Machine Learning for Computational Science and Engineering (2026) 2:43
https://doi.org/10.1007/s44379-026-00092-w

RESEARCH

Decoding microstructure-property relationships in polycrystals: an explainable GNN surrogate for crystal plasticity

Ali R. Safi[1] · Pedro Campos[1] · Frederic E. Bock[1] · Benjamin Klusemann[1,2]

Received: 26 June 2026 / Accepted: 24 August 2026
© The Author(s) 2026

Abstract

In this work, a surrogate modeling approach based on graph neural networks (GNNs) is presented to rapidly predict the homogenized stress-strain response of polycrystalline microstructure volume elements (MVEs). Each grain is represented as a node in a graph, with features such as crystallographic orientation, grain size, and aspect ratio, while grain boundaries are encoded as edges. This graph-based representation enables the GNN to capture both local and long-range interactions that govern the macroscopic mechanical response. A synthetic dataset of MVEs was generated using DREAM.3D, covering a wide range of microstructural variations in grain size distributions, morphological anisotropy, and four crystallographic texture classes. Full-field crystal plasticity (CP) simulations performed on these MVEs provided the ground-truth stress-strain data used to train and validate the GNN model. The results demonstrate high correlation between GNN predictions and CP simulations, with strong agreement across different loading conditions and an inference-time speedup of approximately 55,000× relative to full-field CP, achieved up to the reported accuracies and excluding the one-time training and data-generation overhead. The trained surrogate is then subjected to a comprehensive explainability analysis using integrated gradients, providing grain-level attribution maps that identify the microstructural features most influential to the predicted stress response. Population-level attribution statistics, together with cross-validation against GradientShap, trace the dominant attribution to the statistical representativity of each volume element, whose response dispersion scales inversely with the square root of its grain count. Overall, the proposed framework offers an accurate, efficient, and interpretable tool for microstructure-informed mechanical property prediction in heterogeneous polycrystalline materials.

Keywords Graph neural networks · Surrogate modeling · Crystal plasticity · Polycrystals · Explainable artificial intelligence · Integrated gradients

List of abbreviations

ARDF	Aspect-ratio distribution function
CNN	Convolutional neural network
CP	Crystal plasticity
ESD	Equivalent sphere diameter
FFNN	Feed-forward neural network
GCN	Graph convolutional network
GNN	Graph neural network
GSDF	Grain size distribution function
IG	Integrated Gradients
MSE	Mean squared error
MVE	Microstructure volume element
MPL	Message passing layer
ODF	Orientation distribution function
PDF	Probability density function
ReLU	Rectified linear unit
RMSE	Root-mean-squared error
SD	Standard deviation
SHAP	SHapley Additive exPlanations
ML	Machine learning

✉ Ali R. Safi
ali.safi@hereon.de

[1] Helmholtz-Zentrum Hereon, Institute of Material and Process Design, Solid State Materials Processing, Max-Planck-Straße 1, 21502 Geesthacht, Germany

[2] Leuphana University Lüneburg, Institute for Production Technology and Systems, Universitätsallee 1, 21335 Lüneburg, Germany

1 Introduction

Macroscopic mechanical behavior in polycrystalline materials depends on the collective local interactions of their con-

Published online: 08 September 2026

stituents, where individual grains deform heterogeneously according to their crystallographic orientation, size, morphology, and neighborhood. The full-field Crystal Plasticity (CP) framework is a well-established computational technique that accounts for this heterogeneity by incorporating deformation mechanisms at the intragrain scale into a continuum constitutive description. The governing constitutive models, when solved over spatially resolved representative volume elements using finite elements or spectral solvers, can faithfully reproduce the anisotropic stress-strain response of polycrystalline materials [1]. Despite the rapid development of efficient numerical implementations [2] and the steady increase in available computational resources, full-field CP simulations remain computationally intensive and time-consuming, making them unfeasible to systematically explore large sets of microstructures for the purpose of establishing quantitative (micro-)structure-property linkages.

Consequently, significant efforts have been directed towards making CP frameworks computationally tractable for practical engineering applications. An important and actively developed route is provided by mean field homogenization based crystal plasticity models. Modern self consistent and viscoplastic self consistent formulations account for grain anisotropy, crystallographic texture, and grain interaction effects at a computational cost substantially below that of full field simulations [3, 4]. Such approaches have been successfully applied to predict effective mechanical properties, macroscopic stress strain responses, texture evolution, and anisotropy in polycrystalline materials. Their computational efficiency makes them particularly attractive when the macroscopic response of statistically representative microstructures is the primary quantity of interest. The present work pursues a complementary route by constructing a data driven surrogate of full field CP simulations. Rather than replacing the grain ensemble by a mean field description, the microstructure is retained as an explicit grain resolved representation that includes grain descriptors and neighborhood connectivity. Beyond mean field homogenization approaches, efforts to reduce the cost of full field CP simulations can be broadly grouped into model order reduction techniques [5, 6] and Machine Learning (ML) methods [7–11]. While model order reduction approaches reduce the dimensionality of the problem, ML models can be trained to map microstructures to solutions with high accuracy and precision while reducing the required execution time by orders of magnitude. ML techniques have thus been extensively studied for constructing data driven constitutive models, as reviewed by Fuhg et al. [12]. In terms of microstructure representation, most studies rely either on image based inputs [13–15], where convolutional architectures extract spatial features directly from voxelized microstructure fields, or on compact statistical descriptors [13, 16]. While image based methods can capture local field heterogeneity with high fidelity, both approaches are limited by the large amounts of training data and training time [17, 18].

While the above approaches generally predict a homogenized or local response for a prescribed microstructure, image-based methods such as convolutional neural networks (CNNs) have also been explored as surrogates. Ibragimova et al. [19] demonstrated that CNNs can reproduce local stress and strain distributions in aluminum alloy microstructures with varying grain size, morphology and texture, with excellent agreement to full-field CP simulations, and that the predictions carry over to measured electron backscatter diffraction microstructures that were not represented in the training data. Despite their success, voxelized representations impose a resolution-dependent scaling of the problem size with the number of pixels or voxels rather than with the number of grains, which significantly limits scalability for large or complex microstructures.

In contrast, graph neural networks (GNNs) have by now become a well-established approach within the ML community, distinguished from conventional methods by their ability to operate directly on graph-structured data composed of nodes and edges. Their versatility has led to widespread applications across numerous domains, as comprehensively reviewed by Zhou et al. [20]. GNNs have been successfully applied in fields such as natural language processing [21] and molecular property prediction [22]. In this context, a polycrystalline structure can also be characterized using a graph description where each individual grain is represented by a node and adjacent grains are connected by edges signifying a common grain boundary. Morphological and crystallographic grain properties are stored in the feature vector of each node. By using this microstructural graph representation, it is possible to overcome resolution-dependent limitations as the problem scales with the number of grains in the polycrystal rather than the number of pixels/voxels used to depict it. The graph representation is further advantageous in terms of memory efficiency, as shown by Chung et al. [23], who demonstrated that GNN-based approaches for predicting elastic moduli of digital rock images require significantly less GPU memory than CNN-based counterparts while maintaining superior predictive accuracy on unseen rock samples. Moreover, physically meaningful graph representations of the constitutive response have been explored in geometric machine learning frameworks for encoding time-history data of plastic deformation into low-dimensional Euclidean spaces by Vlassis et al. [24].

Building on this representation, recent studies have shown that grain-level GNN surrogates can be employed in the context of full-field CP models while running orders of magnitude faster. Pagan et al. [25] demonstrated that a two-layer Gaussian-mixture GNN can predict the grain-resolved elastic stresses of both synthetical and experimentally reconstructed microstructures within roughly 5% error. Hu and Latypov

[26] expanded the concept with AnisoGNN, introducing rotational data augmentation so that a single-layer model could capture arbitrary loading directions and reduce test errors by a factor of five to ten compared with orientation-based baselines. Hestroffer et al. [27] showed that minimalistic graphs containing only grain orientation and size still achieve sub-percent errors for stiffness and yield strength while requiring roughly five times less training data than comparable voxel-based CNN surrogates. Temporal behavior has further been addressed by Hu et al. [28] that coupled a graph-convolution layer with linearized minimal-state cells [9, 29] to evolve hidden grain states through strain increments, reproducing full-field crystal-plasticity responses more than a thousand times faster than CP simulations. Extending to fatigue, Sim et al.'s fatigue indicator parameter GNN [30] predicts grain-level Fatemi-Socie indicators with less than 10% error on microstructures significantly larger than those used for training, enabling reliable ranking of microstructure variants for fatigue life assessment. Qin et al. [31] introduced GrainGNN, a dynamic graph surrogate for three-dimensional grain evolution in laser powder-bed fusion that couples transformer long short-term memory encoders with a reduced microstructural graph to reproduce grain statistics at up to 2000-fold speed-up over phase-field calculations. These advances collectively illustrate how microstructural graph representations enable scalable, physics-aware surrogate models for elasticity, plasticity, fatigue as well as time-dependent microstructural evolution. Nevertheless, a systematic explainability framework for grain-resolved GNN surrogates that validates learned structure-property relationships against physical principles at population scale has yet to be established, representing the key gap this study addresses.

In this study, a GNN model is developed to predict the homogenized stress-strain response of Microstructure Volume Elements (MVEs) under mechanical loading. A large synthetic dataset is generated by systematically varying morphological and crystallographic descriptors across four crystallographic texture classes, thereby capturing a wide range of microstructural heterogeneity that causes anisotropic mechanical behavior. This dataset serves as the ground truth for training the GNN surrogate model. The trained model is then subjected to a comprehensive explainability analysis using Integrated Gradients (IGs), providing grain-level attribution maps that quantify the influence of individual microstructural features on the predicted stress response. The analysis spans population-level attribution statistics across large ensembles of MVEs, the evolution of feature importance with applied strain, and cross-validation against GradientShap, collectively establishing physically interpretable structure-property relationships from the GNN's learned representations.

2 Methodology

2.1 Generation of microstructure volume elements

The scarcity of experimental microstructure data requires the utilization of synthetical approaches to construct 3D volume elements of polycrystalline materials that are morphologically and crystallographically realistic. A variety of different microstructure descriptors exist which allow for a microstructure characterization and reconstruction [32]. While high-dimensional microstructure descriptors, such as n-point spatial correlations [33] of symmetrized generalized spherical harmonics, can simultaneously characterize both morphological and crystallographical information at once, they suffer from being numerically costly and difficult to interpret. In contrast, probability density functions (PDF) based on low-dimensional descriptors offer an interpretable set of parameters which can be synthetically defined in a straightforward manner. Examples of low-dimensional descriptors include the grain size distribution function (GSDF), grain aspect ratio distribution function (ARDF), and the orientation distribution function (ODF), which are subject of this study. In this work, DREAM.3D [34] is used to synthetically generate a microstructural dataset using diverse sets of GSDFs, ARDFs and ODFs. The grain morphology and crystallographic state of each grain are described by three statistical distribution functions. Their mathematical definitions are provided in Appendix A.

The grain size is quantified by the equivalent sphere diameter (ESD), d_{ESD}, and its population is described by a lognormal GSDF parametrized by the logarithmic mean μ and its standard deviation (SD) σ. Grain morphological anisotropy is captured through the minimum and maximum aspect ratios $a_{C/A}$ and $a_{B/A}$ of an equivalent ellipsoid with principal semi-axes $A \geq B \geq C$, each distributed according to a Beta-type ARDF. The major axis A of every ellipsoidal grain is aligned with the global Y-axis. Crystallographic texture is represented by the ODF, $f_{\mathrm{ODF}}(\mathbf{g})$, where the grain orientation $\mathbf{g}$ is expressed using Bunge ZXZ Euler angles $(\varphi_1, \phi, \varphi_2)$. In this work, the generated MVEs cover random textures as well as four commonly occurring crystallographic texture types, as shown in Table 1.

Table 1 Crystallographic definition of common texture components in terms of Miller indices and Bunge-Euler angles $(\varphi_1, \phi, \varphi_2)$

Notation	Miller's indices		Euler angle (°)		
	$\{hkl\}$	$\langle uvw \rangle$	φ_1	ϕ	φ_2
Brass	{110}	⟨112⟩	35	45	0
Copper	{112}	⟨111⟩	90	35	45
Cube	{100}	⟨100⟩	0	0	0
Goss	{110}	⟨001⟩	0	45	0

By sampling from prescribed morphological and crystallographical descriptors, DREAM.3D first performs Voronoi tessellations of ellipsoidal structures to fill the volume element with grains leaving no void. To optimize the ODF, a Monte-Carlo type algorithm is employed that may either swap orientations between two grains or introduce new orientations in order to iteratively converge to the ODF of interest. The synthetic MVE generation is shown in Fig. 1.

A total number of 1,000 synthetic polycrystalline MVEs are generated using DREAM.3D, ensuring that the range of grain size distributions and anisotropies cover a wide range of statistical diversity. The generated MVEs can be categorized into three classes according to the descriptors that are varied.

- **Type 1**, 200 microstructures, step-wise variation of μ from 1.0 to 2.0, random texture, equiaxed grains.
- **Type 2**, 200 microstructures, random selection of mean aspect ratios $\overline{a}_{B/A}$ and $\overline{a}_{C/A}$ between 0.1 and 1.0, constant $\mu = 1.5$, random texture.
- **Type 3**, 600 microstructures, random selection of parameters μ, $\overline{a}_{B/A}$ and $\overline{a}_{C/A}$ in the ranges of Type 1 and Type 2, random selection of texture according to Table 1.

For the Type 1 dataset the log-mean μ is varied while log-standard deviation σ is kept fixed. This strategy shifts the mean grain size across the desired range without altering the shape of the log-normal distribution. The distribution width, which is controlled by σ, therefore remains constant. Varying σ instead would alter the width of the distribution rather than its position, which addresses a different design question and is not pursued here. Because each MVE occupies a fixed domain of $16 \times 16 \times 16\ \mu\text{m}^3$, changing μ inevitably changes the number of grains that can be packed into the volume. As μ increases, grains become larger on average and the grain count decreases, ranging from 10 to 271 grains per MVE across the Type 1 set. Over the full dataset, which additionally contains the morphologically anisotropic Type 2 and the randomized Type 3 sets, the grain count spans 7 to 282 per MVE.

2.2 Crystal plasticity and constitutive modeling

In the following, brief introductions to the CP framework and the constitutive model are given. The total deformation

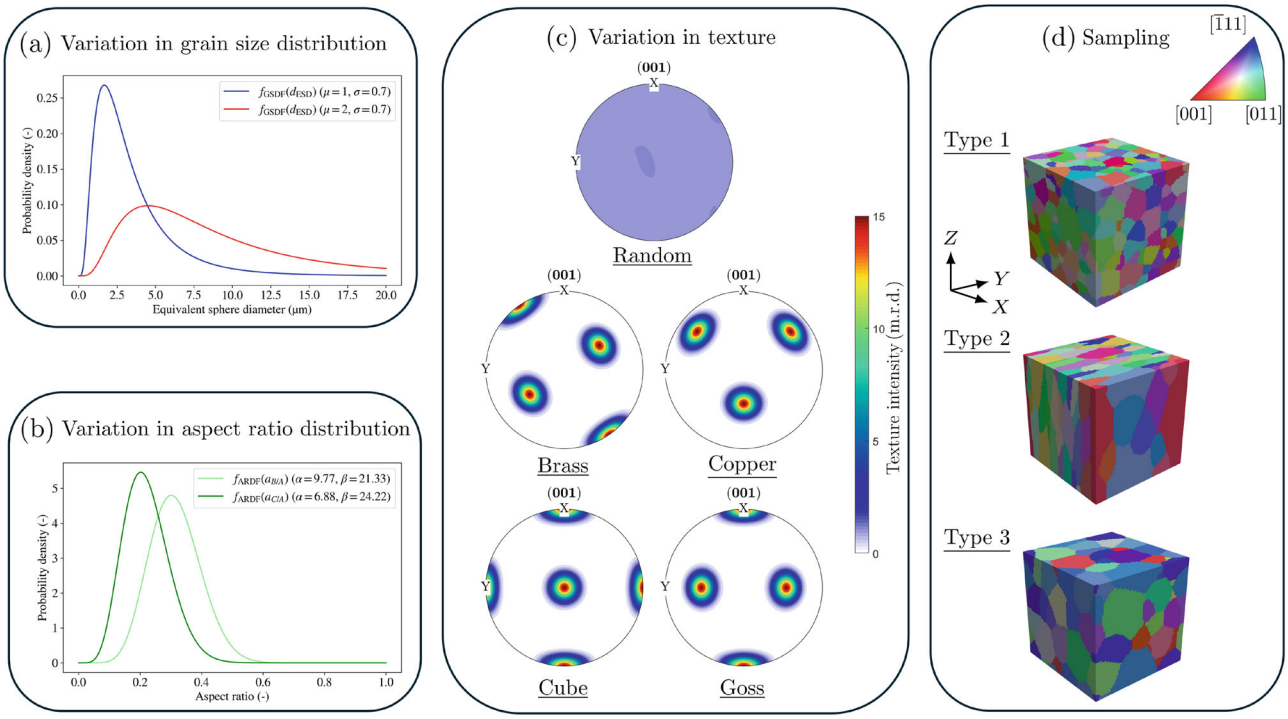

Fig. 1 Illustration of the synthetic microstructure generation process. (**a**) Variation in grain size distribution, $f_{\text{GSDF}}(d_{\text{ESD}})$, shown for two different mean values μ while maintaining a constant SD of $\sigma = 0.7$. (**b**) Variation in aspect ratio distribution, $f_{\text{ARDF}}(a)$, depicted for both the minimum ($a_{C/A}$) and the maximum aspect ratio ($a_{B/A}$). Each line corresponds to a Beta distribution characterized by shape parameters α and β, enabling diverse morphologies from equiaxed to highly elongated grains. (**c**) Variation in crystallographic texture, where the ODFs of four commonly appearing texture components (Brass, Copper, Cube, and Goss) are displayed as pole figures in addition to a random texture reference. (**d**) Representative MVEs of three microstructure classes defined for the synthetic generation, where each type emphasized different variations in the descriptors, namely Type 1 varying the GSDF, Type 2 varying the ARDF, and Type 3 combining variation of GSDF and ARDF while the ODF is randomly selected from the four texture components

gradient, $\mathbf{F}$, can be multiplicatively decomposed into elastic and plastic contributions [35]:

$$\mathbf{F} = \mathbf{F}_e \mathbf{F}_p, \tag{1}$$

where $\mathbf{F}_e$ represents the recoverable elastic deformation gradient and $\mathbf{F}_p$ accounts for the irreversible plastic deformation. The rate of change of the plastic deformation gradient, $\dot{\mathbf{F}}_p$, is further defined as:

$$\dot{\mathbf{F}}_p = \mathbf{L}_p \mathbf{F}_p, \tag{2}$$

where $\mathbf{L}_p$ denotes the plastic velocity gradient. Due to the separation of scales, the definition of $\mathbf{L}_p$ allows for the mean-field incorporation of low-scale deformation mechanisms into the meso-scale. Considering plastic slip as the only active deformation mechanism, $\mathbf{L}_p$ can be determined by accounting for each active slip system, indexed by α, and the associated shear rate $\dot{\gamma}^\alpha$ which is expressed as:

$$\mathbf{L}_p = \sum_{\alpha=1}^{N_{\text{slip}}} \dot{\gamma}^\alpha \left[\mathbf{m}^\alpha \otimes \mathbf{n}^\alpha\right], \tag{3}$$

where the Schmid tensor is formed by the slip direction $\mathbf{m}^\alpha$ and the slip plane normal $\mathbf{n}^\alpha$. A commonly employed phenomenological law for the shear rate is the power-law relation [36, 37]:

$$\dot{\gamma}^\alpha = \dot{\gamma}_0 \left| \frac{\tau^\alpha}{\tau_c^\alpha} \right|^{\frac{1}{m}} \operatorname{sgn}(\tau^\alpha), \tag{4}$$

where $\dot{\gamma}_0$ is the reference shear rate, τ^α is the resolved shear stress on slip system α, and τ_c^α represents the slip resistance/critical resolved shear stress. The parameter $\frac{1}{m}$ is the rate sensitivity exponent, dictating how sensitively the slip rate responds to the applied stress. In parallel with the evolution of the slip rate, the internal state variable τ_c^α evolves according to a hardening law that captures strain hardening, which is described as:

$$\dot{\tau}_c^\alpha = \sum_{\beta=1}^{N_{\text{slip}}} q^{\alpha\beta} \left[h_0 \left[1 - \frac{\tau_c^\beta}{\tau_\infty} \right]^a \right] |\dot{\gamma}^\beta|, \tag{5}$$

where h_0 is the hardening modulus, τ_∞ is the saturation stress, a is a slip hardening parameter, and the coefficients $q^{\alpha\beta}$ represent the interactions between slip system α and β. The Green-Lagrange strain tensor $\mathbf{E}$ is defined as:

$$\mathbf{E} = \frac{1}{2}(\mathbf{F}^{\mathrm{T}}\mathbf{F} - \mathbf{I}), \tag{6}$$

where $\mathbf{I}$ denotes the identity matrix and $\mathbf{F}^{\mathrm{T}}$ is the transpose of $\mathbf{F}$. The deformation gradient can be projected into the stretch and rotation components by polar decomposition:

$$\mathbf{F} = \mathbf{V}\mathbf{R}, \tag{7}$$

where $\mathbf{R}$ is a rotation matrix and $\mathbf{V}$ is the left stretch tensor, out of which it is possible to calculate the logarithmic Hencky-strain tensor, $\boldsymbol{\varepsilon}$, as follows:

$$\boldsymbol{\varepsilon} = \ln(\mathbf{V}). \tag{8}$$

$\boldsymbol{\varepsilon}$ is also known as the true strain tensor. The left polar decomposition is used here since the Hencky strain defined via $\mathbf{V}$ is consistent with the spatial (Eulerian) description adopted for the Cauchy stress $\boldsymbol{\sigma}$. Further, the elastic response is set to follow Hooke's law:

$$\mathbf{S} = \mathcal{C}\,\mathbf{E}_e, \tag{9}$$

with $\mathcal{C}$ being the fourth-order elasticity tensor, $\mathbf{E}_e$ denotes the elastic part of the Green-Lagrange strain tensor, and $\mathbf{S}$ is the second Piola-Kirchhoff stress. The Cauchy stress $\boldsymbol{\sigma}$ can be obtained through the transformation of $\mathbf{S}$:

$$\boldsymbol{\sigma} = \frac{1}{\det(\mathbf{F})}\mathbf{F}\mathbf{S}\mathbf{F}^{\mathrm{T}}. \tag{10}$$

The constitutive parameters for pure aluminum are used in this work and are listed in Table 2.

Each MVE is subjected to CP simulations using the DAMASK [35] fast-Fourier transform solver under three distinct load cases, as shown in Table 3, that are controlled by the rate of the deformation gradient. The simulated load cases include uniaxial tension in X and Z directions and a simple shear loading condition in the XZ plane. Each load case is imposed with an identical two-stage schedule at a constant deformation-gradient rate of $\dot{F} = 10^{-3}\ \mathrm{s}^{-1}$, applied first for 10 s and then continued for a further 60 s, with an output

Table 2 Crystal plasticity constitutive model parameters for pure aluminum

Category	Parameter	Reference
C_{11}, C_{12}, C_{44}	95.22, 57.74, 23.93 GPa	[38]
N_{slip}	12	[39]
a	3.1	[40]
$\dot{\gamma}_0$	$7.5 \times 10^{-5}\ \mathrm{s}^{-1}$	[39]
h_0	170 MPa	[39]
$q^{\alpha\beta}$	1.0 (self/coplanar), 1.4 (non-coplanar)	[40]
m	0.05	[39]
$\tau_{c,0}$	5.0 MPa	[39]
τ_∞	37.5 MPa	[39]

Table 3 Imposed boundary conditions for CP simulations. Components marked with x represent unconstrained degrees of freedom solved for satisfying equilibrium. $\mathbf{P}$ denotes the first Piola-Kirchhoff stress tensor. A zero component indicates a traction-free boundary condition

Case	$\dot{F}$ (s^{-1})	P (MPa)
Uniaxial tension in X	$\begin{pmatrix} 0.001 & 0 & 0 \\ 0 & x & 0 \\ 0 & 0 & x \end{pmatrix}$	$\begin{pmatrix} x & x & x \\ x & 0 & x \\ x & x & 0 \end{pmatrix}$
Uniaxial tension in Z	$\begin{pmatrix} x & 0 & 0 \\ 0 & x & 0 \\ 0 & 0 & 0.001 \end{pmatrix}$	$\begin{pmatrix} 0 & x & x \\ x & 0 & x \\ x & x & x \end{pmatrix}$
Simple shear in XZ	$\begin{pmatrix} 0 & 0 & 0.001 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} x & x & x \\ x & x & x \\ x & x & x \end{pmatrix}$

state written every second increment. This yields a finer strain sampling ($\Delta\varepsilon \approx 5 \times 10^{-4}$) through the elastic-to-yield transition up to approximately 1% strain and a coarser sampling ($\Delta\varepsilon \approx 2 \times 10^{-3}$) into the plastic regime up to approximately 7% strain, giving about 52 output states per simulation. Since the strain tensor subsequently enters the GNN as a continuous input feature, the surrogate interpolates in strain rather than being tied to this sampling grid. The results of the simulations provide the ground-truth stress-strain data in terms of the Cauchy stress and the logarithmic Hencky strain for each MVE, which are subsequently used as the target outputs in the supervised training of the GNN model. The bulk response of each MVE is obtained by homogenizing the local field quantities via volume averaging. For simplicity, throughout the remaining manuscript, the term "stress" and "strain" refer to the homogenized variants of the Cauchy stress and the logarithmic Hencky strain, respectively.

2.3 Graph representation of polycrystalline microstructure volume elements

To enable surrogate models to effectively represent the underlying structure-property relationships in polycrystalline materials, it is required to recast the embedded microstructural information into a latent space representation that captures the key features governing the micromechanical response under applied loads. While the voxelized representations of MVEs facilitate the direct application of CNNs to extract spatial features, such approaches encounter significant scalability issues when processing large-scale images. An alternative strategy is to represent the MVE as an unstructured graph, thereby enabling the use of GNNs. This approach facilitates an efficient latent-space representation of the grain topology and allows for physically meaningful, grain-level model interpretations via IGs.

The dataset for the GNN comprises the microstructure graphs together with the labeled data as calculated from the CP simulations. The microstructure graphs are constructed by mapping individual grains to nodes and connecting nodes if the corresponding grains have a common grain boundary. Node features characterized by crystallographic orientations are described in terms of Euler angles ($\varphi_1, \phi, \varphi_2$), the grain size in terms of the ESD d_{ESD}, the minimum $a_{C/A}$ and maximum $a_{B/A}$ aspect ratios, and the number of neighboring grains N_{nb}. The dataset is split among the MVEs. All strain steps belonging to a given MVE under a given loading condition are kept within the same partition, ensuring no MVE appears in both training and evaluation. This yields 700, 150, and 150 MVEs for training, validation, and testing (70/15/15%) with three simulations each. The test metrics reported below therefore characterize the accuracy attained for unseen strain states of microstructures that are represented in training. The accuracy attained for entirely unseen microstructures classes is a separate question, quantified by the out-of-distribution evaluation reported in Section 3.3.

2.3.1 Graph neural network

In this graph representation, each grain or crystallite within the polycrystal is modeled as a node, and the inherent microstructural heterogeneity is encoded in the node features. Formally, the microstructure graph, $\mathcal{G}$, is defined as:

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X}), \tag{11}$$

where $\mathcal{V} = \{1, 2, \ldots, N_G\}$ is the set of nodes, each representing a grain in the microstructure. Here, N_G is the total number of grains. $\mathcal{E}$ is the set of edges that connect two nodes/grains. An edge $(i, j) \in \mathcal{E}$ indicates that grains i and j share a grain boundary and are considered neighbors. Grain boundaries and their characteristics can thus be encoded edges features. $\mathbf{X} \in \mathbb{R}^{N_G \times N_f}$ is the node feature matrix with N_G nodes and N_f features. Each row $\mathbf{x}_i \in \mathbb{R}^{N_f}$ corresponds to the feature vector of grain i. Given a graph, $\mathcal{G}$, it is further possible to define the connectivity in terms of the adjacency matrix, $\mathbf{A}$. Given that $\mathcal{G}$ is an undirected graph (i.e. the edges $(i, j) = (j, i)$), its elements can be defined as:

$$A_{ij} = \begin{cases} 1 & \text{if } (i, j) \in \mathcal{E} \\ 0 & \text{if } (i, j) \notin \mathcal{E} \\ 0 & \text{if } i = j \end{cases}. \tag{12}$$

To aggregate information from multiple neighborhoods in the graph, a graph convolution network (GCN) is constructed that employs several message passing layers (MPLs). In this context, the graph is input into successive layers that iteratively update the feature representation of each node based on its local neighborhood as shown in Fig. 2. The graph convolution operation is formally described as:

$$\mathbf{H}^{(l)} = \mathrm{g}_{\mathrm{GCN}}\left(\tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{H}^{(l-1)}\mathbf{W}^{(l)}_{\mathrm{GCN}}\right), \tag{13}$$

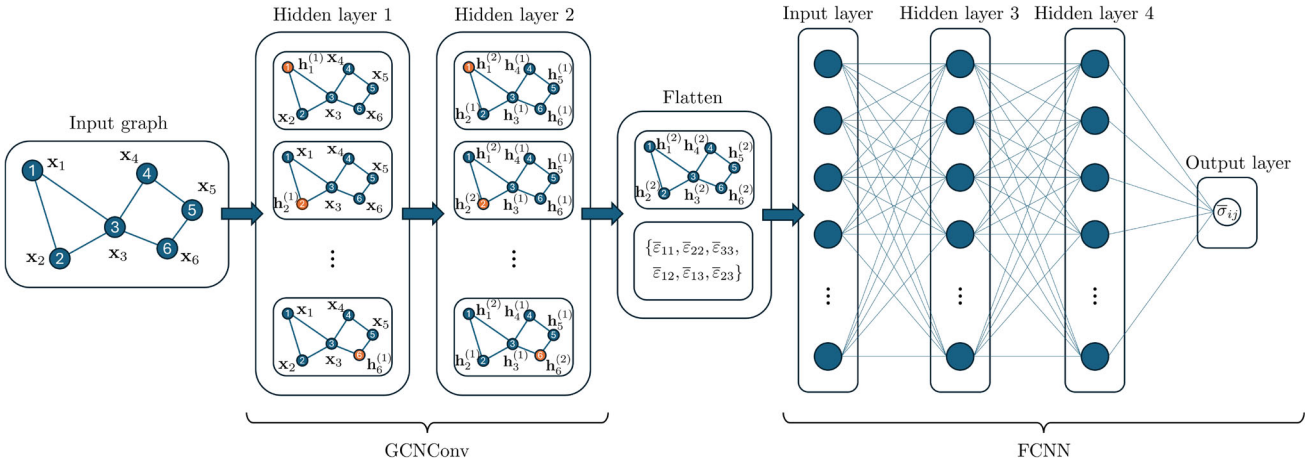

Fig. 2 Schematic illustration of the graph-based surrogate modeling approach. Each grain in the polycrystal is represented as a node (e.g., nodes 1-6), with edges denoting grain boundaries. Node features (e.g., crystallographic orientation, grain size, aspect ratio) are iteratively updated via message passing layers (center panels), where each layer refines the node embeddings based on information from neighboring nodes. The final node embeddings are then flattened and concatenated with the imposed strain components before passing through a feedforward neural network (right). The network's output is the homogenized stress tensor component $\overline{\sigma}_{ij}$

where $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$. $\tilde{\mathbf{D}}$ is a diagonal degree matrix whose i-th diagonal entry is the scalar $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$, with all off-diagonal entries being zero, i.e. $\tilde{D}_{ij} = 0$ for $i \neq j$. $\mathbf{H}^{(l)}$ is the updated embedding matrix after the l-th message passing layer with $\mathbf{H}^{(0)} = \mathbf{X}$. $\mathbf{W}^{(l)}_{\text{GCN}} \in \mathbb{R}^{N_{l-1} \times N_l}$ is a learnable weight matrix with $N_0 = N_f$ and $\mathrm{g}_{\text{GCN}}(\cdot)$ is a nonlinear activation function employed in the GCN. Normalization by the degrees of the nodes is ensured by the symmetric normalization of the modified adjacency matrix, thereby avoiding bias toward high-degree nodes. By stacking MPLs, each node embedding becomes progressively more aware of its neighborhoods through the pooling operations, allowing the final embeddings to capture long-range interactions and grain spatial correlations in the microstructure. After L layers, the final embedding matrix reads:

$$\mathbf{H}^{(L)} = \{\mathbf{h}_1^{(L)}, \mathbf{h}_2^{(L)}, \ldots, \mathbf{h}_{N_G}^{(L)}\}, \tag{14}$$

where $\mathbf{h}_i^{(L)}$ denotes the final feature vector of node i. The embedding matrix is flattened to form a vector, $\mathbf{h}^*$, that captures the microstructural state of the entire polycrystal:

$$\mathbf{h}^* = \mathrm{vec}\left(\mathbf{H}^{(L)}\right) \in \mathbb{R}^{N_G d_L}, \tag{15}$$

where d_L is the dimensionality of the embedding at layer L and $\mathrm{vec}(\cdot)$ denotes the vectorization of a second-order tensor to a first-order tensor. While graph classification problems can be solved directly by using the GCN, it is necessary to include a downstream feedforward neural network (FFNN) to solve the global graph regression problem that are subjected to external forces. In order to account for the imposed global logarithmic Hencky strain $\boldsymbol{\varepsilon}$, we augment the six independent components of the strain tensor by concatenation. The combined feature vector, $\mathbf{x}_{\text{FFNN}}$, becomes:

$$\mathbf{x}_{\text{FFNN}} = \left[\mathbf{h}^*\,;\, \boldsymbol{\varepsilon}\right] \in \mathbb{R}^{N_G d_L + d_\varepsilon}, \tag{16}$$

where $d_\varepsilon = 6$ is the dimensionality of the strain input. This combined vector $\mathbf{x}_{\text{FFNN}}$ defines the input layer of the FFNN. Let the FFNN be composed of K hidden layers, where the operation of the k-th layer is defined as:

$$\mathbf{z}^{(k)} = \mathrm{g}_{\text{FFNN}}\left(\mathbf{W}^{(k)}_{\text{FFNN}}\mathbf{z}^{(k-1)} + \mathbf{b}^{(k)}\right), \quad k = 1, 2, \ldots, K, \tag{17}$$

with $\mathbf{z}^{(0)} = \mathbf{x}_{\text{FFNN}}$. Here, $\mathbf{W}^{(k)}_{\text{FFNN}}$ and $\mathbf{b}^{(k)}$ are the learnable weight matrix and bias vector at the k-th layer, respectively, and $\mathrm{g}_{\text{FFNN}}(\cdot)$ denotes the nonlinear activation function in the FFNN. The final layer of the FFNN is designed to output the prediction of one independent component of the homogenized Cauchy stress $\overline{\boldsymbol{\sigma}}_{ij}$ which can be summarized as:

$$\overline{\sigma}_{ij} = \mathbf{W}^{(K)}_{\text{FFNN}}\mathbf{z}^{(K-1)} + \mathbf{b}^{(K)}. \tag{18}$$

The final prediction task is summarized as:

$$\overline{\sigma}_{ij} = \mathrm{f}_{\text{GNN}}(\mathcal{G}, \overline{\boldsymbol{\varepsilon}}), \tag{19}$$

where $\mathrm{f}_{\text{GNN}}(\cdot)$ denotes the proposed GNN operator that is further schematically illustrated in Fig. 2. In total, six separate GNNs are constructed individually to predict the independent components of the homogenized 3D stress tensor. Training separate models for each component allows the hyperparameter search to adapt to the distinct numerical characteristics

of each stress state. This design comes at the cost of increased total training time but ensures each model is optimally tuned for its target quantity. The six components are predicted independently, and no constraint is imposed to couple them or to enforce a physically admissible stress state. Two conditions nevertheless hold by construction, namely symmetry of the predicted stress tensor, since only the six independent components are modeled, and invariance to grain relabelling, which follows from the permutation-invariant aggregation of the graph convolution. The equilibrium condition $\operatorname{div}\boldsymbol{\sigma} = \mathbf{0}$ is a field relation on the microscopic stress and does not apply to the volume-averaged quantity predicted here, which carries no spatial argument to constrain. It is not thereby irrelevant but located upstream, as equilibrium holds in the CP solutions supplying the training targets, and for a divergence-free field the average stress theorem identifies the volume average with the first moment of the boundary tractions, so the surrogate inherits it through the data rather than through a constraint on its output.

What remains unconstrained is the relationship between the six predicted components. A possible extension could involve a potential based parametrization providing thermodynamic consistency. A systematic assessment of how such a constraint affect model behavior is beyond the scope of the present work. The surrogate further defines a quasi static map from the current strain state to the stress, $\overline{\sigma}_{ij} = \mathrm{f}_{\mathrm{GNN}}(\mathcal{G}, \overline{\boldsymbol{\varepsilon}})$, without an explicit internal state variable. For the monotonic, proportional loading paths considered here, the current strain uniquely parametrizes the response. Genuine history dependent or path dependent behavior is not represented by the present framework and lies outside the scope of this work.

The architecture of the GNN comprises two MPLs that aggregate local neighborhood information through graph convolution operations. For these MPLs, the Rectified Linear Unit (ReLU) activation function is employed while the FFNN employs sigmoid activation functions in the hidden layers and a linear activation function on the output layer. The Adam optimizer [41] is chosen while a batch size of 32 is maintained. The graph node features are min-max normalized with respect to the training dataset. The strain inputs are scaled to $[-1, 1]$, while the stress outputs are scaled to $[1, 5]$. The strictly positive output range of $[1, 5]$ was chosen to keep all normalized targets away from zero, improving gradient conditioning during MSE minimization. The GNN is trained to minimize the mean-squared error (MSE) as the loss function. The root mean-squared error (RMSE) as well as the coefficient of determination R^2 are used as further evaluation metric which are defined as follows:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left[\overline{\sigma}_{ij}^{\mathrm{GNN}} - \overline{\sigma}_{ij}^{\mathrm{CP}}\right]^2, \tag{20}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\overline{\sigma}_{ij}^{\mathrm{GNN}} - \overline{\sigma}_{ij}^{\mathrm{CP}}\right]^2}, \tag{21}$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{n}\left[\overline{\sigma}_{ij}^{\mathrm{CP}} - \overline{\sigma}_{ij}^{\mathrm{GNN}}\right]^2}{\sum_{i=1}^{n}\left[\overline{\sigma}_{ij}^{\mathrm{CP}} - \hat{\overline{\sigma}}_{ij}^{\mathrm{CP}}\right]^2}. \tag{22}$$

A Bayesian hyperparameter optimization was conducted independently for all stress components using Optuna [42] to fine-tune parameters, including the number of neurons in each layer and the learning rate. The resulting optimal hyperparameters for each stress component are summarized in Table 4.

2.3.2 Attribution methods for model explainability

While the GNN can provide accurate predictions for the homogenized stress-strain response, it is often perceived as a black-box model. To obtain physical insights into the learned structure-property relationships and to validate the model's behavior against established mechanical principles, it is essential to employ interpretability techniques in the context of the explainable artificial intelligence paradigm [43]. For this purpose, IGs [44] are utilized as a powerful attribution method that explains model predictions by assigning an importance score to each input feature.

The IG method calculates the integral of the model's output gradients with respect to the input features along a linear path from a baseline (or reference) input to the actual input. This approach satisfies desirable axioms for attribution methods, namely *Sensitivity*, i.e. if a change in a single feature leads to a different output then this feature must have a non-zero attribution, and *implementation invariance*, i.e. if two different functions yield the same output then the feature attribution must be identical. The attribution of a prediction

Table 4 Optimized model parameters, in terms of the optimal number of nodes per hidden layer and the learning rate, for the six individually trained GNNs for each independent stress component $\overline{\sigma}_{ij}$

Hyperparameter	Possible values	$\overline{\sigma}_{11}$	$\overline{\sigma}_{22}$	$\overline{\sigma}_{33}$	$\overline{\sigma}_{12}$	$\overline{\sigma}_{23}$	$\overline{\sigma}_{13}$
Hidden layer 1	[16, 32, 64, 128]	32	32	32	32	32	128
Hidden layer 2	[16, 32, 64, 128]	64	64	128	64	128	64
Hidden layer 3	[256, 512, 1024, 2048]	512	256	1024	1024	256	1024
Hidden layer 4	[32, 64, 128, 256]	128	128	256	256	64	128
Learning rate in 10^{-5}	[5, 10, 50, 75, 100]	75	5	5	10	75	5

from a function $f(\mathbf{x})$ to its i-th input feature x_i is formally defined as:

$$\mathrm{IG}(X_i^k) = \left[X_i^k - X_i'^k\right] \int_{\xi=0}^{1} \frac{\partial\, \mathrm{f_{GNN}}\left(\left(\mathcal{V}, \mathcal{E}, \mathbf{X}' + \xi\left(\mathbf{X} - \mathbf{X}'\right)\right), \overline{\boldsymbol{\varepsilon}}\right)}{\partial X_i^k}\, \mathrm{d}\xi, \tag{23}$$

where X_i^k is the input feature i associated with node k, $X_i'^k$ is the corresponding baseline input, and ξ is the interpolation constant for the path from $X_i'^k$ to X_i^k. The integral is solved via the Gauss-Legendre approximation. In the context of the proposed GNN framework in this work, the model $\mathrm{f_{GNN}}$ maps the graph $\mathcal{G}$ and the applied strain $\overline{\boldsymbol{\varepsilon}}$ to a component of the homogenized stress $\overline{\sigma}_{ij}$. IG is applied to determine the influence of the initial node features $\mathbf{X}$ on the final prediction. The baseline $\mathbf{X}'$ is chosen as a zero matrix, representing a neutral microstructure with no specific grain information. This choice satisfies the completeness axiom of IG, which guarantees that the sum of all feature attributions exactly equals the difference between the model output at the actual input and the prediction at the baseline, providing a fully self-consistent decomposition of the model output.

In the current framework, the resulting IG scores, also referred to as attributions, quantify the contribution of each feature of each grain to the predicted stress component. A high absolute IG value for a specific feature of a given grain indicates that this characteristic was highly influential in the model's prediction for that particular macroscopic strain state. By aggregating and visualizing these attribution scores across all grains, it is possible to generate saliency maps overlaid on the MVEs. These maps highlight the critical grains and local features that govern the macroscopic mechanical response, thereby providing a direct, physically meaningful interpretation of the relationships captured by the GNN.

The quantitative attribution analysis in this work is restricted to the morphological node features. IGs are ill-posed for the crystallographic orientation encoded by the Bunge Euler angles, since orientation is a coupled quantity on $SO(3)$ whose individual angles are not independently interpretable and are non-unique under the 2π periodicity and the cubic crystal symmetry, while the zero baseline corresponds to the special Cube orientation rather than a neutral reference. The straight-line integration path therefore repeatedly traverses crystallographically equivalent orientations, at which the model output returns to near-identical values, and the accumulated integral spuriously inflates the orientation attribution. Morphological descriptors carry no such symmetry degeneracy, so their attributions remain well posed. This limitation, and possible remedies, are discussed in Section 4.

GradientShap [45] provides a sampling-based cross-check on the IG feature rankings. Rather than integrating gradients along a path from a single zero baseline, GradientShap approximates SHAP values as the expected gradient of the model output with respect to the input features, where the expectation is taken over a background distribution of reference microstructures and random noise perturbations. Formally, the attribution of feature i of grain k is approximated as

$$\mathrm{GS}(X_i^k) = \int_{\tilde{\mathbf{X}}} p_{\mathrm{bg}}(\tilde{\mathbf{X}}) \int_{\boldsymbol{\delta}} p_{\delta}(\boldsymbol{\delta})\, \frac{\partial f_{\mathrm{GNN}}(\tilde{\mathbf{X}} + \boldsymbol{\delta})}{\partial X_i^k} \left[X_i^k - \tilde{X}_i^k\right] \mathrm{d}\boldsymbol{\delta}\, \mathrm{d}\tilde{\mathbf{X}}\,, \tag{24}$$

where $\tilde{\mathbf{X}}$ denotes a reference microstructure drawn from the background distribution $p_{\mathrm{bg}}(\tilde{\mathbf{X}})$. In practice this distribution is represented by a finite set of N_{bg} MVEs from which $\tilde{\mathbf{X}}$ is sampled. The vector $\boldsymbol{\delta}$ is a small Gaussian perturbation, $\boldsymbol{\delta} \sim \mathcal{N}\left(\mathbf{0},\ \sigma_{\delta}^2 \mathbf{I}\right)$, with density $p_{\delta}(\boldsymbol{\delta})$, where σ_{δ} denotes the noise SD and is distinct from the stress $\boldsymbol{\sigma}$ or the width σ of the grain size distribution. The inner integral averages the gradient of the GNN output over all possible noise realizations, while the outer integral averages the same quantity over all possible baselines $\tilde{\mathbf{X}}$. Multiplying the gradient by the difference $(X_i^k - \tilde{X}_i^k)$ yields the contribution of feature i of grain k along the straight-line path from the baseline to the target microstructure. Because GradientSHAP integrates over a distribution of baselines rather than a single reference, the resulting attribution is far less sensitive to the particular choice of the baselines.

3 Results

In this section, the predictive capabilities of the proposed GNN framework are systematically evaluated against full-field CP simulations conducted on the synthetically generated MVE dataset. Building on the dataset generation from Section 3.1, where the reference CP simulations under diverse loading conditions are analyzed, the subsequent training and evaluation of the GNN models is described in Section 3.2. By benchmarking the GNN predictions against high-fidelity CP responses across microstructures exhibiting varying grain size distributions, aspect ratios, and all four crystallographic texture classes, the following results provide a comprehensive evaluation of the framework's ability to reproduce effective mechanical responses under tension as well as simple shear. The explainability analysis using IGs is presented in Section 3.4, providing grain-level attribution maps that validate the learned structure-property relationships against established crystal plasticity principles.

3.1 Dataset and simulation setup

Figure 3 shows the resulting stress-strain plots featuring the range of data that exists within the simulated cases. The simulations show significant variability in the mechanical properties that result from the microstructural variations

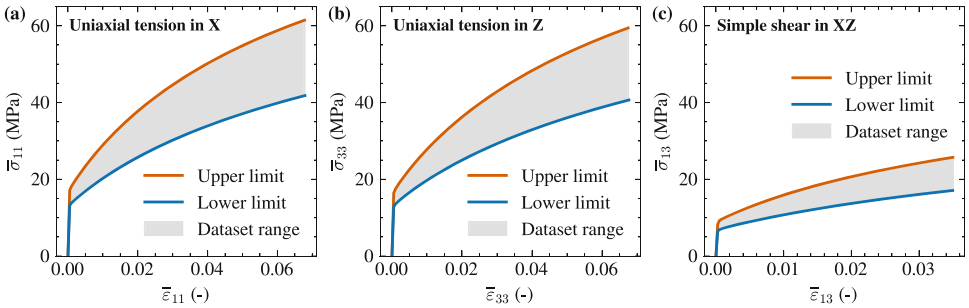

Fig. 3 Overview of the homogenized stress-strain responses observed in the synthetic dataset for three distinct loading conditions, (**a**) uniaxial tension in X, (**b**) uniaxial tension in Z, and (**c**) simple shear in the XZ plane. The red and blue curves represent the upper and lower boundaries of the dataset, respectively, while the gray-shaded region illustrates the full spectrum of stress responses encountered. These bounds highlight the range of mechanical behavior generated by varying microstructural features only

detailed in Section 2.1. For the uniaxial loading cases shown in Fig. 3 (a) and (b), the simulations exhibit a distinct dispersion in flow stress behavior at maximum strains of up to 7 %. In the simple shear loading case, Fig. 3 (c), the evolution of the shear stress component is presented. While all components of the stress tensor may evolve, even under uniaxial or simple shear loading, the analysis herein is restricted to the primary load-bearing components.

3.2 GNN training and performance evaluation

A detailed overview of the training process for the six best performing candidate GNN models used to predict the homogenized stress components is provided in Fig. 4. The training process, conducted on an NVIDIA Tesla V100 GPU (16GB), required an average of 62.5 minutes per 100 epochs, corresponding to approximately 52 hours per model over 5,000 epochs.[1] Once trained, the GNN surrogate predicts the stress response of a single MVE in 0.24 s, compared to 13,334 s (3.7 h) for a full CP simulation on an Intel® Xeon® Platinum 8160 CPU @ 2.10 GHz, leading to a speedup of approximately 55,000×.

The predictive performance of the GNN across training, validation, and test sets for the six independent components of the homogenized Cauchy stress tensor for all simulation data is illustrated in Fig. 5. Each subplot compares predicted and target stress values for specific components, i.e. normal stresses in a-c and shear stresses in d-f. Every subplot pools the three loading cases, so a given normal component is the loaded one only in its own tension case. In the remaining cases it is either traction free, as for $\overline{\sigma}_{11}$ under tension in Z where $P_{11} = 0$ is imposed, or unconstrained, as under simple shear where the deformation gradient is fully prescribed. Its homogenized value is then several orders of magnitude smaller and of either sign, which is the origin of the small negative normal stresses visible in the figure. These are present in the CP reference data and follow from the imposed boundary conditions together with the anisotropy of the grains. The near-unity R^2 values and low RMSE metrics across all data partitions indicate that the model captures both normal and shear components with high accuracy and precision with no signs of overfitting. Notably, the model demonstrates strong predictive performance on the held-out test set. The reported figures derive from a single partition, and a 5-fold cross-validation of the $\overline{\sigma}_{11}$ model, reported in Appendix B, places them at the conservative edge of the fold-wise distribution rather than above it, with an RMSE of 0.086 ± 0.007 MPa across the folds. As this test set is drawn from the same distribution as the training data, the bounds of generalization to genuinely out-of-domain microstructures are examined separately in Section 3.3.

The stress-strain responses are illustrated in Fig. 6, featuring the data obtained via CP simulations and the proposed GNN surrogate models that are compared for six representative test set MVEs covering the full range of microstructure descriptor classes. The GNN model successfully captures the general stress-strain curves with high accuracy and precision for both $\overline{\sigma}_{11}$ and $\overline{\sigma}_{33}$ while showing slight deviations of the shear response around the yield point. This is to be expected according to the training curve shown in Fig. 4 (f) where $\overline{\sigma}_{13}$ shows the highest final validation MSE reported of all the final six models, indicating it was the most challenging component for the model to learn. Several factors may contribute to this larger deviation. Firstly, under simple shear loading significant normal stresses are expected to arise that affect the observed shear stresses in a strongly nonlinear manner. Secondly, shear components often arise from complex, localized grain interactions, crystallographic rotations, and heterogeneous slip system activation rather than being a direct response to the primary applied load. The model

[1] While most models demonstrated stable convergence over the full 5,000 epochs, the GNN for the $\overline{\sigma}_{13}$ component exhibited signs of potential overfitting. To obtain a more robust model, a second training was performed for this component using an early stopping criterion with a patience of 500 epochs, which concluded the training at 14,714 epochs.

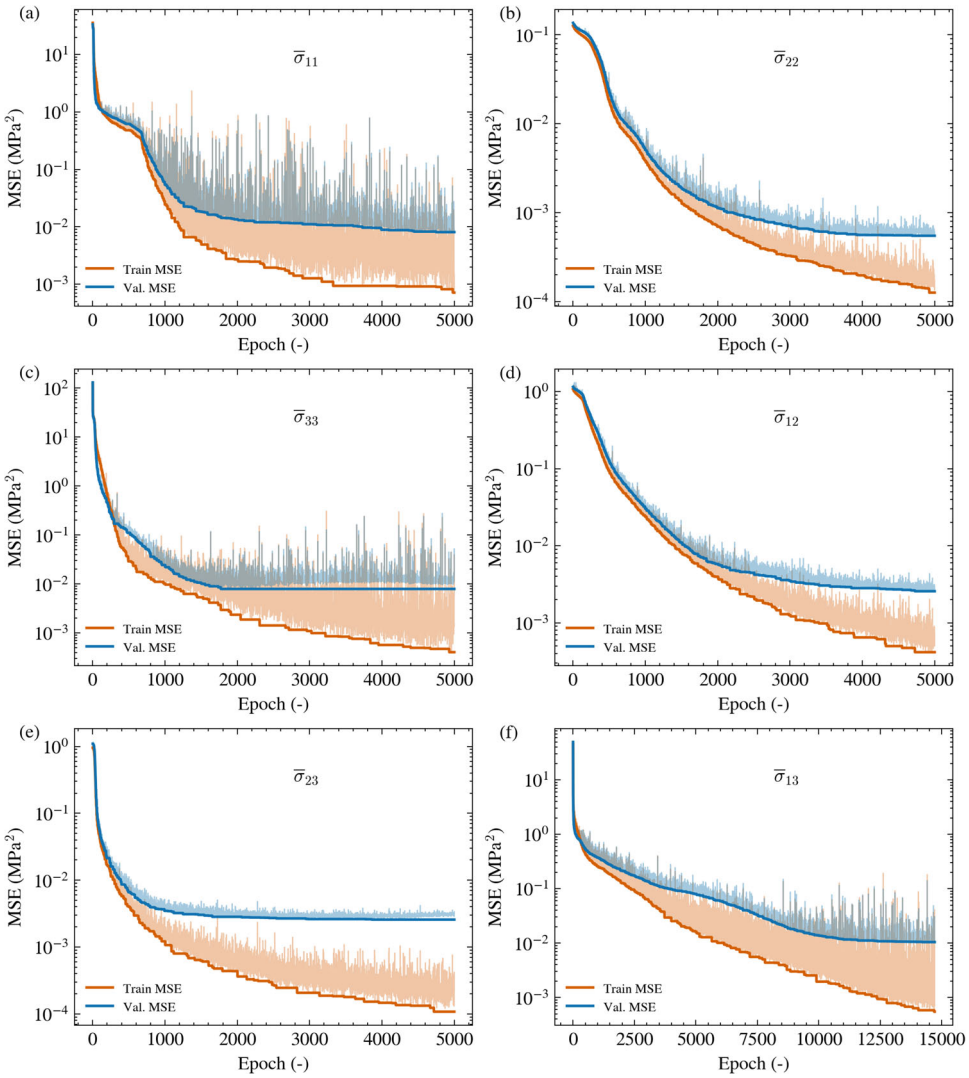

Fig. 4 Evolution of the MSE on a logarithmic scale over 5,000 training epochs for both training (red) and validation (blue) datasets for (**a**) $\overline{\sigma}_{11}$, (**b**) $\overline{\sigma}_{22}$, (**c**) $\overline{\sigma}_{33}$, (**d**) $\overline{\sigma}_{12}$, (**e**) $\overline{\sigma}_{23}$, (**f**) $\overline{\sigma}_{13}$. The light-colored lines indicate the MSE evolution, while the bold lines highlight the lowest MSE achieved during training so far

may therefore struggle to perfectly resolve these effects, particularly in regions where shear localization or sharp stress gradients occur at grain boundaries.

Figure 6 (a) shows the micromechanical response of a Type 1 equiaxed MVE with random texture. The CP reference data shows no mechanical anisotropy between the two uniaxial loading directions, which is evident from the overlapping stress-strain curves, and the GNN captures this symmetric behavior with high fidelity. Figure 6 (b) shows a Type 2 MVE with pronounced morphological anisotropy and random texture, where the two uniaxial loading curves diverge and the GNN successfully reproduces this morphology-induced anisotropy. Figure 6 (c)-(f) present Type 3 MVEs with randomized morphology and each of the four crystallographic texture classes. Depending on the texture, the stress–strain response exhibits anisotropy. The GNN captures with both isotropic and anisotropic responses with strong agreement, demonstrating that the learned node embeddings encode texture-dependent mechanical behavior across all four classes.

3.3 Bounds of generalization

The evaluations reported so far are performed on microstructures that are represented in the training partition. To delimit the range over which the surrogate can be trusted, four hold-out scenarios were constructed for the $\overline{\sigma}_{11}$ component, ordered by the severity of the shift imposed. Each scenario withholds entire microstructures from training, together with all of their loading conditions. In the texture scenario all 149 MVEs carrying the Goss texture are removed, so that the model encounters this texture component for the first time

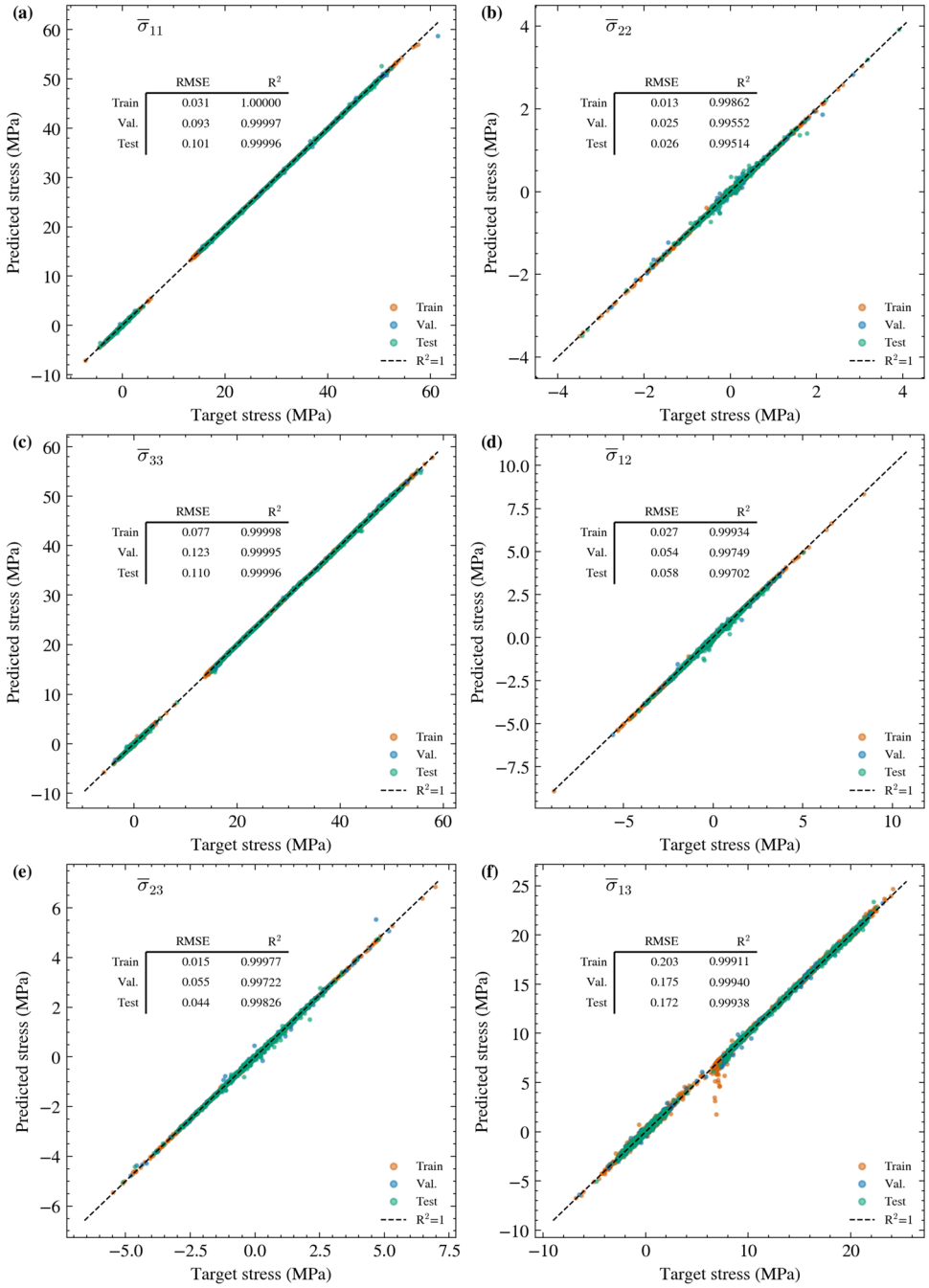

Fig. 5 Comparison of predicted versus target stress values for the six components of the homogenized Cauchy stress for all investigated loading cases, (**a**) $\overline{\sigma}_{11}$, (**b**) $\overline{\sigma}_{22}$, (**c**) $\overline{\sigma}_{33}$, (**d**) $\overline{\sigma}_{12}$, (**e**) $\overline{\sigma}_{23}$, and (**f**) $\overline{\sigma}_{13}$. In each subplot, the red, blue, and green markers represent training, validation, and test sets, respectively. The dashed black line corresponds to perfect agreement (R^2=1). The insets report the RMSE in MPa and the coefficient of determination (R^2) for each dataset, illustrating the model's high fidelity in reproducing both normal and shear stress components across the entire data spectrum

during testing. In the MVE type scenario the 200 MVEs of the morphologically anisotropic Type 2 family are removed in their entirety, so that an entire design family of the dataset is unseen. In the two descriptor extrapolation scenarios the MVEs are ordered by their mean ESD and by their mean minimum aspect ratio $a_{C/A}$, respectively, and the lower and upper deciles of each distribution, 200 MVEs in total, are withheld, so that the model is required to extrapolate beyond the morphological range on which it has been trained.

The four microstructural scenarios are compared with their contrast sets in Fig. 7. Because the in-domain contrast sets are withheld from training in exactly the same way as the shifted sets, and differ from them only in lying inside the training distribution, the difference between the two isolates the effect of the shift from the effect of the microstructure being unseen. For the four microstructural shifts the degradation is modest or absent. Withholding the Goss texture and withholding the entire Type 2 family both leave the out-of-distribution accuracy indistinguishable from the in-domain contrast, at $R^2 = 0.9959$ and 0.9964 against 0.9947 and 0.9951 respectively. The small excess of the out-of-distribution value over its contrast is within the run-to-run

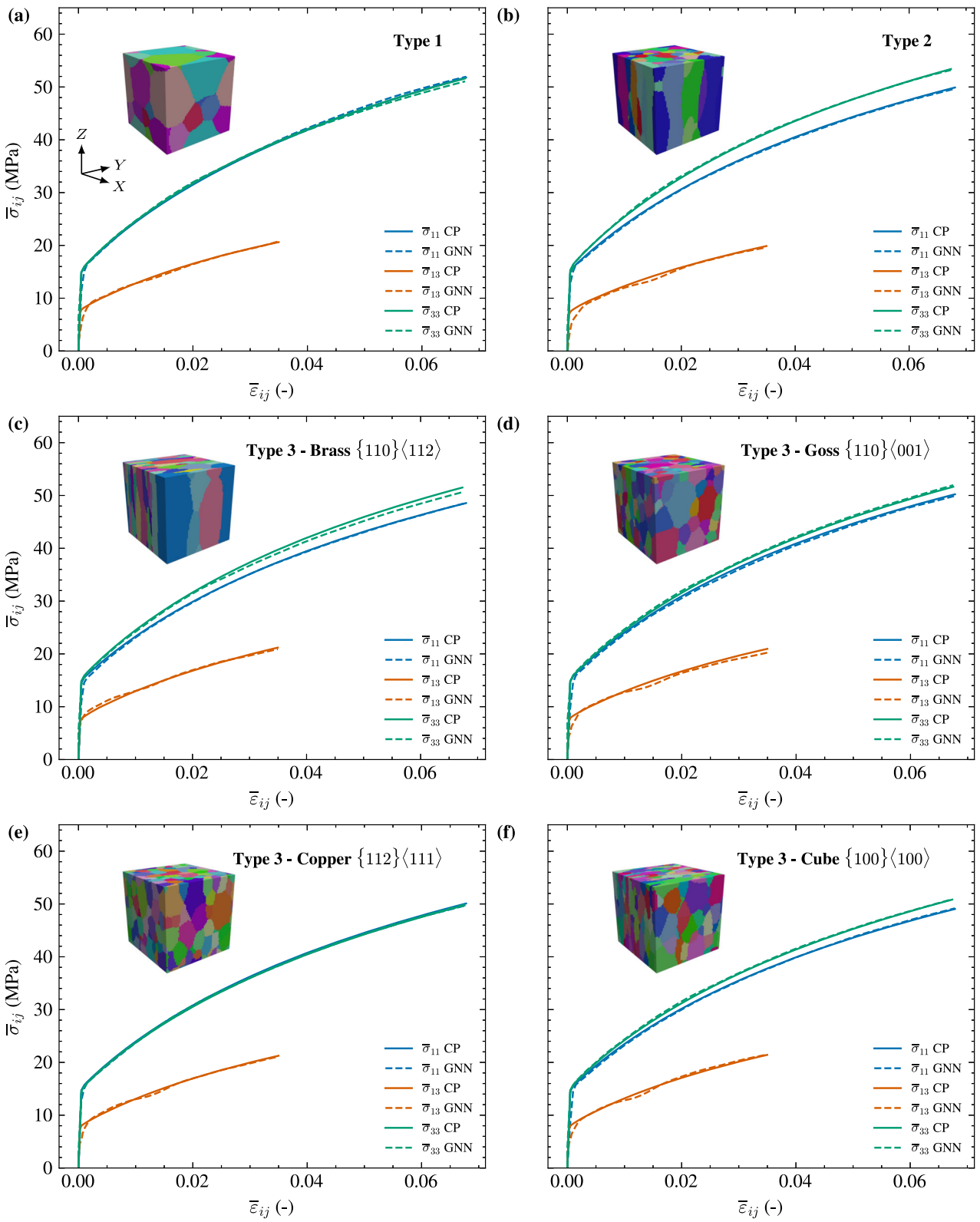

Fig. 6 Comparison of homogenized stress components for six representative microstructures using CP (solid lines) and GNN (dashed lines). Each panel shows stress-strain curves for $\overline{\sigma}_{11}$ (blue), $\overline{\sigma}_{33}$ (green), and $\overline{\sigma}_{13}$ (red). Predictions are shown for (**a**) Type 1 MVE with equiaxed grains and random texture, (**b**) Type 2 MVE with pronounced morphological anisotropy and random texture, and (**c-f**) Type 3 MVEs with randomized morphology and the four crystallographic texture classes

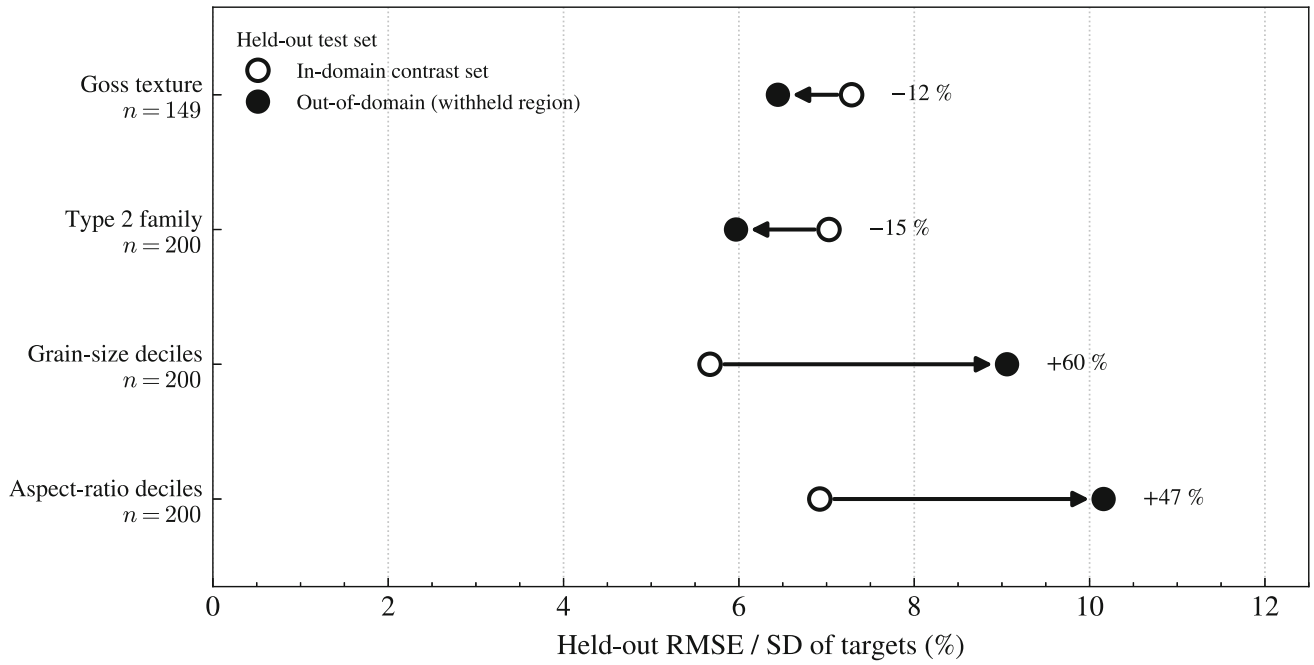

Fig. 7 Cost of a distribution shift for the four microstructural hold-out scenarios. Each row is a single model, trained with the named region removed, and evaluated on two disjoint held-out sets, namely microstructures that are equally absent from training but drawn from the interior of the distribution (open markers), and the withheld region itself, which is both absent from training and outside the training distribution (filled markers). Both sets are equally unseen, so the arrow between them measures the effect of the distribution shift alone, and is annotated with the relative change in error. The error of each set is divided by the SD of its own targets, so that the rows are comparable and the abscissa reads as the error relative to the spread of the quantity being predicted

scatter documented in Appendix C and should not be read as superior performance. Extrapolation to the tails of the morphological distributions does carry a measurable cost, the RMSE rising from 0.90 to 1.46 MPa for grain size and from 1.10 to 1.63 MPa for aspect ratio. The surrogate therefore transfers to an unseen texture component, to an unseen design family and to the tails of the morphological range with a loss of accuracy of at most about 50 % in RMSE, which still leaves it usable for screening.

Within the interpolation regime, for microstructures represented in training, the accuracy reported in Section 3.2 applies and is reliable in the sense established in Appendix B. For a microstructure absent from training, an RMSE of the order of 1 MPa, i.e. a relative error of a few percent, should be expected, and this figure is only moderately sensitive to a distribution shift in texture, MVE type, grain size or grain morphology. Closing this gap requires a larger ensemble of distinct microstructures or a richer graph encoding, as discussed in Section 4.

3.4 Feature interpretation using integrated gradients

The GNN surrogate is interpreted using IGs in an analysis that proceeds from local to global. Single-instance saliency maps identify the features that dominate the predicted stress for individual microstructures, population-level attribution statistics then establish whether those trends hold across the dataset, and a cross-check against GradientShap tests whether the resulting rankings depend on the choice of attribution method.

IG attribution maps are first computed for a single Type 1 MVE under uniaxial tension in the X-direction at 7% strain. The instance shown is the median of the 200-point Type 1 grain-size sweep ($\mu = 1.50$), chosen so that it represents a typical member of that set rather than one of its extremes. The maps, Fig. 8, color each grain by its normalized attribution score for each of the four morphological node features (d_{ESD}, N_{nb}, $a_{C/A}$, $a_{B/A}$), and over the 66 grains of this instance the mean absolute attributions are 1.5×10^{-3} for $a_{B/A}$, 1.1×10^{-3} for d_{ESD}, 1.1×10^{-3} for N_{nb} and 4.7×10^{-4} for $a_{C/A}$. Repeating the analysis for all 200 Type 1 MVEs shows the leading feature to be stable, $a_{B/A}$ ranking first in 142 of them, whereas the two middle features are not. N_{nb} outranks d_{ESD} in 61% of the instances, and the ratio of those two attributions rises with the log-mean grain size at Spearman +0.54, so that neighboring members of the sweep exchange them. Figure 8 is therefore one example of instance-level behavior rather than a template for it.

An instance-level ranking should not be read as a physical ranking of the features. Because IG integrates the gradient along the path from a zero baseline to the observed input,

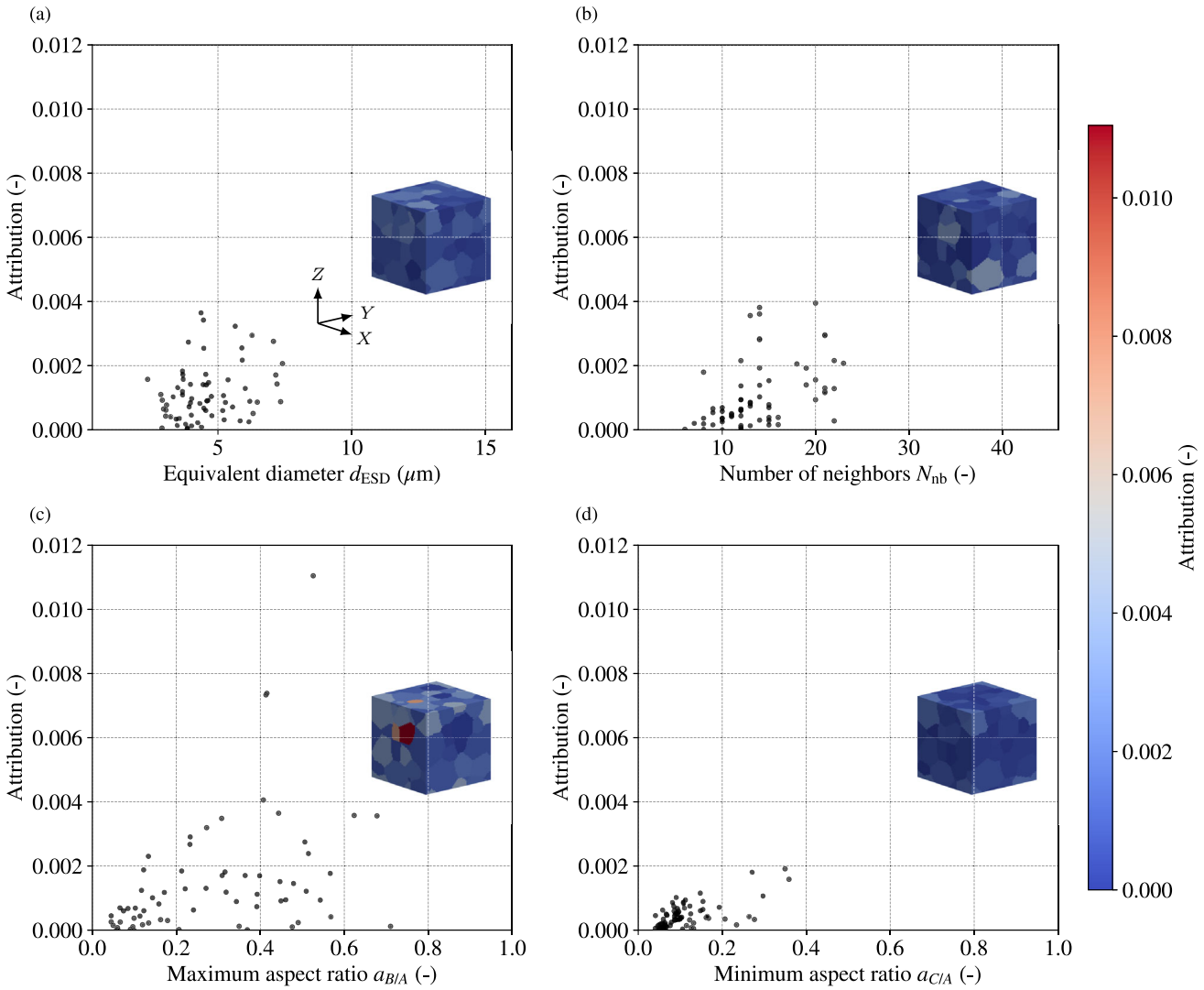

Fig. 8 Saliency maps derived from IG illustrating the feature importance for predicting $\overline{\sigma}_{11}$ in a single Type 1 MVE under uniaxial tension (7% strain), taken at the median of the Type 1 grain-size sweep ($\mu = 1.50$). Grains are colored by their normalized attribution score for (**a**) ESD, (**b**) number of neighboring grains, (**c**) maximum aspect ratio $a_{B/A}$, and (**d**) minimum aspect ratio $a_{C/A}$. The maximum aspect ratio and the equivalent diameter exhibit the highest attributions here, while the minimum aspect ratio is suppressed due to the Y-axis alignment constraint as discussed in the text. The ordering of d_{ESD} and N_{nb} varies between individual MVEs and should not be read from one map, and the statistics over all 200 Type 1 instances are therefore given in the text. The coordinate triad in (a) gives the orientation of the global axes and applies to the MVE renderings in all panels

an attribution measures how far the prediction moves as a feature is carried from zero to its actual value, not the sensitivity of the prediction to the variation of that feature about its mean. What the instance map records is therefore the sensitivity of the model to the grain-to-grain contrast within one microstructure. No mechanism is assigned here to the ordering of the two aspect ratios, since the major grain axis is aligned with the global Y-axis in every MVE while the load is applied along X, so both ratios are referred to a common, fixed denominator and neither offers the orientation contrast that would make such an interpretation meaningful.

To assess whether these trends hold beyond individual instances, IG attributions are computed for all 1,000 MVEs across the three loading conditions, Fig. 10. The maximum aspect ratio $a_{B/A}$ and the equivalent sphere diameter d_{ESD} rank as the most influential morphological features under every loading condition, while the number of neighbors N_{nb} carries the lowest mean attribution of the four. The attribution distributions are right-skewed throughout, indicating that a subset of grains with extreme morphological characteristics drives the macroscopic response disproportionately.

The prominence of d_{ESD} must not be interpreted as a physical grain-size or Hall–Petch effect. The crystal plastic-

ity formulation employed here contains no length scale, so the ground-truth response cannot exhibit a grain-size dependence. This is confirmed directly, in that the 0.2% offset yield strength of the 200 equiaxed Type 1 MVEs shows no correlation with $1/\sqrt{d_{\mathrm{ESD}}}$, Fig. 9 (a). Because every MVE occupies the same fixed volume, the mean grain size is instead an almost deterministic proxy for the number of grains.

Grain count, however, is not mechanically inert, and this is what the attribution recovers. It leaves the mean response untouched, Fig. 9 (b), but it governs the dispersion of the response about that mean, which is the characteristic signature of what is commonly termed the statistical size effect, entirely distinct from the mechanistic, strength-altering size effect ruled out above. Each grain samples the orientation distribution independently, so a volume element containing N grains estimates the polycrystal-average response with a scatter that decays as $N^{-1/2}$, exactly as expected when a finite volume element is used to approximate the response of an infinite, statistically homogeneous polycrystal. The equiaxed, randomly textured Type 1 set isolates this behavior cleanly, grain size being the only descriptor varied across it. Binning by grain count and regressing log(SD) on log N returns an exponent of -0.48 ± 0.07 for the yield strength ($R^2 = 0.89$), recovering the theoretical value of $-1/2$ within one standard error, Fig. 9 (c). At fixed volume, d_{ESD} is consequently the best available predictor of how far an individual MVE departs from the polycrystal-average response, and that departure is what a surrogate operating on individual MVEs must resolve, the average itself being common to all of them. The high attribution assigned to d_{ESD} therefore reflects microstructural representativity rather than strength (Fig. 10).

The pairwise Spearman rank correlation matrices, Fig. 11, reveal that all inter-feature attribution correlations are positive and range from moderate (0.52) to strong (up to 0.93) across the three loading conditions, indicating that the importance of morphological features broadly co-varies across the MVE population. This co-variation reflects the statistical structure of the generated dataset, where microstructures with large grains tend to exhibit more elongated morphologies, so the GNN's sensitivity to one morphological descriptor naturally extends to the others. The notably high correlations observed under $\overline{\sigma}_{11}$ compared to the more moderate and broadly comparable values under $\overline{\sigma}_{33}$ and $\overline{\sigma}_{13}$ suggest that uniaxial tension in X produces the strongest coupling between morphological feature importance under the generated dataset. This is consistent with the preferential alignment of grain elongation along Y, which introduces a directional morphological contrast that is most strongly activated by loading along X.

A complementary question is whether the relative importance of features changes at different deformation stages. To investigate this, IG attributions are tracked as a function of applied strain. Figure 12 shows the mean absolute attribution per feature as a function of strain for the three loading conditions. The attribution trajectories reveal that d_{ESD} is the dominant morphological predictor throughout the entire deformation history for both uniaxial loading cases, with its attribution magnitude growing steadily with applied strain. This growth does not indicate a physical grain-size mechanism, which is absent from the scale-free formulation, but follows from the representativity argument developed above. The dispersion of the homogenized response across MVEs widens as deformation proceeds, from 3.35% of the mean at yield to 4.20% at 5% strain, while the $N^{-1/2}$ grain-count control of that dispersion remains unchanged. Grain size therefore becomes progressively more informative about an individual MVE as the plastic response develops. The oscil-

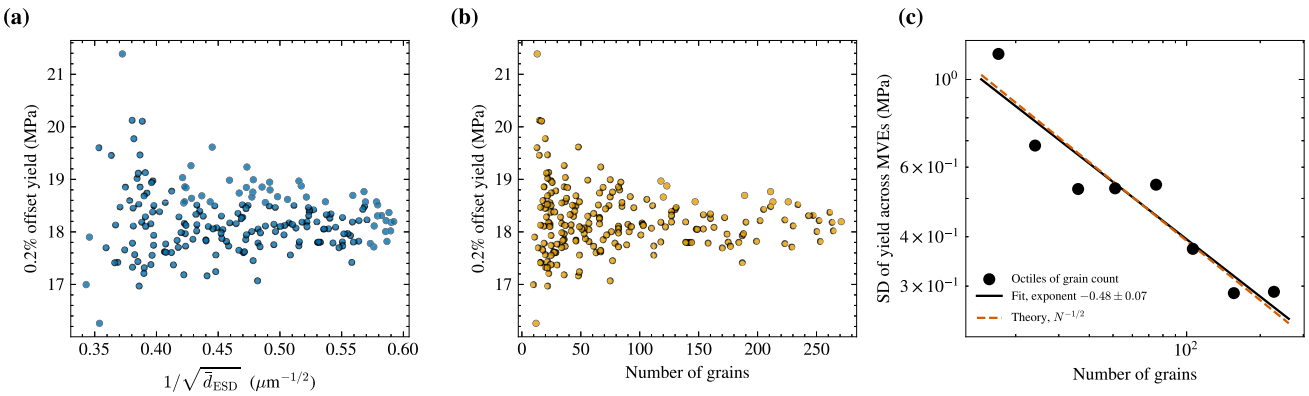

Fig. 9 Test for a physical grain-size effect against the statistical size effect, evaluated over the 200 equiaxed Type 1 MVEs. (**a**) 0.2% offset yield strength against $1/\sqrt{d_{\mathrm{ESD}}}$, which carries no Hall–Petch dependence, the Pearson correlation being -0.08 at $p = 0.29$. (**b**) The same yield strength against the number of grains N, whose mean is equally flat while the scatter about it narrows as N grows. (**c**) SD of the yield strength across octiles of grain count on logarithmic axes, with the fitted exponent -0.48 ± 0.07 at $R^2 = 0.89$ and the $N^{-1/2}$ prediction anchored at the data centroid so that only the slopes are compared

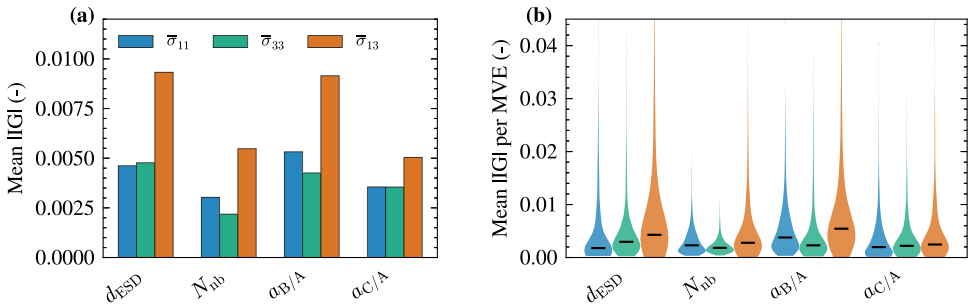

Fig. 10 Population-level IG attribution statistics computed across 1,000 MVEs and three loading conditions. (**a**) Mean absolute attribution ranked by feature, and (**b**) violin plots showing the full attribution distribution per feature

lations superimposed on these trajectories are not assigned a physical interpretation. Although the applied strain is supplied to the GNN as a continuous input, the surrogate was trained on CP responses available only at discrete strain increments. The IG attributions are evaluated independently at the corresponding strain states, and local variations in the learned interpolation and its gradients may therefore appear between adjacent states, particularly across the change in strain-sampling increment at approximately 1% strain. These fluctuations are consequently regarded as a surrogate and attribution sampling effect, while the overall feature rankings and broad strain-dependent trends remain unaffected. N_{nb} attribution remains comparatively stable and low across all strain levels, indicating that local grain connectivity plays a secondary and largely strain-independent role in determining the macroscopic stress response under uniaxial loading. Under simple shear, the attribution trajectories present a significantly different character, where all four morphological features contribute comparably throughout the deformation history, with no single feature emerging as a clear dominant predictor.

A further test of physical consistency is whether the attribution maps differentiate microstructures with known mechanical contrast. IG attributions are therefore compared for the hardest and softest MVEs in the dataset, defined as the top and bottom 5% of the flow-stress distribution at 7% strain. Figure 13 shows the mean absolute attributions for these outlier populations alongside the full-population average. Both hard and soft outlier MVEs exhibit mean attributions approximately two to three times the full-population average across all morphological features and loading conditions. This symmetric elevation reflects that microstructures at both extremes of the flow-stress distribution are also outliers in the morphological feature space, exhibiting strong statistical deviations from the population mean, to which the GNN assigns correspondingly amplified sensitivity irrespective of the direction of the mechanical contrast. Therefore, the systematic elevation of attributions across both extremes indicates that the GNN's learned representations respond to genuine microstructural heterogeneity rather than producing uniformly distributed gradient noise.

As a final validation, GradientShap [45] is applied as an additional attribution method. GradientShap combines gradient computation, similar to IG, with SHAP's sampling framework, providing an independent estimate of feature importance based on cooperative game theory. Figure 14 shows the normalized mean absolute attributions from IG and GradientShap side by side for each morphological feature and loading condition. For both uniaxial loading conditions, the two methods are in qualitative agreement, where $a_{B/A}$

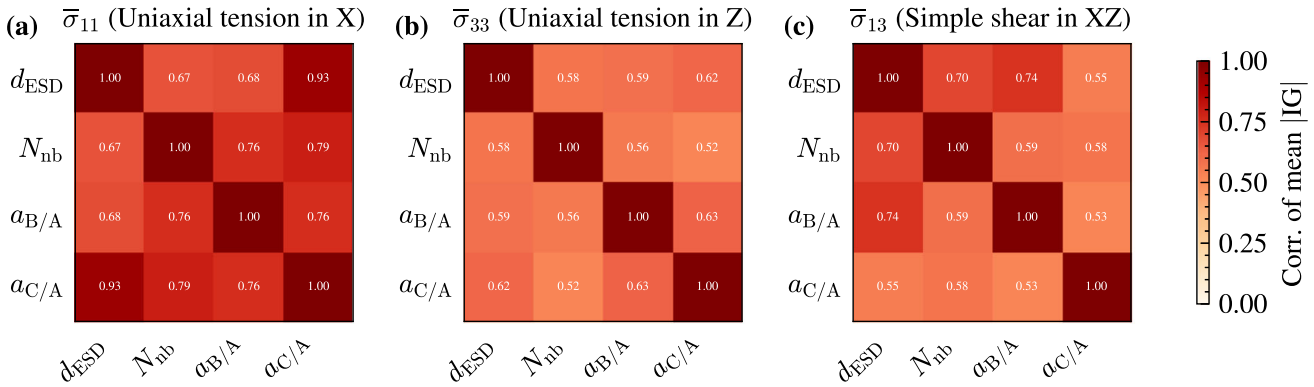

Fig. 11 Pairwise Spearman rank correlations of mean absolute IG attribution between morphological features, computed across 1,000 MVEs for (**a**) $\overline{\sigma}_{11}$, (**b**) $\overline{\sigma}_{33}$, and (**c**) $\overline{\sigma}_{13}$. Each matrix entry gives the Spearman correlation between the mean absolute attributions of two features across the MVE population. High values indicate features whose importance co-varies across microstructures, reflecting coupled morphological effects on the predicted stress response

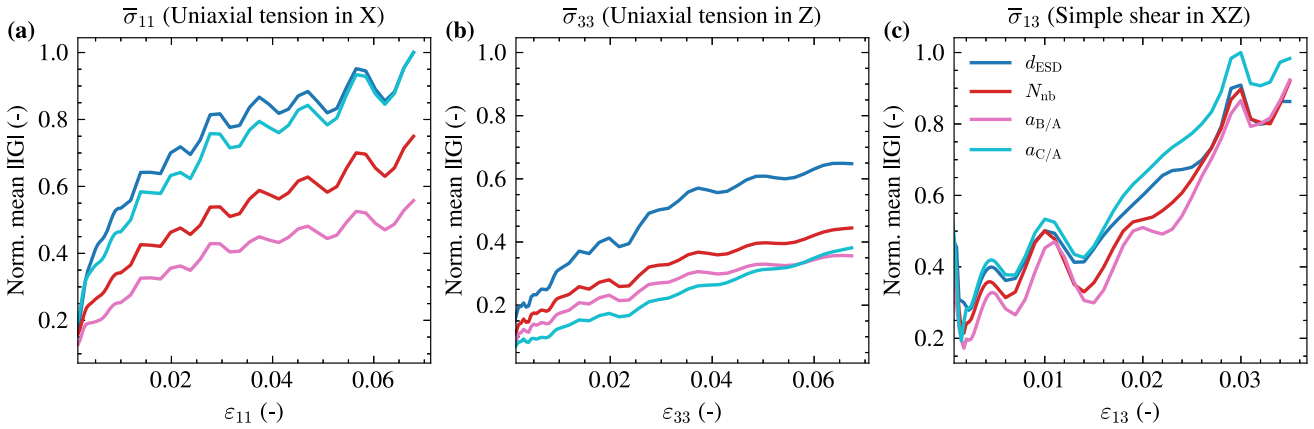

Fig. 12 Evolution of mean absolute IG attribution per feature as a function of applied strain for (**a**) uniaxial tension in X, (**b**) uniaxial tension in Z, and (**c**) simple shear in XZ. Attributions are averaged over 1,000 MVEs at each strain step

and d_{ESD} emerge as the dominant morphological predictors and N_{nb} consistently shows the lowest normalized attribution across both methods. Under simple shear, however, the two methods diverge in their relative feature rankings, with GradientShap assigning comparatively higher importance to $a_{C/A}$ than IG does. This discrepancy likely reflects the sensitivity of attribution estimates to the choice of baseline and sampling strategy under a more complex loading state. Despite this disagreement under shear, the overall consistency between two methods with distinct theoretical foundations supports the conclusion that the reported feature importance reflect genuine properties of the trained GNN rather than artifacts of a particular attribution algorithm.

4 Limitations and implications

The present framework is developed within a well-defined scope. The assumptions and limitations that bound the applicability of the results, and that delineate the directions in which the work can be extended, are collected here.

The surrogate is a quasi-static map from the current strain state to the stress and carries no internal-state variable, so it applies to the monotonic, proportional loading paths on which it was trained. Genuine history- or path-dependent behavior, whether unloading, cyclic or non-proportional loading, is not captured, and reaching those regimes would require temporal or recurrent GNN architectures together with substantially larger training datasets. The same argument applies to the imposed loading rate. Every simulation was run at the single deformation-gradient rate $\dot{F} = 10^{-3}\ \mathrm{s}^{-1}$, whereas the constitutive description is rate sensitive through its power-law flow rule with $m = 0.05$ (Table 2), for which a decade in rate shifts the flow stress by approximately twelve percent. The rate is therefore absorbed into the learned map rather than supplied to it, and the surrogate should not be applied at a different loading rate without retraining. The restriction extends to the loading mode itself.

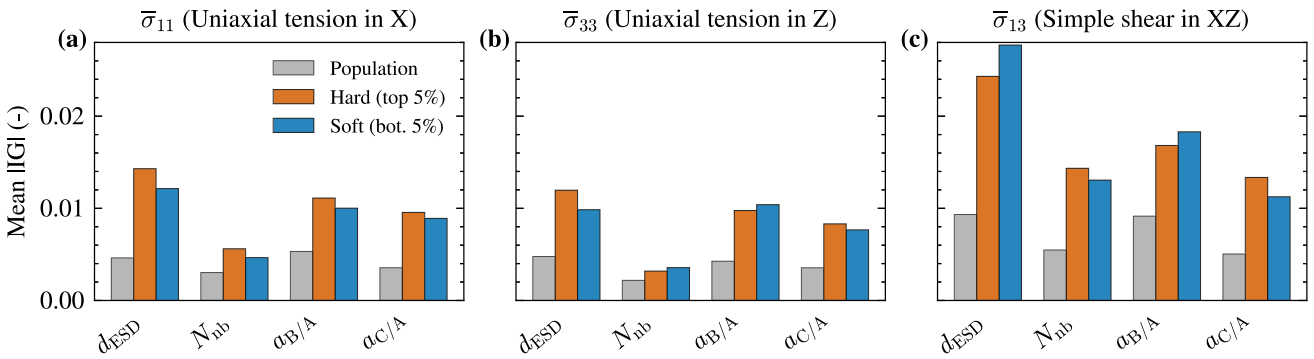

Fig. 13 Mean absolute IG attributions for morphological features comparing the population average (gray), the top-5% flow-stress MVEs (hard, orange), and the bottom-5% flow-stress MVEs (soft, blue), for (**a**) $\overline{\sigma}_{11}$, (**b**) $\overline{\sigma}_{33}$, and (**c**) $\overline{\sigma}_{13}$. Both hard and soft outlier MVEs exhibit substantially elevated attribution magnitudes compared to the full population average, indicating that mechanically extreme samples are characterized by concentrated feature importance regardless of their position in the flow-stress distribution

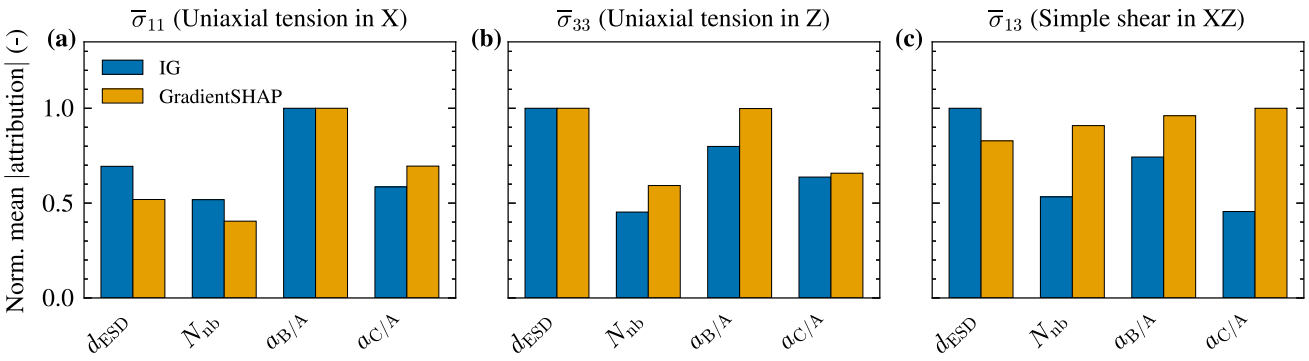

Fig. 14 Comparison of normalized mean absolute attributions from IG (blue) and GradientShap (orange) for each morphological feature under (**a**) uniaxial tension in X, (**b**) uniaxial tension in Z, and (**c**) simple shear in XZ

Only the three loading conditions of Table 3 are represented in the training data, and every loading mode for which the surrogate is to be used must be included in the training campaign, since it is not inferred from the applied strain tensor alone.

For a microstructure absent from training, the error rises to roughly six percent of the SD of the response, against about half a percent for microstructures represented in training, rather than the accuracy reported in Section 3.2, and this holds largely irrespective of whether the microstructure lies inside or outside the training distribution, Section 3.3. The limit is set by the number of distinct microstructures available and by the node-only feature encoding rather than by the optimization. Grain-boundary character, misorientation and energy are not encoded as edge attributes, so the graph captures grain connectivity but not the full grain-boundary state. The microstructures themselves span a deliberately narrow range in two further respects. All are generated in a fixed $16 \times 16 \times 16\ \mu\text{m}^3$ volume and mapped to graphs of up to 282 grains, and the framework has not been tested at markedly larger volumes or grain counts. The constitutive parameters correspond to pure face-centered cubic aluminum, so the model is calibrated to that system and is not expected to transfer to other materials or crystal structures without retraining. On the output side, the six stress components are predicted by independent models with no imposed coupling, so that tensor symmetry holds by construction but no equilibrium or objectivity relation links the components.

The attribution results carry their own qualifications, which follow from the absence of a length scale in the constitutive description and from the homogenized target. Because the CP formulation contains no length scale, the ground truth exhibits no grain-size effect, and the attribution assigned to d_{ESD} reflects the fixed-volume coupling between grain size and grain count, through which grain count sets the representativity of an MVE, rather than a physical size effect. Any conclusion drawn from that attribution is therefore specific to a dataset generated at fixed volume and would not carry over to one in which volume and grain count were varied independently. Relative-size effects within an MVE, namely the volume weighting of each grain in the average and the lower boundary-constraint-to-volume ratio of large grains, are compatible with a scale-free formulation and are not excluded by the present measurements. They act on the internal stress field, which homogenization integrates away, so that such partitioning is absent from the surrogate's target and cannot be recovered by attribution. Resolving it would require a model trained on grain-resolved rather than homogenized stresses. Attribution to crystallographic orientation is ill-posed for a separate reason, namely the crystal-symmetry degeneracy of the Euler-angle representation together with the non-neutral zero baseline. A physically meaningful orientation attribution would require integration along geodesic paths on $SO(3)$ within the fundamental zone, or attribution to a symmetry-invariant, load-specific scalar such as the Schmid or Taylor factor.

5 Conclusion

This work demonstrates that GNNs are powerful and interpretable surrogates for high-fidelity CP simulations of polycrystalline materials. By utilizing an intuitive graph-based representation of microstructures, where each grain is a node and grain interactions are encoded in the adjacency matrix, the proposed GNN model accurately and precisely captures mechanical structure-property relationships, enabling rapid prediction of macroscopic stress-strain responses. The application of IGs allows moving beyond black-box predictions and validates the physical basis of the GNN's decision-making process at both the individual-grain and population levels.

The proposed GNN surrogate accurately reproduces homogenized CP outputs across four crystallographic tex-

ture classes and a broad range of morphological anisotropy, achieving low RMSE on the held-out test set with no signs of overfitting. By predicting the stress response of a single MVE in 0.24 s compared to 13,334 s for a full CP simulation, the model achieves an inference-time speedup of approximately 55,000×, realized up to the reported accuracy relative to full-field CP and excluding the one-time training and data-generation overhead, enabling rapid screening of large microstructure design spaces. The graph-based encoding of grain topology captures both local and long-range grain interactions without the resolution-dependent scaling that constrains pixel- and voxel-based representations.

The comprehensive explainability analysis yields several physically grounded findings. Population-level IG attribution statistics computed across 1,000 MVEs consistently rank grain size (d_{ESD}) and the maximum aspect ratio ($a_{B/A}$) as the most influential features, while N_{nb} shows the lowest mean attribution. Importantly, because the crystal plasticity formulation is scale-free, the prominence of d_{ESD} does not correspond to a physical grain-size or Hall–Petch effect, which is confirmed to be absent from the ground truth. It reflects instead the near-deterministic coupling between grain size and grain count in the fixed-volume MVEs, through which grain size governs how far an individual MVE departs from the polycrystal-average response. The dispersion of that response scales as $N^{-1/2}$ with grain count and widens as deformation proceeds, which accounts for both the rank of d_{ESD} and the growth of its attribution with strain. This demonstrates that the interpretability analysis faithfully recovers the statistical structure that governs the individual response rather than imposing physical relationships that are not present. Cross-validation of IG attributions against GradientShap confirms that the principal feature rankings are largely robust across attribution methods, establishing confidence in the interpretability framework.

The scope boundaries set out in Section 4 define a clear roadmap for future extension, in which temporal GNN architectures can generalize the framework to history-dependent and cyclic loading, and richer edge encodings incorporating grain-boundary misorientation and energy will bring the surrogate closer to a full microstructural description. More broadly, this study lays a strong foundation for interpretable, data-driven materials design tools that combine predictive accuracy with physically meaningful explanations.

Appendix A: Statistical descriptors for synthetic microstructure generation

This appendix provides the mathematical definitions of the three microstructural distribution functions used in Section 2.1 to parametrize the DREAM.3D synthetic MVE generation.

The statistical ensemble of grain sizes is described by a lognormal GSDF, whose probability density is given by

$$f_{\mathrm{GSDF}}(d_{\mathrm{ESD}}, \mu, \sigma) = \frac{1}{d_{\mathrm{ESD}}\sigma\sqrt{2\pi}} \exp\left(-\frac{[\ln d_{\mathrm{ESD}} - \mu]^2}{2\sigma^2}\right), \tag{25}$$

where μ and σ are the mean and SD of the natural logarithm of the ESD d_{ESD}.

The morphological anisotropy of individual grains is captured through the ARDF. Two independent ARDF distributions are defined, one for the minimum aspect ratio $a_{C/A}$ and one for the maximum aspect ratio $a_{B/A}$. Each follows a Beta probability density,

$$f_{\mathrm{ARDF}}(a, \alpha, \beta) = \frac{a^{\alpha-1}(1-a)^{\beta-1}}{B(\alpha, \beta)}, \quad 0 < a < 1, \tag{26}$$

where $B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,\mathrm{d}x$ is the Beta function. The DREAM.3D shape parameters for each distribution are computed from the prescribed mean aspect ratios $\overline{a}_{B/A}$ and $\overline{a}_{C/A}$ as

$$\alpha = \left(1.1 + 28.9\,\overline{a}_{B/A}\right), \qquad \beta = \left(30 - 28.9\,\overline{a}_{C/A}\right), \tag{27}$$

where α and β are the respective shape parameters for the $a_{B/A}$ and $a_{C/A}$ distributions according to DREAM.3D's internal parameterization convention.

The crystallographic state of the grain ensemble is described by the ODF, which gives the probability density of finding a specific crystallographic orientation $\mathbf{g}$ in a volume V:

$$f_{\mathrm{ODF}}(\mathbf{g}) = \frac{1}{V}\frac{\mathrm{d}V}{\mathrm{d}\mathbf{g}}. \tag{28}$$

In this work, $\mathbf{g}$ is represented by Bunge ZXZ Euler angles $(\varphi_1, \phi, \varphi_2)$, and the ODF is discretized and optimized via a Monte-Carlo algorithm within DREAM.3D.

Appendix B: Reliability of the reported model

The performance quoted above derives from a single partition of the dataset. To establish that it is representative rather than a favorable draw, a 5-fold cross-validation was performed for the $\overline{\sigma}_{11}$ component, each fold retrained with the identical architecture, hyperparameters and 5,000-epoch budget of Table 4 and evaluated at the checkpoint of lowest validation error. The results of the analysis are shown in Fig. 15. Across the five folds performance on the held-out set show an RMSE of 0.086 ± 0.007 MPa and a relative error of 0.375 ± 0.018 %, quoted as mean ± SD. The spread is narrow with the RMSE varying by less than 0.02 MPa across folds

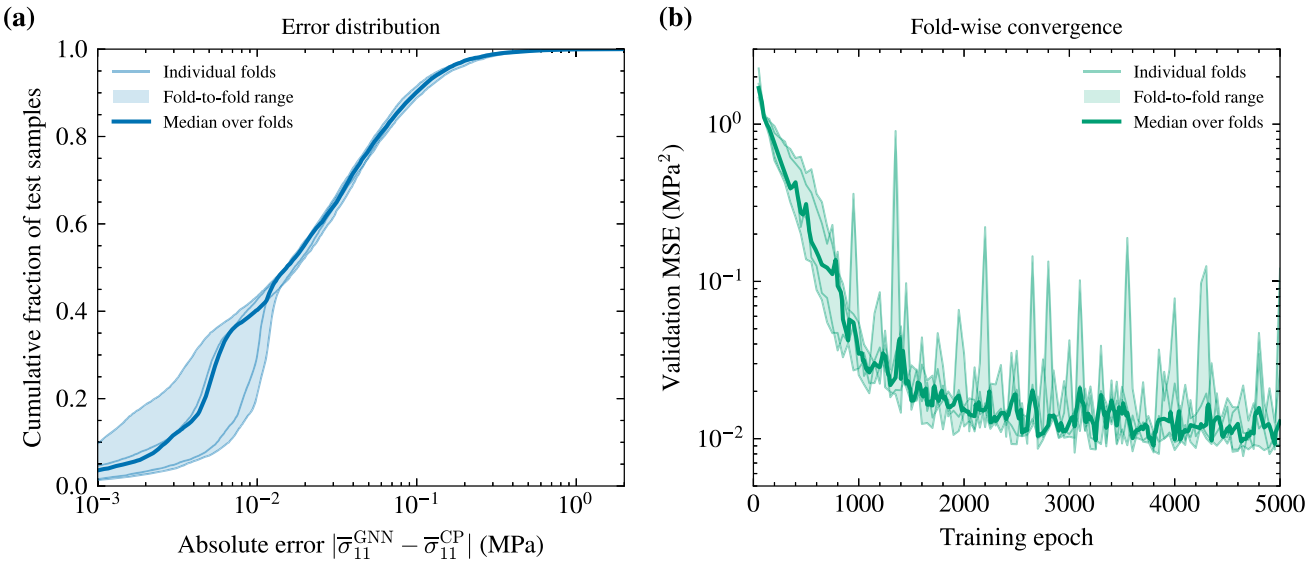

Fig. 15 Reliability of the $\overline{\sigma}_{11}$ model under 5-fold cross-validation. (**a**) Cumulative distribution of the absolute prediction error on the held-out partition and (**b**) validation loss against training epoch, each shown per fold (thin lines) with the median over the folds (thick line) and the fold-to-fold range (shaded). The folds agree across the full error range and not merely in the aggregate statistics, and descend along a common trajectory to comparable validation loss

against a stress range of approximately 62 MPa in the test partition, and the reported production model ($R^2 = 0.99996$, RMSE $= 0.101$ MPa) lies at the conservative edge of this distribution.

Appendix C: Effect of the training-set size

Generating the dataset required approximately 11,100 hours of CP simulation time, so the volume of reference data needed to reach a given accuracy is of direct practical interest. A learning curve was constructed for the $\overline{\sigma}_{11}$ component by retraining the identical architecture, hyperparameters and 5,000 epoch budget on strictly nested subsets of the training partition, ranging from 25 to 100 percent of the available training samples. The held out validation and test partitions were kept fixed and identical for every subset. Figure 16 shows the resulting test error as a function of the fraction of training samples used. Both axes are relative, the ordinate carrying the RMSE of each model divided by that of the model trained on the full partition, so that the curve reads as a convergence study.

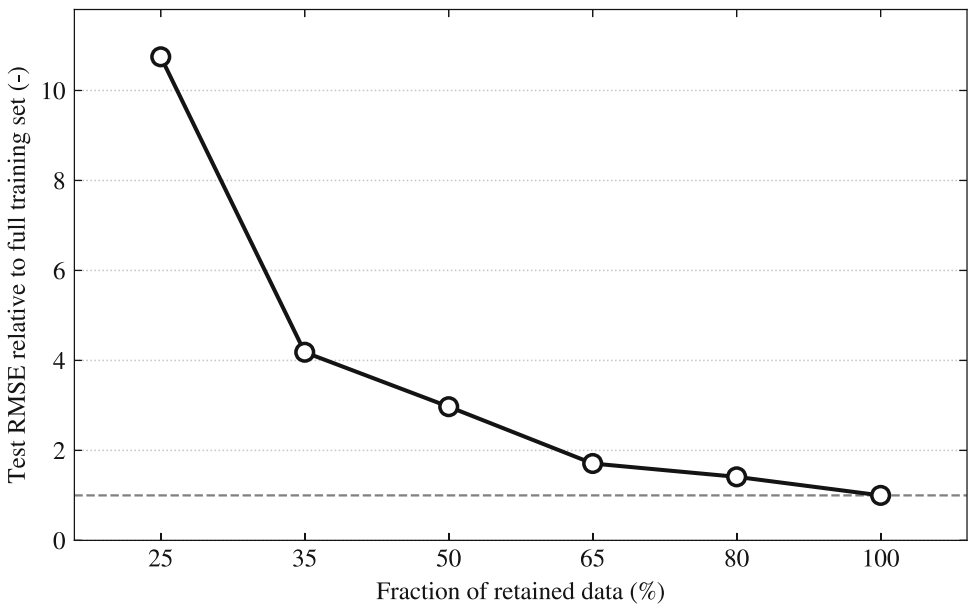

Fig. 16 Learning curve of the $\overline{\sigma}_{11}$ model evaluated on a fixed held out test partition. Each marker represents an independently trained model. The abscissa gives the fraction of retained training data and the ordinate the RMSE of each model divided by that of the model trained on the full partition, the dashed line at unity marking what the present data budget yields. The error falls steeply at small training set sizes and the increments diminish thereafter, without the curve flattening, in that the final third of the training data still removes 41 % of the error remaining at 65 %

Across the range examined, the RMSE decreases by approximately one order of magnitude, with the model trained on one quarter of the data exhibiting nearly eleven times the error of the model trained on the full partition. The reduction is strongly front loaded, with a relative test RMSE of approximately three already reached using half of the available training data. Beyond this point, the curve exhibits clear diminishing returns. Although the final third of the training samples still reduces the remaining error by 41 %, the absolute improvement is small relative to the accuracy already achieved. The present dataset is therefore sufficient to support the reported surrogate accuracy and the intended microstructure screening application.

Acknowledgements This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 101001567).

Author Contributions Ali R. Safi: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing - original draft. Pedro Campos: Data curation, Formal analysis, Methodology, Software, Writing - review & editing. Frederic E. Bock: Formal analysis, Supervision, Writing - review & editing. Benjamin Klusemann: Funding acquisition, Project administration, Supervision, Writing - review & editing.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability The obtained data from this research is online available at Zenodo (https://doi.org/10.5281/zenodo.19220922).

Code Availability The code used in this research is online available at Zenodo (https://doi.org/10.5281/zenodo.19220922).

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Motaman SAH, Haase C. The microstructural effects on the mechanical response of polycrystals: A comparative experimental-numerical study on conventionally and additively manufactured metallic materials. Int J Plast. 2021;140:102941. https://doi.org/10.1016/j.ijplas.2021.102941.
2. Hu F, Niezgoda S, Xue T, Cao J. Efficient GPU-computing simulation platform JAX-CPFEM for differentiable crystal plasticity finite element method. NPJ Comput Mater. 2025;11(1):46. https://doi.org/10.1038/s41524-025-01528-2.
3. Lebensohn RA, Tomé CN. A self-consistent anisotropic approach for the simulation of plastic deformation and texture development of polycrystals: application to zirconium alloys. Acta Metall Mater. 1993;41(9):2611–24.
4. Segurado J, Lebensohn RA, LLorca J. Computational homogenization of polycrystals. Adv Appl Mech. 2018;51:1–114.
5. Chinesta F, Huerta A, Rozza G, Willcox K. Model reduction methods. In: Stein E, Borst R, Hughes TJR (Eds.), Encyclopedia of Computational Mechanics Second Edition, 1st Edition, Wiley; 2017. p. 1–36. https://doi.org/10.1002/9781119176817.ecm2110
6. Gierden C, Kochmann J, Waimann J, Svendsen B, Reese S. A review of FE-FFT-based two-scale methods for computational modeling of microstructure evolution and macroscopic material behavior. Arch Comput Methods Eng. 2022;29(6):4115–35. https://doi.org/10.1007/s11831-022-09735-6.
7. Bock FE, Aydin RC, Cyron CJ, Huber N, Kalidindi SR, Klusemann B. A review of the application of machine learning and data mining approaches in continuum materials mechanics. Frontiers in Materials. 2019;6. https://doi.org/10.3389/fmats.2019.00110.
8. Muhammad W, Brahme AP, Ibragimova O, Kang J, Inal K. A machine learning framework to predict local strain distribution and the evolution of plastic anisotropy & fracture in additively manufactured alloys. Int J Plast. 2021;136:102867. https://doi.org/10.1016/j.ijplas.2020.102867.
9. Bonatti C, Berisha B, Mohr D. From CP-FFT to CP-RNN: Recurrent neural network surrogate model of crystal plasticity. Int J Plast. 2022;158:103430. https://doi.org/10.1016/j.ijplas.2022.103430. https://www.sciencedirect.com/science/article/pii/S074964192200208X
10. Mianroodi JR, Siboni NH, Raabe D. Teaching solid mechanics to artificial intelligence–a fast solver for heterogeneous materials. NPJ Comput Mater. 2021;7(1):1–10. https://doi.org/10.1038/s41524-021-00571-z.
11. Khorrami MS, Mianroodi JR, Siboni NH, Goyal P, Svendsen B, Benner P, et al. An artificial neural network for surrogate modeling of stress fields in viscoplastic polycrystalline materials. NPJ Comput Mater. 2023;9(1):1–10. https://doi.org/10.1038/s41524-023-00991-z.
12. Fuhg JN, Anantha Padmanabha G, Bouklas N, Bahmani B, Sun W, Vlassis NN, et al. A review on data-driven constitutive laws for solids. Arch Comput Methods Eng. 2025;32(3):1841–83. https://doi.org/10.1007/s11831-024-10196-2.
13. Khosravani A, Cecen A, Kalidindi SR. Development of high throughput assays for establishing process-structure-property linkages in multiphase polycrystalline metals: Application to dual-phase steels. Acta Mater. 2017;123:55–69. https://doi.org/10.1016/j.actamat.2016.10.033. https://linkinghub.elsevier.com/retrieve/pii/S135964541630800X
14. Li X, Zhang Y, Zhao H, Burkhart C, Brinson LC, Chen W. A transfer learning approach for microstructure reconstruction and structure-property predictions. Sci Rep. 2018;8(1):13461. https://doi.org/10.1038/s41598-018-31571-7x.
15. Cang R, Li H, Yao H, Jiao Y, Ren Y. Improving direct physical properties prediction of heterogeneous materials from imaging data via convolutional neural network and a morphology-aware generative model. Comput Mater Sci. 2018;150:212–21. https://doi.org/10.1016/j.commatsci.2018.03.074.https://www.sciencedirect.com/science/article/pii/S0927025618302337.
16. Latypov MI, Kühbach M, Beyerlein IJ, Stinville J-C, Toth LS, Pollock TM, et al. Application of chord length distributions and principal component analysis for quantification and representation of diverse polycrystalline microstructures. Mater

Charact. 2018;145:671–85. https://doi.org/10.1016/j.matchar.2018.09.020.https://www.sciencedirect.com/science/article/pii/S1044580318313743.
17. Herriott C, Spear AD. Predicting microstructure-dependent mechanical properties in additively manufactured metals with machine- and deep-learning methods. Comput Mater Sci. 2020;175:109599. https://doi.org/10.1016/j.commatsci.2020.109599.https://linkinghub.elsevier.com/retrieve/pii/S0927025620300902.
18. Cecen A, Dai H, Yabansu YC, Kalidindi SR, Song L. Material structure-property linkages using three-dimensional convolutional neural networks. Acta Mater. 2018;146:76–84. https://doi.org/10.1016/j.actamat.2017.11.053.https://linkinghub.elsevier.com/retrieve/pii/S1359645417310443.
19. Ibragimova O, Brahme A, Muhammad W, Connolly D, Lévesque J, Inal K. A convolutional neural network based crystal plasticity finite element framework to predict localised deformation in metals. Int J Plast. 2022;157:103374. https://doi.org/10.1016/j.ijplas.2022.103374.https://linkinghub.elsevier.com/retrieve/pii/S0749641922001541.
20. Zhou J, Cui G, Hu S, Zhang Z, Yang C, Liu Z, et al. Graph neural networks: A review of methods and applications. AI Open. 2020;1:57–81. https://doi.org/10.1016/j.aiopen.2021.01.001.https://linkinghub.elsevier.com/retrieve/pii/S2666651021000012.
21. Wu L, Chen Y, Shen K, Guo X, Gao H, Li S, Pei J, Long B. Graph neural networks for natural language processing: A survey, version Number: 2. 2021. https://doi.org/10.48550/ARXIV.2106.06090
22. Coley C, Jin W, Rogers L, Jamison TF, Jaakkola TS, Green WH, et al. A graph-convolutional neural network model for the prediction of chemical reactivity. Chem Sci. 2019;10(2):370–7. https://doi.org/10.1039/C8SC04228D.
23. Chung J, Ahmad R, Sun W, Cai W, Mukerji T. Prediction of effective elastic moduli of rocks using Graph Neural Networks. Comput Methods Appl Mech Eng. 2024;421:116780. https://doi.org/10.1016/j.cma.2024.116780.https://www.sciencedirect.com/science/article/pii/S0045782524000367.
24. Vlassis NN, Sun W. Geometric learning for computational mechanics Part II: Graph embedding for interpretable multiscale plasticity. Comput Methods Appl Mech Eng. 2023;404:115768. https://doi.org/10.1016/j.cma.2022.115768.https://www.sciencedirect.com/science/article/pii/S0045782522007241.
25. Pagan DC, Pash CR, Benson AR, Kasemer MP. Graph neural network modeling of grain-scale anisotropic elastic behavior using simulated and measured microscale data. NPJ Comput Mater. 2022;8(1):259. https://doi.org/10.1038/s41524-022-00952-y.
26. Hu G, Latypov MI. AnisoGNN: Graph neural networks generalizing to anisotropic properties of polycrystals. Comput Mater Sci. 2024;243:113121. https://doi.org/10.1016/j.commatsci.2024.113121.https://www.sciencedirect.com/science/article/pii/S0927025624003422.
27. Hestroffer JM, Charpagne M-A, Latypov MI, Beyerlein IJ. Graph neural networks for efficient learning of mechanical properties of polycrystals. Comput Mater Sci. 2023;217:111894. https://doi.org/10.1016/j.commatsci.2022.111894.https://www.sciencedirect.com/science/article/pii/S092702562200605X.
28. Hu Y, Zhou G, Lee M-G, Wu P, Li D. A temporal graph neural network for cross-scale modelling of polycrystals considering microstructure interaction. Int J Plast. 2024;179:104017. https://doi.org/10.1016/j.ijplas.2024.104017. https://www.sciencedirect.com/science/article/pii/S074964192400144X.
29. Heidenreich JN, Mohr D. Extended minimal state cells (EMSC): Self-consistent recurrent neural networks for rate- and temperature dependent plasticity. Int J Plast. 2025;188:104305. https://doi.org/10.1016/j.ijplas.2025.104305. https://www.sciencedirect.com/science/article/pii/S0749641925000646.
30. Sim G-J, Lee M-G, Latypov MI. FIP-GNN: Graph neural networks for scalable prediction of grain-level fatigue indicator parameters. Scripta Mater. 2025;255:116407. https://doi.org/10.1016/j.scriptamat.2024.116407.https://www.sciencedirect.com/science/article/pii/S1359646224004421.
31. Qin Y, DeWitt S, Radhakrishnan B, Biros G. GrainGNN: A dynamic graph neural network for predicting 3D grain microstructure. J Comput Phys. 2024;510:113061. https://doi.org/10.1016/j.jcp.2024.113061.https://www.sciencedirect.com/science/article/pii/S0021999124003103.
32. Bargmann S, Klusemann B, Markmann J, Schnabel JE, Schneider K, Soyarslan C, et al. Generation of 3D representative volume elements for heterogeneous materials: A review. Prog Mater Sci. 2018;96:322–84. https://doi.org/10.1016/j.pmatsci.2018.02.003.https://linkinghub.elsevier.com/retrieve/pii/S0079642518300161.
33. Paulson NH, Priddy MW, McDowell DL, Kalidindi SR. Reduced-order structure-property linkages for polycrystalline microstructures based on 2-point statistics. Acta Mater. 2017;129:428–38. https://doi.org/10.1016/j.actamat.2017.03.009.https://www.sciencedirect.com/science/article/pii/S135964541730188X.
34. Groeber MA, Jackson MA. DREAM.3D: A digital representation environment for the analysis of microstructure in 3D. Integr Mater Manuf Innov. 2014;3(1):56–72. https://doi.org/10.1186/2193-9772-3-5.
35. Roters F, Diehl M, Shanthraj P, Eisenlohr P, Reuber C, Wong SL, et al. DAMASK – The Düsseldorf Advanced Material Simulation Kit for modeling multi-physics crystal plasticity, thermal, and damage phenomena from the single crystal up to the component scale. Comput Mater Sci. 2019;158:420–78. https://doi.org/10.1016/j.commatsci.2018.04.030.https://www.sciencedirect.com/science/article/pii/S0927025618302714.
36. Rice JR. In elastic constitutive relations for solids: An internal-variable theory and its application to metal plasticity. J Mech Phys Solids. 1971;19(6):433–55. https://doi.org/10.1016/0022-5096(71)90010-X.
37. Hutchinson JW, Hill R. Bounds and self-consistent estimates for creep of polycrystalline materials. Proc R Soc Lond A Math Phys Sci. 1997;348(1652):101–27. https://doi.org/10.1098/rspa.1976.0027.
38. Kamm GN, Alers GA. Low-temperature elastic moduli of aluminum. J Appl Phys. 1964;35(2):327–30. https://doi.org/10.1063/1.1713309.
39. Hosford WF, Fleischer RL, Backofen WA. Tensile deformation of aluminum single crystals at low temperatures. Acta Metall. 1960;8(3):187–99. https://doi.org/10.1016/0001-6160(60)90127-9.
40. Kocks UF. The relation between polycrystal deformation and single-crystal deformation. Metallurg Trans. 1970;1(5):1121–43. https://doi.org/10.1007/BF02900224.
41. Kingma DP, Ba J. Adam: A method for stochastic optimization. 2017. arXiv:1412.6980 [cs].
42. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, ACM, Anchorage AK USA; 2019. p. 2623–31. https://doi.org/10.1145/3292500.3330701
43. BarredoArrieta A. N Díaz-Rodríguez, J DelSer, A Bennetot, S Tabik, A Barbado, S Garcia, S Gil-Lopez, D Molina, R Benjamins, R Chatila, F Herrera, Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fus. 2020;58:82–115. https://doi.org/10.1016/j.inffus.2019.12.012. https://www.sciencedirect.com/science/article/pii/S1566253519308103.

44. Sundararajan M, Taly A, Yan Q. Axiomatic attribution for deep networks. 2017. arXiv:1703.01365 [cs].
45. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems. 2017. vol. 30.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.