International Journal of
Applied Data Science in Engineering and Health
https://ijadseh.com

Machine Learning-Assisted Modelling of Void Evolution and Ductile Fracture in Anisotropic Materials: A Critical Review

Sarvenaz Hashem-Sharifi Pishvaei

Department of Continuum Mechanics and Structural Analysis,

Universidad Carlos III de Madrid, 28911 Leganés, Madrid, Spain (Email: sarvesharifi@gmail.com)

Received date: July 29, 2026; Accepted date: August 5, 2026

Abstract

Void evolution and ductile failure in anisotropic metals with tension-compression asymmetry are strongly influenced by the coupled effects of stress triaxiality T, Lode parameter L, material orientation, hardening, and void morphology. Three-dimensional unit-cell simulations can capture these interactions in considerable detail but are computationally expensive for extensive calibration, design-space exploration, and uncertainty quantification. This critical review examines four complementary roles of artificial intelligence: surrogate prediction of finite-element responses, inverse identification of constitutive and damage parameters, physics-informed constitutive learning, and microstructure-aware image- or graph-based modelling. A multi-fidelity workflow is proposed that integrates homogenized models, unit-cell calculations, statistically representative porous volume elements, and experiments, while incorporating constraints related to objectivity, material symmetry, thermodynamic dissipation, physically admissible porosity evolution, and uncertainty calibration. Current evidence most strongly supports the use of AI for repeated forward evaluation and inverse parameter identification. Reliable extension toward broader failure prediction requires validation on unseen loading trajectories and material domains, separate assessment of stress and damage-related variables, and an uncertainty-based abstention or fallback mechanism outside the model's training coverage.

Keywords: Ductile fracture; Void growth; Porous plasticity; Tension-compression asymmetry; Machine learning; Physics-informed AI.

Introduction

Ductile fracture is commonly described as a sequence of void nucleation, growth, plastic-strain localization, and final coalescence. This apparently simple description conceals a multiscale problem: local plastic flow around voids is governed by the constitutive response of the matrix, whereas the macroscopic loss of load-carrying capacity depends on the evolving void population, morphology, spacing, and interactions. Classical micromechanical models and finite-element unit-cell calculations have therefore remained central to the interpretation and prediction of fracture in structural metals. Their predictive capability depends strongly on the consistent representation of the loading path, stress triaxiality T, Lode parameter L, hardening response, and microstructural state (Gurson, 1977; Tvergaard and Needleman, 1984; Benzerga and Leblond, 2010).

The challenge becomes more pronounced for materials that are plastically anisotropic and exhibit a strength differential, namely different yield strengths in tension and compression. Hexagonal close-packed alloys and strongly textured metals are representative examples, while similar directional and tension–compression asymmetric responses have also been reported in some additively manufactured alloys (Cazacu et al., 2006; Ghorbanpour et al., 2020). In such materials, a pressure-insensitive matrix can nevertheless lead to a porous aggregate response that is sensitive to the magnitude and sign of the mean stress and to the third invariant of the deviatoric stress. The CPB06 family of yield criteria was developed to capture the combined effects of orthotropy and tension–compression asymmetry at the matrix level (Cazacu et al., 2006). Subsequent porous extensions and unit-cell studies showed that even moderate tension–compression asymmetry can substantially affect dilatational plastic flow and void growth (Cazacu and Stewart, 2009; Alves and Cazacu, 2015).

Micromechanical and homogenized-model studies from several research groups have established that porous failure in anisotropic solids is governed by the coupled effects of stress state, matrix plasticity, material orientation, and microstructural geometry. Beyond analytical porous formulations, numerical investigations have demonstrated that material orientation can significantly modify void growth in anisotropic solids, while representative volumes containing realistic pore distributions reveal interaction, clustering, and localization patterns that cannot be captured using a single idealized void (Hosseini et al., 2022; Vishnu et al., 2023). Within this broader context, Hashem-Sharifi et al. (2022) provided a controlled three-dimensional unit-cell benchmark for assessing the coupled influence of anisotropy, tension–compression asymmetry, hardening, stress triaxiality, and Lode parameter on macroscopic stress and porosity evolution. Collectively, these studies show that agreement in nominal stress alone is insufficient and that damage-related internal variables, particularly porosity, must be evaluated independently (Benzerga and Leblond, 2010).

Against this background, machine-learning methods can contribute to this problem through several complementary approaches. Forward surrogate models learn mappings from loading, material, and microstructural descriptors to outputs such as stress histories, porosity evolution, critical-event strains, or full-field responses, thereby reducing the cost of repeated high-fidelity simulations. Inverse models estimate constitutive or damage parameters from measured curves or fields and are particularly useful for accelerating calibration. Physics-informed and hybrid models incorporate constraints such as objectivity, material symmetry, thermodynamic consistency, and physically admissible state evolution into the learning process. Microstructure-aware convolutional and graph-based models use images, voxel data, or pore and grain networks to capture spatial morphology and interactions. These approaches differ in their data requirements, interpretability, intended application, and extrapolation risk, and should therefore be selected and validated according to the specific mechanics problem being addressed.

Although a recent comprehensive review surveyed machine-learning applications across fracture-property identification, crack-path prediction, fatigue fracture and life prediction, dynamic fracture, small-scale testing, and multiscale problems (Yu et al., 2026), the specific integration of machine learning with void evolution and porous ductile failure in anisotropic materials exhibiting tension–compression asymmetry remains insufficiently developed in the review literature. In particular, limited attention has been given to the combined influence of stress triaxiality, Lode parameter, material orientation, strength differential, hardening, void morphology, and loading history on the design and validation of data-driven models. The need to assess stress and damage-related internal variables separately, especially porosity, and to account explicitly for discrepancies among homogenized models, unit-cell simulations, representative volume elements with realistic pore distributions, and experiments has also not been examined in a unified framework. The present review addresses these gaps by critically evaluating forward surrogate modelling, inverse parameter identification, physics-informed constitutive learning, and microstructure-aware prediction for this specific class of materials. Its main contribution is the proposal of a physics-

constrained, multi-fidelity, and uncertainty-aware framework, together with validation, applicability-domain, reproducibility, and reporting requirements for credible AI-assisted modelling of void evolution and ductile failure.

Review scope and literature search

This targeted critical review was informed by searches conducted across publisher platforms and proceedings repositories, including ScienceDirect, SpringerLink, Nature Portfolio, Oxford Academic, and the Proceedings of Machine Learning Research. The searches were last updated on 27 July 2026. Search queries combined terms related to artificial intelligence ("machine learning", "artificial intelligence", "physics-informed", and "neural operator"), constitutive and fracture mechanics ("constitutive modelling", "ductile fracture", "void growth", and "porous plasticity"), and material behaviour ("anisotropy" and "tension–compression asymmetry"). Peer-reviewed English-language studies were prioritized when they addressed surrogate modelling, inverse parameter identification, physics-constrained constitutive learning, uncertainty quantification, or microstructure-aware modelling relevant to ductile failure. Earlier foundational studies were retained where necessary to establish porous and asymmetric plasticity, multi-fidelity modelling, and uncertainty-quantification methods. Studies focused exclusively on crack-image detection, structural-health monitoring, or biomedical fracture were excluded unless they introduced a method directly relevant to constitutive or microstructural modelling of ductile failure. The search was targeted and narrative rather than systematic and was not intended to provide exhaustive bibliographic coverage.

The synthesis is narrative rather than systematic; it does not claim exhaustive database coverage and does not include a quantitative meta-analysis. Its objectives are threefold: to identify applications in which AI provides demonstrable advantages over conventional modelling approaches; to propose a physics-constrained workflow tailored to the loading, constitutive, and microstructural variables considered in three-dimensional unit-cell studies; and to define validation requirements that distinguish computational acceleration within the training domain from physically credible generalization to unseen loading and material conditions.

Physics-based foundations and the benchmark problem

Void evolution under multiaxial loading

In porous plasticity, the macroscopic stress state is commonly characterized by the stress triaxiality T and the Lode parameter L, defined as

$$T = \frac{\Sigma_m}{\Sigma_{\mathrm{eq}}}, \qquad L = \frac{2\Sigma_{\mathrm{II}} - \Sigma_{\mathrm{I}} - \Sigma_{\mathrm{III}}}{\Sigma_{\mathrm{I}} - \Sigma_{\mathrm{III}}}, \tag{1}$$

where $\Sigma_m = (\Sigma_{\mathrm{I}} + \Sigma_{\mathrm{II}} + \Sigma_{\mathrm{III}})/3$ is the macroscopic mean stress, Σ_{eq} is the macroscopic von Mises equivalent stress, and $\Sigma_{\mathrm{I}} \geq \Sigma_{\mathrm{II}} \geq \Sigma_{\mathrm{III}}$ are the ordered principal macroscopic stresses. For the adopted definition of the Lode parameter, $L = -1$, $L = 0$, and $L = 1$ correspond to axisymmetric tension, generalized shear, and axisymmetric compression, respectively (Hashem-Sharifi et al., 2022; Hosseini et al., 2022). Stress triaxiality characterizes the hydrostatic contribution that strongly influences void expansion or collapse. The Lode parameter characterizes differences in the

deviatoric stress state associated with the third invariant, which can influence void-shape evolution, plastic-strain localization, and shear-assisted damage. Under proportional loading, T and L, together with the loading orientation relative to the material symmetry axes, provide a compact description of the imposed stress state. Under non-proportional loading, however, their instantaneous values are insufficient because the sequence of stress states and the associated deformation and internal-variable histories must also be retained.

The Gurson yield function introduced porosity as an internal variable that couples the mean stress to the macroscopic yielding of a porous material (Gurson, 1977). Tvergaard-type parameters and coalescence modifications subsequently improved agreement with unit-cell calculations and experimental observations (Tvergaard and Needleman, 1984). Conventional GTN formulations, however, are generally based on an isotropic von Mises matrix and therefore do not directly represent plastic anisotropy or tension–compression asymmetry. Extensions accounting for anisotropic matrix behaviour, non-spherical voids, shear-assisted damage, and void coalescence have progressively broadened the framework (Benzerga and Leblond, 2010). Such extensions typically introduce additional parameters, internal variables, or constitutive functions that require identification and validation, often using a limited set of mechanical tests. Because different parameter combinations may produce similar macroscopic responses, the resulting inverse problem can also be non-unique. The acceleration and regularization of this identification process therefore represent an important application of machine learning.

The porous response of a tension–compression asymmetric matrix cannot generally be represented by merely substituting different tensile and compressive yield stresses into a conventional GTN formulation. The matrix yield criterion and associated flow rule alter the distribution of local plastic strain and, after homogenization, modify the coupling between mean stress and the third deviatoric-stress invariant. Cazacu and Stewart (2009) derived an analytical porous potential for an isotropic matrix exhibiting strength-differential effects. Through three-dimensional micromechanical calculations, Alves and Cazacu (2015) further showed that, at fixed moderate or high triaxiality, changing the sign of the strength-differential parameter can reverse the relative void-growth rates associated with the two opposite axisymmetric Lode states, $L = -1$ and $L = 1$. Consequently, a surrogate trained across varying Lode states, material orientations, or strength-differential parameters, but supplied only with T, cannot uniquely represent the underlying response (Hashem-Sharifi et al., 2022).

Lessons from the 2022 three-dimensional unit-cell study

The 2022 study considered a three-dimensional periodic cubic unit cell containing a central spherical void. Periodic boundary conditions represented an idealized periodic array of voids, while nonlinear loading constraints were used to maintain prescribed macroscopic stress ratios throughout the loading history. The constitutive response of the matrix was described using CPB06, which represented plastic anisotropy and tension–compression asymmetry. The explicit unit-cell response was compared with predictions from the Stewart–Cazacu homogenized porous formulation under matched loading and initial-porosity conditions. This comparison provided a controlled basis for assessing the effects of matrix hardening and the strength-differential parameter and for evaluating the predictive limitations of the homogenized formulation (Hashem-Sharifi et al., 2022).

Three observations from that study are especially relevant to AI model design. First, the problem involves multiple outputs with different degrees of nonlinearity and prediction difficulty. Macroscopic stress may remain relatively smooth and well approximated, whereas porosity evolution can display greater sensitivity and substantially larger model discrepancies as void growth and softening progress. A single averaged error metric may therefore conceal a physically important error in the damage-related output, motivating the separate evaluation of stress and

porosity predictions. Second, matrix hardening affects the stress response, the rate of void growth, and the onset of macroscopic softening. For the higher-hardening cases considered, the original comparison showed substantially larger discrepancies in porosity than in macroscopic stress (Hashem-Sharifi et al., 2022). Third, the response depends jointly on the strength-differential parameter, hardening response, stress triaxiality T, and Lode parameter L. These coupled dependencies motivate structured sampling of the combined parameter space and physics-aware feature design rather than a one-factor-at-a-time dataset.

The idealized spherical-void cell should not be viewed merely as a limitation; it provides a controlled benchmark for attributing changes in response to specific physical variables. Subsequent studies have expanded the range of mechanisms and microstructures considered. Material orientation has been shown to alter void growth in anisotropic solids (Hosseini et al., 2022), while representative volumes with realistic pore distributions reveal interaction, clustering, and localization patterns that are absent from a single-void cell (Vishnu et al., 2023). Initial void aspect ratio and morphology introduce further sensitivities in tension–compression asymmetric matrices (Hashem-Sharifi et al., 2025). Together, these studies motivate a natural multi-fidelity hierarchy: analytical and homogenized constitutive models for low-cost evaluation; idealized unit cells for controlled mechanistic studies; representative volumes with realistic pore distributions for void interactions and localization; and experiments for physical validation. AI models intended for broader failure prediction are more credible when these levels are combined through multi-fidelity training or cross-level validation rather than relying exclusively on a single level.

Computational bottlenecks that motivate AI

High-fidelity simulations that explicitly resolve void geometry and local deformation fields are computationally expensive for three related reasons. First, the combined loading, constitutive, and microstructural parameter space is high dimensional. Second, resolving steep strain gradients, void interactions, and localization requires refined discretization and robust nonlinear solution procedures. Third, several constitutive and microstructural parameters are uncertain or only indirectly identifiable from experiments. A comprehensive sensitivity or design-space study can therefore require hundreds or thousands of simulations across stress triaxiality T, Lode parameter L, material orientation, hardening, initial porosity, void morphology, and loading history. Parameter calibration adds an outer optimization loop, while probabilistic uncertainty propagation requires many additional model evaluations. Crystal-plasticity and image-based representative volume elements further increase the computational cost. These repeated-evaluation settings are particularly suitable for surrogate modelling, active learning, and multi-fidelity strategies, provided that computational acceleration remains accompanied by physics-based validation.

Machine-learning paradigms for ductile failure

Forward surrogate models and reduced-order representations

A forward surrogate approximates the mapping from material, microstructural, loading, and state descriptors to a response that would otherwise require an expensive simulation. A reduced-order model additionally represents a high-dimensional response using a smaller set of coordinates or latent variables. For the unit-cell problem, case-level inputs may include stress triaxiality T, Lode parameter L, the strength-differential parameter, hardening parameters, initial porosity, initial void aspect ratio, and material orientation. For incremental or path-dependent models, additional inputs may include the current strain increment, accumulated plastic strain, current porosity, and void-shape descriptors. Outputs may include scalar-valued histories of macroscopic effective stress, selected stress components, and porosity; clearly defined critical strains for localization or coalescence; or complete displacement, stress, and strain fields.

Feed-forward neural networks, Gaussian processes, random forests, and gradient-boosted trees can be suitable for fixed-length scalar outputs, whereas recurrent or gated architectures are more appropriate for path-dependent response histories. Recurrent models have reproduced anisotropic, path-dependent plasticity under multiaxial loading (Gorji et al., 2020), while experimentally informed neural networks have been used to predict ductile-damage evolution in terms of void fraction (Schowtjak et al., 2022). Broader developments in neural constitutive modelling and the incorporation of physical constraints are reviewed by Dornheim et al. (2024).

For scalar-valued responses, forward surrogates currently provide the most direct and practical form of acceleration. A carefully sampled set of unit-cell calculations can be used to train a surrogate that evaluates stress and porosity histories at a small fraction of the original simulation cost, enabling global sensitivity analysis, optimization, and uncertainty propagation. A major risk is that random pointwise train–test splitting may produce apparently accurate interpolation while failing to assess generalization to unseen loading paths. Validation should therefore hold out complete simulations, trajectories, or contiguous regions of the parameter space rather than individual increments from trajectories already represented during training.

Under non-proportional loading, history dependence requires an explicit representation of the evolving material state. Recurrent and gated neural networks can encode loading history through a hidden state, but this state is difficult to interpret physically and may accumulate error during long recursive predictions. A more auditable alternative is to supply physically meaningful variables, such as accumulated plastic strain, porosity, void-shape descriptors, and the current loading increment, and to use the network only to predict their incremental evolution and the corresponding stress update. Thermodynamically constrained recurrent models can additionally penalize or prevent negative plastic dissipation (Borkowski et al., 2022). Such constraints improve physical admissibility but should be combined with requirements related to objectivity, material symmetry, admissible porosity evolution, and numerical stability.

For full-field outputs, autoencoder-based reduced-order models, graph neural networks, and neural operators are more appropriate. An encoder can compress a stress or strain field into a low-dimensional latent representation, a surrogate can evolve that representation, and a decoder can reconstruct the field. Neural operators instead seek to learn mappings between function spaces, as formalized by DeepONet and general neural-operator frameworks (Lu et al., 2021; Kovachki et al., 2023). They can also be coupled with finite-element solvers in multiscale mechanics (Yin et al., 2022). Material fields, loading data, and boundary conditions can therefore be represented as functions, while specialized formulations may accommodate variations in geometry and discretization. Nevertheless, the sharp gradients, evolving void surfaces, and possible coalescence-related discontinuities involved in porous failure remain challenging for smooth low-dimensional representations. Validation metrics should consequently assess local maxima, hotspot locations, localization bands, and integral damage-related quantities in addition to average field errors.

Inverse identification and automated calibration

Parameter identification is one of the clearest near-term, high-value applications of machine learning in ductile-fracture modelling. Conventional calibration repeatedly solves a forward finite-element problem while an optimization algorithm adjusts the constitutive or damage parameters. The objective-function landscape may be non-convex, different parameters may have compensating effects, and the available measurements may not uniquely identify all model parameters. Machine learning can reduce the repeated computational burden in two principal ways: by learning a forward surrogate that is evaluated within an optimization or Bayesian inference procedure, or by learning an inverse mapping from measured responses directly to parameter estimates. The second approach can provide rapid predictions after training, but it requires particular care when the inverse mapping is non-unique.

Baltic et al. (2021) combined finite-element simulations, global force–displacement histories, DIC-derived displacement fields, and artificial neural networks to identify the parameters of a stress-state-dependent ductile-damage model using one calibration-specimen geometry. Chen et al. (2021) combined a resilient back-propagation neural network with a genetic algorithm for the efficient identification of parameters in a modified GTN model. Yao et al. (2022) used a neural-network-assisted procedure to identify parameters of a semi-coupled fracture formulation for 6061 aluminium alloy from the force–displacement response of a single specimen. Blaschke et al. (2023) trained an inverse model for rapid prediction of TEPLA parameters and assessed the method using synthetic and experimental data. More recently, O'Connor et al. (2024) developed a largely autonomous framework coupling Bayesian optimization with finite-element analysis to reduce manual intervention in ductile-damage calibration.

These studies show that much of the computational burden can be transferred to an offline data-generation and training stage, enabling faster online parameter estimation or surrogate-assisted optimization. Fast inversion, however, does not resolve structural non-identifiability. Different parameter sets may reproduce nearly identical global force–displacement curves while potentially generating different local stress states, porosity histories, or localization fields. Calibration of anisotropic porous materials with tension–compression asymmetry should therefore combine multiple complementary observables, including global force, full-field displacement or strain measurements, local thickness reduction, void-fraction evolution where experimentally accessible, and tensile and compressive tests along multiple material orientations.
When the available observations only weakly constrain the parameters, a single best-fit parameter vector provides an incomplete description of the inverse problem. Bayesian calibration should report posterior parameter distributions, while non-Bayesian inverse models should at least provide multiple admissible solutions or uncertainty-calibrated prediction intervals. Candidate parameter sets should subsequently be assessed using specimen geometries, loading paths, or material orientations that were not used during calibration.

Physics-informed constitutive learning

Purely data-driven constitutive networks can approximate complex stress–strain relations, but unless appropriate restrictions are imposed, they may violate frame indifference, material symmetry, thermodynamic consistency, or, when an explicit yield function is learned, yield-surface convexity. Such violations can produce nonphysical stress responses, inadmissible energy generation, loss of numerical robustness, or poor convergence when the learned model is embedded in a finite-element solver. Mesh dependence associated with local damage and strain softening is a related but distinct problem that generally requires spatial regularization or a nonlocal formulation. Physics-informed constitutive learning addresses these issues by incorporating structural constraints into the network architecture, the loss function, or the parameterization of stored-energy, yield, and dissipation potentials.

Thermodynamics-based artificial neural networks encode relations among free energy, stress, internal variables, and dissipation using automatic differentiation, allowing energy balance and non-negative dissipation to be imposed structurally or through suitable restrictions (Masi et al., 2021). Constitutive Artificial Neural Networks employ mechanically meaningful inputs, modular functional representations, and architectural constraints to preserve relevant invariants and material symmetries and to discover compact constitutive relations (Linka et al., 2021; Linka and Kuhl, 2023). Although these developments have been demonstrated mainly for hyperelastic behaviour, they illustrate how objectivity, symmetry, and energy restrictions can be embedded into constitutive-learning architectures.

Physics-constrained plasticity models have moved further toward learning selected inelastic mechanisms. Vlassis and Sun (2021) used Sobolev training and an evolving level-set representation to learn interpretable stored-energy, yield, plastic-flow, and hardening components.

Meyer and Ekre (2023) embedded neural networks within a conventional plasticity structure to discover evolution laws while retaining thermodynamic consistency and frame invariance. Thermodynamics-informed recurrent models have also been developed for path-dependent granular elastoplasticity (Su et al., 2024). These developments suggest a structure-preserving strategy for CPB06-based porous plasticity: the known transformed-stress representation, material symmetries, strength-differential structure, and admissible state variables can be retained analytically, while only uncertain components of hardening, nucleation, porosity evolution, void-shape evolution, or coalescence are learned.

For the present review, three levels of hybridization are distinguished. In a weak hybrid, the network predicts a bounded correction to the output of a trusted constitutive model, for example a discrepancy in porosity evolution between a homogenized model and a unit-cell benchmark. In a medium hybrid, the analytical yield function and physically meaningful state variables are retained, while the network learns one or more evolution laws, such as hardening, nucleation, void-shape evolution, or coalescence. In a strong hybrid, neural representations replace a substantial part of the constitutive formulation but are constrained by objectivity, material symmetry, thermodynamic dissipation, and appropriate convexity or stability conditions. For applications requiring strong physical auditability, weak and medium hybrids are currently easier to interpret and validate because they preserve recognizable state variables, limiting cases, and established constitutive structure.
Physics-informed neural networks provide a related but distinct route. Whereas constitutive-learning models approximate the material-point update, PINNs can approximate the solution of a boundary-value problem by incorporating equilibrium equations, initial and boundary conditions, and constitutive residuals into the training objective (Raissi et al., 2019). PINNs have been applied to phase-field brittle fracture (Lian et al., 2023), to large-deformation fracture in elastomer-like materials without a gradient-damage formulation (Konale and Srivastava, 2026), and to Lemaitre-based local and nonlocal ductile-damage evolution (Mirzaei et al., 2026). Their ability to approximate fields without a conventional finite-element mesh is attractive, although accurate resolution still requires suitable spatial and temporal collocation. Fracture remains difficult because of sharp gradients, competing loss terms, history dependence, irreversibility, and multiple possible damage paths. For the anisotropic porous-material class considered here, parameter inference or a physics-constrained local constitutive closure embedded within a conventional finite-element solver represents a more readily verifiable near-term application than complete replacement of three-dimensional finite-element analysis.

Microstructure-aware and image-based learning

In reduced descriptions of porous materials, the microstructure is commonly summarized using descriptors such as initial porosity, void-size distribution, void aspect ratio, and void orientation. Such descriptors are useful for controlled unit-cell studies but cannot fully represent realistic materials, in which pore clustering, nearest-neighbour spacing, grain orientation, phase topology, and pore-surface morphology may influence local stress concentration and localization. Convolutional neural networks can learn spatial features from segmented micro-CT volumes, microscopy images, or synthetic voxelized microstructures. Graph neural networks provide an alternative representation in which pores or grains are treated as nodes, while edges encode adjacency, distance, shared-boundary information, or other neighbourhood relationships. Graph representations are particularly useful for irregular microstructures and samples containing variable numbers of defects or grains.

Hestroffer et al. (2023) used grain orientation, size, and neighbour connectivity to predict the effective stiffness and yield strength of polycrystalline α-Ti microstructures. Gao et al. (2025) combined an improved graph-attention network for global stress–strain and yield-strength prediction with a three-dimensional conditional diffusion model for local stress fields. Noh et al. (2026) used a fully convolutional architecture to predict two-dimensional microscale strain

patterns from grain-orientation maps. More recently, Tran et al. (2026) developed a multi-fidelity, physics-guided graph surrogate for predicting strength properties and biaxial yield loci across diverse polycrystalline microstructures. Although these studies primarily address polycrystalline property and field prediction rather than void-driven fracture, they provide transferable representations for encoding irregular microstructures and neighbourhood interactions.

The physical scale at which images are analysed is itself part of the modelling choice. Zheng et al. (2023) showed that machine-learning classification of ductile fracture surfaces depends on image length scale, with the scale yielding the highest classification accuracy varying with the categories being distinguished. Although that study concerned post-fracture surface classification rather than prospective failure prediction, it demonstrates that image scale can materially affect the learned representation. Cropped or resized images should therefore not be treated as scale-free inputs.

For volumetric pore data, the field of view must be sufficiently large to capture relevant pore neighbourhoods and clustering, while the spatial resolution must resolve the defects and microstructural features controlling stress concentration, void growth, or nucleation. Physical voxel size, crop dimensions, and imaging resolution should be retained as model metadata or explicit inputs. Multi-resolution architectures and scale-tagged inputs should consequently be evaluated, with validation performed on microstructures, fields of view, and physical resolutions that were not represented during training.

A tailored AI-enhanced framework for asymmetric porous materials

Multi-fidelity data design

For the material class considered here, the design of an AI workflow should begin with the hierarchy of physical models and available evidence rather than with the selection of a network architecture. At the lowest-cost computational level, the homogenized Stewart–Cazacu formulation can provide a broad, physics-based baseline across stress triaxiality T, Lode parameter L, the strength-differential parameter k, and hardening behaviour. Three-dimensional unit-cell calculations can then be concentrated in regions where the homogenized formulation exhibits large discrepancies or where its accuracy is uncertain. Representative volume elements with realistic pore distributions additionally capture pore interactions, clustering, and spatial variability that are absent from idealized single-void cells. Experiments provide the principal calibration and validation evidence through global force–displacement measurements, full-field displacement or strain data, in situ or interrupted tomography, and post-fracture observations. A portion of the experimental evidence should be reserved exclusively for independent validation.

Multi-fidelity or multi-source learning should preserve the provenance, uncertainty, and systematic discrepancy associated with each information level rather than pooling all samples as equally authoritative observations. Lower-fidelity predictions should therefore be treated as informative but potentially biased. Classical co-kriging provides a statistical framework for relating inexpensive and expensive simulators, while nonlinear information-fusion methods can represent more complex cross-fidelity relationships (Kennedy and O'Hagan, 2000; Perdikaris et al., 2017). When different levels provide non-equivalent outputs, such as macroscopic response histories, local fields, tomographic measurements, and fracture-surface images, the fusion strategy should explicitly define how these observables are linked and how their measurement and modelling uncertainties are represented. Depending on the intended domain of applicability, the candidate design variables should include stress triaxiality T, Lode parameter L, loading direction relative to the material axes, the strength-differential parameter k, hardening and anisotropy parameters, initial porosity, initial void shape and orientation, and an appropriate representation of loading history. For non-proportional loading, the history should be represented through stress or strain increments and evolving state variables rather than only through instantaneous values of T and L.

Table 1. AI methods relevant to void evolution and ductile failure.

Method	Typical inputs and outputs	Principal value	Main risk / recommended use
Scalar surrogate (ANN, Gaussian process, tree ensemble)	Material, microstructural, and loading descriptors → stress or porosity histories and defined event strains	Rapid repeated evaluation for sensitivity analysis, optimization, and uncertainty propagation	Interpolation and extrapolation bias; use within a well-characterized parameter domain
Sequence model (RNN/GRU/state-space network)	Loading increments and physical state → incremental stress and state-variable updates	Represents history dependence and non-proportional loading	Hidden-state drift and error accumulation; retain auditable physical state variables
Field surrogate (autoencoder or neural operator)	Geometry, loading, and boundary fields → stress, strain, or damage-related fields	Rapid full-field prediction and reduced representation of high-dimensional responses	Smoothing or displacement of localization; validate extrema, hotspots, and localization band geometry
Inverse network or simulation-based inference	Measured curves and fields → parameter estimates, admissible solutions, or a posterior distribution	Accelerates calibration and can quantify parameter uncertainty	Non-identifiability; combine complementary experiments and use probabilistic outputs where appropriate
Physics-informed constitutive network	Strain increments, invariants, and internal variables → stress and state-variable updates	Can enforce objectivity, material symmetry, and non-negative dissipation	Numerical instability or misspecified constraints; test material tangents, rotations, unloading, and cyclic histories
CNN / GNN microstructure model	Voxel images, point clouds, or pore/grain graphs → effective properties, local fields, or damage-related indicators	Learns spatial morphology and neighbourhood interactions	Scale, representation, segmentation, and dataset bias; test unseen morphologies and physical resolutions

When realistic microstructures are modelled explicitly, additional descriptors may include pore-size and spacing distributions, clustering measures, nearest-neighbour statistics, phase topology, and crystallographic texture or grain-orientation distributions where grain-scale anisotropy is resolved.

An initial space-filling design may be generated using Latin-hypercube or low-discrepancy sampling within the physically admissible parameter domain, with deliberate inclusion of boundaries, limiting cases, and selected interaction regions. Subsequent enrichment should prioritize regions with high epistemic uncertainty, large predicted cross-fidelity discrepancy, steep response gradients, rapid porosity growth, or indicators of localization, while accounting for the computational cost of the requested fidelity level. Bayesian uncertainty-based acquisition provides one general basis for this adaptive-learning stage (Gal et al., 2017), although fracture-specific acquisition should also consider physical response indicators, sample diversity, and fidelity-dependent cost. Figure 1. summarizes the proposed hierarchy and the feedback loop through which states exceeding a prescribed uncertainty, discrepancy, physics-based, or applicability-domain threshold are returned to high-fidelity simulation and, where feasible, targeted experimentation.

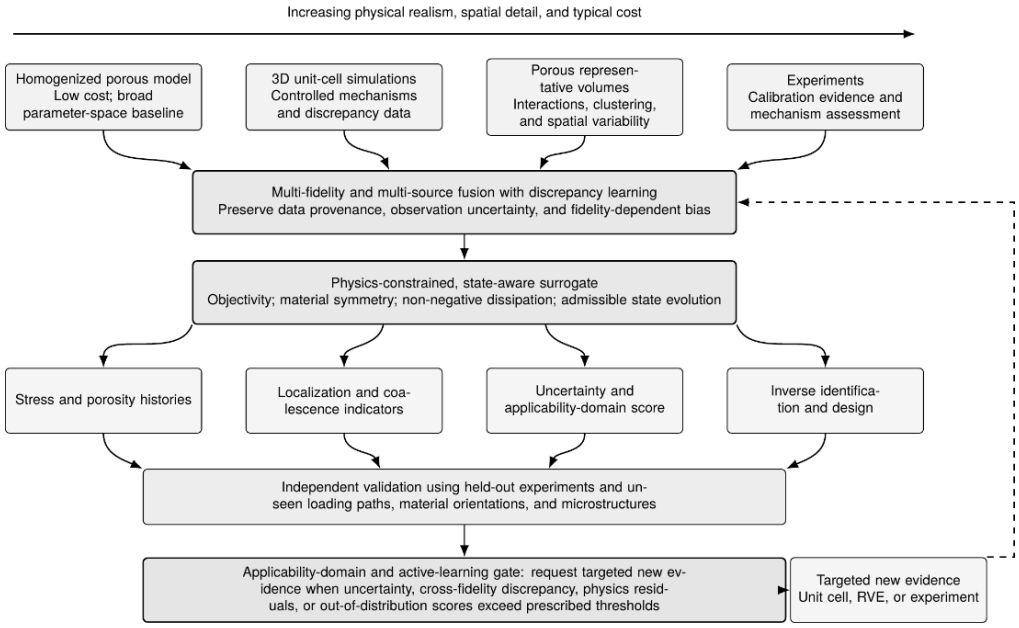

Figure 1. Proposed physics-anchored multi-fidelity and multi-source workflow for AI-enhanced modelling of void evolution and ductile failure.

Model structure and physical constraints

For an extension of the benchmark to path-dependent loading, one practical architecture is a modular, state-aware recurrent surrogate that predicts incremental updates of macroscopic stress and porosity. The model inputs should include the current loading increment, material and loading descriptors, and explicit physical state variables such as accumulated plastic strain, current porosity, and void-shape descriptors. A recurrent hidden state may supplement these variables, but it should not replace physically interpretable state information. An auxiliary event-probability head may be used to predict the onset of localization or coalescence, provided that each event is defined using a reproducible physical or experimental criterion and that corresponding training labels are available.

For idealized spherical-void cells, the microstructure can initially be described using scalar geometric variables. A separate voxel- or graph-based geometry encoder can subsequently be introduced for realistic porous microstructures. The resulting microstructural embedding can be supplied to the same state-update model together with the material and loading variables. This modular strategy allows the constitutive and history-dependent components to be pretrained using controlled spherical-void calculations and later fine-tuned using non-spherical or statistically realistic microstructure data. The training objective should separate physically distinct outputs and constraints rather than combine them within a single mean-square-error term. A representative loss can be written as

$$L = \lambda_\sigma L_\sigma + \lambda_f L_f + \lambda_c L_{\text{event}} + \lambda_{\text{field}} L_{\text{field}} + \lambda_{\text{eq}} L_{\text{eq}} + \lambda_{\text{th}} L_{\text{diss}} + \lambda_{\text{sym}} L_{\text{sym}} + \lambda_u L_{\text{cal}}. \quad (2)$$

Here, L_σ measures errors in the macroscopic stress history or stress increment, while L_f measures errors in porosity evolution and may assign additional importance to regions of rapid void growth or collapse. L_{event} evaluates predictions of clearly defined localization or coalescence events, whereas L_{field} measures errors in predicted local stress, strain, or damage-related fields. When full-field outputs are considered, L_{eq} may additionally penalize local equilibrium residuals.

For a surrogate restricted to macroscopic response histories, the corresponding weight λ_{eq} should be set to zero. The term L_{diss} penalizes negative mechanical dissipation and must be formulated consistently with the retained free energy and internal-variable structure (Masi et al., 2021; Meyer and Ekre, 2023). The term L_{sym} enforces frame indifference and the transformations associated with the adopted material-symmetry group without removing genuine directional dependence. Finally, L_{cal} assesses the calibration of probabilistic predictions or prediction intervals and is applicable only when uncertainty is explicitly modelled.

Before the individual loss terms are combined, their variables and residuals should be nondimensionalized or otherwise normalized. Adaptive or gradient-balanced weights may subsequently be used because the loss components can differ substantially in magnitude and learning rate. Such adaptation should not, however, reduce a physically important constraint to a negligible contribution. Hard constraints are preferable when an exact and differentiable parameterization is available.

For example, predicted porosity should remain within its physically admissible bounds. In a loading domain known to produce monotonic void growth, a positive parameterization may enforce $\Delta f \geq 0$. A broader model covering low-triaxiality, reverse-loading, or compressive states must instead permit physically admissible negative porosity increments associated with void contraction or collapse. The sign of the porosity increment should therefore follow the retained porous-plasticity structure and current material state rather than a simple tensile–compressive classification.
The surrogate should also include an applicability-domain and abstention mechanism. A prediction should be rejected or redirected when epistemic uncertainty, ensemble disagreement, cross-fidelity discrepancy, physics residuals, or an out-of-distribution score exceeds a threshold calibrated using held-out simulations and experiments. The workflow should then select an additional unit-cell calculation, representative volume simulation, or targeted experiment according to the expected information gain and cost, and add the resulting evidence to the training database.

This active-learning loop turns the surrogate into a controlled computational accelerator rather than an opaque replacement for the high-fidelity model. The stored record should include the input state, model version, uncertainty measures, decision threshold, abstention decision, selected fidelity level, and result of the subsequent high-fidelity evaluation.

Validation, uncertainty, and deployment

Random splitting of individual increments is inadequate for a path-dependent damage surrogate because neighbouring increments from the same loading trajectory are strongly correlated. Assigning increments from one trajectory to both the training and test sets can therefore produce overly optimistic error estimates. Complete trajectories should be treated as the basic grouping unit when constructing training, validation, calibration, and test partitions.

Validation should proceed through increasingly demanding levels. A first diagnostic may withhold selected increments from otherwise known trajectories, but this tests only local reconstruction or interpolation. The next levels should include complete trajectories excluded from training, unseen combinations of stress triaxiality T and Lode parameter L within the sampled domain, and controlled extrapolation toward unsampled values of T, L, the strength-differential parameter k, or the hardening parameters. More demanding tests should introduce unseen loading directions, void shapes, void orientations, or microstructures. Experimental observations should provide an external validation level and should not be used simultaneously for model training or hyperparameter selection. Performance limited to withheld increments or trajectories demonstrates trajectory-level interpolation but does not establish constitutive generalization across loading, material, and microstructural domains.

Error reporting should distinguish the different physical outputs. Stress errors should be reported for the relevant components or effective stress measure over the complete loading history, including maximum errors and errors near rapid void evolution or localization. Porosity errors should be assessed separately because accurate stress prediction does not guarantee accurate damage-related evolution. A conventional relative error based on the instantaneous porosity may become unstable when f is small. It is therefore preferable to report the absolute error in porosity together with errors in normalized porosity and normalized porosity change, for example,

$$e_f^{\mathrm{abs}} = \left|\hat{f} - f\right|, \quad e_f^{\mathrm{norm}} = \left|\frac{\hat{f}}{f_0} - \frac{f}{f_0}\right| \qquad (3)$$

where $\hat{f}$ and f denote the predicted and reference porosities, respectively. For loading domains that include void contraction or collapse, these quantities should be described as measures of porosity evolution rather than only void growth.

Critical-event metrics should report the error in the strain or loading increment associated with a clearly defined event. Depending on the available labels, such events may include the onset of rapid porosity evolution, localization, coalescence, or loss of load-carrying capacity. Each event criterion should be stated explicitly and applied consistently to the training and test data. Coalescence errors should not be reported when the reference simulations or experiments do not provide a reproducible coalescence label.

For full-field predictions, a global pixel- or voxel-wise norm should be supplemented by localization-sensitive measures. Relevant metrics include the error in peak principal strain, the distance between predicted and reference hotspots, the position and orientation of the localization band, and the overlap of thresholded high-strain or high-damage regions. Such measures are necessary because a model may achieve a small average field error while smoothing, displacing, or entirely missing the critical localization region.

Predictive uncertainty may contain contributions from aleatory variability in microstructures and experiments, epistemic uncertainty associated with limited training coverage, and model-form uncertainty inherited from the adopted physical approximation. Although model-form uncertainty may be viewed as a component of epistemic uncertainty, representing it separately is useful in a multi-fidelity workflow. Deep ensembles, Bayesian neural networks, and Gaussian process surrogates can provide predictive distributions or uncertainty estimates, while conformal prediction can construct prediction intervals with coverage guarantees under its stated calibration assumptions (Lakshminarayanan et al., 2017; Romano et al., 2019). These methods should not be treated as interchangeable. Deep ensembles provide empirical uncertainty estimates but no universal coverage guarantee. Bayesian neural networks and Gaussian processes depend on the adopted prior, likelihood, kernel, and inference approximation. Conformal prediction provides marginal coverage under exchangeability but does not by itself separate aleatory and epistemic uncertainty. Because increments from a loading trajectory are correlated, conformal calibration should use complete held-out trajectories or appropriately grouped trajectory-level samples rather than randomly selected increments. Calibration should be examined not only globally but also across relevant ranges of T, L, k, hardening, loading direction, and microstructure.

Multi-fidelity discrepancy models are particularly valuable because they allow uncertainty in the homogenized approximation and uncertainty in the learned correction to be represented separately. This decomposition is not automatically identifiable, especially when high-fidelity data are sparse. It should therefore be supported by paired or otherwise well-connected evaluations across fidelity levels, and uncertainty in the discrepancy model should be retained in the final prediction rather than replaced by a deterministic correction. Deployment should be restricted to a declared

applicability domain. Before a surrogate prediction is accepted, the workflow should verify that the input lies within the validated loading, material, and microstructural domain; that predictive uncertainty and out-of-distribution scores remain below prescribed thresholds; and that equilibrium, dissipation, symmetry, and admissibility checks are satisfied where applicable. These thresholds should be selected using held-out trajectories and linked to a predefined acceptable error level rather than chosen only for computational convenience.

When an acceptance criterion is violated, the surrogate should abstain and redirect the case to an appropriate unit-cell calculation, representative-volume simulation, or targeted experiment. Each deployed evaluation should record the model version, input state, predicted outputs, uncertainty measures, physical residuals, acceptance or abstention decision, and any subsequent high-fidelity result. Periodic reassessment on newly acquired data is then required to determine whether the applicability domain can be expanded or whether the model must be retrained.

Comparative assessment and research agenda

Applications with the strongest current evidence

Among the applications reviewed here, the strongest current evidence supports repeated forward evaluation and parameter calibration. These tasks retain a clearly defined numerical or experimental reference and address identifiable computational bottlenecks. For a forward surrogate, predictions can be compared directly with the corresponding finite-element or experimental responses during validation. The model should be accepted only when errors in stress, porosity, and any relevant event indicators satisfy predefined tolerances. During deployment, cases that violate uncertainty, applicability-domain, or physical-admissibility criteria should be redirected to the original finite-element solver rather than assigned an unqualified surrogate prediction.

Machine learning can also reduce the repeated computational cost of inverse identification. After an offline campaign of simulations and model training, a forward surrogate can be embedded within an optimizer, or an inverse model can provide rapid parameter estimates from measured responses. Previous studies have demonstrated such strategies for ductile-fracture loci, modified GTN formulations, semi-coupled damage models, and TEPLA parameters (Baltic et al., 2021; Chen et al., 2021; Yao et al., 2022; Blaschke et al., 2023; O'Connor et al., 2024). The computational advantage is greatest when the same constitutive formulation is calibrated repeatedly for material or processing conditions represented within the trained parameter domain. The initial cost of generating the offline database must nevertheless be reported, and faster inversion should not be interpreted as resolution of parameter non-identifiability. These applications improve engineering productivity without requiring the network to replace or independently infer the complete fracture mechanism.

Validated surrogates are also well suited to global sensitivity analysis and dense exploration of a continuous parameter domain. The 2022 unit-cell study examined selected combinations of matrix hardening, strength differential, stress triaxiality T, Lode parameter L, and loading direction (Hashem-Sharifi et al., 2022). A sufficiently accurate surrogate could extend this analysis by estimating individual and interaction effects, identifying regions of rapid response variation, and locating domains in which the Stewart–Cazacu homogenized formulation departs from the three-dimensional unit-cell response. Such analyses should be performed only within regions where the surrogate has been independently validated, because otherwise the inferred sensitivities may partly reflect surrogate error.

The discrepancy analysis should treat macroscopic stress and porosity as separate outputs. For a vector of loading, material, and microstructural variables x, representative discrepancies may be written as

$$\Delta_\sigma(\mathbf{x}) = \hat{\sigma}_{\mathrm{UC}}(\mathbf{x}) - \sigma_{\mathrm{hom}}(\mathbf{x}),$$
$$\Delta_f(\mathbf{x}) = \hat{f}_{\mathrm{UC}}(\mathbf{x}) - f_{\mathrm{hom}}(\mathbf{x}) \quad (4)$$

where the subscript UC denotes the unit-cell-informed prediction and hom denotes the homogenized response. This separation is necessary because agreement in macroscopic stress does not imply comparable accuracy in porosity evolution.

A discrepancy map could support an adaptive hierarchy of models. The homogenized formulation could be retained in validated regions with small stress and porosity discrepancies. Where systematic differences are present but sufficiently sampled, a unit-cell-informed discrepancy correction could be applied. Cases associated with high correction uncertainty, sparse training coverage, or rapid localization should instead be redirected to the three-dimensional unit-cell solver. A smooth, uncertainty-aware gating strategy is preferable to abrupt switching between models because it reduces the risk of discontinuous predictions at the boundaries of their applicability domains.

Where caution is still required

Claims of broadly transferable constitutive learning remain insufficiently supported for ductile-fracture applications. Available training datasets generally cover only a limited portion of the space of materials, textures, loading paths, material orientations, and microstructures. Broader deployment would introduce further dimensions, including temperature and strain-rate dependence, that are rarely covered simultaneously. A model should therefore be described in terms of its declared applicability domain rather than presented as a universal constitutive representation. Localization creates sharp spatial gradients and rapid changes in the evolution of internal variables. Small errors accumulated during incremental state updates can consequently shift the predicted strain associated with a defined localization, coalescence, or failure event. A network may also reproduce the macroscopic stress response while learning an incorrect damage-related mechanism. This possibility is particularly relevant to the present benchmark because comparisons between the homogenized Stewart–Cazacu formulation and three-dimensional unit-cell calculations showed that agreement in macroscopic stress did not necessarily imply comparable agreement in porosity evolution (Hashem-Sharifi et al., 2022). Validation should therefore assess stress, porosity, void-shape evolution, and event indicators separately.

Extrapolation is particularly hazardous when tension–compression asymmetry is present. A model trained predominantly under tensile and positive-triaxiality conditions may not learn void contraction or collapse, reverse-loading behaviour, or the coupled effects of mean stress, deviatoric stress state, and the strength-differential parameter. These effects may also depend on the initial void shape and orientation (Hashem-Sharifi et al., 2025). Claims involving compression, low-triaxiality loading, or load reversal should therefore be supported by corresponding training and independent validation cases rather than inferred from tensile interpolation.Data augmentation must also respect the actual material symmetry group. Arbitrary rotational or reflection-based augmentation that is appropriate for isotropic data can erase genuine orientation effects in an orthotropic material. The material axes, loading variables, boundary conditions, and tensor-valued outputs must be transformed consistently, and only physically admissible symmetry operations should be treated as equivalent data.

Finally, the designation “physics-informed” does not by itself establish physical correctness. An incomplete governing model, omitted state variable, inappropriate constraint, or poorly balanced loss function can bias the network toward an incorrect response. Physical constraints should

therefore complement, rather than replace, validation against independent simulations and experiments over the declared loading, material, and microstructural domain.

Reproducibility and minimum reporting requirements

Reproducibility remains an important weakness in many mechanics–machine-learning studies. Minimum reporting should include the complete input ranges, physical units, sampling strategy, data provenance and fidelity level, number of independent simulations, number and length of loading trajectories, preprocessing and normalization procedures, and criteria used to exclude failed or non-converged cases. The number of increments should not be presented as the number of independent simulations because increments belonging to the same trajectory are strongly correlated. Similarly, multiple subvolumes, crops, or realizations derived from the same specimen or base microstructure should not be treated as fully independent samples.

The construction of the training, validation, uncertainty calibration, and test sets should be documented at the level of complete trajectories, material-parameter combinations, void geometries, microstructures, and experimental specimens. Related trajectories or microstructural realizations generated from the same parent case should remain within the same partition to avoid information leakage. The uncertainty-calibration set is required only when probabilistic calibration, conformal prediction, or threshold selection is performed and should remain separate from the final test data. The model description should include its architecture, number of trainable parameters, activation functions, optimizer, learning-rate schedule, loss terms and weights, stopping criterion, hyperparameter-selection procedure, random seeds, and the number of independent training repetitions. Performance over repeated runs should be reported using an appropriate measure of central tendency and variability rather than only the best-performing realization. The uncertainty method, applicability-domain definition, and any acceptance, abstention, or fallback thresholds should also be reported, together with the data used to determine those thresholds. The high-fidelity reference model should be documented with comparable detail. Required information includes the element formulation, mesh density and convergence assessment, boundary and loading conditions, increment-control strategy, nonlinear convergence tolerances, constitutive integration procedure, implementation verification, and criteria for accepting or rejecting a simulation. When local fields are used as learning targets, the spatial discretization, field interpolation, registration, and any smoothing or filtering procedures should also be stated. Claims of computational acceleration should report the cost of database generation and model training in addition to online inference time. The hardware, software and solver versions, numerical precision, parallelization strategy, and timing procedure should be documented. Offline and online costs should be reported separately, and speedup factors should be calculated relative to a clearly defined reference workflow using equivalent output requirements and accuracy criteria.

Release of datasets, source code, trained model parameters, fixed data partitions, and evaluation scripts is strongly encouraged. Dataset and model versions should be identified unambiguously so that reported results can be linked to the exact data and implementation used. When proprietary or experimental restrictions prevent full release, a reduced or synthetic benchmark should be provided together with sufficient implementation details to reproduce the principal numerical findings. A synthetic benchmark can support methodological comparison but should not be presented as a substitute for validation against independent experimental observations. The synthetic or reduced benchmark should preserve the main difficulties of the target problem, including path dependence, anisotropy, tension–compression asymmetry, nonlinear porosity evolution, geometric variability, and, where relevant, localization. All target quantities and event labels should be defined operationally. In particular, localization or coalescence labels should be included only when they can be extracted using a reproducible simulation- or experiment-based criterion.

For asymmetric porous materials, a useful community benchmark could contain standardized proportional and non-proportional loading paths over physically admissible combinations of stress triaxiality T and Lode parameter L, multiple signs and magnitudes of the strength-differential parameter k, several hardening responses and initial porosities, multiple material orientations, and spherical, oblate, and prolate voids with selected orientations. More advanced levels could add statistically realistic pore distributions, grain-scale information, and image-based representative volume elements. The benchmark should provide both the input descriptors and the corresponding macroscopic histories, local fields, geometric evolution measures, and event labels that are supported by the reference calculations.

Training and blind-test partitions should be defined before model development. The blind data should remain inaccessible during architecture selection, hyperparameter tuning, loss-weight adjustment, and uncertainty-threshold calibration. Blind evaluation should preferably contain a separate interpolation set, consisting of unseen trajectories and parameter combinations within the sampled domain, and a challenge set containing controlled extrapolation, unseen loading-path families, void geometries, orientations, or microstructures. For continuous scalar parameters, cases outside the convex hull of the training inputs may be identified as one form of extrapolation. However, convex-hull membership alone is insufficient for categorical geometries, high-dimensional images, or complete loading histories. Unseen void classes, material orientations, microstructural families, and loading-path types should therefore be identified separately rather than combined with ordinary interpolation results.

Common evaluation metrics should separately assess stress histories, porosity evolution, void-shape evolution, clearly defined localization or coalescence events, full-field hotspots, uncertainty calibration, and abstention behaviour. Results should be reported separately for interpolation, controlled extrapolation, unseen geometries, and experimental validation. Computational comparisons should distinguish offline database and training costs from online evaluation time. Such a benchmark would make claims of speed, accuracy, uncertainty, and generalization more comparable across methods while ensuring that each claim remains tied to a clearly specified validation domain.

Priority research directions

A staged research agenda for the present material class can be organized around four linked priorities. The first and most direct priority is the development of a multi-fidelity surrogate that learns the discrepancy between the Stewart–Cazacu homogenized formulation and three-dimensional unit-cell simulations. This task is closely connected to the 2022 benchmark, retains a clearly defined low-fidelity baseline, and preserves physical interpretability (Hashem-Sharifi et al., 2022). Established multi-fidelity strategies provide a basis for combining inexpensive but biased predictions with a smaller number of high-fidelity calculations (Kennedy and O'Hagan, 2000; Perdikaris et al., 2017).

The learned discrepancy should be state- and path-dependent and should treat stress and porosity as separate outputs. It should preserve the admissible bounds and limiting behaviour of the retained physical model and should not be extrapolated as an unrestricted deterministic correction. Each corrected prediction should therefore be accompanied by calibrated predictive uncertainty and an applicability-domain. assessment. When uncertainty or disagreement between fidelity levels exceeds a validated threshold, the case should be redirected to an additional unit-cell or representative-volume calculation.

The second priority is probabilistic inverse identification of a selected and demonstrably identifiable subset of CPB06 and porous-damage parameters. The identification database should combine tensile and compressive responses, multiple material orientations, full-field strain information, and, where available, local thickness or void-fraction measurements. A Bayesian

formulation should define the parameter priors, experimental-noise model, likelihood, and treatment of model-form discrepancy before its output is described as a posterior distribution. The posterior should expose parameter correlations and weakly identifiable directions rather than report only one optimal parameter vector. Active experimental design can then rank candidate loading directions, stress states, specimen geometries, or measurement types according to their expected information gain or expected reduction in posterior uncertainty. Experimental cost, feasibility, and sensitivity to measurement noise should be considered together with identifiability. Where simultaneous identification of all parameters remains ill-conditioned, a staged procedure should separate hardening, anisotropy and strength-differential parameters from subsequent porosity and damage-evolution parameters.

The third priority is a controlled extension from a single idealized void to interacting-void systems and statistically realistic microstructures. A credible progression should move from single-void cells to multiple interacting voids, statistically generated representative volumes, and finally image-based microstructures (Vishnu et al., 2023; Hashem-Sharifi et al., 2025). Voxel-based encoders may be used when the complete spatial morphology is available, while graph-based representations may describe variable numbers of pores or grains. In a graph representation, the nodes and edges should have explicit physical definitions. Pores may be represented as nodes, while edges may encode candidate interaction paths using ligament thickness, centre-to-canter spacing, relative size, orientation, and loading direction. Such a graph does not establish the mechanical interaction automatically; its predictive value should be compared with voxel-based and conventional microstructural descriptors. Predictions of localization or coalescence paths should be attempted only when the corresponding events are defined by reproducible criteria and supported by suitable simulation or experimental labels. Scale sensitivity should be evaluated explicitly. The effects of voxel resolution, field of view, representative volume size, segmentation, minimum retained pore size, and microstructural sampling should be reported. Local-field predictions should be validated using hotspot, localization band, and high-damage-region metrics rather than only global field norms.

The fourth priority is robust embedding of learned local updates within a conventional finite-element solver. When an implicit solver is used, the learned constitutive update should provide a consistent algorithmic tangent or a demonstrably robust approximation of it. The model should satisfy objectivity, the adopted material-symmetry group, state-variable bounds, and non-negative dissipation (Masi et al., 2021; Meyer and Ekre, 2023). It should also be tested under loading, unloading, reverse loading, cyclic histories, and changes in loading direction before being used in structural calculations. Solver convergence, time-increment convergence, and mesh objectivity should be evaluated separately. A consistent tangent may improve nonlinear convergence but does not by itself remove the mesh dependence associated with localized softening. When post-localization response is modelled, an appropriate characteristic-length, nonlocal, gradient, or other regularization strategy is required before mesh-independent failure predictions can be claimed.

Physics-constrained hybrid correction models should be prioritized before fully learned yield surfaces or complete constitutive replacements because they preserve recognizable state variables, limiting behaviour, and a direct connection to established material models. The learned correction must nevertheless be constrained and validated so that it does not violate convexity, symmetry, thermodynamic admissibility, or porosity bounds. Uncertainty-aware active learning should operate across all four priorities. When the model encounters an unfamiliar loading path, material state, or microstructure, the workflow should select an additional homogenized calculation, unit-cell simulation, representative-volume analysis, or targeted experiment according to predictive uncertainty, expected information gain, and cost (Gal et al., 2017). The resulting data and the reason for each acquisition should be recorded so that expansion of the model’s applicability domain remains controlled and traceable.

Conclusions

Artificial intelligence can extend the practical use of physics-based modelling of ductile failure, but its credibility depends on a disciplined division of tasks between physical models, high-fidelity evidence, and data-driven methods. Homogenized porous formulations, three-dimensional unit-cell simulations, statistically realistic representative volumes, and experiments provide complementary information at different levels of fidelity. Machine learning can then enable rapid forward evaluation, surrogate-assisted parameter identification, high-dimensional sensitivity analysis, and microstructure-aware prediction. Among the applications considered in this review, the most defensible near-term uses are surrogate-assisted calibration and multi-fidelity correction of established constitutive models rather than their unrestricted replacement.

For anisotropic materials with tension–compression asymmetry, the model inputs and physical constraints should represent the interacting loading, material, and microstructural variables relevant to the declared applicability domain. These may include stress triaxiality T, Lode parameter L, strength-differential parameter k, material orientation, hardening, initial porosity, void morphology, and loading history. Agreement in nominal stress alone is insufficient to demonstrate that the correct damage mechanism has been learned. Porosity evolution should be validated separately, together with void-shape evolution, clearly defined critical events, and localized fields when these quantities are predicted. Random pointwise partitioning should be replaced by trajectory-, geometry-, specimen-, and domain-based validation. Reverse loading, non-proportional paths, unseen morphologies, and controlled extrapolation should be reported as separate challenge cases rather than combined with ordinary interpolation results. A particularly promising next step is an uncertainty-aware, multi-fidelity discrepancy surrogate constructed around the 2022 CPB06 unit-cell benchmark (Hashem-Sharifi et al., 2022). Such a model would retain the Stewart–Cazacu formulation as a physically interpretable baseline while learning state- and path-dependent corrections to its stress and porosity predictions. Preservation of admissible bounds, symmetry, dissipation, and limiting behaviour should be imposed or independently verified rather than assumed. Corrected predictions should also be accompanied by calibrated uncertainty and an applicability-domain assessment.

The broader conclusion of this review is that credible AI for ductile failure should be anchored to established physics, supplied with interpretable state information, validated using independent trajectory- and domain-level tests, and evaluated separately for stress, damage evolution, localization, and uncertainty. When a new loading path, material state, or microstructure lies outside the validated applicability domain, the model should abstain and request an appropriate unit-cell calculation, representative-volume simulation, or experiment. In this role, AI functions as a controlled and traceable computational accelerator rather than an opaque substitute for fracture mechanics.

Declarations

Authorship contribution statement. S. Hashem-Sharifi Pishvaei: Conceptualization, Methodology, Investigation, Writing–original draft, Writing–review and editing, and Visualization.

Funding. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest. The author declares no known competing financial interests or personal relationships that could have influenced the work reported in this article.

Data availability. No new research data were generated for this review. The publications and other sources underlying the analysis are cited in the manuscript.

References

Alves, J. L. and Cazacu, O. (2015). Micromechanical study of the dilatational response of porous solids with pressure-insensitive matrix displaying tension-compression asymmetry. European Journal of Mechanics - A/Solids, 51:44–54.

Baltic, S., Asadzadeh, M. Z., Hammer, P., Magnien, J., Gänser, H.-P., Antretter, T., and Hammer, R. (2021). Machine learning assisted calibration of a ductile fracture locus model. Materials & Design, 203:109604.

Benzerga, A. A. and Leblond, J.-B. (2010). Ductile fracture by void growth to coalescence. In Advances in Applied Mechanics, volume 44, pages 169–305. Elsevier

Blaschke, D. N., Nguyen, T. T. T., Nitol, M. S., O'Malley, D., and Fensin, S. J. (2023). Machine learning based approach to predict ductile damage model parameters for polycrystalline metals. Computational Materials Science, 229:112382.

Borkowski, L., Sorini, C., and Chattopadhyay, A. (2022). Recurrent neural network-based multiaxial plasticity model with regularization for physics-informed constraints. Computers & Structures, 258:106678.

Cazacu, O., Plunkett, B., and Barlat, F. (2006). Orthotropic yield criterion for hexagonal closed packed metals. International Journal of Plasticity, 22(7):1171–1194.

Cazacu, O. and Stewart, J. B. (2009). Plastic potentials for porous aggregates with the matrix exhibiting tensioncompression asymmetry. Journal of the Mechanics and Physics of Solids, 57:325–341.

Chen, D., Li, Y., Yang, X., Jiang, W., and Guan, L. (2021). Efficient parameters identification of a modified GTN model of ductile fracture using machine learning. Engineering Fracture Mechanics, 245:107535.

Dornheim, J., Morand, L., Nallani, H. J., and Helm, D. (2024). Neural networks for constitutive modeling: From universal function approximators to advanced models and the integration of physics. Archives of Computational Methods in Engineering, 31:1097–1127.

Gal, Y., Islam, R., and Ghahramani, Z. (2017). Deep Bayesian active learning with image data. In Proceedings of the 34th International Conference on Machine Learning, volume 70 of Proceedings of Machine Learning Research, pages 1183–1192.

Gao, Z., Zhu, C., Wang, C., Shu, Y., Liu, S., Miao, J., and Yang, L. (2025). Advanced deep learning framework for multi-scale prediction of mechanical properties from microstructural features in polycrystalline materials. Computer Methods in Applied Mechanics and Engineering, 438:117844.

Ghorbanpour, S., Alam, M. E., Ferreri, N. C., Kumar, A., McWilliams, B. A., Vogel, S. C., Bicknell, J., Beyerlein, I. J., and Knezevic, M. (2020). Experimental characterization and crystal plasticity modeling of anisotropy, tension–compression asymmetry, and texture evolution of additively manufactured Inconel 718 at room and elevated temperatures. International Journal of Plasticity, 125:63–79.

Gorji, M. B., Mozaffar, M., Heidenreich, J. N., Cao, J., and Mohr, D. (2020). On the potential of recurrent neural networks for modeling path dependent plasticity. Journal of the Mechanics and Physics of Solids, 143:103972.

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. Journal of Engineering Materials and Technology, 99:2–15.

Hashem-Sharifi, S., Hosseini, N., and Vadillo, G. (2022). 3D numerical simulations and microstructural modeling of anisotropic and tension-compression asymmetric ductile materials. International Journal of Solids and Structures, 256:111936.

Hashem-Sharifi, S., Hosseini, N., and Vadillo, G. (2025). The effect of void shape on the ductile behavior of tensioncompression asymmetric materials. Engineering Fracture Mechanics, 322:111106.

Hestroffer, J. M., Charpagne, M.-A., Latypov, M. I., and Beyerlein, I. J. (2023). Graph neural networks for efficient learning of mechanical properties of polycrystals. Computational Materials Science, 217:111894.

Hosseini, N., Nieto-Fuentes, J. C., Dakshinamurthy, M., Rodríguez-Martínez, J. A., and Vadillo, G. (2022). The effect of material orientation on void growth. International Journal of Plasticity, 148:103149.

Kennedy, M. C. and O'Hagan, A. (2000). Predicting the
output from a complex computer code when fast approximations are available. Biometrika, 87(1):1–13.

Konale, A. and Srivastava, V. (2026). A physics-informed neural network for modeling fracture without gradient damage: Formulation, application, and assessment. Journal of the Mechanics and Physics of Solids, 206:106395.

Kovachki, N., Li, Z., Liu, B., Azizzadenesheli, K., Bhattacharya, K., Stuart, A., and Anandkumar, A. (2023). Neural operator: Learning maps between function spaces with applications to PDEs. Journal of Machine Learning Research, 24(89):1–97.

Lakshminarayanan, B., Pritzel, A., and Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. In Advances in Neural Information Processing Systems, volume 30.

Lian, H., Zhao, P., Zhang, M., Wang, P., and Li, Y. (2023). Physics informed neural networks for phase field fracture modeling enhanced by length-scale decoupling degradation functions. Frontiers in Physics, 11:1152811.

Linka, K., Hillgärtner, M., Abdolazizi, K. P., Aydin, R. C., Itskov, M., and Cyron, C. J. (2021). Constitutive artificial neural networks: A fast and general approach to predictive data-driven constitutive modeling by deep learning. Journal of Computational Physics, 429:110010.

Linka, K. and Kuhl, E. (2023). A new family of constitutive artificial neural networks towards automated model discovery. Computer Methods in Applied Mechanics and Engineering, 403:115731.

Lu, L., Jin, P., Pang, G., Zhang, Z., and Karniadakis, G. E. (2021). Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. Nature Machine Intelligence, 3(3):218–229.

Masi, F., Stefanou, I., Vannucci, P., and Maffi-Berthier, V. (2021). Thermodynamics-based artificial neural networks for constitutive modeling. Journal of the Mechanics and Physics of Solids, 147:104277.

Meyer, K. A. and Ekre, F. (2023). Thermodynamically consistent neural network plasticity modeling and discovery of evolution laws. Journal of the Mechanics and Physics of Solids, 180:105416.

Mirzaei, M. S., Mashayekhi, M., and Mirkhalaf, M. (2026). Physics-informed neural networks for local and nonlocal modeling in continuum damage mechanics. International Journal of Solids and Structures, 337:114052.

Noh, W., Vieira, R. B., Lambros, J., and Chew, H. B. (2026). Microscale strain field predictions from grain microstructure of polycrystalline metals using fully convolutional networks. International Journal of Solids and Structures, 327:113801.

O'Connor, A. N., Mongan, P. G., and O'Dowd, N. P. (2024). A machine learning approach to automate ductile damage parameter selection using finite element simulations. European Journal of Mechanics - A/Solids, 103:105180.

Perdikaris, P., Raissi, M., Damianou, A., Lawrence, N. D., and Karniadakis, G. E. (2017). Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 473(2198):20160751.

Raissi, M., Perdikaris, P., and Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics, 378:686–707.

Romano, Y., Patterson, E., and Candès, E. J. (2019). Conformalized quantile regression. In Advances in Neural Information Processing Systems, volume 32.

Schowtjak, A., Gerlach, J., Muhammad, W., Brahme, A. P., Clausmeyer, T., Inal, K., and Tekkaya, A. E. (2022). Prediction of ductile damage evolution based on experimental data using artificial neural networks. International Journal of Solids and Structures, 257:111950.

Su, M. M., Yu, Y., Chen, T. H., Guo, N., and Yang, Z. X. (2024). A thermodynamics-informed neural network for elastoplastic constitutive modeling of granular materials. Computer Methods in Applied Mechanics and Engineering, 430:117246.

Tran, M. T., Phan, H. C., Jung, J., Kim, S.-J., Kang, S.- H., and Lee, H. W. (2026). Multi-fidelity physics-guided graph neural networks for predicting structure–property relationships in polycrystalline metals. Materials & Design, 267:116310.

Tvergaard, V. and Needleman, A. (1984). Analysis of the cup-cone fracture in a round tensile bar. Acta Metallurgica, 32:157–169.

Vishnu, A. R., Vadillo, G., and Rodríguez-Martínez, J. A. (2023). Void growth in ductile materials with realistic porous microstructures. International Journal of Plasticity, 167:103655.

Vlassis, N. N. and Sun, W. (2021). Sobolev training of thermodynamic-informed neural networks for interpretable elasto-plasticity models with level set hardening. Computer Methods in Applied Mechanics and Engineering, 377:113695.

Yao, D., Pu, S., Li, M., Guan, Y., and Duan, Y. (2022). Parameter identification method of the semi-coupled fracture model for 6061 aluminium alloy sheet based on machine learning assistance. International Journal of Solids and Structures, 254–255:111823.

Yin, M., Zhang, E., Yu, Y., and Karniadakis, G. E. (2022). Interfacing finite elements with deep neural operators for fast multiscale modeling of mechanics problems. Computer Methods in Applied Mechanics and Engineering, 402:115027.

Yu, X., Jin, H., Linforth, S., Fang, X., Liu, J., Wang, P., and Ngo, T. (2026). Machine learning meets fracture mechanics: What can we learn? Archives of Computational Methods in Engineering.

Zheng, X., Battalgazy, B., Molkeri, A., Tsopanidis, S., Osovski, S., and Srivastava, A. (2023). Role of lengthscale in machine learning based image analysis of ductile fracture surfaces. Mechanics of Materials, 181:104661.