



# Stress predictions in polycrystal plasticity using graph neural networks with subgraph training

Hanfeng Zhai<sup>1</sup>

Received: 22 August 2024 / Accepted: 22 January 2025

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2025

## Abstract

Numerical modeling of polycrystal plasticity is computationally intensive. We employ Graph Neural Networks (GNN) to predict stresses on complex geometries for polycrystal plasticity from Finite Element Method (FEM) simulations. We present a novel message-passing GNN that encodes nodal strain and edge distances between FEM mesh cells, and aggregates to obtain embeddings and combines the decoded embeddings with the nodal strains to predict stress tensors on graph nodes. The GNN is trained on subgraphs generated from FEM mesh graphs, in which the mesh cells are converted to nodes and edges are created between adjacent cells. We apply the trained GNN to periodic polycrystals with complex geometries and learn the strain-stress maps based on crystal plasticity theory. The GNN is accurately trained on FEM graphs, in which the  $R^2$  for both training and testing sets are larger than 0.99. The proposed GNN approach speeds up more than 150 times compared with FEM on stress predictions. We also apply the trained GNN to unseen simulations for validations and the GNN generalizes well with an overall  $R^2$  of 0.992. The GNN accurately predicts the von Mises stress on polycrystals. The proposed model does not overfit and generalizes well beyond the training data, as the error distributions demonstrate. This work outlooks surrogating crystal plasticity simulations using graph data.

**Keywords** Polycrystal plasticity · Stress · Graph neural networks · Finite element method · Constitutive behavior

## Contents

1	Introduction . . . . .
2	Crystal plasticity . . . . .
2.1	Mechanistic model and problem formulation . . . . .
2.2	Material parameters and data generation . . . . .
3	Message-passing graph neural networks . . . . .
3.1	Message-passing on edges for nodal inference . . . . .
3.2	Model framework and training algorithm . . . . .
3.2.1	Subgraph sampling and training . . . . .
3.2.2	Training algorithm . . . . .
4	Results and discussions . . . . .
4.1	Training and testing results . . . . .
4.2	Analysis on finite element meshes . . . . .
4.3	Deployment on validation dataset . . . . .
4.4	Limitations of the proposed method . . . . .
5	Summary and conclusions . . . . .
Appendix A	Data preparation . . . . .
Appendix A.1	Graph conversion methods . . . . .
Appendix A.2	Finite element implementation of crystal plasticity . . . . .
Appendix B	Graph neural networks . . . . .

Appendix B.1	Training procedure . . . . .
Appendix B.2	Model characterization . . . . .
Appendix C	Additional results . . . . .
Appendix C.1	Predictions from GNN . . . . .
Appendix C.2	Discussions on loading-coupled directions . . . . .
Appendix C.3	Error distribution for stress components . . . . .
References	. . . . .

## List of symbols

$\bar{v}$	Imposed velocity vector on the surface
$F$	Deformation gradient tensor
$f$	Body force
$p$	Material parameters
$\oplus$	Aggregation operator
$x$	Current configuration
$\dot{\gamma}^\alpha$	Shearing rate of the $\alpha$ slip system
$\dot{\gamma}_0$	Fixed-state strain rate scaling coefficient
$\dot{g}^\alpha$	Strength of the $\alpha$ slip system
$\ell_{ij}$	Edge link length between mesh cells
$\epsilon$	Lagrangian strain tensor
$\epsilon^e$	Elastic strain
$\epsilon^p$	Plastic strain
$\hat{D}^p$	Plastic deformation rate
$\hat{L}^p$	Plastic velocity gradient

✉ Hanfeng Zhai  
hzhai@stanford.edu

<sup>1</sup> Department of Mechanical Engineering, Stanford University, Stanford, CA 94305, USA

$\hat{\mathbf{m}}^\alpha$	Normal to the slip plane for the $\alpha$ slip system
$\hat{\mathbf{p}}^\alpha$	Symmetric part of the Schmid tensor for the $\alpha$ slip system
$\hat{\mathbf{q}}^\alpha$	Skew part of the Schmid tensor for the $\alpha$ slip system
$\hat{\mathbf{s}}^\alpha$	Slip direction for the $\alpha$ slip system
$\hat{\mathbf{W}}^p$	Plastic spin
$\mathbb{C}$	Stiffness tensor
$M$	Number of edges in a graph
$N$	Number of nodes in a graph
$\mathbf{c}^{\text{ind}}$	Connection indices of the subgraph
$\mathbf{F}^e$	Elastic portion of deformation gradient tensor
$\mathbf{F}^p$	Plastic portion of deformation gradient tensor
$\mathbf{h}_{ij}$	Node information on edge $i-j$
$\mathbf{h}_i$	Node information on node $i$
$\mathbf{r}^*$	Lattice rotation
$\mathbf{v}$	Velocity vector of a point in the current configuration
$\mathbf{v}^e$	Elastic stretch
$\mathbf{X}$	Reference configuration
$\mathcal{G}$	Graph object
$\mathcal{L}$	Loss function for the optimization problem
$\mathcal{M}_{ij}$	Message information obtained on edges
$\mathcal{N}(i)$	Neighboring node list for node $i$
$\Phi$	MLP outside the message-passing layer on nodes
$\phi$	MLP within the message-passing layer
$\text{MLP}$	Multi-layer perceptron
$\text{MSG}$	Message function to pass message from nodes to edges
$\sigma_{\text{vM}}$	von Mises stress
$\tau$	Shear stress
$\Xi$	Output decoded information on nodes
$\xi_{\text{train}}$	Subgraph ratio of the full graph
$E$	Edges of graph
$g_0$	Initial slip system strength
$g_s$	Saturation strength
$h_0$	Strength hardening rate coefficient
$m$	Fixed-state strain rate sensitivity
$n$	Nonlinear Voce hardening exponent
$N_G$	Number of training graphs
$V$	Vertices of graph
$V_{\text{sub}}$	Vertices (nodes) of subgraph

## 1 Introduction

Plasticity refers to the permanent deformation of solid materials under external load, of which has been researched for more than 100 years. The earliest efforts include works of von Mises [1] and Huber [2] to phenomenologically capture yield criteria. The *flow* process describes the post-yielding behavior, in which dislocation plays a significant role. Accurate predictions of plastic deformation are crucial for various practical applications, such as optimizing metal forming processes [3], designing materials with specific properties, e.g.,

fatigue resistance [4]), controlling semiconductor interconnects [5], and controlling metal 3D printing processes [6]. These applications demand accurate and efficient digital twins characterizing crystal plasticity.

Due to decades of effort in understanding plasticity, constructing the constitutive model for polycrystals is still an active and ongoing research area due to (1) There are various ways to pose plasticity mechanisms characterizing the plastic deformation in continuum models such as temperature and rate dependence, anisotropy, etc. [7–9]; (2) By nature, plasticity is a multiscale problem, where numerous mechanisms contribute to the overall plastic behavior, such as single crystal dislocation [10], inter-grain friction [11], and grain boundary interactions [12], making it a challenging task to craft plasticity models integrate phenomena occurring at various scales; (3) The high computational expense associated with accurately simulating polycrystal plasticity using numerical methods such as finite elements [13–15]. The computational cost is mainly attributed to the path dependence and nonlinear nature of plasticity.

The recent developments of data-driven modeling for physical models could potentially task the high computational cost and surrogate plasticity models, considering their demonstrated success in fluid mechanics [16, 17], heat transfer [18, 19], and design optimization [20–22]. In the subgrain scale, mechanical responses can be predicted by combining machine learning and dislocation simulation data [23, 24]. It has also been shown convolutional neural networks (CNN), graph neural networks (GNN), and general regression methods (e.g., Ridge, Lasso) can be applied to molecular dynamics (MD) simulations to predict mechanical responses and corresponding material properties [25–27]. Based on grain representations, GNN has been used to predict the magnetic properties of polycrystalline graphs [28]. Using experimental data, mechanical responses can be predicted from 2D images of static structures and CNN [29] or graph convolutional networks [30]. Pagan et al. [31] demonstrated that GNN can learn the anisotropic elastic response of alloys. In the continuum scale, CNN has been used to predict stress-strain responses [32]. GNN has been used to learn mesh-based time-dependent PDEs [33]. Notably, Mozaffar et al. [34] demonstrate that recurrent neural networks can learn path-dependent plasticity, and Fuhg et al. [35] show that partially input convex neural networks can predict plane stress macroscopic yield as a function of crystallographic texture.

This paper aims to demonstrate the capability of GNN in capturing the mechanical responses of polycrystals in the plastic regime, with two key innovations: (1) Handling complex geometry in polycrystal plasticity, and (2) Leveraging subgraph training for more efficient learning. Using open-source polycrystal generation and finite element method (FEM) software, *Neper & FEPX* to generate meshes and conduct numerical simulations of periodic polycrystals [36,

[37], the goal is to develop accurate and generalizable surrogate plasticity models based on finite element calculations [38, 39]. Developing such surrogate models has three main challenges: (1) The generated finite element meshes have varying degrees of freedom (DoF) for different polycrystal geometries. Traditional regression tools, such as Gaussian processes or neural networks, typically rely on a fixed number of training points. (2) The spatial connectivity between finite element mesh cells preserves important physical features, i.e., physical properties are passed between adjacent mesh cells during the finite element calculations. Matrix-based data struggles to preserve such geometric relationships. (3) The data size is large; each 10-grain polycrystal mesh contains approximately 10,000 mesh cells, making the training a computationally expensive task.

To tackle the first problem, we propose using GNN to handle data with different DoFs. Since GNN can be trained on graphs with different numbers of nodes & edges, meshes with different sizes can be potentially handled. This also helps us solve the second problem since GNN can handle the connectivity within the data. The connections between mesh cells preserve the spatial feature of the polycrystals. We generate graphs in which cells are converted to nodes where the adjacent cells are connected. To tackle the third problem, we propose training the GNN on subgraphs of finite element meshes. In this paper, we hope to combine the proposed approaches and explore stress predictions on polycrystals using GNN.

The paper is arranged as follows: In Sect. 2 we present the formulation of the crystal plasticity model and the data generation process. In Sect. 3 we present the mathematical model and training details of the message-passing GNNs, explaining the details of subgraph sampling and training, with additional explanations of the mesh-graph data conversion process. In Sect. 4 we present the results of the predictions on training and testing sets, analysis of comparing GNN with FEM, and further deployment on unseen datasets as validation. We briefly conclude the paper in Sect. 5.

## 2 Crystal plasticity

### 2.1 Mechanistic model and problem formulation

Crystal plasticity models are employed [40], in which we use the general theory following Han et al. and the FEM implementation in *FEPX* [37, 41]. We begin with the deformation gradient tensor, defined as  $\mathbf{F} = \frac{\partial \mathbf{x}}{\partial \mathbf{X}}$ , can be decomposed into elastic and plastic parts:

$$\mathbf{F} = \mathbf{F}^e \mathbf{F}^p = \mathbf{v}^e \mathbf{r}^* \mathbf{F}^p$$

where the elastic gradient tensor can be decomposed to lattice rotation  $\mathbf{r}^*$  and elastic stretch  $\mathbf{v}^e$ .  $\mathbf{F}^p$  pertains plastic slip.  $\mathbf{x}$  is the current configuration and  $\mathbf{X}$  is the reference configuration. The general schematic of the theory is illustrated in Fig. 1.

The polycrystal will generate a stress field distribution  $\sigma$ . Under loading, the local form of the equilibrium equation writes:

$$\nabla \cdot \sigma + \mathbf{f} = 0$$

where  $\sigma$  is the Cauchy stress (or simply termed “stress”).  $\mathbf{f}$  is the body force vector, in our implementation  $\mathbf{f} = 0$ . The relationship between the Cauchy stress and the shear stress writes:

$$\tau = (\det(\mathbf{v}^e)) \sigma$$

For elastic deformations, the stress-strain relationship can be expressed as the generalized Hooke’s law, which can be written as the

$$\sigma = \mathbb{C} \epsilon^e \quad (1)$$

where  $\mathbb{C}$  is the elastic moduli tensor (or stiffness tensor).  $\mathbb{C} = [\mathcal{C}_{ij}]$  contains elastic constants to be specified in the simulation.

After yield, the stress contributes to plastic flow, which can be described by restricted slip. Here,  $\hat{\mathbf{L}}^p$  is the plastic velocity gradient, which can be written in terms of the plastic slip:

$$\hat{\mathbf{L}}^p = (\dot{\mathbf{F}}^p) (\mathbf{F}^p)^{-1} \quad (2)$$

The Lagrangian strain tensor contains both the elastic and plastic contributions and can be expressed in terms of elastic and plastic strain tensors:

$$\begin{aligned} \epsilon &= \frac{1}{2} \left( \mathbf{F}^{e\top} \mathbf{F}^e - (\mathbf{F}^p)^{-\top} (\mathbf{F}^p)^{-1} \right) \\ &= \epsilon^e + \epsilon^p \end{aligned}$$

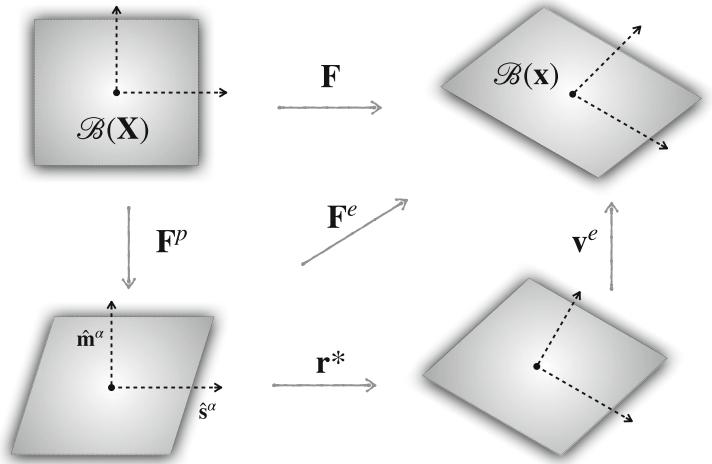
Using Schmid tensor’s symmetric and skew part, the plastic deformation gradient in Eq. (2) can be written in terms of slip using Schmid tensor’s symmetric and skew parts:

$$\hat{\mathbf{L}}^p = \hat{\mathbf{D}}^{p'} + \hat{\mathbf{W}}^p$$

where

$$\hat{\mathbf{D}}^{p'} = \sum_{\alpha} \dot{\gamma}^{\alpha} \hat{\mathbf{p}}^{\alpha}, \quad \text{and} \quad \hat{\mathbf{W}}^p = (\dot{\mathbf{r}}^*) (\mathbf{r}^*)^{\top} + \sum_{\alpha} \dot{\gamma}^{\alpha} \hat{\mathbf{q}}^{\alpha} \quad (3)$$

**Fig. 1** Schematic diagram for the decomposition in different configurations in crystal plasticity formulation. The visualization is inspired by Refs. [37, 41]



Here,  $\hat{\mathbf{p}}^\alpha$  and  $\hat{\mathbf{q}}^\alpha$  are defined as

$$\begin{aligned}\hat{\mathbf{p}}^\alpha &= \hat{\mathbf{p}}^\alpha(\mathbf{q}) = \text{sym}(\hat{\mathbf{s}}^\alpha \otimes \hat{\mathbf{m}}^\alpha) \\ \hat{\mathbf{q}}^\alpha &= \hat{\mathbf{q}}^\alpha(\mathbf{q}) = \text{skw}(\hat{\mathbf{s}}^\alpha \otimes \hat{\mathbf{m}}^\alpha)\end{aligned}\quad (4)$$

where  $\hat{\mathbf{s}}^\alpha$  and  $\hat{\mathbf{m}}^\alpha$  are the slip directions obtained after the kinetic decomposition visualized in Fig. 1. Note that the symmetric and skew parts are expressed as:

$$\begin{aligned}\text{sym}(\cdot) &= \frac{1}{2}[(\cdot) + (\cdot)^\top] \\ \text{skw}(\cdot) &= \frac{1}{2}[(\cdot) - (\cdot)^\top]\end{aligned}$$

$\dot{\gamma}^\alpha$  is the slip system shearing rate. Here, the shearing rate relates to the resolved shear stress  $\tau^\alpha$  via an assumed power law relationship:

$$\dot{\gamma}^\alpha = \dot{\gamma}_0 \left( \frac{|\tau^\alpha|}{g^\alpha} \right)^{\frac{1}{m}} \text{sgn}(\tau^\alpha) \quad (5)$$

where  $\dot{\gamma}_0$  is the fixed-rate strain rate scaling coefficient,  $m$  is the rate sensitivity exponent. The resolved shear stress  $\tau^\alpha$  is the projection of the crystal stress tensor onto the slip plane (in that particular slip direction) obtained via the Schmid tensor's symmetric part (Eq. (4)):

$$\tau^\alpha = \text{tr}(\hat{\mathbf{p}}^\alpha \tau')$$

The evolution of slip system strength  $g^\alpha$  can be characterized by hardening modulus  $h_0$  and the initial strengths following a power law:

$$\dot{g}^\alpha = h_0 \left( \frac{g_s(\dot{\gamma}) - g^\alpha}{g_s(\dot{\gamma}) - g_0} \right)^n \dot{\gamma}$$

where  $n$  is the nonlinear Voce hardening exponent.  $g_s(\dot{\gamma})$  is the initial slip system saturation strength.  $g_0$  is the initial slip

system strength.  $\dot{\gamma}$  is calculated as the summation of the slip shearing rates, related to resolved shear stresses (Eq. (5)):

$$\dot{\gamma} = \sum_\alpha |\dot{\gamma}^\alpha|$$

For the FEM implementations of this method, some numerical details are summarized in [Appendix A.2](#).

The boundary conditions (B.C.s) can be specified via

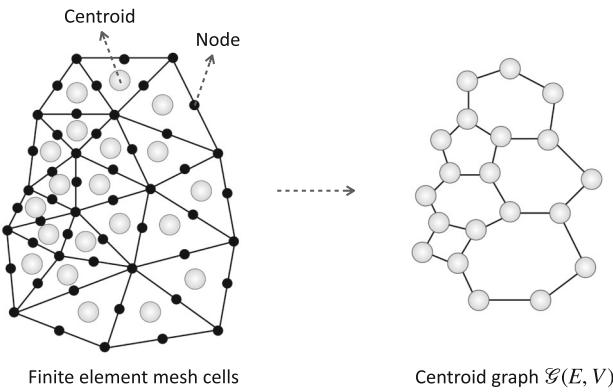
$$\mathbf{v}(\mathbf{x}) = \bar{\mathbf{v}}$$

as the velocity B.C.s. In our implementation in *FEPX* [37], we apply fixed strain rate in  $x$ -direction,  $\dot{\epsilon}_{xx} = 10^{-3} \text{ s}^{-1}$ , which only acts on the  $v_x$  components. The applied strain rates in other directions are all set to be zero.

## 2.2 Material parameters and data generation

The rate sensitivity exponent in Eq. (5) is set to be  $m = 0.02$ . We employ an isotropic hardening type. The fixed-rate strain rate is  $\dot{\gamma}_0 = 1$ . The simulation targets a total strain of  $\epsilon_{xx} = 0.01$ , where the strain increment per step is 0.001. We generate 90 10-grain periodic polycrystals, in which the mesh is generated via *Neper* [36]. We used body-centered cubic (BCC) crystals with elastic constants  $C_{11} = 236.9$  [GPa],  $C_{12} = 140.6$  [GPa], and  $C_{44} = 116.0$  [GPa]. The hardening modulus  $h_0 = 391.90$  [MPa] and the slip strengths are  $g_0 = 200$  &  $g_s = 335$  [MPa], respectively. The nonlinear Voce hardening exponent is  $n = 1$ . The 90 simulation results are then converted to graphs, of which 80% (72 graphs) are selected for the training, and the remaining (18 graphs) are considered as the testing sets. See Refs. [37, 42] for details and related FEM implementation.

In our formulation, we hypothesize that the finite element meshes can be formulated as a graph  $\mathcal{G} = \mathcal{G}(V, E)$ .  $V \in \mathbb{R}^N$



**Fig. 2** Schematic illustration of the procedures for converting FEM meshes to graphs. FEM element cell centroids are treated as nodes, and neighboring cells (sharing 3 common nodes for ‘tetra10’ elements) share an edge

&  $E \in \mathbb{R}^M$  are vertices and edges of the graph,<sup>1</sup> where  $V = V(\epsilon_i \mapsto \sigma_i)$  are the node features (on the finite element mesh node  $i$ ) and  $E = E(\ell_{ij})$  (Euclidean distances of mesh cells, on edge  $i-j$  that connects nodes  $i$  &  $j$ ).  $M$  &  $N$  are the number of edges and nodes. Each cell of the finite element mesh is considered a node. The edges are constructed according to the connectivity of the nodes. To enhance training efficiency, the strain data is rescaled by multiplying  $10^4$ , and the Euclidean norms are rescaled by multiplying  $10^3$ , making all the feed in-out data have a similar scale with the stress data ( $\sim 10^3$ ). Figure 2 indicates how the finite element meshes are converted to graph data for training. For the created tetra10 mesh for polycrystals, the centroids were converted to graph nodes (or vertices). For two adjacent mesh cells sharing 3 common nodes (or one mesh edge), an edge is being created on the graph.<sup>2</sup> Similar approaches are also employed for multiscale plasticity and topology optimization [43, 44]. Some discussions on this graph conversion method are provided in Appendix A.1.

Here, we aim to learn the nodal map from total strain to stress, i.e.,  $\mathfrak{M} : \epsilon \in \mathbb{R}^6 \mapsto \sigma \in \mathbb{R}^6$ . We want to use the GNN to surrogate the model  $\mathfrak{M}$ . By converting the mesh to graphs, one can construct mapping for polycrystals of irregular complex geometries since the learning is independent of the dimensions of the data. The overall strain-stress map for the physics-based model can be simplified in a form:

$$\sigma_i^{FEM} \equiv \sigma_i(\mathbf{x}) = \mathfrak{M}([\epsilon_i(\mathbf{x}), \mathbf{F}]; \mathbf{p}) \quad (6)$$

where  $\mathbf{p} = (\mathbb{C}, g_0, g_s, h_0, m, n, \dots)$  subsumes all the related material parameters used in the simulation. The model  $\mathfrak{M}(\cdot)$  takes strain  $\epsilon_i$  and the configurational map  $\mathbf{F}$ , and  $\mathbf{p}$  as input

<sup>1</sup> We are using the terms *vertex* and *node* interchangeably in this manuscript.

<sup>2</sup> Note that this also respects the conformality in finite element mesh.

and predict stress  $\sigma_i$  according to the equations presented above.

In Eq. (6),  $\sigma_i$  contains six stress elements are defined as  $\{\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6\} \equiv \{\sigma_{11}, \sigma_{12}, \sigma_{13}, \sigma_{22}, \sigma_{23}, \sigma_{33}\} \in \mathbb{R}^6$  (same for the strain components) for the overall stress & strain tensor elements in the FEM implementation. Note that  $\{x, y, z\}$  and  $\{1, 2, 3\}$  are used interchangeably in the subscripts.

### 3 Message-passing graph neural networks

#### 3.1 Message-passing on edges for nodal inference

Message-passing GNN learns the data relationship on graphs by passing the message from edges to nodes (vertex) and conducting nonlinear regression (using multi-layer perceptron, MLP) in the feature space. One begins with preparing the “messages” on edges, where the accumulated nodal message  $\tilde{\mathcal{M}}_{ij}^\epsilon$  (on edge) and edge message  $\tilde{\mathcal{M}}_{ij}^\ell$  for edge  $i-j$  can be written as:

$$\tilde{\mathcal{M}}_{ij}^\epsilon = \text{MSG}^{(n)}(\{\epsilon_i^{in}, \epsilon_j^{in}\}), \quad \tilde{\mathcal{M}}_{ij}^\ell = \text{MSG}^{(e)}(\{\ell_{ij}\}) \quad (7)$$

where  $\tilde{\mathcal{M}}_{ij}^\epsilon \in \mathbb{R}^{M \times 6}$  (passing the information of strains  $\epsilon$ ) and  $\tilde{\mathcal{M}}_{ij}^\ell \in \mathbb{R}^{M \times 1}$  (information of Euclidean distances  $\ell$ ). Here, 6 and 1 are the feature space dimensions for nodes and edges.  $\epsilon_i^{in}$  and  $\epsilon_j^{in}$  are the nodal strains (input property) for nodes  $i$  &  $j$ ; and  $\ell_{ij}$  are the edge input property, i.e., the mesh link length of edge  $i-j$ .

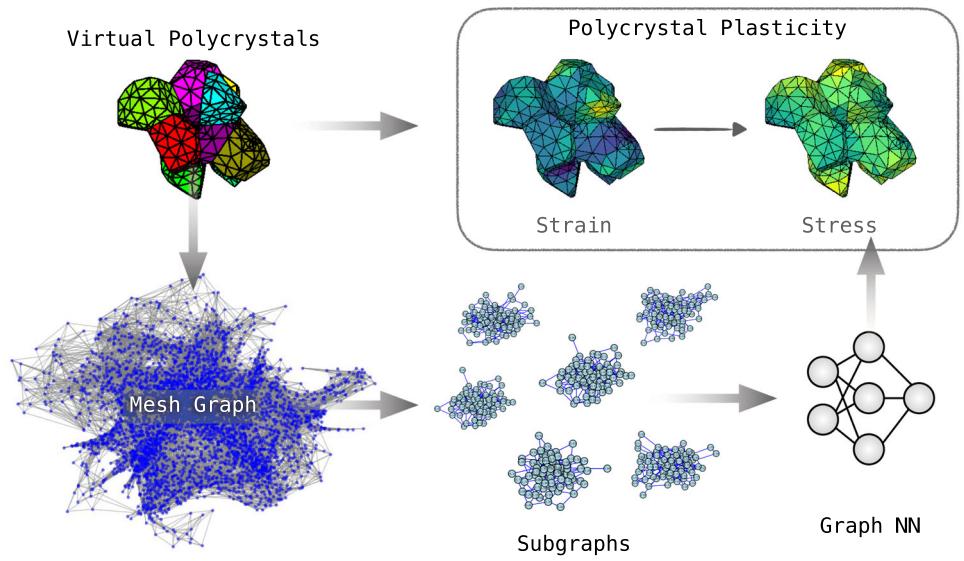
The nodal end edge messages are passed to the embedding dimension via two separate MLPs and output  $\mathbf{h}_{ij}^{(n)}$  and  $\mathbf{h}_{ij}^{(e)}$  with dimension  $\mathbb{R}^{M \times \text{emb}}$ .  $\oplus_j$  is the aggregation operator that sums over the neighboring node and edge information for node  $i$ . The predicted embedding output is then concatenated into  $\tilde{\mathbf{h}}_{ij}$ , with dimension  $\mathbb{R}^{M \times (2\text{emb})}$ :

$$\begin{aligned} \mathbf{h}_{ij}^{(n)} &= \text{MLP}^{(N-ENC)}(\tilde{\mathcal{M}}_{ij}^\epsilon), & \mathbf{h}_{ij}^{(e)} &= \text{MLP}^{(E-ENC)}(\tilde{\mathcal{M}}_{ij}^\ell), \\ \tilde{\mathbf{h}}_i^{(n)} &= \bigoplus_{j \in \mathcal{N}(i)} \left( \left\{ \mathbf{h}_{ij}^{(n)} \right\} \right), & \tilde{\mathbf{h}}_i^{(e)} &= \bigoplus_{j \in \mathcal{N}(i)} \left( \left\{ \mathbf{h}_{ij}^{(e)} \right\} \right) \end{aligned} \quad (8)$$

where  $\mathcal{N}(i)$  stands for the neighboring node list for node  $i$ . To pass the prediction on nodes, aggregation is being conducted for the decoded information in the message-passing layer. The output is then concatenated and fed into the decoding MLP:

$$\Xi_i = \text{MLP}^{(\text{DEC})} \left( \left\{ \tilde{\mathbf{h}}_i^{(n)}, \tilde{\mathbf{h}}_i^{(e)} \right\} \right)$$

**Fig. 3** General schematic of the workflow for using GNN to learn polycrystal plasticity. Virtual polycrystals are generated using *FEPX*. It is converted to mesh graphs based on finite element cells. The subgraphs are extracted to train the GNN. The GNN is then deployed to surrogate polycrystal plasticity simulations



This process is the *message-passing* from edges to nodes. This aggregation is being done in the embedding dimension and the output  $\Xi_i \in \mathbb{R}^{N \times 1}$  is the properties on the node. The prediction follows an “equation-MLP” using the given nodal information and decoded edge information (on the node):

$$\tilde{\sigma}_i = \text{MLP}^{(\text{EQN})} \left( \left\{ \epsilon_i^{in}, \Xi_i \right\} \right) \quad (9)$$

where  $\tilde{\sigma}_i \in \mathbb{R}^{N \times 6}$  are the final stress predictions for the supervised learning target, which is the optimization goal  $\tilde{\sigma}_i \sim \sigma_i$ .

The training uses mean-squared error (MSE) as the objective  $\mathcal{L}$  parameterized by trainable variables  $\Theta$  for supervised training.  $\dim()$  denotes the dimension of the given data. The optimization problem writes:

$$\arg \min_{\Theta} \mathcal{L}(\Theta), \quad (10)$$

$$\mathcal{L} = \frac{1}{\dim(\sigma)} \sum_{i \in \dim(\sigma)} (\tilde{\sigma}_i - \sigma_i)^2$$

where MSE is being calculated on all the nodes on the graph,  $\dim(\sigma) = N$ . For a graph,  $\dim()$  indicates the number of nodes.

To summarize, from Eqs. (7)~(9), the overall model can be simplified as a surrogate model for the  $\epsilon \mapsto \sigma$  map:

$$\sigma_i^{GNN} \equiv \tilde{\sigma}_i = \Phi \left( \epsilon_i^{in}, \bigoplus_{j \in \mathcal{N}(i)} \phi \left( \left\{ \epsilon_i^{in}, \epsilon_j^{in} \right\}, \ell_{ij} \right) \right) \quad (11)$$

where  $\Phi$  and  $\phi$  represent (combination of) different MLPs. This formula follows the generalized formula for message-passing GNN. The prediction is estimated based on the comparison of stresses predicted by GNN and FEM denoted

in Eqs. (6) and (11). The performance of the model is evaluated using the coefficient of determination (denoted as  $R^2$ ), which quantifies how well the predicted stress values from the GNN model match the true FEM values. The  $R^2$  score is calculated as:

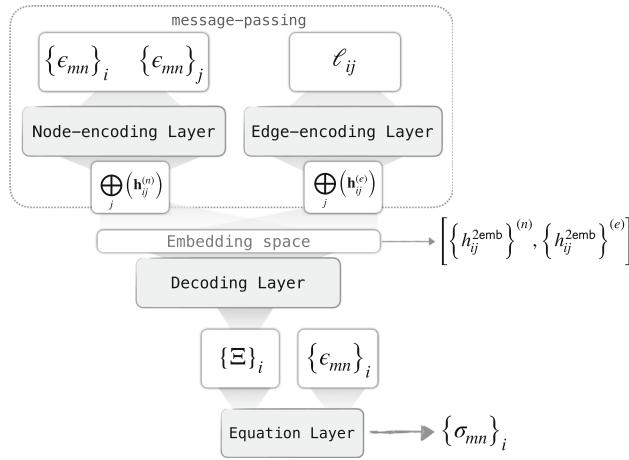
$$R^2 = 1 - \frac{\sum_{i=1}^N (\sigma_i^{FEM} - \sigma_i^{GNN})^2}{\sum_{i=1}^N (\sigma_i^{FEM} - \bar{\sigma}^{FEM})^2}$$

where  $\sigma_i^{FEM}$  are the stress values predicted by the FEM model,  $\sigma_i^{GNN}$  are the stress values predicted by the GNN model,  $\bar{\sigma}^{FEM}$  is the mean value of the true FEM stress values, and  $N$  is the number of data points. The  $R^2$  value typically ranges from 0 to 1, with a value of 1 indicating perfect agreement between the predicted and true values, and a value closer to 0 indicating poor prediction performance.

### 3.2 Model framework and training algorithm

Figure 4 illustrates the general architecture of our message-passing GNN. The node and edge-encoding layer takes in the nodal and edge properties on edge  $i-j$ , and output  $\mathbf{h}_{ij}^{(n)}$  and  $\mathbf{h}_{ij}^{(e)}$ , which are then being operated by the aggregation operator to pass the properties from edges to nodes to obtain  $\tilde{\mathbf{h}}_i^{(n)}$  and  $\tilde{\mathbf{h}}_i^{(e)}$  in the embedding dimension (Eq. (8)). The concatenated output  $\left\{ \tilde{\mathbf{h}}_i^{(n)}, \tilde{\mathbf{h}}_i^{(e)} \right\} (\in \mathbb{R}^{N \times 2\text{emb}})$  is then sent to the decoding layer to  $\Xi_i \in \mathbb{R}^{N \times 1}$ .  $\Xi_i$  and node input property  $\epsilon_i$  are then being concatenated and input to the equation layer to give the prediction that aims to approximate  $\sigma_i$  (Eq. (9)). The subscripts  $(mn)$  denote the elements in the stress & strain tensors, and  $(ij)$  denotes the connection of nodes in graphs.

There is 1 hidden layer for the *message-passing MLP* ( $\phi$ ) and *equation MLP* ( $\Phi$ ) respectively. The hidden dimension



**Fig. 4** The general architecture for the GNN. The node-encoding layer takes the strains on neighboring nodes for the input edges, and the edge-encoding layer takes the mesh cell link length (i.e. Euclidean norm of mesh cells). The combined outputs are then fed input to the embedding space ( $\mathbb{R}^{2\text{emb}}$ ).  $\{h_{ij}^{\text{emb}}\}^{(n)}$  and  $\{h_{ij}^{\text{emb}}\}^{(e)}$  are the decoded nodal and edge information in the embedding space. The output data is then fed input to the decoding layer that maps  $\mathbb{R}^{2\text{emb}}$  to  $\mathbb{R}^4$ . The output of the decoding layer is then passed to the message-passing operator (i.e.  $\oplus$ ), where the decoded messages are passed on nodes. The edge information on nodes  $\{\epsilon_{mn}\}_i$  are then combined to put into the equation layer, to predict the corresponding stress components  $\{\sigma_{mn}\}_i$

of the MLP  $\text{emb} = 31$ . As noted, the input and output dimensions are 6, that is,  $\epsilon_i \mapsto \sigma_i, i = 1, 2, \dots, 6$ . ReLU activation function is used for  $\Phi$  and tanh activation function is used for  $\phi$ . Note that data is “activated” twice in  $\phi$  for node- and edge-encoding layers respectively. Since there is only one hidden layer, the simple and light GNN model also effectively prevents the issue of oversmoothing [45]. More details of the model can be seen in Appendix B.

### 3.2.1 Subgraph sampling and training

We propose training GNN on subgraphs to learn the mapping. Let  $\mathcal{G}_{\text{sub}} = \mathcal{G}_{\text{sub}}(V_{\text{sub}}, E_{\text{sub}})$  denote the subgraph extracted from the full graph  $\mathcal{G}$  with randomly selected nodes, where  $V_{\text{sub}} \subseteq V$  &  $E_{\text{sub}} \subseteq E$ . Within  $V_{\text{sub}}$ , let  $\hat{V}_{\text{sub}}$  be all the subgraph nodes that preserve full edges compared with  $\mathcal{G}$ .  $\hat{V}_{\text{sub}} \setminus V_{\text{sub}}$  are the nodes that lose edges during the subgraph extraction process. Each finite element mesh graph contains  $\sim 10^5$  nodes in our implementation. For effective and efficient training of GNN, we propose training GNN on the subgraph, in which only the  $\hat{V}_{\text{sub}}$  and the connected edges are considered in the loss calculation,  $\mathcal{L} = \text{MSE}(\mathcal{G}_{\text{sub}}(\hat{V}_{\text{sub}}))$ .

$\hat{V}_{\text{sub}}$  can be selected by comparing the number of edges per node of  $\mathcal{G}_{\text{sub}}$  and  $\mathcal{G}$  (based on the global node index). The filtered node indices (termed as *connection indices*) are denoted as  $\mathbf{c}^{\text{ind}}$ . One then uses  $\mathbf{c}^{\text{ind}}$  to select “active nodes” to train based on the loss function.

Figure 5 illustrates the details of the subgraph training method to capture mapping on nodes. Based on a full graph converted from FEM meshes (Fig. 2), one first extracts the subgraph from full graph,<sup>3</sup> and then filter out the active nodes according to the connection index  $\mathbf{c}^{\text{ind}}$ , exemplified as  $i, j, k$  in Fig. 5 to train according to the loss function. To help the GNN to be more comprehensively trained according to this method, we propose presample the subgraphs before training, as shown in Fig. 6. For a given full graph  $\mathcal{G}$  to be trained on  $N$  epochs, one sample of all the subgraphs in the training sets for the  $i$ -th epoch as  $\mathcal{G}_{\text{sub}}^{(i)}$ . Essentially, for a full graph  $\mathcal{G}$  being trained, the GNN is “seeing” new subgraphs for each epoch, in which it preserves the local feature (i.e., the constitutive map from strain to stress in our case) of the full graph.

### 3.2.2 Training algorithm

The training algorithm is shown in Algorithm 1. Finite element mesh-based graphs are stored in the form of Torch Geometric tensors, containing nodal ( $\epsilon$  &  $\sigma$ ) and edge properties ( $\ell$ ). Subgraphs are extracted based on the training ratio  $\xi_{\text{train}}$ , specifying the ratio of the number of nodes selected from the full graph  $\mathcal{G}$ . We use  $\xi_{\text{train}} = 0.5$  for our training.<sup>4</sup> We pre-sample a set of subgraphs in the training set for each epoch and prepare a list of sampled subgraphs  $\mathcal{B}_G$  for training implementation. Under each epoch, the unique subgraph sets per that epoch will be selected, in which the active nodes are selected based on  $\mathbf{c}^{\text{ind}}$  and fed into the loss function (Eq. (10)). Adam optimizer is selected for gradient-based optimization. The model is being trained on the 72 graphs for 1000 epochs. Accompanying our subgraph sampling and training method, we use a double loop structure to first loop the subgraph batch sampled per epoch and then loop over each subgraph to learn the local feature map (strain to stress) on the subgraphs. This hierarchical training enables the GNN to learn on the subgraph generated from different polycrystals at each model evaluation step, i.e., iteration.

## 4 Results and discussions

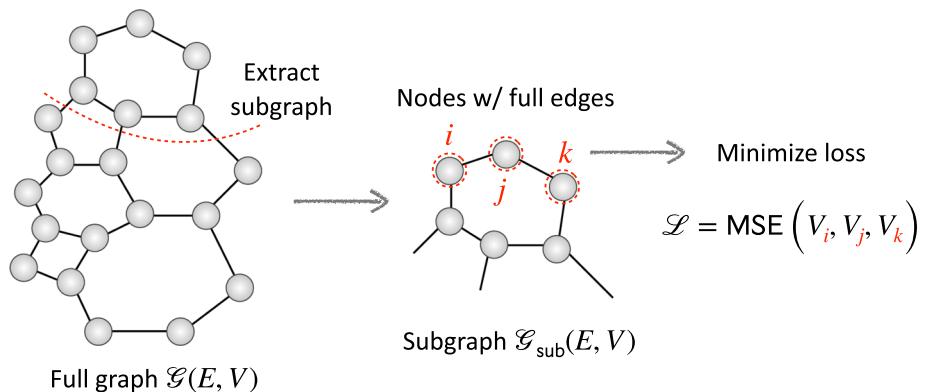
### 4.1 Training and testing results

Figure 7 displays the overall results of the model training and testing. The left subfigure shows the prediction evaluations on the training set, while the right subfigure shows the prediction evaluations on the testing set. Both subfigures illustrate a high degree of correlation between the predicted and benchmark values, with  $R^2$  values of 0.993 and Pearson correlation coefficients of 0.996. The red dashed lines represent the ideal

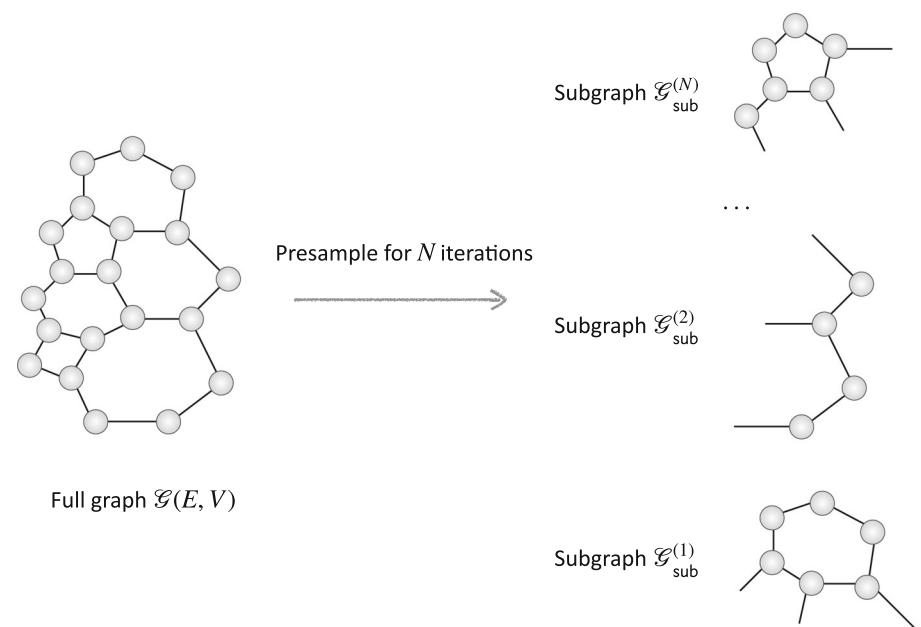
<sup>3</sup> For details one could refer to `torch_geometric.utils.subgraph`.

<sup>4</sup> Half of the graph nodes are sampled from the full graph.

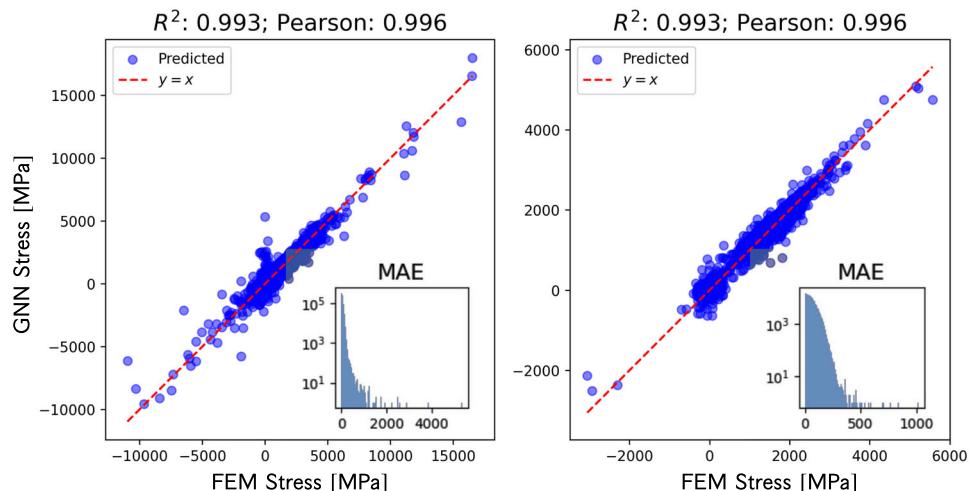
**Fig. 5** Schematic illustration for the proposed subgraph training method using the subgraph extracted from the full graph. The subgraph  $\mathcal{G}_{\text{sub}}$  are extracted from the sampled nodes from the full graph  $\mathcal{G}$ , in which the nodes containing full-edge information (e.g.,  $i$ ,  $j$ , &  $k$ ) were considered in the loss function during training



**Fig. 6** Schematic illustration of the presampling methods for subgraphs training to improve training efficiency and prediction accuracy. Subgraphs  $\mathcal{G}_{\text{sub}}^N$  are presampled for  $N$  epochs for the full graph  $\mathcal{G}$ . All the FEM graphs in the training sets are presampled for  $N$  epochs before training



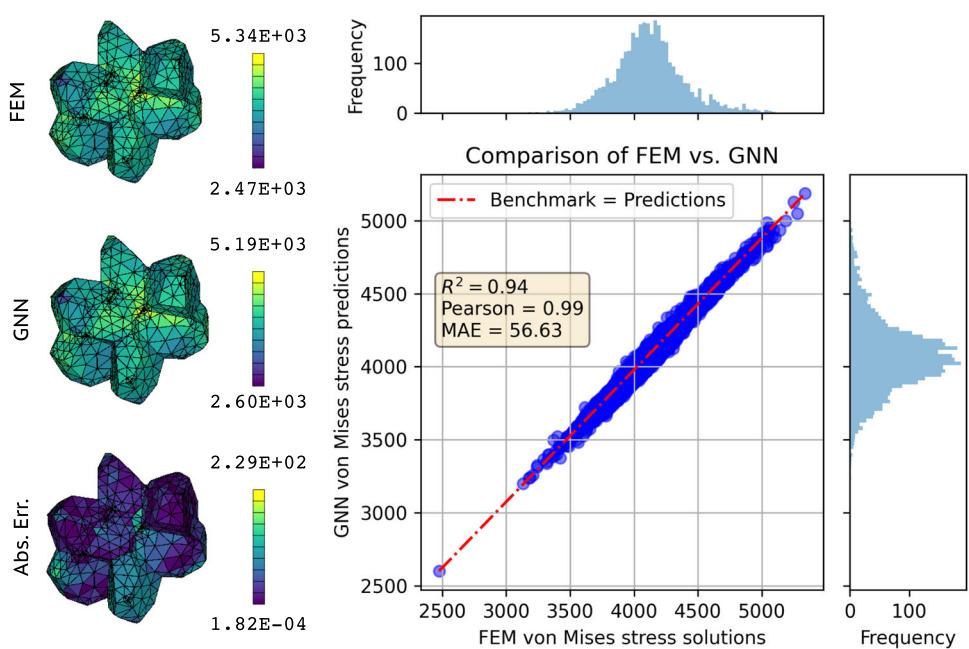
**Fig. 7** The predictions of the trained model on the training & testing sets for all the stress components. The left subfigure is the prediction on the training set and the right subfigure is on the testing set. The distribution of MAE is then visualized in the right-bottom corners



“ $y = x$ ” line, indicating predictions equal to benchmark values. The insets in each subfigure show the mean absolute error (MAE) distributions, further highlighting the model’s

performance. These results demonstrate the model’s robust predictive accuracy on both training and testing datasets.

**Fig. 8** The comparison between FEM and GNN predictions, with absolute errors (visualized on elements with marked color bars) on von Mises stress. The unit for stress is [MPa]. The evaluation of the prediction quality on von Mises stress for the example polycrystal. The unit for stress is [MPa]



#### Algorithm 1 Training algorithms for message-passing GNN

**Require:** Graph data files containing  $\mathcal{G}(V, E)$  converted from finite element meshes, stored in the form of Torch Geometric tensors; mapping the nodal input to outputs  $\epsilon \in \mathbb{R}^{N \times 6} \oplus \ell \in \mathbb{R}^{M \times 1} \mapsto \sigma \in \mathbb{R}^{N \times 6}$ . Number of training graphs  $N_G$  ( $= 72$ ).

**Hyperparameters:** Subgraph ratio:  $\xi_{\text{train}}$ ; The embedding dimension emb; Number of epochs Epochs; Select optimizer Adam(); pre-sampled subgraphs list  $\mathcal{B}_G(\text{Epochs})$  from the training graphs.

Load pretrained GNN model  $\text{GNN}[\cdot]$ .  $\triangleright$  optional based on existence for  $ep < \text{Epochs}$  do

```

 $\mathcal{B}_G^{(ep)} \leftarrow \mathcal{B}_G(ep)$ 
for  $\text{ID}_G$  in  $N_G$  do
     $\mathcal{G}_{\text{sub}} \leftarrow \mathcal{B}_G^{(ep)}(\text{ID}_G)$             $\triangleright$  obtain pre-sampled subgraph
     $\mathbf{c}^{\text{ind}} \leftarrow \mathcal{F}(\mathcal{G}_{\text{sub}}, \mathcal{G})$ .           $\triangleright \mathcal{F}(\cdot)$ : filtering function to sort
connection indices
     $\tilde{\sigma} \leftarrow \text{GNN}[\mathcal{G}_{\text{sub}}(\epsilon, \ell)]$   $\triangleright$  based on defined GNN in Sect. 3.2.
     $\mathcal{L} \leftarrow \text{MSE}(\tilde{\sigma}[\mathbf{c}^{\text{ind}}], \sigma[\mathbf{c}^{\text{ind}}])$   $\triangleright$  only active nodes are included
in the loss.
     $\text{GNN}(\Theta) \xleftarrow{\text{backward}} \mathcal{L}$ .            $\triangleright$  backpropagation
    Clips gradient norm, & optimization:  $\arg \min_{\Theta} \mathcal{L}$ .  $\triangleright$  Adam()
end for
 $ep += 1$ 
end for
Save the trained GNN model  $\text{GNN}$ .  $\triangleright$  requires the specified device
for testing.

```

## 4.2 Analysis on finite element meshes

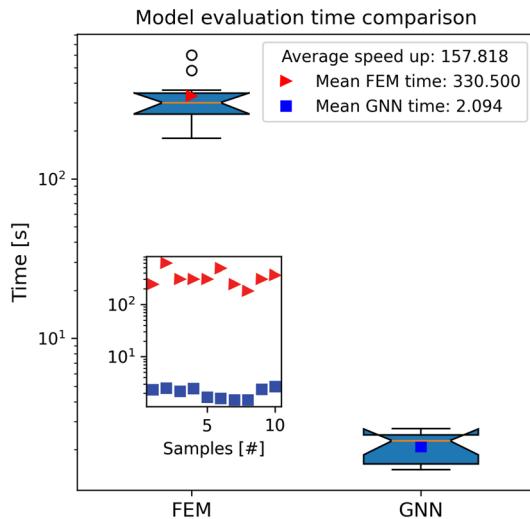
The von Mises stresses are calculated on each cell as<sup>5</sup>

$$\sigma_{\text{vM}} = \sqrt{\frac{1}{2} [(\sigma_1 - \sigma_4)^2 + (\sigma_4 - \sigma_6)^2 + (\sigma_6 - \sigma_1)^2 + 6(\sigma_2^2 + \sigma_3^2 + \sigma_5^2)]} \quad (12)$$

<sup>5</sup> Using the stress notation introduced in Sect. 2.2.

are visualized on the virtual polycrystals (Fig. 8) comparing FEM and GNN, accompanied by the absolute errors. The general stress distribution trends are well learned on the meshes, demonstrated by the stress data distribution. Figure 8 quantitatively verifies this observation with a high  $R^2$  value of 0.94 and Pearson coefficient of 0.99. The overall MAE for the von Mises stress for this polycrystal is 56.63 [MPa], verifying and quantifying the low deviation of the GNN predictions from the benchmark. Combined analysis from Figs. 18 and 8 detailedly illustrates GNN's effective learning. Note that because the GNN is trained directly on the stress tensor components, accurate predictions of the von Mises stress are inherently expected due to the model's efficient learning capabilities.

One of the main advantages of the proposed approach is that it reduces the computational burden for plasticity modeling. Figure 9 presents the speed-up evaluation comparing the FEM and GNN computational time on 10 randomly selected polycrystals in the testing sets. From the subfigure, one observes that the time does not vary much for the 10 polycrystal samples (blue & red dots). The average speed-up is estimated at 158, showing that the proposed GNN plasticity can significantly accelerate plasticity modeling with high-accuracy predictions. Several reasons could contribute to this speed-up: (i) In the FEM model, the solver updates stress fields iteratively for each step. This involves nested loops to account for the nonlinear plasticity model, resulting in a significantly increased computational load [36]. (ii) The forward evaluation is computationally efficient in PyTorch [46]. (iii) Our model size is compact (Eq. (6) with small embedding size);



**Fig. 9** Speed up time comparing constitutive model evaluations of FEM and GNN from 10 randomly selected polycrystal samples. The models are evaluated on a single CPU node on the Sherlock system [47]

this lightweight nature further contributes to the high-speed evaluation mentioned in (ii).

### 4.3 Deployment on validation dataset

To thoroughly analyze the generalizability of the proposed GNN method, we extend our evaluation beyond the testing sets by running 30 unseen simulations with newly generated polycrystals as the validation set and estimating the prediction quality of the GNN. Figure 10 provides an analysis of another polycrystal, achieving an overall  $R^2$  value of 0.993 for stress components. The von Mises stresses are predicted with high accuracy, as demonstrated by the qualitative observations on the left and quantitative comparisons on the right, yielding  $R^2$  values of 0.94 and 0.96. It demonstrates the GNN-based plasticity method generalizes well beyond the training and testing sets, maintaining high-quality predictions on unseen polycrystals. Notably, the polycrystal meshes used in the training, testing, and validation datasets have varying dimensions. Conducting inference on such samples would be nearly impossible with traditional regression methods like vanilla MLPs or CNNs, highlighting the effectiveness of the GNN approach. The overall  $R^2$  score for the validation set is 0.992 and a Pearson coefficient of 0.996 (see Fig. 10).

Figure 11 presents an error analysis that compares training and testing sets with validation data sets in all stress components ( $\sigma_1$  to  $\sigma_6$ ), by directly visualizing the MAE distribution. The distributions for the training and testing sets closely resemble those of the validation datasets, validating that the proposed GNN plasticity method does not overfit

and generalizes well to the stress distributions across various polycrystals.

### 4.4 Limitations of the proposed method

With the demonstrated fast, accurate, and generalizable predictability of our method on polycrystal plasticity, there are still several limitations. (1) The model is not able to predict the loading path, instead, we are demonstrating the map between strain-stress snapshots that are learnable from our GNN model. (2) Currently, the model is not agnostic to the given material parameters (i.e., elastic moduli, hardening coefficients, etc.).<sup>6</sup> The test cases are material-dependent. (3) Uncoupled stress components with the loading orientation are not accurately captured. By taking all the stress components as a full dataset, since loading coupled stress components  $\{\sigma_{ij} \mid i = 1, j = 1, 2, 3\}$  (using general tensor notation) and uncoupled stress components  $\{\sigma_{ij} \mid i \neq 1, j = 1, 2, 3\}$  are not on the same scale, it is quite challenging for the GNN to accurately predict all the stress components. These are valuable future directions that continue with our current model.

## 5 Summary and conclusions

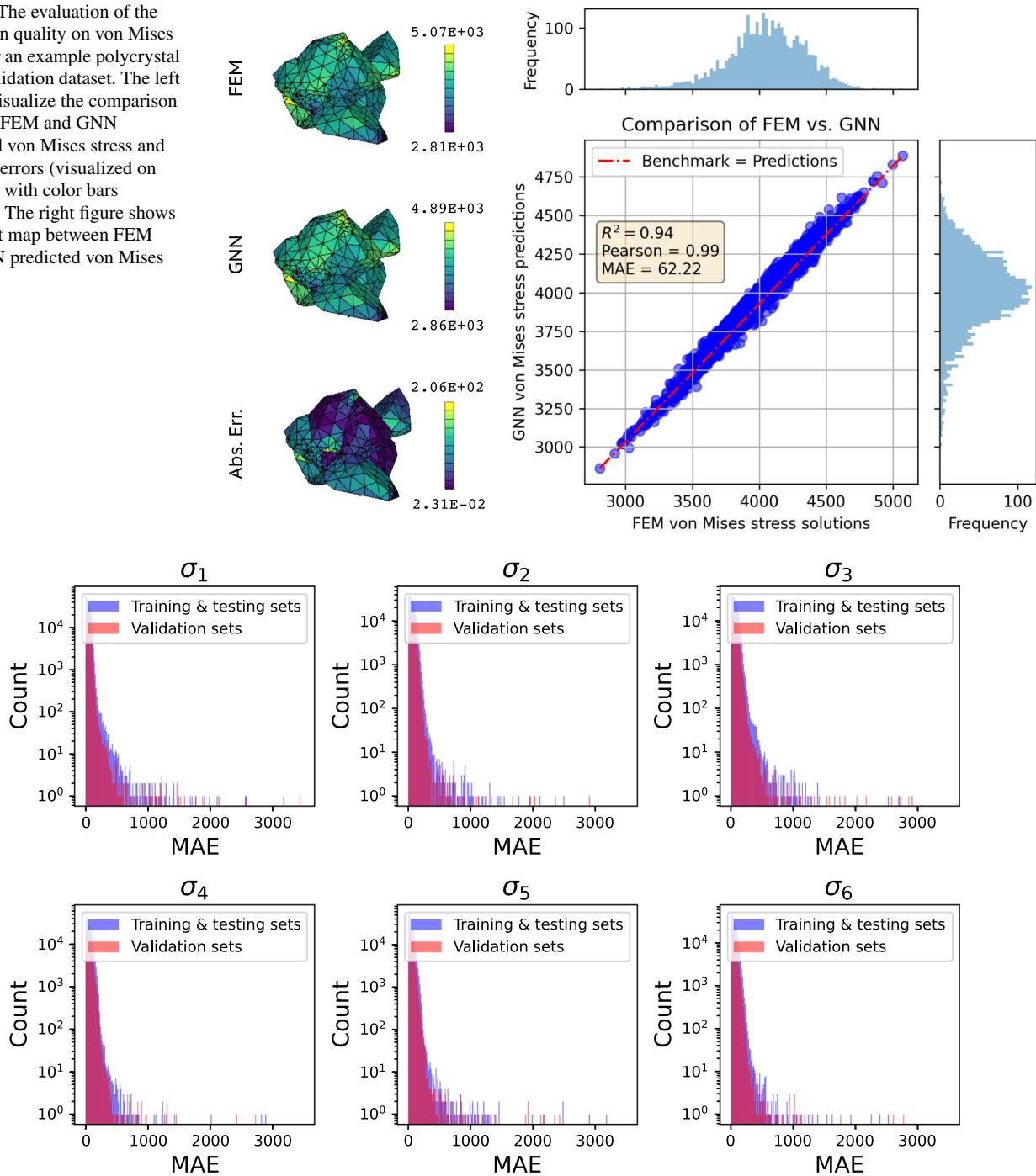
In this paper, we introduce a novel approach for stress predictions using graph neural networks with subgraph training in polycrystal plasticity. The key advantages of our method are: (1) Handling data with varying dimensions—our GNN model accommodates different node counts generated from various polycrystal meshes, allowing for flexible input data; (2) Efficient subgraph training—by randomly sampling subgraphs from polycrystals containing  $\sim 10^5$  nodes and edges,<sup>7</sup> we reduce the computing memory requirements; (3) Preserving geometric features—our GNN model incorporates nodal and edge information, preserving the spatial distribution of stress and strain, thereby enhancing the “learnability” of the data.

Our numerical experiments demonstrate that the GNN model accurately predicts stress components, achieving  $R^2$  scores greater than 0.99 on the training, testing, and validation datasets. Additionally, the von Mises stress predictions for the polycrystals indicate that the proposed GNN method accurately captures von Mises stress features. The model generalizes well beyond the training and testing data, as evidenced by the similar MAE distribution across the training, testing, and validation datasets. The proposed GNN method speeds up stress predictions in the plastic regime more than 150 times compared with the benchmark finite element methods.

<sup>6</sup> According to the data is generated for a defined material in Sect. 2.2.

<sup>7</sup> I.e., on the order of.

**Fig. 10** The evaluation of the prediction quality on von Mises stress for an example polycrystal in the validation dataset. The left figures visualize the comparison between FEM and GNN predicted von Mises stress and absolute errors (visualized on elements with color bars marked). The right figure shows the direct map between FEM and GNN predicted von Mises stresses



**Fig. 11** Comparison of absolute error distributions between the training, testing, and validation datasets

We also briefly outline the limitations of our framework: stress components that are uncoupled from the loading direction are not accurately captured. Only the map between stress-strain snapshots is the learning target for our GNN. However, these uncoupled stress components will not evidently affect the effectiveness of the GNN method in mechanical analysis, particularly when estimating the critical stress

under plastic deformation as the von Mises stresses are accurately captured.

This work outlooks surrogate modeling of polycrystal plasticity using graphs to demonstrate the transient strain-stress map can be learned by GNNs. Future work could include incorporating physics-informed features into the framework and tackling more path-dependent plasticity modeling tasks, leaving open space to the field.

## Appendix A Data preparation

### Appendix A.1 Graph conversion methods

The training graphs were converted from the FEM mesh cells and the neighboring connections (Sect. 2). There were three main considerations for converting the graph in such a way: (1) The numerical values from FEM are solved on mesh cells. Hence, directly converting mesh elements to nodes makes defining the map from strain to stress much more straightforward. (2) Such FEM graphs are agnostic to the order of the test functions used in our FEM calculations. (3) In some sense, only the “first-order” connections between the mesh cells are created as edges, making the conversion process much faster and more efficient. The FEM graphs converted from this method are inspired by respecting the conformality of FEM, in which two common mesh elements share the edge. One of the other intuitive ways is to use the FEM node as a graph vertex directly, where nodal connections are edges. The left subfigure of Fig. 12 (Node graph) illustrates this graph conversion method. As mentioned previously, the drawbacks are that it is hard to define the strain-stress map on the graphs, that elements of higher order introduce redundant nodes, and that the edges do not preserve the connections between element cells. The right subfigure of Fig. 12 (Mesh graph) illustrates the graph using the mesh cell method we proposed.

### Appendix A.2 Finite element implementation of crystal plasticity

Here we briefly discuss the implementation of crystal plasticity using FEM using *FEPX*. Note that only the main steps are summarized herein for a clearer understanding of the manuscript; for numerical implementation details, please refer to Refs. [37, 42]. For the elastic deformation, the cubic symmetry for the stiffness matrix is employed, representing the stress-strain relationship following:

$$\begin{bmatrix} \tau_{11} \\ \tau_{22} \\ \tau_{33} \\ \tau_{23} \\ \tau_{13} \\ \tau_{12} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{12} \\ C_{12} & C_{11} & C_{12} \\ C_{12} & C_{12} & C_{11} \\ & & C_{44} \\ & & C_{44} \\ & & C_{44} \end{bmatrix} \begin{bmatrix} e_{11} \\ e_{22} \\ e_{33} \\ 2e_{23} \\ 2e_{13} \\ 2e_{12} \end{bmatrix} \quad (\text{A.1})$$

where  $e_{ij}$  and  $\tau_{ij}$  are the shear strain and stress.

In the kinematic evolution, the motion can be split into volumetric and deviatoric parts,<sup>8</sup> in which the elasticity equation

relating the Kirchhoff stress and elastic strain writes:

$$\tau_r\{\tau\} = \frac{\kappa}{3}\tau_r\{\epsilon^e\}, \quad \{\tau'\} = [\mathbb{C}']\{\epsilon^{e'}\} \quad (\text{A.2})$$

where  $\kappa$  is bulk modulus, following the relationship  $\kappa = 3(C_{11} + 2C_{12})$ . For the components in  $[\mathbb{C}']$ , they relates to  $[\mathbb{C}]$  as  $C'_{11} = C_{11} - C_{12}$ ,  $C'_{22} = C_{11} - C_{12}$ , and  $C'_{33} = C'_{44} = C'_{55} = C_{44}$  based on cubic symmetry.

The spatial time-rate difference of the elastic strain can be expressed in finite difference scheme:

$$\{\dot{\epsilon}^e\} = \frac{1}{\Delta t} (\{\epsilon^e\} - \{\epsilon_0^e\}) \quad (\text{A.3})$$

where  $\{\epsilon^e\}$  and  $\{\epsilon_0^e\}$  are the elastic strains at the end and beginning of a time step.  $\Delta t$  is the time step.

For the deviatoric portion of the motion, the deformation rate can be expanded in the form (from Eq. (3))

$$\{\mathbf{D}'\} = \frac{1}{\Delta t} \{\epsilon^{e'}\} + \{\hat{\mathbf{D}}^p\} + [\hat{\mathbf{W}}^p] \{\epsilon^{e'}\} - \frac{1}{\Delta t} \{\epsilon_0^{e'}\} \quad (\text{A.4})$$

The deviatoric portion of the Cauchy stress can be updated following:

$$\{\sigma'\} = [\mathbf{s}] (\{\mathbf{D}'\} - \{\mathbf{h}\}) \quad (\text{A.5})$$

where

$$\begin{aligned} [\mathbf{s}]^{-1} &= \frac{\beta}{\Delta t} [\mathbb{C}']^{-1} + \beta [\mathbf{m}] \\ \{\mathbf{h}\} &= [\hat{\mathbf{W}}^p] \{\epsilon^{e'}\} - \frac{1}{\Delta t} \{\epsilon_0^{e'}\} \end{aligned} \quad (\text{A.6})$$

in which  $\beta = \det(\mathbf{v}^e)$ , and  $[\mathbf{m}]$  is the map from the deviatoric Kirchhoff stress  $\tau'$  to  $\hat{\mathbf{D}}^p$ .

The interpolation function  $[\mathbf{N}(\xi, \eta, \zeta)]$  interpolates the nodal coordination points and the velocity fields via  $\{\mathbf{x}\} = [\mathbf{N}(\xi, \eta, \zeta)] \{\mathbf{X}\}$  and  $\{\mathbf{v}\} = [\mathbf{N}(\xi, \eta, \zeta)] \{\mathbf{V}\}$  for the global assembly. Under the Galerkin formulation, weight functions were used to construct the residual, where the weight function  $\psi$  can be interpolated in the same way:

$$\{\psi\} = [\mathbf{N}(\xi, \eta, \zeta)] \{\Psi\} \quad (\text{A.7})$$

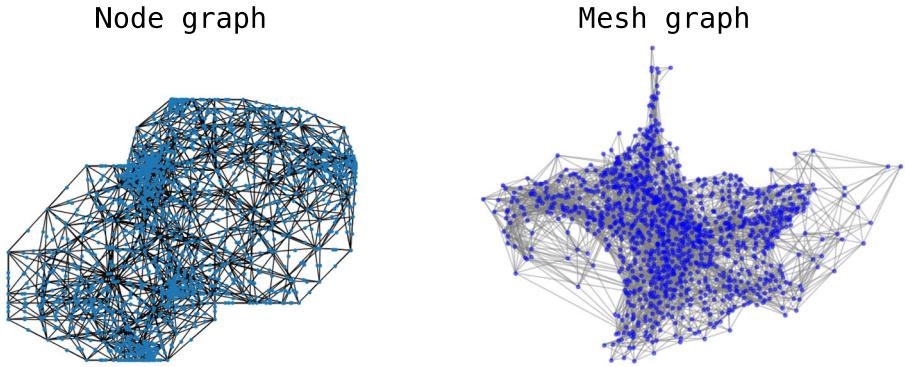
To achieve equilibrium for the system, one would solve the global weighted residual equation:

$$R_u = \int_{\mathcal{B}} \psi \cdot (\nabla \cdot \sigma + \mathbf{f}) d\mathcal{B} = 0 \quad (\text{A.8})$$

where  $\mathcal{B}$  is the continuum body to be solved.

<sup>8</sup> For convenience of numerical implementation.

**Fig. 12** Schematic illustration for different graphs obtained from FEM mesh nodes and cells



After the global assembly, one could solve the nonlinear system to obtain the velocity field  $V$  from<sup>9</sup>:

$$[[K_d] + [K_v]] \{V\} = \{F_a\} + \{F_d\} + \{F_v\} \quad (\text{A.9})$$

where the details of calculating the elements in the global stiffness matrix and force vectors can be checked in Ref. [37].

## Appendix B Graph neural networks

### Appendix B.1 Training procedure

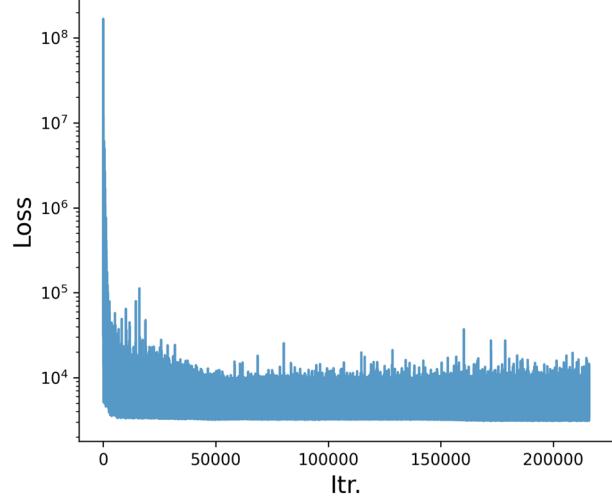
Since the model is directly trained on the stress data ( $\sim 10^3$ ), the loss is the MSE loss between the predicted & benchmark stress, hence displaying a large loss also on the scale of  $10^3$ . Based on our observations, the training loss shows a minimal decrease beyond 1000 epochs, indicating that the model has effectively converged. To further investigate this, we extended the training to 3000 epochs, and the corresponding loss curve is presented in Fig. 13.

It is important to note that the GNN is iteratively trained on 72 graphs for each epoch. Thus, training for 3000 epochs corresponds to approximately 216,000 iterations. The figure shows that the loss stabilizes after approximately 5000 iterations, plateauing to fluctuating values without significant further reduction. This behavior suggests that 1000 epochs are sufficient for the GNN to achieve “convergence”.

We believe this demonstrates that the choice of 1000 epochs is appropriate for training the GNN, as it ensures convergence without unnecessary computational overhead.

### Appendix B.2 Model characterization

Figure 14 shows the model prediction quality with and without message-passing (MP) layers tested with full graph ratio. Interestingly, for this test, both models perform well with high-quality predictions. However, it can be seen that the



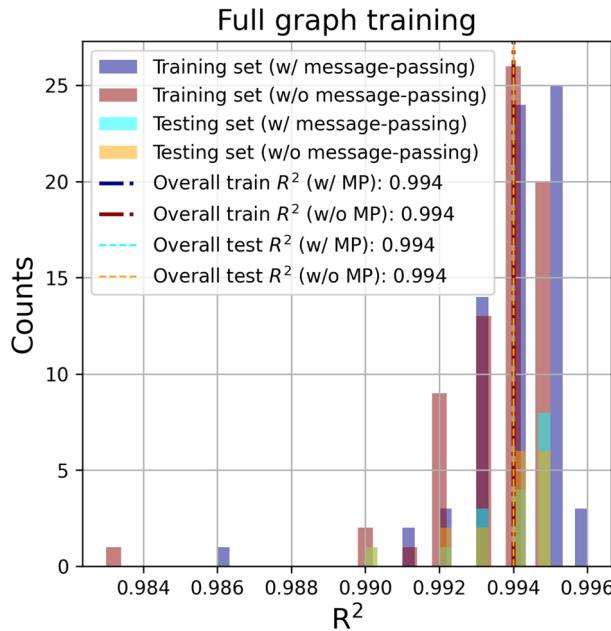
**Fig. 13** Loss curve w.r.t. iterations during the training process for 3000 epochs

predictions with MP have more high-quality predictions by comparing the blue and red bars and comparing the light blue with the yellow bars at the relatively lower  $R^2$  scores regime ( $< 0.99$ ). This suggests that the message-passing mechanism improves GNN’s predictability but does not significantly improve the model.

Figure 15a visualizes the Pearson correlations between the FEM calculated stress components, indicating the correlations between  $\sigma_{1j}$ ,  $j = 1, 2, 3$  is much higher than that of other stress components. Figure 15b visualizes the Pearson correlations between the stress components predicted by GNN and FEM. It can be observed that GNN predicts much more accurately for stress components  $\sigma_{1j}$ . This suggests that the loading-coupled stress components, whose both numerical values and correlations are higher, are much more well-predicted by the GNN, which is consistent with the physical scenario in which 1-directional loading is applied.

Figure 16 presents the  $R^2$  scores for GNN predictions trained with varying subgraph ratios. The results indicate that a subgraph ratio of 0.25 does not yield accurate predictions. However, increasing the ratio to 0.5 allows the GNN to achieve an overall  $R^2 = 0.994$  for both training and test-

<sup>9</sup> Nonlinear in the sense that  $[K_d]$  and  $[K_v]$  depends on  $[V]$ .



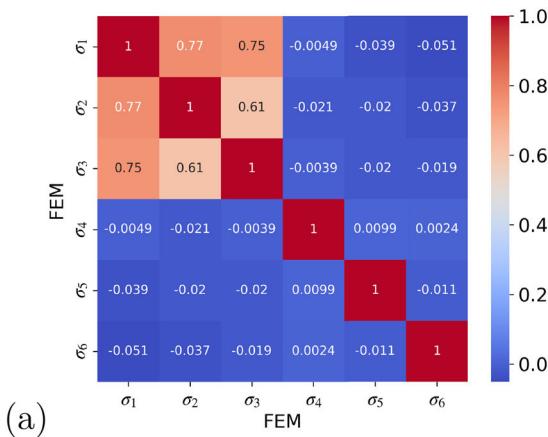
**Fig. 14**  $R^2$  values comparing GNN with and without message-passing mechanisms for full-graph training

ing sets, with no low-accuracy samples observed. Per our standard,  $R^2 > 0.99$  signifies a high-accuracy model. These findings suggest that a subgraph ratio of 0.5 is a decent balance, enabling the GNN to maintain high accuracy while preserving the essential information of the graph.

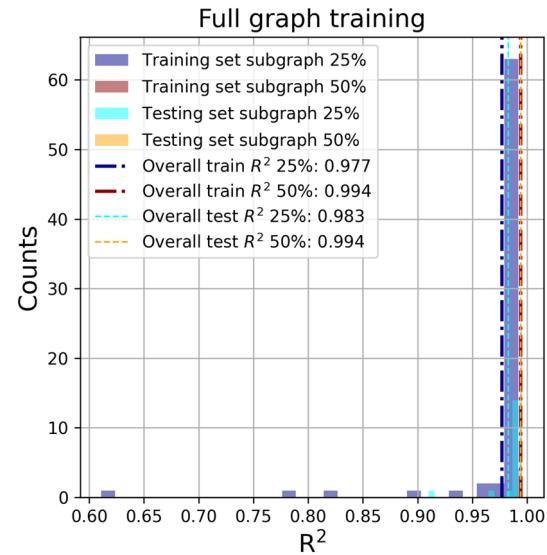
## Appendix C Additional results

### Appendix C.1 Predictions from GNN

Accompanying the high  $R^2$  values, to directly verify the high-quality predictions using GNN, the stress values on each



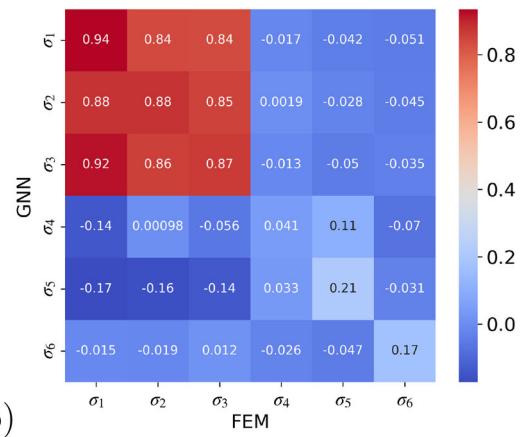
**Fig. 15** Pearson correlation between the stress components. **a** Self-correlations for the FEM calculated stress components. **b** Correlations for calculated stress components predicted by GNN and FEM

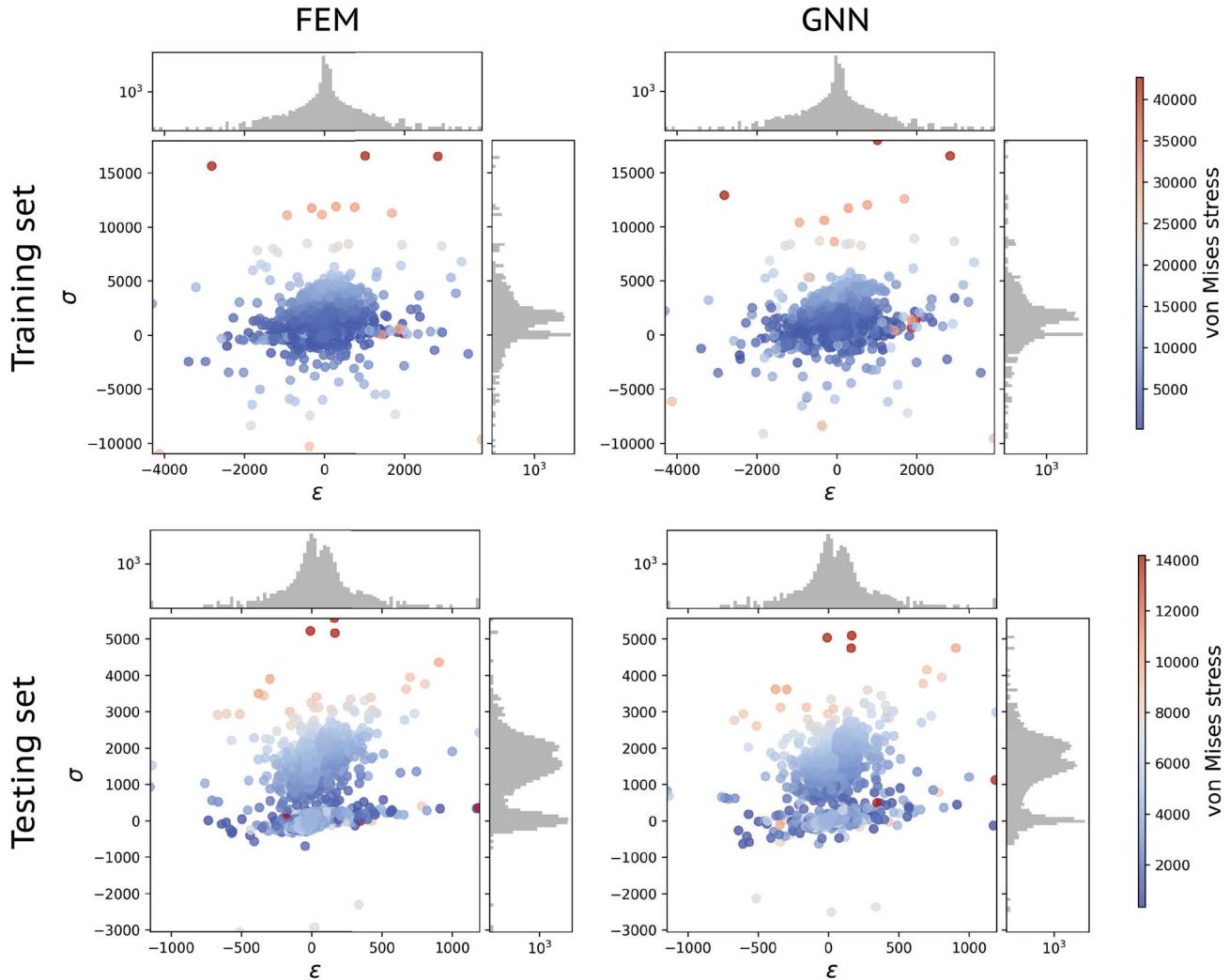


**Fig. 16**  $R^2$  scores from GNN predictions based on different subgraph ratios

finite element cell for both the training and testing sets are visualized (Fig. 17). The general data trends are well captured. With von Mises stress as label marks, the predictions preserve the stress distribution among mesh cells for both the training and testing sets, indicating no overfitting for the proposed method. Figure 7 demonstrates the quality of the predictions with direction prediction data visualization and high  $R^2$  scores.

Associated with the visualization of the data (Fig. 17), we directly visualize the stress distributions on the virtual polycrystals (Fig. 18) for a test polycrystal with  $R^2$  of 0.994. Data scales are well captured by the GNN for each stress component  $\sigma_i$ , with comparably small absolute errors (Fig. 18). GNN predicts much more accurately in the active loading directions by learning different stress value ranges for each component.





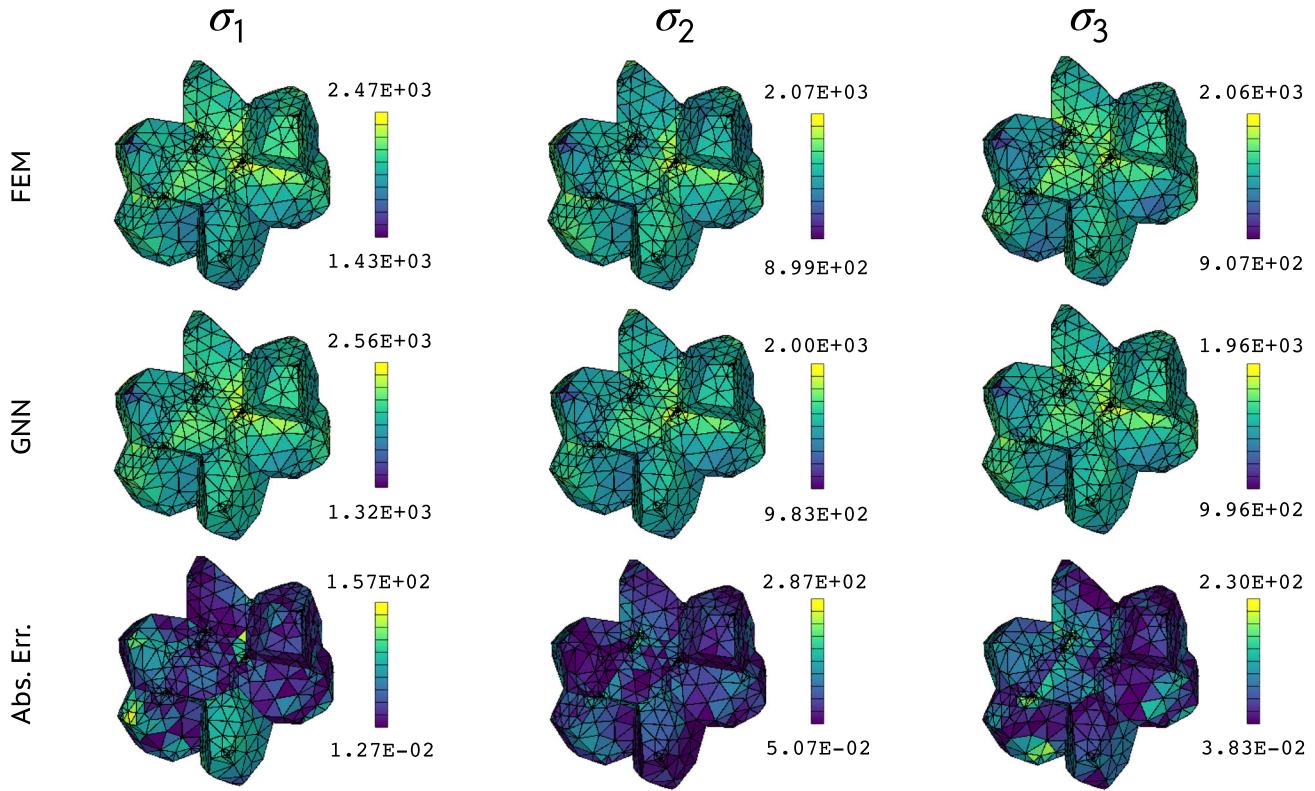
**Fig. 17** The comparison between the ground truth (by *FEPX*) and predictions on the strain-stress maps for all the stress & strain components. The upper figures correspond to the training sets. The bottom figures

correspond to the testing sets. The data points are visualized according to the calculated von Mises stresses

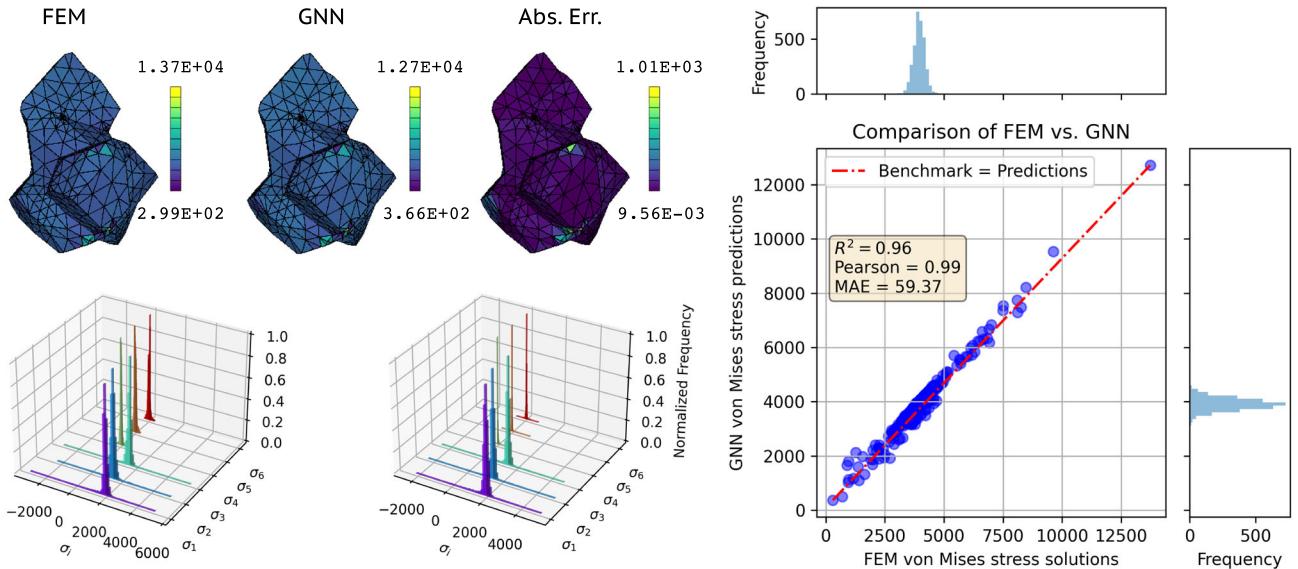
Following the analysis procedure, we demonstrate the effective learning of the GNN by analyzing two other polycrystals with overall  $R^2$  values of 0.992 and 0.991 from inferences, respectively (Figs. 19 and 20). From the upper left figures, one observes very similar von Mises ranges are predicted by FEM and GNN, accompanied by low absolute errors. The quantitative comparison in the right figures confirms the qualitative observation for the von Mises stress inferences, with  $R^2$  and Pearson coefficients of 0.96 and 0.99 for both the two polycrystals, respectively. Also, the two methods both predict similar stress component ranges demonstrated in the bottom left figures, illustrated by different color histograms. To summarize, these results demonstrate a few aspects of the robustness of the proposed GNN plasticity modeling: (1) overall stress components are predicted well by the high  $R^2$  values, with no overfitting

for testing sets; (2) general trends of stress components are captured; (3) von Mises stress are well learned verified both qualitatively and quantitatively, demonstrated via similar stress distribution and high  $R^2$  and Pearson coefficients. Specifically, von Mises is not introduced (or constrained) in the training process. Additionally, the plasticity model is effectively learned from a limited amount of training data.

Figure 20 also reflects the potential limitation of the proposed method: the distribution of the stress components  $\sigma_4$ ,  $\sigma_5$ , and  $\sigma_6$  are not fully captured by the GNN. Qualitatively, one may argue that the variance of the data distribution around zero is not learned via GNN. This can be explained by the low-stress value range for the related stress components uncoupled with the loading direction (i.e., 1-direction). However, as can be visually observed and with Eq. (12), the stresses in the uncoupled directions do not significantly con-

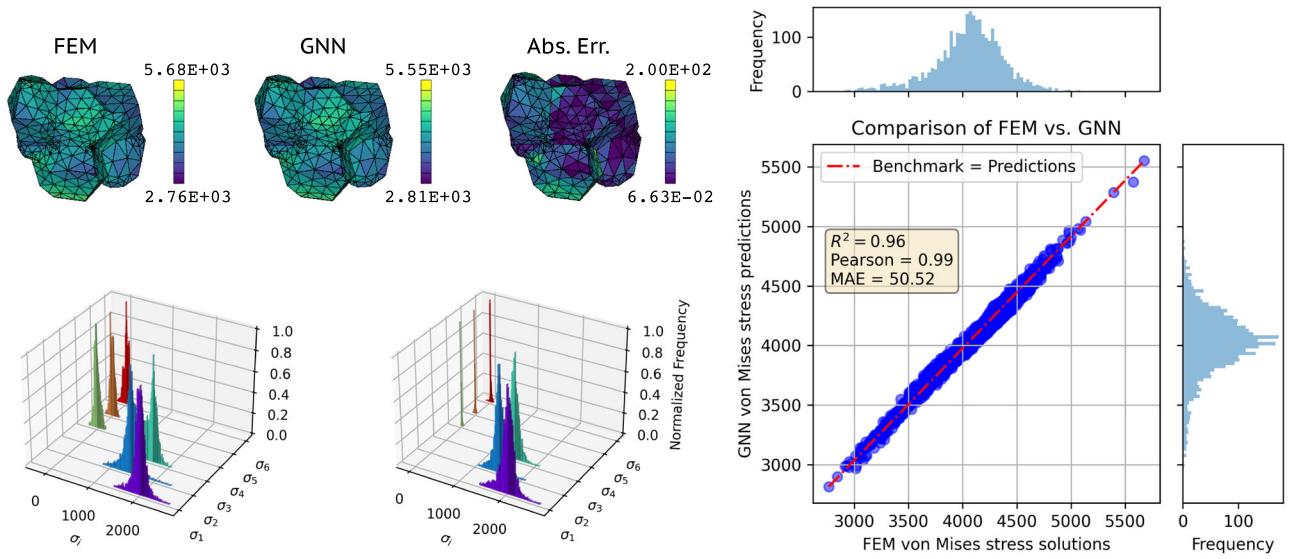


**Fig. 18** The comparison between FEM and GNN predictions, with absolute errors of stress components on one example grain in the testing sets for stress components  $\sigma_1$ ,  $\sigma_2$  and  $\sigma_3$  (visualized on elements with marked color bars). The unit for stress is [MPa]



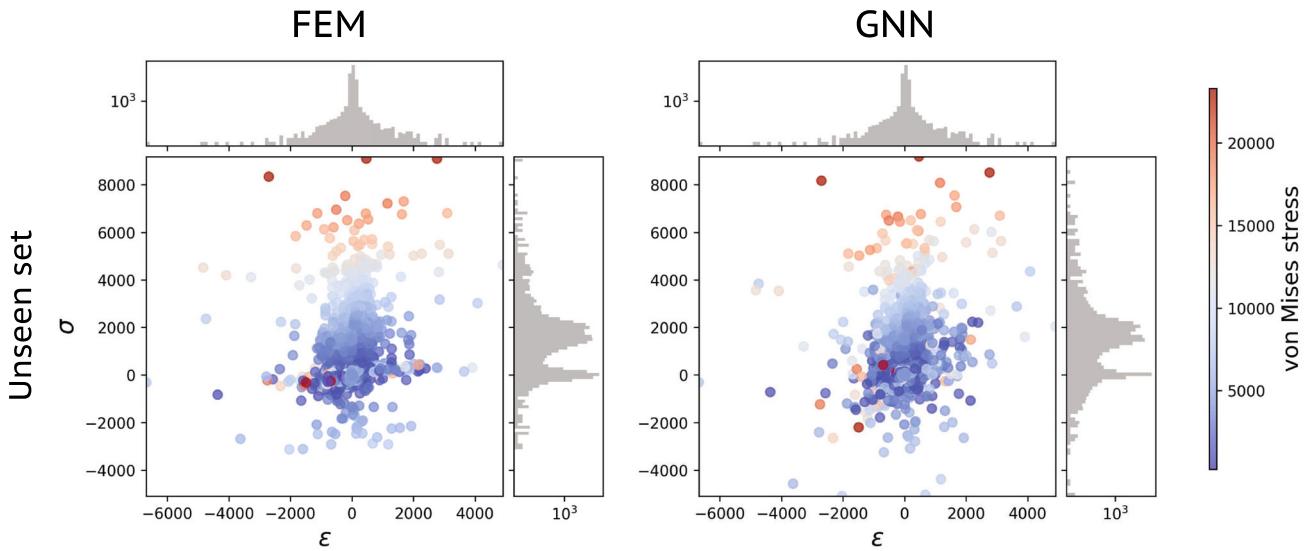
**Fig. 19** The evaluation of the prediction quality on von Mises stress for another example polycrystal. The upper left figures visualize the comparison between FEM and GNN-predicted von Mises stress and absolute errors (visualized on elements with color bars marked). The

bottom left figures visualize the distributions of different stress components. The right figure shows the direct map between FEM and GNN predicted von Mises stresses



**Fig. 20** The evaluation of the prediction quality on von Mises stress for another example polycrystal. The upper left figures visualize the comparison between FEM and GNN predicted von Mises stress and absolute errors (visualized on elements with colorbars marked). The

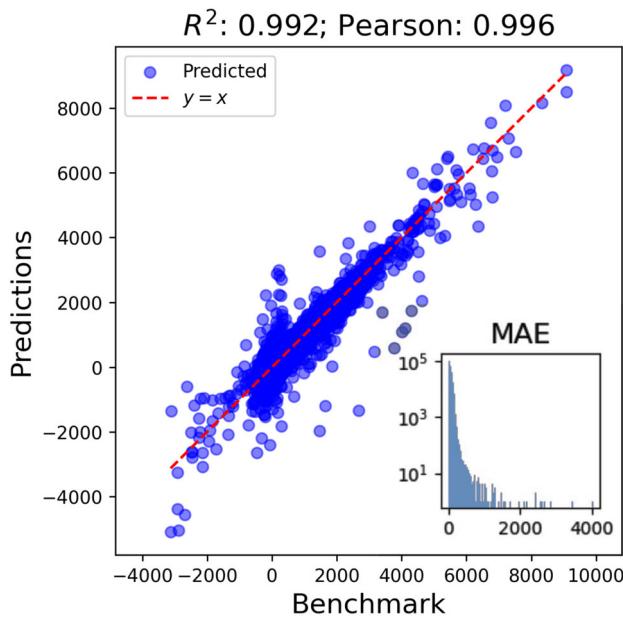
bottom left figures visualize the distributions of different stress components. The right figure shows the direct map between FEM and GNN predicted von Mises stresses



**Fig. 21** The comparison between the ground truth (by *FEPX*) and predictions on the strain-stress maps for all stress-strain components. The data points are visualized according to the calculated von Mises stresses

tribute to the von Mises stresses, considering the high-quality predictions on the von Mises stresses (Figs. 19 and 20). The stress components correlated to the loading direction match well with the benchmark as illustrated in the left-bottom figure.

Figure 21 shows the comparison of the stress components predictions per cell for the overall  $\epsilon \mapsto \sigma$  and  $\epsilon_{11} \mapsto \sigma_{11}$  maps, respectively. The GNN inferences effectively preserve the stress data trends, including both the data distribution and the von Mises stress values. Interestingly, one may



**Fig. 22** The overall prediction results on the validation dataset and the absolute error distribution

discern qualitatively higher discrepancies between the two methods in the “low-stress regime.” This observation aligns with the discussion on the limitations highlighted in Fig. 20: the values for stress components around zero are not accurately captured. Nonetheless, the model demonstrates high performance and provides good overall predictions on the strain-stress maps.

Figure 22 presents the overall predictions of stress components for the 30 unseen simulations, demonstrating accurate predictions with an  $R^2$  value of 0.992 and a Pearson coefficient of 0.996. These results indicate that the proposed GNN

method not only generalizes well within the provided training and testing sets (i.e., interpolation) but also effectively extrapolates to data outside the given data regime.

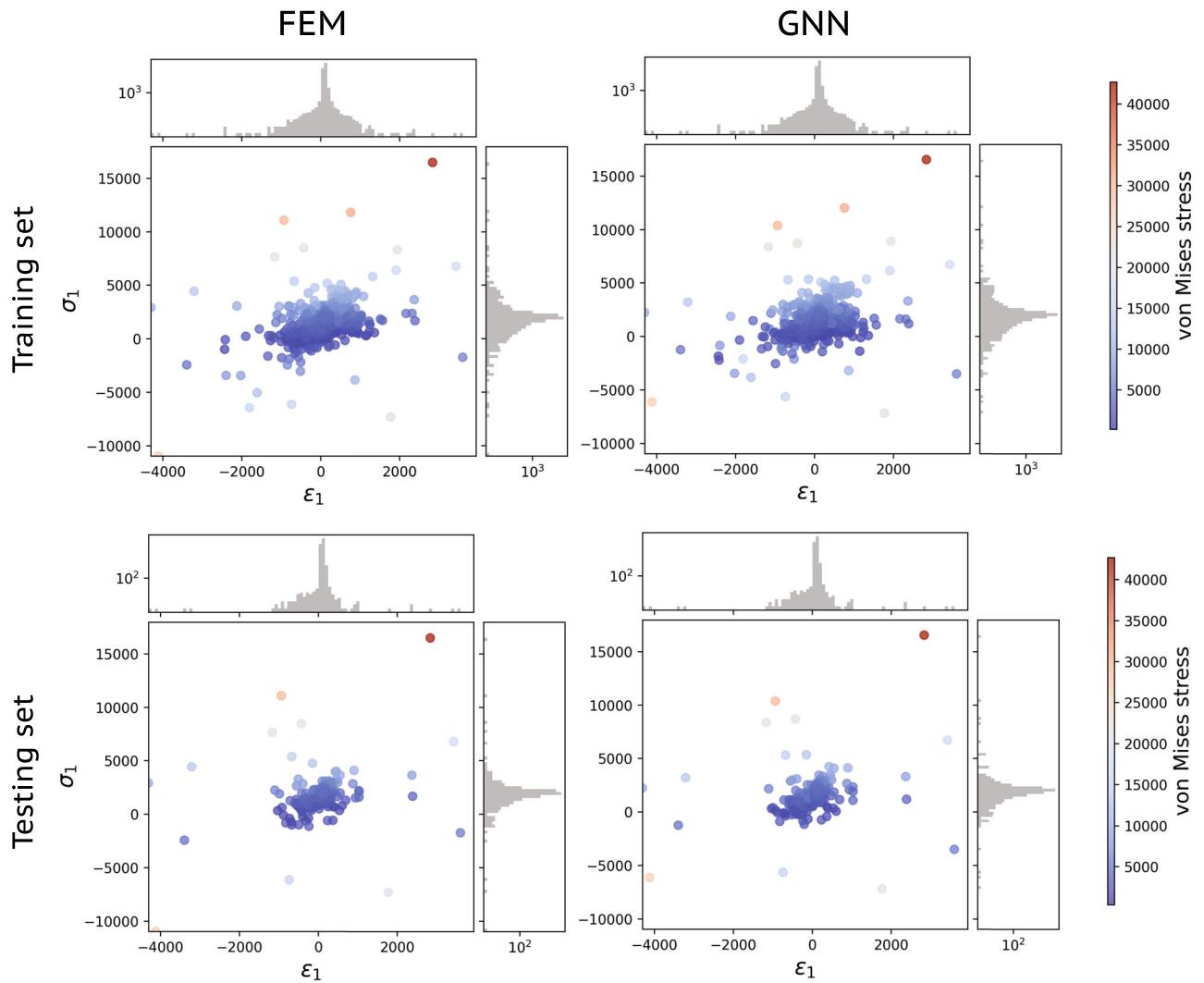
### Appendix C.2 Discussions on loading-coupled directions

Figures 23 and 24 illustrate the 1-directional strain-to-stress mapping, comparing FEM and GNN predictions alongside Figs. 17 and 21. The results show that the GNN achieves significantly more accurate predictions for these maps. We hypothesize that this improved accuracy arises from the larger value range in the loading direction (and its coupled components), which may enhance the model’s ability to predict the map more effectively.

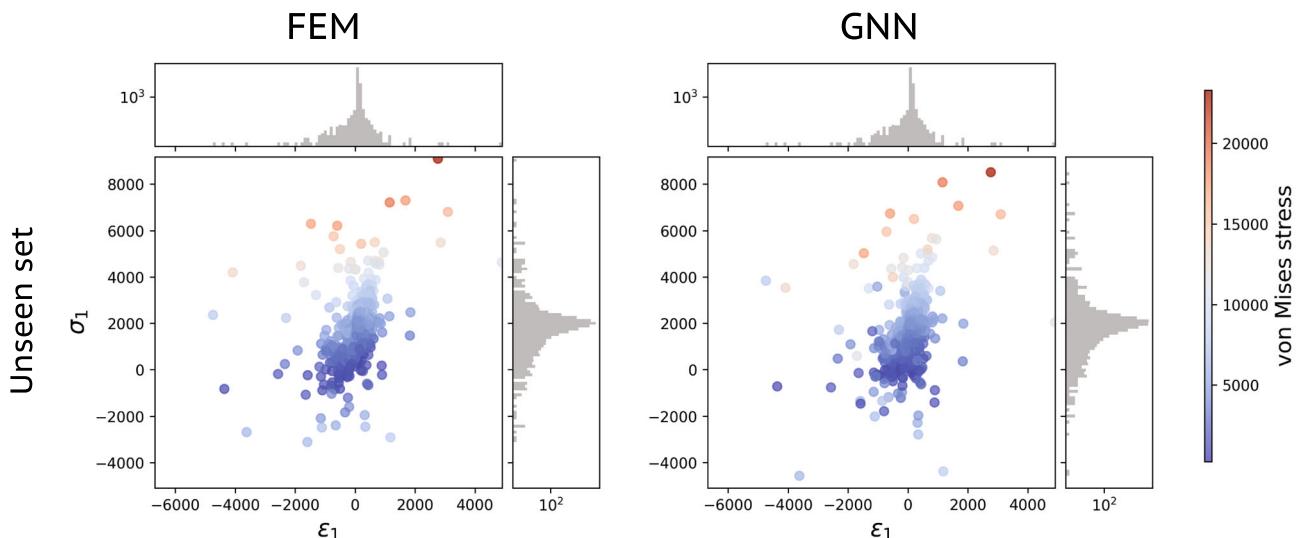
To further elaborate, Fig. 25 presents the predicted stress components  $\sigma_3$ ,  $\sigma_4$ , and  $\sigma_5$  (uncoupled from the loading direction), corresponding to Fig. 18. The predictions for these loading-direction-uncoupled components are observably less accurate. This visualization supports and reinforces our earlier discussions.

### Appendix C.3 Error distribution for stress components

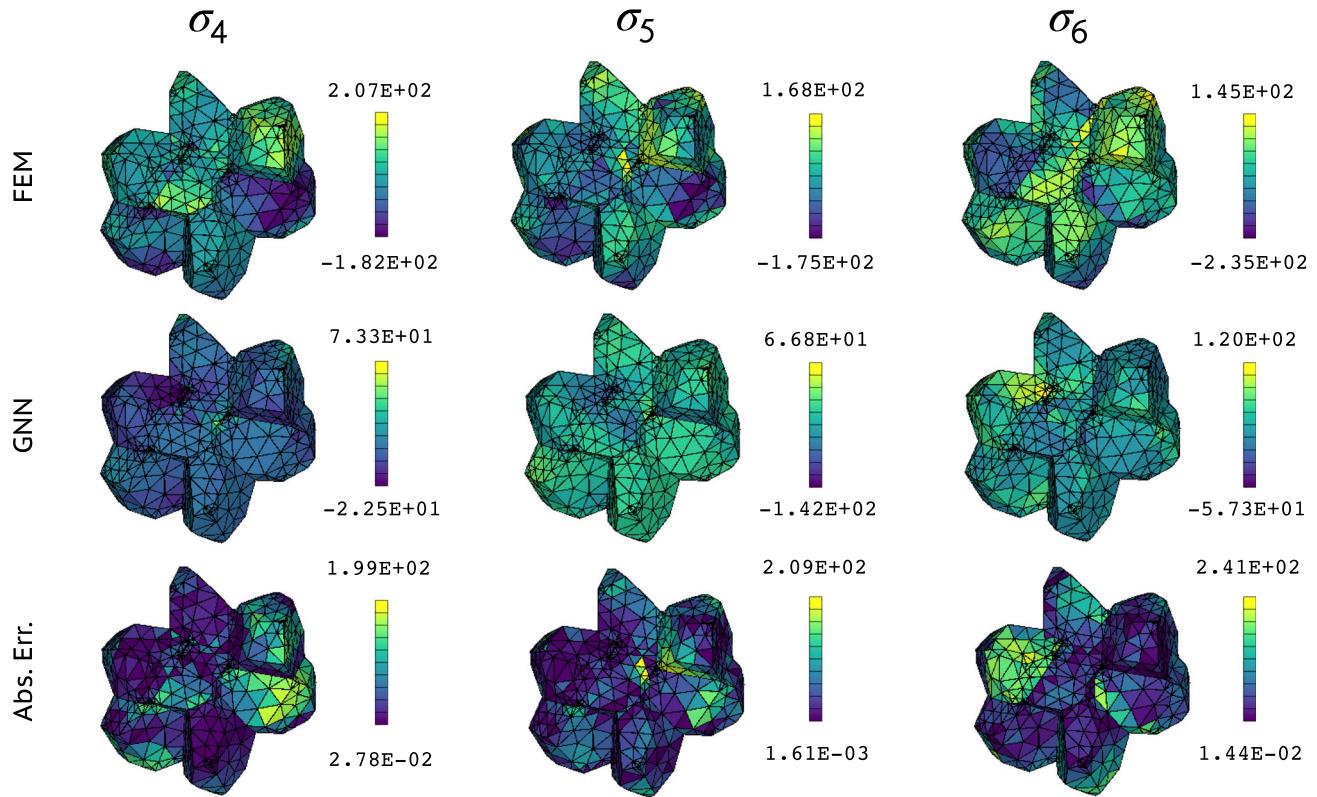
Figure 26 visualizes the prediction error distribution corresponding to Fig. 11, comparing the error distributions across different stress components. The results confirm that the errors remain consistent across the training, testing, and validation datasets. Additionally, the data range shows minimal variation between the training and testing sets and the validation set. This serves as further evidence supporting Fig. 11, demonstrating the strong generalizability of the GNN.



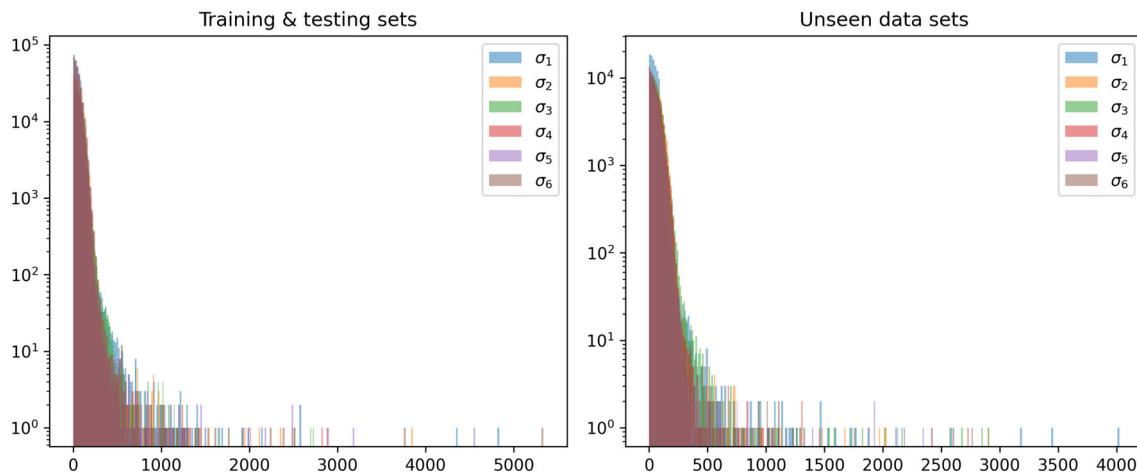
**Fig. 23** Comparison on FEM and GNN predictions on the  $\epsilon_1$ - $\sigma_1$  mapping between the training and testing sets corresponding to Fig. 7



**Fig. 24** Comparison on FEM and GNN predictions on the  $\epsilon_1$ - $\sigma_1$  mapping for the validation sets corresponding to Fig. 21



**Fig. 25** The comparison between FEM and GNN predictions for the loading uncoupled directions, with absolute errors of stress components on one example grain in the testing sets for stress components  $\sigma_4$ ,  $\sigma_5$  and  $\sigma_6$  (visualized on elements with marked color bars). The unit for stress is [MPa]



**Fig. 26** Error distribution between the training & testing sets and the validation dataset (marked as unseen) for comparing the six stress components

**Acknowledgements** The author acknowledges support from the Enlight Foundation Graduate Fellowship via Leland Stanford Junior University. The author thanks Myung Chul Kim of Stanford University for discussions on presampling algorithms for training GNNs, Romain Quey of CNRS for discussions on the implementation of *Neper* and *FEPX* for simulation visualizations, and Matthew Kasemer of The University of Alabama for the general comments on GNN, discussions on plasticity theory, and on the *FEPX* implementation. The author also thanks the anonymous reviewers for their invaluable comments, which significantly improved the manuscript.

**Data Availability** The associated codes and data are available at <https://gitlab.com/hanfengzhai2/GNN-FEM-PolyPlas>. All data and code are published under [MIT License](#). The virtual polycrystals are generated and visualized using *Neper*, accessible at <https://neper.info/>. The FEM plasticity simulations utilize the open-source software package *FEPX*, publicly available at <https://fepx.info/index.html> [37, 42].

## Declarations

**Conflicts of Interest** None.

## References

- von Mises R (1913) Mechanik der festen Körper im plastisch-deformablen Zustand. Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-Physikalische Klasse 1913:582–592
- Huber MT (1904) Właściwa praca odkształcenia jako miara wyteżenia materiału, Czasopismo Techniczne, vol 22
- Azushima A, Kopp R, Korhonen A, Yang D, Micari F, Lahoti G, Groche P, Yanagimoto J, Tsuji N, Rosochowski A, Yanagida A (2008) Severe plastic deformation (SPD) processes for metals. CIRP Ann 57:716–735. <https://doi.org/10.1016/j.cirp.2008.09.005>
- Kreml E (1997) Design for fatigue resistance. ASM International, vol 20, pp 516–532. <https://doi.org/10.31399/asm.hb.v20.a0002469>
- Jiang T, Wu C, Spinella L, Im J, Tamura N, Kunz M, Son H-Y, Gyu Kim B, Huang R, Ho PS (2013) Plasticity mechanism for copper extrusion in through-silicon vias for three-dimensional interconnects. Appl Phys Lett. <https://doi.org/10.1063/1.4833020>
- Hu D, Grilli N, Yan W (2023) Dislocation structures formation induced by thermal stress in additive manufacturing: multiscale crystal plasticity modeling of dislocation transport. J Mech Phys Solids 173:105235. <https://doi.org/10.1016/j.jmps.2023.105235>
- Mitchell JA (2011) A nonlocal, ordinary, state-based plasticity model for peridynamics, Technical Report SAND2011-3166, Sandia National Laboratories
- Drucker DC, Prager W (1952) Soil mechanics and plastic analysis for limit design. Q Appl Math 10:157–165
- Dou W, Xu Z, Hu H, Huang F (2021) A generalized plasticity model incorporating stress state, strain rate and temperature effects. Int J Impact Eng 155:103897. <https://doi.org/10.1016/j.ijimpeng.2021.103897>
- Sills RB, Bertin N, Aghaei A, Cai W (2018) Dislocation networks and the microstructural origin of strain hardening. Phys Rev Lett. <https://doi.org/10.1103/PhysRevLett.121.085501>
- Luan B, Robbins MO (2021) Friction and plasticity in contacts between amorphous solids. Tribol Lett. <https://doi.org/10.1007/s11249-021-01429-7>
- Chen J, Furushima T (2024) Effects of intergranular deformation incompatibility on stress state and fracture initiation at grain boundary: experiments and crystal plasticity simulations. Int J Plasticity 180:104052. <https://doi.org/10.1016/j.ijplas.2024.104052>
- Alharbi HF, Kalidindi SR (2015) Crystal plasticity finite element simulations using a database of discrete Fourier transforms. Int J Plasticity 66:71–84. <https://doi.org/10.1016/j.ijplas.2014.04.006>
- Needleman A, Asaro R, Lemonds J, Peirce D (1985) Finite element analysis of crystalline solids. Comput Methods Appl Mech Eng 52:689–708. [https://doi.org/10.1016/0045-7825\(85\)90014-3](https://doi.org/10.1016/0045-7825(85)90014-3)
- Kalidindi S, Bronkhorst C, Anand L (1992) Crystallographic texture evolution in bulk deformation processing of FCC metals. J Mech Phys Solids 40:537–569. [https://doi.org/10.1016/0022-5096\(92\)80003-9](https://doi.org/10.1016/0022-5096(92)80003-9)
- Raissi M, Yazdani A, Karniadakis GE (2020) Hidden fluid mechanics: learning velocity and pressure fields from flow visualizations. Science 367:1026–1030. <https://doi.org/10.1126/science.aaw4741>
- Zhai H, Zhou Q, Hu G (2022) Predicting micro-bubble dynamics with semi-physics-informed deep learning. AIP Adv. <https://doi.org/10.1063/5.0079602>
- Guo Z, Roy Chowdhury P, Han Z, Sun Y, Feng D, Lin G, Ruan X (2023) Fast and accurate machine learning prediction of phonon scattering rates and lattice thermal conductivity. npj Comput Mater. <https://doi.org/10.1038/s41524-023-01020-9>
- Guo Z, Han Z, Feng D, Lin G, Ruan X (2024) Sampling-accelerated prediction of phonon scattering rates for converged thermal conductivity and radiative properties. npj Comput Mater. <https://doi.org/10.1038/s41524-024-01215-8>
- Zhai H, Yeo J (2022) Computational design of antimicrobial active surfaces via automated Bayesian optimization. ACS Biomater Sci Eng 9:269–279. <https://doi.org/10.1021/acsbiomaterials.2c01079>
- Zhai H, Yeo J (2023) Controlling biofilm transport with porous metamaterials designed with Bayesian learning. J Mech Behav Biomed Mater 147:106127. <https://doi.org/10.1016/j.jmbbm.2023.106127>
- Zhai H, Hao H, Yeo J (2024) Benchmarking inverse optimization algorithms for materials design. APL Mater. <https://doi.org/10.1063/5.0177266>
- Yang Z, Papanikolaou S, Reid ACE, Liao W-k, Choudhary AN, Campbell C, Agrawal A (2020) Learning to predict crystal plasticity at the nanoscale: deep residual networks and size effects in uniaxial compression discrete dislocation simulations. Sci Rep. <https://doi.org/10.1038/s41598-020-65157-z>
- Salmenjoki H, Alava MJ, Laurson L (2018) Machine learning plastic deformation of crystals. Nat Commun. <https://doi.org/10.1038/s41467-018-07737-2>
- Miñkowski M, Laurson L (2023) Predicting elastic and plastic properties of small iron polycrystals by machine learning. Sci Rep. <https://doi.org/10.1038/s41598-023-40974-0>
- Yang Z, Buehler MJ (2022) Linking atomic structural defects to mesoscale properties in crystalline solids using graph neural networks. npj Comput Mater. <https://doi.org/10.1038/s41524-022-00879-4>
- Amigo N, Palominos S, Valencia FJ (2023) Machine learning modeling for the prediction of plastic properties in metallic glasses. Sci Rep. <https://doi.org/10.1038/s41598-023-27644-x>
- Dai M, Demirel MF, Liang Y, Hu J-M (2021) Graph neural networks for an accurate and interpretable prediction of the properties of polycrystalline materials. npj Comput Mater. <https://doi.org/10.1038/s41524-021-00574-w>
- Fan Z, Ma E (2021) Predicting orientation-dependent plastic susceptibility from static structure in amorphous solids via deep learning. Nat Commun. <https://doi.org/10.1038/s41467-021-21806-z>
- Thomas A, Durmaz AR, Alam M, Gumbesch P, Sack H, Eberl C (2023) Materials fatigue prediction using graph neural networks on microstructure representations. Sci Rep. <https://doi.org/10.1038/s41598-023-39400-2>

31. Pagan DC, Pash CR, Benson AR, Kasemer MP (2022) Graph neural network modeling of grain-scale anisotropic elastic behavior using simulated and measured microscale data. *npj Comput Mater.* <https://doi.org/10.1038/s41524-022-00952-y>
32. Yee Tung S, Ee Siang Tan (2023) Influence of curing agent on thermal conductivity and mechanical properties of epoxy composite with silicon nitride nanofillers. In: IOP conference series: materials science and engineering, vol 1284, p 012052. <https://doi.org/10.1088/1757-899X/1284/1/012052>
33. Huang W, Zhu J-Y, Song C-Y, Sun L, Zheng J-P (2024) Mesh-based GNN surrogates for time-independent PDEs. *Sci Rep* 14:53185. <https://doi.org/10.1038/s41598-024-53185-y>
34. Mozaffar M, Bostanabad R, Chen W, Ehmann K, Cao J, Bessa MA (2019) Deep learning predicts path-dependent plasticity. *Proc Natl Acad Sci* 116:26414–26420. <https://doi.org/10.1073/pnas.1911815116>
35. Fuhg JN, van Wees L, Obstalecki M, Shade P, Bouklas N, Kasemer M (2022) Machine-learning convex and texture-dependent macroscopic yield from crystal plasticity simulations. *Materialia* 23:101446. <https://doi.org/10.1016/j.mtla.2022.101446>
36. Quey R, Dawson P, Barbe F (2011) Large-scale 3d random polycrystals for the finite element method: generation, meshing and remeshing. *Comput Methods Appl Mech Eng* 200:1729–1745. <https://doi.org/10.1016/j.cma.2011.01.002>
37. Dawson PR, Boyce DE (2015) Fepx—finite element polycrystals: Theory, finite element formulation, numerical implementation and illustrative examples. [arXiv:1504.03296](https://arxiv.org/abs/1504.03296)
38. Maurizi M, Gao C, Berto F (2022) Predicting stress, strain and deformation fields in materials and structures with graph neural networks. *Sci Rep.* <https://doi.org/10.1038/s41598-022-26424-3>
39. Gulakala R, Markert B, Stoffel M (2023) Graph neural network enhanced finite element modelling. *PAMM.* <https://doi.org/10.1002/pamm.202200306>
40. Aifantis EC (1987) The physics of plastic deformation. *Int J Plasticity* 3:211–247. [https://doi.org/10.1016/0749-6419\(87\)90021-0](https://doi.org/10.1016/0749-6419(87)90021-0)
41. Han C-S, Gao H, Huang Y, Nix WD (2005) Mechanism-based strain gradient crystal plasticity-I. Theory. *J Mech Phys Solids* 53:1188–1203. <https://doi.org/10.1016/j.jmps.2004.08.008>
42. Quey R, Kasemer M (2022) The neper/fepx project: free/open-source polycrystal generation, deformation simulation, and post-processing. In: IOP conference series: materials science and engineering, vol 1249, p 012021. <https://doi.org/10.1088/1757-899X/1249/1/012021>
43. Vlassis NN, Sun W (2023) Geometric learning for computational mechanics part II: graph embedding for interpretable multiscale plasticity. *Comput Methods Appl Mech Eng* 404:115768. <https://doi.org/10.1016/j.cma.2022.115768>
44. Gavris GB, Sun W (2024) Topology optimization with graph neural network enabled regularized thresholding. *Extreme Mech Lett* 71:102215. <https://doi.org/10.1016/j.eml.2024.102215>
45. Rusch TK, Bronstein MM, Mishra S (2023) A survey on oversmoothing in graph neural networks. [arXiv:2303.10993](https://arxiv.org/abs/2303.10993)
46. Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, Lin Z, Desmaison A, Antiga L, Lerer A (2017) Automatic differentiation in PyTorch. In: NIPS 2017 Workshop Autodiff, pp –. <https://openreview.net/forum?id=BJJsrmfCZ>, Accessed 28 Oct 2017 (modified: 28 Oct 2017)
47. University S (2024) Sherlock cluster documentation. <https://www.sherlock.stanford.edu/docs>. Accessed: 05 08 2024

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.