B Wang *et al*. *J. Mater. Inf.* **2026**;*6*:[Accepted]
DOI: 10.20517/jmi.2025.85

Journal of materials Informatics

Review

Open Access

Multi-fidelity learning in materials informatics: methodologies, applications, and outlook

Bo Wang[1], Yangyang Xu[2,*], Yumei Zhou[1], Dezhen Xue[1]

[1]State Key Laboratory for Mechanical Behavior of Materials, Xi'an Jiaotong University, Xi'an 710049, Shaanxi, China.

[*]**Correspondence to:** Dr. Yangyang Xu, MOE Key Laboratory for Nonequilibrium Synthesis and Modulation of Condensed Matter, School of Physics, Xi'an Jiaotong University, Xi'an 710049, Shaanxi, China.E-mail: xuyangyang@xjtu.edu.cn; Prof. Yumei Zhou, Prof. Dezhen Xue, State Key Laboratory for Mechanical Behavior of Materials, Xi'an Jiaotong University, Xi'an 710049, Shaanxi, China. E-mail: zhouyumei@xjtu.edu.cn; xuedezhen@xjtu.edu.cn

How to cite this article: Wang B, Xu Y, Zhou Y, Xue D. Multi-fidelity learning in materials informatics: methodologies, applications, and outlook. *J Mater Inf* 2026;6:[Accept]. http://dx.doi.org/10.20517/jmi.2025.85

Received: 8 October 2025 **Revised:** 24 December 2025 **Accepted:** 14 January 2026

Abstract

Data-driven methods are transforming materials design by accelerating the discovery of new compounds and optimization of existing systems. Yet the progress of such approaches is often constrained by the scarcity of high-fidelity data from experiments and advanced simulations. Multi-fidelity learning has emerged as a powerful strategy to address this challenge by integrating information from diverse data sources that vary in accuracy and cost. In this review, we provide a systematic overview of the major methodologies for multi-fidelity learning, including statistical and parametric models, machine learning models with fidelity features, correction-based models such as co-kriging, deep learning frameworks, and active learning frameworks. We discuss the strengths, limitations, and typical applications of each method in materials science, with illustrative examples spanning electronic structure modeling, alloy design, and interatomic potential development. Cross-cutting issues are also examined, including the bias-variance trade-off, data requirements for nested versus non-nested designs, and computational scalability. Finally, we highlight outstanding challenges and outline emerging opportunities, such as physics-informed and generative multi-fidelity models, standardized datasets, and integration with autonomous laboratories. Together, these perspectives define a roadmap for advancing multi-fidelity learning as a core enabler of next-generation materials discovery.

Keywords: Multi-fidelity learning, materials informatics, surrogate modeling, active learning, data-driven materials design

ABBREVIATIONS

MF	Multi-fidelity
SF	Single-fidelity
HF	High-fidelity
LF	Low-fidelity
DFT	Density functional theory
$F(\cdot)$	Mapping model/function
ε	Noise
σ^2	Variance
θ	Parameters of a parametric model
$\mathcal{N}(\cdot)$	Normal distribution
GP	Gaussian process
e_f	One-hot encoding vector of the fidelity level
α	Reduced absorption coefficient
TL	Transfer Learning
FE	Fidelity-embedding
GNN	Graph neural network
DNN	Deep neural network
GFN	Generative flow network
$C(\cdot)$	Correction operator
AR1	First-order autoregressive
ρ	Scale factor
$\delta(\cdot)$	Discrepancy Gaussian process
RMSE	Root-mean-square error
MAE	Mean absolute error
γ	Interface energy
LER	Lower-error region
EXP	Experimental
BO	Bayesian optimization
$\alpha(\cdot)$	Acquisition function
$c(\cdot)$	Cost function
COFs	Covalent organic frameworks
GPU	Graphic Processing Unit
PBE	Perdew-Burke-Ernzerhof
HSE	Heyd-Scuseria-Ernzerhof
$O(\cdot)$	The upper bound of the computational complexity of an algorithm
B	Batch size
P	Parameter count
E	Epoch count

1. INTRODUCTION

Data-driven methodologies are increasingly being utilized to accelerate the design and discovery of new materials.[1–11] At the core of these approaches lies data, which provides the foundation for training predictive models and guiding design strategies.[12–17] Yet the accelerating pace of materials discovery is constrained by a persistent data bottleneck, where high-fidelity data are scarce while low-fidelity data are abundant.[18–23] To formalize this disparity in data quality, we introduce the notion of data fidelity. Here, "fidelity" refers to the quality of a data source or model, encompassing its systematic bias relative to the desired high-truth target, its noise level or predictive variance, and the completeness/integrity of the recorded variables. High-fidelity sources, such as density functional theory (DFT) calculations at the hybrid functional level[9,24–27] or experimental measurements of structural, electronic, and mechanical properties[28–30], are regarded as the gold standard. Although these methods provide accurate and reliable information, they are costly in terms of computational resources, experimental effort, or wall-clock time, which limits their scalability across the vast chemical and structural design space.[31] By comparison, low-fidelity data, including semi-empirical simulations, classical force fields, and simplified experiments, are much more accessible, but inevitably less accurate and possibly biased.[28–30,32,33] Bridging this divide between scarce, high-quality data and plentiful, low-quality data remains a central challenge for materials informatics. This challenge has motivated the development of strategies capable of jointly leveraging information across fidelity levels.

Among these strategies, multi-fidelity learning has emerged as a powerful approach to overcome such a bottleneck. In essence, multi-fidelity learning refers to a class of machine learning approaches that integrate information from datasets of varying accuracy and cost, such as combining low-fidelity simulations with high-fidelity experimental measurements.[34–36] The central idea is to exploit the complementary strengths of different data sources: low-fidelity data provide broad coverage of the design space and capture general trends in materials performance, while high-fidelity data anchor the models with reliable accuracy.[28,37,38] By learning correlations and systematic discrepancies between fidelity levels, multi-fidelity approaches can effectively transfer knowledge across datasets, yielding

predictions that are both accurate and data-efficient.[34,39–45] This paradigm is particularly promising for materials design, where the cost of exhaustive high-fidelity characterization is prohibitive, yet the demand for predictive accuracy is high.[28–30,32,33] Key methodological ideas include statistical and parametric models, machine learning models with fidelity features, correction-based models such as co-kriging, deep learning frameworks, and active learning frameworks, each offering distinct trade-offs in terms of accuracy, scalability, and interpretability.[46–50] Together, these strategies enable the construction of surrogate models with higher accuracy and lower costs, guiding to acceleration of materials discovery.[29,51–53]

Multi-fidelity learning has broad practical relevance across materials research, spanning high-throughput screening pipelines[42,54], interatomic potential predictions[38–40], and data-driven design of new materials[37,55]. In steel design, it couples fast thermo-micromechanical surrogates with strategically chosen high-fidelity finite element evaluations to optimize strain-hardening performance with both efficiency and accuracy.[53] In halide perovskites, multi-fidelity machine learning integrates semi-local and hybrid-functional density functional theory data with experimental measurements to predict decomposition energies, guiding the genetic algorithm-driven discovery of stable photovoltaic alloys.[56] In polymers, a multi-fidelity co-training neural network fuses all-atom molecular dynamics simulation data (low-fidelity) with scarce experimental measurements (high-fidelity) to accurately predict phthalonitrile melting points, enabling low-cost screening of low-melting candidates for improved processability.[57] In geomaterials, an active-learning multi-fidelity residual Gaussian process framework was trained on abundant low-fidelity data collected across multiple sites, with predictive uncertainty used as an acquisition function to guide high-fidelity sampling. This approach demonstrated robust extrapolation in sparsely sampled regions and significantly reduced high-fidelity data requirements compared to single-fidelity models.[58] In additively manufactured alloys, multi-fidelity physics-informed frameworks have been developed that combine physics-guided low-fidelity data generation with transfer learning on limited high-fidelity measurements, achieving improved accuracy and physical consistency in fatigue-life prediction.[59] More broadly, as labs juggle mixed-cost measurements and computational proxies, best-practice guidelines for multi-fidelity Bayesian optimization codify when and how to couple low-fidelity (LF) and high-fidelity (HF) sources to maximize discovery per unit budget and avoid failure modes in practical materials discovery campaigns.[60]

Building on these diverse applications, the purpose of this review is to provide a comprehensive overview of methodologies for multi-fidelity learning in materials design. We first classify and analyze the principal strategies, illustrating each with representative case studies. We then discuss cross-cutting themes, including the handling of bias-variance trade-offs, data co-location requirements and computational scaling. Finally, we highlight the challenges and opportunities that will shape the future of this field, from the integration with autonomous discovery systems to the development of physics-informed, generative frameworks. By consolidating methodological insights and materials applications, this review aims to guide both machine learning researchers and materials scientists toward more effective multi-fidelity approaches. The principal methodological strategies are summarized in Figure 1, which also serves as a roadmap for the sections that follow.

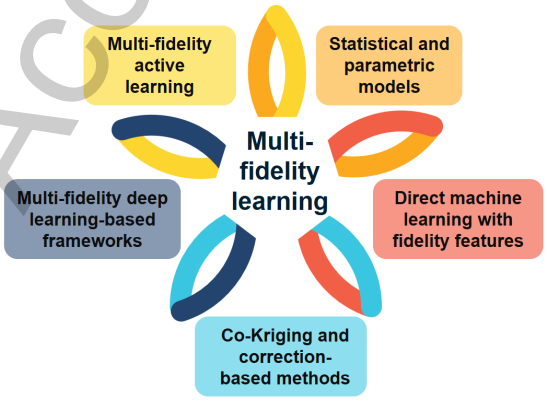

Figure 1. Taxonomy of multi-fidelity learning methodologies in materials design. Five principal strategies are highlighted: (i) Statistical and parametric models, (ii) Direct machine learning with fidelity features, (iii) Co-Kriging and correction-based methods, (iv) Multi-fidelity deep learning-based frameworks, and (v) Multi-fidelity active learning and adaptive sampling. These categories provide the structural roadmap for section 2, where each approach is introduced, illustrated with representative case studies, and evaluated in terms of strengths, limitations, and domains of applicability.

2. STRATEGIES AND APPLICATIONS FOR MULTI-FIDELITY LEARNING IN MATERIALS DESIGN

We classify multi-fidelity learning methodologies into five broad categories (Figure 1): statistical and parametric models, direct machine learning with fidelity features, co-kriging and correction-based methods, multi-fidelity deep learning-based frameworks, and multi-fidelity active learning and adaptive sampling. Each of these strategies embodies a distinct way of integrating information across fidelity levels, and in the subsections below we describe their core ideas, illustrate their use with representative examples, and evaluate their strengths, weaknesses, and appropriate domains of application.

2.1. Statistical and Parametric Models

A classical route for incorporating multi-fidelity (MF) information is to assume an explicit functional form grounded in physical or statistical reasoning. Data at different fidelity levels are represented by distinct noise variances, and these variances are used to fit the functional form. HF data are accurate (small noise variance) but scarce, while LF data are abundant but noisy (large noise variance). Different data are therefore weighted according to their noise levels during the fitting process, typically achieved by maximum likelihood under fidelity-dependent noise.[61,62]

A common strategy is to posit a parametric mapping $y = F(x;\theta)$ and estimate the parameters θ by maximum likelihood under fidelity-dependent noise. Concretely, one assumes

$$y_i = F(x_i;\theta) + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2_{f(i)}), \tag{1}$$

where $\sigma^2_{f(i)}$ denotes the noise variance associated with the fidelity level $f(i)$ of sample i. The corresponding log-likelihood is

$$\log p(y \mid \theta) = -\tfrac{1}{2} \sum_i \left[(y_i - F(x_i;\theta))^2 / \sigma^2_{f(i)} + \log \sigma^2_{f(i)} \right] + C, \tag{2}$$

where the fidelity-dependent noise variance directly enters the likelihood. Maximizing this fidelity-weighted Gaussian log-likelihood with respect to θ is equivalent to solving a weighted least squares problem,

$$\hat{\theta} = \arg\min_\theta \sum_i w_i \left(y_i - F(x_i;\theta)\right)^2, \quad w_i = 1/\sigma^2_{f(i)}. \tag{3}$$

This formulation makes the role of fidelity explicit: higher-variance (lower-fidelity) points are down-weighted, while HF points with smaller variance exert stronger influence on the parameter estimates.[61,62] Thus, multi-fidelity fusion is naturally achieved by adjusting the statistical weight of each fidelity level in the estimation process.

The same variance-weighting principle also appears in Gaussian process (GP) and Kriging frameworks, where predictive means and variances from multiple fidelities are combined *via* inverse-variance weighting, and hyperparameters are fitted by maximum likelihood.[46,63] In these approaches, fidelity is again treated statistically, with higher-fidelity data associated with smaller predictive variance, thereby exerting stronger influence on the posterior model. This parallel highlights a unifying view: fidelity enters the modeling process through its impact on the variance structure.

The parametric and weighted least squares treatment is attractive because it preserves interpretability and transparency. The functional form F can be chosen to encode physical insights (*e.g.*, linear or low-order polynomial response surfaces), while multi-fidelity fusion is handled cleanly through the likelihood-based weighting scheme. It is also computationally efficient and, with projection-enabled linear regression, can accommodate non-overlapping MF datasets where HF and LF data are observed at different inputs.[64] When domain knowledge suggests a credible approximation form for F, parametric MF models are parsimonious, fast, and robust, with a principled mechanism to privilege HF data while still leveraging LF samples to capture global trends.

In practice, parametric MF models are most effective in applications where domain expertise provides guidance for a plausible functional form. For example, in hypersonic vehicle surface pressure modeling, MF linear regression improved predictive accuracy by up to 12% compared to single-fidelity regression, even with as few as 3-10 HF samples.[64] In materials science, heteroscedastic Gaussian process regression (HGPR) has been applied to porous two-phase microstructures generated by finite element simulations. In this setting, input-dependent noise captured inherently noisy regimes, linking geometric descriptors (ellipse axes, porosity, spacing) to effective stress in representative volume elements (60%-40% split for training and validation sets).[65] This study illustrates that heteroscedasticity in materials datasets, traditionally regarded as a nuisance, can instead be leveraged as an informative signal to guide design and optimization.

These advantages, however, come with trade-offs. Parametric MF models often suffer from limited representational capacity: linear or polynomial structures may be inadequate for highly nonlinear, high-dimensional behaviors. They can also degrade when LF data introduce systematic bias or when fidelity relationships deviate significantly from simple mappings. Furthermore, mis-specified fidelity-dependent variances can undermine the intended HF emphasis. Such issues can be mitigated by re-estimating variance terms *via* maximum likelihood or by extending the framework to include fidelity-specific offsets and scaling factors.[46]

From a practical standpoint, for parametric multi-fidelity models, a key design choice is the selection of the functional form. A practical strategy is to start from the simplest model justified by physical insight and progressively increase complexity as needed. Candidate models can be evaluated using cross-validation, and the model that exhibits the best generalization performance is selected.

2.2. Direct Machine Learning with Fidelity Features

A second strategy for multi-fidelity learning is to directly encode fidelity information as an input feature to a unified model. Rather than building separate models at each fidelity or linking them hierarchically, all data are pooled and a single model is trained with an additional "fidelity feature". For each sample (x, y) with fidelity $f \in \{1 \text{ (highest)}, \ldots, K \text{ (lowest)}\}$, a K-dimensional one-hot vector e_f is concatenated with the material descriptors x_i to form the augmented input $\tilde{x}_i = [\, x_i \,\|\, e_{f(i)} \,]$. At inference, setting $e_f = [1, 0, \ldots, 0]$ yields predictions at the highest fidelity, while other encodings correspond to lower fidelities (Figure 2a). This approach is flexible because it does not require one-to-one correspondence of data across fidelities or a fixed model hierarchy.[39,40]

Formally, each observation can be written as

$$y_i = F\left(x_i, e_{f(i)};\, \theta\right) + \varepsilon_i, \tag{4}$$

where x_i denotes the material descriptors, $e_{f(i)}$ is the one-hot encoding of the fidelity level, and ε_i represents observation noise. Unlike the parametric formulations discussed in section 2.1, the functional form F is not prescribed a priori. Instead, F is estimated directly from data using machine learning models such as neural networks, support vector machines, or Gaussian processes. The fidelity encoding therefore guides the model to integrate information across fidelities within a unified representation, rather than weighting data points through fidelity-dependent noise variances.

Direct multi-fidelity learning with fidelity features has been successfully applied in several areas of materials modeling. Ko and Ong integrated fidelity information into the "global state" of a graph neural network (M3GNet) to build a multi-fidelity interatomic potential. Training on combined datasets of LF and HF simulations (*e.g.*, silicon and water), the multi-fidelity M3GNet achieved the accuracy of a high-fidelity-only model using only ~10% of the HF data, representing an order-of-magnitude saving in computational cost for generating HF data (Figure 2b).[39] Kim *et al.* developed SevenNet-MF, an equivariant GNN trained on mixed generalized gradient approximation (GGA, LF) and meta-generalized gradient approximation (meta-GGA, HF) calculations using one-hot fidelity features. With just 10% HF data relative to the LF set, SevenNet-MF achieved $< 10\%$ error in Li-ion conductivity and $R^2 \approx 0.98$ for mixing energies on validation sets (90%-10% split for training and validation sets), outperforming fine-tuning and Δ-learning baselines.[40] Together, these studies demonstrate that fidelity features enable models to expand HF datasets effectively by leveraging LF trends.

The approach has also been extended beyond interatomic potentials. Ibrahim and Ataca developed a GNN for

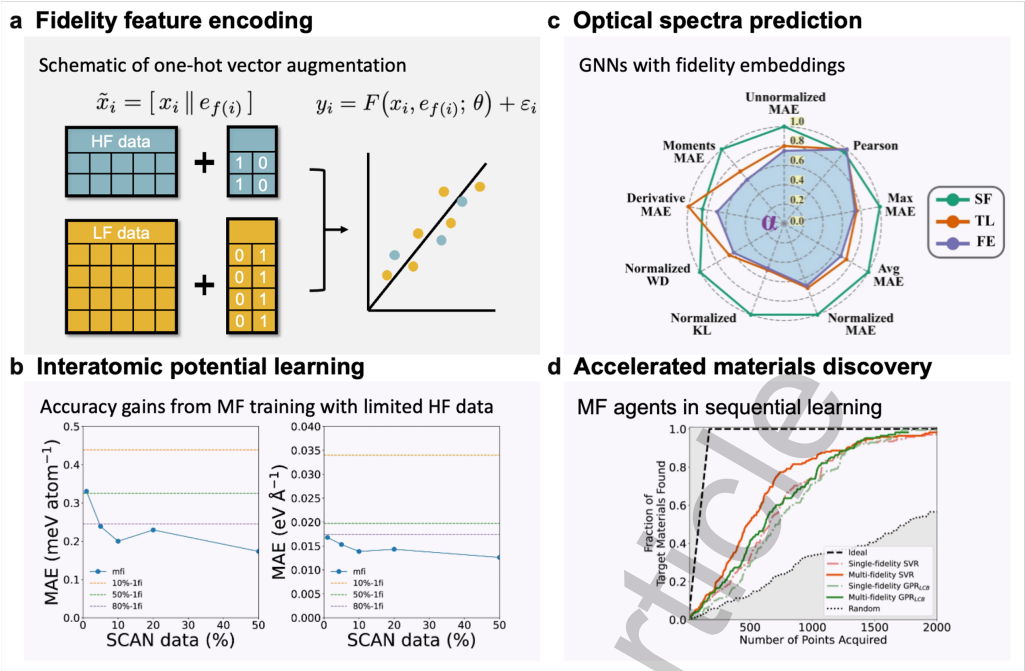

Figure 2. Examples of direct multi-fidelity learning with fidelity features. a. Schematic of the one-hot encoding fidelity approach: each sample is augmented with a fidelity indicator vector e_f, allowing a single unified model to learn across low- and high-fidelity data. b. Interatomic potential for water: test MAEs of a multi-fidelity model (mfi) trained with varying fractions of HF (SCAN) data (left: energy, right: force). Dashed lines indicate the MAEs of single-fidelity (1fi) baselines using the same fractions of HF data, highlighting the efficiency of MF learning.[39] c. Frequency-dependent optical spectra (reduced absorption coefficient α): radar charts show normalized median errors on the HF test set for three GNNs-single-fidelity (SF), transfer learning (TL), and fidelity-embedding (FE), demonstrating the gain from fidelity-aware embeddings.[66] d. Simulated inorganic materials discovery under sequential learning: the x-axis shows experimental acquisition budget, and the y-axis shows the fraction of ideal materials (with visible-spectrum band gaps) discovered. MF agents with fidelity indicators accelerate discovery relative to SF agents.[47]

predicting frequency-dependent optical spectra by embedding fidelity information as a learnable vector. Joint training on LF and HF spectra (34327 LF data and 14560 HF data) significantly reduced errors on test sets (80%-5%-15% split for training-validation-test sets) in dielectric response predictions compared to HF-only training, and even outperformed transfer-learning baselines (Figure 2c).[66] Similarly, in the inverse design of halide perovskites, multi-fidelity surrogates with fidelity indicators guided genetic algorithms to yield experimentally stable compositions by combining generalized gradient approximation Perdew-Burke-Ernzerhof (GGA-PBE, LF, 570 data), Heyd-Scuseria-Ernzerhof hybrid functional (HSE06, HF, 347 data), and experimental measurements (alternative HF, 97 data).[56] In high-throughput screening, multi-output Gaussian processes incorporated fidelity information by concatenating fidelity tags with material descriptors, enabling the fusion of experimental and computational measurements within a single workflow and improving both efficiency and uncertainty quantification.[54] Mora *et al.* encode data sources as one-hot vectors and feed them into a Bayesian neural network that maps fidelity indicators onto a low-dimensional continuous fidelity manifold, through which uncertainty is propagated to the output layer. This formulation imposes no restriction on the number of fidelities and avoids a predefined hierarchy. The network outputs both predictive means and variances, enabling quantification of prediction uncertainty as well as fidelity-related uncertainty. Applied to multi-fidelity damage simulations of porous metallic components and to binding-energy prediction in hybrid organic-inorganic perovskites, the learned manifold disentangles heterogeneous data sources and yields well-calibrated uncertainty estimates.[67] Beyond deterministic surrogates, Foumani *et al.* introduced a cost-aware Bayesian optimization framework with a multi-fidelity emulator that one-hot-encodes fidelity as a categorical variable, learns across fidelities, and automatically detects when LF sources are too biased to be useful.[68] Palizhati *et al.* further showed that MF agents encoding fidelity features (DFT as LF, experiments as HF) accelerated early discovery by 20-60% compared to single-fidelity baselines (Figure 2d).[47]

Overall, direct multi-fidelity machine learning provides a powerful and flexible strategy for exploiting heterogeneous datasets under limited high-fidelity budgets. They have already enabled the cost-effective development of interatomic potentials approaching coupled-cluster accuracy, predictors of complex optical properties, and efficient high-

throughput surrogates. It can reduce HF requirements by up to an order of magnitude while matching or even surpassing the accuracy of HF-only models.[39,40] Compared to correction-based approaches introduced in section 2.3, direct methods do not require overlapping data across fidelities and instead integrate complementary information within a unified representation.[39,66]

Despite these strengths, several limitations remain. Encoding fidelity as a one-hot tag does not explicitly capture cross-fidelity correlations (LF→HF mapping), which reduces interpretability relative to autoregressive or correction-based models. The approach also does not inherently account for heteroscedasticity or calibration: noisy LF data can dilute HF signal and compromise uncertainty quantification. In addition, severe class imbalance, where only a small number of HF samples are available, may lead to overfitting to LF trends or spurious correlations between fidelity and input features. Thus, while direct multi-fidelity learning with fidelity features is highly flexible and data-efficient, it is most effective when LF data are reasonably correlated with HF targets and when a sufficient number of HF points are available to anchor the learning process.

From a practical standpoint, performance is also sensitive to model-specific hyperparameter choices. Hyperparameter tuning in direct multi-fidelity learning depends on the underlying machine learning model. For neural networks, the primary hyperparameters include the network architecture, learning rate, batch size, and regularization strength. For Gaussian processes, the dominant choices are the kernel (covariance function), its associated hyperparameters, and the noise level.

2.3 Co-Kriging and Correction-Based Methods

Correction-based methods, most prominently co-kriging, form one of the most established strategies in multi-fidelity learning.[48] They extend GP regression by modeling correlations between LF and HF data, thereby leveraging the broad coverage of inexpensive LF data to complement the accuracy of HF data. The basic idea, illustrated in Figure 3a, is to learn a statistical correction that maps LF outputs to HF outputs, effectively "lifting" cheap predictions into the accuracy regime of expensive ones. Formally, one seeks a correction operator C such that

$$f_{\mathrm{HF}}(x) = C(f_{\mathrm{LF}}(x)), \tag{5}$$

where C is learned from data. In co-kriging, this mapping is embedded within a GP framework so that LF and HF functions are modeled as correlated random processes, and the learned correction enables the surrogate to follow HF trends while exploiting LF coverage. The GP framework naturally provides predictive uncertainty across the input space.[41,48,69–71] In GP regression, the posterior prediction at a new input x^* is Gaussian with a closed-form mean and variance, which naturally provides calibrated predictive uncertainty. Here, uncertainty reflects both data noise (aleatoric uncertainty) and limited knowledge of the underlying function (epistemic uncertainty). In materials informatics, such uncertainty-aware predictions are essential for risk-sensitive decision-making and adaptive sampling, where uncertainty estimates guide the selection of new simulations or experiments. GP-based co-kriging therefore enables principled quantification and propagation of uncertainty across fidelity levels.

A particularly common instantiation of this correction framework is the first-order autoregressive (AR1) formulation introduced by Kennedy-O'Hagan and later popularized by Forrester.[71] In this approach, the HF function is expressed as[71]

$$f_{\mathrm{HF}}(x) = \rho f_{\mathrm{LF}}(x) + \delta(x), \tag{6}$$

where ρ is a scale factor and $\delta(x)$ is an independent GP that captures the systematic LF→HF discrepancy. This formulation can be viewed as a specific parametric choice for the correction operator $C(\cdot)$, with LF predictions rescaled and then adjusted by an additive discrepancy term. The resulting joint GP induces a block covariance structure linking LF and HF observations, thereby enabling interpolation of HF data while rigorously quantifying predictive uncertainty. Despite its popularity, the AR1 co-kriging formulation relies on several implicit assumptions. First, the mapping between low- and high-fidelity responses is assumed to be approximately linear, such that a single scaling factor captures the dominant cross-fidelity relationship. When the LF→HF relationship is strongly nonlinear or state dependent, this linear autoregressive assumption may break down, motivating the use of more

flexible nonlinear correction operators. Second, both the low-fidelity process $f_{\mathrm{LF}}(x)$ and the discrepancy term $\delta(x)$ are typically modeled as stationary Gaussian processes, meaning that their mean functions are constant and their covariance structures depend only on the separation between inputs. This co-stationarity assumption can be restrictive in systems where cross-fidelity correlations vary significantly across the input space.[41,48,70,72–75] When these linearity or stationarity assumptions are violated, alternative multi-fidelity strategies may be more appropriate. Examples include nonlinear correction-based formulations, composite or deep Gaussian process models, and direct machine learning approaches with fidelity features that do not impose explicit parametric structure on LF-HF mappings. In practice, diagnosing the degree of linearity and stationarity in cross-fidelity relationships provides useful guidance for choosing between classical co-kriging, nonlinear GP variants, and deep learning-based frameworks.

There are also alternative forms of the general correction operator $C(\cdot)$. Recursive co-kriging extends the autoregressive framework to more than two fidelity levels by chaining GPs across coarse, intermediate, and fine datasets.[48,70,72] Composite and deep GP models enrich the operator $C(\cdot)$ to capture nonlinear mappings between fidelities that cannot be represented by a simple affine relation.[41,73–75] Other formulations, such as those based on radial-basis functions or multi-output GP surrogates, have likewise been introduced, further broadening the family of correction-based strategies.[76]

Co-kriging has already enabled significant advances in materials design.[32,51,77–81] In crystal plasticity, a co-kriging model combining 100 LF orientation-distribution-function (ODF) data with 10 HF crystal plasticity finite element simulations (CPFEM) data was used to calibrate slip and twinning parameters of Ti-7Al against tensile data, markedly improving stress-strain predictions relative to the ODF-calibrated model (Figure 3b).[77] In precipitation modeling of Mg alloys, co-kriging identified low-error regions more efficiently than HF-only surrogates (Figure 3c).[28] In electronic-structure applications, it has been used to correct bandgaps computed with semi-local DFT using the Perdew-Burke-Ernzerhof (PBE) functional to the accuracy of the Heyd-Scuseria-Ernzerhof (HSE) hybrid functional, enabling HF-quality predictions for thousands of materials at a fraction of the computational cost (Figure 3d).[34,42,52] In molecular benchmarks, autoregressive GPs trained across multiple fidelities, including HF DFT, second-order Møller-Plesset perturbation theory (MP2), and coupled-cluster with singles, doubles, and perturbative triples [CCSD(T)], showed steadily decreasing errors as each fidelity level was incorporated (Figure 3e).[78] For molecular crystal prediction, co-kriging enabled re-ranking of candidate polymorphs with hybrid-DFT accuracy using only a small subset of direct calculations.[33] In alloy design, LF molecular dynamics simulations with a machine-learned potential have been combined with select HF DFT data to efficiently map composition-property relationships.[82] Battery applications have likewise benefited, where LF data from related cells informed extrapolation of long-cycle HF degradation profiles.[37] Finally, in process engineering, analytical melt-pool models have been corrected with finite-element simulations to predict melt-pool dimensions in additive manufacturing, while in fracture mechanics small sets of HF experiments (6 points) have been amplified by co-kriging with inexpensive theoretical predictions (22 points).[83,84] Together, these examples highlight the broad applicability of correction-based methods across length scales, materials classes, and property types.

Beyond specific applications, co-kriging offers several general advantages. By jointly modeling LF and HF data, it can deliver HF-level accuracy at a fraction of the cost, particularly when abundant LF data are available to constrain global trends. Uncertainty quantification is inherent to the GP framework, which is especially valuable for downstream tasks such as active learning and Bayesian optimization. By contrast, most deep learning-based multi-fidelity approaches do not provide calibrated uncertainty estimates unless explicitly augmented with probabilistic outputs, such as Bayesian neural networks or ensemble methods. Moreover, co-kriging is versatile: whenever cross-fidelity correlations exist, it can fuse heterogeneous sources, including multiple levels of simulation, simplified analytical models, and experimental measurements. Compared with multivariate probability distribution-based multi-fidelity models, co-kriging explicitly learns functional input-output relationships across fidelities, which is critical for capturing complex structure-property mappings.

Nevertheless, several practical limitations remain. Reliable calibration of the LF→HF mapping often requires overlapping (nested) data, that is, the same inputs observed at multiple fidelity levels. When this condition cannot be satisfied, more flexible alternatives, such as direct machine learning with fidelity features or probabilistic deep learning models, may be better suited for non-nested or heterogeneous datasets. Beyond data requirements, computational scalability poses a further challenge. Training cost scales cubically with the number of samples, and the joint covariance structure grows with the number of fidelities, making large-scale applications computationally expensive. Interpretability can also be limited, since the discrepancy term $\delta(x)$ is a statistical adjustment that lacks

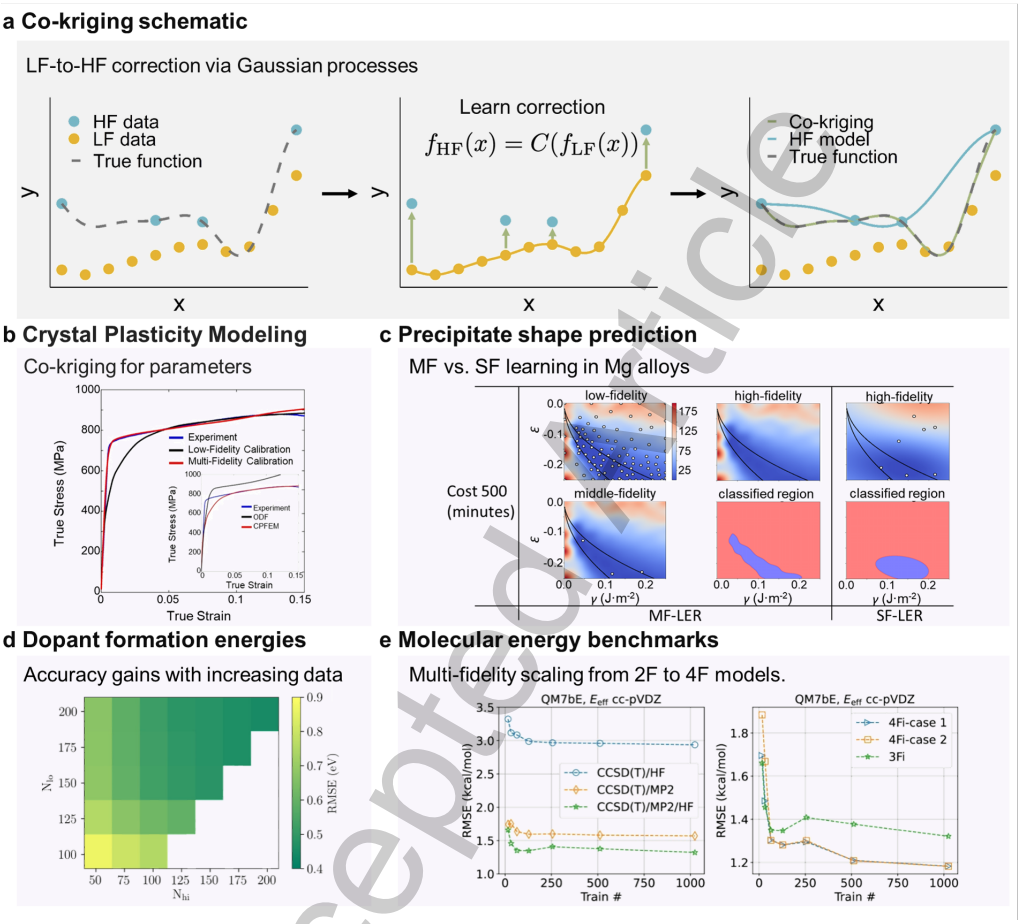

Figure 3. Examples of correction-based multi-fidelity learning using co-kriging. (a) Schematic illustration of the co-kriging approach, where a GP model learns a correction that maps low-fidelity LF to HF responses (adapted from[71]). (b) Crystal plasticity modeling of Ti-7Al alloy: resulting stress-strain responses indicate the MF calculation perfectly reproduce the experimental result. Inset shows HF (CPFEM) and LF (ODF) data[77]. (c) Precipitate shape simulation in Mg alloys: heatmaps show predicted discrepancies between simulated and experimental precipitate aspect ratios. White points mark observed samples, black lines the actual lower-error region (LER). Binary images are predicted LERs from co-kriging (MF) and HF-only (SF) models, demonstrating that additional LF data improves efficiency[28]. The ε denotes the crystal mismatch between the precipitate and matrix phases while γ is interface energy. (d) DFT dopant formation energies in hafnia: prediction accuracy of co-kriging models improves as the number of HF (N_{hi}) and LF (N_{lo}) training points increases[34]. (e) Atomization energy benchmarks: root-mean-square error (RMSE) decreases as more fidelities are combined. Two-fidelity models use CCSD(T)/HF or CCSD(T)/MP2, while three- and four-fidelity models (3Fi, 4Fi) show progressively better accuracy[78]. Overall, these case studies highlight the versatility of correction-based surrogates in bridging simulations and experiments across diverse materials applications.

direct physical meaning. Finally, simple linear autoregressive assumptions may fail when LF→HF relationships are nonlinear or strongly state-dependent, motivating the development of deep or composite GP corrections. These trade-offs are important to consider when deciding between correction-based and alternative multi-fidelity strategies.

From a practical perspective, successful application of co-kriging also depends on careful hyperparameter selection and calibration. The dominant hyperparameters of co-kriging largely mirror those of Gaussian processes, including the choice of kernel (covariance function), its associated hyperparameters, and the noise level. In addition, for linear autoregressive formulations, the scaling factor ρ plays a critical role. The simplest choice is a constant ρ, which can be generalized to an input-dependent $\rho(x)$ (e.g., linear or polynomial in x) to capture more complex corrections. Common kernel choices include exponential, radial basis function (RBF), Matérn family, and periodic kernels. If the data exhibit repeating patterns, periodic kernels are appropriate; if a smooth response is expected, exponential or RBF kernels are often preferred; conversely, for rougher behavior, Matérn 3/2 or 5/2 kernels provide reasonable starting points. Kernel hyperparameters are typically tuned via grid search or gradient-based optimization combined with cross-validation. When data are noisy or heteroscedastic across fidelities, separate noise parameters should be retained and jointly optimized.

2.4 Multi-Fidelity Deep Learning-Based Frameworks

Recent advances in multi-fidelity modeling have increasingly turned to deep learning architectures as flexible tools for linking LF and HF datasets in materials science, as schematically shown in Figure 4a. Unlike co-kriging or other correction-based models, neural networks can capture highly nonlinear and non-smooth relationships, making them particularly attractive for property landscapes where traditional surrogates struggle.[49,85,86]

One important strategy is *transfer learning* between simulation tiers or between simulations and experiments. A common paradigm pre-trains a deep neural network (DNN) on abundant LF data and then fine-tunes selected layers on limited HF labels, sometimes with physics-informed constraints.[87–89] For example, Chakraborty proposed a multi-fidelity physics-informed deep neural network (MF-PIDNN) that encoded approximate physical laws at LF and updated the model with scarce HF data, demonstrating reliable predictions in reliability analysis where experiments were limited but governing physics was known.[88] In materials applications, this approach has been used for mechanical-property inference from small-punch tests, polymer property prediction with heterogeneous data sources, and formation-energy prediction where DFT pre-training accelerates learning on scarce experimental labels (Figure 4b).[35,43,57] Similar paradigms have been applied to fatigue-life prediction with small datasets[90] and even cross-domain transfer from inorganic datasets to metal-organic frameworks, where pre-training on one materials domain accelerates discovery in another.[55] The latter study provides a concrete validation of the generalization capability of multi-fidelity transfer learning, demonstrating that latent representations learned from one class of materials can be reused in chemically distinct domains. Such transfer substantially reduces high-fidelity data requirements when exploring new materials families and highlights the robustness of multi-fidelity representations beyond narrowly defined training distributions.

While transfer learning leverages sequential training, an alternative is to integrate multiple fidelities simultaneously within a *composite neural network*.[49] Islam et al.[91] demonstrated this concept for molecular dynamics, where coarse time-step data served as LF inputs and fine time-step results provided HF corrections, yielding accurate pressure predictions with reduced computational time cost. To assess not only accuracy but also reliability, the authors evaluated model stability under repeated training. For each training-set size, the networks were retrained multiple times using different random selections of high-fidelity data while keeping the low-fidelity dataset fixed (10,000 samples for E^* or 100,000 samples for σ_y). The error bars in Figure 4c report the standard deviation of the resulting mean absolute percentage error (MAPE) across these repeated trainings, thereby quantifying variability in generalization performance. The substantially narrower error bars observed for the multi-fidelity models indicate more stable predictions and higher reliability under changes in the high-fidelity training subset. This example illustrates that, beyond improving mean accuracy, multi-fidelity learning can significantly enhance robustness and reduce sensitivity to data selection.[30] Similar frameworks have been applied to predict residual strength of graded composites and recover stress-strain fields from digital image correlation in welds, all by coupling simulated and experimental data within a unified deep network.[36,92] These studies show how composite neural networks exploit correlations between fidelities to deliver robust surrogates that outperform HF-only training at much lower cost.

A related approach is *correction learning*, where a network is trained directly on the discrepancy between LF and HF predictions. In band-gap benchmarks, this strategy achieved lower errors than both transfer learning and composite networks, since the model concentrated its capacity on systematic LF bias.[93] However, correction learning requires

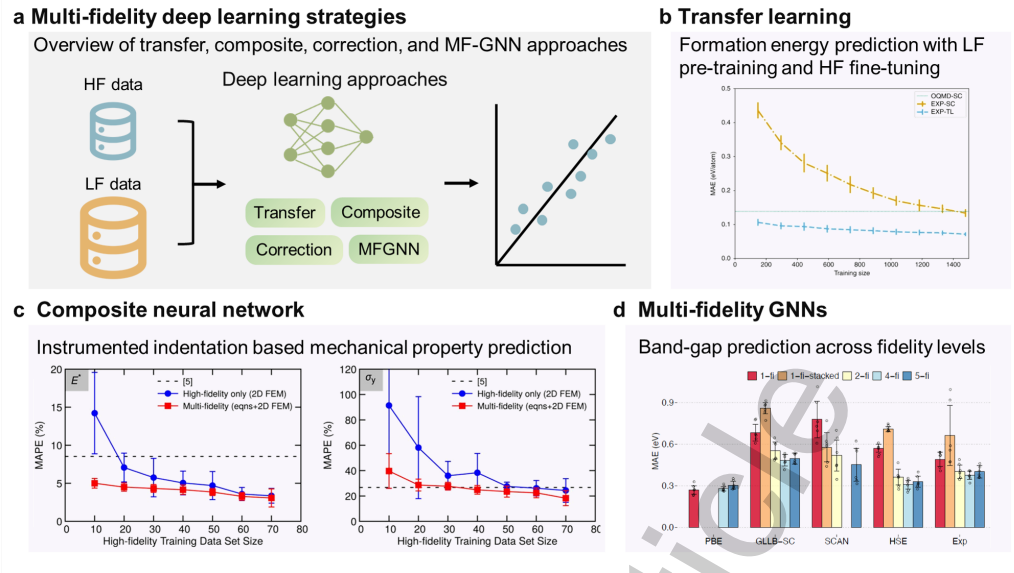

Figure 4. Deep learning-based multi-fidelity frameworks for materials science. (a) Schematic overview of representative strategies: transfer learning, composite neural networks, correction learning, and multi-fidelity graph neural networks (MFGNN). (b) Transfer learning for formation energy prediction: performance comparison of single-fidelity models trained on OQMD (LF, 341k entries) and experimental data (HF) with a transfer-learned model (EXP-TL). Incorporating LF knowledge reduces mean absolute error (MAE) across training sizes; results are obtained by 10-fold cross-validation, and error bars indicate the standard deviation (confidence interval) across folds.[35] (c) Composite neural network for extracting mechanical properties from indentation: Mean average percentage error (MAPE) as a function of training dataset size for HF only model and MF model, in comparison with LF baseline (noted as [5]).[30] Reproduced from[30], © (2020) National Academy of Sciences. Distributed under the Creative Commons Attribution–NonCommercial–NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license; image reused without modification.(d) Multi-fidelity graph neural network for band-gap prediction: comparison of MAE distributions across single-fidelity (1-fi), correction (1-fi-stacked), two-fidelity (2-fi), and multi-fidelity (4-fi, 5-fi) models. LF data comprise PBE band gaps, while HF datasets include GLLB-SC, SCAN, HSE, and experimental values. Increasing fidelity integration systematically lowers prediction error.[29] Reproduced from[29] under the Creative Commons Attribution–ShareAlike (CC BY-SA 4.0) license.Together, these examples highlight the versatility of deep learning architectures in capturing nonlinear correlations and leveraging heterogeneous datasets for improved accuracy in multi-fidelity learning.

nested LF/HF pairs for training and typically demands an LF evaluation for every new query, which can limit applicability when LF costs are significant.

Building on these approaches, the most ambitious efforts extend deep learning into *multi-fidelity graph neural networks (GNNs)*. Chen and co-workers incorporated fidelity tags into the Materials Graph Network (MEGNet) to learn across band-gap datasets ranging from the PBE (LF) functional to the HSE hybrid functional and experimental measurements (HF).[29] By embedding fidelity as a global state, their model integrated heterogeneous datasets and captured correlations across fidelity levels. As shown in Figure 4d, multi-fidelity GNNs consistently outperformed single-fidelity and stacking models. Performance is evaluated on randomly segmented test sets (80%-10%-10% split for training-validation-test sets, repeating 6 runs), where the error bars show one standard deviation across runs (confidence intervals) and the dots denote the individual model errors. Beyond the common two-level low- and high-fidelity setting, the effect of introducing multiple intermediate fidelity levels has been systematically examined. As additional fidelities are incorporated, prediction errors generally decrease and uncertainty bands narrow, reflecting improved regularization of the shared latent representation. However, adding a fidelity level with very limited data or weak correlation can degrade performance, as observed when a small SCAN dataset was included. This illustrates that increasing the number of fidelities is beneficial only when the added levels contribute sufficient data volume and maintain meaningful correlation with existing fidelities. These graph-based frameworks are particularly suited to materials discovery, as they naturally represent atomic topology while allowing fidelity-specific corrections.

Deep learning-based multi-fidelity frameworks therefore provide remarkable flexibility in handling heterogeneous and non-nested data, while capturing strongly nonlinear trends that challenge traditional surrogates. At the same time, they are computationally intensive, sensitive to hyperparameter choices, and often lack interpretability, which motivates continued integration of physics-based constraints to mitigate small-data challenges.[94] Overall, these

methods are emerging as powerful tools in materials science, particularly where properties are highly nonlinear, datasets are heterogeneous, and rapid discovery is essential. By coupling the adaptability of deep learning with domain knowledge, multi-fidelity neural frameworks are poised to become indispensable for next-generation materials design.

From a practical standpoint, the effectiveness of deep learning-based multi-fidelity frameworks depends strongly on careful hyperparameter selection and training strategy. In practice, effective deployment of deep learning-based multi-fidelity frameworks requires strategy-specific hyperparameter tuning. For transfer learning, a common approach is to stabilize the shared representation using abundant LF data, followed by limited fine-tuning with HF labels; when HF data are scarce, freezing early layers is often beneficial. For composite neural networks, loss weighting should prioritize HF accuracy while still retaining LF trend learning; a practical strategy is to begin with HF-weighted objectives and gradually relax toward a balanced multi-task loss. Throughout training, both accuracy and probabilistic metrics (e.g., negative log-likelihood and prediction-interval coverage) should be monitored to ensure that uncertainty calibration does not degrade as LF information is incorporated. For correction learning, residual blocks should be kept lightweight, with capacity sufficient to capture systematic LF→HF discrepancies without overfitting. For multi-fidelity graph neural networks, fidelity embeddings should remain compact and be jointly tuned with the shared trunk so that they modulate predictions without overwhelming structural features. In general, modest, budgeted hyperparameter searches focused on learning rate, weight decay, and loss weights (guided by HF validation and uncertainty calibration) tend to yield larger gains than aggressive architectural changes, particularly when HF data are limited.

2.5 Multi-Fidelity Active Learning and Adaptive Sampling

Active learning and adaptive sampling extend multi-fidelity modeling into a dynamic, decision-making context. Rather than passively integrating heterogeneous datasets, these approaches sequentially decide not only which candidate material to evaluate next, but also at which fidelity level (Figure 5a). The central objective is to maximize expected information gain per unit cost, thereby accelerating discovery while respecting finite computational and experimental budgets.

A central paradigm is *multi-fidelity Bayesian optimization (MF-BO)*, in which GP surrogates or deep-kernel variants model correlations between fidelity levels (Table 1).[50] Acquisition functions are adapted to weigh both evaluation cost and expected improvement, enabling the algorithm to alternate between inexpensive LF approximations and costly HF evaluations.[50] For example, MUlti-task Max-value Bayesian Optimization (MUMBO) employs a cost-aware acquisition function that selects the candidate-fidelity pair expected to deliver the largest reduction in uncertainty about the global optimum per unit cost, scaling effectively to multi-task and multi-fidelity settings.[50] Let (x, z) denote a material candidate x evaluated at fidelity level z. The next query (x_{n+1}, z_{n+1}) is chosen according to[50]

$$(x_{n+1}, z_{n+1}) = \arg \max_{(x,z)\in\mathcal{X}\times\mathcal{Z}} \frac{\alpha_n(x, z)}{c(x, z)}, \tag{7}$$

where $\alpha_n(x, z)$ is an acquisition function quantifying the expected improvement of the target property at iteration n, and $c(x, z)$ is a cost function associated with fidelity z.[50] This explicit balance between information gain and evaluation cost allows the algorithm to dynamically allocate resources across fidelities, alternating between LF approximations and HF evaluations.[50]

Applications in materials science highlight the versatility of these strategies. In materials screening, multi-output GPs have been used to fuse experimental and computational data on-the-fly, replacing rigid stagewise funnels with progressive, budget-aware exploration and cutting optimization costs by nearly threefold in benchmark studies.[54] In covalent organic frameworks (COFs) optimization, MF-BO achieved lower regret at a given budget compared to single-fidelity baselines, as shown in Figure 5b. To assess not only average performance but also reliability, both multi-fidelity (MF) and single-fidelity (SF) active learning strategies were evaluated using 20 independent random seeds, corresponding to different initial samples and acquisition trajectories. In Figure 5b, the shaded regions represent the standard deviation of simple regret at each budget level, capturing the spread of optimization outcomes across repeated campaigns. The visibly narrower shaded bands for MF strategies compared to their SF counterparts indicate reduced run-to-run variability and therefore higher reliability. These results highlight that

Table 1. Pseudo code of multi-fidelity Bayesian optimization (MF-BO).

1:	Given design space $\mathcal{X}$, fidelity set $\mathcal{Z}$ (e.g., LF, HF), initial dataset $D_0 = \{(x_i, z_i, y_i)\}$, total budget B, cost function $c(x, z)$, surrogate model class $\mathcal{S}$, and acquisition rule $\alpha_n(x, z)$.
2:	**Start** MF-BO loop.
3:	$n \leftarrow 0$, $D \leftarrow D_0$, $\text{spent\_cost} \leftarrow 0$.
4:	**while** $\text{spent\_cost} < B$ and stopping criterion not met **do**
5:	Fit multi-fidelity surrogate $s_n \leftarrow \mathcal{S}.\text{fit}(D)$
6:	**for** many candidate pairs (x, z) sampled from $\mathcal{X} \times \mathcal{Z}$ **do**
7:	Use s_n to obtain the predictive distribution of the target property at (x, z).
8:	Compute acquisition value $\alpha_n(x, z)$ based on this predictive distribution.
9:	Compute cost-aware score $\text{score}(x, z) \leftarrow \alpha_n(x, z)/c(x, z)$.
10:	**end for**
11:	Select next evaluation pair $(x_{\text{next}}, z_{\text{next}}) \leftarrow \arg\max_{(x,z)} \text{score}(x, z)$.
12:	Query simulator / experiment at chosen fidelity
13:	$y_{\text{next}} \leftarrow \text{evaluate\_material}(x_{\text{next}}, z_{\text{next}})$.
14:	Update dataset and budget
15:	$D \leftarrow D \cup \{(x_{\text{next}}, z_{\text{next}}, y_{\text{next}})\}$.
16:	$\text{spent\_cost} \leftarrow \text{spent\_cost} + c(x_{\text{next}}, z_{\text{next}})$.
17:	$n \leftarrow n + 1$.
18:	**end while**

multi-fidelity active learning not only accelerates convergence but also yields more reproducible and dependable discovery trajectories.[60]

Sequential learners extend this idea by combining regressors such as GPs, random forests, and support vector machines into active loops that adaptively decide whether to query a fast DFT simulation (LF) or a costly experiment (HF), with successful demonstrations in two-fidelity band-gap discovery agents.[47] In structural materials, MF-BO has been applied to dual-phase steel design, where LF micromechanical models (isostrain, isostress, isowork, secant, elastic-constraint) were queried far more often than costly FEM (with representative volume element, RVE) simulations, especially in early iterations.[53,95] Similarly, in interatomic potential development under extreme conditions, adaptive multi-fidelity sampling efficiently identified promising parameter regimes that would have been infeasible to explore exhaustively.[38] Recent innovations further embed these principles into generative frameworks. In the multi-fidelity generative flow network (MF-GFN), the acquisition function is treated as a reward signal that guides stochastic generation of candidate-fidelity tuples, enabling discovery of diverse, high-performing materials while maintaining cost efficiency and outperforming random or single-fidelity baselines (Figure 5c).[96]

The strengths of multi-fidelity active learning lie in their ability to allocate resources efficiently. By dynamically balancing exploration of uncertain regions with exploitation of promising candidates, these methods accelerate discovery trajectories compared to static or fixed-fidelity sampling. They are particularly well suited for autonomous experimental laboratories and adaptive simulation campaigns, where real-time decisions about the next measurement are essential.

The main challenges arise from implementation complexity and the need for robust uncertainty quantification. Acquisition functions must be carefully calibrated to avoid over-exploitation of biased LF data and to maintain trustworthy uncertainty estimates in high-dimensional, structured search spaces. Surrogate models that combine diverse fidelities can also become computationally demanding when embedded in sequential loops. Studies have shown that while approximate LF solvers can accelerate discovery, they may also mislead active learning if correlations with HF targets are not explicitly modeled.[97] Dimensionality is another bottleneck. Recent advances such as adaptive active-subspace methods embedded within MF-BO have improved efficiency by reducing search to informative subspaces in process-structure-property optimization.[95] Overall, multi-fidelity active learning and adaptive sampling represent a critical step toward autonomous, closed-loop materials discovery, where algorithms not only predict outcomes but also decide the most efficient way to generate new data.

From a practical perspective, effective deployment of multi-fidelity active learning benefits from staged hyper-parameter tuning. A recommended workflow is to first stabilize the multi-fidelity surrogate and cost models to ensure reliable predictions and uncertainty estimates. Next, the acquisition function should be calibrated to balance

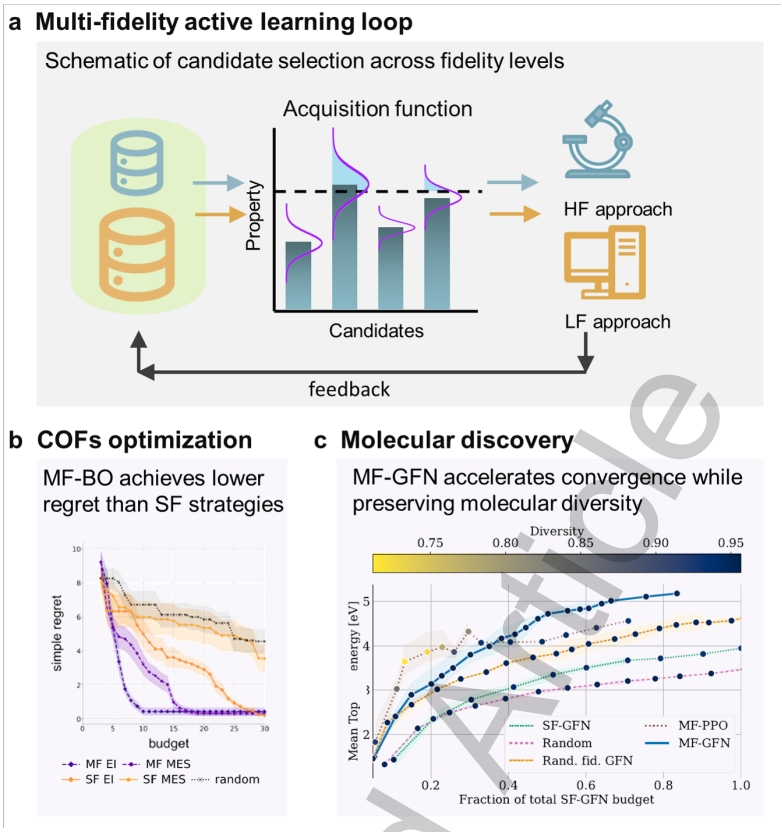

Figure 5. Multi-fidelity active learning and adaptive sampling in materials science. (a) Active-learning loop: schematic of candidate selection and fidelity allocation, where promising candidates are evaluated with costly HF methods and less valuable ones with inexpensive LF approximations. (b) COF optimization: MF active learning achieves lower simple regret for xenon–krypton selectivity compared to single-fidelity (SF) and random baselines.[60] (c) Molecular discovery: MF generative flow networks (MF-GFN) accelerate convergence toward diverse molecules with desirable electron affinity compared with SF and random-fidelity approaches.[96] Together, these examples illustrate how MF active learning dynamically balances exploration, exploitation, and cost to accelerate discovery across molecular and porous materials.

expected information gain against evaluation expense. Finally, scheduling parameters such as stopping criteria, fidelity budgets, or query ratios can be defined. In practice, a useful starting point is to pair a well-calibrated multi-fidelity surrogate with a simple cost model, for example based on computational time or dataset size.

The choice of acquisition strategy should align with the problem constraints and fidelity structure. Information-gain-per-cost criteria (e.g., entropy-based methods or multi-fidelity knowledge-gradient) are robust when fidelities differ strongly in accuracy and cost, whereas Expected Improvement (EI)-per-cost or Upper Confidence Bound (UCB)-per-cost strategies are effective when costs are relatively stationary across the domain. Stopping criteria can be organized into three practical tiers. Budget-based termination (e.g., fixed wall-time, total cost, or number of high-fidelity evaluations) is the default in autonomous campaigns with predefined resource limits. Progress-based criteria monitor the learning signal, such as halting when the maximum EI (or EI-per-cost) remains below a threshold for several iterations or when posterior uncertainty in the region of interest falls below a prescribed margin. Finally, value-of-information (VoI)-based stopping rules compare the expected knowledge gain per unit cost against a minimum acceptable return and terminate when further evaluations are no longer justified. The VoI-based strategies, including cost-aware knowledge-gradient variants, are particularly suitable when cost, risk, and uncertainty must be explicitly balanced.

3. CROSS-CUTTING THEMES

3.1 Bias vs. Variance

Compared with high-fidelity data, low-fidelity sources typically exhibit larger errors, which can be decomposed into bias and variance. Bias refers to systematic deviations introduced by differences in data-generation pipelines, such as experiment versus simulation, simplified physical models, alternative algorithms (e.g., HSE versus PBE), or choices of simulation hyperparameters.[29,32–34,52,82,91,95,98,99] Variance, by contrast, reflects random measurement noise and is generally lower in high-fidelity data. Examples include measurements made under the same protocol but carried out by novice versus experienced researchers, or using worn versus newly calibrated instruments.[100]

The relative contributions of bias and variance strongly influence the choice of multi-fidelity methodology, as show in Table 2. When systematic bias dominates, correction-based methods are particularly effective because they explicitly model cross-fidelity discrepancies. Deep learning approaches or models with explicit fidelity indicators (*e.g.*, one-hot encoding) can play a similar role by learning fidelity-specific corrections. When variance is the primary concern, statistical approaches with heteroscedastic modeling or inverse-variance weighting provide a principled treatment. In practice, especially in experimental materials data, bias and variance almost always coexist.[101] In such mixed settings, kriging-based methods (*e.g.*, co-kriging) that simultaneously capture cross-fidelity correlations and account for heteroscedastic errors are attractive options.[48] Active learning strategies also provide robust solutions, as they can adaptively down-weight low-fidelity inputs in non-critical regions without requiring explicit assumptions about error type.

Overall, recognizing whether bias, variance, or both dominate a multi-fidelity dataset is essential for selecting an appropriate strategy, and remains a central cross-cutting theme in the design of effective materials informatics workflows.

Table 2. Comparison of multi-fidelity strategies across advantages, limitations, bias–variance focus, and data requirements.

Strategy	Advantages	Limitations	Bias/Variance	Data Requirements
Statistical weighting	Simple, interpretable	Limited expressiveness	Variance	Non-nested, flexible
Direct ML with fidelity features	Flexible, leverages features	May underperform with sparse HF data	Bias	Works with non-nested
Correction-based (e.g., co-kriging)	Captures correlations, uncertainty quantification	Needs nested data, heavy compute	Both	Best with nested
Deep learning frameworks	Captures complex relations	Needs large datasets, less interpretable	Bias	Non-nested, but data-hungry
Active learning / adaptive sampling	Sample-efficient, reduces cost	Complex acquisition design	Both	Few samples; can create nested adaptively

Note: In this review, we use the terms *nested* and *non-nested* to describe whether the same inputs are observed at multiple fidelity levels. This is sometimes referred to as *co-located* vs. *non-co-located* in the literature; both describe the same concept of overlapping vs. non-overlapping input sets.

3.2 Data Requirements: Nested vs. Non-Nested

In addition to error characteristics, the feasibility of multi-fidelity methods is strongly influenced by their data requirements, particularly the distinction between nested and non-nested designs. In a nested design, the same input x is observed at multiple fidelity levels, allowing direct comparison and correction between low- and high-fidelity outputs. By contrast, in a non-nested design, most inputs are observed at only one fidelity level, leaving little or no overlap across fidelities.

Correction-based methods such as co-kriging are especially sensitive to this distinction, as they typically assume a nested structure in order to learn systematic deviations between fidelities. In practice, however, materials datasets are often non-nested, arising from disparate sources such as separate DFT databases, experimental measurements, or high-throughput simulations conducted with different parameter sets. This lack of overlap makes direct discrepancy modeling challenging and can reduce the effectiveness of classical co-kriging.

To address these challenges, several extensions have been proposed that relax the nesting requirement. Specifically, recursive variants of classical co-kriging can propagate correlations across multiple fidelity levels without requiring strict nesting of input locations.[48] In the Augmented Bayesian Treed Co-Kriging (ABTCK) framework, missing cross-fidelity pairings are statistically imputed through hierarchical partitioning of the input space, creating an augmented dataset in which the joint posterior can be factorized. This approach enables fully Bayesian predictive inference under non-nested designs without enforcing explicit co-location.[102] Beyond ABTCK, the Generalized Co-Kriging (GCK) framework aggregates a calibrated low-fidelity Kriging model with a stochastic discrepancy

Kriging term, such that the low-fidelity predictive distribution (mean and variance) is propagated to locations where only high-fidelity observations are available. This probabilistic coupling permits separate parameter estimation for each fidelity and enables robust fusion under both nested and non-nested conditions, improving accuracy and stability relative to classical AR1 formulations when LF-HF correlations are imperfect.[103] These methods are particularly useful in materials science, where integrating heterogeneous datasets, such as semi-local DFT with hybrid-functional calculations, or simulations with experiments, rarely yields perfectly aligned input sets.

It is important to note, however, that not all multi-fidelity strategies depend on nested data. For example, statistical weighting and direct ML with fidelity features can operate on non-nested datasets. Deep learning frameworks provide additional flexibility, as architectures such as hierarchical neural operators or graph-based encoders can learn shared latent spaces even when input sets across fidelities do not overlap; however, such approaches typically require larger training datasets to generalize reliably. In contrast, active learning and adaptive sampling are particularly suitable for few-sample scenarios, as they strategically select high-fidelity evaluations to maximize information gain and can dynamically create nested datasets when needed. As summarized in Table 2, the degree of nesting required varies widely across methodologies, underscoring the need for approaches that can exploit both nested and non-nested data in real-world materials design.

3.3 Computational Cost and Scalability

The computational demands of multi-fidelity algorithms vary widely, and scalability is often a decisive factor in method selection. Kriging-based approaches, including co-kriging, require inversion of an $n \times n$ covariance matrix, where n is the number of data points. As a result, training time scales as $O(n^3)$ and memory as $O(n^2)$.[104] In practice, this cubic scaling quickly becomes prohibitive, making covariance-matrix inversion the dominant computational bottleneck and limiting the applicability of co-kriging once multi-fidelity datasets reach thousands of entries.[48]

By contrast, deep learning-based multi-fidelity models can be trained using mini-batch stochastic gradient descent, for which the per-iteration computational cost scales approximately linearly with the batch size B and the number of trainable parameters P (i.e., $O(B \cdot P)$). The total training cost is therefore $O(E \cdot n \cdot P/B)$ for E epochs over n samples. This structure maps naturally onto modern hardware accelerators: data-parallel training across multiple GPUs can achieve near-linear speedup (subject to communication overhead), while mixed-precision training and gradient checkpointing substantially reduce memory footprint and wall-time. At inference time, prediction cost scales as $O(P)$ per query and can be efficiently batched, enabling deployment on datasets that are orders of magnitude larger than those tractable with Gaussian process or co-kriging models.

In summary, while kriging-based methods provide strong uncertainty quantification and interpretability, their poor scalability necessitates caution when applied to large datasets. Neural networks and other scalable surrogates offer practical alternatives in data-rich regimes, making computational cost a central criterion in choosing among multi-fidelity methodologies.

4. OUTLOOK: CHALLENGES AND OPPORTUNITIES

This section outlines both the challenges that constrain current progress and the opportunities that may define the future of multi-fidelity learning in materials design.

Despite notable progress, several fundamental challenges remain. First, data scarcity and imbalance are persistent barriers. High-fidelity data, whether from advanced simulations or experiments, remain expensive and unevenly distributed across material classes. The absence of standardized benchmarks further complicates fair comparison among methodologies, slowing methodological innovation and adoption.

High dimensionality further exacerbates effective data scarcity in multi-fidelity learning. As the number of compositional, structural, and processing descriptors grows, the design space expands rapidly, increasing the amount of data required for reliable surrogate modeling. This challenge is particularly pronounced for multi-fidelity methods, where accurate learning of LF-HF correlations often relies on co-located observations in high-dimensional spaces. As a result, simply adding more low-fidelity data may yield only limited benefit. This motivates multi-fidelity strategies that aim to reduce the effective dimensionality of the problem, for example through physics-informed feature engineering, active-subspace or manifold identification, and learned latent representations.

Beyond data volume and dimensionality, incomplete or partially observed feature sets also induce an implicit degradation of data fidelity. When key compositional, microstructural, processing, or environmental descriptors are missing, physically distinct material states may collapse onto similar representations in feature space, inflating epistemic uncertainty in poorly characterized regions. In heterogeneous materials databases, records with more complete descriptor sets can therefore be regarded as higher-fidelity observations, whereas samples with missing variables represent degraded-fidelity views that should be treated explicitly rather than discarded. For example, Ching and Phoon employed a hierarchical Bayesian framework with Gibbs sampling to infer missing feature values prior to multi-fidelity fusion, demonstrating improved robustness and predictive performance when only partial variables are available. This perspective highlights a fundamental connection between data completeness and fidelity and motivates future multi-fidelity frameworks that jointly address missing-data inference and fidelity-aware learning.[105]

Second, transferability across material classes and design domains is limited. Models trained on multi-fidelity datasets for alloys or oxides, for example, may not generalize to polymers or energy-storage materials without extensive retraining. This lack of universality reflects both intrinsic differences in physics across classes and the difficulty of encoding fidelity relationships in a transferable manner.

Third, interpretability and trust in black-box methods remain major obstacles. Deep-learning frameworks can achieve impressive accuracy in multi-fidelity prediction tasks, but their opaque decision-making makes it difficult to diagnose when low-fidelity biases are propagating unchecked. Without greater transparency and interpretability, the adoption of these methods in high-stakes applications, such as structural materials or energy technologies, will remain limited.

A particularly salient theme is the integration of multi-fidelity methods with automated laboratories and closed-loop workflows. On the one hand, current approaches face significant barriers: to guide experiments in real time, models must not only be accurate but also computationally efficient, uncertainty-aware, and robust to noisy data streams. In addition, practical deployment requires standardized data-exchange protocols that are still lacking. On the other hand, overcoming these challenges would make multi-fidelity learning the core intelligence layer of autonomous discovery platforms. Such integration promises to reduce costs, accelerate timelines, and open pathways to discovery that are inaccessible through traditional trial-and-error workflows.

These key challenges are summarized schematically in Figure 6, which also highlights the corresponding opportunities. At the same time, these challenges create avenues for innovation. Physics-informed multi-fidelity models represent a particularly promising direction, embedding domain knowledge into statistical or neural architectures to improve both accuracy and interpretability. Generative models, including diffusion models and variational autoencoders, offer new possibilities for sampling candidate materials across fidelities and for data augmentation when high-fidelity samples are scarce.

Another important frontier is multi-objective and multi-property optimization. Real-world materials design rarely targets a single property, and extending multi-fidelity frameworks to balance competing objectives, such as strength and ductility in alloys or conductivity and stability in batteries, will substantially expand their practical relevance.

Standardization of open multi-fidelity datasets is also crucial. Community-wide benchmarks, analogous to those in computer vision or natural language processing, would accelerate progress by enabling direct comparison of algorithms under reproducible conditions. Such datasets could also serve as training grounds for transfer learning across material classes.

Reducing cost in multi-fidelity learning typically entails fewer high-fidelity evaluations, which can increase epistemic uncertainty, particularly in regions of the input space that are sparsely sampled by high-fidelity data. Multi-fidelity models partially mitigate this effect by using abundant low-fidelity data to constrain the hypothesis space, but they cannot fully substitute for missing high-fidelity information. This inherent trade-off between cost and uncertainty motivates the development of cost-aware and uncertainty-aware strategies, such as multi-fidelity active learning, which aim to maximize uncertainty reduction per unit cost rather than minimizing cost alone.

In summary, while multi-fidelity learning in materials design faces significant challenges, it also offers a unique confluence of opportunities. Together, these challenges and opportunities define a roadmap for advancing multi-

fidelity learning as a central tool in next-generation materials discovery.

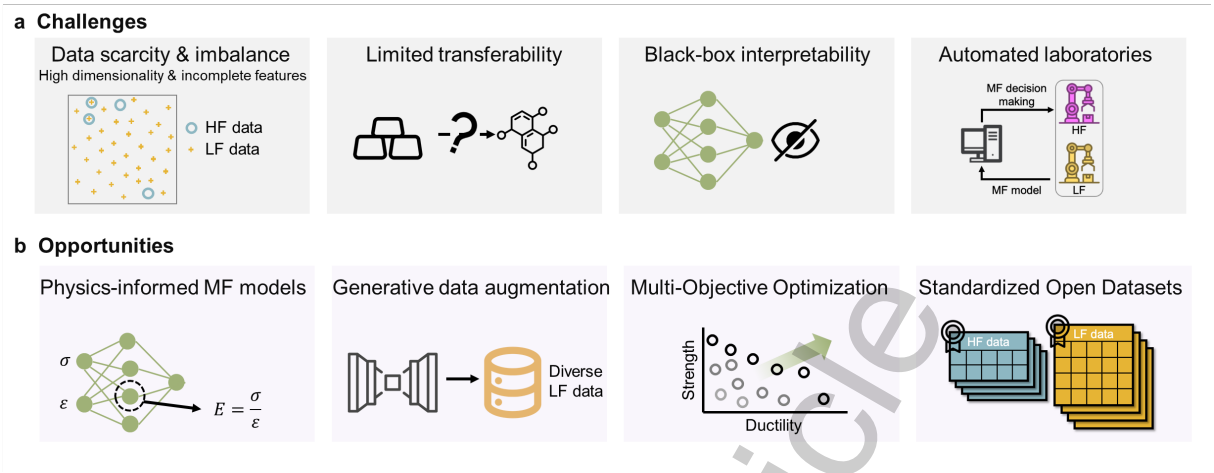

Figure 6. Conceptual summary of challenges and opportunities for multi-fidelity learning in materials design. a. **Challenges**: key barriers include data scarcity and imbalance between HF and LF datasets, limited transferability across material domains, the difficulty of interpreting black-box models, and integration with automated laboratories. b. **Opportunities**: promising directions involve physics-informed MF models that embed domain knowledge, generative approaches for data augmentation, multi-objective optimization frameworks, standardized open datasets of HF and LF data, and coupling with autonomous discovery platforms. Together, these challenges and opportunities define a roadmap for advancing multi-fidelity learning as a central tool in next-generation materials discovery.

5. CONCLUSION

This review has provided a structured overview of multi-fidelity learning methodologies for accelerating data-driven materials modeling, design, and discovery. We summarized major classes of multi-fidelity approaches, including statistical and parametric models, direct machine learning with fidelity features, correction-based methods such as co-kriging, deep learning frameworks, and multi-fidelity active learning and Bayesian optimization. Across these methods, we highlighted how heterogeneous data sources can be integrated to exploit cross-fidelity correlations, improve predictive accuracy, and reduce reliance on costly high-fidelity data. Key practical considerations were also discussed, including data organization, uncertainty quantification, scalability, and integration with adaptive sampling. Looking ahead, advances in physics-informed modeling, generative methods, standardized multi-fidelity datasets, and autonomous experimentation are expected to further expand the impact of multi-fidelity learning. Together, these developments position multi-fidelity approaches as a central component of efficient, reliable, and closed-loop materials discovery workflows.

DECLARATIONS

Authors' contributions

B.W. and D.X. prepared the initial draft of the manuscript. B.W., Y.X., Y.Z., and D.X. contributed to the conception and design of the review, and all authors participated in revising and improving the manuscript.

Availability of data and materials

Not applicable.

Financial support and sponsorship

This work was supported by National Key Research and Development Program of China (2021YFB3802100), National Natural Science Foundation of China (Nos. 52573255, 52173228, 52271190, and 524B200280), Innovation Capability Support Program of Shaanxi (2024ZG-GCZX-01(1)-06) and Natural Science Foundation Project of Shaanxi Province (Grant No. 2022JM-205).

Conflicts of interest

Dezhen Xue is a Youth Editorial Board Member of the Journal of Materials Informatics, but was not involved in any steps of editorial processing, notably including reviewer selection, manuscript handling, and decision making, while the other authors have declared that they have no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

REFERENCES

1. Batra R, Song L, Ramprasad R. Emerging materials intelligence ecosystems propelled by machine learning [Journal Article]. *Nature Reviews Materials* 2021;6:655–78. DOI
2. Rao Z, Tung PY, Xie R, Wei Y, Zhang H, et al. Machine learning–enabled high-entropy alloy discovery [Journal Article]. *Science* 2022;378:78–85. Available from: https://www.science.org/doi/abs/10.1126/science.abo4940. DOI
3. Sohail Y, Zhang C, Xue D, Zhang J, Zhang D, et al. Machine-learning design of ductile FeNiCoAlTa alloys with high strength [Journal Article]. *Nature* 2025;643:119–24. Available from: https://doi.org/10.1038/s41586-025-09160-2. DOI
4. Wang T, Pan R, Martins ML, Cui J, Huang Z, et al. Machine-learning-assisted material discovery of oxygen-rich highly porous carbon active materials for aqueous supercapacitors [Journal Article]. *Nature Communications* 2023;14:4607. Available from: https://doi.org/10.1038/s41467-023-40282-1. DOI
5. Wu J, Torresi L, Hu M, Reiser P, Zhang J, et al. Inverse design workflow discovers hole-transport materials tailored for perovskite solar cells [Journal Article]. *Science* 2024;386:1256–64. Available from: https://www.science.org/doi/abs/10.1126/science.ads0901. DOI
6. Tian Y, Hu B, Dang P, Pang J, Zhou Y, et al. Noise-Aware Active Learning to Develop High-Temperature Shape Memory Alloys with Large Latent Heat [Journal Article]. *Advanced Science* 2024:11:2406216. Available from: https://advanced.onlinelibrary.wiley.com/doi/abs/10.1002/advs.202406216. DOI
7. Xian Y, Dang P, Tian Y, Jiang X, Zhou Y, et al. Compositional design of multicomponent alloys using reinforcement learning [Journal Article]. *Acta Materialia* 2024;274:120017. Available from: https://www.sciencedirect.com/science/article/pii/S1359645424003690. DOI
8. Wei Q, Cao B, Yuan H, Chen Y, You K, et al. Divide and conquer: Machine learning accelerated design of lead-free solder alloys with high strength and high ductility [Journal Article]. *npj Computational Materials* 2023 10;9. DOI
9. Gong X, Louie S, Duan W, Xu Y. Generalizing deep learning electronic structure calculation to the plane-wave basis [Journal Article]. *Nature Computational Science* 2024;4:752–60. DOI
10. Dang L, He X, Tang D, Xin H, Wu B. A fatigue life prediction framework of laser-directed energy deposition Ti-6Al-4V based on physics-informed neural network [Journal Article]. *International Journal of Structural Integrity* 2025 02;16:327–54. Available from: https://doi.org/10.1108/IJSI-10-2024-0170. DOI
11. Rasul A, Karuppanan S, Perumal V, Ovinis M, Iqbal M. An artificial neural network model for determining stress concentration factors for fatigue design of tubular T-joint under compressive loads [Journal Article]. *International Journal of Structural Integrity* 2024;15:633–52. Available from: https://doi.org/10.1108/IJSI-02-2024-0034. DOI
12. Yuan R, Wang B, Li J, Sun P, Liu Z, et al. Machine learning-enabled design of ferroelectrics with multiple properties via a Landau model [Journal Article]. *Acta Materialia* 2025;286:120760. Available from: https://www.sciencedirect.com/science/article/pii/S1359645425000539. DOI
13. Wang W, Jiang X, Tian S, Liu P, Dang D, et al. Automated pipeline for superalloy data by text mining [Journal Article]. *npj Computational Materials* 2022;8:9. Available from: https://doi.org/10.1038/s41524-021-00687-2. DOI
14. Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science [Journal Article]. *Nature* 2018;559:547–55. Available from: https://doi.org/10.1038/s41586-018-0337-2. DOI
15. Raccuglia P, Elbert KC, Adler PDF, Falk C, Wenny MB, et al. Machine-learning-assisted materials discovery using failed experiments [Journal Article]. *Nature* 2016;533:73–76. Available from: https://doi.org/10.1038/nature17439. DOI
16. Dinic F, Singh K, Dong T, Rezazadeh M, Wang Z, et al. Applied Machine Learning for Developing Next-Generation Functional Materials [Journal Article]. *Advanced Functional Materials* 2021;31:2104195. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/adfm.202104195. DOI
17. Rickman JM, Lookman T, Kalinin SV. Materials informatics: From the atomic-level to the continuum [Journal Article]. *Acta Materialia* 2019;168:473–510. Available from: https://www.sciencedirect.com/science/article/pii/S1359645419300667. DOI
18. Hong T, Chen T, Jin D, Zhu Y, Gao H, et al. Discovery of new topological insulators and semimetals using deep generative models [Journal Article]. *npj Quantum Materials* 2025;10. DOI

19. Xie J. Prospects of materials genome engineering frontiers [Journal Article]. *Materials Genome Engineering Advances* 2023;1:e17. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/mgea.17. DOI
20. Jiang X, Fu H, Bai Y, Jiang L, Zhang H, et al. Interpretable Machine Learning Applications: A Promising Prospect of AI for Materials [Journal Article]. *Advanced Functional Materials* 2025. DOI
21. Hart GLW, Mueller T, Toher C, Curtarolo S. Machine learning for alloys [Journal Article]. *Nature Reviews Materials* 2021;6:730–55. DOI
22. Hippalgaonkar K, Li Q, Wang X, Fisher III JW, Kirkpatrick J, et al. Knowledge-integrated machine learning for materials: lessons from gameplaying and robotics [Journal Article]. *Nature Reviews Materials* 2023;8:241–60. DOI
23. Tabor DP, Roch LM, Saikin SK, Kreisbeck C, Sheberla D, et al. Accelerating the discovery of materials for clean energy in the era of smart automation [Journal Article]. *Nature Reviews Materials* 2018;3:5–20. DOI
24. Gong X, Li H, Zou N, Xu R, Duan W, et al. General framework for E(3)-equivariant neural network representation of density functional theory Hamiltonian [Journal Article]. *Nature Communications* 2023;14. DOI
25. Li H, Wang Z, Zou N, Ye M, Xu R, et al. Deep-learning density functional theory Hamiltonian for efficient ab initio electronic-structure calculation [Journal Article]. *Nature Computational Science* 2022;2:367–77. DOI
26. Cohen AJ, Mori-Sanchez P, Yang W. Challenges for Density Functional Theory [Journal Article]. *Chemical Reviews* 2012;112:289–320. DOI
27. Janesko BG. Replacing hybrid density functional theory: motivation and recent advances [Journal Article]. *Chemical Society Reviews* 2021;50:8470–95. DOI
28. Takeno S, Tsukada Y, Fukuoka H, Koyama T, Shiga M, et al. Cost-effective search for lower-error region in material parameter space using multifidelity Gaussian process modeling [Journal Article]. *Physical Review Materials* 2020;4:083802. Available from: https://link.aps.org/doi/10.1103/PhysRevMaterials.4.083802. DOI
29. Chen C, Zuo Y, Ye W, Li X, Ong SP. Learning properties of ordered and disordered materials from multi-fidelity data; 2021. Available from: https://doi.org/10.48550/arXiv.2005.04338. DOI
30. Lu L, Dao M, Kumar P, Ramamurty U, Karniadakis GE, et al. Extraction of mechanical properties of materials through deep learning from instrumented indentation [Journal Article]. *Proceedings of the National Academy of Sciences* 2020;117:7052–62. Available from: https://www.pnas.org/doi/abs/10.1073/pnas.1922210117. DOI
31. Jain A, Hautier G, Moore CJ, Ong SP, Fischer CC, et al. A high-throughput infrastructure for density functional theory calculations [Journal Article]. *Computational Materials Science* 2011;50:2295–310. DOI
32. Pilania G, Gubernatis JE, Lookman T. Multi-fidelity machine learning models for accurate bandgap predictions of solids [Journal Article]. *Computational Materials Science* 2017;129:156–63. Available from: https://www.sciencedirect.com/science/article/pii/S0927025616306188. DOI
33. Egorova O, Hafizi R, Woods DC, Day GM. Multifidelity Statistical Machine Learning for Molecular Crystal Structure Prediction [Journal Article]. *The Journal of Physical Chemistry A* 2020;124:8065–78. Available from: https://doi.org/10.1021/acs.jpca.0c05006. DOI
34. Batra R, Pilania G, Uberuaga BP, Ramprasad R. Multifidelity Information Fusion with Machine Learning: A Case Study of Dopant Formation Energies in Hafnia [Journal Article]. *ACS Applied Materials & Interfaces* 2019;11:24906–18. Available from: https://doi.org/10.1021/acsami.9b02174. DOI
35. Jha D, Choudhary K, Tavazza F, Liao Wk, Choudhary A, et al. Enhancing materials property prediction by leveraging computational and experimental data using deep transfer learning [Journal Article]. *Nature Communications* 2019;10:5316. Available from: https://doi.org/10.1038/s41467-019-13297-w. DOI
36. Peng B, Zhang M, Ye D. DIC/DSI based studies on the local mechanical behaviors of HR3C/T92 dissimilar welded joint during plastic deformation [Journal Article]. *Materials Science and Engineering: A* 2022;857:144073. Available from: https://www.sciencedirect.com/science/article/pii/S0921509322014526. DOI
37. Valladares H, Li T, Zhu L, El-Mounayri H, Hashem AM, et al. Gaussian process-based prognostics of lithium-ion batteries and design optimization of cathode active materials [Journal Article]. *Journal of Power Sources* 2022;528:231026. Available from: https://www.sciencedirect.com/science/article/pii/S0378775322000507. DOI
38. Lindsey RK, Bastea S, Hamel S, Lyu Y, Goldman N, et al. ChIMES Carbon 2.0: A transferable machine-learned interatomic model harnessing multifidelity training data [Journal Article]. *npj Computational Materials* 2025;11:26. Available from: https://doi.org/10.1038/s41524-024-01497-y. DOI
39. Ko TW, Ong SP. Data-efficient construction of high-fidelity graph deep learning interatomic potentials [Journal Article]. *npj Computational Materials* 2025;11:65. Available from: https://doi.org/10.1038/s41524-025-01550-4. DOI
40. Kim J, Kim J, Kim J, Lee J, Park Y, et al. Data-Efficient Multifidelity Training for High-Fidelity Machine Learning Interatomic Potentials [Journal Article]. *Journal of the American Chemical Society* 2025;147:1042–54. Available from: https://doi.org/10.1021/jacs.4c14455. DOI
41. Perdikaris P, Raissi M, Damianou A, Lawrence ND, Karniadakis GE. Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling [Journal Article]. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 2017;473:20160751. Available from: https://royalsocietypublishing.org/doi/abs/10.1098/rspa.2016.0751. DOI
42. Liu M, Gopakumar A, Hegde VI, He J, Wolverton C. High-throughput hybrid-functional DFT calculations of bandgaps and formation energies and multifidelity learning with uncertainty quantification [Journal Article]. *Physical Review Materials* 2024;8:043803. Available from: https://link.aps.org/doi/10.1103/PhysRevMaterials.8.043803. DOI
43. Yang ZN, Zou J, Huang L, Yang R, Zhang JY, et al. Machine learning-based extraction of mechanical properties from multi-fidelity small punch test data [Journal Article]. *Advances in Manufacturing* 2025;13:511–24. Available from: https://doi.org/10.1007/s40436-024-00540-x. DOI

44. Yi J, Ferreira BP, Bessa MA. Single-to-multi-fidelity history-dependent learning with uncertainty quantification and disentanglement: Application to data-driven constitutive modeling [Journal Article]. *Computer Methods in Applied Mechanics and Engineering* 2026;448:118479. Available from: https://www.sciencedirect.com/science/article/pii/S0045782525007510. DOI
45. Yi J, Cheng J, Bessa MA. Practical multi-fidelity machine learning: fusion of deterministic and Bayesian models; 2025. Available from: https://arxiv.org/abs/2407.15110. DOI
46. Cheng M, Jiang P, Hu J, Shu L, Zhou Q. A multi-fidelity surrogate modeling method based on variance-weighted sum for the fusion of multiple non-hierarchical low-fidelity data [Journal Article]. *Structural and Multidisciplinary Optimization* 2021;64:3797–818. Available from: https://doi.org/10.1007/s00158-021-03055-2. DOI
47. Palizhati A, Torrisi SB, Aykol M, Suram SK, Hummelshøj JS, et al. Agents for sequential learning using multiple-fidelity data [Journal Article]. *Scientific Reports* 2022;12:4694. Available from: https://doi.org/10.1038/s41598-022-08413-8. DOI
48. Le Gratiet L. Multi-fidelity Gaussian process regression for computer experiments [Thesis]; 2013.
49. Meng X, Karniadakis GE. A composite neural network that learns from multi-fidelity data: Application to function approximation and inverse PDE problems [Journal Article]. *Journal of Computational Physics* 2020;401:109020. Available from: https://www.sciencedirect.com/science/article/pii/S0021999119307260. DOI
50. Moss HB, Leslie DS, Rayson P. MUMBO: MUlti-task Max-Value Bayesian Optimization. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). vol. 12459 LNAI; 2021. pp. 447–62. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85103303615. DOI
51. Xu X, Zhao W, Hu Y, Wang L, Lin J, et al. Discovery of thermosetting polymers with low hygroscopicity, low thermal expansivity, and high modulus by machine learning [Journal Article]. *Journal of Materials Chemistry A* 2023;11:12918–27. Available from: http://dx.doi.org/10.1039/D2TA09272G. DOI
52. Patra A, Batra R, Chandrasekaran A, Kim C, Huan TD, et al. A multi-fidelity information-fusion approach to machine learn and predict polymer bandgap [Journal Article]. *Computational Materials Science* 2020;172:109286. Available from: https://www.sciencedirect.com/science/article/pii/S0927025619305853. DOI
53. Molkeri A, Khatamsaz D, Couperthwaite R, James J, Arróyave R, et al. On the importance of microstructure information in materials design: PSP vs PP [Journal Article]. *Acta Materialia* 2022;223:117471. Available from: https://www.sciencedirect.com/science/article/pii/S1359645421008508. DOI
54. Fare C, Fenner P, Benatan M, Varsi A, Pyzer-Knapp EO. A multi-fidelity machine learning approach to high throughput materials screening [Journal Article]. *npj Computational Materials* 2022;8:257. Available from: https://doi.org/10.1038/s41524-022-00947-9. DOI
55. He Y, Cubuk ED, Allendorf MD, Reed EJ. Metallic Metal-Organic Frameworks Predicted by the Combination of Machine Learning Methods and Ab Initio Calculations [Journal Article]. *Journal of Physical Chemistry Letters* 2018;9:4562–69. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85050867632. DOI
56. Yang J, Manganaris P, Mannodi-Kanakkithodi A. Discovering novel halide perovskite alloys using multi-fidelity machine learning and genetic algorithm [Journal Article]. *The Journal of Chemical Physics* 2024;160. Available from: https://doi.org/10.1063/5.0182543. DOI
57. Xu B, Hu X, Lan H, Wang T, Xu XY, et al. Phthalonitrile melting point prediction enabled by multi-fidelity learning [Journal Article]. *Journal of Materials Informatics* 2024;4:21. Available from: https://www.oaepublish.com/articles/jmi.2024.27. DOI
58. He GF, Zhang P, Yin ZY. Active learning inspired multi-fidelity probabilistic modelling of geomaterial property [Journal Article]. *Computer Methods in Applied Mechanics and Engineering* 2024;432:117373. Available from: https://www.sciencedirect.com/science/article/pii/S0045782524006285. DOI
59. Wang L, Zhu SP, Wu B, Xu Z, Luo C, et al. Multi-fidelity physics-informed machine learning framework for fatigue life prediction of additive manufactured materials [Journal Article]. *Computer Methods in Applied Mechanics and Engineering* 2025;439:117924. Available from: https://www.sciencedirect.com/science/article/pii/S0045782525001963. DOI
60. Sabanza-Gil V, Barbano R, Gutiérrez DP, Luterbacher JS, Hernández-Lobato JM, et al.. Best Practices for Multi-Fidelity Bayesian Optimization in Materials and Molecular Research; 2025. Available from: https://arxiv.org/abs/2410.00544. DOI
61. Seber GAF, Wild CJ. Nonlinear Regression. Wiley Series in Probability and Statistics. Wiley; 2005. Available from: https://books.google.ca/books?id=YBYlCpBNo.cC. DOI
62. Bishop CM. Pattern Recognition and Machine Learning. Information Science and Statistics. Springer; 2006. Available from: https://books.google.ca/books?id=kTNoQgAACAAJ.
63. Yang B, Chen B, Liu Y, Chen J. Gaussian process fusion method for multi-fidelity data with heterogeneity distribution in aerospace vehicle flight dynamics [Journal Article]. *Engineering Applications of Artificial Intelligence* 2024;138:109228. Available from: https://www.sciencedirect.com/science/article/pii/S0952197624013861. DOI
64. Sella V, Pham J, Chaudhuri A, Willcox KE. In: Projection-based multifidelity linear regression for data-poor applications. AIAA SciTech Forum. American Institute of Aeronautics and Astronautics; 2023. Available from: https://doi.org/10.2514/6.2023-0916. DOI
65. Ozbayram O, Olivier A, Graham-Brady L. Heteroscedastic Gaussian Process Regression for material structure–property relationship modeling [Journal Article]. *Computer Methods in Applied Mechanics and Engineering* 2024;431:117326. Available from: https://www.sciencedirect.com/science/article/pii/S0045782524005814. DOI
66. Ibrahim A, Ataca C. Prediction of Frequency-Dependent Optical Spectrum for Solid Materials: A Multioutput and Multifidelity Machine Learning Approach [Journal Article]. *ACS Applied Materials & Interfaces* 2024;16:41145–56. Available from: https://doi.org/10.1021/acsami.4c07328. DOI
67. Mora C, Eweis-Labolle JT, Johnson T, Gadde L, Bostanabad R. Probabilistic neural data fusion for learning from an arbitrary number of multi-fidelity data sets [Journal Article]. *Computer Methods in Applied Mechanics and Engineering* 2023;415:116207. Available from:

https://www.sciencedirect.com/science/article/pii/S0045782523003316. DOI
68. Zanjani Foumani Z, Shishehbor M, Yousefpour A, Bostanabad R. Multi-fidelity cost-aware Bayesian optimization [Journal Article]. *Computer Methods in Applied Mechanics and Engineering* 2023;407:115937. Available from: https://www.sciencedirect.com/science/article/pii/S0045782523000609. DOI
69. Toal DJJ. Some considerations regarding the use of multi-fidelity Kriging in the construction of surrogate models [Journal Article]. *Structural and Multidisciplinary Optimization* 2015;51:1223–45. Available from: https://doi.org/10.1007/s00158-014-1209-5. DOI
70. Gratiet LL, Garnier J. Recursive Co-Kriging Model for Design of Computer Experiments with Multiple Levels of Fidelity [Journal Article]. *International Journal for Uncertainty Quantification* 2014;4:365–86. DOI
71. Forrester AIJ, Sóbester A, Keane AJ. Multi-fidelity optimization via surrogate modelling [Journal Article]. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 2007;463:3251–69. Available from: https://royalsocietypublishing.org/doi/abs/10.1098/rspa.2007.1900. DOI
72. Xiao M, Zhang G, Breitkopf P, Villon P, Zhang W. Extended Co-Kriging interpolation method based on multi-fidelity data [Journal Article]. *Applied Mathematics and Computation* 2018;323:120–31. Available from: https://www.sciencedirect.com/science/article/pii/S0096300317307646. DOI
73. Lu CK, Shafto P. Conditional Deep Gaussian Processes: Multi-Fidelity Kernel Learning [Journal Article]. *Entropy* 2021;23:1545. Available from: https://www.mdpi.com/1099-4300/23/11/1545. DOI
74. Cutajar K, Pullin M, Damianou A, Lawrence N, González J. Deep Gaussian Processes for Multi-fidelity Modeling; 2019. Available from: https://arxiv.org/abs/1903.07320. DOI
75. Raissi M, Karniadakis G. Deep Multi-fidelity Gaussian Processes; 2016. Available from: https://arxiv.org/abs/1604.07484. DOI
76. Song X, Lv L, Sun W, Zhang J. A radial basis function-based multi-fidelity surrogate model: exploring correlation between high-fidelity and low-fidelity models [Journal Article]. *Structural and Multidisciplinary Optimization* 2019;60:965–81. Available from: https://doi.org/10.1007/s00158-019-02248-0. DOI
77. Acar P. Crystal Plasticity Model Calibration for Ti-7Al Alloy with a Multi-fidelity Computational Scheme [Journal Article]. *Integrating Materials and Manufacturing Innovation* 2018;7:186–94. Available from: https://doi.org/10.1007/s40192-018-0120-0. DOI
78. Shah AA, Leung PK, Xing WW. Rapid high-fidelity quantum simulations using multi-step nonlinear autoregression and graph embeddings [Journal Article]. *npj Computational Materials* 2025;11:57. Available from: https://doi.org/10.1038/s41524-024-01479-0. DOI
79. Pan W, Sun L, Guo C, Yang X, Sun J, et al. An efficient/accurate multi-scale fatigue prediction method for Metal-Polymer hybrid (MPH) interface [Journal Article]. *International Journal of Fatigue* 2024;184:108304. Available from: https://www.sciencedirect.com/science/article/pii/S0142112324001622. DOI
80. Saunders RN, Teferra K, Elwany A, Michopoulos JG, Lagoudas D. Metal AM process-structure-property relational linkages using Gaussian process surrogates [Journal Article]. *Additive Manufacturing* 2023;62:103398. Available from: https://www.sciencedirect.com/science/article/pii/S2214860423000118. DOI
81. Morse L, Khodaei ZS, Aliabadi MH. A multi-fidelity modelling approach to the statistical inference of the equivalent initial flaw size distribution for multiple-site damage [Journal Article]. *International Journal of Fatigue* 2019;120:329–41. Available from: https://www.sciencedirect.com/science/article/pii/S0142112318305255. DOI
82. Tran A, Tranchida J, Wildey T, Thompson AP. Multi-fidelity machine-learning with uncertainty quantification and Bayesian optimization for materials design: Application to ternary random alloys [Journal Article]. *Journal of Chemical Physics* 2020;153:8. Available from: ⟨GotoISI⟩://WOS:000619306200001https://aip.scitation.org/doi/pdf/10.1063/5.0015672. DOI
83. Menon N, Mondal S, Basak A. Multi-Fidelity Surrogate-Based Process Mapping with Uncertainty Quantification in Laser Directed Energy Deposition [Journal Article]. *Materials* 2022;15:2902. Available from: https://www.mdpi.com/1996-1944/15/8/2902. DOI
84. Wiangkham A, Aengchuan P, Kasemsri R, Pichitkul A, Tantrairatn S, et al. Improvement of Mixed-Mode I/II Fracture Toughness Modeling Prediction Performance by Using a Multi-Fidelity Surrogate Model Based on Fracture Criteria [Journal Article]. *Materials* 2022;15:8580. Available from: https://www.mdpi.com/1996-1944/15/23/8580. DOI
85. Howard AA, Perego M, Karniadakis GE, Stinis P. Multifidelity deep operator networks for data-driven and physics-informed problems [Journal Article]. *Journal of Computational Physics* 2023;493:112462. Available from: https://www.sciencedirect.com/science/article/pii/S0021999123005570. DOI
86. Motamed M. A multi-fidelity neural network surrogate sampling method for uncertainty quantification; 2020. Available from: https://arxiv.org/abs/1909.01859. DOI
87. Song DH, Tartakovsky DM. Transfer Learning on Multi-Fidelity Data; 2021. Available from: https://arxiv.org/abs/2105.00856. DOI
88. Chakraborty S. Transfer learning based multi-fidelity physics informed deep neural network [Journal Article]. *Journal of Computational Physics* 2021;426:109942. Available from: https://www.sciencedirect.com/science/article/pii/S0021999120307166. DOI
89. Snapp KL, Silverman S, Pang R, Tiano TM, Lawton TJ, et al. A physics-informed impact model refined by multi-fidelity transfer learning [Journal Article]. *Extreme Mechanics Letters* 2024;72:102223. Available from: https://www.sciencedirect.com/science/article/pii/S2352431624001032. DOI
90. Chen D, Li Y, Liu K, Li Y. A physics-informed neural network approach to fatigue life prediction using small quantity of samples [Journal Article]. *International Journal of Fatigue* 2023;166:107270. Available from: https://www.sciencedirect.com/science/article/pii/S0142112322005205. DOI
91. Islam M, Thakur MSH, Mojumder S, Hasan MN. Extraction of material properties through multi-fidelity deep learning from molecular dynamics simulation [Journal Article]. *Computational Materials Science* 2021;188. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85097093631. DOI

92. Liu B, Yang Y, Wang G, Li Y. A Method for Calculating Residual Strength of Crack Arrest Hole on Tungsten-Copper Functionally Graded Materials by Phase-Field Gradient Element Combined with Multi-Fidelity Neural Network [Journal Article]. *Materials* 2025;18:1973. Available from: https://www.mdpi.com/1996-1944/18/9/1973. DOI
93. Pierre-Paul De B, Grégoire H, Gian-Marco R. Accurate experimental band gap predictions with multifidelity correction learning [Journal Article]. *Journal of Materials Informatics* 2022;2:10. Available from: http://dx.doi.org/10.20517/jmi.2022.13. DOI
94. Cleeman J, Agrawala K, Nastarowicz E, Malhotra R. Partial-physics-informed multi-fidelity modeling of manufacturing processes [Journal Article]. *Journal of Materials Processing Technology* 2023;320:118125. Available from: https://www.sciencedirect.com/science/article/pii/S0924013623002704. DOI
95. Khatamsaz D, Molkeri A, Couperthwaite R, James J, Arróyave R, et al. Adaptive active subspace-based efficient multifidelity materials design [Journal Article]. *Materials & Design* 2021;209:110001. Available from: https://www.sciencedirect.com/science/article/pii/S0264127521005566. DOI
96. Hernández-García A, Saxena N, Jain M, Liu CH, Bengio Y. Multi-Fidelity Active Learning with GFlowNets. *Transactions on Machine Learning Research* 2024. Expert Certification. Available from: https://openreview.net/forum?id=dLaazW9zuF.
97. Koutsourelakis PS. Accurate Uncertainty Quantification Using Inaccurate Computational Models [Journal Article]. *SIAM Journal on Scientific Computing* 2009;31:3274–300. Available from: https://epubs.siam.org/doi/abs/10.1137/080733565. DOI
98. Wang Z, Liu X, Chen H, Yang T, He Y. Exploring Multi-Fidelity Data in Materials Science: Challenges, Applications, and Optimized Learning Strategies [Journal Article]. *Applied Sciences* 2023;13. Available from: https://www.mdpi.com/2076-3417/13/24/13176. DOI
99. Greenman KP, Green WH, Gómez-Bombarelli R. Multi-fidelity prediction of molecular optical peaks with deep learning [Journal Article]. *Chemical Science* 2022;13:1152–62. Available from: http://dx.doi.org/10.1039/D1SC05677H. DOI
100. Burdick RK, Borror CM, Montgomery DC. A Review of Methods for Measurement Systems Capability Analysis [Journal Article]. *Journal of Quality Technology* 2003;35:342–54. Available from: https://doi.org/10.1080/00224065.2003.11980232. DOI
101. Giannoukou K, Marelli S, Sudret B. Uncertainty-aware multi-fidelity surrogate modeling with noisy data; 2024. Available from: https://arxiv.org/abs/2401.06447. DOI
102. Konomi BA, Karagiannis G. Bayesian analysis of multifidelity computer models with local features and non-nested experimental designs: Application to the WRF model; 2019. Available from: https://arxiv.org/abs/1910.08063. DOI
103. Zhou Q, Wu Y, Guo Z, Hu J, Jin P. A generalized hierarchical co-Kriging model for multi-fidelity data fusion [Journal Article]. *Structural and Multidisciplinary Optimization* 2020;62:1885–904. Available from: https://doi.org/10.1007/s00158-020-02583-7. DOI
104. Rasmussen CE, Williams CKI. Gaussian Processes for Machine Learning. The MIT Press; 2005. Available from: https://doi.org/10.7551/mitpress/3206.001.0001. DOI
105. Ching J, Phoon KK. Constructing a Site-Specific Multivariate Probability Distribution Using Sparse, Incomplete, and Spatially Variable (MUSIC-X) Data [Journal Article]. *Journal of Engineering Mechanics* 2020;146:04020061. Available from: https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29EM.1943-7889.0001779. DOI