Component-Based Machine Learning Paradigm for Discovering Rate-Dependent and Pressure-Sensitive Level-Set Plasticity Models

Nikolaos N. Vlassis
Postdoctoral Research Scientist
Department of Civil Engineering and Engineering
Mechanics,
Columbia University,
New York, NY 10027
e-mail: nnv2102@columbia.edu

WaiChing Sun[1]
Associate Professor
Department of Civil Engineering and Engineering
Mechanics,
Columbia University,
New York, NY 10027
e-mail: wsun@columbia.edu

Conventionally, neural network constitutive laws for path-dependent elastoplastic solids are trained via supervised learning performed on recurrent neural networks, with the time history of strain as input and the stress as output. However, training neural networks to replicate path-dependent constitutive responses requires significantly more data due to the path dependence. This demand on diversity and abundance of accurate data, as well as the lack of interpretability to guide the data generation process, could become major roadblocks for engineering applications. In this study, we attempt to simplify these training processes and improve the interpretability of the trained models by breaking down the training of material models into multiple supervised machine learning programs for elasticity, initial yielding, and hardening laws that can be conducted sequentially. To predict pressure sensitivity and rate dependence of the plastic responses, we reformulate the Hamliton–Jacobi equation such that the yield function is parametrized in a product space spanned by the principal stress, the accumulated plastic strain, and time. To test the versatility of the neural network meta-modeling framework, we conduct multiple numerical experiments where neural networks are trained and validated against (1) data generated from known benchmark models, (2) data obtained from physical experiments, and (3) data inferred from homogenizing sub-scale direct numerical simulations of microstructures. The neural network model is also incorporated into an offline FFT-FEM model to improve the efficiency of the multiscale calculations. [DOI: 10.1115/1.4052684]

1 Introduction

One of the century-old challenges for mechanics researchers is to formulate plasticity theory that predicts the relationship among strain history, plastic deformation, and stress for materials governed by different deformation mechanisms at smaller scales. As plastic deformation accumulates, the dissipation and plastic work may lead the yielding criteria to evolve and cause a variety of hardening/softening mechanisms due to the evolution of microstructures, such as twinning [1], dislocation [2], pore collapse [3], void nucleation [4], and rearranging of particles [5]. Generations of scholars including Coulomb [6], von Mises [7], and Drucker and Prager [8] spent decades to create new plasticity theories to incorporate new causality relations and hypotheses for path-dependent materials. In stress-based plasticity theories, yield function is expressed as a function of stress, internal variables, and the hardening laws (e.g., isotropic, kinematic hardening, rotational, mixed-mode, hardening) as deduced from experimental observations and sub-scale micro-mechanical simulations (e.g., dislocation dynamic, molecular simulations).

In the past decades, new plasticity models are often generated by modifying existing models with different expressions of the yield functions or the hardening laws. For instance, a search in Google Scholar for "modified Johnson-Cook model" and "modified Gurson model" reveals more than 632,000 and 8,350 results, respectively.[2] A majority of these published works are dedicated to manually modifying the original model with new evolution laws or shapes of the yield surface that accompany new physics, new materials, or new insights for more precise predictions. While this conventional workflow has led to numerous improvements in modeling, the more sophisticated models are often inherently harder to tune due to the expansion of parametric space. This expansion does not only make it less feasible to determine the optimal mathematical expressions through a manner trial-and-error effort (even after the causality of the yielding and hardening is known [9]) but also require a more complicated inverse problems to identify the material parameters [10–12].

The recent success of deep neural networks has inspired a new trend where one simply builds a forecast engine by training a network with a pair of strain and stress histories [13,14]. To replicate the history dependence of the plastic deformation, an earlier neural network approach work would employ strain and stress from multiple previous time-steps to predict new stress states [15], whereas more recent works such as Refs. [16,17] employ recurrent neural networks, such as the long short-term memory (LSTM) and gated recurrent neural networks to introduce memory effects. The promise and expectation are that the continuous advancement of neural networks or other machine learning techniques might one day replace the modeling paradigm currently

---

[1]Corresponding author.

[2]as of 7/8/2021.

employed in the engineering industry with superior accuracy, efficiency, and robustness [16,18]. However, the early success in the 1990s and the recent resurrection of optimism about neural network predictions on constitutive responses so far have a limited impact on industry applications. This reluctance is not entirely unjustified. In fact, recent studies and workshops conducted by the US Department of Energy have cited the lack of domain awareness, interpretability, and robustness as some of the major technical barriers for the revolution of artificial intelligence (AI) for scientific machine learning [19]. To facilitate changes in the industry, the trustworthiness of the predictions is necessary and interpretability is a necessary condition to overcome these obstacles [20].

As such, our focus in this work is to explore the possibility of building an interpretable machine learning paradigm capable of serving as the interface between plasticity modelers and artificial intelligence. As such, our focus has shifted from solely using AI to make predictions to building AI to create plasticity theories compatible with domain knowledge, easily interpreted, and capable of not only improving the accuracy but also the robustness of existing models. We propose the training of elastoplasticity models through multiple supervised learning procedures to generate the model components of knowledge separately (i.e., elastic stored energy, yielding function, and hardening laws). The resultant model is the composition of these machine-generated knowledge components and is fully interpretable by modelers.

To achieve this goal, we have recast both rate-independent and rate-dependent plasticity as a Hamilton–Jacobi problem in a parametric space spanned by the principal stress, the accumulated plastic strain, the plastic strain rate, and the real time. Meanwhile, the anisotropy of plastic yielding is achieved by mapping the yield function level sets of the same material under different orientations through a supervised neural network training on a chosen yield function projection basis. These treatments enable us to create a general mathematical description for a large number of existing plasticity models, including von Mises plasticity, Drucker–Prager, and Cam-clay models combined with any possible hardening law as merely special cases of the level set plasticity model. Instead of solving the level set extension problem in the parametric space, we formulate a supervised learning problem to generate the constitutive updates from neural computation and to speed up calculation compared to classical hierarchical multiscale computation [21–24].

More importantly, this new AI-enabled framework represents a new paradigm shift where the goal of machine learning has shifted from merely generating forecast engines for mechanistic predictions to creating interpretable mathematical models that inherently obey physical laws with the assistance of machine learning. The resultant model does not require the usage of recurrent neural networks, it is easier to train, and provides more robust results for blind predictions.

2 Level Set Plasticity

The goal of this section is to extend the previous published work [25] to incorporate pressure sensitivity and rate dependence. The mathematical framework is similar except that the introduction of the pressure dependence and the rate dependence may lead to a higher dimensional space for the Hamilton–Jacobi problem and therefore higher demand on data. These new implications are highlighted in this section. Details of the initial implementation of the level set plasticity model can be referred to Ref. [25]. The algorithm used to generate the yield function level sets and trained the plasticity model neural networks is summarized in Algorithm 1.

Here, we formulate a new learning problem for plasticity that does not complete the training in a single supervised learning, but splits the learning task into multiple smaller components (predicting a stored elasticity energy functional, predicting a yield surface, and introducing a mapping for anisotropy), each constituting one neural network trained for a sub-goal. Then, complex behaviors can be predicted by integrating these networks in a level set plasticity framework. This treatment does not only improve the predictions but

also introduces a learning structure where the casual relations of the individual components are defined without losing the generality of individual model predictions. As pointed out by recent work for interpretable machine learning [20,26], this component design helps promoting both *simulatability* (the ability to internally simulate and reason about the overall predictions) and *modularity* (the ability to interpret portions of the predictions independently) of the AI-generated models.

We introduce a new concept of treating the yield surfaces in the parametric space composed of stress, accumulated plastic strain, and strain rate as a level set. We also discuss the importance of leveraging material symmetries to reduce the data demand for the supervised machine learning problem. Previously, Refs. [25,27] have introduced NURB and machine learning-based interpolations, respectively, to generate yield surfaces with isotropic rate-independent plasticity. The key departure here is the new capacity to generalize the learning algorithm for anisotropic rate-dependent/independent plasticity.

An important factor that dictates whether the training of the machine learning model with limited data could be successful is how material symmetry can be leveraged. For example, the data collection can be significantly reduced for isotropic plasticity, as the principal strain and stress are co-axial. Another important aspect to consider is how to leverage material symmetry to select the coordinate system that represents the same data in the parametric space. For instance, a Euclidean space spanned by the values of the three principal stresses could be sufficient for an isotropic yield function and hence leads to a simpler supervised learning problem than those that use all six stress components. Furthermore, the choice of the coordinate system may affect how one plans to collect data and vice versa. For instance, while it is possible to formulate the level set problem with the principal stress as the Cartesian basis, i.e., $(\sigma_1, \sigma_2, \sigma_3)$, it might be even more efficient to consider the usage of (q, p) stress for experimental data obtained from conventional triaxial tests where only two distinctive principal stress can be controlled. In this latter case, the anisotropy and the dependence of all three invariants of the constitutive responses could not be sufficiently captured from the data gathered by the set of the experiments alone. Hence, increasing the dimensions of the parametric space for the elasticity energy and the yield function would not be beneficial. In the numerical experiments we conducted, we adopt the cylindrical coordinates (see Eq. (1)) for the π-plane orthogonal to the hydrostatic axis, where $\sigma_1 = \sigma_2 = \sigma_3$ (compared to Ref. [28]). This treatment enables us to detect any symmetry on the π-plane that might reduce the dimensions of the data and potentially simplify the training of the neural network with less data.

$$
\begin{Bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \end{Bmatrix} = \boldsymbol{R}' \boldsymbol{R}'' \begin{Bmatrix} \sigma_1'' \\ \sigma_2'' \\ \sigma_3'' \end{Bmatrix}
$$
$$
= \begin{bmatrix} \sqrt{2}/2 & 0 & \sqrt{2}/2 \\ 0 & 1 & 0 \\ -\sqrt{2}/2 & 0 & \sqrt{2}/2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sqrt{2/3} & 1/\sqrt{3} \\ 0 & -1/\sqrt{3} & \sqrt{2/3} \end{bmatrix} \begin{Bmatrix} \sigma_1'' \\ \sigma_2'' \\ \sigma_3'' \end{Bmatrix} \quad (1)
$$

Before translating the yield surface f_Γ data points into a yield function level set, we reduce the dimensionality of the stress point **x** representation. In the case of the isotropic pressure-dependent plasticity, we can reduce the stress representation from six dimensions $\mathbf{x}(\sigma_{11}, \sigma_{22}, \sigma_{33}, \sigma_{12}, \sigma_{23}, \sigma_{13})$ (already reduced from 9 due to balance of angular momentum) to an equivalent three stress invariant representation $\hat{\mathbf{x}}(p, \rho, \theta)$. In this representation, p is the mean pressure and ρ and θ are the Lode's radius and angle, respectively.
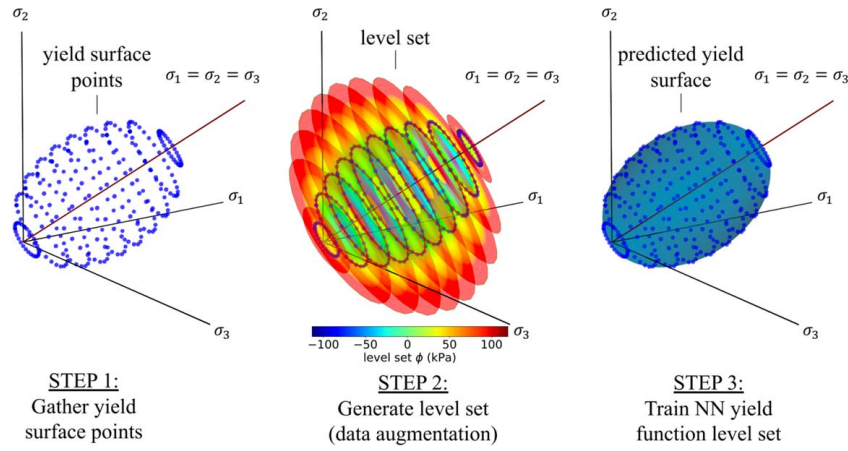
Fig. 1 Universal training process for level set yield functions: (1) gather yield surface data points, (2) generate level set through the initialization process, and (3) train neural network on the level set data (the zeroth level of the predicted level set is the approximated yield surface)

The yield function is then postulated to be a signed distance function defined as follows:

$$\phi(\widehat{\mathbf{x}}, \xi, \dot{\xi}, t) = \begin{cases} d(\widehat{\mathbf{x}}) & \text{outside } f_\Gamma \text{(inadmissible stress)} \\ 0 & \text{on } f_\Gamma \text{(yielding)} \\ -d(\widehat{\mathbf{x}}) & \text{inside } f_\Gamma \text{(elastic region)} \end{cases} \quad (2)$$

where $d(\widehat{\mathbf{x}})$ is the minimum Euclidean distance between any point $\widehat{\mathbf{x}}$ of the solution domain Ω of the stress space where the signed distance function is defined and the yield surface $f_\Gamma = \{\widehat{\mathbf{x}} \in \mathbb{R}^3 | f(\widehat{\mathbf{x}}) = 0\}$, defined as follows:

$$d(\widehat{\mathbf{x}}) = \min(|\widehat{\mathbf{x}} - \widehat{\mathbf{x}}_\Gamma|) \quad (3)$$

where $\widehat{\mathbf{x}}_\Gamma$ is the yielding stress for a given value of accumulated plastic strain ξ and its rate $\dot{\xi}$ at time t. The plastic internal variable ξ is monotonically increasing and represents the history-dependent behavior of the material. The time t signifies a snapshot of the current state of the level set ϕ for the current value of the plastic internal variable ξ and its rate $\dot{\xi}$.

2.1 Data Augmentation Through Signed Distance Function Generation. We can now preprocess the stress point cloud of the yield surface for a given ξ and $\dot{\xi}$ by solving the Eikonal equation $|\nabla^{\widehat{\mathbf{x}}}\phi| = 1$, while prescribing the signed distance function to 0 at $\widehat{\mathbf{x}} \in f_\Gamma$. For every stress point in the yield surface data set, we generate a discrete number of auxiliary points that construct a signed distance function. In the context of level set theory, this can be seen as solving the level set initialization problem. In the context of machine learning, the signed distance function construction can be interpreted as a method of data augmentation—a large number of auxiliary data samples, where $f_\Gamma \neq 0$ are introduced to improve the training performance as well as the accuracy and robustness of both the learned function f_Γ and equally importantly its stress gradient $\partial f_\Gamma / \partial \sigma_{ij}$. A schematic of the yield surface data preprocessing into a signed distance function is demonstrated in Fig. 1. The color is the value of the signed distance yield function. It is negative in the elastic region and positive in the inadmissible stress region. The material yields if the current stress is at the location where the value of yield function equals to zero. It is noted that the signed distance function has been selected as the preferred level set function due to the simplicity of the implementation—the yield function can be formulated as other implicit functions, which will be considered in the future work.

2.2 Hardening as a Level Set Extension Problem. After preprocessing the yield surface f_Γ data points for a sequence of internal variable values ξ and rates $\dot{\xi}$ into a level set by solving the level set initialization problem, we will recover the velocity function of a Hamilton-Jacobi equation of a level set extension problem to describe the temporal evolution of the level set. A general Hamilton-Jacobi equation reads:

$$\frac{\partial \phi}{\partial t} + \mathbf{v} \cdot \nabla^{\widehat{\mathbf{x}}} \phi = 0 \quad (4)$$

where $\mathbf{v}$ is the normal velocity field that describes the geometric evolution of the boundary (yield surface f_Γ). In the context of plasticity, the velocity field corresponds to the observed hardening mechanism. The velocity vector field can be described by a magnitude scalar function F and a direction vector field $\mathbf{n} = \nabla^{\widehat{\mathbf{x}}}\phi / |\nabla^{\widehat{\mathbf{x}}}\phi|$ such that:

$$\mathbf{v} = F \cdot \mathbf{n} \quad (5)$$

Substituting into Eq. (4):

$$\frac{\partial \phi(\xi, \dot{\xi})}{\partial t} + F(\xi, \dot{\xi})|\nabla^{\widehat{\mathbf{x}}}\phi(\xi, \dot{\xi})| = 0,$$

$$\text{where } F_i \approx \frac{\phi_{i+1}(\xi_{i+1}, \dot{\xi}_{i+1}) - \phi_i(\xi_i, \dot{\xi}_i)}{\Delta t} \quad (6)$$

In Eq. (6), $F_i(p, \rho, \theta, \xi, \dot{\xi}) = F(p, \rho, \theta, \xi, \dot{\xi}, t_i)$ for $i = 0, 1, 2, \ldots, n + 1$ is the finite difference approximated scalar velocity (hardening) function that corresponds to the preprocessed collection of signed distance functions $\{\phi_0, \phi_1, \ldots, \phi_{n+1}\}$ at time $\{t_0, t_1, \ldots, t_{n+1}\}$. Thus, we have recast a yield function f into a signed distance function ϕ, such that $f(p, \rho, \theta, \xi, \dot{\xi}) = \phi(p, \rho, \theta, \xi, \dot{\xi})$. We can now formulate a machine learning problem to approximate the level set yield function f with its neural network yield function $\widehat{f} = \widehat{f}(p, \rho, \theta, \xi, \dot{\xi}|\mathbf{W}, \mathbf{b})$ counterpart, parametrized by weights $\mathbf{W}$ and biases $\mathbf{b}$ to be optimized during training.

The training objective for the neural network optimization is to minimize the following loss function at training samples $(\hat{\mathbf{x}}, \xi, \dot{\xi}, t)$ for $i \in [1, \ldots, N]$:

$$W', b' = \underset{W,b}{\text{argmin}} \left(\frac{1}{N} \sum_{i=1}^{N} \left(\left\| f_i - \widehat{f}_i \right\|_2^2 + w_p \, \text{sign}\left(-\sum_{A=1}^{3} \sigma_{A,i} \frac{\partial \widehat{f}_i}{\partial \sigma_{A,i}} \right) \right) \right) \quad (7)$$

where we have added a penalty term, weighted by a factor w_p, that will activate when the yield function is not obeying convexity during training.

It is noted that the Hamilton-Jacobi equation described in this section will not be solved numerically—while theoretically possible (e.g., fast marching solver). Its solution will be directly predicted by a neural network. The zeroth level of the neural network predicted level set is the yield surface. The neural network approximated velocity field is the data-driven hardening mechanism.

Remark 1 Rescaling of the training data . In every loss function in this study, we have introduced scaling coefficients γ_α to remind the readers that it is possible to change the weighting to adjust the relative importance of different terms in the loss function. These scaling coefficients may also be viewed as the weighting function in a multi-objective optimization problem. In practice, we have normalized all data to avoid the vanishing or exploding gradient problem that may occur during the back-propagation process [29]. As such, normalization is performed before the training as a preprocessing step. The X_i sample of a measure X is scaled to a unit interval via

$$\overline{X_i} := \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \tag{8}$$

where $\overline{X_i}$ is the normalized sample point. $X_{\min}$ and $X_{\max}$ are the minimum and maximum values of the measure X in the training data set such that all different types of data used in this paper (e.g., energy, stress, stress gradient, stiffness) are all normalized within the range [0, 1]. ∎

Algorithm 1   Training of a pressure and rate-dependent isotropic yield function level set neural network.

Require: Data set of N samples: stress measures $\boldsymbol{\sigma}$ at yielding, accumulated plastic strain $\overline{e}_p$, and accumulated plastic strain rate $\dot{\overline{e}}_p$, a L_{levels} number of levels (isocontours) for the constructed signed distance function level set (data augmentation), and a parameter $\zeta > 1$ for the radius range of the constructed signed distance function.

1. Project stress onto π-plane
   Initialize empty set of π-plane projection training samples (ρ_i, θ_i, p_i) for i in $[0, ..., N]$.
   for i in $[0,...,N]$ **do**
       Spectrally decompose $\boldsymbol{\sigma}_i = \sum_{A=1}^{3} \sigma_{A,i} \boldsymbol{n}_i^{(A)} \otimes \boldsymbol{n}_i^{(A)}$.
       Transform $(\sigma_{1,i}, \sigma_{2,i}, \sigma_{3,i})$ into $(\sigma''_{1,i}, \sigma''_{2,i}, \sigma''_{3,i})$ via Eq. (1)
       $\rho_i \leftarrow \sqrt{\sigma''^2_{1,i} + \sigma''^2_{2,i}}$
       $\theta_i \leftarrow \tan^{-1}\left(\frac{\sigma''_{2,i}}{\sigma''_{1,i}}\right)$
       $p_i \leftarrow \frac{3}{\sqrt{3}} \sigma''_{3,i}$
   end for
2. Construct yield function level set (data augmentation)
   Initialize empty set of augmented training samples $(\rho_m, \theta_m, p_m, \overline{e}_{p,m}, \dot{\overline{e}}_{p,m}, f_m)$ for m in $[0, ..., N \times L_{\text{levels}}]$.
   $m \leftarrow 0$.
   for i in $[0, ..., N]$ **do**
       for j in $[0, ..., L_{\text{levels}}]$ **do**
       $\rho_m \leftarrow \left(\frac{\zeta j}{L_{\text{levels}}}\right)\rho_i$ ▷ the signed distance function is constructed for a radius range of $[0, \zeta\rho_i]$
       $\theta_m \leftarrow \theta_i$
       $p_m \leftarrow p_i$
       $\overline{e}_{p,m} \leftarrow \overline{e}_{p,i}$
       $\dot{\overline{e}}_{p,m} \leftarrow \dot{\overline{e}}_{p,i}$
       $f_m \leftarrow \left(\frac{\zeta j}{L_{\text{levels}}}\right)\rho_i - \rho_i$ ▷ the signed distance function value range is $[-\rho_i, (\zeta - 1)\rho_i]$
       Rescale $(\rho_m, \theta_m, p_m, \overline{e}_{p,m}, \dot{\overline{e}}_{p,m}, f_m)$ into $(\overline{\rho_m}, \overline{\theta_m}, \overline{p_m}, \overline{\overline{e}_{p,m}}, \overline{\dot{\overline{e}}_{p,m}}, \overline{f_m})$ via Eq. (8).
       $m \leftarrow m + 1$
       end for
   end for
3. Train neural network $\widehat{f}(\overline{\rho_m}, \overline{\theta_m}, \overline{\overline{e}_{p,m}}, \overline{\dot{\overline{e}}_{p,m}})$ with loss function Eq. (7).
4. Output trained yield function $\widehat{f}$ neural network and exit.

2.3   Higher-Order Sobolev Training. In this study, we distinguish between the material's elastic and plastic behaviors by training two different neural network model components—a hyperelastic energy functional and a yield function level set that evolves according to accumulated plastic strain. These components are then combined in a specific form of return mapping algorithm (Algorithm 1) that may take an arbitrary elasticity model and a yield function with generic hardening law to generate the constitutive update for the class of inelastic materials that has a distinct elastic region defined in a parametric space. The hyperelastic network counterpart is expected to have interpretable derivatives —the first derivative of the energy functional with respect to the strain should be a valid stress tensor and the second derivative a valid stiffness tensor. We adopt a Sobolev training objective, first introduced in Ref. [30], and we extend it to higher-order constraints to train the energy functional approximator $\widehat{\psi}^{\text{e}}(\boldsymbol{\epsilon}^{\text{e}} \mid \boldsymbol{W}, \boldsymbol{b})$ using the following loss function:

$$W', b' = \underset{W,b}{\text{argmin}} \left(\frac{1}{N} \sum_{i=1}^{N} \left(\|\psi_i^{\text{e}} - \widehat{\psi}_i^{\text{e}}\|_2^2 + \left\|\frac{\partial \psi_i^{\text{e}}}{\partial \epsilon_i^{\text{e}}} - \frac{\partial \widehat{\psi}_i^{\text{e}}}{\partial \epsilon_i^{\text{e}}}\right\|_2^2 \right.\right.$$
$$\left.\left. + \left\|\frac{\partial^2 \psi_i^{\text{e}}}{\partial \epsilon_i^{\text{e}} \otimes \partial \epsilon_i^{\text{e}}} - \frac{\partial^2 \widehat{\psi}_i^{\text{e}}}{\partial \epsilon_i^{\text{e}} \otimes \partial \epsilon_i^{\text{e}}}\right\|_2^2 \right)\right) \tag{9}$$

A benefit of using Sobolev training is the notable data efficiency. Sobolev training has been shown to produce more accurate and smooth predictions for the energy, stress, and stiffness fields for the same amount of data compared to classical L^2 norm approaches that would solely constrain the predicted energy values [25].

3   Numerical Experiments

In this section, we demonstrate the AI's capacity to rediscover plasticity models from the literature, we explore the model's ability to capture highly complex new hardening modes and, finally, showcase how the AI can discover the yield surface for a new polycrystal material and replace the plasticity model in a finite element simulation. To test whether the machine learning approach can be generalized, we purposely test the AI against a wide range of material data sets for soil, rock, polycrystal, and steel. In particular, we employ three types of data sets, (1) data generated from known literature models, (2) data obtained from experiments, and (3) data obtained from sub-scale direct numerical simulations of microstructures. The first type of data is used as a benchmark to verify whether the neural network can correctly deduce the correct plastic deformation mechanisms (yield surface and hardening) when given the corresponding data. The second and third types of data are used to validate and examine the AI's ability to discover new plastic deformation mechanisms with a geometrical interpretation in the stress space.

3.1   Verification Examples. The first set of examples   is used to showcase the capacity of the algorithm to reproduce the modeling capacity of classical plasticity theory. We first demonstrate our algorithms ability to recover yield surface and hardening mechanisms from the classical plasticity literature. We then demonstrate the frameworks capacity to make predictions calibrated on experimental data for pressure-dependent and rate-dependent plasticity.

3.1.1   Verification on Classical Plasticity Theories. The proposed AI can readily reproduce numerous yield function models from the plasticity literature, following the same universal data preprocessing and neural network training algorithm. For this benchmark experiment, we generate synthetic data sets for four initial yield surfaces of increasing shape complexity: the J2 [7] (cylinder), Drucker–Prager [8] (cone), Modified Cam-Clay [31] (oval), and Argyris [32] (ovoid with triangular cross section) yield surfaces. We simultaneously study four common hardening mechanisms
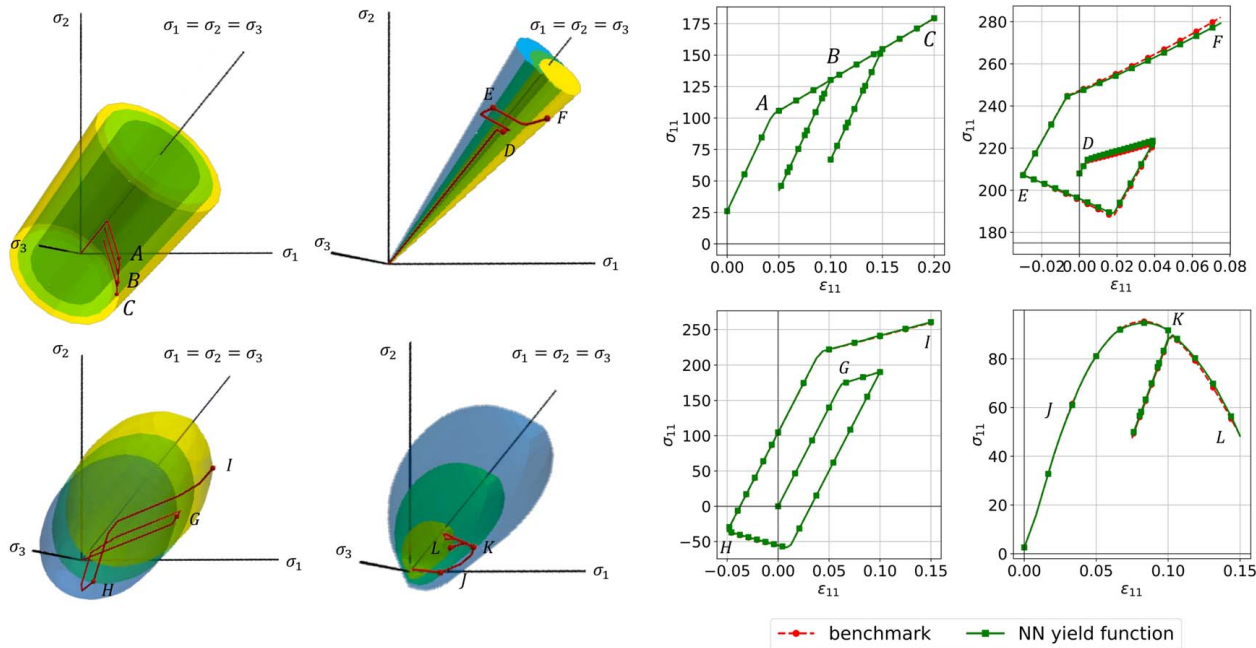
Fig. 2 AI can rediscover classical plasticity models: J2 plasticity model with isotropic hardening (top left), Drucker–Prager model with rotational hardening (top right), MCC model with kinematic hardening (Bauschinger effect) (bottom left), and Argyris model with softening (bottom right). The corresponding benchmark and predicted strain–stress curves are also demonstrated. The stress measure is in kPa.

that transform and/or translate these surfaces in the 3D stress space: isotropic hardening (cylinder dilation), rotational hardening (cone rotation), kinematic hardening (translation along the hydrostatic axis), and softening (shrinking).

The data sets for these yield surfaces are populated by sampling from the above-mentioned literature yield functions. The sampling was performed as a uniform grid of the stress invariants and the accumulated plastic strain. We sample 50 data points along the mean pressure axis, 100 data points along the angle axis, and 10 data points along the accumulated plastic strain axis (a total of 50,000 data samples per yield function data set). The yield surface data points are preprocessed into a signed distance function level set database through the level set initialization procedure. For each yield surface, 15 levels are constructed: the yielding level, 7 in the elastic region, and 7 in the region of inadmissible stress. After data augmentation, the training data set consists of 750,000 level set sample points.

For each level set database, we train a feed-forward neural network to approximate the initial yield function and its evolution. The yield function neural networks consist of a hidden dense layer (100 neurons/ReLU), followed by two multiply layers, then another hidden dense layer (100 neurons/ReLU) and an output dense layer (linear). The use of multiply layers was first introduced in Ref. [25] to increase the continuity of the activation functions of neural network functional approximators. They were shown to allow greater control over the network's higher-order derivatives and the application of higher-order Sobolev constraints in the loss function. The layers' kernel weight matrix was initialized with a Glorot uniform distribution and the bias vector with a zero distribution. All the models were trained for 2000 epochs with a batch size of 128 using the NAdam optimizer, set with default values of the Keras library.[3]

The neural network predicted yield surfaces are demonstrated in Fig. 2. For each model, three surfaces are shown for three different levels of accumulated plastic strain. It is highlighted that, given an

accumulated plastic strain value, we can recover the entire yield locus.

3.1.2 Level Set Plasticity Model Discovery for Rate-Dependent and Anisotropic Materials.
In this section, we test the framework capacity to make predictions on rate-dependent and anisotropic data.

To test the trained neural network prediction of rate-dependent responses, we incorporate data from the published work [33] for steel that exhibits different yielding stress under different strain rates. In the numerical experiments, we use experimental data collected at strain rates ranging from 0 to $0.02\,\mathrm{s}^{-1}$ as the training data, sampled in a uniform grid of ten strain rate increments. The yield surface is sampled at 25 points along the mean pressure axis, at 100 points along the angles axis, and at 10 points along the accumulated plastic strain axis (a total of 250,000 sample points). The data are preprocessed into signed distance functions of 15 levels, generating 3,750,000 training sample points. The
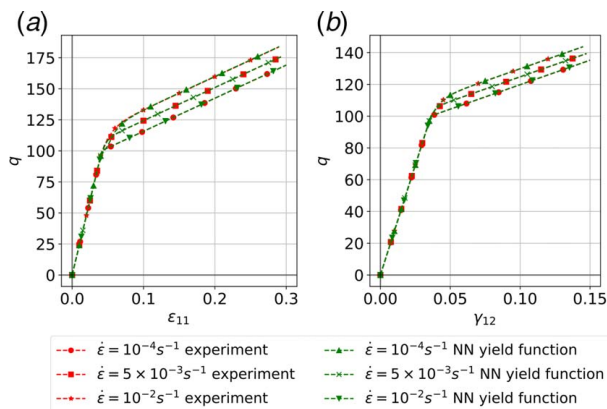

Fig. 3 Neural network predicted viscoplastic response for increasing loading strain rates for a tension (a) and shear (b) loading test performed on mild-steel beams (experimental data obtained from Ref. [33])
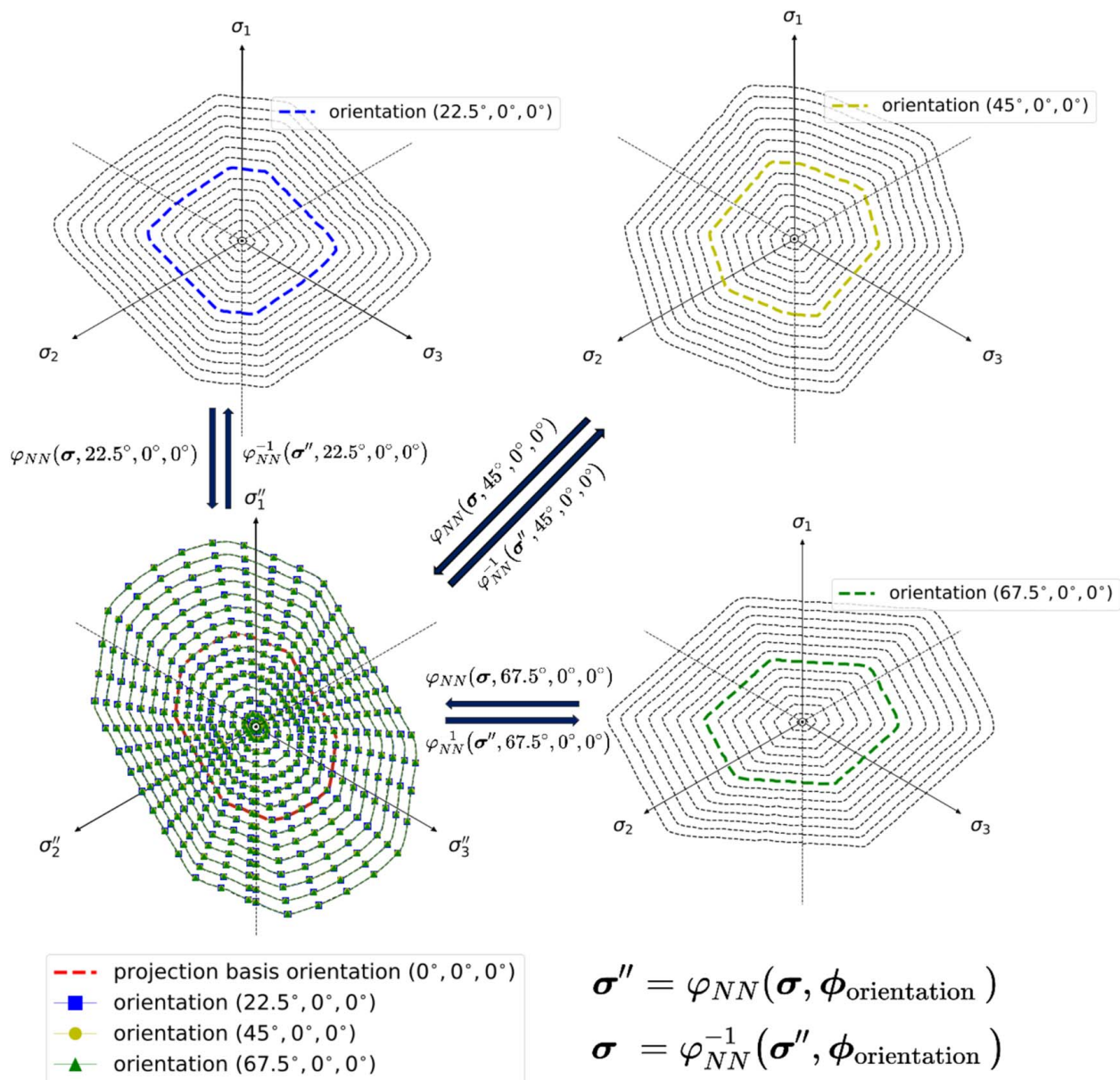

Fig. 4 The framework can capture anisotropic responses by projecting anisotropic yield surfaces onto a master projection basis curve using a neural network stress space mapping φ_{NN}

neural network used for this viscoplastic model training follows the same architecture as the yield function neural networks described in Sec. 3.1.1.

We use the experimental data collected at strain rates 10^{-4}, 5×10^{-1}, and $0.02\ \text{s}^{-1}$ to validate the ability of the model to make blind predictions for unseen events. Figure 3(a) shows the results of the six predictions that the AI generated for unseen data. The left figure shows the stress–strain predictions on the uniaxial tensile tests of three different loading rates, while the right figure shows the stress–strain predictions on the simple shear test counterpart. In both cases, the predictions match well with the unseen benchmark data that is excluded from the training data set.

As for the anisotropic predictions, Fig. 4 shows the machine learning generated mapping that predicts how the yield surface in the principal stress space evolves for different material orientations. The data we employed in this second experiment are generated from an FFT solver that simulates the polycrystal plasticity of a specimen composed of FCC crystal grains. We sample the material constitutive behavior at ten microstructure orientations at 150 lode angle sampling directions and preprocess the data into signed distance functions of 15 levels, generating 22,500 training sample points for the projection mapping neural network. Working on the pressure-independent stress space, the network inputs the true stress invariants and the microstructural orientation information that describes the anisotropy—in the case of polycrystals studied in this work, the polycrystal orientations as three Euler angles, and outputs the reference stress space invariants. The network has the following layer structure: dense layer (200 neurons/ReLU), multiply layer, dense layer (200 neurons/ReLU), multiply layer, followed by three more dense layers (200 neurons/ReLU), and an output dense layer (linear). The layers' kernel weight matrix was initialized with a Glorot uniform distribution and the bias vector with a zero distribution. The model was trained for 2000 epochs with a batch size of 256 using the NAdam optimizer. The predictions of the mapping suggest that it is possible to generate a single mapping function that maps all yield surfaces obtained from different polycrystal specimens of different orientation onto a reference stress domain, denoted as (σ_1'', σ_2'', σ_3'').
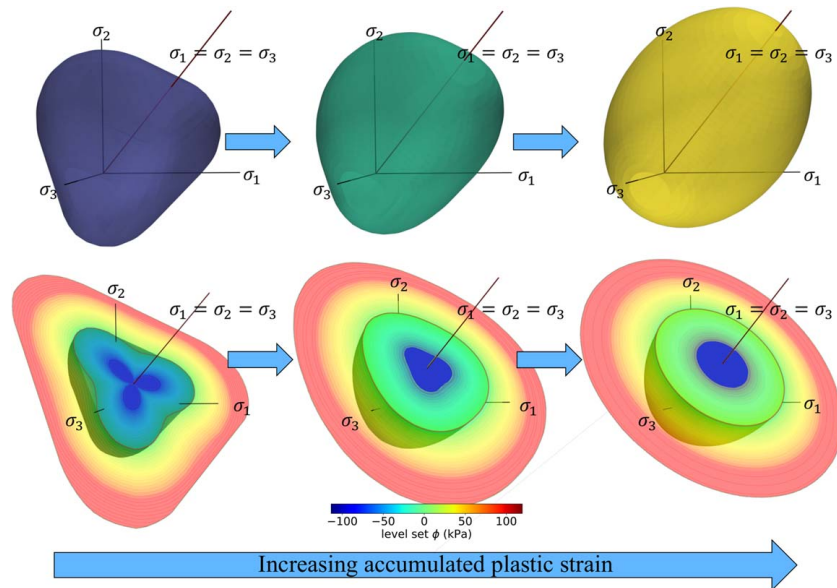
Fig. 5 AI discovered yield surfaces and hardening mechanisms as evolving level sets from synthetic data. The yield surface and corresponding yield function level set are evolved according to the accumulated plastic strain.

3.2 Demonstration of Model Discovery Capacity. Yield surface discovery in the literature has been limited by the difficulty of deriving mathematical expressions for higher-complexity geometrical shapes that represent them. Additional obstacles arise when there is need to describe the smooth transition from the shape of the initial yield surface to that of a state with more accumulated plasticity. The algorithm's capability to discover new yield surfaces and hardening mechanisms automatically directly from the data overcomes these impediments.

To test this, we construct a fictitious yield surface database that is based on the Argyris model [32] and combines the modified Cam-Clay [31] hardening mechanism along with a transformation of the elastic region's cross section from a triangular shape to a circle. The yield surface is sampled at a total of 50,000 points and preprocessed to generate 750,000 level set sample points. The predictions for the yield surface and underlying level set for increasing accumulated plastic strain are demonstrated in Fig. 5. Deriving a mathematical expression for this data set is not straightforward. Even if the derivation is successful, the resultant mathematical expression might require additional material parameters that lacks physics underpinning. The capability of the neural network to approximate arbitrary function therefore offers us a flexible and simple treatment to handle the evolution of yield function.

To analyze the sensitivity with respect to the random neural network weight initialization, we have repeated the supervised learning for the synthetic yield function problem showcased in Fig. 5 five times, each with a different random seed. The results shown in Fig. 6, indicate that the training and the resultant losses for both the training and testing cases are close. This result suggests that the training is not very sensitive to the random seeds. Furthermore, the small difference in the training and testing loss of the yield function also suggests that there is no significant overfitting.

Our proposed algorithm also automates the discovery of yield surfaces for new materials. We generate a yield surface database for a randomly generated polycrystal microstructure through efficient data sampling of the invariant stress space with FFT solver elastoplastic simulations. To gather the yield surface data points for the polycrystal material, we subdivide the π-plane uniformly at 140 Lode's angles and sample the stress space with monotonic loading simulations at each angle direction. The yield surface data points are gathered as soon as yielding is first detected, recording the stress response and the accumulated plastic strain. The FFT simulations provide 157,500 sample points that are preprocessed into 2,362,500 level set sample points. It is noted that the material was observed to be pressure independent. Thus, sampling on the π-plane at a constant mean pressure was enough to capture the entire stress response.

The yield surface data points are preprocessed into a level set data base, and the results of the trained polycrystal neural network yield function are demonstrated in Fig. 5. The neural network parameters for the new model training in this section remain identical as previously described. Investing the modeling effort to describe the complex yielding behavior of a material could be proven futile, especially if the material is highly heterogeneous. Conceiving a new yield function for every new material studied can become rather impractical and automation in yield surface generation can accelerate the plasticity study of novel materials.

3.3 Offline Multiscale FFT-FEM Numerical Experiments. In engineering practice, a constitutive law is seldom used as a standalone forecast engine but is often incorporated into a solver that provides a discretized numerical solution. Here, we test whether the AI-generated models can be deployed into an existing finite element solution. The yield surface neural networks combined with a hyperelastic energy functional neural network can be readily plugged into a strain space return mapping algorithm to make strain–stress predictions. In this study, we utilize a linear elasticity energy functional as the neural network that will provide the elastic response in the algorithm. We train a two layer feed-forward neural network that inputs the elastic volumetric ϵ_v^e and deviatoric ϵ_s^e strain invariants to approximate the hyperelastic energy funtional ψ^e. The network is trained on 2500 data points sampled from a uniform grid of $\left(\epsilon_v^e, \epsilon_s^e\right)$ pairs. The architecture consists of a hidden dense layer (100 neurons/ReLU), followed by two multiply layers, then another hidden dense layer (100 neurons/ReLU), and an output dense layer (linear). The models were trained for 1000 epochs with a batch size of 32 using the NAdam optimizer [34], set with default values in the Keras library. By using a Sobolev training framework, the model was optimized with a higher-order H^2 training objective—the loss function constrains the predicted energy, stress, and stiffness similar to (9). The resulting stress predictions for the literature yield surfaces for random cyclic loading
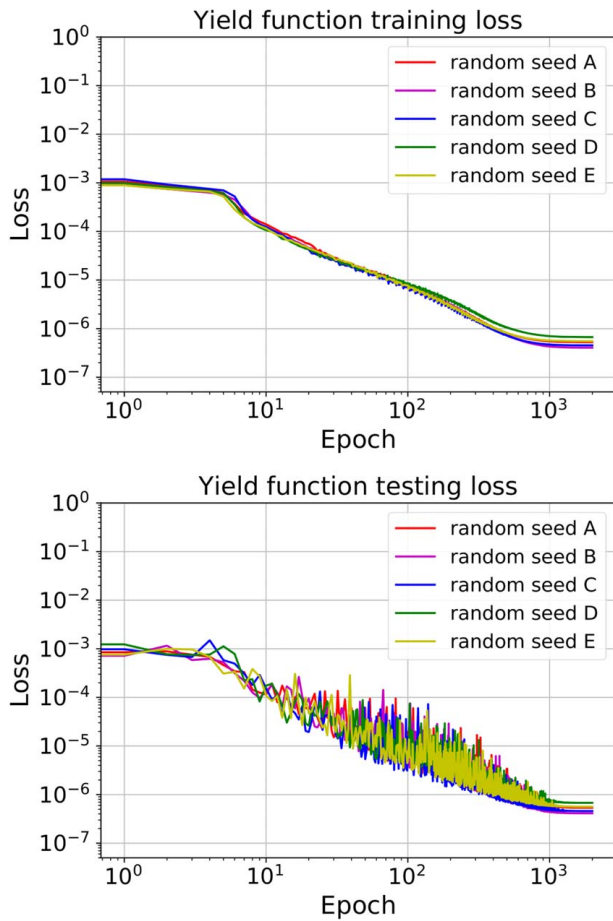
Fig. 6 Loss versus epoch for the synthetic yield function shown in Fig. 5 for the training data set (top) and the testing data set for cross-validation (bottom). The test data set and the training data set are mutually exclusive.

and unloading strain paths are demonstrated in Fig. 2 for each approximated yield surface model.

We have also successfully incorporated the trained neural network plasticity model into a finite element solver to deliver an excellent match with the high-cost FFT-FEM predictions for unseen loading paths not included in the training data set. The discovered yield function for a randomly generated polycrystal microstructure is demonstrated in Fig. 7. In Fig. 8, the polycrystal
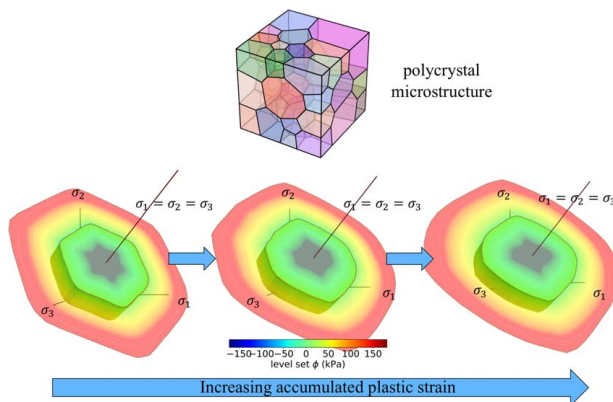
Fig. 7 Yield function level set of a new polycrystal microstructure with increasing accumulated plastic strain
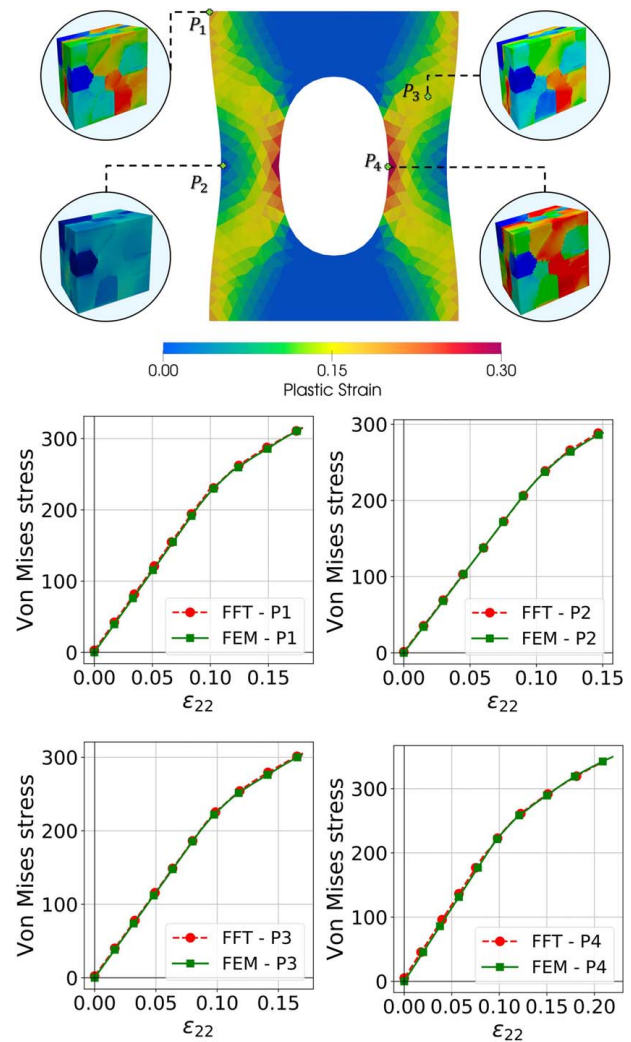

Fig. 8 The discovered yield function can be readily implemented in FEM simulations, replacing the FFT solver. The accumulated plastic strain profile for an FEM simulation and the predicted stress responses at different points of the domain against the FFT benchmark simulations are also shown. The stress measure is in kPa.

plasticity model trained by a neural network is used to replace the FFT solver that provides the constitutive updates from DNS simulations at the sub-scale level. The simulation is performed on a square plate with a circular hole supported on frictionless rollers on the top and bottom surfaces. Results shown in Fig. 8 indicate that the NN-FEM model is capable of replacing the computationally heavy FFT-FEM simulations (compared to Ref. [35]) with a fraction of the cost.

In this offline multiscale problem, the finite element contains 960 elements with 2880 integration points. An FFT-FEM framework may take an average of 11,110 (approximately 3.85 s per integration point) seconds to complete the incremental constitutive updates for all integration points, whereas the neural network counterpart require an average of 230 s (approximately 0.08 s per integration point) to finish the same task in a MacBook Pro with 8-core CPU. As for the overhead cost to generate the training data from the FFT polycrystal simulations, the time to generate the training data set for the polycrystal yield function (157,500 yield function sample points) is approximately 5 h.

4 Discussion

The proposed algorithm provides a general approach of discovering complex yield surface shapes and their evolution directly from data. In Sec. 3 of this work, all yield functions and hardening mechanisms are predicted by neural networks without any specific modeler intervention with hand-crafted derivations. All models in this work, be it models from the plasticity literature or models designed for new materials, followed identical data preprocessing, neural network training, and return mapping implementation procedures.

Our neural network yield functions provide a unique advantage in crafting interpretable data-driven plasticity models. The capacity to predict and visualize the entire yield locus at every time-step of an elastoplastic simulation allows for the anticipation of elastic or plastic responses and the inspection of thermodynamic consistency (e.g., convexity). Especially, by adopting a low-dimensional stress representation (Lode's coordinates), not only is the model complexity reduced but also a transparent yield surface data sampling scheme becomes possible. The alternative of random sampling of strain paths comes with the uncertainty of sufficiently visiting the yield surface in the entire stress space.

4.1 Physics Underpinning for the Partition of Elastic and Plastic Strain.
Decomposing the elastoplastic behavior prediction into two simple feed-forward neural networks—a hyperelastic energy functional and a yield function—is central to the algorithm's interpretability and allows for a clear-cut distinction of elastic and plastic behavior. This is not necessarily true with the classical recurrent network approach, such as the common LSTM or GRU [16,36]. When training neural networks with these architectures, the elastic and plastic constitutive responses are often indistinguishable. This treatment does not only cause issues with interpretability but also renders the black-box predictions vulnerable to erroneous causality or correlation structures. For instance, experimental data of the nonlinear elasticity response may actually affect the yielding response as there is no explicit mechanism to distinguish the two. The models trained on monotonic loading data can readily predict nonmonotonic constitutive responses due to the explicitly defined elastic range, whereas the black-box alternative cannot (see Fig. 2). Furthermore, the recurrent network's dependency on the input strain rate, the importance of the sampling frequencies in the time domain, and the more difficult training due to the vanishing or exploding [37] are rarely addressed in the machine learning plasticity literature.

Note that the machine learning algorithm proposed here does not exhibit better interpretability than the hand-crafted counterpart, but is easier to interpret than the RNN and the multi-step ANN approaches that do not provide definite distinction between the elastic region and the yielding. An exception is the recent work by Xu et al. [38], in which the neural network partitioned the total strain into elastic and plastic components via a partition-of-unity function. Nevertheless, when a continuous weighting function (such as sigmoid function) is used to partition for the elastic and plastic strain, it may introduce a transition zone where the materials are considered both path independent and path dependent.

4.2 Representation of Parametric Space and Geometrical Interpretation of Elastoplasticity Models.
Another advantage of the interpretable machine learning approach is that the geometrical interpretation is helpful for determining the optimal data exploration strategies. Given the fact both real experiments and direct numerical simulations are often costly to conduct, a Monte Carlo simulation to randomly sample the parametric space for path-dependent materials is too costly to be feasible [39]. By introducing the level set to define the yielding criterion, however, we can conceptualize the elastic range as a multidimensional object in a Euclidean space.

This feature may help us to visual the abstract concept of yielding on a Euclidean space and help us estimating the sufficiency of the data by defining a proper metrics in the parametric vector space and to decide the distribution of data that helps us better capture important features such as replicating sharp gradient, determining convexity by checking the Hessian and ensuring connectivity of the learned models. These tasks are not necessarily impossible but are difficult to achieve with a black-box model.

4.3 Smoothness of the Machine Learning Plasticity Model.
Training the neural networks of this work with a higher degree of continuity activation functions and higher-order Sobolev loss function constraints allows one to control the prediction accuracy of the derivatives of the approximated functionals. This control of stress gradient to the yield function is crucial, whereas the automatic differentiation used in the back-propagation can help us generate sufficiently smooth elastoplastic tangent operators suitable for PDE solvers. Conversely, classical black-box neural network elastoplasticity approaches usually do not control the quality of the derivatives of the trained functions. While finite difference methods can be used to approximate the tangent tensor obtained from the neural network without Sobolev training if necessary [40], the smoothness and accuracy of the approximated tangent cannot be guaranteed. Furthermore, the Sobolev training and higher-order activation functions allow controlling the smoothness and continuity of the yield surface. This can be a more efficient alternative to the current practice where a plasticity model with a nonsmooth yield surface either requires specific algorithmic algorithm to generate incremental constitutive updates [41] or modified manually into a smoothed version to bypass the numerical barrier [42,43].

In principle, the approach may generate a sufficiently smooth yield surface in parametric space of different dimensions (e.g., principal stress space, strain space, porosity-stress space). However, if the yield surface is nonsmooth for physical reasons, then both (1) specific supervised learning algorithms that detect the singular point and (2) the corresponding specific treatment to handle the bifurcated stress gradient of the yield surface are necessary. Furthermore, unlike the classical hand-crafted model or models generated from geometric learning (see Ref. [44]) that are designed for an entire class of materials of similar but distinctive microstructures, the proposed algorithm is designed to generate a surrogate model specifically tailored for one RVE or specimen.

4.4 Comparison With Parameter Identification of Predetermined Models.
Note that, while both parameter identification and supervised machine learning involve solving inverse problems and, in many cases, multi-objective optimization, the proposed approach does not assume specific forms of equations a priori for the hyperelasticity energy functional and yield function. With a sufficient neural network architecture, the neural network approach may offer more flexibility in finding the optimal forms of equations (see universal approximation theorem [45]). However, this flexibility comes at the expense of having to deal with the Banach space (compared with the studies by Parhi and Nowak [46] and Weinan and Wojtowytsch [47]) of much higher dimensions (of the neural network learned function) than the Euclidean space for a typical parameter identification problem.

A similar analogy can be drawn between nonparametric/symbolic regression and polynomial regression where the lack of predetermined form of the former approach not only offers greater flexibility but also increases the difficulty of the inverse problem. As demonstrated in the previous work (compared with Wang et al. [48]), even in the case where the inverse problem is merely used to determine the optimal set of choices among a handful of predetermined components of the elastoplasticity model, the additional effort and cost to solve the combinatorial optimization on top of the CPU time required to identify the parameter identification process can be enormous.

This complexity motivates us to propose this alternative paradigm that enable us to learn the elastoplasticity problem in a divide-and-conquered manner, i.e., (1) learning the elasticity first, (2) then the initial yield function, and (3) the hardening/softening rules that evolve the yield function, all with multilayer perceptrons. In the cases we demonstrated here, there is no need to use recurrent neural networks that are more difficult to train well and apply regularization than the simpler multilayer perceptrons [49]. In the future, we may explore proper ways to generate more complex rules for the yield function evolution with recurrent neural networks, but this is out of the scope of this study.

5 Conclusion

We propose a generalized machine learning paradigm capable of generating pressure-sensitive and rate-dependent plasticity models consisting of interpretable components. The component approach enables geometrical interpretation of the hyperelastic energy and yield function in the corresponding strain and stress spaces. This treatment allows us to examine thermodynamic constraints through geometrical interpretation (e.g., convexity) and provide a higher degree of modularity and simulatability required to interpret mechanisms of plastic deformation. In the numerical experiments presented in this article, we first verify the capacity of the paradigm to recover existing plasticity models with the corresponding data. Then, we provide additional examples to show that the revised Hamilton-Jacobi solution formulated for rate-dependent plasticity may generate models from experimental data for steel. Finally, the machine learning paradigm is used to generate a macroscopic elastoplasticity surrogate model from FFT simulations of a polycrystal consisting of FCC grains. The resultant macroscopic surrogate model is tested against FFT direct numerical simulations at the Gauss point level. From the results of the numerical experiments, we observe that the generated models are not only able to recover old plasticity laws but also capable of deducing new ones, with a reasonable level of predictive and descriptive accuracy for the given amount of data. This interpretability is necessary for ensuring trustworthiness for engineering applications.

Conflict of Interest

There are no conflicts of interest.

Data Availability Statement

The datasets generated and supporting the findings of this article are obtainable from the corresponding author upon reasonable request. The authors attest that all data for this study are included in the paper.

References

[1] Cheng, Z., Zhou, H., Lu, Q., Gao, H., and Lu, L., 2018, "Extra Strengthening and Work Hardening in Gradient Nanotwinned Metals," Science, **362**(6414), p. eaau1925.
[2] Van der Giessen, E., and Needleman, A., 1995, "Discrete Dislocation Plasticity: A Simple Planar Model," Modell. Simul. Mater. Sci. Eng., **3**(5), p. 689.
[3] Aydin, A., Borja, R. I., and Eichhubl, P., 2006, "Geological and Mathematical Framework for Failure Modes in Granular Rock," J. Struct. Geol., **28**(1), pp. 83–98.
[4] Gurson, A. L., 1977, "Continuum Theory of Ductile Rupture by Void Nucleation and Growth: Part I—Yield Criteria and Flow Rules for Porous Ductile Media," J. Eng. Mater. Technol., **99**(1), pp. 2–15.
[5] Martin, C., Bouvard, D., and Shima, S., 2003, "Study of Particle Rearrangement During Powder Compaction by the Discrete Element Method," J. Mech. Phys. Solids., **51**(4), pp. 667–693.
[6] Coulomb, C. A., 1776, Essai Sur Une Application Des Regles De Maximis Et Minimis Quelques Problemes De Statique, Relatifs Lárchitecture, Memoires de Mathematique de l'Academie Royale de Science 7, Paris.
[7] Mises, R. v., 1913, "Mechanik Der Festen Körper Im Plastisch-deformablen Zustand," Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse, **1913**, pp. 582–592.
[8] Drucker, D. C., and Prager, W., 1952, "Soil Mechanics and Plastic Analysis or Limit Design," Q. Appl. Math., **10**(2), pp. 157–165.
[9] Sun, X., Bahmani, B., Vlassis, N. N., Sun, W., and Xu, Y., 2021, "Data-Driven Discovery of Interpretable Causal Relations for Deep Learning Material Laws With Uncertainty Propagation," Granular Matter.
[10] Ehlers, W., and Scholz, B., 2007, "An Inverse Algorithm for the Identification and the Sensitivity Analysis of the Parameters Governing Micropolar Elasto-Plastic Granular Material," Arch. Appl. Mech., **77**(12), pp. 911–931.
[11] Wang, K., Sun, W., Salager, S., Na, S., and Khaddour, G., 2016, "Identifying Material Parameters for a Micro-Polar Plasticity Model Via X-Ray Micro-Computed Tomographic (CT) Images: Lessons Learned From the Curve-Fitting Exercises," Int. J. Multiscale Comput. Eng., **14**(4), pp. 389–413.
[12] Jang, J., and Smyth, A. W., 2017, "Model Updating of a Full-Scale Fe Model With Nonlinear Constraint Equations and Sensitivity-Based Cluster Analysis for Updating Parameters," Mech. Syst. Signal. Process., **83**, pp. 337–355.
[13] Ghaboussi, J., Pecknold, D. A., Zhang, M., and Haj-Ali, R. M., 1998, "Autoprogressive Training of Neural Network Constitutive Models," Int. J. Numer. Methods Eng., **42**(1), pp. 105–126.
[14] Ghaboussi, J., Garrett Jr, J., and Wu, X., 1991, "Knowledge-Based Modeling of Material Behavior With Neural Networks," J. Eng. Mech., **117**(1), pp. 132–153.
[15] Lefik, M., and Schrefler, B. A., 2003, "Artificial Neural Network as an Incremental Non-linear Constitutive Model for a Finite Element Code," Comput. Methods. Appl. Mech. Eng., **192**(28–30), pp. 3265–3283.
[16] Mozaffar, M., Bostanabad, R., Chen, W., Ehmann, K., Cao, J., and Bessa, M., 2019, "Deep Learning Predicts Path-Dependent Plasticity," Proc. Natl. Acad. Sci. USA, **116**(52), pp. 26414–26420.
[17] Wang, K., and Sun, W., 2018, "A Multiscale Multi-Permeability Poroplasticity Model Linked by Recursive Homogenizations and Deep Learning," Comput. Methods. Appl. Mech. Eng., **334**, pp. 337–380.
[18] Chi, H., Zhang, Y., Tang, T. L. E., Mirabella, L., Dalloro, L., Song, L., and Paulino, G. H., 2021, "Universal Machine Learning for Topology Optimization," Comput. Methods. Appl. Mech. Eng., **375**, p. 112739.
[19] Baker, N., Alexander, F., Bremer, T., Hagberg, A., Kevrekidis, Y., Najm, H., Parashar, M., Patra, A., Sethian, J., Wild, S., and Willcox, K., 2019, "Workshop Report on Basic Research Needs for Scientific Machine Learning: Core Technologies for Artificial Intelligence, Technical Report, USDOE Office of Science (SC) Washington, DC.
[20] Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., and Yu, B., 2019, "Definitions, Methods, and Applications in Interpretable Machine Learning," Proc. Natl. Acad. Sci. USA, **116**(44), pp. 22071–22080.
[21] Liu, Y., Sun, W., Yuan, Z., and Fish, J., 2016, "A Nonlocal Multiscale Discrete-Continuum Model for Predicting Mechanical Behavior of Granular Materials," Int. J. Numer. Methods Eng., **106**(2), pp. 129–160.
[22] Wang, K., and Sun, W., 2016, "A Semi-Implicit Discrete-Continuum Coupling Method for Porous Media Based on the Effective Stress Principle at Finite Strain," Comput. Methods. Appl. Mech. Eng., **304**, pp. 546–583.
[23] Feyel, F., and Chaboche, J.-L., 2000, "Fe2 Multiscale Approach for Modelling the Elastoviscoplastic Behaviour of Long Fibre Sic/ti Composite Materials," Comput. Methods. Appl. Mech. Eng., **183**(3–4), pp. 309–330.
[24] Hartmaier, A., Buehler, M. J., and Gao, H., 2005, "Multiscale Modeling of Deformation in Polycrystalline Thin Metal Films on Substrates," Adv. Eng. Mater., **7**(3), pp. 165–169.
[25] Vlassis, N. N., and Sun, W., 2021, "Sobolev Training of Thermodynamic-Informed Neural Networks for Interpretable Elasto-Plasticity Models With Level Set Hardening," Comput. Methods. Appl. Mech. Eng., **377**, p. 113695.
[26] Molnar, C., Casalicchio, G., and Bischl, B., 2018, "iml: An R Package for Interpretable Machine Learning," J. Open Source Softw., **3**(26), p. 786.
[27] Coombs, W. M., and Motlagh, Y. G., 2017, "Nurbs Plasticity: Yield Surface Evolution and Implicit Stress Integration for Isotropic Hardening," Comput. Methods. Appl. Mech. Eng., **324**, pp. 204–220.

[28] Borja, R. I., 2013, *Plasticity: Modeling & Computation*, Springer Science & Business Media, Berlin/Heidelberg.

[29] Bishop, C. M., 1995, *Neural Networks for Pattern Recognition*, Oxford University Press, Oxford, UK.

[30] Czarnecki, W. M., Osindero, S., Jaderberg, M., Świrszcz, G., and Pascanu, R., 2017, "Sobolev Training for Neural Networks," Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), Curran Associates Inc., Red Hook, NY, pp. 4281–4290.

[31] Roscoe, K., and Burland, J., 1970, "On The Generalized Stress-Strain Behavior of "Wet" Clay: 60. K. H. Roscoe and J. B. Burland. Engineering Plasticity (Papers for a Conference Held in Cambridge, Mar. 1968), Cambridge, University Press, 535–609 (1968)," J. Terramechanics, **7**(2), pp. 107–108.

[32] Argyris, J., Faust, G., Szimmat, J., Warnke, E., and Willam, K., 1974, "Recent Developments in the Finite Element Analysis of Prestressed Concrete Reactor Vessels," Nucl. Eng. Des., **28**(1), pp. 42–75.

[33] Cowper, G. R., and Symonds, P. S., 1957, "Strain-Hardening and Strain-Rate Effects in the Impact Loading of Cantilever Beams, Brown Univ Providence Ri, Technical Report, Office of Naval Research.

[34] Dozat, T., 2016, "Incorporating Nesterov Momentum Into Adam," http://cs229.stanford.edu/proj2015/054_report.pdf

[35] Kochmann, J., Wulfinghoff, S., Reese, S., Mianroodi, J. R., and Svendsen, B., 2016, "Two-Scale FE-FFT-and Phase-Field-Based Computational Modeling of Bulk Microstructural Evolution and Macroscopic Material Behavior," Comput. Methods. Appl. Mech. Eng., **305**, pp. 89–110.

[36] Fuchs, A., Heider, Y., Wang, K., Sun, W., and Kaliske, M., 2021, "Dnn2: A Hyper-Parameter Reinforcement Learning Game for Self-Design of Neural Network Based Elasto-Plastic Constitutive Descriptions," Comput. Struct., **249**, p. 106505.

[37] Pascanu, R., Mikolov, T., and Bengio, Y., 2013, "On the Difficulty of Training Recurrent Neural Networks," International Conference on Machine Learning, Atlanta, GA, June 17–19, PMLR, pp. 1310–1318.

[38] Xu, K., Huang, D. Z., and Darve, E., 2021, "Learning Constitutive Relations Using Symmetric Positive Definite Neural Networks," J. Comput. Phys., **428**, p. 110072.

[39] Giunta, A., Wojtkiewicz, S., and Eldred, M., 2003, "Overview of Modern Design of Experiments Methods for Computational Simulations," 41st Aerospace Sciences Meeting and Exhibit, Reno, NV, Jan. 6–9, p. 649.

[40] Hashash, Y., Jung, S., and Ghaboussi, J., 2004, "Numerical Implementation of a Neural Network Based Material Model in Finite Element Analysis," Int. J. Numer. Methods Eng., **59**(7), pp. 989–1005.

[41] de Souza Neto, E. A., Peric, D., and Owen, D. R., 2011, *Computational Methods for Plasticity: Theory and Applications*, John Wiley & Sons, Hoboken, NJ.

[42] Abbo, A., and Sloan, S., 1995, "A Smooth Hyperbolic Approximation to the Mohr-Coulomb Yield Criterion," Comput. Struct., **54**(3), pp. 427–441.

[43] Matsuoka, H., and Nakai, T., 1974, "Stress-Deformation and Strength Characteristics of Soil Under Three Different Principal Stresses," Proceedings of the Japan Society of Civil Engineers, no. 232, Japan Society of Civil Engineers, pp. 59–70.

[44] Vlassis, N. N., Ma, R., and Sun, W., 2020, "Geometric Deep Learning for Computational Mechanics Part I: Anisotropic Hyperelasticity," Comput. Methods. Appl. Mech. Eng., **371**, p. 113299.

[45] Scarselli, F., and Tsoi, A. C., 1998, "Universal Approximation Using Feedforward Neural Networks: A Survey of Some Existing Methods, and Some New Results," Neural Netw., **11**(1), pp. 15–37.

[46] Parhi, R., and Nowak, R. D., 2021, "Banach Space Representer Theorems for Neural Networks and Ridge Splines," J. Mach. Lear. Res., **22**(43), pp. 1–40.

[47] Weinan, E., and Wojtowytsch, S., 2020, "On the Banach Spaces Associated With Multi-Layer ReLU Networks: Function Representation, Approximation Theory and Gradient Descent Dynamics," arXiv Preprint.

[48] Wang, K., Sun, W., and Du, Q., 2019, "A Cooperative Game for Automated Learning of Elasto-plasticity Knowledge Graphs and Models With AI-Guided Experimentation," Comput. Mech., **64**(2), pp. 467–499.

[49] Zaremba, W., Sutskever, I., and Vinyals, O., 2014, "Recurrent Neural Network Regularization," arXiv Preprint.