REVIEW Open access

Representation Learning for Materials Microstructures — Conceptual Advances and Interpretability Challenges: A Review Study

Hiroshi Nakamura[1*], Yuta Kato[1], Aiko Morita[2]

Abstract

The field of materials science has witnessed a transformative shift with the advent of representation learning techniques, particularly for analyzing complex microstructures. This review synthesizes recent conceptual advances in representation learning, including deep neural networks, autoencoders, and vision transformers, applied to microstructure data for tasks such as property prediction, inverse design, and evolution modeling. We explore how these methods extract latent features from high-dimensional microstructure images, enabling efficient computation and discovery of structure-property relationships. However, interpretability remains a significant challenge, as black-box models often obscure the physical meaning of learned representations, hindering trust and scientific insight. We discuss strategies for enhancing interpretability, such as attention mechanisms, heat maps, and post-hoc explanations, drawing from recent studies in alloy microstructures and additive manufacturing. The review highlights the integration of domain knowledge to disentangle representations and address data scarcity issues. By examining case studies in metals, ceramics, and composites, we identify gaps in current approaches, including bias in learned features and limited generalizability across materials classes. Ultimately, this review aims to guide future research toward interpretable representation-learning frameworks that accelerate materials design and foster a deeper understanding of microstructural phenomena.

Keywords Interpretability, Inverse design, Representation learning, Materials microstructures, Deep learning, Structure-property relationships

*Correspondence:
Hiroshi Nakamura
hiroshi.nakamura@gmail.com
[1] Department of Intelligent Materials Engineering, Faculty of Engineering, Nagoya University, Nagoya, Japan
[2] Department of AI Materials Systems, Faculty of Engineering, Tohoku University, Sendai, Japan

Introduction

Materials microstructures—the internal arrangement of phases, grains, defects, and interfaces at the microscale—play a decisive role in governing macroscopic material behavior, including mechanical strength, thermal and electrical transport, chemical stability, and long-term degradation pathways [1]. The sensitivity of these properties to subtle variations in microstructural morphology has long motivated detailed characterization and analysis. Classical approaches to microstructure analysis have historically relied on manual or semi-automated characterization techniques, such as optical microscopy, scanning and transmission electron microscopy, and diffraction-based methods, combined with low-order statistical descriptors including grain size distributions, phase volume fractions, aspect ratios, and orientation textures [2]. While these descriptors have proven valuable for well-controlled and near-equilibrium systems, they are fundamentally

limited in their ability to capture spatial heterogeneity, long-range correlations, hierarchical organization, and non-linear feature interactions that increasingly define modern materials.

These limitations are particularly pronounced in advanced materials systems such as high-entropy alloys, architected and multifunctional composites, and additively manufactured components. In such systems, complex processing routes—often involving rapid solidification, extreme thermal gradients, or layer-by-layer fabrication—induce non-equilibrium microstructures characterized by anisotropy, stochastic defects, and coupled features across multiple length scales [3, 4]. Traditional descriptor-based analyses struggle to represent these complexities in a manner that is both compact and predictive. Concurrently, the rapid maturation of high-throughput characterization techniques—including electron backscatter diffraction (EBSD), focused ion beam (FIB) serial sectioning, and synchrotron-based X-ray tomography—has led to an unprecedented growth in microstructure datasets, not only in volume but also in dimensionality and resolution [5, 6]. This data abundance has exposed a widening gap between data acquisition capabilities and the analytical tools available to extract physically meaningful insight, motivating the integration of advanced data-driven methodologies.

Representation learning, a central paradigm within modern machine learning, has emerged as a powerful response to this challenge by enabling the automated extraction of informative features directly from raw, high-dimensional data [7]. Rather than relying on hand-crafted descriptors defined a priori, representation learning seeks to learn transformations that map microstructure data—such as two-dimensional images, three-dimensional voxelized volumes, or graph-based encodings—into compact latent embeddings that preserve salient structural information relevant to downstream tasks. Deep learning architectures, most notably convolutional neural networks (CNNs) and graph neural networks (GNNs), have driven major conceptual advances in this area due to their capacity to model spatial locality, hierarchical feature composition, and relational structure intrinsic to materials microstructures [8, 9]. These models can, in principle, discover complex descriptors that are difficult or impossible to formalize analytically, offering new avenues for microstructure quantification.

The adoption of representation learning has enabled a broad range of applications in materials science, including microstructure classification and segmentation, similarity search and clustering, anomaly and defect detection, and direct structure–property prediction. These capabilities support emerging paradigms such as inverse materials design, accelerated materials discovery, and closed-loop experimentation, in which predictive models guide synthesis or processing decisions in real time [10, 11]. Despite these successes, representation learning models often operate as black boxes, learning latent features that lack clear physical interpretation. This opacity raises fundamental concerns regarding interpretability, physical consistency, and generalizability—particularly when models are applied across different processing conditions, materials classes, or data regimes that differ from those encountered during training.

The objectives of this review are threefold. First, it surveys recent conceptual and methodological advances in representation learning for materials microstructures, with particular emphasis on innovations in model architectures, learning paradigms (including supervised, self-supervised, and hybrid approaches), and data representations spanning images, volumes, and graphs [12, 13]. Second, it critically examines interpretability challenges, including the disconnect between learned representations and established physical descriptors, and reviews emerging strategies for embedding physical constraints, invariances, and explanatory mechanisms into representation learning frameworks [14, 15]. Third, it identifies open research directions aimed at integrating domain knowledge more effectively, improving robustness under data scarcity and distribution shift, and balancing predictive performance with scientific interpretability and trustworthiness [16]. Focusing on peer-reviewed studies, this review provides a timely synthesis of the field, situating representation learning within broader epistemic and methodological debates in materials artificial intelligence and highlighting pathways toward more physically grounded, generalizable, and actionable models.

Main Text

Conceptual advances in representation learning for microstructures

Representation learning for materials microstructures has undergone a decisive shift from manually engineered descriptors toward fully data-driven paradigms, in which models learn hierarchical feature representations directly from raw microstructure data [17-19]. Early computational approaches relied on hand-crafted features—such as texture statistics, Fourier descriptors, or two-point correlation functions—which encoded domain intuition but were inherently limited in scalability, adaptability, and representational richness [20]. These methods often struggled to generalize across materials systems or processing conditions, as their expressiveness was constrained by predefined assumptions about relevant structural features.

Deep learning has fundamentally transformed this landscape by enabling end-to-end learning of multi-scale representations that adapt to the complexity of microstructures. Through layered architectures, models can simultaneously encode fine-scale features, such as grain boundaries or precipitate morphologies, and higher-order spatial arrangements, including phase connectivity and mesoscale heterogeneity [21]. This hierarchical abstraction represents a conceptual advance, shifting microstructure analysis from descriptor selection to representation discovery.

A central development in this evolution is the widespread adoption of convolutional neural networks (CNNs) for microstructure segmentation, classification, and feature extraction. CNN-based architectures have been successfully optimized for the semantic segmentation of complex metallic microstructures, such as steels, enabling the accurate identification of phases including ferrite, bainite, and martensite under varied processing conditions [22]. Building on this foundation, vision transformers (ViTs) have emerged as a complementary paradigm that replaces locality-biased convolutional filters with global self-attention mechanisms. By explicitly modeling long-range spatial dependencies, ViTs have demonstrated superior performance in tasks such as property prediction from simulated microstructure datasets, particularly where non-local interactions dominate material behavior [18, 23].

Generative representation learning constitutes another major conceptual advance. Variational autoencoders (VAEs) and generative adversarial networks (GANs) have enabled the synthesis of statistically realistic microstructures that preserve salient structural characteristics of experimental data [24, 25]. Beyond data augmentation, these models offer latent spaces that can be interrogated to explore microstructure variability and continuity, thereby supporting exploratory design and sensitivity analysis. Such generative embeddings move representation learning beyond prediction toward hypothesis generation. The evolution of representation learning paradigms for materials microstructures, including their core assumptions and limitations, is summarized in **Table 1**.

Table 1. Representation learning paradigms for materials microstructures

Representation paradigm	Primary data form	Core conceptual contribution	Strengths	Key limitations
Hand-crafted descriptors	Images, diffraction data	Encodes prior physical intuition	Interpretable, low data demand	Limited expressiveness, poor generalization
CNN-based representations	2D/3D images	Hierarchical multi-scale feature discovery	Strong performance, spatial locality	Local bias, limited long-range reasoning
Vision transformers (ViTs)	Images, patches	Global dependency modeling via attention	Captures non-local interactions	Data-hungry, opaque attention semantics
Graph neural networks (GNNs)	Grain/phase graphs	Explicit relational and topological encoding	Physically aligned for networks	Graph construction complexity
Generative latent models (VAE/GAN)	Images, volumes	Continuous latent spaces for synthesis	Data augmentation, design exploration	Latent variables lack physical meaning
Multi-fidelity	Images +	Fusion of experimental	Robustness, noise	Epistemic mismatch risks

representations	simulations	and synthetic data	mitigation	

Recent work has further extended representation learning by integrating multi-fidelity data sources. By jointly learning from low-resolution experimental images and high-fidelity simulation outputs, models can leverage complementary information to improve robustness and reduce sensitivity to noise or missing data [26]. In parallel, inverse design frameworks have reframed representation learning as a bidirectional mapping problem, where models learn not only structure–property relationships but also property-to-structure mappings. Reinforcement learning and constrained optimization techniques have been employed to navigate these latent spaces, enabling the design of microstructures that satisfy competing performance criteria [27]. Collectively, these advances have been applied across diverse materials classes, including additively manufactured alloys and fiber-reinforced composites, underscoring the generality of representation learning as a conceptual tool in materials science [28, 29].

Interpretability challenges in representation learning

Despite substantial progress in representation learning for materials microstructures, interpretability remains a fundamental and unresolved challenge. Learned representations frequently lack direct correspondence to physically meaningful descriptors—such as dislocation density, phase transformation pathways, or damage initiation mechanisms—making them difficult to reconcile with established materials science theories and experimental intuition [30]. This opacity limits the scientific explanatory value of model outputs. It undermines confidence in their deployment, particularly in safety-critical contexts including aerospace components, nuclear materials, and structural integrity assessment, where erroneous inferences can carry severe consequences [1, 2].

A central interpretability challenge concerns feature attribution: determining which specific microstructural elements, spatial regions, or relational patterns drive a given prediction. High-capacity models can achieve impressive predictive accuracy while relying on subtle, non-obvious correlations that are inaccessible to human inspection. Such behavior is especially problematic when training data are imbalanced or have limited coverage, as underrepresented microstructural states or rare but critical defect modes may be systematically overlooked, leading to biased representations and fragile generalization under distribution shifts [3, 4].

Post-hoc interpretability techniques have been widely adopted to address these concerns partially. Methods such as saliency maps, gradient-based attribution, and layer-wise relevance propagation aim to visualize regions of microstructure images that most strongly influence model outputs [5, 6]. While these techniques can provide qualitative insights and aid debugging, they offer limited explanatory depth and do not establish causal or mechanistic validity. More recent attention-based architectures introduce inherent interpretability by explicitly weighting spatial regions or graph components during inference, as demonstrated in defect-detection studies of welded joints and additively manufactured materials [7, 8]. However, attention weights themselves may be unstable, sensitive to training perturbations, or difficult to interpret in terms of physically grounded microstructural features.

Domain-informed approaches seek to move beyond purely statistical explanations by embedding physical constraints and materials knowledge directly into representation learning frameworks. Physics-informed neural networks and constraint-aware architectures incorporate governing equations, symmetry principles, or thermodynamic consistency into the learning process, encouraging representations that are more closely aligned with physical reality [9, 10]. Despite their conceptual appeal, these methods face practical limitations, including increased computational complexity, challenges in scaling to high-dimensional microstructure data, and difficulties in maintaining transferability across materials systems governed by distinct physical regimes [11, 12]. Existing interpretability strategies and their epistemic limitations in microstructure representation learning are compared in **Table 2.**

Table 2. Interpretability strategies in microstructure representation learning

Interpretability approach	Mechanism	What it explains	What it cannot explain	Epistemic status

Saliency maps	Gradient sensitivity	Influential image regions	Physical causality	Descriptive
Layer-wise relevance propagation	Back-propagated attribution	Feature contribution paths	Mechanistic validity	Descriptive
Attention mechanisms	Learned weighting	Spatial or relational focus	Physical meaning of weights	Weakly explanatory
Physics-informed constraints	Embedded laws/symmetries	Consistency with physics	Emergent mechanisms	Partially explanatory
Hybrid ML–physics models	Coupled inference	Structure–mechanism links	Full causal closure	Proto-mechanistic

Figure 1 conceptually illustrates the trade-off between representational expressiveness and scientific interpretability, highlighting the role of physics-aware integration in narrowing the epistemic gap.

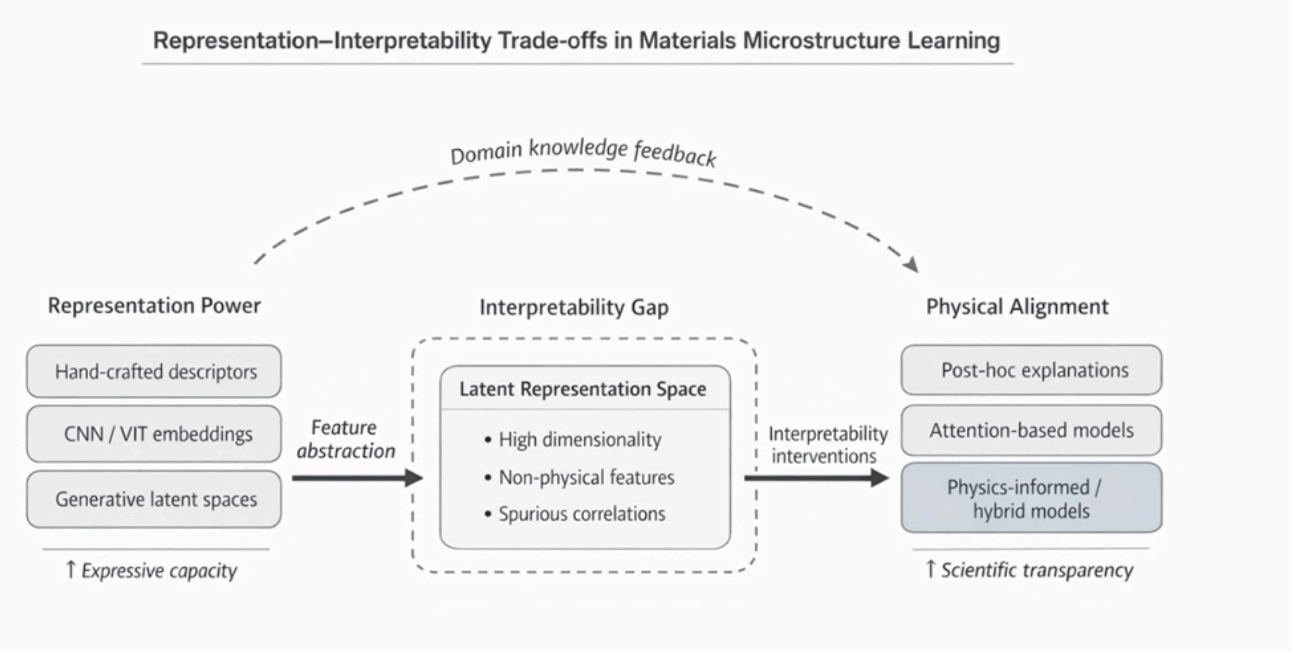

Figure 1. Representation–interpretability trade-offs in materials microstructure learning

Taken together, these challenges expose a central tension in representation learning for materials microstructures: the trade-off between expressive modeling capacity and scientific transparency. Resolving this tension requires frameworks that do not treat interpretability as an afterthought, but rather as an intrinsic design objective. Ongoing efforts toward hybrid, physics-aware, and epistemically grounded representation learning reflect a growing recognition that predictive success alone is insufficient for advancing trustworthy and scientifically meaningful materials artificial intelligence.

Applications in materials classes

Representation learning has been extensively applied across major materials classes, with methodological choices strongly shaped by differences in microstructural complexity, characteristic length scales, and dominant physical mechanisms. While a common objective across classes is to learn compact, predictive representations of structure, the form of these representations and the learning paradigms employed vary substantially.

In metallic systems, convolutional neural networks (CNNs) remain the dominant approach for microstructure analysis, owing to the relatively regular spatial organization of phases and grains and the availability of large, labeled datasets. CNN-based models have been successfully applied to tasks such as phase identification, grain size distribution estimation, and defect classification, as well as to regression problems involving yield strength, fatigue life, creep resistance, and phase stability [13, 14]. Hierarchical feature extraction in deep CNNs aligns naturally with the multiscale nature of metallic microstructures, where local features, such as dislocation structures or precipitates, interact with mesoscale grain morphology to determine macroscopic properties. More recently, self-supervised and contrastive learning approaches have been introduced to reduce reliance on labeled data, enabling models to learn transferable microstructural descriptors that generalize across alloy systems and processing conditions. These advances are facilitated by relatively mature processing–structure–property (PSP) relationships in metallurgy, which provide a physical scaffold for both model training and validation.

In ceramic materials, microstructural representation learning faces distinct challenges due to strong anisotropy, complex grain boundary networks, and the brittle nature of failure mechanisms. Fracture and strength are often governed not by local pixel-level features, but by long-range interactions among grains, pores, and interfaces. Graph neural networks (GNNs) have therefore emerged as a powerful alternative to grid-based CNNs, representing grains as nodes and grain boundaries or interfaces as edges [15, 16]. Such graph-based representations explicitly encode topology, connectivity, and neighborhood relations, enabling improved prediction of crack initiation, propagation paths, and catastrophic failure events. Extensions incorporating edge attributes (e.g., boundary misorientation, chemistry, or energy) further enhance physical fidelity. However, constructing accurate graphs from experimental microstructure data remains nontrivial, particularly in three dimensions, where segmentation errors and incomplete boundary information can propagate through the learning pipeline. Scalability is another concern, as realistic ceramic microstructures may contain tens of thousands of grains. These limitations have motivated recent work on hierarchical GNNs, coarse-grained graph abstractions, and hybrid models that couple graph representations with learned latent embeddings [24].

For composite materials, representation learning increasingly adopts multi-modal and multi-scale frameworks to account for the intrinsic heterogeneity of constituent phases and architectures. Fiber-reinforced polymers, particle-reinforced composites, and architected metamaterials often require the fusion of disparate data types, including microscopy images, volumetric scans, processing parameters, mechanical test data, and textual or symbolic descriptions of constituent materials. Deep learning architectures that integrate CNNs, recurrent networks, transformers, or attention-based fusion modules have demonstrated promise in capturing interactions between reinforcement geometry, matrix behavior, and interfacial properties [17, 18]. These representations have been applied to design optimization problems such as stiffness–weight trade-offs, damage-tolerance prediction, and impact resistance, where performance arises from complex interactions across length scales. Despite these successes, interpretability remains a significant challenge: learned features often lack clear correspondence to physically meaningful composite descriptors, such as load transfer efficiency or interfacial debonding mechanisms, complicating model validation and adoption in safety-critical applications [19, 20, 25]. Differences in representation requirements and interpretability challenges across materials classes are summarized in **Table 3**.

Table 3. Materials-class–specific representation learning characteristics

Materials class	Dominant representation	Governing microstructural features	Main success	Persistent challenge
Metals	CNNs, self-supervised embeddings	Grains, phases, precipitates	Strong PSP prediction	Dataset bias, transfer limits
Ceramics	GNNs, graph abstractions	Grain networks, pores	Fracture modeling	Graph scalability, noise

Composites	Multi-modal fusion models	Interfaces, architectures	Design optimization	Poor physical interpretability
Additive manufacturing	Hybrid CNN + attention	Defects, anisotropy	Defect detection	Rare-event underrepresentation

Across all materials classes, case studies consistently report improved predictive performance for tasks including defect detection, fatigue life estimation, and property screening when representation learning is employed. At the same time, they reveal persistent gaps in physical interpretability, particularly in systems with hierarchical or highly irregular microstructures. In such cases, latent representations may encode spurious correlations or dataset-specific artifacts, limiting confidence in extrapolative predictions and hindering integration with physics-based modeling frameworks.

Data and computational considerations

Data scarcity remains a central constraint in microstructure representation learning, especially for emerging alloys, advanced ceramics, and bespoke composite systems, where experimental characterization is costly and time-consuming. Labeled datasets are often small, imbalanced, and biased toward specific processing routes or property regimes. To mitigate these limitations, transfer learning has become a dominant strategy, with models pre-trained on large natural image datasets or materials-specific repositories and subsequently fine-tuned on limited microstructure datasets [21, 22]. While this approach can substantially improve convergence rates and predictive accuracy, it raises concerns regarding representation mismatch. Features learned from non-physical image domains may encode biases—such as texture or contrast cues—that are irrelevant or even detrimental to materials-specific inference, particularly when subtle morphological differences govern material behavior [26].

Recent computational advances have partially alleviated these challenges. The development of efficient neural architectures, including lightweight CNN variants, sparse and dilated convolutional networks, and memory-efficient GNN implementations, has enabled training on high-resolution two-dimensional and three-dimensional microstructure data without prohibitive computational cost [23, 24]. In parallel, GPU-accelerated simulation pipelines—such as phase-field, cellular automata, and crystal plasticity models—have facilitated the generation of large synthetic microstructure datasets for pretraining and data augmentation. These synthetic datasets allow systematic exploration of processing and structural parameter spaces but introduce new questions regarding epistemic validity, as simulated microstructures may fail to capture experimental noise, defects, or unmodeled physical phenomena [27].

Despite methodological and computational progress, the field lacks standardized benchmarks for evaluating representation quality independently of downstream task performance. Most studies rely on task-specific metrics, making it difficult to compare representations across models, datasets, and materials classes. Emerging proposals advocate benchmark suites that assess properties such as invariance to rotation and scaling, robustness to distribution shifts, sensitivity to physically meaningful perturbations, and alignment with known microstructural descriptors [25, 26, 28]. Without such standards, progress remains fragmented, limiting reproducibility and slowing the accumulation of transferable scientific insight across the materials informatics community.

Results and Discussion

The conceptual advances in representation learning for materials microstructures demonstrate substantial potential to accelerate materials discovery and design. Yet, they simultaneously expose deep epistemic and interpretability challenges that warrant critical scrutiny [1–3]. A primary concern is the opacity of deep learning models, where extracted representations may correlate strongly with target properties while lacking explicit correspondence to underlying physical mechanisms, such as dislocation motion, diffusion pathways, or phase transformation kinetics [4, 5]. This disconnect risks reinforcing spurious correlations, particularly in datasets with limited diversity or narrow processing windows, as observed in alloy microstructure studies that failed to generalize across unseen thermomechanical conditions [6, 7, 29].

The high dimensionality and heterogeneity of microstructure data further exacerbate computational and interpretive difficulties. Hybrid dimensionality reduction approaches—such as coupling autoencoders with linear techniques like principal component analysis—offer partial relief but may obscure fine-scale features critical to physical interpretation [8, 9]. Interpretability strategies, while increasingly sophisticated, remain constrained in scalability and explanatory depth. Post-hoc attribution methods, including SHAP-based analyses applied to CNN predictions, provide localized feature importance but do not establish causal relationships or mechanistic validity [10, 11]. In contrast, inherently interpretable models that embed symmetry constraints or physics-informed inductive biases align more closely with materials science principles, yet often incur increased computational complexity and reduced flexibility in real-time or adaptive settings [12, 13, 30].

Empirical studies in additive manufacturing highlight both the promise and limitations of interpretable representation learning. While such models can uncover meaningful process–structure linkages, their reliability is frequently undermined by biased training data, particularly the underrepresentation of rare but critical defect modes [14, 15]. Multi-modal integration—combining microstructural imagery with spectroscopic, mechanical, and processing data—offers a pathway toward more robust and context-aware representations. Still, persistent issues in data standardization and interoperability impede widespread adoption [16, 17, 28].

Beyond technical considerations, ethical and governance implications are increasingly salient. Automated microstructure analysis is beginning to influence materials qualification and certification workflows, raising concerns about transparency, accountability, and the risk of erroneous decisions in safety-critical applications [18, 19]. Comparative analyses across materials classes further reveal structural inequities in data availability: metals benefit from decades of accumulated datasets, while ceramics and composites lag due to their intrinsic heterogeneity and experimental complexity, necessitating specialized, often bespoke embedding strategies [20, 21, 27].

Looking forward, the field would benefit from a principled shift toward hybrid frameworks that integrate machine learning representations with physics-based simulations and causal reasoning. Such approaches hold promise for improving both predictive accuracy and epistemic transparency, supporting representation learning systems that not only perform well but also produce scientifically interpretable and ethically defensible insights [22, 23, 30].

Conclusion

In summary, representation learning has ushered in a new era for materials microstructure analysis, with conceptual advances enabling unprecedented efficiency in property prediction and inverse design. However, interpretability challenges continue to impede widespread adoption, demanding innovative solutions that bridge computational prowess with physical intuition. By addressing data scarcity through generative techniques and fostering interdisciplinary collaborations, the field can advance toward trustworthy AI-driven materials science. Ultimately, overcoming these hurdles will not only refine current methodologies but also pave the way for transformative applications in sustainable materials development.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 25 Aug 2025 Revised: 30 Aug 2025 Accepted: 11 Nov 2025
Published online: 18 January 2026

References

De Breuck P-P, Evans ML, Hautier G. Machine learning in materials research: Developments over the last decade and challenges for the future. Curr Opin Solid State Mater Sci. 2024;28:101139.
https://doi.org/10.1016/j.cossms.2024.101139.

Damewood JR, Karaguesian J, Lunger JR, Tan AR, Xie M, Peng J, et al. Representations of materials for machine learning. Annu Rev Mater Res. 2023;53:399-426.
https://doi.org/10.1146/annurev-matsci-080921-085947.

Che L, He Z, Zheng K, Si T, Ge M, Cheng H, et al. Deep learning in alloy material microstructures: Application and prospects. Mater Today Commun. 2023;37:107531.
https://doi.org/10.1016/j.mtcomm.2023.107531.

Goossens N, Gong Y, Zhang C, Liu B, Luan X. A deep learning approach for complex microstructure inference. Nat Commun. 2021;12:6105.
https://doi.org/10.1038/s41467-021-26565-5.

Pei Z, Rozman KA, Doğan ÖN, Wen Y, Gao N, Holm EA, et al. Machine-learning microstructure for inverse material design. Adv Sci. 2021;8:2101207.
https://doi.org/10.1002/advs.202101207.

Wang X, Su H, Li N, Chen Y, Yang Y, Meng H. A deep learning labeling method for material microstructure image segmentation. Processes. 2023;11:3272.
https://doi.org/10.3390/pr11123272.

Zhong X, Gallagher B, Liu S, Kailkhura B, Hiszpanski A, Han TYJ. Explainable machine learning in materials science. npj Comput Mater. 2022;8:204.
https://doi.org/10.1038/s41524-022-00884-7.

Li X, Liu W, Khonsari MM, Xie L, Chen J, Sun C. Machine learning-based identification of interpretable process-structure linkages in metal additive manufacturing. Addit Manuf. 2023;71:103589.
https://doi.org/10.1016/j.addma.2023.103589.

Lu Z, Li Y, Zhang J, Wu J, Liu Y. Interpretable machine-learning strategy for soft-magnetic property and thermal stability in Fe-based metallic glasses. npj Comput Mater. 2020;6:187.
https://doi.org/10.1038/s41524-020-00460-7.

Kondo R, Yamashita T, Uesugi M, Aratani N. Interpretable deep learning approach for extracting nanoparticle features from transmission electron microscopy images. Sci Rep. 2022;12:4274.
https://doi.org/10.1038/s41598-022-06597-5.

Chowdhury A, Kautz E, Yener B, Lewis D. Image driven machine learning methods for microstructure recognition. Comput Mater Sci. 2020;171:109283.
https://doi.org/10.1016/j.commatsci.2019.109283.

DeCost BL, Francis T, Holm EA. High-throughput 3D segmentation of large-scale heterogeneous microstructure data using convolutional neural networks. Microsc Microanal. 2021;27:1-12.
https://doi.org/10.1017/S1431927620024695.

Stan T, Thompson ZT, Voorhees PW. Optimizing convolutional neural networks to perform semantic segmentation on digitized histological images of microstructures. J Intell Manuf. 2020;31:595-610.
https://doi.org/10.1007/s10845-018-1456-4.

Furat O, Petrich L, Finegan DP, Schmidt V. Fusing feature engineering and deep learning: A case study for volumetric segmentation of high-resolution porosimetry images. J Mater Sci. 2022;57:1402-27.
https://doi.org/10.1007/s10853-021-06710-7.

Ling J, Hutchinson B, Lumley R, Jimenez-Melero E. Microstructure classification of steels using deep convolutional neural networks: Addressing data variability and explainability. Mater Charact. 2021;177:111173.
https://doi.org/10.1016/j.matchar.2021.111173.

DeCost BL, Lei B, Francis T, Holm EA. High-throughput feature extraction for measuring attributes of deformed microstructure images. Comput Mater Sci. 2020;172:109308.
https://doi.org/10.1016/j.commatsci.2019.109308.

Ma B, Ban X, Huang H, Liu Y, Yang C. Deep learning-based image classification for microstructure characterization of complex metal alloys. Mater Des. 2020;185:108199.
https://doi.org/10.1016/j.matdes.2019.108199.

Chen C, Zuo Y, Ye W, Li X, Ong SP. Learning properties of ordered and disordered materials from multi-fidelity data. Nat Comput Sci. 2021;1:46-53.
https://doi.org/10.1038/s43588-020-00002-x.

Jacobs R, Liu T, Wang H, Markl M, Dumpala S, Morgan D. Microstructural feature extraction by deep learning for oligocrystalline materials during tensile testing. J Mater Res. 2021;36:658-70.
https://doi.org/10.1557/s43578-020-00027-3.

Chen Z, Zhang J, Arjovsky M, Bottou L. Symmetries and invariances in representation learning for materials science. J Chem Phys. 2022;156:124503.
https://doi.org/10.1063/5.0083418.

Ovadia O, Kahana A, Ziskind D, Yaari E, Sternlieb T, Bucher M, et al. Interpretable deep learning model for microstructural defect detection in TSS 2205 duplex stainless steel welds. J Mater Sci. 2024;59:3832-50.
https://doi.org/10.1007/s10853-024-09425-6.

Venugopal A, Kumar S, Hamedani HA, Agarwal A. Deep learning for microstructure segmentation and analysis in additive manufactured Ti-6Al-4V. Addit Manuf. 2022;49:102479.
https://doi.org/10.1016/j.addma.2021.102479.

Shen C, Wang C, Huang X, Roberts N, Zhang Z. Deep learning assisted phase retrieval for X-ray ptychography of microstructures. Opt Express. 2020;28:17376-91.
https://doi.org/10.1364/OE.393221.

Chan H, Cherukara MJ, Narayanan B, Loeffler TD, Sankaranarayanan SKRS, Cherukara MJ. Machine learning enabled autonomous microstructural characterization in 3D samples. npj Comput Mater. 2022;8:55.
https://doi.org/10.1038/s41524-022-00736-4.

Damewood J, Ward L, Blaiszik B. Deep language models for interpretative and predictive materials science. APL Mach Learn. 2023;1:010901.
https://doi.org/10.1063/5.0133585.

Hsu Y-C, Yu C-H, Buehler MJ. What can machine learning help with microstructure-informed materials modeling and design? MRS Bull. 2024;49:1036-46.
https://doi.org/10.1557/s43577-024-00797-4.

Tang M, Zou Y. Interpretable machine learning approach for predicting process-structure-property relationships in metal additive manufacturing. Addit Manuf. 2024;87:104263.
https://doi.org/10.1016/j.addma.2024.104263.

Hasson M, Andersen UL, Winther O, Alstrøm BL, Castelli IE, Hummel F, et al. Compact representations of microstructure images using triplet networks. npj Comput Mater. 2020;6:157.
https://doi.org/10.1038/s41524-020-00423-2.

Zheng X, Watanabe I, Paik J, Li J, Guo X, Naito M. Text-to-microstructure generation using generative deep learning. Small. 2024;20:2402685.
https://doi.org/10.1002/smll.202402685.

Wan J, Wang H, Ding F, Zhang S, Xu B, Wang D, et al. Advancing materials science through next-generation machine learning. Mater Today. 2024;74:123-39.
https://doi.org/10.1016/j.mattod.2024.02.002.