereby ensuring viscoelastic mechanical predictions that satisfy mechanics-informed constraints, without resorting to explicit analytical derivations of hyperelastic parameters or viscous relaxation functions. Another key advantage of the proposed method is its data efficiency. Only a limited number of simple material-level experiments, SFA-level tests, and a small amount of augmented numerical data generated from RVE simulations are required. The loading durations of the numerical and experimental protocols can be flexibly adjusted to match the characteristic relaxation timescale of the material, while avoiding excessive experimental effort.

The practical applicability of the proposed framework is demonstrated through its implementation as a UMAT subroutine in ABAQUS. The resulting finite element simulations show good agreement with experimental measurements, confirming the robustness and predictive capability of the proposed framework. Furthermore, for composite structures subjected to complex loading conditions, the proposed approach exhibits significantly reduced computational cost compared with DNS, while maintaining comparable accuracy for mechanical responses well beyond the training loading durations. In addition, the proposed framework has strong potential for extension to other viscoelastic soft structures.

Although the present work focuses on SFAs, the proposed framework is expected to be transferable to other viscoelastic soft structures. Future work will address the incorporation of material anisotropy, as well as damage evolution and stress softening mechanisms, to further extend the applicability of the method.

CRediT authorship contribution statement

Yicheng Lu: Writing – review & editing, Writing – original draft, Validation, Software, Project administration, Methodology, Formal analysis, Data curation. **Xinjie Hu:** Validation, Investigation, Data curation. **Yichen Pu:** Validation, Investigation, Data curation. **Zefeng Yu:** Validation, Investigation, Formal analysis. **Ning An:** Validation, Supervision, Resources, Investigation, Funding acquisition. **Shan Tang:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization. **Xu Guo:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization.

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.

Acknowledgments

The financial support from the National Natural Science Foundation of China (12372194), the National Key Research and Development Program of China (2024YFB3310402, 2024YFB3310401), the Key Science and Innovation Program of Liaoning Province (2024JH1/11700046), and the Open Project of State Key Laboratory of Structural Analysis for Industrial Equipment, Dalian University of Technology (GZ22120) are gratefully acknowledged.

Appendix A. Conventional models for preliminary evaluation of the Sil 950 material parameters

The instantaneous hyperelastic parameters are evaluated using the Yeoh and Arruda–Boyce models, which can be expressed as:

$$W = \gamma_1(\bar{I}_1 - 3) + \gamma_2(\bar{I}_1 - 3)^2 + \gamma_3(\bar{I}_1 - 3)^3 + \frac{1}{D_m}(J-1)^2 \tag{39}$$

and

$$W = \mu\left[\frac{1}{2}(\bar{I}_1 - 3) + \frac{1}{20\lambda_m^2}(\bar{I}_1^2 - 9) + \frac{11}{1050\lambda_m^4}(\bar{I}_1^3 - 27) + \frac{19}{7000\lambda_m^6}(\bar{I}_1^4 - 81) + \frac{519}{673750\lambda_m^8}(\bar{I}_1^5 - 243)\right] + \frac{1}{D_m}\left(\frac{J^2-1}{2} - \ln J\right) \tag{40}$$

respectively, where W is the strain energy potential, $\bar{I}_1 = J^{-\frac{2}{3}}(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)$ is the first deviatoric strain invariant, and γ_i, μ, λ_m and D_m are the parameters to be evaluated.

The Prony series model used to evaluate the viscous parameters is expressed as a series of exponential terms as follows:

$$\begin{cases} \alpha_G \dot{g}_R(t) = \dfrac{1}{G_0}\displaystyle\sum_{i=1}^{n_G} \frac{G_i}{\tau_i^G} e^{-t/\tau_i^G} \\ \alpha_K \dot{k}_R(t) = \dfrac{1}{K_0}\displaystyle\sum_{i=1}^{n_K} \frac{K_i}{\tau_i^K} e^{-t/\tau_i^K}, \end{cases} \tag{41}$$

where n_G and n_K represent the numbers of terms associated with the shear and bulk response modes, respectively. For each ith term, G_i and K_i denote the corresponding shear and bulk moduli, while τ_i^G and τ_i^K characterize the associated relaxation times.

Appendix B. The detailed derivation of $\partial \boldsymbol{F}/\partial \boldsymbol{C}$ required in Eq. (31)

At any material point, the right Cauchy–Green tensor $\boldsymbol{C}$ can be expressed in the principal space,

$$\boldsymbol{C} = C_i^2 \boldsymbol{N}_i \otimes \boldsymbol{N}_i \ , \tag{42}$$

where C_i^2 $(i = 1,2,3)$ are the eigenvalues of $\boldsymbol{C}$ and $\boldsymbol{N}_i$ $(i = 1,2,3)$ are the associated eigenvector.

The derivative of deformation gradient $\boldsymbol{F}$ with respect to right Cauchy–Green tensor $\boldsymbol{C}$ is derived as follows:

$$\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{C}} = \frac{\partial \boldsymbol{R}\cdot\boldsymbol{U}}{\partial \boldsymbol{C}} = \boldsymbol{R}\cdot\frac{\partial \boldsymbol{U}}{\partial \boldsymbol{C}} \ , \tag{43}$$

where $\boldsymbol{U} = C_i \boldsymbol{N}_i \otimes \boldsymbol{N}_i$ is the right stretch tensor, and $\boldsymbol{R} = \boldsymbol{F}\cdot\boldsymbol{U}^{-1}$ is the rigid body rotation tensor. According to Gullerud et al. [102], $\boldsymbol{U}$ can be expressed by the function of $\boldsymbol{C}$ as follows:

$$\boldsymbol{U} = \beta_1(\beta_2 \boldsymbol{1} + \beta_3 \boldsymbol{C} - \boldsymbol{C}^2) \ , \tag{44}$$

where the β coefficients defined by the invariants of U ($I_U = C_1 + C_2 + C_3$, $II_U = C_1C_2 + C_1C_3 + C_2C_3$, $III_U = C_1C_2C_3$),

$$\beta_1 = 1/(I_U II_U - III_U), \beta_2 = I_U III_U, \beta_3 = I_U^2 - II_U. \tag{45}$$

Therefore, $\frac{\partial \boldsymbol{U}}{\partial \boldsymbol{C}}$ in Eq. (43) can be derived as follows:

$$\frac{\partial \boldsymbol{U}}{\partial \boldsymbol{C}} = \sum_{i=1}^{3}\left(\frac{\partial(\beta_1\beta_2)}{\partial C_i}\frac{\partial C_i}{\partial \boldsymbol{C}}\boldsymbol{1} + \frac{\partial(\beta_1\beta_3)}{\partial C_i}\frac{\partial C_i}{\partial \boldsymbol{C}}\boldsymbol{C} + \beta_1\beta_3\boldsymbol{1}\otimes\boldsymbol{1} - \frac{\partial \beta_1}{\partial C_i}\frac{\partial C_i}{\partial \boldsymbol{C}}\boldsymbol{C}^2 - \beta_1(\boldsymbol{1}\otimes\boldsymbol{C} + \boldsymbol{C}\otimes\boldsymbol{1})\right), \tag{46}$$

where the derivative of the above β coefficients with respect to the eigenvalues of $\boldsymbol{U}$ can be obtained as follows:

$$\begin{cases} \dfrac{\partial(\beta_1\beta_2)}{\partial C_i} = \dfrac{C_j^2 C_k^2\left(2C_i + C_j + C_k\right)}{\left(C_i + C_k\right)^2\left(C_i + C_j\right)^2\left(C_j + C_k\right)} \\ \dfrac{\partial(\beta_1\beta_3)}{\partial C_i} = -\dfrac{(C_j^2 + C_k^2)\left(2C_i + C_j + C_k\right)}{\left(C_i + C_k\right)^2\left(C_i + C_j\right)^2\left(C_j + C_k\right)} (i \neq j \neq k = 1,2,3). \\ \dfrac{\partial \beta_1}{\partial C_i} = -\dfrac{2C_i + C_j + C_k}{\left(C_i + C_k\right)^2\left(C_i + C_j\right)^2\left(C_j + C_k\right)} \end{cases} \tag{47}$$

Additionally, the derivative of the eigenvalues of $\boldsymbol{U}$ with respect to the right Cauchy–Green tensor C can be derived as follows:

$$\begin{aligned} &\frac{\partial C_i}{\partial \boldsymbol{C}} = \left(2C_i\right)^{-1}\boldsymbol{N}_i\otimes\boldsymbol{N}_i \quad (i = 1,2,3) && \text{for} \quad C_1 \neq C_2 \neq C_3, \\ &\left.\begin{aligned} &\frac{\partial \lambda}{\partial \boldsymbol{C}} = (2\lambda)^{-1}\left(\boldsymbol{1}\otimes\boldsymbol{1} - \boldsymbol{N}_3\otimes\boldsymbol{N}_3\right) \\ &\frac{\partial C_3}{\partial \boldsymbol{C}} = \left(2C_3\right)^{-1}\boldsymbol{N}_3\otimes\boldsymbol{N}_3 \end{aligned}\right\} && \text{for} \quad \lambda = C_1 = C_2 \neq C_3, \\ &\frac{\partial \lambda}{\partial \boldsymbol{C}} = \sum_{i=1}^{3}(2\lambda)^{-1}\boldsymbol{N}_i\otimes\boldsymbol{N}_i = (2\lambda)^{-1}\boldsymbol{1}\otimes\boldsymbol{1} && \text{for} \quad \lambda = C_1 = C_2 = C_3. \end{aligned} \tag{48}$$

Finally, $\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{C}}$ can be obtained by substituting Eqs. (46)–(48) into Eq. (43).

Appendix C. The detailed derivation of Eq. (33)

According to the chain rule, $\frac{\partial \boldsymbol{f}_{\text{ANN}}}{\partial \boldsymbol{F}}$ in Eq. (33) can be obtained as follows:

$$\frac{\partial \boldsymbol{f}_{\text{ANN}}}{\partial \boldsymbol{F}} = \frac{\partial \boldsymbol{f}_{\text{ANN}}}{\partial \boldsymbol{B}} : \frac{\partial \boldsymbol{B}}{\partial \boldsymbol{F}} \ , \tag{49}$$

where the derivative of ANN's output (the instantaneous hyperelastic Kirchhoff stress) $\boldsymbol{f}_{\text{ANN}}$ with respect to ANN's input (left Cauchy–Green tensor) $\boldsymbol{B}$ can be computed by the backpropagation algorithm of ANN [92,103], and $\frac{\partial \boldsymbol{B}}{\partial \boldsymbol{F}}$ can be expressed as:

$$\frac{\partial B_{ij}}{\partial F_{kl}} = \delta_{ik}F_{jl} + F_{il}\delta_{jk} \ , \tag{50}$$

with δ_{ik} is the Kronecker delta.

Similarly, $\frac{\partial \boldsymbol{f}_{\text{TKL-CNN1}}}{\partial \boldsymbol{F}^n}$ in Eq. (33) can be obtained as follows:

$$\frac{\partial \boldsymbol{f}_{\text{TKL-CNN1}}}{\partial \boldsymbol{F}^n} = \frac{\partial \boldsymbol{f}_{\text{TKL-CNN1}}}{\partial \hat{\boldsymbol{\tau}}^n_{h,dev}} : \frac{\partial \hat{\boldsymbol{\tau}}_{h,dev}}{\partial \hat{\boldsymbol{\tau}}_h} : \frac{\partial \boldsymbol{f}_{\text{ANN}}}{\partial \boldsymbol{F}} + \frac{\partial \boldsymbol{f}_{\text{TKL-CNN1}}}{\partial \overline{\boldsymbol{F}}^n} : \frac{\partial \overline{\boldsymbol{F}}}{\partial \boldsymbol{F}} \ , \tag{51}$$

where $\frac{\partial \hat{\boldsymbol{\tau}}_{h,dev}}{\partial \hat{\boldsymbol{\tau}}_h}$ and $\frac{\partial \overline{\boldsymbol{F}}}{\partial \boldsymbol{F}}$ can calculated by follows:

$$\begin{cases} \left(\dfrac{\partial \hat{\boldsymbol{\tau}}_{h,dev}}{\partial \hat{\boldsymbol{\tau}}_h}\right)_{ijkl} = \left(\dfrac{\partial \text{dev}(\bullet)}{\partial(\bullet)}\right)_{ijkl} = \delta_{ik}\delta_{jl} - \dfrac{1}{3}\delta_{ij}\delta_{kl} \\ \left(\dfrac{\partial \overline{\boldsymbol{F}}}{\partial \boldsymbol{F}}\right)_{ijkl} = -\dfrac{1}{3}J^{-\frac{1}{3}}F^{-1}_{lk}F_{ij} + J^{-\frac{1}{3}}\delta_{ik}\delta_{jl}. \end{cases} \tag{52}$$

Additionally, the derivative of TKL-CNN1's output (the deviatoric part of viscous Kirchhoff stress) $\boldsymbol{f}_{\text{TKL-CNN1}}$ with respect to the components of TKL-CNN1's inputs (the deviatoric part of instantaneous Kirchhoff stress) $\hat{\boldsymbol{\tau}}^n_{h,dev}$ and (the distortional deformation gradient) $\overline{\boldsymbol{F}}^n$ can be computed by the backpropagation algorithm of TKL-CNN1 as:

$$\frac{\partial \boldsymbol{f}_{\text{TKL-CNN1}}}{\partial \hat{\boldsymbol{\tau}}^n_{h,dev}} = \frac{\partial \text{dev}(\bullet)}{\partial(\bullet)} : \left(k_{\text{TKL-CNN1}}(1)\Delta t\,\boldsymbol{1}\otimes\boldsymbol{1}\right), \tag{53}$$

and

$$\frac{\partial \boldsymbol{f}_{\mathrm{TKL-CNN1}}}{\partial \overline{\boldsymbol{F}}^{n}} = \frac{\partial \mathrm{dev}(\bullet)}{\partial(\bullet)} : \sum_{\zeta=1}^{\eta} \left[\mathbf{1} \otimes \mathbf{1} \cdot \left(\overline{\boldsymbol{F}}^{-1}\right)^{n-\zeta} \cdot \hat{\boldsymbol{\tau}}_{h}^{n-\zeta} \cdot \left(\overline{\boldsymbol{F}}^{-T}\right)^{n-\zeta} \cdot \left(\overline{\boldsymbol{F}}^{T}\right)^{n} \right.$$
$$\left. + \left(\overline{\boldsymbol{F}}\right)^{n} \cdot \left(\overline{\boldsymbol{F}}^{-1}\right)^{n-\zeta} \cdot \hat{\boldsymbol{\tau}}_{h}^{n-\zeta} \cdot \left(\overline{\boldsymbol{F}}^{-T}\right)^{n-\zeta} \cdot \mathbf{1} \otimes \mathbf{1} \right] k_{\mathrm{TKL-CNN1}}(\zeta+1)\Delta t, \tag{54}$$

with

$$\begin{cases} n < N_G & \Longrightarrow \quad \eta = n - 1 \\ \text{Otherwise} & \Longrightarrow \quad \eta = N_G - 1. \end{cases} \tag{55}$$

Finally, $\frac{\partial \boldsymbol{f}_{\mathrm{TKL-CNN2}}}{\partial \boldsymbol{F}^{n}}$ in Eq. (33) can be obtained as follows:

$$\frac{\partial \boldsymbol{f}_{\mathrm{TKL-CNN2}}}{\partial \boldsymbol{F}^{n}} = \frac{\partial \boldsymbol{f}_{\mathrm{TKL-CNN2}}}{\partial \hat{\boldsymbol{\tau}}_{h,vol}^{n}} : \frac{\partial \hat{\boldsymbol{\tau}}_{h,vol}^{n}}{\partial \hat{\boldsymbol{\tau}}_{h}} : \frac{\partial \boldsymbol{f}_{\mathrm{ANN}}}{\partial \boldsymbol{F}}, \tag{56}$$

where $\frac{\partial \hat{\boldsymbol{\tau}}_{h,vol}}{\partial \hat{\boldsymbol{\tau}}_{h}}$ can calculated as:

$$\left(\frac{\partial \hat{\boldsymbol{\tau}}_{h,vol}^{n}}{\partial \hat{\boldsymbol{\tau}}_{h}}\right)_{ijkl} = \frac{1}{3}\delta_{ij}\delta_{kl}. \tag{57}$$

In addition, the derivative of TKL-CNN2's output (the volumetric part of viscous Kirchhoff stress) $\boldsymbol{f}_{\mathrm{TKL-CNN2}}$ with respect to the component of TKL-CNN2's inputs (the volumetric part of instantaneous Kirchhoff stress) $\hat{\boldsymbol{\tau}}_{h,vol}^{n}$ can be computed by the backpropagation algorithm of TKL-CNN2 as:

$$\frac{\partial \boldsymbol{f}_{\mathrm{TKL-CNN2}}}{\partial \hat{\boldsymbol{\tau}}_{h,vol}^{n}} = k_{\mathrm{TKL-CNN2}}(1)\Delta t\, \mathbf{1} \otimes \mathbf{1}. \tag{58}$$

As such, Eq. (33) can be computed based on the above results.

Appendix D. Two-phase soft cantilever beam actuator with random particle inclusions

A two-phase soft cantilever beam actuator with random particle inclusions is investigated, with the particle and matrix phases modeled using two different conventional viscoelastic materials (Table D.3). The explicit formulations of Ogden and Neo-Hookean models are expressed as:

$$W = \frac{2\mu}{\alpha^2}(\bar{\lambda}_1^{\alpha} + \bar{\lambda}_2^{\alpha} + \bar{\lambda}_3^{\alpha} - 3) + \frac{1}{D_m}(J-1)^2 \tag{59}$$

and

$$W = \frac{\mu}{2}(\bar{I}_1 - 3) + \frac{1}{D_m}(J-1)^2 \tag{60}$$

respectively, where $\bar{\lambda}_i = J^{-\frac{1}{3}}\lambda_i$ are the deviatoric principal stretches, and α, μ and D_m are the parameters to be evaluated. Notably, compared with the conventional viscoelastic models adopted for preliminary evaluation of the Sil 950 material parameters, the particle and matrix phases employ different hyperelastic formulations and higher-order Prony series representations.

As shown in Fig. 10(a), the beam is fixed at one end, while a combined loading consisting of a complex cyclic force F_y and a linear torque M_x, as illustrated in Fig. 10(b), is applied at the other end. To better characterize the time-dependent viscoelastic response, the applied force F_y follows a cyclic loading protocol, in which it is ramped to a local peak within 450 s, maintained at that level for 450 s, and then unloaded to zero over the next 450 s. This cycle is repeated until the maximum peak force of 0.45 N is attained at 5850 s. Direct numerical simulation (DNS) and the proposed fMDN model are employed in this study. As shown in Fig. 10(c), the DNS is conducted using three-dimensional ten-node elements (C3D10), discretizing the beam into 259,238 nodes and 184,164 elements. In contrast, the fMDN approach relies on a training dataset generated by loading a RVE with particle inclusions at the same volume fraction (7.29%) using the loading paths and protocols listed in Table 1. For the seven viscous loading paths marked with *, the loading duration is limited to 3600 s, which is much shorter than the applied beam loading duration of 5850 s. Following the pre-training strategy described in Section 3, the fMDN is trained to obtain an effective homogenized material model. Since volumetric relaxation effect are considered, the trained fMDN model yields $N_G = 278$ and $N_K = 232$, further indicating that the relaxation effect differs between the deviatoric and volumetric components of the material. Compared with the DNS approach, the proposed fMDN approach represents the effects of particle inclusions through a homogenization procedure. The resulting homogenized structure is discretized into 25,856 nodes and 22,500 elements using three-dimensional eight-node elements with reduced integration (C3D8R).

Notably, in all examples involving the fMDN model presented in this work, the maximum allowable increment size is set to be no larger than the fixed training time increment Δt_f. When interpolation is performed over a relatively large time increment, the linear interpolation strategy may neglect a number of intermediate values of the originally trained network parameters and estimate the updated parameter value solely from the two endpoint values associated with the large increment. To overcome this limitation, an interpolation-based homogenization strategy is introduced. Specifically, when the elapsed time $t^n - t^i$ between the current loading step n and the ith historical loading increment spans several fixed kernel time grids, the corresponding viscoelastic kernel value is evaluated by accumulating all trained kernel values over this interval on their original fixed-grid time scale and subsequently averaging the accumulated contribution over the elapsed time. In this manner, for relatively large time increments, the information carried by all trained kernel values traversed within the corresponding time interval can be retained, rather than being discarded by endpoint-only interpolation.

The influence of interpolation with different time-increment sizes on the predicted results is further discussed. Fig. 10(d) compares the time-dependent rotation θ_x and bending displacement U_y predicted by the DNS and by the fMDN models with different maximum increment sizes of $0.1\Delta t_f$, Δt_f, and $10\Delta t_f$. For all interpolation time-increment sizes considered, the two approaches show close agreement in both the torsional and bending responses, with only minor deviations observed at the onset and termination of the force plateaus. Fig. 10(e) presents the relative errors of the rotation θ_x and displacement U_y between the DNS and the fMDN models with different maximum increment sizes of $0.1\Delta t_f$, Δt_f, and $10\Delta t_f$. It can be observed that high accuracy is maintained for all interpolation time-increment sizes when volumetric relaxation effects are included, even for loading durations extending beyond the training range (1.625 times longer), with quantitatively consistent predictive performance (maximum relative error $< 7\%$) compared with that within the training time range. These results demonstrate the strong robustness of the proposed fMDN framework with respect to interpolation using different time-increment sizes. In addition, these findings further substantiate that the proposed fMDN framework does not exhibit significant error accumulation when predicting mechanical responses beyond the training dataset, indicating a robust and well-demonstrated intrinsic extrapolation capability.

To further evaluate the fMDN predictions at different time instants, equivalent stress contours at t = 3600 s, 4500 s, and 5850 s obtained from the two approaches are compared in Fig. 11. Despite minor local differences induced by homogenization, especially near particle inclusions, the overall equivalent stress distributions predicted by the fMDN model agree well with those from DNS. Finally, the element types, total number of nodes, total number of elements, time step, and computational cost of the two approaches are compared and summarized in Table D.4. Both approaches are implemented on the same hardware (Intel Xeon Gold 5220R CPU @ 2.20 GHz, 64 RAM). The results indicate that the fMDN approach requires 2.16 h, whereas the DNS requires 20.8 h, corresponding to an approximately one-order-of-magnitude reduction in computational time. These findings

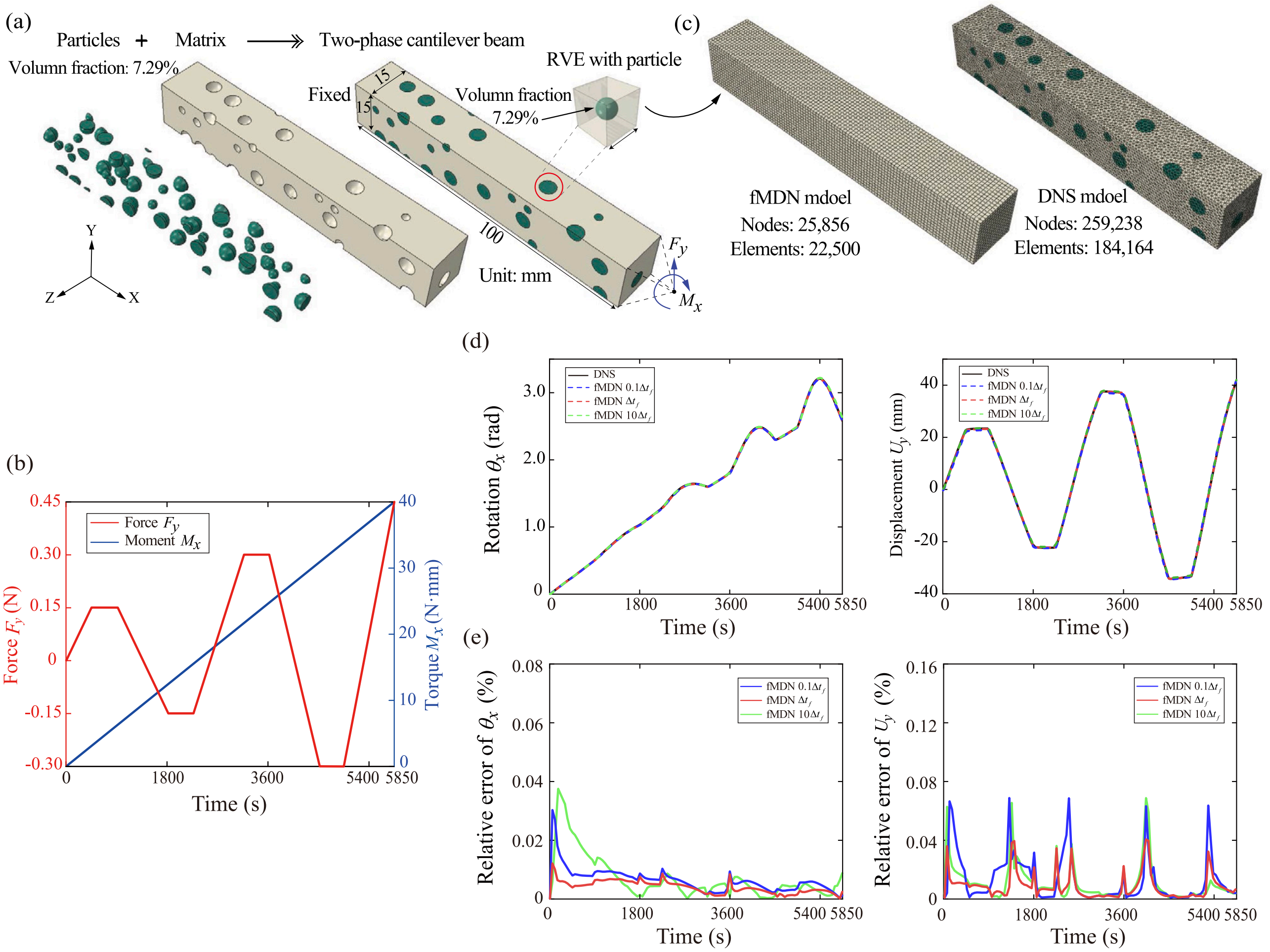

Fig. 10. Finite element analysis of a two-phase soft cantilever beam actuator with random particle inclusions: (a) Geometry and boundary conditions of the particle and matrix phases; (b) Applied complex cyclic force F_y and linear torque M_x; (c) Meshes of the fMDN and DNS models with corresponding node and element counts; (d) Comparison of the time-dependent rotation θ_x and displacement U_y responses calculated by the DNS and by the fMDN models with different maximum increment sizes of $0.1\Delta t_f$, Δt_f, and $10\Delta t_f$; (e) Relative errors of the rotation θ_x and displacement U_y between the DNS and the fMDN models with different maximum increment sizes of $0.1\Delta t_f$, Δt_f, and $10\Delta t_f$.

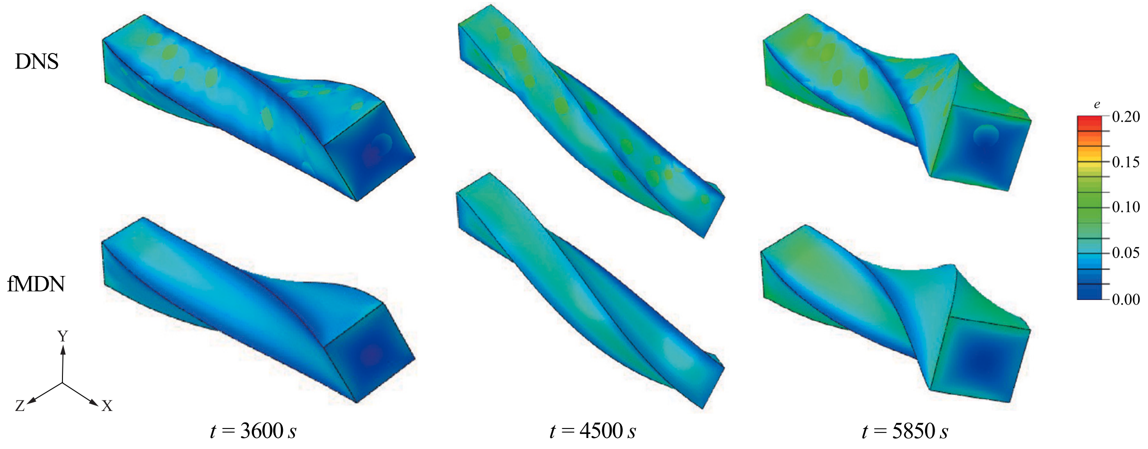

Fig. 11. Comparison of equivalent stress contour plots computed by the DNS and fMDN models at different time instants (3600 s, 4500 s, and 5850 s)

Table D.3
The preliminarily evaluated material parameters of Sil 950 using the material-level experimental data.

Particles	Ogden model (instantaneous)			Prony series model (viscous)			
	μ	α	D_m	i	g_i	k_i	τ_i
Value	0.5	7	0.2	1	0.2	0.1	5
				2	0.12	0.03	20
				3	0.05	0.05	50
				4	0.02	0.02	100
				5	0.02	0.03	300
				6	0.003	0.003	1000
Matrix	Neo-Hookean model (instantaneous)			Prony series model (viscous)			
	μ		D_m	i	g_i	k_i	τ_i
Value	0.3		0.34	1	0.25	0.15	6
				2	0.15	0.1	10
				3	0.07	0.03	40
				4	0.02	0.03	200
				5	0.01	0.01	600

Table D.4
Comparison of element types, total number of nodes, total number of elements, time step, and computational cost between the fMDN approach and DNS.

	fMDN	DNS
Element type	C3D8R	C3D10
Number of nodes	25,856	259,238
Number of elements	22,500	184,164
Time step	Automatic	Automatic
Computational cost	2.16 h	20.8 h

demonstrate the effectiveness of the proposed fMDN approach for efficient and accurate prediction of viscoelastic behavior in composite structures. For all analysis examples presented in this work using the proposed fMDN model, the average number of iterations per increment remains below 3, demonstrating excellent convergence behavior.

Appendix E. LLM-inspired design principles of the fMDN framework

The proposed fMDN framework represents a mechanics-informed adaptation of LLM architectures' underlying principles for modeling viscoelastic materials, rather than directly adopting or implementing standard LLM architectures.

In terms of input representation, the fMDN framework treats the deformation history of viscoelastic materials as an input sequence to predict the material response at the current time step. This is conceptually analogous to autoregressive modeling in LLMs, where the prediction of the next token depends on the preceding sequence. In principle, LLMs estimate the conditional probability distribution $P(x_{n+1}|x_1, x_2, \dots, x_n)$, while the proposed framework predicts the current material response conditioned on the loading history.

In terms of intermediate processing, the fMDN model captures the time-dependent behavior of viscoelastic materials by quantifying the influence of historical deformation states on the current response. This is achieved through the evaluation of the decaying kernel values $\boldsymbol{k}_{\mathrm{TKL-CNN1}}$ and $\boldsymbol{k}_{\mathrm{TKL-CNN2}}$, which are constrained by the mechanics-informed constraints (e.g., consistency thermodynamics and memory decay effect). This is analogous to the self-attention strategy in LLMs [104], where the influence of past inputs on the current state is evaluated. In addition, the concept of kernel length in TKL-CNNs is analogous to the notion of context length (or block size) in classical n-gram models [105], as both determine the effective range of historical information that influences the current prediction. Furthermore, the mapping of fixed temporal kernels to non-uniform time increments in UMAT implementations is conceptually similar to tokenization in natural language processing.

In terms of output prediction, the fMDN framework adopts a three-stage training paradigm consisting of pre-training, fine-tuning, and inference. The pre-training stage learns general constitutive behavior from a limited yet readily available dataset generated by conventional constitutive models, while the fine-tuning stage adapts the model to more complex and application-relevant experimental observations. This is analogous to the transfer learning paradigm widely adopted in LLMs. At the inference stage, the proposed framework demonstrates measurable temporal extrapolation capability and strong generalization ability, enabling accurate prediction of long-term viscoelastic behavior. Moreover, it shows potential applicability to a broad class of soft viscoelastic structures beyond SFAs. This behavior is consistent with the design philosophy of LLM-inspired learning strategies, particularly their ability to generalize and transfer across diverse tasks and domains.

Data availability

Data will be made available on request.

References

[1] Runciman M, Darzi A, Mylonas GP. Soft robotics in minimally invasive surgery. Soft Robot 2019;6(4):423–43.

[2] Hu X, Chen A, Luo Y, Zhang C, Zhang E. Steerable catheters for minimally invasive surgery: a review and future directions. Comput Assist Surg 2018;23(1):21–41.

[3] Ansari Y, Manti M, Falotico E, Mollard Y, Cianchetti M, Laschi C. Towards the development of a soft manipulator as an assistive robot for personal care of elderly people. Int J Adv Robot Syst 2017;14(2):1729881416687132.

[4] Kalita B, Leonessa A, Dwivedy SK. A review on the development of pneumatic artificial muscle actuators: Force model and application. In: Actuators, vol. 11, (10):MDPI; 2022, p. 288.

[5] Achilli GM, Amici C, Dragusanu M, Gobbo M, Logozzo S, Malvezzi M, Tiboni M, Valigi MC. Soft, rigid, and hybrid robotic exoskeletons for hand rehabilitation: roadmap with impairment-oriented rationale for devices design and selection. Appl Sci 2023;13(20):11287.

[6] Guo W, Katiyar SA, Davis S, Nefti-Meziani S. Overview: A comprehensive review of soft wearable rehabilitation and assistive devices, with a focus on the function, design and control of lower-limb exoskeletons. Machines 2025;13(11):1020.

[7] De Greef A, Lambert P, Delchambre A. Towards flexible medical instruments: Review of flexible fluidic actuators. Precis Eng 2009;33(4):311–21.

[8] Shintake J, Cacucciolo V, Floreano D, Shea H. Soft robotic grippers. Adv Mater 2018;30(29):1707035.

[9] Deimel R, Brock O. A compliant hand based on a novel pneumatic actuator. In: 2013 IEEE international conference on robotics and automation. IEEE; 2013, p. 2047–53.

[10] Alam Z, Kaushik P, Sharma AK, Wu B, Chen W. Obliquely propagating incident SH waves in periodic hard-magnetic soft laminates. Int J Mech Sci 2025;287:109945.

[11] Sharma AK, Kosta M, Shmuel G, Amir O. Gradient-based topology optimization of soft dielectrics as tunable phononic crystals. Compos Struct 2022;280:114846.

[12] Alam Z, Sharma AK, Ramachandran VP. Design of hard-magnetic soft laminated composites for wide longitudinal wave band gaps using topology optimization. Int J Solids Struct 2025;320:113493.
[13] Alam Z, Sharma AK. Topology optimization of hard-magnetic soft laminates for wide tunable SH wave bandgaps. Compos Struct 2025;366:119157.
[14] Alam Z, Sharma AK. Topology optimization of hard-magnetic soft phononic structures for wide magnetically tunable band gaps. J Appl Mech 2024;91(10):101009.
[15] Padmanabhan S, Alam Z, Sharma AK. Tunable anti-plane wave bandgaps in 2D periodic hard-magnetic soft composites. Int J Mech Sci 2024;261:108686.
[16] Alam Z, Padmanabhan S, Sharma AK. Magnetically tunable longitudinal wave band gaps in hard-magnetic soft laminates. Int J Mech Sci 2023;249:108262.
[17] Gul JZ, Sajid M, Rehman MM, Siddiqui GU, Shah I, Kim KH, Lee JW, Choi KH. 3D printing for soft robotics–a review. Sci Technol Adv Mater 2018;19(1):243–62.
[18] Wallin T, Pikul J, Shepherd RF. 3D printing of soft robotic systems. Nat Rev Mater 2018;3(6):84–100.
[19] Xavier MS, Fleming AJ, Yong YK. Finite element modeling of soft fluidic actuators: Overview and recent developments. Adv Intell Syst 2021;3(2):2000187.
[20] Hu W, Alici G. Bioinspired three-dimensional-printed helical soft pneumatic actuators and their characterization. Soft Robot 2020;7(3):267–82.
[21] Connolly F, Walsh CJ, Bertoldi K. Automatic design of fiber-reinforced soft actuators for trajectory matching. Proc Natl Acad Sci 2017;114(1):51–6.
[22] Pu Y, Zheng S, Hu X, Tang S, An N. Robotic skins inspired by auxetic metamaterials for programmable bending of soft actuators. Mater Des 2024;246:113334.
[23] Mosadegh B, Polygerinos P, Keplinger C, Wennstedt S, Shepherd RF, Gupta U, Shim J, Bertoldi K, Walsh CJ, Whitesides GM. Pneumatic networks for soft robotics that actuate rapidly. Adv Funct Mater 2014;24(15):2163–70.
[24] Ogden RW, Saccomandi G, Sgura I. Fitting hyperelastic models to experimental data. Comput Mech 2004;34(6):484–502.
[25] Dorfmann A, Ogden R. A pseudo-elastic model for loading, partial unloading and reloading of particle-reinforced rubber. Int J Solids Struct 2003;40(11):2699–714.
[26] Tawk C, Alici G. Finite element modeling in the design process of 3D printed pneumatic soft actuators and sensors. Robotics 2020;9(3):52.
[27] Schapery RA. On the characterization of nonlinear viscoelastic materials. Polym Eng Sci 1969;9(4):295–310.
[28] Zapas L, Craft T. Correlation of large longitudinal deformations with different strain histories. J Res Natl Bur Stand Sect A, Phys Chem 1965;69(6):541.
[29] Arruda EM, Boyce MC. A three-dimensional constitutive model for the large stretch behavior of rubber elastic materials. J Mech Phys Solids 1993;41(2):389–412.
[30] Bagley RL, Torvik PJ. Fractional calculus-a different approach to the analysis of viscoelastically damped structures. AIAA J 1983;21(5):741–8.
[31] Simo JC. On a fully three-dimensional finite-strain viscoelastic damage model: formulation and computational aspects. Comput Methods Appl Mech Engrg 1987;60(2):153–73.
[32] Bergström JS, Boyce MC. Large strain time-dependent behavior of filled elastomers. Mech Mater 2000;32(11):627–44.
[33] Lu T, Wang J, Yang R, Wang T. A constitutive model for soft materials incorporating viscoelasticity and mullins effect. J Appl Mech 2017;84(2):021010.
[34] Shepherd RF, Ilievski F, Choi W, Morin SA, Stokes AA, Mazzeo AD, Chen X, Wang M, Whitesides GM. Multigait soft robot. Proc Natl Acad Sci 2011;108(51):20400–3.
[35] Suzumori K, Endo S, Kanda T, Kato N, Suzuki H. A bending pneumatic rubber actuator realizing soft-bodied manta swimming robot. In: Proceedings 2007 IEEE international conference on robotics and automation. IEEE; 2007, p. 4975–80.
[36] Polygerinos P, Wang Z, Galloway KC, Wood RJ, Walsh CJ. Soft robotic glove for combined assistance and at-home rehabilitation. Robot Auton Syst 2015;73:135–43.
[37] Martinez RV, Branch JL, Fish CR, Jin L, Shepherd RF, Nunes R, Suo Z, Whitesides G. Robotic tentacles with three-dimensional mobility based on flexible elastomers. 2013.
[38] Chen Y, Yang Y, Liu Y. A neural network peridynamic method for modeling rubber-like materials. Int J Mech Sci 2024;273:109234.
[39] Salahshoor H, Ortiz M. Application of data-driven computing to patient-specific prediction of the viscoelastic response of human brain under transcranial ultrasound stimulation. Biomech Model Mechanobiol 2024;23(4):1161–77.
[40] Ling S, Wu Z, Mei J, Lv S. An efficient machine learning-based model for predicting the stress-strain relationships of thermoplastic polymers with limited testing data. Compos Part B: Eng 2024;283:111600.
[41] Cappelli L, Balokas G, Montemurro M, Dau F, Guillaumat L. Multi-scale identification of the elastic properties variability for composite materials through a hybrid optimisation strategy. Compos Part B: Eng 2019;176:107193.
[42] Fuhg JN, Anantha Padmanabha G, Bouklas N, Bahmani B, Sun W, Vlassis NN, Flaschel M, Carrara P, De Lorenzis L. A review on data-driven constitutive laws for solids. Arch Comput Methods Eng 2024;1–43.
[43] Gorji MB, Mozaffar M, Heidenreich JN, Cao J, Mohr D. On the potential of recurrent neural networks for modeling path dependent plasticity. J Mech Phys Solids 2020;143:103972.
[44] Xu D, Zhou H, Jie T, Zhou Z, Yuan Y, Jemni M, Quan W, Gao Z, Xiang L, Gusztav F, et al. Data-driven deep learning for predicting ligament fatigue failure risk mechanisms. Int J Mech Sci 2025;301:110519.
[45] Flaschel M, Steinmann P, De Lorenzis L, Kuhl E. Convex neural networks learn generalized standard material models. J Mech Phys Solids 2025;200:106103.
[46] Liu X, Tian S, Tao F, Yu W. A review of artificial neural networks in the constitutive modeling of composite materials. Compos Part B: Eng 2021;224:109152.
[47] Maghami A, Stender M, Papangelo A. Pull-off force prediction in viscoelastic adhesive hertzian contact by physics augmented machine learning. 2025, arXiv preprint arXiv:2505.11685.
[48] Fantoni F, Bacigalupo A, Gnecco G, Gambarotta L. Multi-objective optimal design of mechanical metafilters based on principal component analysis. Int J Mech Sci 2023;248:108195.
[49] Sbriglio C, Ptak M, Kwiatkowski A. Advances in computational modelling of head injury biomechanics–a comprehensive review. Arch Comput Methods Eng 2025;1–43.
[50] González D, García-González A, Chinesta F, Cueto E. A data-driven learning method for constitutive modeling: application to vascular hyperelastic soft tissues. Materials 2020;13(10):2319.
[51] Gonzalez-Saiz E, Garcia-Gonzalez D. Model-driven identification framework for optimal constitutive modeling from kinematics and rheological arrangement. Comput Methods Appl Mech Engrg 2023;415:116211.
[52] Salahshoor H, Ortiz M. Model-free data-driven viscoelasticity in the frequency domain. Comput Methods Appl Mech Engrg 2023;403:115657.
[53] Sharma AK, Joglekar MM. A computationally efficient locking free numerical framework for modeling visco-hyperelastic dielectric elastomers. Comput Methods Appl Mech Engrg 2019;352:625–53.
[54] Nandan S, Sharma D, Sharma AK. Viscoelastic effects on the nonlinear oscillations of hard-magnetic soft actuators. J Appl Mech 2023;90(6):061001.
[55] Sharma AK, Joglekar M. A numerical framework for modeling anisotropic dielectric elastomers. Comput Methods Appl Mech Engrg 2019;344:402–20.
[56] Khurana A, Kumar A, Raut SK, Sharma AK, Joglekar MM. Effect of viscoelasticity on the nonlinear dynamic behavior of dielectric elastomer minimum energy structures. Int J Solids Struct 2021;208:141–53.
[57] Hussain A, Sakhaei AH, Shafiee M. A data-driven constitutive model for 3D lattice-structured material utilising an artificial neural network. Appl Mech 2024;5(1):212–32.
[58] Hamel CM, Long KN, Kramer SL. Calibrating constitutive models with full-field data via physics informed neural networks. Strain 2023;59(2):e12431.
[59] Ladevèze P, Chamoin L. Data-driven material modeling based on the constitutive relation error. Adv Model Simul Eng Sci 2024;11(1):23.
[60] Kirchdoerfer T, Ortiz M. Data-driven computational mechanics. Comput Methods Appl Mech Engrg 2016;304:81–101.
[61] Al-Haik M, Hussaini M, Garmestani H. Prediction of nonlinear viscoelastic behavior of polymeric composites using an artificial neural network. Int J Plast 2006;22(7):1367–92.
[62] Jordan B, Gorji MB, Mohr D. Neural network model describing the temperature- and rate-dependent stress-strain response of polypropylene. Int J Plast 2020;135:102811.
[63] Jung S, Ghaboussi J. Neural network constitutive model for rate-dependent materials. Comput Struct 2006;84(15–16):955–63.
[64] Basistov YA, Yanovsky YG, Danilin AN, Karnet YN. Dynamic neural network as a model of viscoelastic media. Compos: Mech Comput Appl: Int J 2018;9(4).
[65] Graf W, Freitag S, Sickert JU, Kaliske M. Structural analysis with fuzzy data and neural network based material description. Computer-Aided Civ Infrastruct Eng 2012;27(9):640–54.
[66] Yılmaz İ, Arslan E, Kızıltaş Ec, Cavdar K. Development of a prediction method of Rayleigh damping coefficients for free layer damping coatings through machine learning algorithms. Int J Mech Sci 2020;166:105237.
[67] Daareyni A, Baghani M, Ghezelbash F, Zakerzadeh MR. Data-driven modeling of an elastomer bushing system under various visco-hyperelastic deformations. Comput Mater Sci 2022;214:111710.
[68] Long R, Mayumi K, Creton C, Narita T, Hui CY. Time dependent behavior of a dual cross-link self-healing gel: Theory and experiments. Macromolecules 2014;47(20):7243–50.
[69] Klein DK, Fernández M, Martin RJ, Neff P, Weeger O. Polyconvex anisotropic hyperelasticity with neural networks. J Mech Phys Solids 2022;159:104703.
[70] Flaschel M, Kumar S, De Lorenzis L. Unsupervised discovery of interpretable hyperelastic constitutive laws. Comput Methods Appl Mech Engrg 2021;381:113852.
[71] Huang S, He Z, Reina C. Variational onsager neural networks (VONNs): A thermodynamics-based variational learning strategy for non-equilibrium PDEs. J Mech Phys Solids 2022;163:104856.
[72] Hernandez Q, Badias A, Chinesta F, Cueto E. Thermodynamics-informed graph neural networks. 2022, arXiv preprint arXiv:2203.01874.

[73] Lu Y, Sun Z, Elkhodary K, Xiao H, Tang S, Guo X. Introducing memory decay network for microstructured viscoelastic composites. Compos Struct 2025;354:118792.

[74] Xu K, Tartakovsky AM, Burghardt J, Darve E. Learning viscoelasticity models from indirect data using deep neural networks. Comput Methods Appl Mech Engrg 2021;387:114124.

[75] Bozkurt MO, Tagarielli VL. Data-driven homogenisation of viscoelastic porous elastomers: Feedforward versus knowledge-based neural networks. Int J Mech Sci 2025;286:109824.

[76] Mojumder S, Tao L, Li Y, Liu WK. Mechanistic data science for modeling and design of aerospace composite materials. 2021, arXiv preprint arXiv:2112.00968.

[77] Dekhovich A, Turan OT, Yi J, Bessa MA. Cooperative data-driven modeling. Comput Methods Appl Mech Engrg 2023;417:116432.

[78] Yuan Z, Niu MQ, Ma H, Gao T, Zang J, Zhang Y, Chen LQ. Predicting mechanical behaviors of rubber materials with artificial neural networks. Int J Mech Sci 2023;249:108265.

[79] Abdolazizi KP, Linka K, Cyron CJ. Viscoelastic constitutive artificial neural networks (vCANNs)–A framework for data-driven anisotropic nonlinear finite viscoelasticity. J Comput Phys 2024;499:112704.

[80] Liu F, Peng P, Tian P, Ding X, Ma T, Ye F. Mesostructure-performance analysis and dynamic modulus prediction of asphalt mixture based on a meso-characteristics-data analytics framework. Constr Build Mater 2025;498:143976.

[81] Case JC, White EL, Kramer RK. Soft material characterization for robotic applications. Soft Robot 2015;2(2):80–7.

[82] Libby J, Somwanshi AA, Stancati F, Tyagi G, Patel A, Bhatt N, Rizzo J, Atashzar SF. What happens when pneu-net soft robotic actuators get fatigued? In: 2023 international symposium on medical robotics. ISMR, IEEE; 2023, p. 1–6.

[83] Hajhashemkhani M, Hematiyan MR, Goenezen S. Identification of hyper-viscoelastic material parameters of a soft member connected to another unidentified member by applying a dynamic load. Int J Solids Struct 2019;165:50–62.

[84] Brinson HF, Brinson LC, et al. In: Polymer engineering science and viscoelasticity. An introduction, vol. 99, Springer; 2008, p. 157.

[85] As’ad F, Farhat C. A mechanics-informed deep learning framework for data-driven nonlinear viscoelasticity. Comput Methods Appl Mech Engrg 2023;417:116463.

[86] Moseley P, Florez JM, Sonar HA, Agarwal G, Curtin W, Paik J. Modeling, design, and development of soft pneumatic actuators with finite element method. Adv Eng Mater 2016;18(6):978–88.

[87] Diani J, Fayolle B, Gilormini P. A review on the mullins effect. Eur Polym J 2009;45(3):601–12.

[88] Mullins L. Effect of stretching on the properties of rubber. Rubber Chem Technol 1948;21(2):281–300.

[89] Sanchez-Palencia E, Zaoui A. Homogenization techniques for composite media. Homog Tech Compos Media 1987;272.

[90] Hori M, Nemat-Nasser S. On two micromechanics theories for determining micro–macro relations in heterogeneous solids. Mech Mater 1999;31(10):667–82.

[91] Yvonnet J, Gonzalez D, He QC. Numerically explicit potentials for the homogenization of nonlinear elastic heterogeneous materials. Comput Methods Appl Mech Engrg 2009;198(33–36):2723–37.

[92] Yang H, Guo X, Tang S, Liu WK. Derivation of heterogeneous material laws via data-driven principal component expansions. Comput Mech 2019;64:365–79.

[93] Yang H, Qiu H, Xiang Q, Tang S, Guo X. Exploring elastoplastic constitutive law of microstructured materials through artificial neural network—a mechanistic-based data-driven approach. J Appl Mech 2020;87(9):091005.

[94] Liu D, Yang H, Elkhodary K, Tang S, Liu WK, Guo X. Mechanistically informed data-driven modeling of cyclic plasticity via artificial neural networks. Comput Methods Appl Mech Engrg 2022;393:114766.

[95] Liu D, Yang H, Elkhodary K, Tang S, Guo X. Cyclic softening in nonlocal shells—A data-driven graph-gradient plasticity approach. Extrem Mech Lett 2023;60:101995.

[96] Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In: Computer vision–ECCV 2014: 13th European conference, Zurich, Switzerland, September 6-12, 2014, proceedings, part i 13. Springer; 2014, p. 818–33.

[97] LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, Jackel LD. Backpropagation applied to handwritten zip code recognition. Neural Comput 1989;1(4):541–51.

[98] Garro BA, Vázquez RA. Designing artificial neural networks using particle swarm optimization algorithms. Comput Intell Neurosci 2015;2015(1):369298.

[99] Holzapfel GA. Nonlinear solid mechanics: A continuum approach for engineering science. Kluwer Academic Publishers Dordrecht; 2002.

[100] Gorissen B, Reynaerts D, Konishi S, Yoshida K, Kim JW, De Volder M. Elastic inflatable actuators for soft robotic applications. Adv Mater 2017;29(43):1604977.

[101] Polygerinos P, Correll N, Morin SA, Mosadegh B, Onal CD, Petersen K, Cianchetti M, Tolley MT, Shepherd RF. Soft robotics: Review of fluid-driven intrinsically soft devices; manufacturing, sensing, control, and applications in human-robot interaction. Adv Eng Mater 2017;19(12):1700016.

[102] Gullerud A, Koppenhoefer K, Roy A, Dodds, jr. R. WARP3D-release 13.12. 3-D dynamic nonlinear fracture analysis of solids using parallel computers and workstations. Res Rep Struct Ser 2001;(607).

[103] Andrew AL, Tan RC. Computation of derivatives of repeated eigenvalues and the corresponding eigenvectors of symmetric matrix pencils. SIAM J Matrix Anal Appl 1998;20(1):78–100.

[104] Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I, et al. Language models are unsupervised multitask learners. OpenAI Blog 2019;1(8):9.

[105] Bengio Y, Ducharme R, Vincent P, Jauvin C. A neural probabilistic language model. J Mach Learn Res 2003;3(Feb):1137–55.