IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 23 April 2024, accepted 12 June 2024, date of publication 19 June 2024, date of current version 26 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3416872

RESEARCH ARTICLE

Bridging Diverse Physics and Scales of Knee Cartilage With Efficient and Augmented Graph Learning

SEYED SHAYAN SAJJADINIA[1], BRUNO CARPENTIERI[1], AND GERHARD A. HOLZAPFEL[2,3]

[1]Faculty of Engineering, Free University of Bozen-Bolzano, 39100 Bozen-Bolzano, Italy
[2]Institute of Biomechanics, Graz University of Technology, 8010 Graz, Austria
[3]Department of Structural Engineering, Norwegian University of Science and Technology, 7034 Trondheim, Norway

Corresponding author: Bruno Carpentieri (bruno.carpentieri@unibz.it)

This work was supported by the Open Access Publishing Fund of the Free University of Bozen-Bolzano, Italy.

ABSTRACT Articular cartilage (AC) is essential for minimizing friction in the human knee, but its healthy function is highly influenced by biomechanical factors such as weight bearing. Non-invasive biomechanical and numerical simulations are widely used to study AC but often require complex and costly numerical approximations. Machine learning (ML) provides a more efficient alternative and uses data from these numerical methods for training. Hybrid ML models (HML) complemented by reduced-order numerical models can achieve similar outcomes with minimal data input but may have problems with generalizability across different scales. In this study, we present an extended HML framework (EHML) for developing a multiscale surrogate model specifically tailored for knee cartilage simulations. Our approach is based on integrating hybrid graph neural networks (GNNs) with tissue-scale data and aims to achieve remarkable few-shot learning and potential zero-shot generalizability for large-scale analysis. The main proposed idea is a physics-constrained data augmentation (DA) technique coupled with a set of pre-processing and customization algorithms to bridge the scales. Specifically, we integrate feature transformations, resampling, and cost-sensitive functions to manage the observed data imbalances, all within a customized, memory-efficient training framework. Our rigorous testing using an advanced multi-physics cartilage model demonstrates the viability of our approach. Comparative analyses underscore the significant role of pre-processing and DA methods in enhancing generalizability and efficiency. They helped reduce the normalized mean squared errors to 0.1 or less (compared to the ablated model with its error of 2 or higher). Therefore, this work represents an important step towards addressing the challenges of limited generalizability and efficiency of existing ML-based surrogate models and opens new possibilities for their application in more complex simulations.

INDEX TERMS Biomechanics, cartilage, knee, finite element analysis, graph neural networks, machine learning.

I. INTRODUCTION

Knee biomechanics and the critical underlying structure of articular cartilage (AC), which facilitates lubrication between the articulating bones [1], have been the subject of extensive research [2], [3], [4] due to the prevalence of cartilage damage and the complex nature of the knee [5]. The increasing incidence of such damage significantly affects both the older and younger populations, especially due to biomechanical factors such as obesity and physical inactivity, placing a financial burden on the healthcare system [6]. Beyond the complex biomechanical interactions between the knee and

The associate editor coordinating the review of this manuscript and approving it for publication was Weiping Ding.

© 2024 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

its substructures, the cartilage phases, such as the fibrillar network and osmotic pressure, are strongly influenced by damage [7], making them a focal of biomechanical and orthopedic studies [8].

Physics-based and numerical algorithms, particularly finite element (FE) analysis, have been recognized for their usefulness in noninvasive and advanced biomechanical analysis of knee and cartilage mixture modeling [9], [10], [11]. These algorithms typically involve discretizing the complex modeling domain into smaller, manageable elements, but at the expense of longer computational time, making them inefficient for clinical use [12]. On the other hand, data-driven algorithms such as machine learning (ML) offer potential for similar analyses by mining biomedical datasets [13], [14], [15]. Nevertheless, the scarcity of these datasets and the costly nature of model training present challenges comparable to those of FE methods [12].

An alternative approach is to use surrogate models that employ samples of numerically generated data to train fast ML models. Although these methods have shown promise in cartilage and knee studies [16], [17], [18], [19], they are not without limitations, particularly related to the cost of training and dataset creation [12]. We recently developed a hybrid ML (HML) algorithm to address these challenges by integrating a simplified numerical model that has been experimentally demonstrated to facilitate accelerated few-shot learning of cartilage biomechanics [20]. However, this method was originally developed for single-scale studies and requires significant improvements to enable surrogate modeling of knee cartilage with all of its primary biomechanical phases at multiple scales. This need forms the motivation for our study.

In this paper, we introduce the extended HML (EHML) framework, which aims to enable efficient and multiscale surrogate modeling in the context of advanced cartilage biomechanics. Our main contributions are compared with the other related research in Section II and detailed in Section III. In short, EHML first integrates the efficiency of hybrid multi-fidelity modeling with a novel data augmentation (DA) method that aims to enable zero-shot generalization across scales while maintaining performance comparable to that of single-scale multi-physics few-shot learning. As discussed in Section IV, the training data are highly complex, imbalanced, and biased, which affects both the performance and the fairness of the evaluations. To address these issues, we explore various techniques including subgraph resampling, various loss functions, and feature compression. Moreover, we developed a customized, memory-efficient training loop to accommodate the increased complexity. Section V concludes the study and highlights the future implications of our contributions. For educational purposes and to support further research in this area, we have made the research data available as open source at github.com/shayansss/ehml.

II. RELATED WORK

A. GENERALIZABLE SURROGATE MODELING

The generalization capabilities of ML-driven surrogate models, especially those dealing with different topologies, have been evaluated in a few studies [21], [22]. They achieved this by integrating theory-based or physics-constrained features into these models. In particular, theoretically inspired feature engineering was employed to transition a model from structured to irregular mesh configurations [23], and a domain-specific feature definition was used for the analysis of truss structures, allowing adaptation to different topologies [24]. More general alternatives are graph neural networks (GNN) models that exploit relational inductive biases [25]. In particular, the message-passing implementation allows aggregating local information from one scale to predict in the other scale [25], [26].

However, training GNNs can be a significant challenge, primarily due to the need for a large number of message-passing blocks, which complicates training and can lead to oversmoothing problems in deeper layers, where signals become indistinguishable and ineffective [27]. A possible solution is to keep the number of message passings small by utilizing multi-level graph resolutions [28], [29], [30]. We hypothesize that the HML strategy could present an alternative solution without the need for multi-level input meshes. While message passings are reduced, it might not affect accuracy as the hybrid paradigm might distribute global data; thus, there is no need for a large number of message passings to incorporate these data. In addition, we implement a custom memory-efficient training algorithm along with an upstream autoencoder (AE), a method frequently employed in ML for model order and noise reduction, to potentially further improve our training.

B. AUGMENTED GRAPH LEARNING

DA techniques play a pivotal role when the available datasets are small. Among the widely used methods, the Mixup technique stands out, which interpolates new data through the convex combination of sample pairs and has applications in different fields such as natural language processing [31], [32] and image analysis [33], [34] and graph-based modeling [35], [36]. In particular, it is widely used in graph classification, but less so in regression-based problems such as surrogate modeling, for which our study provides a practical solution.

In terms of physics or engineering applications, DA has often been implemented through generative models, noise injection, and specific prepossessing transformations such as rotations to integrate configuration dependencies into tensors [37], [38], [39]. Moreover, problem-specific algorithms have been developed, e.g., to augment datasets, e.g., by using pairwise feature differences, if applicable [40] or the application of signal decomposition [41]. Additionally, in certain multi-fidelity studies, low-fidelity (LF) training samples were added to the training sets for augmentation [42]. Our research presents a novel DA specifically tailored to

multiscale numerical problems where extensive training data is not available.

C. HANDLING DATA IMBALANCE

Finally, we hypothesize that the multiscale cartilage data are highly biased and imbalanced, similar to other biomedical data [43]; therefore, it requires special treatments. We then experiment with several data-preparation transformations such as cubic and Yeo-Johnson transformations [44], [45], which effectively stabilize the variance and reduce the distribution asymmetry. This avoids the need to generate uninformative representations focused on the simpler nodes prior to training and evaluation.

Furthermore, in our surrogate-based ML problems, we adapt techniques mainly used in graph classification tasks to deal with data imbalance or data distortion [46], [47], [48]. An example is subgraphing (i.e., resampling the graph partition) with the sampling probability determined either statically by the initial distribution of nodal features or dynamically by the distribution of node errors during training. Other examples include the implementation of maximal loss, which focuses on the highest error percentages [49] or weighted loss, where weights are assigned statically or dynamically, as explained in [50] and [51].

III. METHODOLOGY

A. PRELIMINARIES

Representing first- and second-order tensors in bold, we define $\mathbf{X}$ and $\mathbf{x}$ as position vectors before and after deformation, respectively, applied to a model such as cartilage. Furthermore, the deformation gradient tensor $\mathbf{F}$ can be determined, which estimates the nonlinear deformation as

$$\mathbf{F} = \frac{\partial \mathbf{u}}{\partial \mathbf{X}} + \mathbf{I}, \tag{1}$$

where $\mathbf{u} = \mathbf{x} - \mathbf{X}$ is the deformation vector and $\mathbf{I}$ is the identity tensor. This tensor can be used to determine the updated fibril direction, e.g., for the fibril bundle I, to be determined as

$$\mathbf{n}^I = \frac{\mathbf{F}\mathbf{N}^I}{\|\mathbf{F}\mathbf{N}^I\|_2}. \tag{2}$$

Here, $\mathbf{N}^I$ and $\mathbf{n}^I$ represent the initial and updated fibril directions, respectively, and $\|\bullet\|_2$ denotes the l_2-norm of $\bullet$. Furthermore, the determinant of the deformation gradient, i.e., $J = \det \mathbf{F}$ estimates the volume ratio and can be used to obtain changes in material fractions due to deformation and fluid diffusion, i.e.

$$\rho^{\mathrm{S}} = \frac{\rho_0^{\mathrm{S}}}{J}, \quad \text{and} \quad \rho^{\mathrm{COL}} = \frac{\rho_0^{\mathrm{COL}}}{J}. \tag{3}$$

Here, ρ denotes the material fraction, where the index 0 represents its initial value, while the superscripts S and COL correspond to the solid and total collagen fractions, respectively. Moreover, we can decompose the deformation gradient into the rotation tensor $\mathbf{R}$ with orthogonality $\mathbf{R}^{\mathrm{T}}\mathbf{R} = \mathbf{I}$ and the left stretch tensor $\mathbf{V}$, which gives another deformation metric that excludes rigid body rotation. However, for computational reasons, we may prefer to use the left Cauchy-Green deformation tensor $\mathbf{B}$ as follows

$$\mathbf{B} = \mathbf{F}\mathbf{F}^{\mathrm{T}} = \mathbf{V}^2. \tag{4}$$

Thus, the logarithmic strain $\boldsymbol{\epsilon}$, a commonly used metric of deformation, can be defined by

$$\boldsymbol{\epsilon} = \ln \mathbf{B}. \tag{5}$$

This tensor can also be restricted to a specific fibril direction by

$$\epsilon^I = \log(\lambda^I). \tag{6}$$

Here $\lambda^I = \sqrt{\mathbf{n}^I \cdot \mathbf{B} \cdot \mathbf{n}^I}$ represents the fibril stretch. These deformation metrics will be used in the constitutive equations to determine the true stress tensor $\boldsymbol{\sigma}$ or one of its variants, such as the Kirchhoff stress tensor $\boldsymbol{\tau} = J\boldsymbol{\sigma}$ [52]. For theoretical details of these equations, readers are referred to [53].

FE methods can simulate the cartilage biomechanics after selecting a constitutive equation that defines the accuracy of the model. ML can serve as a surrogate for such a cartilage model trained on FE-generated data samples. In addition, for few-shot learning, the HML method presented in our previous work [20] is also used here, which uses a simplified FE model behind the ML model. The simplification can be achieved by deliberately ignoring certain aspects of constitutive behavior.

Now, by restricting the ML approach to a message-passing GNN where each node repeatedly aggregates edge representations and *vice versa*, we can learn the local behavior to be used for multiscale modeling. This approach may be feasible with the extensions described in the following section.

B. EFFICIENT AND GENERALIZABLE LEARNING

To ensure fair evaluation and enhance the training process, our approach first centralizes and balances the dataset. This can be achieved by employing the normalization and Yeo-Johnson transform, and probably another complementary nonlinear transformation after an initial analysis of the dataset. By implementing cost-sensitive approaches and resampling techniques, we ensure that the training process pays more attention to the more complex nodes or less common nodal values. A straightforward method is to apply the maximal loss [49], which focuses on the highest node error of 10%.

For more advanced methods, we first start by discretizing the range of values into n equally-distanced bins along each feature dimension. The next step is to calculate the frequency f_i of each bin i, which is the inverse of the size of the bin. These calculated frequencies are then summed to one by normalization, yielding the weight w_i, which is assigned to the values within the bin as follows

$$w_i = \frac{f_i}{\sum_{j=1}^{n} f_j}. \tag{7}$$

These static weights can then be used for applying resampling (by sampling the subgraphs) or cost-sensitive

methods (implemented by multiplying the weights with the sample-wise errors in the regular l_2 loss). This approach emphasizes the rarer values in the dataset that may still exist after the above-mentioned transformations. Alternatively, weights can be determined dynamically by calculating the normalized frequency of loss errors in the same way, focusing training on the most complex nodes.

For subgraphing [54], [55], [56], we define a subgraph for each node before training by aggregating the indices of adjacent nodes and edges depending on the number of message passings. During training, a group of subgraphs is first sampled, where except the central node (as illustrated in Fig. 1), other nodes are masked in the loss function since their semantics are affected by subgraphing. The above-mentioned weights can now be viewed as resampling probabilities. Likewise, this process can be carried out both statically and dynamically.

Next, we describe the DA strategy that involves physics-constrained nodal perturbation. In the context of the conventional graph-based Mixup method, a new nodal feature can be generated by interpolating two adjacent nodal features, $\mathbf{T}_i$ and $\mathbf{T}_j$, using a random variable $K \in [0, 1]$, expressed as $K\mathbf{T}_i+(1-K)\mathbf{T}_j$. This approach allows us to randomly select a temporal frame between recorded frames and then interpolate each nodal representation from the representations of the closest temporal frames. The use of this new inner time frame can also be interpreted as a shift in the nodal representation along the time dimension [57].

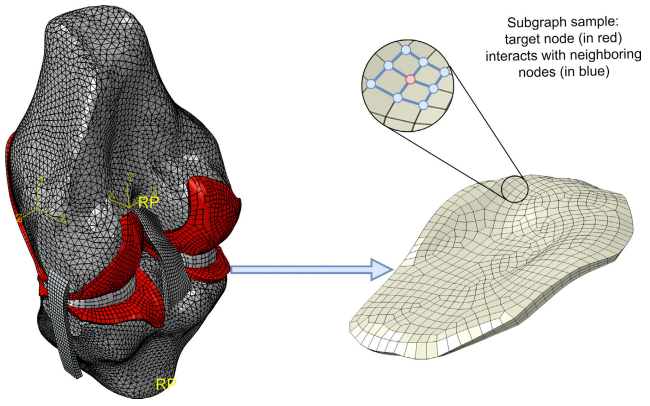

FIGURE 1. A one-hop subgraph node (in red) and its neighboring nodes and edges (in blue) within an FE cartilage mesh. In this way, the central (target) node of a subgraph still maintains the semantics.

Regarding geometric perturbations, our DA technique used a similar approach but additionally guaranteed the generation of valid geometries. This is achieved by first randomly selecting an *allowable* connecting edge for each node. As illustrated in Fig. 2, the selected edge dictates the new position of the node and ensures that it does not modify the boundary. Subsequently, the nodes are only allowed to move along one of these allowed edges at a random location between two connected nodes, which is used for a similar linear interpolation of the new nodal features. Identifying the allowable edges is straightforward in the basic geometry utilized for DA, as explained in the following section.

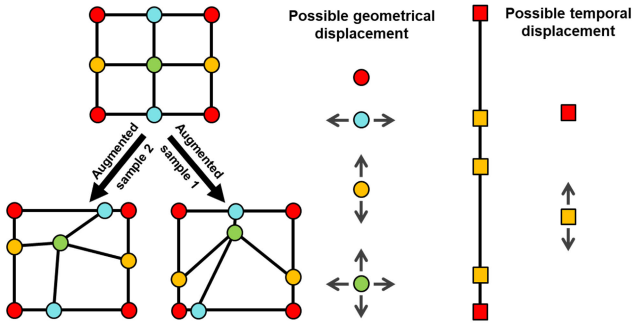

FIGURE 2. An example of DA across distinct spatiotemporal dimensions, constrained so that the nodal displacements are limited to allowable directions that have no impact on the physics, e.g., by creating an invalid shape boundary.

To complement our DA, it is also necessary to consider the effects of configuration rotations between scales by randomly rotating the corresponding features. With a random rotation matrix $\mathbf{R}$, any feature vector $\mathbf{T}_1$ (except one-hot vectors) is rotated to $\mathbf{T}_1\mathbf{R}$, whereas an arbitrary (symmetric) second-order feature tensor $\mathbf{T}_2$ is rotated to $\mathbf{R}^{\mathrm{T}}\mathbf{T}_2\mathbf{R}$. By implementing all these DA methods, new samples can be created, as displayed in Fig. 3. Despite the potential acceleration of GNN training through our hybrid approach, as the LF solver distributes boundary information, thereby reducing required message passing [58], these transformations and extensions could still introduce significant bottleneck performance.

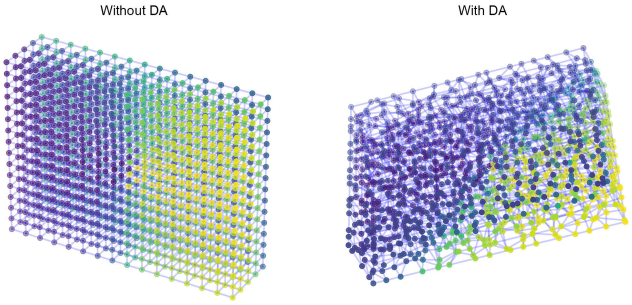

FIGURE 3. Sample of a graph before and after DA. The validity of data after augmentation can be verified by applying the inverse of augmentation transformations.

In this context, we design a custom training acceleration algorithm, as shown in Fig. 4. The trainer performs pre-processing and early stopping mechanism [59] during each training epoch. This is complemented by a subsequent inner loop dedicated to training that incorporates an additional criterion for early stopping. Introducing new data only when deemed necessary speeds up the training process. However, due to this incremental learning strategy, there is a possibility of overshooting. To counteract this, the learning rate, denoted as α, shrinks tenfold after each epoch. It is then

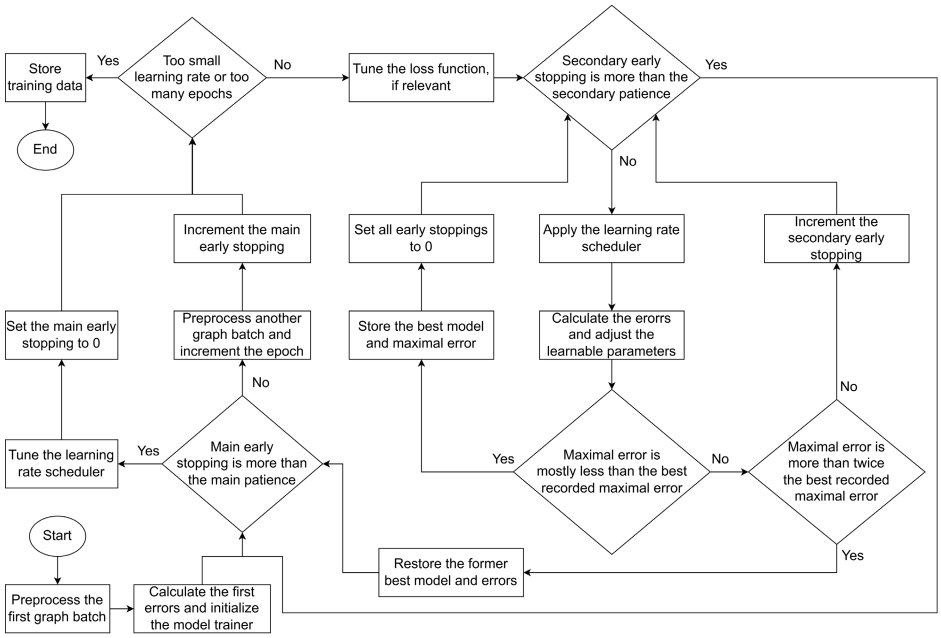

FIGURE 4. Main steps of our custom training loop involve incrementally pre-processing large sets of graphs or subgraphs in each epoch. This is executed via an inner loop that processes each batch. This speeds up training because pre-processing is performed on demand. Training efficiency is further enhanced through the use of early-stop algorithms and learning rate schedulers that help reduce overshooting and overfitting.

gradually increased in the inner loop by

$$\alpha = 10^{-i_{\text{outer}}} \left[\alpha_{\text{init}} + \min\left(\frac{i_{\text{inner}}}{N_{\text{inner}}}, 1 \right) (\alpha_{\text{final}} - \alpha_{\text{init}}) \right]. \quad (8)$$

Here i_{inner} represents the iteration number within the inner loop and N_{inner} is the patience threshold for the nested loop. The term i_{outer} indicates the number of completed epochs in which pre-processing is implemented. The parameters α_{init} and α_{final} are the initial and final values of the learning rate set during the first epoch. If an overshoot is detected, the algorithm returns to the last stored state, while the training process is stopped if the learning rate drops significantly. In this way, efficient experimental executions become possible, even on CPUs.

C. SIMULATION EXPERIMENTS

We are developing a series of cartilage simulation tests to investigate the role of DA and other extensions, particularly with regard to efficient generalization while maintaining appropriate biomechanical complexity. First, the high-fidelity (HF) and LF constitutive equations are described for use in the generation of subsequent multiscale datasets. The training experiments and the evaluation strategy are then described in detail.

1) MULTI-FIDELITY CONSTITUTIVE MODELING

Following our previous HML surrogate modeling technique [20], we can use a multiphasic model that incorporates all relevant governing equations such as the equilibrium equation, the continuity equation and Darcy's law [52], [60], [61], for HF modeling. Conversely, for solid LF modeling, a single-phase model is used, which focuses on the fundamental solid mechanical properties and omits cartilage phases along with associated equations such as the continuity equation to speed up simulation. However, in this instance, in addition to our prior HF model, we also consider the fibrillar contribution, leading to the following HF stress formulation [62], [63], [64]

$$\boldsymbol{\sigma}^{\text{HF}} = \boldsymbol{\sigma}^{\text{COL}} + \boldsymbol{\sigma}^{\text{MAT}} - \boldsymbol{\sigma}^{\text{GAG}} - P\mathbf{I}, \quad (9)$$

where P represents the fluid pressure within the cartilage's porous structure. The superscripts COL, MAT, and GAG denote the stress contributions from the fibrillar collagen network, the non-fibrillar extracellular matrix, and the osmotic pressure (induced by the electrostatic forces of entrapped glycosaminoglycans), as determined by [64], [65], [66], [67], [68], [69], and [7]:

$$\boldsymbol{\sigma}^{\text{COL}} = \frac{\phi_0^{\text{S}}}{J} \sum_{I=1}^{9} \max\left(\rho_{\text{C}}^{I} \lambda^{I} \epsilon^{I} (E_1 + E_2 \epsilon^{I}),\, 0 \right) \mathbf{m}^{I}, \quad (10)$$

$$\boldsymbol{\sigma}^{\text{MAT}} = \phi_0^{\text{S}} G_m \frac{1 - \rho_0^{COL}}{J} \left[\mathbf{B} - J^{2/3}\mathbf{I} + \frac{\ln J}{2} \left(\frac{1}{3} + \frac{J + \phi_0^{\text{S}}}{J - \phi_0^{\text{S}}} - \frac{\phi_0^{\text{S}} J \ln J}{(J - \phi_0^{\text{S}})^2} \right) \mathbf{I} \right], \quad (11)$$

$$\boldsymbol{\sigma}^{\text{GAG}} = \alpha_1 J^{-\alpha_2} \mathbf{I}. \quad (12)$$

Here, $\mathbf{m}^{I}$ is the structural tensor, i.e. the dyadic product of the updated fibril direction with itself, α_1 and α_2 are material constants that govern the exponential function of the

osmotic pressure formula, G_m is a material constant of the modified neo-Hookean model, E_1 and E_2 are constants that control the nonlinearity of fibrils, and ρ_{C}^{I} denotes the volume fraction of the associated fibril bundle. Seven isotropic fibril directions are assumed to exist at each point, but two others are organized anisotropically by the so-called arcade organization [67], in which the fibrils align parallel to the split lines of the cartilage surface and rotate perpendicular to the subchondral bones.

The typically depth-dependent material parameters are adopted from our calibrated 2D model [70], which is extendable to a 3D simulation [4] using first-order brick elements with pore pressure FEs for solution approximation in Abaqus soil consolidation analysis [71]. Initial conditions, which are influenced by pre-existing tissue stresses and are therefore unknown, are estimated through an optimization process that compares pre-stressed conditions with experimental observations [72]. This is implemented by a Python script in Abaqus and Fortran subroutines for constitutive equations, as detailed in [4], [70], and [12]. After the initial conditions are established, subsequent HF analyses are performed using the mechanical and biochemical boundary conditions.

As for the LF analysis, Hook's law was adopted in our previous work [20], which assumes single-phase elastic and isotropic behavior. Accordingly, using the Young's modulus E and Poisson's ratio ν, the corresponding elastic Kirchhoff stress tensor becomes [52]

$$\boldsymbol{\tau}^{\mathrm{EL}} = J\frac{1+\nu}{E}\left(\boldsymbol{\epsilon} + \frac{\nu}{1-2\nu}\mathrm{tr}\boldsymbol{\epsilon}\mathbf{I}\right). \tag{13}$$

This time, to obtain the LF stress, we also apply a simple viscosity level by using the material constants $\mathcal{K}$, $\mathcal{G}$ and $\mathcal{T}$ to include the temporal response (without direct inclusion of fluid variation), as supported by Abaqus [71], [73], [74], i.e.

$$\boldsymbol{\sigma}_t^{\mathrm{LF}} = \frac{1}{j_t}\Bigg[\boldsymbol{\tau}_t^{\mathrm{EL}} - \frac{\mathcal{K}}{3\mathcal{T}}\left(\int_0^t \mathrm{tr}(\boldsymbol{\tau}_{t-s}^{\mathrm{EL}})e^{-s/\mathcal{T}}\,\mathrm{d}s\right)\mathbf{I} - \frac{\mathcal{G}}{\mathcal{T}}\left(\int_0^t \bar{\mathbf{F}}_t\bar{\mathbf{F}}_{t-s}^{-1}\left(\boldsymbol{\tau}_{t-s}^{\mathrm{EL}}\right)^D\bar{\mathbf{F}}_{t-s}^{-T}\bar{\mathbf{F}}_t^T e^{-s/\mathcal{T}}\,\mathrm{d}s\right)^D\Bigg]. \tag{14}$$

Here we use time subscripts to distinguish parameters at different time frames, while superscript D represents the deviatoric parts of the stress tensors and $\bar{\mathbf{F}} = J^{-1/3}\mathbf{F}$ is the distortion part. The first term of this equation represents the instantaneous elastic response, while the other terms decay the solution by the introduced distortion gradients to account for configuration rotation. With the discretized integrals, the solution can also be approximated using a regular quasi-static Abaqus step and its linear (purely displacement-based) brick FEs.

Finally, by applying the mechanical boundary conditions used in HF analyses, the LF simulations can be executed with (14), adapting the material parameters of [75]. We therefore set $\mathcal{G} = 0.744$, $\mathcal{K} = 0.978$, $\mathcal{T} = 13.3$ sec, and $\nu = 0.47$, whereas the elasticity is set to a large value of $E = 37.8$ MPa to improve numerical convergence and also to better distribute the boundary information within the model, thereby improving the message passing in our hybrid GNN implementation.

2) MULTISCALE DATASET GENERATION

We first generate 50 training samples from modeling of a $7.5 \times 7.5 \times 1.9\,\mathrm{mm}^3$ cubic mesh (modeling cartilage) and a semi-elliptic mesh (modeling a meniscus of an indenter shape) with maximum and minimum radii of 2 and 0.5 mm, respectively. The dataset is divided into three equal subsets for training, validation, and in-distribution evaluation. These data need to be augmented by our DA technique, for which we can either define the allowed edges manually or do it automatically by first collecting the indices of all connected edges for each node and then ordering them by their number. This method ensures that nodes with fewer connections (e.g., corner nodes) are not allowed to move during DA, which means that all their connection edges are considered disallowed. Conversely, the inner nodes with the most connections can treat all their connections as allowable edges, allowing nodal movements on any of them during the DA. In general, it is assumed that edges connected to lower-rank nodes are allowed to ensure that the boundaries remain unaffected during DA.

To sample the input variables of the FE models, we use uniform distributions U. Accordingly, the indenter compresses the tissue with axial displacements $U(0,\ 0.6\,\mathrm{mm})$ applied to a reference point coupled to the planar surface or indenter mesh. Likewise, linear displacements parallel to the flat surface of the indenter are imposed by $U(0.5,\ 0.5\,\mathrm{mm})$, while the rotational displacements are imposed by $U(-0.1,\ 0.1\,\mathrm{rad})$. In addition, the bottom of the cubic mesh is fully constrained to mimic the subchondral bones. However, these only generate a training dataset to be augmented, while the out-of-distribution evaluation is carried out with the open-knee [76] tibiofemoral joint (as displayed in Fig. 5). Application to the human data is legalized as indicated in the source [77].

In the large-scale analysis, we use the boundary conditions similar to our previous study [20]. The only exception is our assumption that no fluid diffusion occurs at the surface. We are aware that this, combined with firm assumptions about the geometry and material behavior of the indenter, can have (at least) an impact on the accuracy of boundary modeling. However, this approach allows us to efficiently collect small datasets while incorporating constitutive complexity for subsequent surrogate modeling assessment.

We simulate the effects of human weights on the knee using a uniform distribution $U(-1000,\ 0\,\mathrm{N})$, while the simulation time is sampled from $U(0.001,\ 100\,\mathrm{sec})$ in all simulations on both scales. We further randomize the resulting strain rates by randomly weighting the applied loads and boundary conditions. Four equally distanced weighting multipliers are then created from $U(0,\ 1)$ and interpolated over the entire

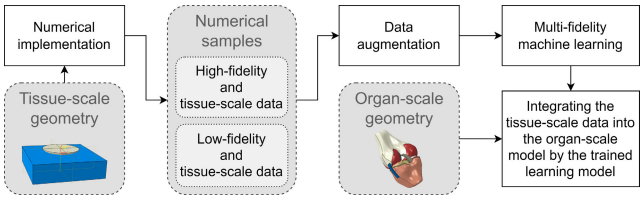

FIGURE 5. Main steps of data generation for training and evaluations.

time range using a scaled cubic spline interpolator [78] to minimize extreme changes. With these inputs, we first carry out the HF simulations and at the same time store the time points of the recorded frames. These points are used to synchronize the interpolated output frames of the calculated LF data. The HF outputs, which represent the stress contributions of cartilage phases, are considered nodal targets. The mesh topology, primary fibril orientations, and the LF stress-strain results are used to encode input graph features as implemented in our previous work [20].

More specifically, the weight is modeled as a concentrated load on the femur. While this approach does not consider non-uniform weight distribution, it still provides a reasonable approximation of human weight during standing. Previously calibrated and validated constitutive models are applied, and simulations are run until the most possible convergence is achieved within a reasonable runtime, making it suitable for experimentation. A limited number of small-scale training samples are used to mimic data scarcity in biomechanical analyses. Although these factors might affect the biomechanical accuracy, we adopted conditions similar to those used in previous studies to minimize inaccuracies. More application-oriented and accurate analyses are left for future studies (as, here, we simplify modeling while retaining the biomechanical complexity).

Finally, due to the high computational costs of knee simulations, only six samples outside the distribution are managed to be randomly generated to estimate surrogate generalizability. To speed up the simulation process, in Abaqus we assign an arbitrarily high threshold for the maximum variation in fluid pressure during each increment. This approach is occasionally used in similar large-scale cartilage studies, see, e.g., [79].

3) TRAINING AND EVALUATION

Once the datasets are collected, they are prepared for training by first finding the hyperparameters for normalization and applying the Yeo-Johnson transformations, which are approximated using training data samples. During training, our on-the-fly pre-processing system creates graphs containing all temporal frames using the nodal and target features as well as the Euclidean nodal distance and its value as edge features. Next, temporal sampling is done by randomly selecting one or two frames – either close together or interpolated – with also 2 and 5 message passing blocks, respectively. Furthermore, we apply DA to these graph frames and also include the relative time intervals as an additional node feature. A large graph batch is then assembled for training purposes, combining these graphs through disjoint unions of 100 sampled graph frames. Then, while the edge features are only prepared by the normalization transformation, the other features are fully transformed for training or inference.

In our experiments, we first address the application of various loss functions and subgraph resampling, which are described in detail in the following section. Furthermore, we introduce a slight Gaussian noise (with a standard deviation of 0.01) to the node features and explore the utility of an AE with a single compression layer. We will report the compression ratio (CR), which indicates the proportion of the size of the input node features to the dimensionality of the latent layers. Then, by setting the primary and secondary patience values to 2 and 100, respectively, training begins with a learning rate of 0.1 and ends when it falls below 10^{-5}. For experiments with resampling, we set the hyperparameters to collect 1 000 subgraphs per batch, using 50 bins in dynamic resampling and 2 bins in static resampling to mitigate the curse of dimensionality. Other hyperparameters remain the same as the ones in our previous work [20].

Evaluation metrics include mean squared errors (MSEs) and pointwise MSEs (PMSEs) introduced in our previous work [80] for transformed targets, allowing a fair comparison of all outputs. In fact, the original feature distributions are very imbalanced because most of the data revolves around a certain common value, as will be shown in the next section, which can negatively impact both learning and evaluation. Pre-processing transformations align the ranges of feature values to comparable normalized or even uniform distributions, with cubic transformations further refining error sensitivity to small output errors. Given the high cost and inherent approximations in numerical analysis, we aim for a margin error similar to the errors recorded in our previous work [20], i.e., between 0.1 and 0.001.

For test evaluation, the in-distribution batch size is 100 graphs, while for out-of-distribution evaluation, it is limited to 5 to accommodate memory limitations for large-scale analyses. This adjustment may randomly affect assessment results under certain conditions. However, our analysis focuses primarily on general trends across a variety of conditions rather than individual sample errors. Moreover, our model assumptions – including hyperparameters, training settings, and configurations for numerical analysis – could be further diversified, particularly for smaller batches, to improve biomechanical fidelity. However, our experimental setup is specifically designed to efficiently assess extensions while achieving sufficiently small errors. Therefore, this is not considered a significant limitation.

IV. RESULTS AND DISCUSSION

This study addresses a significant challenge in surrogate modeling, with a particular emphasis on knee cartilage modeling, namely zero-shot learning across different scales. To improve few-shot learning with small-scale training

TABLE 1. Means and standard deviations of runtimes in minutes for each group of numerical models. Relative runtime differences between 15 and 38 were recorded for different fidelities. Furthermore, while the runtimes of the large-scale models were still simplified to collect some evaluation samples, they were 2 to 6 times slower than their small-scale counterparts, justifying the use of multiscale surrogate models.

Model fidelity	Small-scale simulation	Large-scale simulation
High	21.21 ± 2.16	118.31 ± 6.42
Low	1.31 ± 0.69	3.07 ± 0.27

datasets, we used our previously developed HML framework [20]. This framework integrated an upstream, simplified FE model into the surrogate structure. A key feature of our hybrid model is its ability to distribute non-local data across the LF model, reducing the need to employ resource-intensive GNN architectures. This approach, combined with a memory-efficient training loop, enables optimized and efficient ML training and inference processes.

Specifically, rotational and Mixup-based DA, constrained by physics, were employed along with other extensions. A hybrid approach was implemented, utilizing simpler models without regularization. The effects of the extended version were also minimized to compare it with the basic HML model (without extensions). While the modeling conditions and parameterization have been described, readers interested in further details are encouraged to consult the open-source research code and its documentation.

In the next section, we focus on describing the runtimes of the numerical analyses, which represent the main bottleneck of this study. This bottleneck is particularly detailed because our numerical analyses involved a sophisticated cartilage model that includes fibrils and other phases. We will demonstrate the computational efficiency achieved by limiting training to small-scale analysis while the evaluations are carried out on a limited selection of randomly generated large-scale samples. We then present an empirical evaluation of the extensions we added to our HML framework. It was conducted at two different scales and is in alignment with the multiscale aspect of this research. The section ends with a discussion on the inherent limitations of our study and future research directions.

A. COMPUTATIONAL GAIN

A series of cartilage simulations at both tissue and tibiofemoral scales were performed with a standard setting (using a Core i5 CPU) and at various fidelities, as listed in Table 1. Due to the use of a more complex cartilage model, a more significant decrease in runtime was recorded compared to our previous study [20].

Training of the experiments was then implemented on a GPU within a few hours, which is particularly efficient compared to the hours to days of training typical for regular GNN-based surrogates [81]. In parallel, pre-processing was also carried out on a CPU. However, both training and pre-processing could have been executed sequentially on a single standard CPU.

B. TISSUE-SCALE SIMULATION

A series of data preparation transformations were applied to address the imbalance and skewness in the nodal data of the inputs and outputs, which is consistent with our original hypothesis about such data characteristics, as illustrated in Fig. 6. These transformations were performed on the ablated pipeline both individually and in combination before the training phase. The effectiveness of these transformations is obvious, if handled improperly, it could have resulted in biased training with misleading evaluation [82].

The normalization transformation successfully reduced data imbalance but was less effective in correcting the skewness presented in the data. On the other hand, the Yeo-Johnson transformation significantly reduced data skewness, particularly for input features. Combined with normalization, this showed promise in terms of normalizing the dataset and reducing skewness. Despite these improvements, the imbalance remained, mostly close to zero, likely indicating unaffected cartilage areas. The addition of a cubic transformation alongside these methods appeared to further refine the dataset. The combined transformation not only eliminated the remaining imbalances but also resulted in a more evenly distributed dataset. This could prevent training and evaluation from being excessively controlled by the majority of simple, unaffected nodes.

Next, the surrogate models were trained through a set of different loss functions and resampling strategies with or without upstream AEs, as shown in Figs. 7 and 8. In the appendix, the test errors were also averaged and listed in Tables 2 and 3. It was noted that in most cases a single graph batch generated in the first epoch and trained iteratively through an inner loop was sufficient to significantly reduce errors to the desired value range. However, for models with lower CR and increased complexity, using more training batches could be beneficial, but at the expense of more time-consuming preprocessing.

Our results showed no obvious improvement by applying AE, which is probably due to the lack of hyperparameter search. Interestingly, traditional loss functions performed better, especially in complex models (with the lower CR and higher number of message-passings). It is important to recognize that these improvements do not necessarily correspond to improved generalization abilities, for which out-of-distribution evaluation (on large-scale samples) will be discussed. Overall, the similar range of converged results between all cases strengthens our hypothesis that HML enables the use of simpler, more efficient models. Further simplifying the architecture does not necessarily mean better generalization [83], but accelerates training [84].

C. KNEE-SCALE SIMULATION

Figure 9 shows a detailed analysis of out-of-distribution MSEs, further detailed in Tables 4 to 11 in the appendix. Due to memory limitations, a limited batch size was used for the knee-scale simulation. This could lead to some

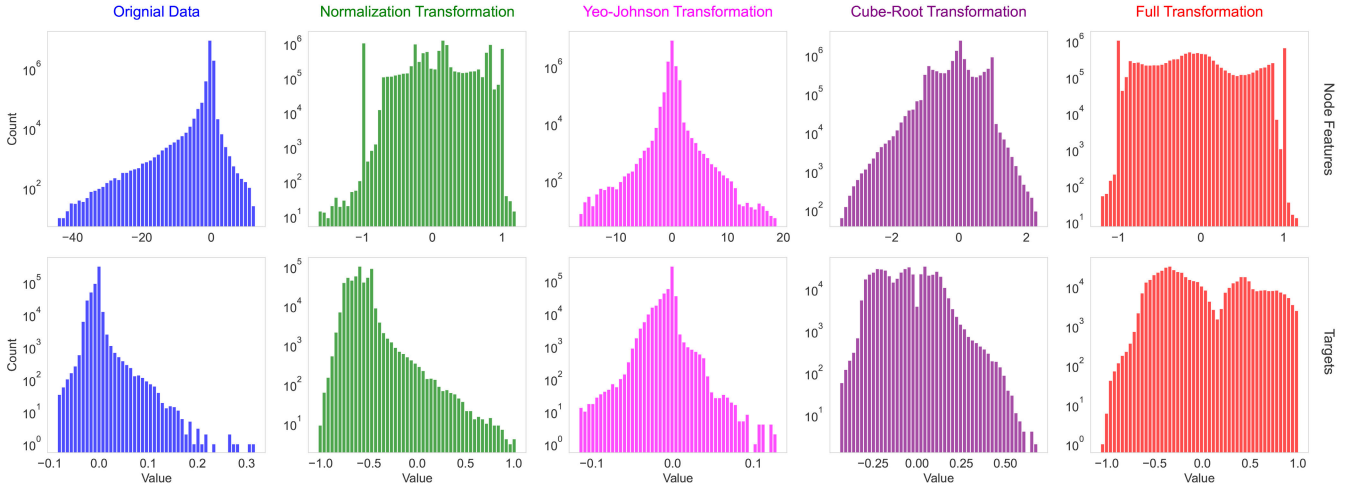

FIGURE 6. Comparison of the effects of various transformations applied individually to each nodal input and output feature dimension, while they are summarized here for visualization. The transformations helped achieve a more standardized normalized distribution when used individually and a more uniform distribution when combined. Such data preparation is critical because it addresses the problem that average evaluation errors are unduly influenced by a majority of simpler nodes rather than a minority of complex nodes, which are typically located in boundary areas with the largest values. These are areas in which the cartilage is exposed to concentrated effects of applied loads and deformations, especially at high loading rates, e.g., at the contact surfaces.

bias in the recorded errors due to the variability of the sampling distribution. Nevertheless, a clear trend can be seen in the data collected if the outliers are ignored. As before, the use of AEs did not result in significant performance improvements, whereas DA had a remarkably positive effect by training on more diverse mesh configurations. Comparing all implementations, we find that the augmented models using a weighted loss approach consistently have the smallest MSEs with significantly lower variance.

We further investigated the performance of our proposed graph-based DA method by analyzing the averaged PMSEs, as shown in Fig. 10. We discovered that the standard loss function l_2 resulted in the highest error rates, whereas using weighted losses significantly reduced these errors, making them almost negligible in comparison. While the maximal loss approach may be beneficial in simpler surrogate models, it requires a considerably large amount training batches. Furthermore, methods based on subgraph resampling were not better than those employing weighting strategies, which may be due to the need for hyperparameter optimization. Dynamic variants of these techniques, requiring minimal tuning, demonstrated improved performance compared to their static equivalents when combined with DA. Regardless of the technique used to resolve data imbalances, the inclusion of our DA consistently resulted in a significant reduction in errors, further validating its usefulness.

D. LIMITATIONS AND FUTURE WORK

Here we need to report multiple limitations in our experiments. First, the specificity of the subgraphing implementation represents a potential limitation. In particular, we limit resampling to each training epoch rather than each step to avoid performance bottlenecks. Although this technique was effective in our specific tests, it still has room for further optimization and acceleration [85]. Second, although we successfully modeled the surrogate of the HF cartilage multiphasic behavior, the numerical modeling and the applied boundary conditions were specifically tailored to circumvent the challenges of generating costly HF assessment data. While this is a restrictive limitation, it still allowed us to generate datasets of advanced biomechanical simulations for our experiments, effective for our particular assessments.

In the future, we can design and determine various strategies to create efficient and informative training datasets for more advanced simulations accelerated using the EHML technique. Sample-by-sample validation of the trained surrogate model may not be generalizable to unseen data, especially as expensive large-scale analysis may exclude HF test samples. While employing a more diverse testing and evaluation strategy could be a feasible solution, ideally we may need extensive *in vitro* or *in vivo* tests to assess generalizability. These go far beyond the scope of this research, which focused on efficiently evaluating the benefits of the EHML method. Nevertheless, such biomechanical experiments are essential for the evaluation of more sensitive and application-oriented studies.

V. CONCLUSION

In this research, we proposed and empirically experimented the EHML technique, extending our previous HML [20] to enable efficient training of multiscale surrogate modeling in advanced cartilage biomechanics. The first major enhancement of EHML involves the integration of all primary tissue phases, including the fibril component, into surrogate modeling. We then present a novel DA method that generates new graphs with nodal perturbation while preserving the

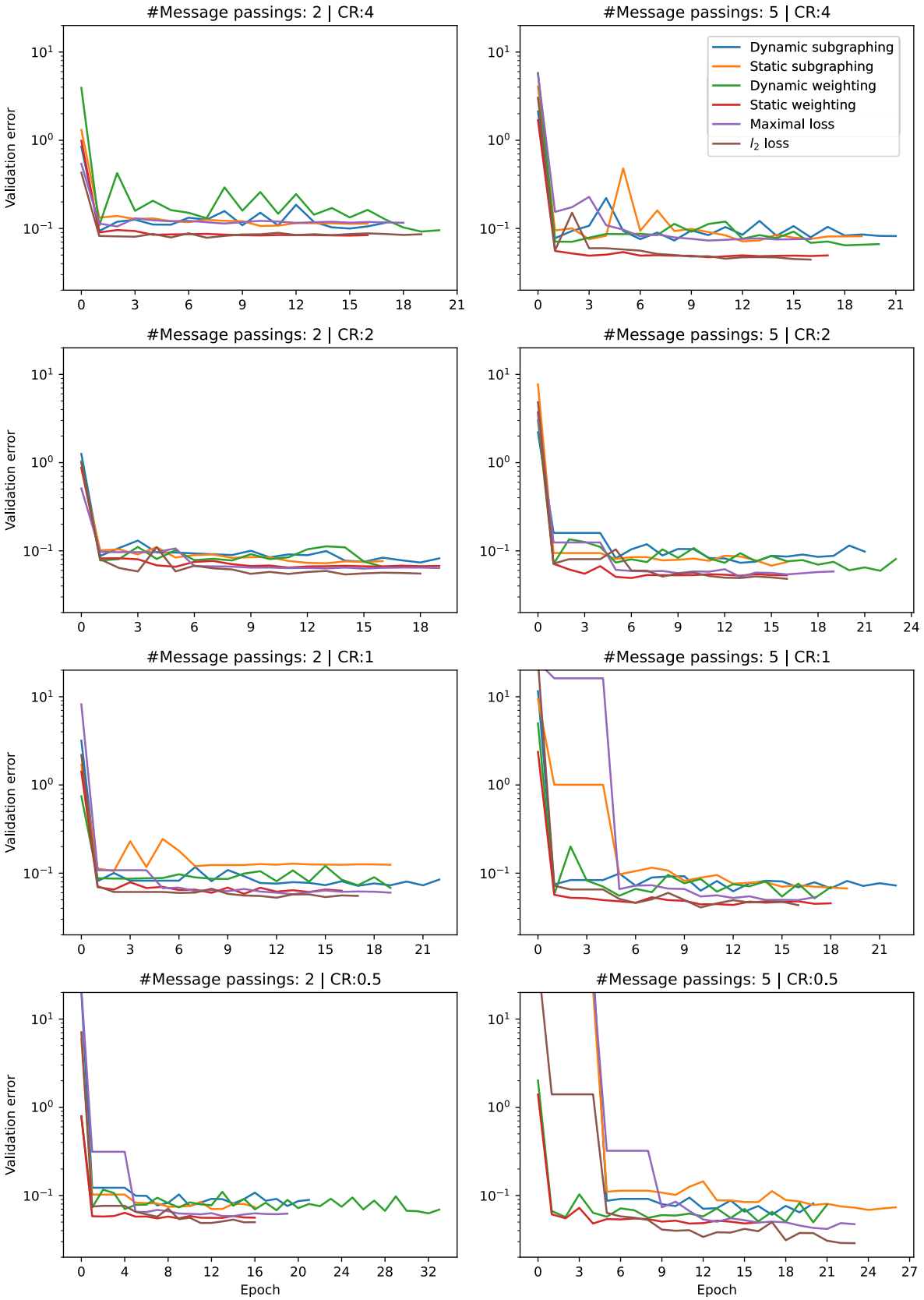

FIGURE 7. Comparing validation (mean-squared) errors vs. epochs for surrogates with AE.

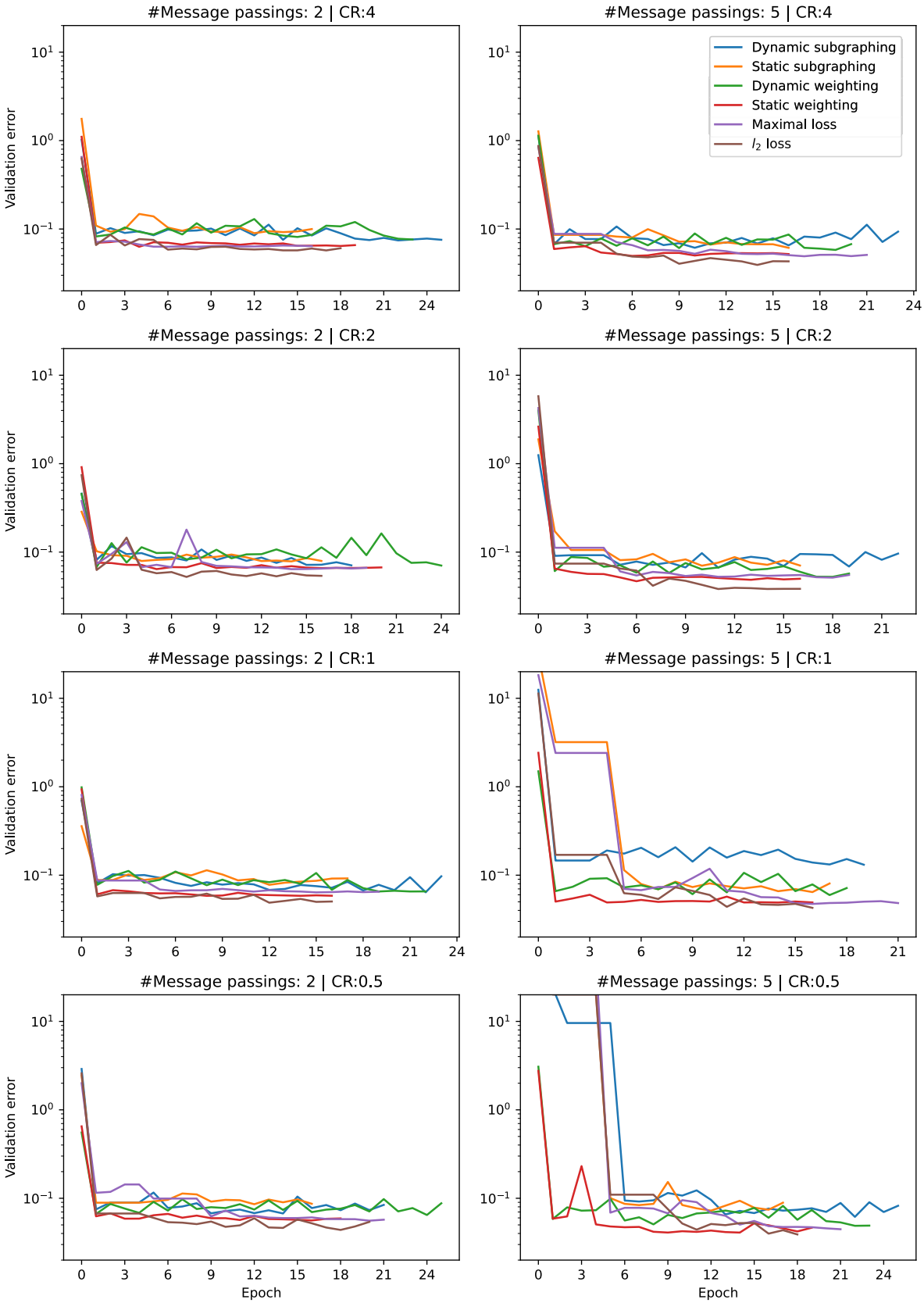

FIGURE 8. Comparing validation (mean-squared) errors vs. epochs for surrogates without AE.

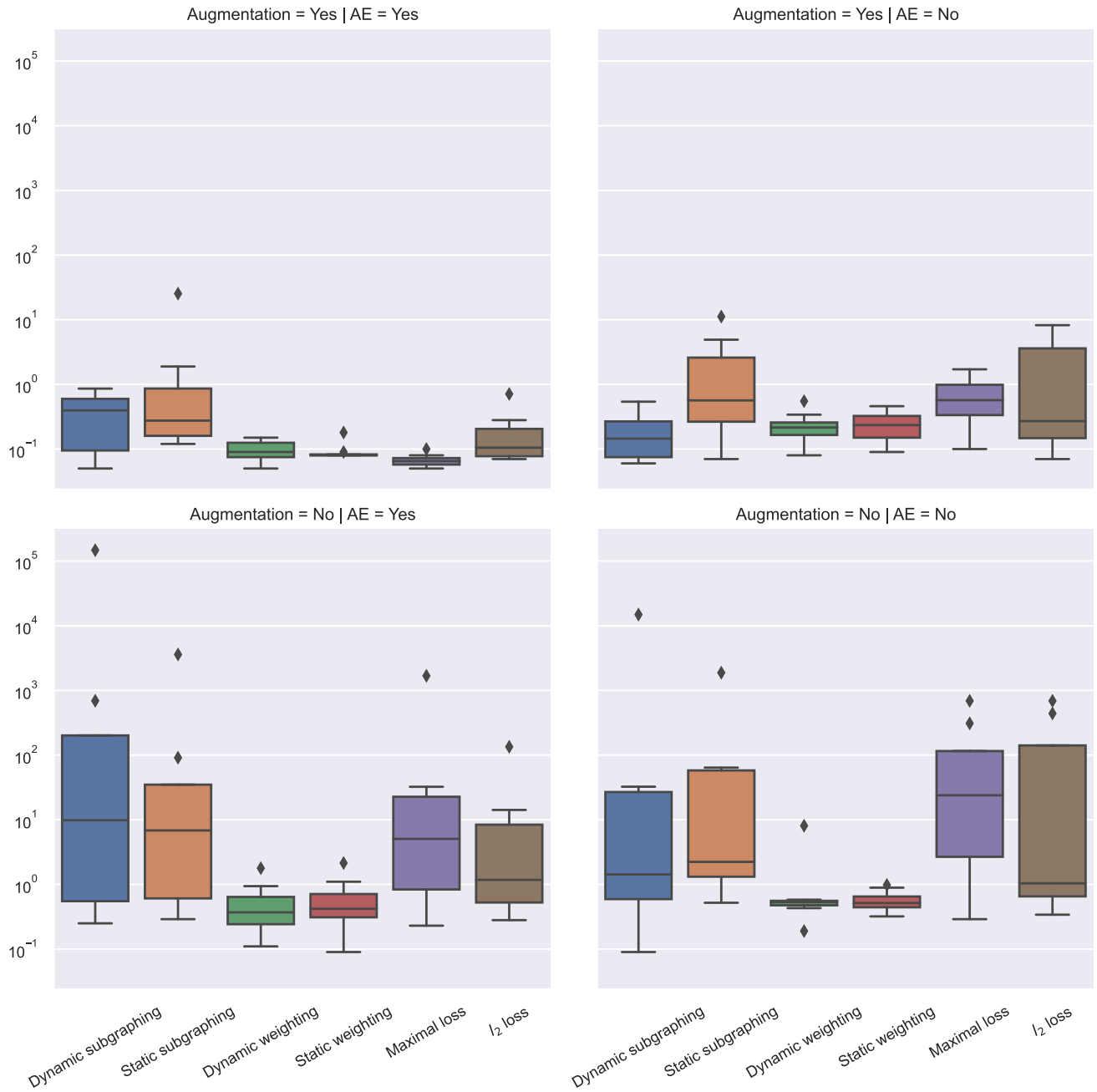

FIGURE 9. Averaged knee-scale (out-of-distribution) test errors for different models trained with different resampling and loss function implementations.

TABLE 2. In-distribution MSE of the small-scale model with message passings: 2.

Model type	Ratio of compression: 4		Ratio of compression: 2		Ratio of compression: 1		Ratio of compression: 0.5	
	Without AE	With AE	Without AE	With AE	Without AE	With AE	Without AE	With AE
Dynamic subgraphing	0.08	0.13	0.11	0.08	0.07	0.07	0.07	0.08
Static subgraphing	0.15	0.17	0.09	0.25	0.10	0.15	0.10	0.99
Dynamic weighting	0.09	0.27	0.07	0.19	0.08	0.40	0.07	0.53
Static weighting	0.08	0.21	0.06	0.31	0.07	0.41	0.07	0.34
Maximal loss	0.08	0.17	0.07	1.10	0.08	0.76	0.08	1.44
l_2 loss	0.07	0.13	0.08	0.20	0.08	0.18	0.05	0.34

shape of the model. In this way, we could approximate the element configuration of the large-scale model from small-scale data, enabling zero-shot generalization. Additionally, our results revealed significant data imbalances and biases,

FIGURE 10. Averaged pointwise knee-scale (out-of-distribution) errors when testing samples for different surrogates trained with various resampling and loss functions. Models are particularly compared by ablating important data augmentation blocks, among others, showing the most significant influence.

TABLE 3. In-distribution MSE of the small-scale model with message passings: 5.

Model type	Ratio of compression: 4		Ratio of compression: 2		Ratio of compression: 1		Ratio of compression: 0.5	
	Without AE	With AE	Without AE	With AE	Without AE	With AE	Without AE	With AE
Dynamic subgraphing	0.06	0.10	0.08	0.09	0.17	0.06	0.10	0.07
Static subgraphing	0.10	0.41	0.10	0.30	0.09	0.28	0.10	1.11
Dynamic weighting	0.09	0.26	0.07	0.30	0.07	0.22	0.08	0.23
Static weighting	0.08	0.20	0.08	0.32	0.08	0.16	0.07	0.19
Maximal loss	0.06	1.80	0.06	0.33	0.07	7.48	0.07	1.10
l_2 loss	0.06	0.36	0.09	0.16	0.08	26.07	0.07	0.19

TABLE 4. Out-of-distribution MSE of the full model with CR 4 and number of message passings 2.

Model type	With AE and DA	Without AE but with DA	With AE but without DA	Without AE and DA
Dynamic subgraphing	0.13	0.10	0.45	0.16
Static subgraphing	0.47	0.16	1.26	0.94
Dynamic weighting	0.09	0.19	0.43	14.83
Static weighting	0.12	0.67	0.55	0.99
Maximal loss	0.09	0.16	0.51	19.73
l_2 loss	0.13	0.11	1.08	1.52

TABLE 5. Out-of-distribution MSE of the full model with CR 4 and number of message passings 5.

Model type	With AE and DA	Without AE but with DA	With AE but without DA	Without AE and DA
Dynamic subgraphing	1.28	0.15	7.42	2.52
Static subgraphing	0.17	0.59	18.65	2.68
Dynamic weighting	0.21	0.30	0.22	0.96
Static weighting	0.11	0.24	0.16	1.94
Maximal loss	0.11	2.35	0.47	0.78
l_2 loss	0.11	0.43	12.06	1.15

TABLE 6. Out-of-distribution MSE of the full model with CR 2 and number of message passings 2.

Model type	With AE and DA	Without AE but with DA	With AE but without DA	Without AE and DA
Dynamic subgraphing	0.84	0.09	0.62	1.22
Static subgraphing	0.25	0.53	0.71	4.61
Dynamic weighting	0.15	0.37	0.40	0.80
Static weighting	0.13	0.50	2.05	0.72
Maximal loss	0.09	0.79	1.87	0.52
l_2 loss	0.11	0.24	0.55	2.32

TABLE 7. Out-of-distribution MSE of the full model with CR 2 and number of message passings 5.

Model type	With AE and DA	Without AE but with DA	With AE but without DA	Without AE and DA
Dynamic subgraphing	0.09	0.14	66.11	56.07
Static subgraphing	2.83	2.62	163.23	100.30
Dynamic weighting	0.10	0.34	1.69	0.89
Static weighting	0.25	0.28	0.99	0.51
Maximal loss	0.10	0.20	3018.96	1279.40
l_2 loss	0.39	4.29	26.25	1299.88

TABLE 8. Out-of-distribution MSE of the full model with CR 1 and number of message passings 2.

Model type	With AE and DA	Without AE but with DA	With AE but without DA	Without AE and DA
Dynamic subgraphing	0.14	0.33	0.97	0.65
Static subgraphing	0.33	0.08	0.44	1.80
Dynamic weighting	0.13	0.46	0.88	0.86
Static weighting	0.12	0.47	0.59	0.99
Maximal loss	0.13	0.90	6.06	68.72
l_2 loss	0.09	0.19	0.44	0.53

which we addressed through a combination of subgraph resampling and various loss functions. While this increases the complexity of the pre-processing and training stages, it was effectively handled by our custom, memory-efficient training loop.

In addition, the trained surrogate models were able to demonstrate sufficient accuracy even with significantly simplified architectures. These simplifications include reducing the number of message-passing steps and time frames. Additional efforts to further improve our method included implementing upstream AE. However, this did not bring any substantial improvement. In contrast, our other improvements, especially DA, have markedly improved the scalability and generalizability of the surrogates, successfully bridging the scales. With the increasing acceptance of similar multiscale models in biomechanics, we hope that our open-source method provides a new template for efficient implementation in related studies.

TABLE 9. Out-of-distribution MSE of the full model with CR 1 and number of message passings 5.

Model type	With AE and DA	Without AE but with DA	With AE but without DA	Without AE and DA
Dynamic subgraphing	0.78	0.72	1159.24	25001.22
Static subgraphing	37.67	15.85	5983.59	3685.13
Dynamic weighting	0.17	0.20	0.41	0.84
Static weighting	0.11	0.17	0.76	0.61
Maximal loss	0.12	0.82	12.27	554.70
l_2 loss	0.89	15.37	2.90	792.63

TABLE 10. Out-of-distribution MSE of the full model with CR 0.5 and number of message passings 2.

Model type	With AE and DA	Without AE but with DA	With AE but without DA	Without AE and DA
Dynamic subgraphing	0.58	0.77	29.05	2.83
Static subgraphing	0.18	0.85	6.17	3.51
Dynamic weighting	0.10	0.68	0.68	0.82
Static weighting	0.11	0.36	0.48	0.78
Maximal loss	0.07	1.14	35.30	5.81
l_2 loss	0.16	7.86	1.05	1.05

TABLE 11. Out-of-distribution MSE of the full model with CR 0.5 and number of message passings 5.

Model type	With AE and DA	Without AE but with DA	With AE but without DA	Without AE and DA
Dynamic subgraphing	0.58	0.25	276541.50	46.19
Static subgraphing	0.75	6.94	29.73	118.65
Dynamic weighting	0.17	0.27	3.32	0.37
Static weighting	0.13	0.18	4.06	1.72
Maximal loss	0.08	1.92	60.22	94.65
l_2 loss	0.25	0.32	250.52	76.40

APPENDIX
DETAILS OF EVALUATION ERRORS

Tables 2 and 3 describe in-distribution errors, while Tables 4 to 11 detail out-of-distribution errors.

REFERENCES

[1] X. L. Lu and V. C. Mow, "Biomechanics of articular cartilage and determination of material properties," *Med. Sci. Sports Exerc.*, vol. 40, no. 2, pp. 193–199, 2008, doi: 10.1249/mss.0b013e31815cb1fc.

[2] S. B. Kwon, H.-S. Han, M. C. Lee, H. C. Kim, Y. Ku, and D. H. Ro, "Machine learning-based automatic classification of knee osteoarthritis severity using gait data and radiographic images," *IEEE Access*, vol. 8, pp. 120597–120603, 2020, doi: 10.1109/ACCESS.2020.3006335. https://doi.org/10.1109/ACCESS.2020.3006335

[3] T. M. Said, A. M. Khateeb, and A. M. Gody, "Development of a theoretical microwave model to predict the dielectric properties of articular cartilage tissues," *IEEE Access*, vol. 9, pp. 161030–161037, 2021, doi: 10.1109/ACCESS.2021.3132691. https://doi.org/10.1109/ACCESS.2021.3132691

[4] S. S. Sajjadinia, B. Carpentieri, and G. A. Holzapfel, "Large-scale finite element modeling of pre-stress in articular cartilage," in *Computer Methods in Biomechanics and Biomedical Engineering II*. Cham, Switzerland: Springer, 2024, pp. 105–112, doi: 10.1007/978-3-031-55315-8_12.

[5] J. P. Goldblatt and J. C. Richmond, "Anatomy and biomechanics of the knee," *Operative Techn. Sports Med.*, vol. 11, no. 3, pp. 172–186, Jul. 2003, doi: 10.1053/otsm.2003.35911.

[6] J. Puig-Junoy and A. Ruiz Zamora, "Socio-economic costs of osteoarthritis: A systematic review of cost-of-illness studies," *Seminars Arthritis Rheumatism*, vol. 44, no. 5, pp. 531–541, Apr. 2015, doi: 10.1016/j.semarthrit.2014.10.012.

[7] S. S. Sajjadinia, M. Haghpanahi, and M. Razi, "Computational simulation of the multiphasic degeneration of the bone-cartilage unit during osteoarthritis via indentation and unconfined compression tests," *Proc. Inst. Mech. Engineers, Part H, J. Eng. Med.*, vol. 233, no. 9, pp. 871–882, Sep. 2019, doi: 10.1177/0954411919854011.

[8] V. Klika, E. A. Gaffney, Y.-C. Chen, and C. P. Brown, "An overview of multiphase cartilage mechanical modelling and its role in understanding function and pathology," *J. Mech. Behav. Biomed. Mater.*, vol. 62, pp. 139–157, Sep. 2016, doi: 10.1016/j.jmbbm.2016.04.032.

[9] D. Shriram, G. Yamako, E. Chosa, and K. Subburaj, "Biomechanical evaluation of isotropic and shell-core composite meniscal implants for total meniscus replacement: A nonlinear finite element study," *IEEE Access*, vol. 7, pp. 140084–140101, 2019, doi: 10.1109/ACCESS.2019.2943689.

[10] F. P. Nikolopoulos, E. I. Zacharaki, D. Stanev, and K. Moustakas, "Personalized knee geometry modeling based on multi-atlas segmentation and mesh refinement," *IEEE Access*, vol. 8, pp. 56766–56781, 2020, doi: 10.1109/ACCESS.2020.2982061.

[11] S. S. Sajjadinia and M. Haghpanahi, "A parametric study on the mechanical role of fibrillar rotations in an articular cartilage finite element model," *Scientia Iranica*, vol. 1, no. 3, pp. 1–15, May 2020, doi: 10.24200/sci.2020.51785.2362.

[12] S. Sajjadinia, B. Carpentieri, and G. Holzapfel, "Hybrid data-driven and numerical modeling of articular cartilage," in *Big Data Analysis and Artificial Intelligence for Medical Sciences*. Hoboken, NJ, USA: Wiley, 2024, pp. 181–203, doi: 10.1002/9781119846567.ch9.

[13] M. Antico, F. Sasazawa, Y. Takeda, A. T. Jaiprakash, M.-L. Wille, A. K. Pandey, R. Crawford, G. Carneiro, and D. Fontanarosa, "Bayesian CNN for segmentation uncertainty inference on 4D ultrasound images of the femoral cartilage for guidance in robotic knee arthroscopy," *IEEE Access*, vol. 8, pp. 223961–223975, 2020, doi: 10.1109/ACCESS.2020.3044355.

[14] M. R. Karim, J. Jiao, T. Döhmen, M. Cochez, O. Beyan, D. Rebholz-Schuhmann, and S. Decker, "DeepKneeExplainer: Explainable knee osteoarthritis diagnosis from radiographs and magnetic resonance imaging," *IEEE Access*, vol. 9, pp. 39757–39780, 2021, doi: 10.1109/ACCESS.2021.3062493.

[15] A. Rehman, A. Raza, F. S. Alamri, B. Alghofaily, and T. Saba, "Transfer learning-based smart features engineering for osteoarthritis diagnosis from knee X-ray images," *IEEE Access*, vol. 11, pp. 71326–71338, 2023, doi: 10.1109/ACCESS.2023.3294542.

[16] G. Paiva, S. Bhashyam, G. Thiagarajan, R. Derakhshani, and T. Guess, "A data-driven surrogate model to connect scales between multi-domain biomechanics simulations," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2012, pp. 3077–3080, doi: 10.1109/EMBC.2012.6346614.

[17] V. Arbabi, B. Pouran, G. Campoli, H. Weinans, and A. A. Zadpoor, "Determination of the mechanical and physical properties of cartilage by coupling poroelastic-based finite element models of indentation with artificial neural networks," *J. Biomechanics*, vol. 49, no. 5, pp. 631–637, Mar. 2016, doi: 10.1016/j.jbiomech.2015.12.014.

[18] V. Arbabi, B. Pouran, H. Weinans, and A. A. Zadpoor, "Combined inverse-forward artificial neural networks for fast and accurate estimation of the diffusion coefficients of cartilage based on multi-physics models," *J. Biomechanics*, vol. 49, no. 13, pp. 2799–2805, Sep. 2016, doi: 10.1016/j.jbiomech.2016.06.019.

[19] F. S. Egli, R. C. Straube, A. Mielke, and T. Ricken, "Surrogate modeling of a nonlinear, biphasic model of articular cartilage with artificial neural networks," *PAMM*, vol. 21, no. 1, Dec. 2021, Art. no. e202100188, doi: 10.1002/pamm.202100188.

[20] S. S. Sajjadinia, B. Carpentieri, D. Shriram, and G. A. Holzapfel, "Multi-fidelity surrogate modeling through hybrid machine learning for biomechanical and finite element analysis of soft tissues," *Comput. Biol. Med.*, vol. 148, Sep. 2022, Art. no. 105699, doi: 10.1016/j.compbiomed.2022.105699.

[21] S. Taghizadeh, F. D. Witherden, and S. S. Girimaji, "Turbulence closure modeling with data-driven techniques: Physical compatibility and consistency considerations," *New J. Phys.*, vol. 22, no. 9, Sep. 2020, Art. no. 093023, doi: 10.1088/1367-2630/abadb3.

[22] J. G. Hoffer, A. B. Ofner, F. M. Rohrhofer, M. Lovrić, R. Kern, S. Lindstaedt, and B. C. Geiger, "Theory-inspired machine learning—Towards a synergy between knowledge and data," *Weld. World*, vol. 66, no. 7, pp. 1291–1304, Jul. 2022, doi: 10.1007/s40194-022-01270-z.

[23] H. Gao, L. Sun, and J.-X. Wang, "PhyGeoNet: Physics-informed geometry-adaptive convolutional neural networks for solving parameterized steady-state PDEs on irregular domain," *J. Comput. Phys.*, vol. 428, Mar. 2021, Art. no. 110079, doi: 10.1016/j.jcp.2020.110079.

[24] M. Nourbakhsh, J. Irizarry, and J. Haymaker, "Generalizable surrogate model features to approximate stress in 3D trusses," *Eng. Appl. Artif. Intell.*, vol. 71, pp. 15–27, May 2018, doi: 10.1016/j.engappai.2018.01.006.

[25] P. W. Battaglia et al., "Relational inductive biases, deep learning, and graph networks," 2018, *arXiv:1806.01261*.

[26] A. Sanchez-Gonzalez, J. Godwin, T. Pfaff, R. Ying, J. Leskovec, and P. W. Battaglia, "Learning to simulate complex physics with graph networks," 2020, *arXiv:2002.09405*.

[27] X. Zhang, Y. Xu, W. He, W. Guo, and L. Cui, "A comprehensive review of the oversmoothing in graph neural networks," in *Computer Supported Cooperative Work and Social Computing*, 2024, pp. 451–465.

[28] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar, "Multipole graph neural operator for parametric partial differential equations," 2020, *arXiv:2006.09535*.

[29] M. Fortunato, T. Pfaff, P. Wirnsberger, A. Pritzel, and P. Battaglia, "MultiScale MeshGraphNets," 2022, *arXiv:2210.00612*.

[30] Y. Cao, M. Chai, M. Li, and C. Jiang, "Efficient learning of mesh-based physical simulation with bi-stride multi-scale graph neural network," in *Proc. 40th Int. Conf. Mach. Learn.*, vol. 202, 2023, pp. 3541–3558.

[31] H. Guo, "Nonlinear mixup: Out-of-manifold data augmentation for text classification," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 4, pp. 4044–4051.

[32] A. Jindal, A. Ghosh Chowdhury, A. Didolkar, D. Jin, R. Sawhney, and R. R. Shah, "Augmenting NLP models using latent feature interpolations," in *Proc. 28th Int. Conf. Comput. Linguistics*, Dec. 2020, pp. 6931–6936, doi: 10.18653/v1/2020.coling-main.611.

[33] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "Mixup: Beyond empirical risk minimization," 2017, *arXiv:1710.09412*.

[34] H. Guo, Y. Mao, and R. Zhang, "Mixup as locally linear out-of-manifold regularization," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 3714–3722.

[35] Y. Wang, W. Wang, Y. Liang, Y. Cai, and B. Hooi, "Mixup for node and graph classification," in *Proc. Web Conf.*, Apr. 2021, pp. 3663–3674, doi: 10.1145/3442381.3449796.

[36] X. Han, Z. Jiang, N. Liu, and X. Hu, "G-mixup: Graph data augmentation for graph classification," 2022, *arXiv:2202.07179*.

[37] G. Dong, J. C. Wong, L. Lestandi, J. Mikula, G. Vastola, M. H. Jhon, M. H. Dao, U. Kizhakkinan, C. S. Ford, and D. W. Rosen, "A part-scale, feature-based surrogate model for residual stresses in the laser powder bed fusion process," *J. Mater. Process. Technol.*, vol. 304, Jun. 2022, Art. no. 117541, doi: 10.1016/j.jmatprotec.2022.117541.

[38] A. Mumuni and F. Mumuni, "Data augmentation: A comprehensive survey of modern approaches," *Array*, vol. 16, Dec. 2022, Art. no. 100258, doi: 10.1016/j.array.2022.100258.

[39] H. Lam Cheung, P. Uvdal, and M. Mirkhalaf, "Augmentation of scarce data—A new approach for deep-learning modeling of composites," 2023, *arXiv:2311.14557*.

[40] X. Zhao, Z. Gong, J. Zhang, W. Yao, and X. Chen, "A surrogate model with data augmentation and deep transfer learning for temperature field prediction of heat source layout," *Structural Multidisciplinary Optim.*, vol. 64, no. 4, pp. 2287–2306, Oct. 2021, doi: 10.1007/s00158-021-02983-3.

[41] S. S. Parida, S. Bose, and G. Apostolakis, "Earthquake data augmentation using wavelet transform for training deep learning based surrogate models of nonlinear structures," *Structures*, vol. 55, pp. 638–649, Sep. 2023, doi: 10.1016/j.istruc.2023.05.122.

[42] V. Sella, J. Pham, A. Chaudhuri, and K. E. Willcox, "Projection-based multifidelity linear regression for data-poor applications," in *Proc. AIAA SciTech Forum*, 2023, p. 0916.

[43] M. Rauschenberger and R. Baeza-Yates, "How to handle health-related small imbalanced data in machine learning?" *I-COM*, vol. 19, no. 3, pp. 215–226, Jan. 2021, doi: 10.1515/icom-2020-0018.

[44] S. Bhattacharya, V. Rajan, and H. Shrivastava, "ICU mortality prediction: A classification algorithm for imbalanced datasets," in *Proc. AAAI Conf. Artif. Intell.*, 2017, vol. 31, no. 1, pp. 1–16.

[45] I.-K. Yeo, "A new family of power transformations to improve normality or symmetry," *Biometrika*, vol. 87, no. 4, pp. 954–959, Dec. 2000, doi: 10.1093/biomet/87.4.954.

[46] W.-L. Chiang, X. Liu, S. Si, Y. Li, S. Bengio, and C.-J. Hsieh, "Cluster-GCN: An efficient algorithm for training deep and large graph convolutional networks," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*. New York, NY, USA: Association for Computing Machinery, Jul. 2019, pp. 257–266, doi: 10.1145/3292500.3330925.

[47] H. Zeng, H. Zhou, A. Srivastava, R. Kannan, and V. Prasanna, "Graph-SAINT: Graph sampling based inductive learning method," in *Proc. Int. Conf. Learn. Represent.*, 2020, pp. 1–13.

[48] Y. Wang, Y. Zhao, N. Shah, and T. Derr, "Imbalanced graph classification via graph-of-graph neural networks," in *Proc. 31st ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2022, pp. 2067–2076, doi: 10.1145/3511808.3557356.

[49] S. Shalev-Shwartz and Y. Wexler, "Minimizing the maximal loss: How and why," in *Proc. 33rd Int. Conf. Mach. Learn.*, vol. 48, 2016, pp. 793–801.

[50] M. Steininger, K. Kobs, P. Davidson, A. Krause, and A. Hotho, "Density-based weighting for imbalanced regression," *Mach. Learn.*, vol. 110, no. 8, pp. 2187–2211, Aug. 2021, doi: 10.1007/s10994-021-06023-5.

[51] K. R. M. Fernando and C. P. Tsokos, "Dynamically weighted balanced loss: Class imbalanced learning and confidence calibration of deep neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 7, pp. 2940–2951, Jul. 2022, doi: 10.1109/TNNLS.2020.3047335. https://doi.org/10.1109/TNNLS.2020.3047335

[52] G. Holzapfel, *Nonlinear Solid Mechanics: A Continuum Approach for Engineering*. Hoboken, NJ, USA: Wiley, 2000.

[53] M. Marino, "Constitutive modeling of soft tissues," in *Encyclopedia of Biomedical Engineering*. Amsterdam, The Netherlands: Elsevier, 2019, pp. 81–110.

[54] Y. You, T. Chen, Y. Sui, T. Chen, Z. Wang, and Y. Shen, "Graph contrastive learning with augmentations," 2020, *arXiv:2010.13902*.

[55] F.-Y. Sun, J. Hoffmann, V. Verma, and J. Tang, "InfoGraph: Unsupervised and semi-supervised graph-level representation learning via mutual information maximization," 2019, *arXiv:1908.01000*.

[56] P. Velič ković, W. Fedus, W. L. Hamilton, P. Lia, Y. Bengio, and R. Devon Hjelm, "Deep graph infomax," 2018, *arXiv:1809.10341*.

[57] C. Oh, S. Han, and J. Jeong, "Time-series data augmentation based on interpolation," *Proc. Comput. Sci.*, vol. 175, pp. 64–71, Sep. 2020, doi: 10.1016/j.procs.2020.07.012.

[58] K. Um, R. Brand, Y. Fei, P. Holl, and N. Thuerey, "Solver-in-the-loop: Learning from differentiable physics to interact with iterative PDE-solvers," 2020, *arXiv:2007.00016*.

[59] L. Prechelt, "Early stopping—But when?" in *Neural Networks: Tricks of the Trade*, 2nd ed. Springer, 2012, pp. 53–67.

[60] K. Terzaghi, *Theoretical Soil Mechanics*. Hoboken, NJ, USA: Wiley, 1943.

[61] F. Dullien, *Porous Media: Fluid Transport and Pore Structure*. Amsterdam, The Netherlands: Elsevier, 1979.

[62] V. C. Mow, S. C. Kuei, W. M. Lai, and C. G. Armstrong, "Biphasic creep and stress relaxation of articular cartilage in compression: Theory and experiments," *J. Biomechanical Eng.*, vol. 102, no. 1, pp. 73–84, Feb. 1980, doi: 10.1115/1.3138202.

[63] S. M. Klisch, "Internally constrained mixtures of elastic continua," *Math. Mech. Solids*, vol. 4, no. 4, pp. 481–498, Dec. 1999, doi: 10.1177/108128659900400405.
[64] W. Wilson, C. C. van Donkelaar, B. van Rietbergen, and R. Huiskes, "A fibril-reinforced poroviscoelastic swelling model for articular cartilage," *J. Biomechanics*, vol. 38, no. 6, pp. 1195–1204, Jun. 2005, doi: 10.1016/j.jbiomech.2004.07.003.
[65] M. D. Buschmann and A. J. Grodzinsky, "A molecular model of proteoglycan-associated electrostatic forces in cartilage mechanics," *J. Biomechanical Eng.*, vol. 117, no. 2, pp. 179–192, May 1995, doi: 10.1115/1.2796000.
[66] G. A. Ateshian, N. O. Chahine, I. M. Basalo, and C. T. Hung, "The correspondence between equilibrium biphasic and triphasic material properties in mixture models of articular cartilage," *J. Biomechanics*, vol. 37, no. 3, pp. 391–400, Mar. 2004, doi: 10.1016/S0021-9290(03)00252-5.
[67] W. Wilson, C. C. van Donkelaar, B. van Rietbergen, K. Ito, and R. Huiskes, "Stresses in the local collagen network of articular cartilage: A poroviscoelastic fibril-reinforced finite element study," *J. Biomechanics*, vol. 37, no. 3, pp. 357–366, Mar. 2004, doi: 10.1016/s0021-9290(03)00267-7.
[68] W. Wilson, J. M. Huyghe, and C. C. van Donkelaar, "Depth-dependent compressive equilibrium properties of articular cartilage explained by its composition," *Biomechanics Model. Mechanobiology*, vol. 6, nos. 1–2, pp. 43–53, Jan. 2007, doi: 10.1007/s10237-006-0044-z.
[69] M. E. Stender, C. B. Raub, K. A. Yamauchi, R. Shirazi, P. Vena, R. L. Sah, S. J. Hazelwood, and S. M. Klisch, "Integrating qPLM and biomechanical test data with an anisotropic fiber distribution model and predictions of TGF-β1 and IGF-1 regulation of articular cartilage fiber modulus," *Biomechanics Model. Mechanobiology*, vol. 12, no. 6, pp. 1073–1088, Nov. 2013, doi: 10.1007/s10237-012-0463-y.
[70] S. S. Sajjadinia, B. Carpentieri, and G. A. Holzapfel, "A backward prestressing algorithm for efficient finite element implementation of in vivo material and geometrical parameters into fibril-reinforced mixture models of articular cartilage," *J. Mech. Behav. Biomed. Mater.*, vol. 114, Feb. 2021, Art. no. 104203, doi: 10.1016/j.jmbbm.2020.104203.
[71] Abaqus. (2021). *Dassault Systemes*. [Online]. Available: https://www.simulia.com
[72] X. Wang, T. S. E. Eriksson, T. Ricken, and D. M. Pierce, "On incorporating osmotic prestretch/prestress in image-driven finite element simulations of cartilage," *J. Mech. Behav. Biomed. Mater.*, vol. 86, pp. 409–422, Oct. 2018, doi: 10.1016/j.jmbbm.2018.06.014.
[73] J. C. Simo, "On a fully three-dimensional finite-strain viscoelastic damage model: Formulation and computational aspects," *Comput. Methods Appl. Mech. Eng.*, vol. 60, no. 2, pp. 153–173, Feb. 1987.
[74] B. Fazekas and T. J. Goda, "Determination of the hyper-viscoelastic model parameters of open-cell polymer foams and rubber-like materials with high accuracy," *Mater. Des.*, vol. 156, pp. 596–608, Oct. 2018, doi: 10.1016/j.matdes.2018.07.010.
[75] K. E. Keenan, S. Pal, D. P. Lindsey, T. F. Besier, and G. S. Beaupre, "A viscoelastic constitutive model can accurately represent entire creep indentation tests of human patella cartilage," *J. Appl. Biomechanics*, vol. 29, no. 3, pp. 292–302, Jun. 2013, doi: 10.1123/jab.29.3.292.
[76] A. Erdemir, "Open knee: Open source modeling and simulation in knee biomechanics," *J. Knee Surgery*, vol. 29, no. 2, pp. 107–116, Oct. 2015, doi: 10.1055/s-0035-1564600.
[77] SIMTK. (2023). *Open Knee: A Three-dimensional Finite Element Representation of the Knee Joint*. Accessed: Apr. 23, 2023. [Online]. Available: https://simtk.org/home/openknee
[78] C. De Boor, "A practical guide to splines," *Appl. Math. Sci.*, vol. 78, p. 392, 1978.
[79] M. E. Mononen, A. Paz, M. K. Liukkonen, and M. J. Turunen, "Atlas-based finite element analyses with simpler constitutive models predict personalized progression of knee osteoarthritis: Data from the osteoarthritis initiative," *Sci. Rep.*, vol. 13, no. 1, p. 8888, Jun. 2023, doi: 10.1038/s41598-023-35832-y.
[80] S. Sajjadinia, B. Carpentieri, and G. Holzapfel, "A pointwise evaluation metric to visualize errors in machine learning surrogate models," in *Proc. CECNet*, vol. 345, 2021, pp. 26–34.
[81] T. Pfaff, M. Fortunato, A. Sanchez-Gonzalez, and P. W. Battaglia, "Learning mesh-based simulation with graph networks," 2020, *arXiv:2010.03409*.
[82] S. Rezvani and X. Wang, "A broad review on class imbalance learning techniques," *Appl. Soft Comput.*, vol. 143, Aug. 2023, Art. no. 110415, doi: 10.1016/j.asoc.2023.110415.
[83] R. Novak, Y. Bahri, D. A. Abolafia, J. Pennington, and J. Sohl-Dickstein, "Sensitivity and generalization in neural networks: An empirical study," 2018, *arXiv:1802.08760*.
[84] M. Tan and Q. V. Le, "EfficientNetV2: Smaller models and faster training," 2021, *arXiv:2104.00298*.
[85] Z. Xue, Y. Yang, and R. Marculescu, "SUGAR: Efficient subgraph-level training via resource-aware graph partitioning," *IEEE Trans. Comput.*, vol. 72, no. 11, pp. 1–11, Oct. 2023, doi: 10.1109/TC.2023.3288755.

SEYED SHAYAN SAJJADINIA is currently pursuing the Ph.D. degree in computer science at the Faculty of Engineering, Free University of Bozen-Bolzano, with a focus on biomedical engineering and computational mechanics. He is also a Senior Data Scientist at CAEmate S.R.L., Italy, contributing to the development of the industrial digital twins. His current research interests include integration of data science and numerical modeling, particularly using deep learning and finite element analysis, in computer simulations.

BRUNO CARPENTIERI received the Laurea degree in applied mathematics from Bari University, in 1997, and the Ph.D. degree in computer science, Toulouse, France. He also worked as a Post-Doctoral Researcher at the University of Graz, an Assistant Professor at the University of Groningen, and a Reader at Nottingham Trent University. Since May 2017, he has been an Associate Professor of applied mathematics at the Faculty of Engineering, Free University of Bozen-Bolzano. His main research interests include numerical linear algebra and high-performance computing.

GERHARD A. HOLZAPFEL is currently a Professor and the Head of the Institute of Biomechanics, Graz University of Technology. He is also an Adjunct Professor at the Norwegian University of Science and Technology and a Visiting Professor at the University of Glasgow. He has made significant contributions in both computational and experimental biomechanics, with a focus on soft biological tissues. He has received numerous awards and honors for his work such as the 2021 William Prager Medal and the 2021 Warner T. Koiter Medal, and he was listed in 'The World's Most Influential Scientific Minds: 2014' by Thomson Reuters. He is also the Co-Founder and the Co-Editor-in-Chief of the journal of *Biomechanics and Modeling in Mechanobiology*.

Open Access funding provided by 'Libera Università di Bolzano' within the CRUI CARE Agreement