Nano Research

Open Access

Review Article

Machine learning-driven material intelligence research and development

Huijie Zhou[1,2], Jiming Xu[1,4], Xinyu Qin[2]✉, Jing Zhang[1], Wenjiang Zou[1,5], Mohsen Shakouri[3], Jiang Xu[1], Lvzhou Li[1], Jianning Ding[1]✉, and Huan Pang[1,2]✉

[1] *Institute of Technology for Carbon Neutralization, Yangzhou University, Yangzhou 225127, China*
[2] *School of Chemistry and Chemical Engineering, Yangzhou University, Yangzhou 225009, China*
[3] *Canadian Light Source, University of Saskatchewan, Saskatoon, Saskatchewan S7N 2V3, Canada*
[4] *School of Chemistry and Chemical Engineering, Chongqing University of Science and Technology, Chongqing 401331, China*
[5] *School of Environmental Science, Nanjing Xiaozhuang University, Nanjing 211171, China*

Cite this article: *Nano Research*, **2025**, *18*, 94908095. https://doi.org/10.26599/NR.2025.94908095

ABSTRACT: Machine learning (ML) is transforming material research and development (R&D), driving a fundamental shift from experience-driven approaches to data-driven frameworks. This review systematically highlights the transformative breakthroughs brought by machine learning throughout the entire process of intelligent material innovation. And it provides a comprehensive full chain analysis, from atomic scale design to macroscopic applications, emphasizing multi-scale modeling that combines physical mechanisms with data-driven methods, running through all stages of material innovation. In the design phase, ML promotes performance-oriented structural optimization through inverse design systems and generative models. For synthesis and processing, closed-loop autonomous systems and green controllable synthesis strategies significantly improve efficiency and sustainability. In terms of advanced representation, ML-powered techniques can help proactively tackle key challenges of complex structures. Performance prediction models enable precise correlations between material properties and extreme properties (such as auxiliary structures) by revealing catalytic descriptors and decoding biological interface mechanisms. Ultimately, these ML-driven advancements are unlocking practical applications in key fields, such as energy, biomedicine, environmental remediation, and structural engineering. This article aims to provide a comprehensive technological roadmap for the next generation of smart material development by integrating cross scale insights and autonomous strategies, and to outline future directions for this rapidly developing paradigm.

KEYWORDS: nanomaterial design, inverse design, autonomous synthesis, machine learning, intelligent characterization, multi-scale modeling

1 Introduction

The traditional paradigm of material research and development (R&D) has long been constrained by bottlenecks, including high trial and error costs, ambiguously defined structure–property relationships, and disconnected cross-scale mechanisms [1–7].

Received: August 28, 2025; **Revised:** September 15, 2025
Accepted: September 17, 2025

✉Address correspondence to Xinyu Qin, xinyuqin@snu.ac.kr; Huan Pang, huanpangchem@hotmail.com; Jianning Ding, dingjn@yzu.edu.cn

With rapid breakthrough in artificial intelligence (AI) technology, machine learning (ML) is deeply reconstructing the entire chain of material research and development. The full chain design of machine learning, from atomic scale structural design to macroscopic device performance optimization, forming a closed-loop system of "data-driven design- autonomous synthesis-intelligent characterization- application landing" [6–13]. In the field of intelligent design, ML has promoted innovation in the reverse design paradigm. For example, reinforcement learning-based Monte Carlo tree search achieves global optimization of protein ultra compressed structures, whereas generative models such as ViNAS Pro construct a virtual material library using standardized protein data bank (PDB) annotations, thereby supporting reverse

mapping from target performance to candidate structures [14, 15]. Furthermore, cross scale modeling approaches, such as physical LassoLars interaction potential (PLIP)-based quantification of interface polarization effects, further effectively couple physical mechanisms to constraint and guide the generation of mesoscopic structures [16, 17].

The breakthroughs in the intelligent synthesis focus on achieving a balance between efficiency and sustainability [18–21]. Closed loop autonomous systems, such as microfluidic Bayesian optimization (BO) platforms, have compressed the quantum dot synthesis cycle to the minute level and achieved cross material system process transfer through transfer learning (TL) [22, 23]. Green and controllable preparation relies on multi-objective optimization algorithms (such as NSGA-II Pareto front) to synchronously improve yield, environmental compatibility, and scalability. For example, continuous flow synthesis has successfully produced kilogram-level antibacterial ZnO nanoparticles (NPs) [2, 24].

Advanced representation technologies are breaking through the limits of spatiotemporal resolution by intergrating ML [25, 26]. Real-time analysis of dynamic interface behavior (such as the "endpoint fusion" growth mechanism in concave nanocubes) has been enabled by *in-situ* electron microscopy combined deeply with ML (such as U-Net liquid-phase nanoparticle tracking) [27, 28]. Intelligent morphology analysis (such as spectral inversion engine gold nanorod (AuNR) spectral morphology analysis (SMA)) can directly extract geometric parameters from absorption spectra, replacing traditional time-consuming microscopic statistics [26, 29]. However, comprehensive three-dimensional (3D) reconstruction and dynamic evolution quantification of complex systems still depend on the integration of multimodal data (such as four-dimensional scaning transmission electron microscopy (4D-STEM) combined with Raman spectroscopy) [30, 31].

The innovation of performance prediction models lies in their ability to decode interpretable structure–activity relationships [32–36]. For example, catalytic descriptors (such as the d-band center and ΔG^*_{OOH} adsorption energy) are quantified to trace the origin of activity through SHapley Additive exPlanations (SHAP) analysis [37, 38]. Research on understanding biological interface mechanisms (such as GraphBNC framework) reveals the dynamic evolution of protein coronation and clarifies delivery pathways within the main conductor [39, 40]. For the prediction of extreme material properties, ML enables potential function simulation for negative thermal expansion materials, and crosses the scale gap of traditional simulations [31, 41]. Together, these breakthroughs provide trustworthy design foundations for application scenarios ranging from energy storage devices and targeted drug delivery to environmental remediation materials and lightweight structures. Nevertheless, cross-scale correlation mechanisms (such as atomic defect evolution and macroscopic fracture behavior) remain the core challenge [42, 43] (Fig. 1). This article aims to integrate innovations across the entire chain and promote the next generation paradigm shift in material research and development towards "physical mechanism-driven, cross-scale closed loop, and intelligent system autonomy".

2 New paradigm of intelligent design

The new paradigm of intelligent design marks a fundamental shift in material research and development from traditional "trial and error" approaches toward "goal-driven", with the core aim of constructing a reverse mapping of "performance–structure–process" [24, 28, 44–61]. This paradigm shift is developing rapidly along two mutually supportive main lines (Fig. 2): One is to build an efficient and robust reverse design system to directly generate the optimal structure and process parameters for target performance, and the other is to develop a cross scale simulation-driven computational framework to provide deep physical mechanism support and verification for reverse design, which jointly promote materials research and development into the intelligence era.

2.1 Reverse design system

The architecture of the reverse engineering system elucidates the deep coupling mechanism between "performance-oriented physical verification" and "virtual base innovation engine", highlighting the core paradigm shift of ML-driven material development chain. Within this framework, the reverse design system addresses the challenge of "how to accurately achieve predefined goals", whereas the generative model overcomes the limitation of "how to create unknown possibilities".

2.1.1 Performance-oriented structural synthesis

The performance-oriented structure generation field of reverse system design presents a development trend of multi-scale collaborative innovation. Lutz et al. [14] innovatively proposed a reinforcement learning-driven "top-down" protein design paradigm, employing Monte Carlo tree search to globally optimize protein conformation under complex geometric constraints (Fig. 3(a)). This approach successfully constructed ultra compressed icosahedral and disc-shaped nanopore structures not found in nature. Its capability to precisely locate spatial arrangements of surface functional domains opens up new paths for vaccine design. The global optimization concept was extended to the field of chiral optical materials in Dai et al.'s research (Fig. 3(b)) [62]. The constructed closed-loop ML framework established a quantitative mapping between synthesis parameters and circularly polarized luminescence performance (glum value of 0.15) through a decision tree algorithm, revealing the multivariate structure–activity relationship of solvent ions in synergistic modulation of assembly chirality, and realizing the reverse mapping from performance targets to synthesis parameters. Furthermore, Jiang et al. [63] developed a "data-driven evolution" paradigm by integrating

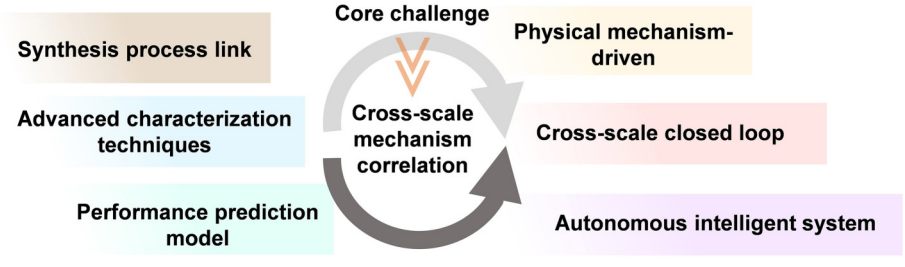

Figure 1 Overview of the full chain process and ML methods for synthesizing nanomaterials.

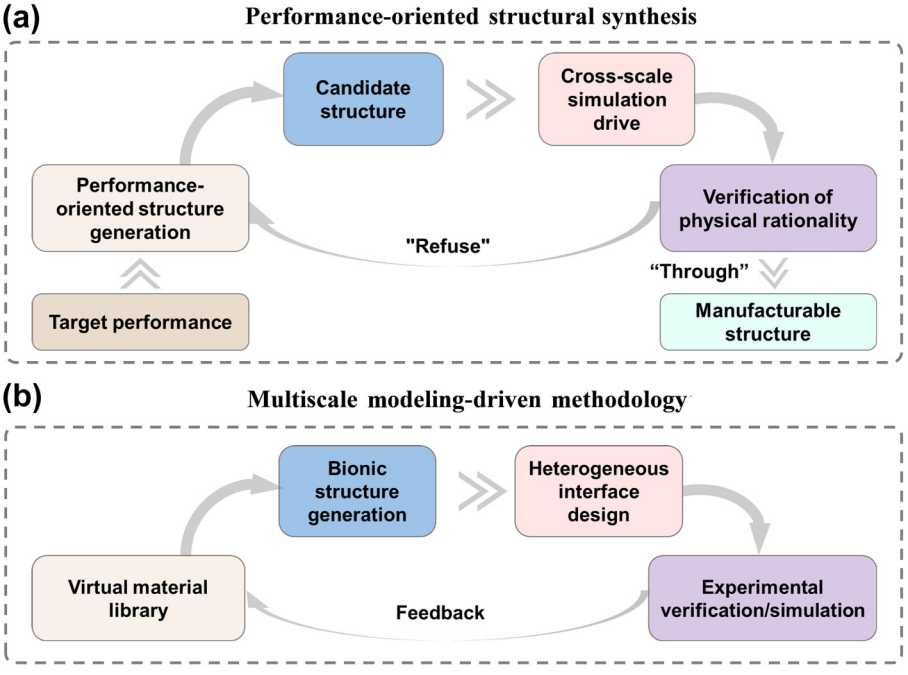

Figure 2 Design schematic diagram of reverse engineering system (performance-oriented and cross scale simulation-driven).

literature mining generated multi-enzyme genotype phenotype databases with quantum mechanics/molecular mechanics (QM/MM) path optimization algorithms (Fig. 3(c)). This enabled the reverse engineering of $MnCuCo_7O_{12}$ solid solution nanozymes, overcoming bottlenecks caused by multi-enzyme activity inhibition (increasing the turnover number K_{cat} by 3–8 times), demonstrating the advantages of evolutionary algorithms for cross scale reverse design.

The development of platforms, concepts, and models has greatly strengthened the complex reverse design process mentioned above. For example, the ViNAS Pro platform developed by Wang et al. [15] provides critical infrastructure by employing standardized PDB format nanostructure annotations, enabling machine-readable atomic level descriptor computation. By combining this with the AutoNanoML toolkit to construct a virtual nanolibrary, the platform significantly accelerates the closed-loop iteration of forward prediction of "structural properties" and backward generation of "performance structures". This platform-centered thinking approach has been deepened in Li et al.'s multi-objective security design study [64]. They established a Pareto matrix for the phototoxicity of sunscreen factors using a random forest (RF) model, enabling direct reverse generation of TiO_2 nanoparticle parameter combinations that satisfy sun protection factor (SPF) > 30 and reactive oxygen species (ROS) < 5% (such as (001) aspect ratio > 60%), quantifying the trade-off relationship between performance conflicts. Similarly, Cui et al. [65] leveraged the CatBoost model to achieve rapid reverse screening of two-dimensional (2D) material work functions (103 times faster than density functional theory (DFT) calculations). In combination with SHAP analysis, the structure performance mechanism of electronic structure of the spatial group ($Pmn2_1$) was analyzed, guiding the screening of high-performance organic light-emitting diode (OLED) anode materials with a transmittance > 90%. In the end, Kioumourtzoglou et al. [66] proposed the concept of "nanomaterials-as-a-service" (NaaS), achieving on-demand synthesis (batch difference < 5%) via a Bayesian optimization-driven microfluidic platform. This shortened the optimization cycle for silver nanoplate synthesis to 48 h, marking an engineering level transformation from reverse design toward intelligent manufacturing.

2.1.2 *Multi-scale modeling-driven methodology*

ML is driving the evolution of material simulation from a single scale approach to a multi-scale coupled paradigm, providing comprehensive physical support throughout the entire reverse design. For instance, PLIP developed by Salazar et al. [17] accurately quantifies the surface polarization effect in zinc oxide nanocrystals (NCs) at the atomic scale for the first time. This work revealed that interfacial energy dominates the nucleation of metastable body-centered tetragonal (BCT) phases under deep undercooling conditions (critical size < 50 atoms), providing key physical constraints for the generation of mesoscopic structures. However, analyzing scale structures experimentally still faces the challenge of incomplete diffraction information for nanocrystals. To addressing this bottleneck, Guo et al. [67] proposed the diffusion model PXRDnet, which directly reconstructs 10 Å scale atomic structures (coordinate error < 7%) from low signal-to-noise ratio diffraction patterns ($Q < 8.2$ Å^{-1}) using Langmuir dynamics (Fig. 3(d)). Its end-to-end generative mechanism effectively compensates for peak broadening effect induced by small dimensions. The integration of this model with Salazar's work forms a closed loop of "mechanism simulation structural reconstruction". The PLIP function validates the energy rationality of the generated structures, whereas PXRDnet produces experimentally validated structural output for the study of nucleation pathways.

When system complexity increases to disordered nanoclusters, the ML MotEx method developed by Anker et al. [68] quantifies the contribution weight of structural motifs within pair distribution function (PDF) data using SHAP values, enabling motif recognition of 256 atomic systems with a 1015 fold increase in computational efficiency. This pre-screening mechanism of this technology complements the probability generation of PXRDnet: The former

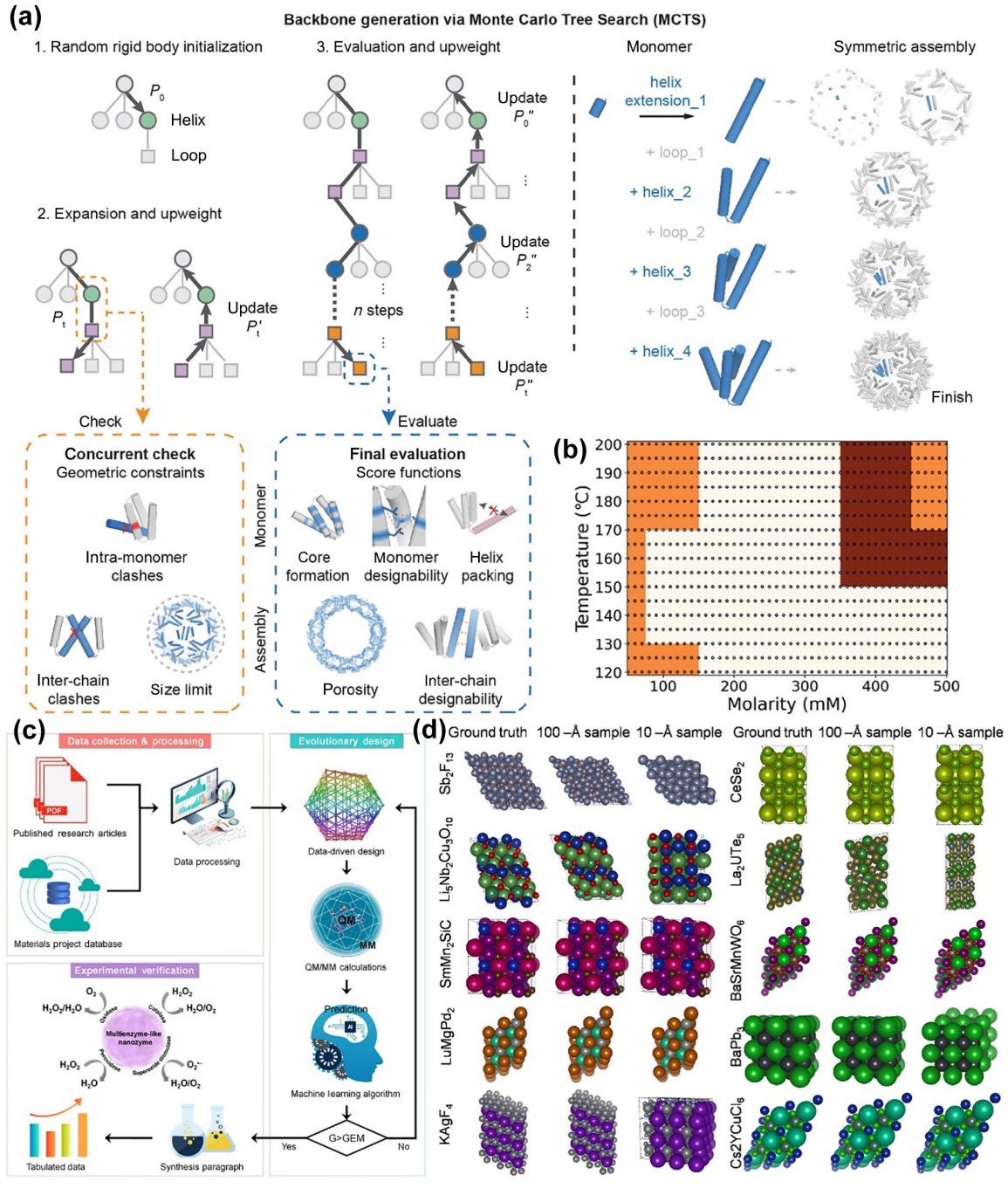

Figure 3 (a) Top-down design strategy and computational pipeline flowchart. Reproduced with permission from Ref. [14], © Lutz, I. D. et al. 2023. (b) Prediction decision distribution diagram. Reproduced with permission from Ref. [62], © Wiley-VCH GmbH 2023. (c) Data-driven evolutionary design process flowchart for multi-enzyme nanozymes. Reproduced with permission from Ref. [63], © American Chemical Society 2024. (d) PXRDnet structure prediction (the leftmost column shows the real crystal structure, whereas the other columns show the reconstructed crystal structure of PXRDnet from simulated nanocrystal powder X-ray diffraction (PXRD) modes with diameters of 10 and 100 Å). Reproduced with permission from Ref. [67], © Guo, G. B. et al. 2025.

provides important sampling guidance for the Langmuir dynamics of the latter, whereas the MP-20-PXRD dataset of PXRDnet can optimize the initial model construction of ML MotEx, jointly solving the challenge of uncertainty transfer across scales. At the mesoscale, Chen et al. [69] designed a physics-informed deep homogenization neural network (DHN) that solves the Gurtin–Murdoch equation incorporating surface energy through an adaptive weighting strategy, accurately predicting a 45% modulus attenuation from a 30 nm aperture. By explicitly encoding interfacial physics, the DHN extends Salazar's atomic level interface effect modeling to the continuous media, whereas its Fourier morphology descriptor provide a universal interface for Gupta's anisotropic system modeling.

For complex assembly systems, the computational approach for structure generation of anisotropic particles (CASGAP) algorithm developed by Gupta et al. [70] integrates the von Mises–Fisher distribution to enable probabilistic control over the orientation ordering of non-spherical particles (such as the optimal time

domain structure with f_{cut} = 0.5). The mapping of "processing parameters–microstructure" established in this work corresponds to Salazar's "undercooling–crystal selection" mechanism across scales, while the generated structures can be directly used as input boundary condition for the DHN to verify nanocomposite properties. However, current cross scale integration still faces three challenges: (1) The atomic reconstruction of PXRDnet requires the fusion of ML MotEx's dynamic motif tracking capability. (2) The random pore modeling in DHN requires coupling with the morphology probability generation provided by CASGAP. (3) The polarization effect of PLIP potential function should be embedded in the mesoscopic constitutive models. Addressing these key issues would enable a full chain simulation of "atomic nucleation mechanism cluster structure evolution mesoscopic performance response", thereby promoting the reverse design from empirical exploration to a physics-driven paradigm.

2.2 Generative design model

2.2.1 Construction of virtual material library

ML-driven virtual material libraries are overcoming the static limitations of traditional databases by establishing a "design validation" loop through standardized structural coding and cross scale validation mechanisms. The ViNAS Pro platform developed by Wang et al. [15] pioneered a machine-readable nanostructure annotation paradigm, converting 14 types of nanomaterials (such as gold nanoparticles and carbon nanotubes (CNTs)) into standardized PDB files (Fig. 4(a)). This atomic level structure encoding (including chemical bonds/coordinates) serves as a precise input for molecular dynamics simulations. Its key innovation include: (1) integrating SMILES ligand labeling to achieve surface modification digitization; (2) generating virtual nanolibraries and predicting multiple properties (such as biological activity) using the AutoNanoML modules; and (3) employing the visual molecular dynamics (VMD) visualization tool to facilitate structure performance correlation analysis. This platform aims to solve the core bottleneck of the lack of machine readability in nanomaterials, but the reliability of the virtual library depends on experimental verification. To tackle this, Choi et al. [71] developed a hydrogel particle analysis system.

The current technology integration needs to overcome three barriers: (1) The PDB annotation of ViNAS Pro needs to be expanded to multiphase soft material systems such as hydrogels. (2) Fluorescence coding rules for hydrogels should be integrated into the structure generation constraints of the virtual database. (3) The attention mechanism of YOLOv5 can optimize the visualization module of ViNAS Pro. This multi-level encoding system of "atomic coordinate micrometer pattern" will promote the evolution of virtual material libraries from a single scale to a cross scale digital twin paradigms, providing a self validating infrastructure for generative design and development. As discussed in the following section, the establishment of this cross scale digital twin infrastructure paves the way for the rational design of complex biomimetic and heterogeneous material systems.

2.2.2 Biomimetic/heterogeneous structure design

Based on the digital infrastructure of virtual material libraries, ML has further promoted the transformation of biomimetic structure design from experience-based imitation to mechanism-based prediction mode, enabling precise control of multi-scale heterogeneous interfaces [72, 73]. For example, the water lily inspired photothermal actuator developed by Jiang et al. [74] achieves real-time feedback of the action state of the resistance signal and combines it with a convolutional neural network to realize object recognition. This work reveals the structure–activity relationship between biomimetic structures and electrical signal response, however, its adaptability in complex environments remains constrained by the stability of heterogeneous interfaces. In response to this challenge, Abay et al. [75] proposed a hybrid scale electronic textile strategy (Fig. 4(b)). The macroscopic thermally bonded polyester layer provides mechanical stability, whereas the microscopic spun lace cellulose layer forms a fluffy porous sensing network. A 3D conductive heterojunction interface is construted by spraying MXene/graphene oxide silver nanowires (GO AgNWs), decoupling temperature/pressure signal via ML with > 98.9% accuracy. This work proves that the synergistic effect between 2D MXene sheets and one-dimensional (1D) AgNWs is the key to inhibit stacking and improve the sensing sensitivity (33 kPa^{-1}). Its hierarchical assembly logic and cascade drive of hydrogels establishes a scale bionic echo.

At the molecular scale heterojunction interface design level, Jeindl et al. [76] integrated DFT calculations with ML to analyze the nonintuitive assembly mechanism of functionalized molecules on Ag (111) surface (Fig. 4(c)). Their findings revealed that the size of the molecular skeleton determines the selection of assembly motif ($B_2O/A_2O/P_2O$) by regulating three antagonistic driving forces: the adsorbate substrate interaction (E_{ads}), the intermolecular interaction, (E_{int}), and the steric hindrance (E_{steric}). The quantitative model they established is

$$\Delta G = \alpha E_{ads} + \beta E_{int} + \gamma E_{steric} (\alpha, \beta, \gamma \in [0,1]) \quad (1)$$

This model provides molecular level design principles for enhancing interfacial strength between electronic textiles and hydrogels. Specifically, the anchoring mechanism of MXene on polyester fiber is similar to that of quinone molecules competing for adsorption sites on silver surface, whereas the interpenetrating molecular chain design of hydrogel avoids the steric effect.

The current integration effort needs to overcome three barriers (Fig. 5): (1) The molecular scale force balance model should be embedded within the optimization algorithm used for designing hydrogel cross-linking network. (2) The generation of porous structures in electronic textiles requires the incorporation of a photothermal driving strategy inspired by water lily actuators. (3) The synthetic data augmentation technology of YOLOv5 can improve the noise robustness of molecular assembly prediction. This cross scale biomimetic design framework of "molecule micro-nano macro" quantifies the mechanism of energy/information transmission in nature through ML, providing a universal paradigm for heterogeneous interface engineering for the next generation of intelligent materials.

3 Synthesis process optimization

ML-driven synthesis process optimization is shifing from discrete parameter tuning to a paradigm of full chain autonomy and sustainability [23, 77–86]. This section systematically analyzes the two core paradigms: closed-loop autonomous synthesis and green controllable backup, highlighting their breakthroughs in traditional R&D bottlenecks through cross scale closed-loop architecture and multi-objective collaborative mechanisms. The former enables an

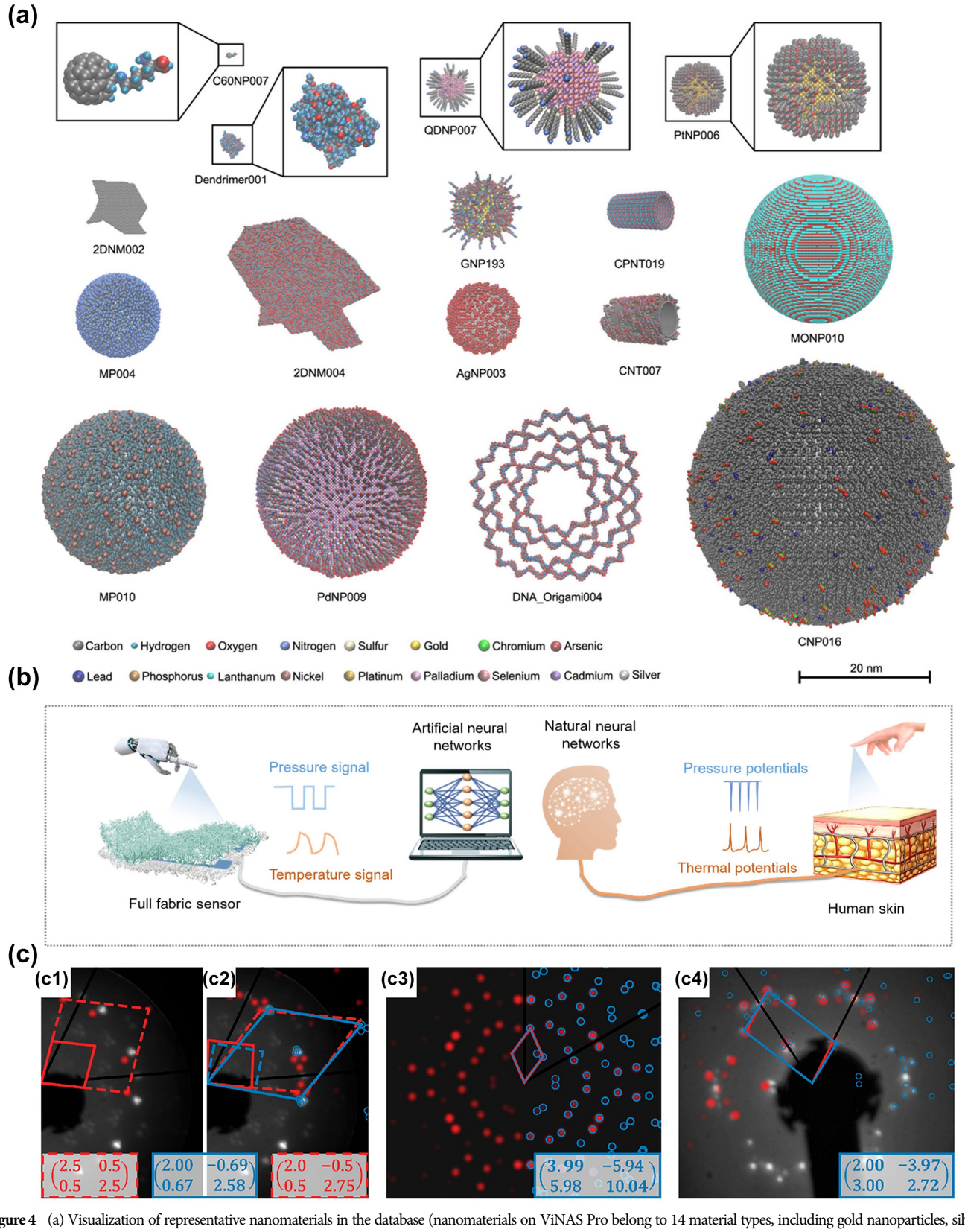

Figure 4 (a) Visualization of representative nanomaterials in the database (nanomaterials on ViNAS Pro belong to 14 material types, including gold nanoparticles, silver nanoparticles, platinum nanoparticles, palladium nanoparticles, Buckminster fullerenes, carbon nanoparticles, carbon nanotubes, dendritic macromolecules, DNA origami nanoparticles, metal oxide nanoparticles, quantum dots, cyclic peptide nanotubes, two-dimensional nanomaterials, and microplastics). Reproduced with permission from Ref. [15], © Wang, T. et al. 2024. (b) Schematic diagram of temperature and pressure detection by biological skin sensing system and artificial sensing system. Reproduced with permission from Ref. [75], © Elsevier Ltd. 2025. (c) Comparison of theoretical low energy electron diffraction (LEED) patterns (red) obtained through kinematic diffraction theory with STM images (blue) and FFT fitting of LEED images. Reproduced with permission from Ref. [76], © Jeindl, A. et al. 2021.

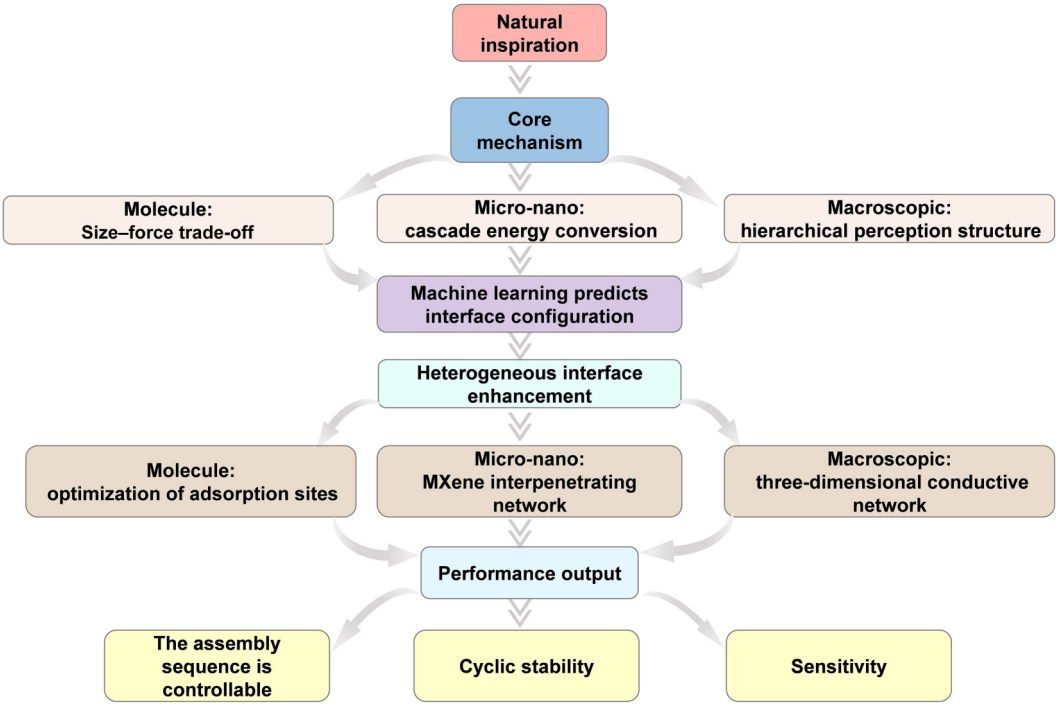

Figure 5 A schematic diagram of the cross scale biomimetic design framework from molecule to micro-nano and then to macro, and the process of quantifying the energy/information transmission mechanism in nature through ML.

intelligent upgrade from molecular mechanism analysis to instruction-driven manufacturing, whereas the latter constructs new standards for green synthesis under the triple constraints of performance and environment economy, jointly accelerating the evolution of material manufacturing into an intelligent ecosystem of "on-demand design, autonomous execution, and sustainable output".

3.1 Closed loop autonomous synthesis

3.1.1 Robot experimental system

Robot experimental systems reconstruct the material synthesis paradigm through a "perception decision execution" closed loop, with their technological breakthrough reflected in cross scale mapping and advances in universality ability.

In the field of material intelligence research and development, robot experimental systems are driving a paradigm shift in synthesis by multidimensional technological innovation. For example, the "artificial chemist" system developed by Epps's team was the first to construct a closed-loop architecture of "microfluidic reactor *in-situ* spectral monitoring Bayesian optimization" [22], achieving multi-objective collaborative optimization of quantum dot bandgap, quantum yield, and emission linewidth. Its knowledge transfer mechanism doubled the optimization efficiency for new precursor system and significantly reduced batch differences (Fig. 6(a)). The core breakthrough of this system lies in establishing a cross scale mapping relationship between synthesis parameters, structural features, and optoelectronic performance. By using Bayesian optimization to analyze parameter spaces with more than 7 dimensions, the negative correlation contradiction between photoluminescence quantum yield (PLQY) and low polydispersity (E_{FWHM}) is synchronously regulated, compressing the single experimental period to 2 min. However, its adaptability to more complex nanosystems such as core–shell structures remains constrained by the precision of microscopic characterization methods.

To address the bottleneck of material universality, Fanourgakis's team proposed a new paradigm based on atomic type descriptors [87], which replaces traditional structural descriptors with chemical essential features (such as hybrid states), allowing ML models to migrate across different material systems (such as metal-organic framework (MOF) to covalent organic framework (COF)) (Fig. 6(b)). This method reduces the data requirement for gas adsorption prediction to 1/10 of that needed by traditional approaches. Its universal design complements the knowledge transfer mechanism of "artificial chemists". The former provides a theoretical basis for cross material characterization, whereas the latter achieves closed-loop optimization of synthesis processes.

In the dimension of stability optimization, Jing et al. [88] achieved a key breakthrough through ML-guided ligand engineering (Fig. 6(c)). Based on Bayesian optimization, they precisely identified the optimal synthesis window for the zwitterionic ligand ASC-16 within two rounds of iteration. This enabled the realization of a PLQY of 192% for ytterbium-doped perovskite quantum cutting materials, while simultaneously maintaining extreme environmental stability. This work deepens the structure–activity relationship down to the molecular scale (sulfonic acid coordination strength and lattice stability), and its ultra efficient optimization strategy further validates the value of closed-loop systems in addressing the "stability performance" trade-off.

Finally, Li et al. [89] integrated the above technology framework into a full stack intelligent operating system (material acceleration operation system (MAOS)), constructing a "virtual reality (VR) training collaborative robot reinforcement learning" trinity architecture to achieve on-demand synthesis of quantum dots (Fig. 6(d)). Within this system, reinforcement learning is used to invert the theory of nucleation dynamics, shortening the time of a

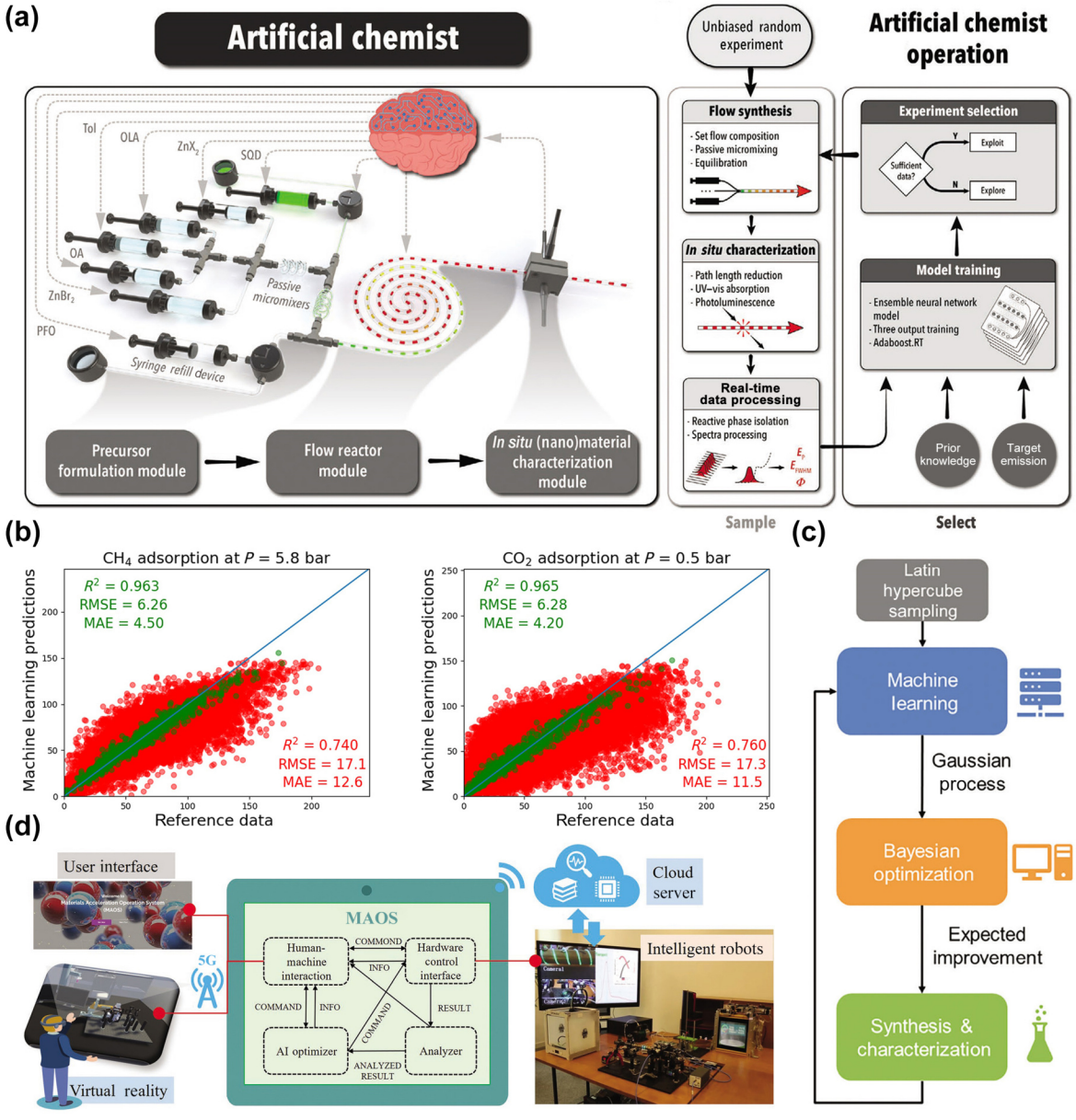

Figure 6 (a) Schematic diagram of an intelligent modular fluid microprocessor developed for autonomous synthesis path discovery and optimization of colloidal quantum dots, along with a detailed process flowchart illustrating its operation. Reproduced with permission form Ref. [22], © WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim 2020. (b) Application examples of RF in CH_4 at P = 5.8 bar and CO_2 at P = 0.5 bar. Reproduced with permission from Ref. [87], © American Chemical Society 2020. (c) Optimizing the ML and BO active learning workflow of PLQY for NCs, starting from LHS initial sampling, and then conducting model training to achieve BO sample extraction. Reproduced with permission from Ref. [88], © Wiley-VCH GmbH 2024. (d) Schematic diagram of MAOS and its integrated modules, including human–computer interaction, hardware control interface, analyzer, and AI optimizer. Reproduced with permission from Ref. [89], © Li, J. G. et al. 2020.

single experiment to 1/20 of that of traditional methods, marking the transition from closed-loop autonomous synthesis to instruction-driven manufacturing. However, the current challenges remain, including achieving cross system transferability of multi-component synthesis strategies, integrating high-temperature and high-pressure platforms, and precisely modeling complex material degradation mechanisms, all of which present key directions that be urgently addressed by the next generation smart research and development systems.

3.1.2 *Microfluidic real-time decision-making*

In closed-loop autonomous synthesis systems, the deep integration of microfluidic technology and ML is promoting real-time decision-making capabilities towards molecular mechanism analysis. Pan et al. [90] developed a machine learning framework combining artificial neural networks (ANN) and partial least-squares regression (PLSR) to tackle the core challenge of non-characteristic spectral analysis during the synthesis of platinum nanoparticles. By dynamically correlating subtle changes in ultraviolet (UV) spectra with reaction processes, the framework enables real-time monitoring of Pt concentration in ionic liquid solvent systems with an error rate less than 4%. This work reveals for the first time the disruptive influence of solvent cation fine-tuning (1-butyl-1-methylpyrrolidinium triflate (BMPYRR-OTf) and 1-butyl-2-methylpyridinium triflate (BMPY-OTf)) on reaction pathways. Pyrrolidine cations achieve a high yield of 94% through ordered

assembly pathways. Pyridine cations trigger disordered side reactions, causing the yield to drop sharply to 10%. This discovery combines the unique advantages of convex microfluidic platforms and ML in capturing the kinetics of solvent effects, converting transient reaction order changes that are difficult to capture with traditional offline characterization into quantifiable decision parameters.

Landry's team expanded the dimension of real-time decision-making from the perspective of biological nano interfaces [91]. They developed a random forest model that accurately predicts the adsorption tendency of proteins on carbon nanotubes, achieving an accuracy of 78% by decoding implicit structural flexibility features in protein sequences, such as the solvent exposed glycine content and the proportion of non-secondary structural domains. This model not only replaces high cost mass spectrometry experiments for screening, but also clarifies the structure–activity relationship mechanism. Flexible proteins match the curvature of carbon nanotubes through adaptive deformation, whereas rigid structures dominated by beta folding inhibit adsorption. In line with the methodology demonstrated in Pan's work [90], both studies use ML to solve the challenge of "featureless response". The former analyzed the synthesis process of nanoparticles without plasma resonance peaks, and the latter predicted the adsorption behavior of proteins without clear optical labels. The two studies jointly constructed a closed-loop decision chain of "real-time monitoring mechanism inversion decision optimization".

The current challenge lies in the coupled modeling of multi-scale dynamic processes. Firstly, the instantaneous impact of solvent molecule dynamic arrangement on the nucleation energy barrier in Pt synthesis has not yet been incorporated into real-time decision-making. Secondly, the conformational relaxation dynamics that occur during protein adsorption need to be combined with molecular simulations to further improve prediction accuracy. Moving forward, it is necessary to develop ML architectures that span physical domains and cansynchronously integrate multimodal data streams ranging from chemical kinetics to biomolecule behavior in microfluidic systems. Such advances will enable closed-loop autonomous decision-making, thereby bridging up the gap from nano synthesis to biological interface, engineering and driving the development of a new generation of smart manufacturing systems.

3.2 Green controllable backup

The core of green controllable synthesis is to achieve a Pareto frontier breakthrough of "performance environment economy" empowered by ML. The innovation covers macro production optimization and micro precision regulation.

3.2.1 Multi-objective optimization of processes

In the field of green and controllable material preparation, ML is balancing performance, sustainability, and scalability through multi-objective collaborative optimization mechanisms. Jose et al. [92] pioneered an integrated platform combing a high shear microreactor and a multi-objective algorithm (Thompson sampling efficient multi-objective (TSEMO) algorithm) high-throughput analysis integrated platform, achieving kilogram level continuous synthesis of antibacterial ZnO nanoparticles (spatiotemporal yield of 62.4 $kg{\cdot}m^{-3}{\cdot}day^{-1}$) (Fig. 7(a)). This work optimized both yield and antibacterial activity (comparable to commercial products) with fewer than 100 experiments, revealing that the morphological functional structure–activity relationship of nano star/rod structures that synergistically improve antibacterial efficiency via increased surface area and anisotropy. The breakthrough lies in incorporating green chemistry principles (low solvent consumption) and engineering scaling requirements (numbering strategy) into the optimization objectives directly, but it requires combining the steep Pareto front in high-yield areas with crystal growth simulations to improve accuracy.

Mekki-Berrada et al. [93] proposed a two-step optimization framework to address micro precision control challenges (Fig. 7(b)). In this approach, BO quickly locks the target plasma spectrum (such as the peak of 520 nm), whereas deep neural network (DNN) inversion reveals the nonlinear impact of sodium citrate concentration on peak shift (with a contribution rate of 68%). This strategy achieved spectral matching within 120 experiments, demostrating an efficiency 10 times higher than traditional screening. Its human–machine collaborative design mode (visual BO path and DNN feature weight) provides a new paradigm for on-demand optical material development. Complementarity with Jose's macro production optimization, this work highlights advances in customization of micro optical properties.

Building on this direction, Wahl's team further advanced multi-objective optimization towards high-dimensional complex systems (Fig. 7(c)) [94]. Their closed-loop system driven by Bayesian optimization precisely locates the dual phase interface in an octahedral alloy system (Au, Ag, Cu, Co, Ni, Pd, Sn, and Pt), successfully synthesizing a six phase nanoheterostructure (diameter < 50 nm) with significantly enhanced interface stability when lattice mismatch < 5%. This study overcomes the limitations of chemical intuition and verifies the synergistic ability of the algorithm for multiple objectives of composition structure stability with a success rate of 18/19 iterations. Nonetheless, predicting metastable phase dynamics and scaling up production to kilogram remain key challenges for practical deployment.

In response to environmental governance needs, Al-Gheethi et al. [95] developed a green synthesis strategy optimized by an adaptive neuro-fuzzy inference system (ANFIS). Using watermelon peel extract as a sustainable precursor, they synthesized defective ZnO nanoparticles (eco-green (EG)-ZnO NPs), and the adaptive neural fuzzy system (R^2 = 0.9967) was used to optimize the photocatalytic degradation conditions of metronidazole (170 $mg{\cdot}L^{-1}$, pH = 5), achieving a removal rate of 78.14% and toxicity elimination (EC50 increased by 5 times). Its core contribution lies in the establishment of a quantitative chain from defect density (photoluminescence (PL) peak intensity ratio) to ·OH yield and degradation efficiency, but the interference of actual water anions requires the model to enhance environmental robustness.

Shao et al. [96] focused on resource recycling by utilizing waste glass powder (WGP) as a carrier for the *in-situ* growth of CNTs via microwave pyrolysis. They innovatively applied image enhanced clustering algorithm to quantitatively evaluate dispersion (clustering index < 0.15). Through the ML, they established a correlation among growth parameters, dispersion quality, and the resulting structure–activity relationship, which reduced the electrical resistivity of conductive cement by 80%. This work highlights the dual potential for "waste value-added performance improvement", along side Al-Gheethi's research, that combining green precursor selection with process optimization algorithms can systematically reduce the environmental footprint of the preparation process.

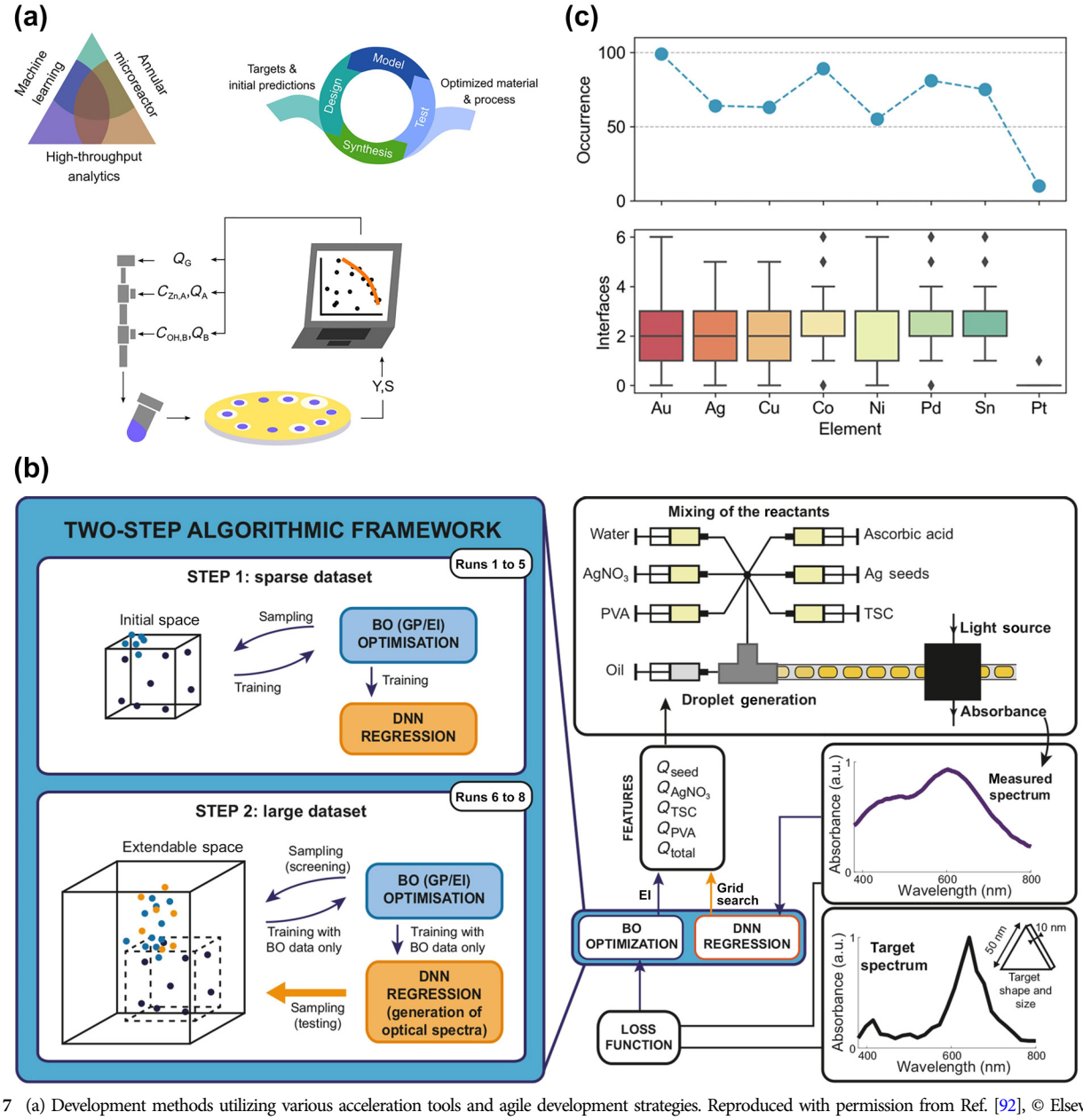

Figure 7 (a) Development methods utilizing various acceleration tools and agile development strategies. Reproduced with permission from Ref. [92], © Elsevier B.V. 2021. (b) Algorithm framework for high-throughput experimental circuits. Reproduced with permission from Ref. [93], © Mekki-Berrada, F. et al. 2021. (c) The performance results of the single interface particle collection algorithm were simulated using existing datasets in two different scenarios with different batch sizes, and compared with a randomly collected baseline, as well as the interface count diagrams corresponding to each element. Reproduced with permission from Ref. [94], © Wahl, C. B. et al. 2021.

3.2.2 *Small sample synthesis strategy*

In the field of green controllable backup, ML is breaking through the constraints of small samples through innovative algorithms to achieve precise design and risk prediction of material synthesis. Mahmood et al. [97] developed a multiple imputation using chained equations (MICE) to solve the micro control problem of the morphology control of the silicon shell of gold nanorods, and constructed a highly robust decision tree model using only 30 sets of experimental data. This work innovatively established a quantitative relationship between the localized surface plasmon resonance/transmission surface plasmon resonance (LSPR/TSPR) spectral shift (peak width difference > 38 nm) and SiO_2 coating morphology (leaf-like/uniform), and achieved accurate synthesis of the target structures through a reverse design workflow. The core breakthrough is to convert the *in-situ* spectral feedback into morphology decision basis, but the complexity of the process requires the algorithm to further integrate prior knowledge of reaction kinetics to compress the optimization cycle.

Cai's team proposes a TL-driven cross domain knowledge reuse framework for porous material screening scenarios [98]. By sharing geometric/energy descriptors (such as pore size distribution of 4–6 Å and equivalent adsorption heat) between the MOF database (over 50,000 structures) and the COF system (over 7000 structures), the Xe/Kr separation performance ($R^2 > 0.85$) can be predicted with only 300 COF samples, reducing the data requirements by 95% compared with traditional methods. This strategy aligns the feature

space through the maximum mean difference (MMD) domain adaptation technique, revealing that the adsorption enthalpy difference (SHAP value > 0.7) is the selective dominant factor. Its descriptor generalization mechanism complements Mahmood's research. The former reuses cross material knowledge, whereas the latter mines hidden features of a single system.

Chatterjee et al. [99] further expanded the small sample paradigm to the field of nanosafety assessment. The quantitative cross reading algorithm they developed is based on Euclidean distance and kernel function similarity measurement (Gaussian kernel/Laplacian kernel). It only requires ≤ 20 sets of toxicity data to predict the biological effects of metal oxide nanoparticles ($Q^2_{ext\_F2}$ > 0.85). By optimizing the distance threshold (0.4–0.5) and feature matching rules (such as giving priority to the correlation between zeta potential/solubility), the algorithm automatically selects weighted prediction values of 2–5 of structural analogues, promoting the risk assessment from experience-driven to computation-driven priority sorting. This work and the aforementioned research joingly constructed a closed loop of "data efficiency improvement knowledge transfer risk avoidance", highlighting the universal value of the small sample strategy in material lifecycle management.

4 Advanced characterization analysis

ML is driving a paradigm shift in material characterization from static observation to dynamic mechanism analysis, with the core being to break through the cross scale correlation bottleneck of "microscopic behavior–mesoscopic morphology–macroscopic performance" [100–108]. This section systematically elaborates on the two technical pillars of cross scale dynamic monitoring and intelligent morphology analysis (Fig. 8). Cross scale dynamic monitoring achieves quantitative decoding of interface dynamics through *in-situ* multidimensional data fusion, and intelligent morphology analysis constructs a quantitative mapping of morphology and performance through intelligent inversion of physical constraints, jointly promoting the theoretical upgrade of material characterization from empirical description to mechanism prediction, and providing underlying support for intelligent research and development closed-loop.

4.1 Cross-scale dynamic monitoring

4.1.1 In-situ electron microscopy technique

The deep intergration of *in-situ* characterization with ML is breaking through traditional spatiotemporal resolution limits, enabling multi-scale dynamic analysis from nanosecond level interfacial dynamics to complex environmental behaviors. Chen et al. [109] tackled the critical challenge of understanding carbon based catalyst deactivation mechanism by combining *in-situ* characterization with ML quantification, revealing for the first time the essence of the self inhibitory effect of polymer deposition (Fig. 9(a)). Their study showed that high degree of polymerization (DP) products are continuously generated at a constant energy barrier, with their DP values increasing due to enhanced hydrophobic interactions and van der Waals forces (ML quantification contribution > 85%). This results in tight polymer adhesion to the catalyst surface, competitive blocking of active sites, and impeded electron transfer. This work innovatively establishes a cross scale structure–activity relationship of "DP–interfacial force strength–adhesion stability", providing a theoretical foundation for regeneration strategy design. However, real-time monitoring of the *in-situ* transformation process of polymers is still limited by the compatibility of electron microscopy in liquid phase environment.

Yao's team focused on the liquid-phase dynamics of nanoparticles and developed the U-Net intelligent analytical framework to overcome traditional characterization bottlenecks (Fig. 9(b)) [110]. By using generative adversarial networks (GANs) to synthesize training data, this approach achieves precise segmentation of nanoparticles in liquid phase transmission electron microscopy (TEM) videos with signal-to-noise ratio of < 2 (an accuracy of 95%), and improving efficiency by a factor of 100 compared to manual analysis. This technology captures for the first time the "endpoint fusion chain growth" mechanism of concave nanocubes (activation energy of 12 kT), and quantifies the exponential decay relationship between curvature radius and etching rate (R^2 = 0.96). Its high-throughput dynamic tracking capability forms a technological synergy with Chen's research [110], with the former analyzing the solid–liquid interface deactivation process and the latter uncovering the self-assembly mechanisms at liquid–liquid interface, jointly constructing a cross scale interfacial dynamics database.

Landry et al. [91] further extended dynamic monitoring into the dimension of predicting biological molecular behavior (Fig. 9(c)). Their random forest model accurately predicts the adsorption tendency of flexible proteins on carbon nanotubes (with an accuracy rate of 78%) by decoding protein sequence features such as solvent exposure glycine content and the proportion of non-secondary domain, thereby elucidating the universal mechanism of "structural adaptability-driven interface adsorption". Although *in-situ* electron microscopy was not directly used, this work employed ML to invert dynamic conformational changes in biomolecules, overcoming the limitations of *in-situ* techniques for observing non-rigid molecules, such as the difficulty in capturing protein deformation. This approach directly echos the methodological challenges addressed in Yao's research. Both

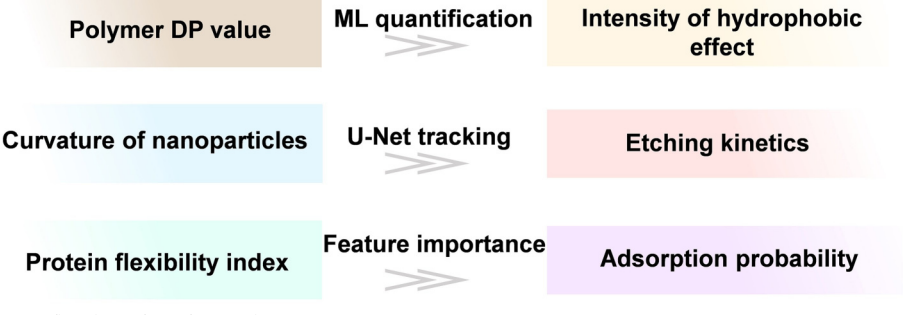

Figure 8 Quantitative schematic flowchart of interface mechanism.

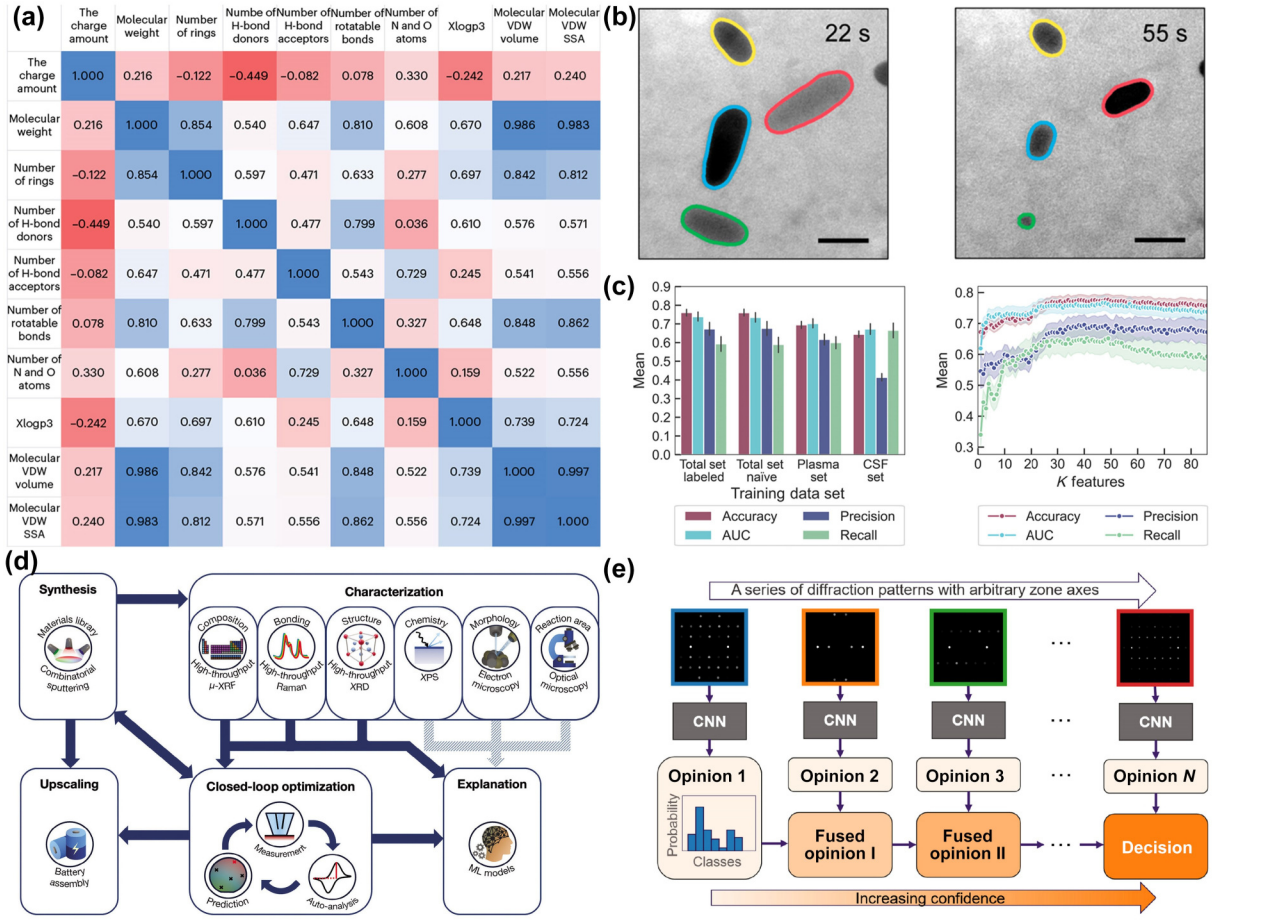

Figure 9 (a) ML assisted sedimentation mechanism analysis of relevant heat maps with different parameters. Reproduced with permission from Ref. [109], © Duan, P. J. et al. 2025. (b) Delayed liquid phase TEM images of nanorod boundary coverage predicted using a trained U-Net model (each color representing a different nanorod). Reproduced with permission from Ref. [110], © American Chemical Society 2020. (c) The performance results of the classifier are tested on different biological fluid training datasets and different protein feature inputs. Reproduced with permission from Ref. [91], © Ouassil, N. et al. 2022. (d) An interpretable ML model. Reproduced with permission from Ref. [111], © Sanin, A. et al. 2025. (e) A framework for automatically identifying crystal structures from any DPs with axes. Reproduced with permission from Ref. [112], © Chen, J. et al. 2023.

methods compensate for the technical blind spots of *in-situ* observations through computational simulation and algorithmic analysis. The specific shared challenges and potential pathways for breakthrough are summarized in Table 1.

4.1.2 *Multidimensional data fusion*

In the material intelligence research and development system, multidimensional data fusion cross scale for dynamic monitoring has become a key path for analyzing complex structure–activity relationships. Sanin et al. [111] demonstrated dynamic analysis of ternary thin film anodes (Si-Ge-Sn) across temporal and spatial scales by constructing a three-in-one platform of "autonomous electrochemical exploration multimodal characterization ML closed-loop optimization" (Fig. 9(d)). Their integrated scanning droplet cell (SDC) and sequence learning algorithm overcame the temporal and spatial constraints of traditional high-throughput experiments, enabling real-time capture of electrochemical responses under compositional gradients. By further integrating Raman (short-range bonding order), X-ray diffraction (XRD, long-range crystal evolution), and X-ray photoelectron spectroscopy (XPS, surface chemistry) data, they constructed an interpretable ML model that quantitatively revealed the synergistic effects of Ge/Sn components on ion migration energy barriers and volume expansion (such as Si Ge bonding enhancing ion diffusion and Sn suppressing lattice distortion). Of particular importance, this work successfully bridged microscale thin film screening with practical device validation through an ML-driven performance transfer model (error < 8%), allowing the optimized component ($Si_{60}Ge_{25}Sn_{15}$) to achieve a specific capacity five times that of graphite and an increase of 30% in energy density in coin cells.

Table 1 Common challenges and breakthrough paths of electron microscopy technology

Challenge dimension	Collaborative solution	Case support
Spatio-temporal resolution gap	GAN generates cross-scale training data	Liquid-phase nanoparticle collision simulation (Yao)
Ambiguity of interface function	SHAP/feature importance analysis dominant factor	DP value → adhesion force glycine content → adsorption capacity
Complex environmental interference	Domain adaptation enhances the robustness of the model	Prediction of environmental adsorption in plasma/cerebrospinal fluid

This multidimensional fusion paradigm has been further expanded in the field of structural representation. Chen et al. [112] developed a multiview opinion fusion ML (MVOF-ML) framework to tackle the problem of random orientation in electron diffraction patterns (DPs) (Fig. 9(e)). By utilizing convolutional networks with physical constraints to extract radial/angular symmetric features of Bragg spots, and through deep learning, the uncertainty of single view prediction can be quantified. By integrating multi-angle DP data, the model maintains a crystal system classification accuracy of 0.94 under strong noise interference, significantly improving the reliability of high-throughput 4D-STEM analysis. Notably, the commonality with Sanin's research lies in their shared use of physical information ML to embed domain knowledge and utilize uncertainty quantification techniques to achieve dynamic monitoring closed-loop [111]. Together, they underscore the central value of multidimensional data fusion in bridging cross scale correlation, from Sanin's dynamic transmission of "atomic bonding–crystal evolution–battery performance" to Chen's spatial integration of "single view diffraction–multi view consensus–crystal system classification".

However, current technologies still face common bottlenecks in cross scale modeling. Sanin's work reveals that the influence of surface oxides (XPS data) on cycling stability has yet to be fully incorporated into ML models, whereas Chen's framework needs to be further extended to deeper structural analysis, such as space group identification. In the future, it is necessary to develop adaptive fusion architectures that incorporate interface evolution under dynamic conditions (such as electrode electrolyte interface) and defect dynamics into a unified computational framework to overcome the full chain decoding problem of "micro mechanism–mesoscopic behavior–macro performance".

4.2 Intelligent shape analysis

4.2.1 Image recognition quantification

The intelligent morphology analysis of advanced material characterization is undergoing a paradigm shift from passive observation to active deconstruction, with the core being image quantification technology that integrates deep learning and physical constraints. Cho et al. [113] constructed a high-throughput deep statistical characterization framework that enables cross scale decoding of the growth kinetics of Co_3O_4 nanocrystals (Fig. 10(a)). Using convolutional neural network (CNN) segmentation of 441,067 high-resolution TEM (HRTEM) images, they quantified size/shape evolution with sub nanometer accuracy, and for the first time, identified the critical "starting radius" (~ 5 nm) that triggers the transition from thermodynamic control (Wulff polyhedra) to dynamic control (concave structure). This work not only establishes a quantitative structure–activity relationship of "size morphology synthesis conditions", but also reveals the defect mediated

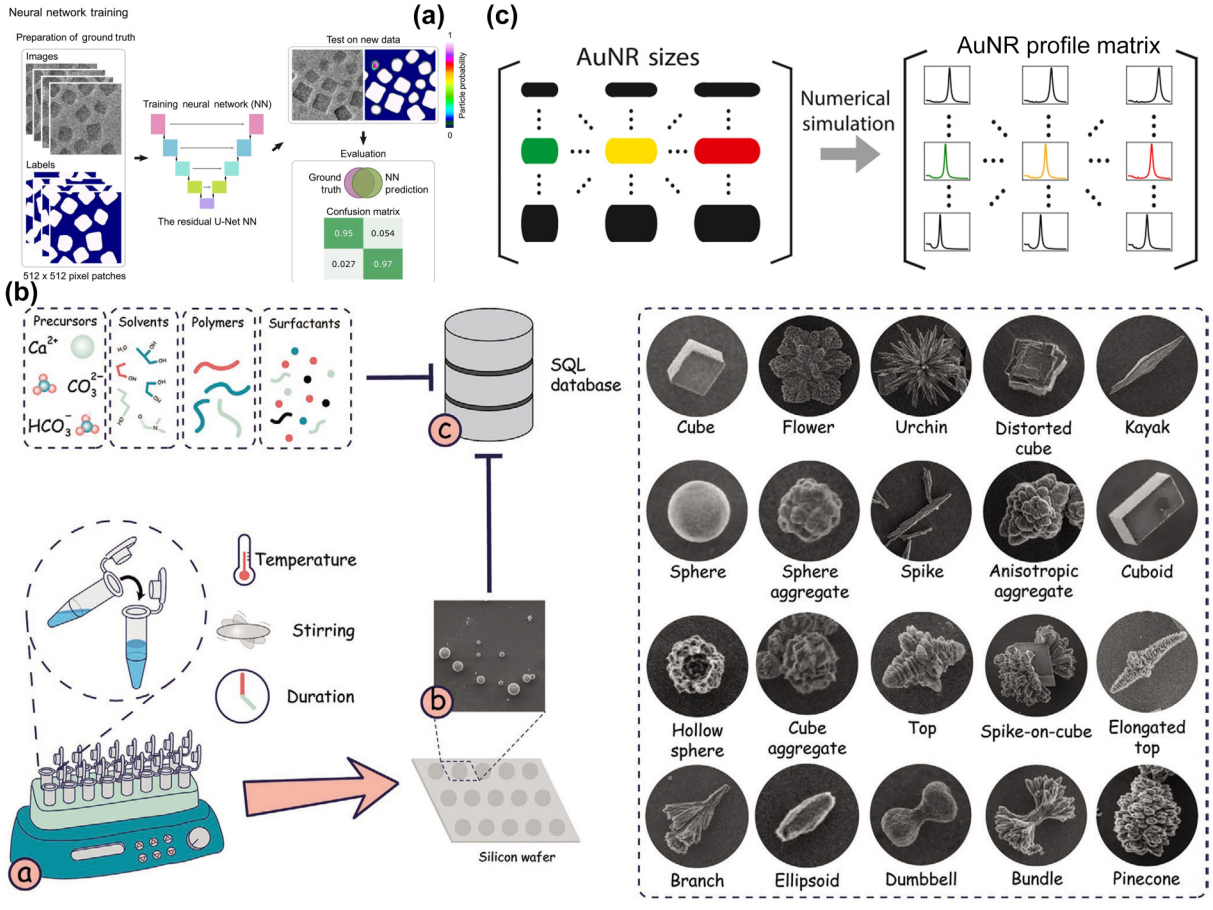

Figure 10 (a) Overview of CNN training, testing, and evaluation process. Reproduced with permission from Ref. [113], © Cho, M. G. et al. 2024. (b) Random high-throughput multi-parameter synthesis and database development of calcium carbonate-based nanomaterials. Reproduced with permission from Ref. [105], © Wiley-VCH GmbH 2022. (c) Showing how to numerically simulate a 2D matrix of AuNR size and how to enter a matrix of the same size for SMA contour display. Reproduced with permission from Ref. [102], © Gleason, S. P. et al. 2024.

mechanism for growth mode switching, providing a theoretical basis for morphology-controlled synthesis.

To break through the semantic limitations of traditional morphological descriptions, Serov et al. [105] innovatively proposed a reverse morphological search paradigm, using transfer learning (VGG16) to encode hand drawn sketches into morphological descriptors (Fig. 10(b)). They extracted fractal dimensions (1.2–1.8 quantifying calcium carbonate branching degree) through Canny edge detection to achieve scanning electron microscopy (SEM)/TEM image reverse retrieval with > 90% similarity. This technology closed the loop of "morphology design and synthesis verification", significantly enhancing the controllability of complex multi-level structures (such as flower shaped $CaCO_3$). However, it still faces inherent challenges related to 3D projection distortion.

In response to the quantization bottleneck posed by low signal-to-noise ratio microscopy images, Sun et al. [107] developed a lightweight real-time segmentation architecture, NSNet, which integrates a channel attention (dual attention module (DAM)) and a multi-scale perception module (light-weight atrous spatial pyramid pooling (LASPP)). This framework achieves a segmentation accuracy of 86.2% (11 frames/s) in dense packing/high noise samples, and combines a watershed algorithm to extract Blaschke shape coefficients. Its core value is in improving manual statistics efficiency by a hundred times, but the recognition of < 10 nm particles still requires further optimization. To complement this, Mahjoubi et al. [106] focused on 2D material thickness analysis. They applied adaptive gamma correction to eliminate overexposure in optical microscopy images and designed layered CNN that fuses texture/semantic features. Additionally, they improved minority class recognition through weak learning strategies (7–10 layer graphene mIoU reached 59%), providing a key quality inspection tool for thickness electrical performance correlation in wafer level manufacturing.

In the future, it will be necessary to develop dynamic morphology tracking networks that coupled with *in-situ* electron microscopy with multiphysics field simulations, enabling full chain decoding from nanocrystal growth mechanism switching to macroscopic material properties, such as graphene thickness effect.

4.2.2 *Spectral inversion technique*

Through the 3D integration of physical principles algorithm analysis and innovation application closed loop, the transformative role of intelligent spectral inversion technology in multi-scale morphological analysis of nanomaterials is revealed, providing a theoretical support for the evolution of ML-driven representation paradigms (Fig. 11).

Intelligent spectral characterization is transforming nanomaterial morphology analysis from indirect inference to direct deconstruction, with its core being the integration of physical models and deep learning inversion techniques. Gleason et al. [102] developed a spectral morphology inversion physics engine (AuNR SMA) to address the challenge of quantifying the geometric morphology of AuNRs (Fig. 10(c)). By establishing a strict mapping between the peak/line width and size distribution of longitudinal LSPR, the length (mean ± standard deviation: 45.2 ± 4.8 nm) and aspect ratio (3.2 ± 0.3) were directly extracted from the absorption spectrum, achieving an accuracy comparable to TEM statistics (error < 5%). This model is integrated into an autonomous synthesis platform, forming a closed loop of "spectral inversion–data generation–ML reverse design". The size dataset obtained from inversion is used to train a random forest model, enabling precise control of reaction parameters (such as ascorbic acid concentration) to achieve target morphology (R^2 = 0.96), and mining size spectral correlations not reported in historical literature to expand the synthesis knowledge base.

When the research object is extended to the sub-nanometer scale, the traditional spectral structural correlation becomes blurred due to quantum effects. Chen et al. [57] made a breakthrough in decoding the "featureless" spectra of gold nanoclusters (AuNCs) using deep convolutional neural networks. By using a 1D CNN architecture to learn the electronic transition fingerprints hidden in the UV–visible (UV–vis) absorption spectra (such as the quantitative relationship between 300 and 500 nm shoulder peak and the number of core atoms), the accurate molecular formulas (such as $Au_{15}(SR)_{13}$, mean absolute error (MAE) = 0.0053) can be directly predicted from broad peak. This model further realizes the reverse generation of chemical composition spectra. The predicted spectrum of the target molecular formula $Au_{25}(SG)_{18}$ is consistent with the experimental results, with R^2 > 0.98, providing high-throughput screening tools for nanocluster systems that are incompatible with mass spectrometry.

In the future, it is necessary to develop graph neural networks with multiple physical constraints, integrate LSPR theory (long-range electromagnetic response) and quantum chemical computation (short-range electron correlation) into a unified architecture, and realize full spectrum intelligent inversion from nanorod size effects to nanocluster quantum effects.

5 Performance prediction and validation

The core of the ML-driven performance prediction and verification paradigm for intelligent research and development of reconstructed materials lies in establishing a multi-scale and computable mapping of "atom mesoscopic system", overcoming the limitations of traditional empirical trial and error [114–123]. This section reveals how dynamic descriptors evolve across scales by decoding structure–activity relationships, and combines extreme

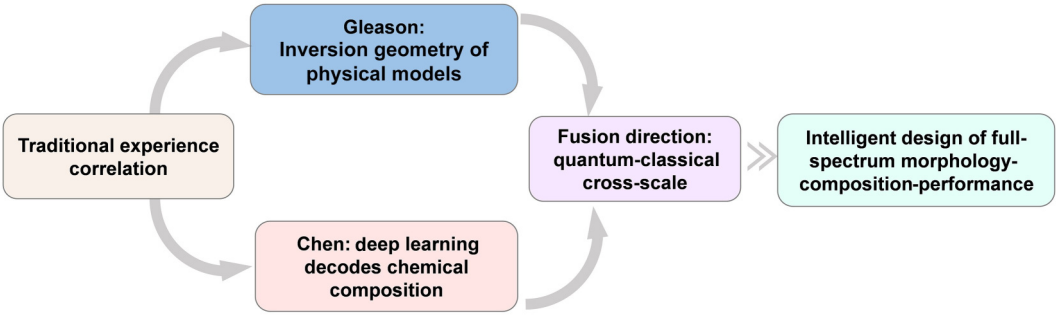

Figure 11 A schematic diagram of the three-dimensional analysis process from physical principles to algorithm innovation and application closed-loop.

performance prediction to crack the complex mechanisms of mechanical/fracture behavior and environmental toxicity. Ultimately, this enables the construction of a full verification chain system from chemical bond breakage to biological interface response. However, bottlenecks such as real-time modeling of dynamic processes, coupling of multiple physical fields, and cross scale correlations still urgently require the development of *in-situ* ML with dynamic graph neural networks to bridge the theoretical experimental gap.

5.1 Decoding of structure activity relationships

Through the 3D analysis of descriptor innovation, scale crossing, and dynamic coupling, this system elucidates the technical route of catalytic descriptor discovery, evolving from static parameters extraction to dynamic mechanisms, and from a single scale to biochemical cross interface evolution. Together, it establishes a theoretical framework for the full chain design of catalytic materials driven by ML (Fig. 12).

5.1.1 *Catalytic descriptor mining*

The rational design of nanocatalysts is shifting from empirical screening to descriptor-driven paradigms, with a central focus on the establishment of computable mappings for multi-scale structure–activity relationships. Wei et al. [124] pioneered an interpretable dual-mode deep learning framework that integrates over 1200 heterogeneous data samples to classify nanozyme activity types (peroxidase (POD)/oxidase (OXD)/superoxide dismutase (SOD)/catalase (CAT), an accuracy of 90.6%) and predict intensity (R^2 = 0.80) (Fig. 13(a)). SHAP sensitivity analysis revealed that the transition metal d-band center is the key descriptor (contribution weight > 35%), overturning the traditional "metal non-metal synergy" cognition and guiding the rational design of Ru-based nanozyme peroxidase activity with 3.2 times enhancement. However, this model remains limited in biological interface scenarios, such as the surface protein crown effect not being included in the descriptor system.

In response to this challenge, Ferdosi et al. [39] innovatively proposed a biological interface competition descriptor (protein/nanoparticle (P/NP) ratio) (Fig. 13(b)). By tuning the ratio of protein to nanoparticle surface area to enhance binding competitiveness, they increased the number of proteins identified on individual nanoparticle surface by 1.2–1.7 times and expanded plasma proteome coverage by 3 times. This descriptor is also associated with the fate of nanomaterials *in vivo*, filling the gap between the structure and activity chain of "synthetic surface–biological interface–*in vivo* behavior". However, its applicability in catalytic scenarios needs to be further deepened by combining chemical bond descriptors.

At the chemical bond scale, Dang et al. [116] discovered an interface bond sequence descriptor (Ni–O–Co bond density). Ni/$CoMoO_4$ heterostructures were constructed through interface engineering, and *in-situ* Fourier transform infrared (FTIR) and DFT confirmed that the Ni–O–Co bond was a dual enzyme active site, which reduced the electron transfer energy barrier (0.32 eV) and increased the oxidase activity by 46.2 times. This descriptor was integrated into a deep learning assisted colorimetric platform to achieve high-sensitivity detection of organic phosphorus (limit of detection (LOD) = 1 μM), but its universality was limited by specific metal combinations.

To overcome the limitations of materials, Gao et al. [120] developed a white box ML framework that integrates DFT calculated OOH adsorption free energy (ΔG_{OOH}) and bandgap characteristics to establish a 2D nanozyme activity prediction model (R^2 = 0.91) (Fig. 13(c)). The MoS_2 edge sites with ΔG_{*OOH} = 0.8 eV obtained increased the H_2O_2 decomposition efficiency by 80%, validating the universal value of the adsorption energy descriptors in tumor catalytic therapy.

Compared with the above work techniques (Table 2), in the future, it is necessary to develop dynamic operation descriptors and combine them with *in-situ* spectroscopic ML technology, to capture the dynamic process from chemical bond breaking (such as Dang's Zr blocking mechanism) to protein crown reconstruction (such as Ferdosi's Vroman effect) in real time, and fianlly establish a full chain catalytic activity prediction model for the "chemical bond interface biological environment".

5.1.2 *Biological interface mechanisms*

In the study of nanomaterial biological interface mechanisms, ML provides a powerful analytical paradigm for systematically reconstructing structure–activity relationships. The GraphBNC framework was the first to achieve accurate atomic level prediction of gold nanoclusters blood protein interactions by integrating multi-scale optimization strategies with graph neural networks (Fig. 13(d)) [125]. Based on molecular dynamics simulation data combined with Monte Carlo structural search, this approach overcomes the limitations of traditional experimental dependency and reveals the key regulatory rules of ligand reconstruction on binding energy (such as p-MBSA carboxyl orientation enhancing apolipoprotein E binding). This "structural topology dynamic interface functional output" decoding logic has been widely

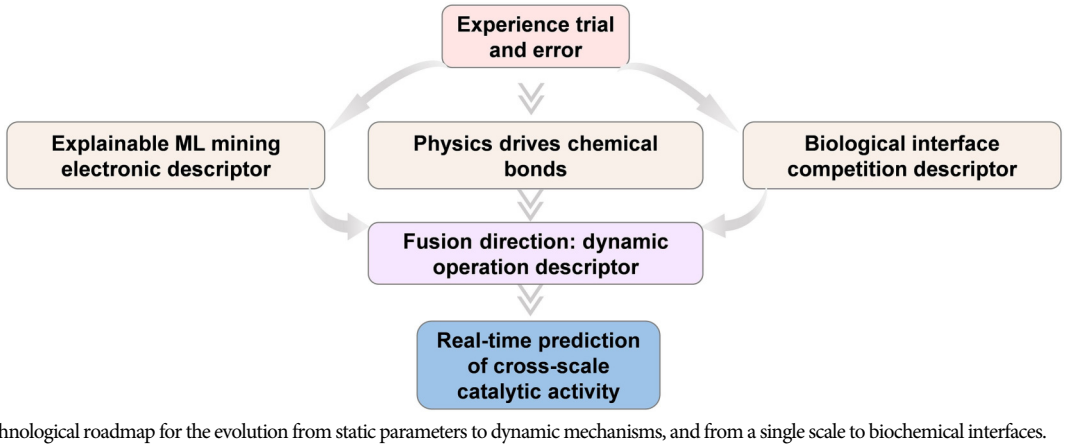

Figure 12 A technological roadmap for the evolution from static parameters to dynamic mechanisms, and from a single scale to biochemical interfaces.

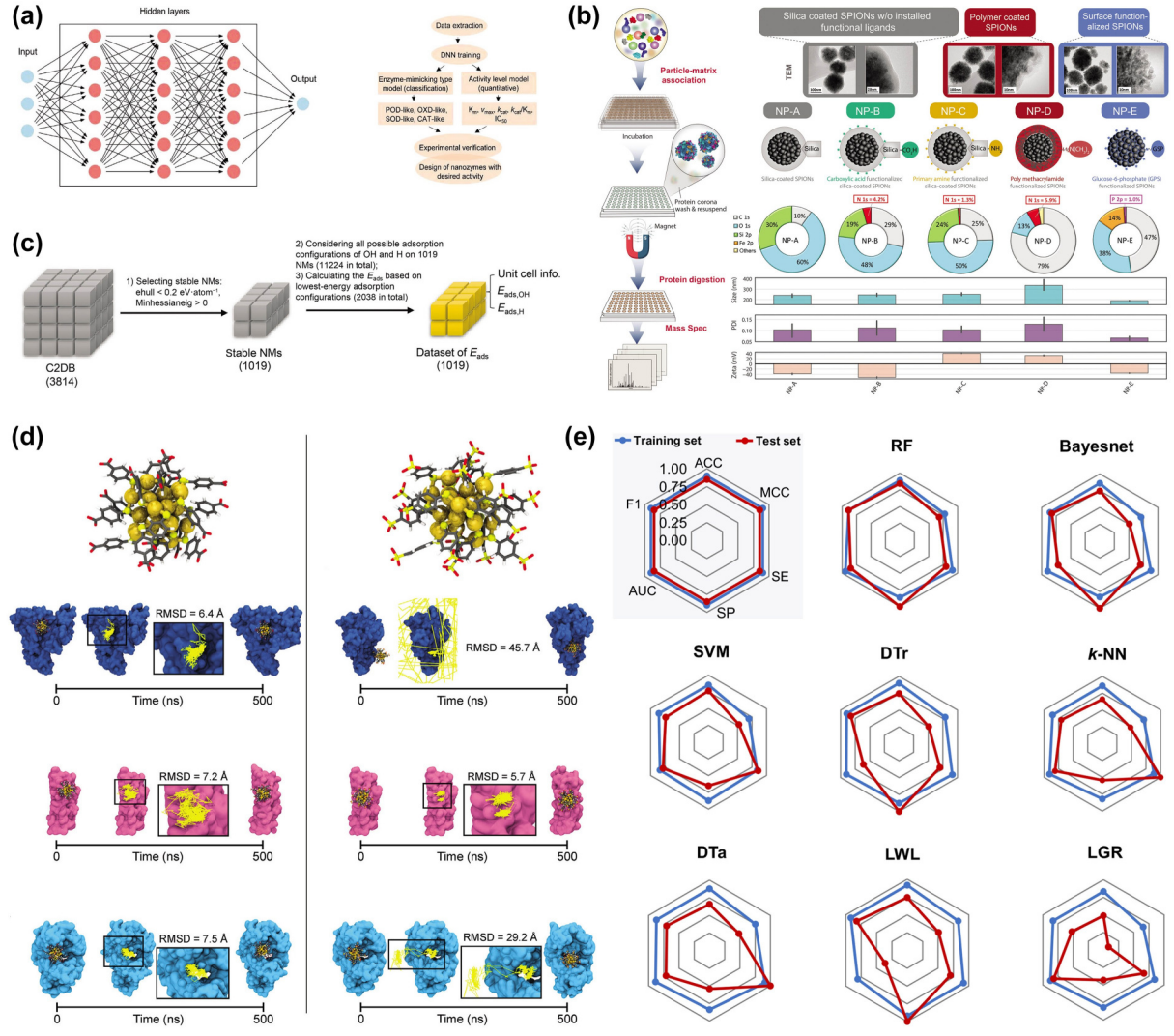

Figure 13 (a) Schematic diagram of the construction of a ML model for predicting enzyme activity in nanomaterials. Reproduced with permission from Ref. [124], © Wiley-VCH GmbH 2022. (b) NP proteomics workflow and characterization. Reproduced with permission from Ref. [39], © Ferdosi, S. et al. 2022. (c) Constructing an adsorption energy database based on high-throughput DFT calculations using a two-dimensional material library (C2DB). Reproduced with permission from Ref. [120], © Wiley-VCH GmbH 2023. (d) Validation of the predicted $Au_{25}NC$ protein complex by GraphBNC. Reproduced with permission from Ref. [125], © Pihlajamäki, A. et al. 2024. (e) The performance of models constructed using eight ML algorithms. Reproduced with permission from Ref. [114], © Huang, Y. et al. 2025.

Table 2 Comparison table of technological breakthroughs in various works

Descriptor type	Prediction accuracy	Experimental verification of effectiveness
d belt center	Active intensity $R^2 = 0.80$	Ru-based enzyme activity ×3.2
ΔG_{*OOH}	Bandgap correlation $R^2 = 0.91$	MoS_2 efficiency +80%
Ni–O–Co density	The activity has increased by 46.2 times	Detection limit: 1 μM
P/NP ratio	Protein detection ×3	The prediction of *in vivo* delivery was consistent

developed in the protein crown research of Lazarovits et al. [126]. This work uses a mass spectrometry deep learning closed-loop system to quantitatively correlate the dynamic adsorption protein combination patterns (such as IgG/complement ratio) with organ distribution, demonstrating that the synergistic effects of multi protein (rather than single protein) dominates the fate *in vivo*, with a double-blind prediction accuracy of 94%. Furthermore, Sengoteyan et al. [127] constructed a triple descriptor system (core coating protein crown) and quantified for the first time the dominant contribution of protein crown composition to zeta potential in human serum environment (a weight of 0.62), overturning the traditional notion of "core chemistry dominance" and revealing a strong correlation between albumin/apolipoprotein ratio and potential reversal ($r = -0.75$).

When the research scale extends to chronic pathological mechanisms, a multimodal feature fusion model integrates 87 cross scale interface features to establish a quantitative prediction chain (area under the curve (AUC) = 0.89) linking "physicochemical properties (zeta potential/solubility)–lysosomal damage–cytokine cascade (IL-1β/TGF-β1)–pulmonary fibrosis" [114], revealing that

macrophage nicotinamide adenine dinucleotide hydride (NADH) depletion is the core pathway of fibrosis (Fig. 13(e)). Correspondingly, a dual isotope fingerprint tracing starts from the intrinsic materials properties of and uses ML to explore the hidden correlation between silicon oxygen isotope fingerprints and manufacturing sources (a classification accuracy of 93.3%), confirming that the inherent physical properties of nanoparticles can decode the attribution of biological interface behavior in complex environments.

5.2 Extreme performance prediction

5.2.1 Mechanics/fracture behavior

In the field of predicting extreme mechanical properties, ML interatomic potential (MLIP) is becoming a core tool to break through the bottleneck of traditional simulations, especially for high complexity systems (porous structures and low symmetry materials) and unconventional properties (negative thermal expansion and ultra-low thermal conductivity), thus demonstrating paradigm advantages. Mortazavi's team systematically validated the cross scale prediction capability of MLIP (Figs. 14(a)–14(c)) [128–131]. By passively training moment tensor potential (MTP) instead of density functional theory (DFT), they enabled efficient simulations of kiloatomic level systems such as fullerene networks [132], carbon nitride [130], and graphitic alkynes [133], achieving a 10–20 fold increase in computational efficiency and elimination of imaginary frequencies (Fig. 14(d)) [134]. The key breakthrough lies in establishing a quantitative structure–property relationship between atomic configurations and extreme physical properties.

(1) Mechanical strength mechanism: Revealing that the ultra-high in-plane strength (91 GPa) of γ-graphitic alkyne originates from the sp-sp^2 hybrid network [133], whereas the strength attenuation (670 to 480 MPa) of Ti_3C_2 MXene shows a critical threshold response to oxygen vacancy concentration [135].

(2) Abnormal thermomechanical behavior: Quantifying the correlation between the negative thermal expansion coefficient of porous materials (such as graphene reaching 2.3×10^{-5} K^{-1}, 10 times higher than graphene) and out of plane wrinkle strengthening, and analyzing the dominant effect of porosity on ultra-low thermal conductivity (C_3N_6 thermal conductivity < 1 W·mK^{-1}).

(3) Multi-functional coupling design: By regulating the connection groups, the synergistic optimization of the strength (67 GPa) and carrier mobility of C_6N_7 based materials is achieved [131], providing a theoretical model for flexible optoelectronic devices.

The above work corresponds to the methodology of Sengottiyan's triple descriptor framework, as both approaches integrate multi-level structural features (nucleus coating protein crown vs. atomic bonding pore stacking sequence) to decipher the structure–activity relationships of complex systems. Moreover, Firestein's *in-situ* experimental verification (TEM/atomic force microscopy (AFM)) has strengthened the reliability of MLIP

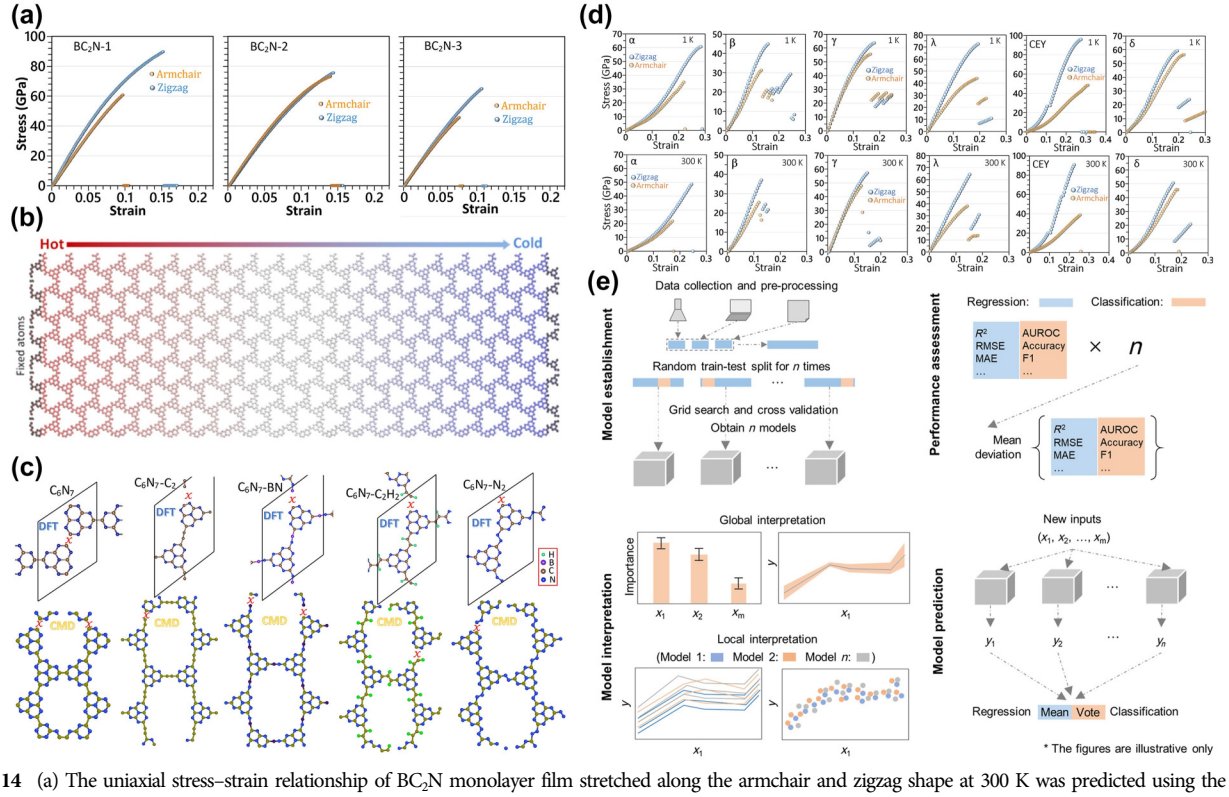

Figure 14 (a) The uniaxial stress–strain relationship of BC_2N monolayer film stretched along the armchair and zigzag shape at 300 K was predicted using the MTP-based classical molecular dynamics (CMD) method. Reproduced with permission from Ref. [128], © Elsevier Ltd. 2021. (b) Schematic diagram of non-equilibrium molecular dynamics (NEMD) simulation calculation of lattice thermal conductivity. Reproduced with permission from Ref. [130], © Elsevier Ltd. 2020. (c) The failure mechanism of C_6N_7 monolayer film stretched along the zigzag direction was analyzed using CMD (the second row) models based on DFT (the first row) and MTP (the second row). The "*x*" symbol highlights the area where the initial key fracture occurred. Reproduced with permission from Ref. [131], © Elsevier Ltd. 2022. (d) Predicting the uniaxial stress–strain response of graphene monolayer films using MTP-based molecular dynamics simulations at 300 K in the first and second rows. Reproduced with permission from Ref. [133], © Elsevier B.V. 2022. (e) The average strategy for ML interpretation tasks based on small datasets. Reproduced with permission from Ref. [137], © American Chemical Society 2023.

prediction [135], such as the thickness strength power-law decay ($\sigma \propto d^{-0.5}$) of Ti_3C_2 cross scale experiments and computational cross validation.

5.2.2 *Environmental toxicity assessment*

In the field of environmental toxicity prediction for nanomaterials, ML is systematically solving the quantitative correlation problem between complex biological interfaces and extreme toxicity responses through multi-scale feature integration and interpretable algorithms. Labouta's team pioneered a cross material toxicity prediction paradigm using a meta-analysis framework [136]. Based on 2896 heterogeneous databases, they constructed a decision tree model that revealed the hierarchical influence of "material chemistry > concentration > size > cell type" (information gain > 0.85), and quantified the positive correlation between surface charge (|zeta| > 20 mV) and cell uptake (r = 0.79). This study provides a universal ranking of dominant toxicity factors, but the lack of modeling for dynamic exposure process was specifically addressed in follow-up work by Yu et al. [137] in their plant toxicity research (Fig. 14(e)). By using a small sample averaging strategy (10 fold stratified training) and the RuleGrid visualization tool, they analyzed the regulatory weight (0.58) of the maize root mucin layer on ZnO solubility and found a nonlinear synergistic effect between specific surface area and concentration (high Brunauer–Emmett–Teller (BET) and low concentration promote transport), establishing the first interpretable rule set for agricultural nanosafety design.

When investigating specific material systems, Shirokii et al. [138] innovatively introduced atomic descriptors (electronegativity/ionic radius), overcoming the limitations of traditional classification within LightGBM models and achieving quantitative prediction of concentration dependent toxicity of inorganic nanomaterials (Q^2 = 0.86). A key strength of their approach is its extrapolation capability, and atomic level features allow the model to generalize to unknown materials (with an error of less than 12.2%) and guide the design of lower-toxicity zinc oxide (IC50 increased by three times). In response, Ma et al. [139] constructed a large-scale toxicity database containing 986 cell viability data points for graphene. Using a random forest model, they demonstrated that exposure dose was the most dominant factor influencing toxicity (weight 32.7%), achieving accurate IC50 predictions (R^2 = 0.903) and revealing a triple optimization criterion of "dose particle size surface chemistry".

6 Application scenario implementation

ML-driven material intelligence research and development is achieving a full chain connection from fundamental mechanism analysis to terminal system integration, with the core of establishing a multi-scale collaborative paradigm of "atomic design structure regulation system optimization" [131, 140–143]. This section focuses on the collaborative innovation such as energy storage/optoelectronic devices for energy materials, targeted delivery/intelligent diagnosis closed-loop construction for biomedical applications, intelligent design of pollutant detection/remediation materials for environmental management, and cross scale optimization of composite materials/lightweight structural materials, together forming a deep coupling of "mechanism performance scenario" [144–151]. Nevertheless, critical challenges remain, including limited accuracy in non-equilibrium dynamics, incomplete cross scale validation of faults, and insufficient robustness under extreme operating conditions. Filling in these gaps will require the development of dynamic *in-situ* ML and multiphysics coupling models to promote laboratory breakthroughs and industrial implementation [152–162].

6.1 Energy materials

6.1.1 *Energy storage devices*

Within the intelligent research and development system for energy materials, ML promotes performance breakthroughs in energy storage device by decoupling multi-scale structure–activity relationships. The collaboration between ML and ligand engineering is driving a paradigm shift in rational design of energy storage materials. The study of vanadium-based MOFs uses Br polar ligands to regulate the electronic structure, combined with orthogonal unfolding ML models to decode the nonlinear mapping between electrode thickness/current density and performance, revealing the mechanism of charge redistribution reducing the Zn^{2+} migration energy barrier (Fig. 15(a)) [1]. This paradigm is further applied in manganese based systems, where 1,4-DHAQ (DHAQ = dihydroxy anthraquinone) para substitution is used to construct π-conjugated systems (Fig. 15(b)) [3]. The homologous ML model quantifies the coating parameter capacity relationship from sparse data, confirming that electron delocalization improves charge transfer efficiency. The two works jointly demonstrate the universality of ML in small sample multivariate coupling problems, and optimize charge transfer dynamics through ligand electronic engineering (synergistic induction/conjugation effects) to provide a dual drive design strategy for flexible energy storage devices.

Innovation at the basic simulation layer began with the development of accurate potential functions. Salazar et al. [17] constructed a PLIP ML potential function, which qualified for the first time the effect of surface polarization of ZnO nanoparticles on the competitive nucleation of polymorphs, revealing the critical size effect of the transition from metastable BCT phase (interface energy dominant) in the deep undercooled region to stable Wurtzite (WRZ) phase. This atomic scale mechanism analysis lays a physical foundation for the design of nano energy storage materials, and its long-range interaction capture capability has been extended to research on mechanical energy storage of carbon nanowires by the Zhao et.al (Fig. 15(c)) [163]. The ML potential function discovered an ultra-high twist transfer energy density of 2680 kJ·kg^{-1} for anti configuration nanowires, which is 100 times more efficient than DFT, and also revealed the failure mechanism of uneven strain distribution due to the increase in beam diameter.

The material design layer focuses on high-throughput screening and structural optimization. The forward prediction framework proposed by Liu et al. [164] achieves cross screening of lithium/zinc/aluminum ion batteries positive electrode materials (MAE ≤ 47.34 Wh·kg^{-1}) through the autoencoder Coulomb matrix and CNN model. Its standardized "database descriptor prediction" process significantly accelerated the discovery of new materials. This strategy is further deepened in the field of two-dimensional materials. Long et al. [165] combined ML and *in-situ* characterization to elucidate the process of densification assembly regulated by capillary force on the surface functional groups of graphene/MXene, resolving the contradiction between ion transport and volumetric capacity in high-density electrodes, and providing a theoretical leverage for compact energy storage devices.

The intelligent upgrade of the system application layer is

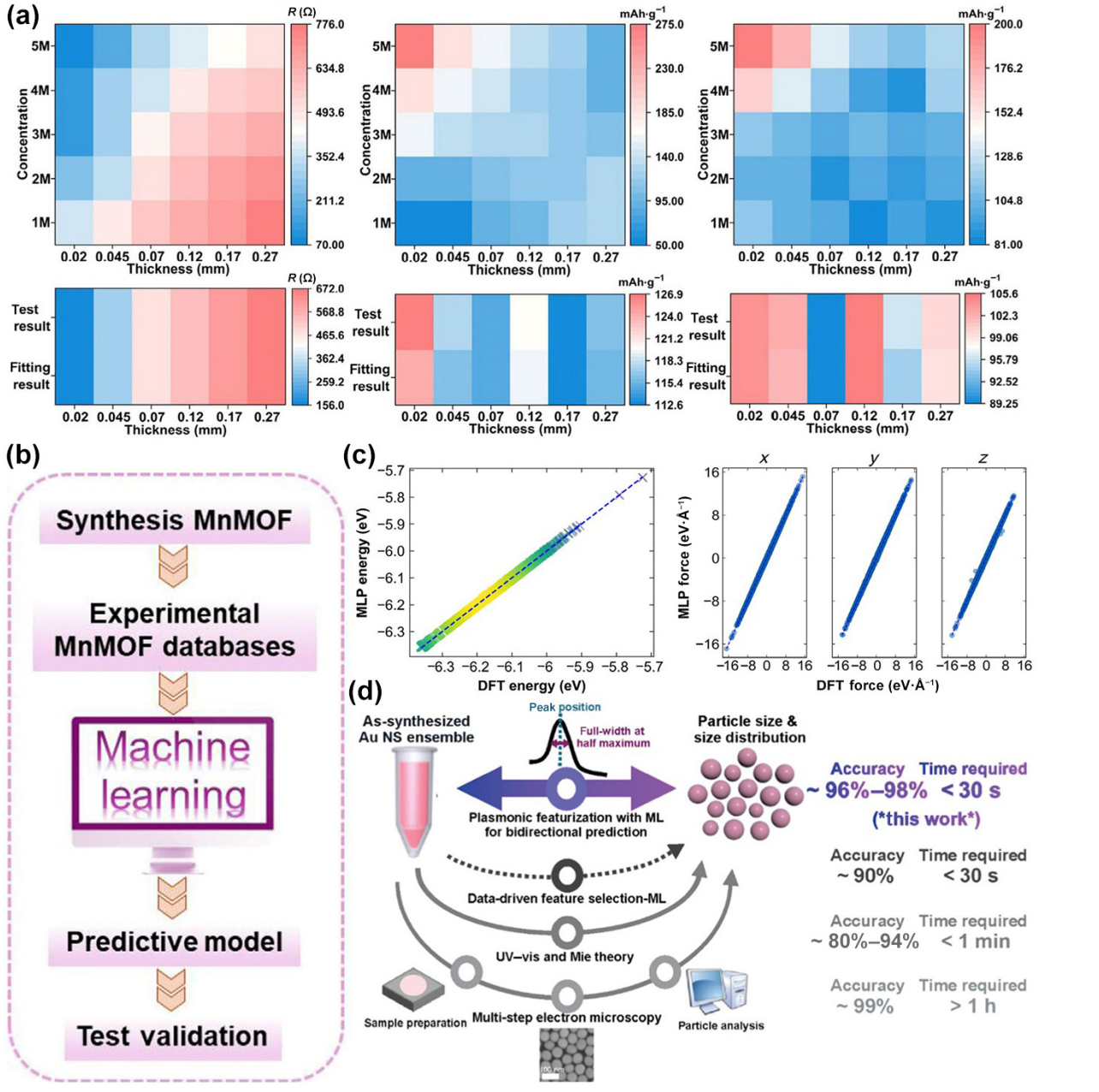

Figure 15 (a) Key electrochemical performance parameters in the first row and ML assisted experimental verification of VBr-180 in the second line. Reproduced with permission from Ref. [1], © Wiley-VCH GmbH 2025. (b) Workflow adopted to build ML models of MnMOCs (MOC = metal-organic coordination). Reproduced with permission from Ref. [3], © Wiley-VCH GmbH 2025. (c) The relationship between the energy predicted by the ML model and the energy calculated using DFT in the test set, as well as the relationship between the atomic force predicted by the ML model and the atomic force based on DFT level. Reproduced with permission from Ref. [163], © Zhao, L. N. et al. 2023. (d) The use of plasma knowledge to guide FE enables us to achieve bidirectional ML methods for the size and size distribution of gold nanospheres from experimentally obtained UV–vis extinction spectra at unprecedented speeds (in seconds) and accuracies (~ 96% to 98%). Reproduced with permission from Ref. [170], © The Royal Society of Chemistry 2022.

reflected in the optimization of environmental adaptability. Kiaghadi's team used the auto-regressive integrated moving average (ARIMA) model to accurately predict the melting dynamics of phase change materials (mean absolute percentage error (MAPE) = 0.0003) [166], verifying that the fin structure improves the photovoltaic efficiency to 13.993% under the gravity field. Ramesh Srenyvasan et al. [167] used support vector machine/long-short-term memory (SVM/LSTM) dual model collaboration to compress the prediction error of the ZnMnO4 coated solar distiller to < 3%, and established a quantitative decision chain from meteorological parameters to thermal insulation efficiency to daily water production. Cross scale functional integration reached a new height in the research of Lin et al. [168]. Based on the wide-kernel deep CNN (WDCNN) model of the moisture-enabled energy generation (MEG) device, the fluid classification (an accuracy of 100%) and continuous power generation were achieved synchronously. Its hydrophilic and hydrophobic microchannel design endows the device with dual functional characteristics of "sensing power supply", opening up a new path for self-driven energy storage systems.

Current challenges focus on the accuracy of non-equilibrium dynamics simulations (such as colloid assembly processes), the robustness of extreme operating condition models (such as photovoltaic thermal management), and cross scale data governance (such as the scarcity of rare microbial data). In the future, it is necessary to integrate active learning, enhance the generalization ability of potential functions, combine generative AI to optimize the construction of material database, and ultimately achieve an intelligent closed-loop of "mechanism analysis material creation device optimization".

6.1.2 *Optoelectronic devices*

Through in-depth research on material synthesis, structural analysis, and system control in the field of optoelectronic devices, this study systematically elucidates the empowering role of ML in the entire chain of "atomic interface control (ligand engineering)–micro morphology design (wrinkles/plasma)–macro device optimization (distillation/ light-emitting diode (LED))", highlighting the core value of multi-scale modeling and dynamic regulation.

ML drives breakthroughs in optoelectronic device performance through precise multi-scale regulation. The core of quantum material design layer lies in ligand engineering and stability optimization. Jing's team developed a zwitterionic ASC-16 ligand and combined it with the Bayesian optimization framework to achieve two iterations to lock the optimal synthesis conditions [88]. The near-infrared PLQY of ytterbium-doped perovskite quantum cutting materials reached 192% ± 5%, and triple stability was achieved through strong chelation (UV irradiation > 100 h/ 85 °C annealing > 24 h/high humidity environment > 3 months). This strategy is being continuously deepened in the field of perovskite LEDs.

The nanostructure analysis layer focuses on the quantitative correlation between morphology and function. An et al. [169] combined electron tomography with Gaussian mixture model (GMM) to quantitatively divide 151 nanofolds polyamide films into four morphology groups (dome/depression/pancake/cluster) for the first time, revealing the power-law of monomer concentration regulating spatial arrangement through Turing instability, and established a closed-loop structure activity model of the local Young's modulus methanol permeability of the fold morphology. This morphological analysis paradigm has been innovatively applied in plasma materials, such as the "knowledge-driven characterization" strategy proposed by Tan's team [170], which extracts key parameters such as dipole/quadrupole resonance peaks from extinction spectra, constructs a bidirectional B-spline model, and accurately predicts the size (an error of 2.3%) and distribution (an error of 1.0 nm) of gold nanospheres, providing a basis for the reverse design of optical field control devices (Fig. 15(d)).

The intelligent control layer of the system can achieve dynamic performance optimization. Gandhi et al. [171] developed the online sequential extreme learning machine (OSELM) adaptive model and combined it with the binary search tree (BST) algorithm to optimize the heat transfer process of the SiO_2/TiO_2 coating ladder distiller in real time. Compared with the traditional LSTM, the evaporation rate increased by 49.21%, the unit production cost decreased by 35%, and the response speed increased by 3 times. This dynamic control logic has been upgraded in LED thermal management.

The current challenges of this related research focus on adaptability to extreme operating conditions (such as predicting perovskite phase transitions under high humidity environments) and multi-scale correlation modeling (from nanowrinkle dynamics to membrane system permeability). We need to develop strategies that combine dynamic *in-situ* characterization with ML, such as real-time GMM clustering using liquid-phase atomic force microscopy, and explore the potential of generative AI in cross scale structural reverse design, thus ultimately achieving intelligent collaborative optimization of "quantum efficiency light field regulation thermal management".

6.2 Biomedical sciences

6.2.1 *Targeted delivery*

ML is reconstructing biomedical delivery systems by decoding multi-scale interface and optimizing dynamic process. By focusing on the space–time in the cell transport layer, it enables dynamic analysis and targeted immune regulation. For instance, Yang et al. [172] developed the LungVis 1.0 AI-3D imaging platform, which applies an adaptive NNU Net to segment 20 level bronchial trees in mouse lungs. This system quantitatively revealed for the first time the alveolar deposition advantage of aerosol administration (ventilator-assisted aerosol delivery (VAAD)) (B/A ratio of 0.07 vs. liquid administration 1.1–1.4), and discovered that tissue resident macrophages (TRM) actively patrol via Kohn pores to mediate 83%–93% nanoparticle clearance, overturning the traditional passive clearance cognition. This dynamic mechanism has also been theoretically validated in plant nano interactions (Fig. 16(a)). Awere's study demonstrated that carbon dots (CDs) induce oxidative stress, thereby activating antioxidant pathways [173], whereas a multilayer perceptron (MLP) model accurately correlated stress parameters with metabolite accumulation (R^2 = 0.999). This provides a universal analytical approach for understanding how nanomaterials modulate cellular defense responses.

The core breakthrough at the molecular interaction layer lies in the precise prediction of nano biological interface. Pihlajamäki et al. [125] developed the GraphBNC framework, which innovatively integrates graph neural networks with Monte Carlo structure search. Relying solely on molecular dynamics simulation data, this framework predicts dynamic binding sites between gold nanoclusters (Au_{25}/Au_{102}) and blood proteins (such as p-MBSA ligand carboxyl orientation enhanced apolipoprotein E binding), solving the problem of reconstructing non-static interfaces that are difficult to capture experimentally. This strategy has been further extended to predicting the fate of the body. Lazarovits's team established a mass spectrometry deep learning closed-loop model [126], demonstrating that multi protein synergistic effects (such as IgG/complement ratio) dominate the distribution of nanoparticle organs (Fig. 16(b)). By reversely engineering the protein crown in the host environment, they achieved a 50%–70% reduction in liver and spleen uptake, marking a paradigm transition from "single protein knockout" to "multi protein synergistic regulation".

At the clinical diagnosis and treatment layer, an intelligent closed-loop of "diagnosis treatment" has been realized. Huang et al. [174] developed a CNN-ANN hybrid model to classify ear endoscopic images with an accuracy of 95% (sensitivity of 95%/specificity of 100%) while synchronously optimizing the poly(lactic-co-glycolic acid) (PLGA) drug-loaded nanoparticle dosing scheme, tripling the concentration at the middle ear lesion and improving the cure rate by 40%. The diagnosis and treatment integration paradigm was

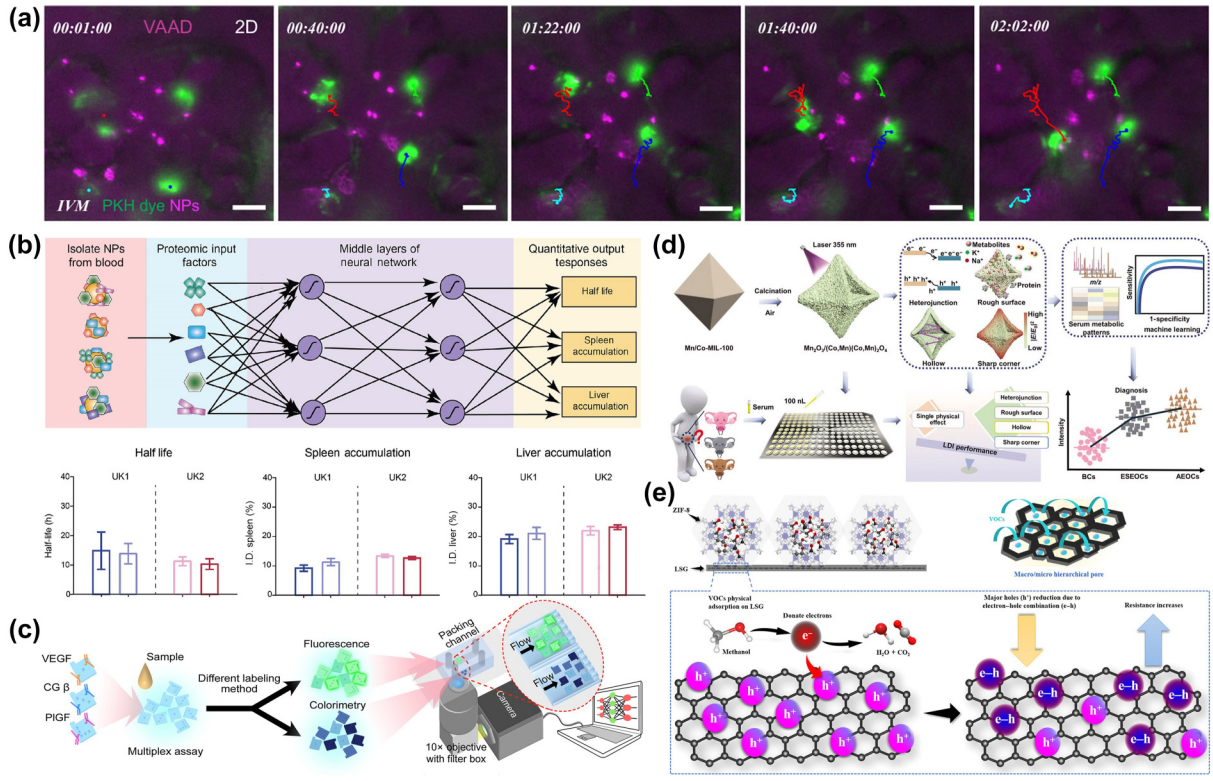

Figure 16 (a) Pulmonary *in vivo* microscopy (IVM) shows that PKH-labeled macrophages dynamically move towards NPs deposited on the surface of alveolar epithelium by patrolling. $N = 3$, VAAD-0 h lung, independent biological replicates. Scale: 20 μm. Reproduced with permission from Ref. [172], © Yang, L. et al. 2024. (b) Proteins separated from the surfaces of 8, 15, 35, 50, and 80 nm gold nanoparticles in the loop were used as inputs for the neural network at five time points. Reproduced with permission from Ref. [126], © American Chemical Society 2019. (c) Describing the multiplex immunoassay and microscopy imaging process with particle filled channels for fluorescence and colorimetric labeling methods. Reproduced with permission from Ref. [71], © American Chemical Society 2023. (d) A workflow diagram of MO/CMO assisted LDI ms extraction of serum metabolic patterns was prepared using Mn/Co-MIL-100 as a template. Reproduced with permission from Ref. [175], © Wiley-VCH GmbH 2023. (e) ZIF-8@LSG (LSG = laser-scribedgra-phene) schematic diagram of sensor capturing VOCs, sensing mechanism based on LSG sensor, and schematic diagram of macroscopic/microscopic layered porous structure with VOC gas transport. Reproduced with permission from Ref. [179], © The Royal Society of Chemistry 2024.

further upgraded in particle analysis by Choi's team, who used a YOOv5 model combined with synthetic data enhancement to achieve rapid identification of drug loaded gel particles (mean average precision (mAP) = 0.88, 0.11 s/frame (Fig. 16(c)) [71]. Their triple immunoassay flux was 50 times higher than enzyme-linked immunosorbent assay (ELISA), providing a real-time monitoring tool for personalized delivery.

Current challenges of this related research focus on cross-scale modeling bottlenecks (such as GraphBNC not covering the temporal evolution of protein coronations) and clinical translation barriers (mouse to human B/A ratio validation, *in vivo* real-time monitoring accuracy). To address this, researchers need to develop dynamic *in-situ* characterization strategies coupled with active learning (such as combining liquid-phase atomic force microscopy with real-time graph neural networks), and explore federated learning approaches to address privacy constraints in medical data, ultimately achieving a precise control loop of "molecular design cell transport clinical response".

6.2.2 *Intelligent diagnosis*

ML reconstructs the biomedical diagnostic system through multimodal signal enhancement and closed-loop clinical decision-making. The core breakthrough of the molecular diagnostic layer lies in the precise capture of trace biomarkers. Pei et al. [175] designed a concave octahedral Mn_2O_3/(Co,Mn)(Co,Mn)$_2O_4$ (MO/CMO) heterojunction material, synergistically regulating the surface roughness and band structure, thereby increasing the signal intensity of laser desorption/ionization mass spectrometry (LDI-MS) by 10–48 times (Fig. 16(d)). Combined with a sparse learning algorithms, early diagnosis of ovarian cancer (AUC = 0.986) was achieved, breaking the dependence of traditional mass spectrometry on serum sample preprocessing. This enhancement strategy was further deepened in single atom detection. Hao's team developed a functional nanomaterial FSVNet deep learning architecture [176], which combines click chemical signal amplification with convolutional neural network time-frequency domain feature fusion, breaking through the threshold limit of signal-to-noise ratio and achieving the detection of 0.2 amol·L^{-1} copper ion (single atom level) bivalent states, with a sensitivity 100 times higher than traditional electrochemical methods. Yu et al. [177] developed a label free microbial classification strategy based on microbially synthesized AuNPs and ML. By characterizing the particle size, surface plasmon resonance spectroscopy, and surface potential of AuNPs synthesized by different microorganisms (bacteria/fungi), combined with the random forest algorithm, a multi-level classification model (covering kingdoms/orders/genera/species) was constructed, which can achieve a classification accuracy of 100%, which is consistent with traditional gene identification.

The wearable sensing layer can focus on flexible electronics and dynamic signal analysis. Kwon et al. [178] developed biocompatible

solderable graphene (functionalized conductive graphene (FCG)) and combined it with fully printed nanofilm technology to construct a wireless electromyography sensing system (all-printed nanomembrane hybrid electronics (p-NHE)). By optimizing the three channel electrode layouts through deep learning, the accuracy of seven gesture recognition reached 99%, subverting the traditional high-density electrode array requirements. This dynamic perception paradigm has been innovatively applied in gas diagnosis. Tran's team constructed a ZIF-8@laser engraved graphene biomimetic sensor (ZIF = zeolitic imidazolate framework) [179], combined with wavelet packet transform and MLP, and achieved a detection limit of 0.8 ppm and a classification accuracy of 96.5% for lung cancer respiratory markers at room temperature, solving the bottleneck of traditional sensor high-temperature operation (Fig. 16(e)).

The real-time diagnostic layer promotes the implementation of clinical scenarios. Liu et al. [180] developed a smartphone integrated electrochemiluminescence (ECL) platform that enhanced the signal by modifying perovskite Ag^{+}@UiO-66-NH_2, and embedded an ANN model to automatically analyze ECL images, thus achieving on-site detection of nitrofuranone (LOD = 0.09 nM). This portable diagnostic logic has been upgraded in fluid analysis. The Au/Au@Pt nanoprobe array designed by Wang's team captures urinary protein through surface charge gradient [181]. The principal component analysis (PCA)-SVM model synchronously completes early diagnosis of bladder cancer and protein classification with an accuracy of > 95%, and solves the time-consuming problem of traditional electrophoresis within 10 min.

The current challenges in related research focus on actual sample interference (such as respiratory humidity affecting volatile organic compound (VOC) recognition) and individual difference compensation (model drift caused by fluctuations in urine metabolism) (Table 3). Researchers need to develop adaptive online learning strategies (such as incremental SVM to dynamically update classification boundaries) and explore federated learning to protect medical privacy data, ultimately achieving the precision medical closed-loop of "trace capture dynamic perception bedside diagnosis".

6.3 Environmental governance

6.3.1 *Pollutant detection*

Intelligent pollutants detection technology in environmental governance is undergoing a paradigm shift from single mechanism sensing to multimodal collaborative sensing. Henriquez et al. [182] first integrated chemical resistors (ZnO/CuO) and calorimetric sensors (Pt catalysis/thermal conductivity) through a multi-transducer microelectromechanical system (MEMS) sensor array (Fig. 17(a)). They utilized local *in-situ* nano synthesis technology to achieve ultra-low power operation of 7 mW·device^{-1}, and combined CNN transient analysis strategies to synergistically decode the dynamic response of flammable/toxic gases, breaking through the selective limitations of a single sensing mechanism (a classification accuracy of 97.95%, and a concentration regression error of 14%). Shao et al. [183] further deepened the multi-mechanism fusion strategy. By constructing a 3D van der Waals heterojunction of Pt-COF@SnO_2@C and combining it with the PCA-SVM ML framework, the contribution of core–shell components was quantified. It was revealed for the first time that the Pt COF shell dominated 72% of the TEA response signal at room temperature and high humidity environments (the SnO_2 core enhanced electron conduction contributed 28%), providing theoretical guidance for gas monitoring in complex environments.

For monitoring pollutants in the food supply chain, ML-driven sensor design is shifting towards multifunctional integration and enhanced anti-interference capabilities. Zhao et al. [184] developed an ionic liquid tuned anthocyanin/UiO-66 colorimetric array (Fig. 17(b)). A superhydrophobic interface (contact angle > 130°) was achieved by nano SiO_2 coating, significantly improving the stability in high humidity environments. The correlation between cross reactive RGB signal polyamines and material porosity (> 800 $m^2 \cdot g^{-1}$) was analyzed using a VGG16 deep convolutional network. Ammonia sensitivity (r = 0.89) successfully quantified the pork spoilage stage with a classification accuracy of 99.16%. Similarly, Abbas et al. [185] used random forest regression to optimize the synthesis parameters of poly(3,4-ethylenedioxythiophene) functionalized carbon matrix suspended Cu nanoparticles (PEDOT-C@CuNPs) electrodes (accurate prediction of pH = 7.2/drying time of 8 min), the exposure of highly active Cu (111) crystal faces (accounting for > 60%), and high recovery detection of nitrite in complex food matrices (98.5%, LOD = 3.91 μM), confirming the analytical advantages of ML over traditional orthogonal experiments in the structure–activity relationships of materials (such as PEDOT thickness–electron transfer rate volcano curves).

The common technical challenges focus on the suppression of interference from complex matrix and the verification of long-term stability. All four research works point out the need to address key issues such as the cross interference of mixed gases (such as multi-component toxic gases in industrial scenarios and coexistence of food spoilage polyamines), extreme environmental adaptability (high humidity of 90%RH (RH = relative humidity)), and long-term service stability of materials (> 6 months). There is an urgent need to develop adaptive ML algorithms and robust nano coating technology.

6.3.2 *Repair material*

Through a progressive framework of repairing material design, analyzing pollutant mechanisms, and improving environmental adaptability, this system elucidates the core role of ML in bridging the full chain of "waste resource utilization–adsorption mechanism decoding–complex scene implementation". Collectively, these

Table 3 Key scientific logic and technological evolution path

Diagnostic level	Core breakthrough	Technological extensibility
Molecular diagnosis	MO/CMO enhanced mass spectrometry and sparse learning for early screening of ovarian cancer	Upgrade to single-atom electrochemical detection
Wearable perception	Three-channel deep learning gesture recognition	Extend to bionic gas sensing
Instant diagnosis	On-site detection of smart phone ECL-ANN	Upgrade to the three-in-one analysis of urine protein
Cross-level collaborative hub	The surface charge gradient regulates the protein adsorption capacity	Guide the collaborative design of "materials-interfaces-algorithms" for diagnostic devices

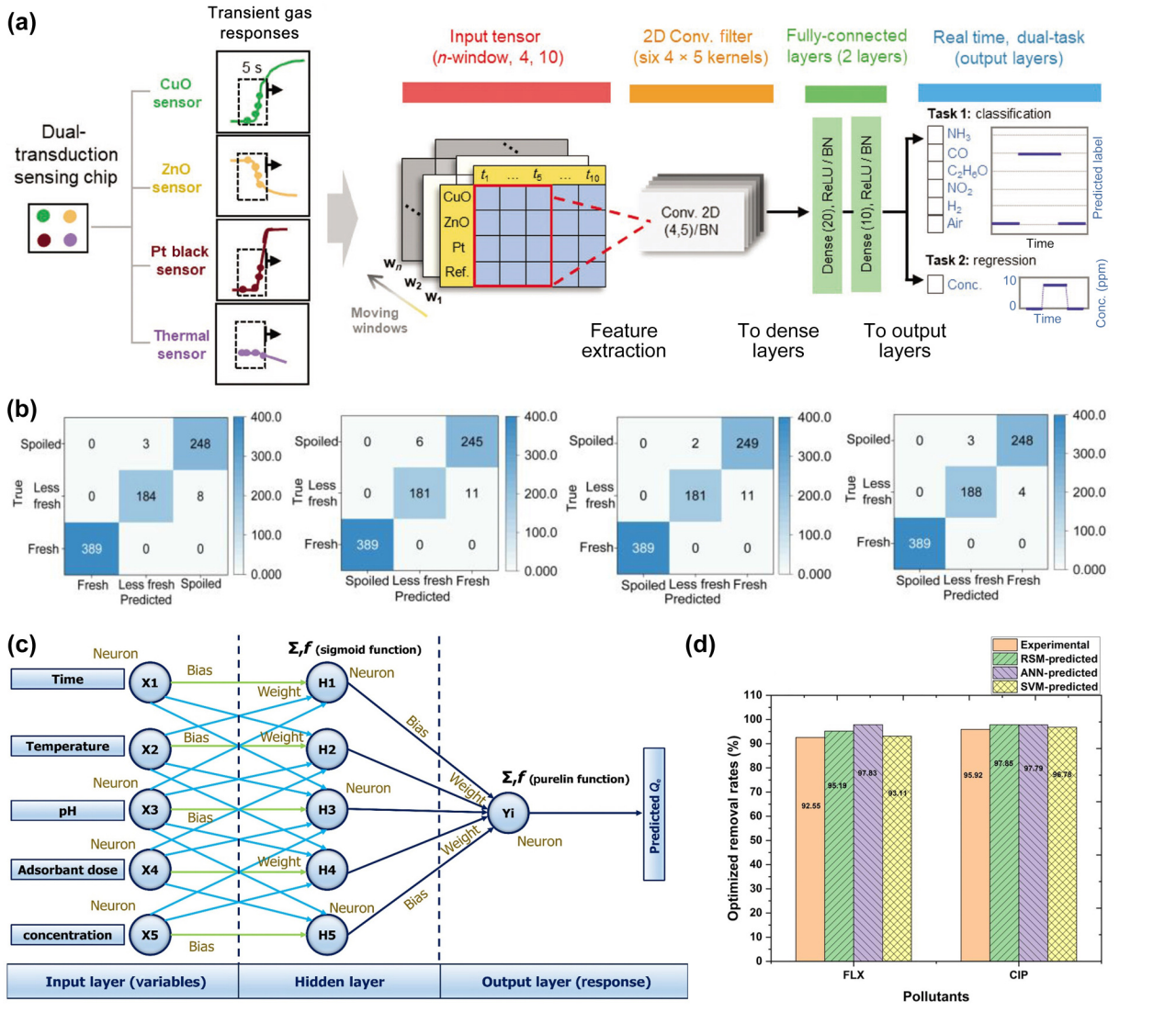

Figure 17 (a) A data processing strategy for real-time identification and quantification of target gas concentration using four types of sensor signals. Reproduced with permission from Ref. [182], © Wiley-VCH GmbH 2023. (b) The deep CNN (DCNN) model predicts the confusion matrix of pork freshness using 832 test datasets. Reproduced with permission from Ref. [184], © Elsevier Ltd. 2024. (c) Artificial neural network model architecture. Reproduced with permission from Ref. [187]. © Elsevier B.V. 2024. (d) The bar chart compares the optimized removal rates of response surface methodology (RSM), ANN, and SVM with the experimentally measured values. Reproduced with permission from Ref. [189], © Elsevier B.V. 2025.

studies demonstrate a technological evolution of "single pollutant treatment–multi pollutant collaborative removal–cross disciplinary functional expansion".

ML is driving the evolution of environmental remediation materials towards a trinity paradigm of "precision design, intelligent optimization, and functional expansion". Al-Nowaiser et al. [186] pioneered an electrocoagulation/GO@ZIF-7 adsorption synergistic system, utilizing an XGBoost model to analyze multi-parameter coupling effect (R^2 = 0.98) and to quantify the linear correlation between ZIF-7 pore size (3.4–4.0 Å) and Pb^{2+} adsorption energy (R^2 = 0.93). This approach achieved a remarkable removal efficiency of 98% while elucidating the synergistic mechanism between aluminum electrode flocculation and MOF confinement. Extending this data-driven strategy to biomass-based materials, Belcaid et al. [187] used cassava peel biochar loaded with MnO_2/γ-Fe_2O_3 nanoparticles, developing an ANN model that precisely predicts cobalt/chromium adsorption capacity (349–546 mg·g^{-1}) (Fig. 17(c)). Through feature engineering (FE) of pH/temperature interactions (root mean squared error (RMSE) < 5%), this study solved the problem of performance fluctuations caused by insufficient porosity of agricultural waste.

For complex pollutant systems, ensemble learning and cross scale modeling have become crucial for overcoming the bottleneck of multi-mechanism collaboration. BinMakhashen et al. [188] designed PGB-LTH (PGB = polyethyleneimine, GO, and bentonite, and LTH = layered triple hydroxide) layered nanocomposites and used the Extra Trees integrated model (test set R^2 = 0.99) to quantify the reverse adsorption mechanism of anionic/cationic dyes (methyl orange ΔH = −41.23 kJ·mol^{-1} vs. crystal violet ΔH = 19.33 kJ·mol^{-1}). The feature importance analysis revealed that pH was the decisive parameter (Gini index > 0.35), whereas the Bagging strategy effectively overcame the risk of overfitting in small samples. Similarly, Ebere Enyoh et al. [189] combined artificial neural networks with molecular docking technology to synthesize carbon quantum dots using discarded polyethylene terephthalate (PET) (Fig. 17(d)), simultaneously optimizing the adsorption

efficiency (> 95%) and biological activity (SERT binding energy of −8.7 kcal·mol^{-1}) for fluoxetine/ciprofloxacin. A multitask learning framework was adopted to address the challenge of cross domain data heterogeneity in environmental pharmaceuticals.

A scientific challenge remains the practical adaptation of cross scale mechanism correlations for industrialization. The challenges of coexisting ion competition adsorption, material regeneration cycling, and complex matrix interference urgently demand the development of adaptive reinforcement learning algorithms and multiphysics coupling models to promote the transformation of laboratory materials into engineering applications [190].

6.4 Structural materials

6.4.1 *Composite materials*

Through the technical chain of precise performance prediction, quantifiable interface mechanism, and systematic sustainable design, the core value of ML across the entire "component design–process optimization–service performance" cycle of structural composite materials is systematically explained. The evolution from single performance prediction models (such as Sun's CNT strength model) to multi-objective collaboration frameworks [191, 192], and further toward cross lifecycle optimization, highlights the disruptive innovation of intelligent R&D paradigm on traditional material development pathways [193].

ML is driving the transformation of structural composite material R&D towards a three-in-one paradigm of "multi-objective collaborative optimization interface precise control sustainable design". Alabduljabbar et al. [191] and Alyousef et al. [194] used gene expression programming (GEP) models to accurately predict the compressive strength (R^2 > 0.92) and electrical resistivity (MAE reduced by 82%) of graphene reinforced concrete, combining them with the NSGA-II multi-objective algorithm to generate Pareto optimized material parameters (Fig. 18(a)). SHAP interpretability analysis quantified the core impact mechanisms of graphene dosage (contribution > 50%) and curing age (weight 45%), overcoming the limitations of traditional empirical formulas for synergistic optimizing electrical and mechanical properties. This data-driven strategy has been further extended to multi-component composite

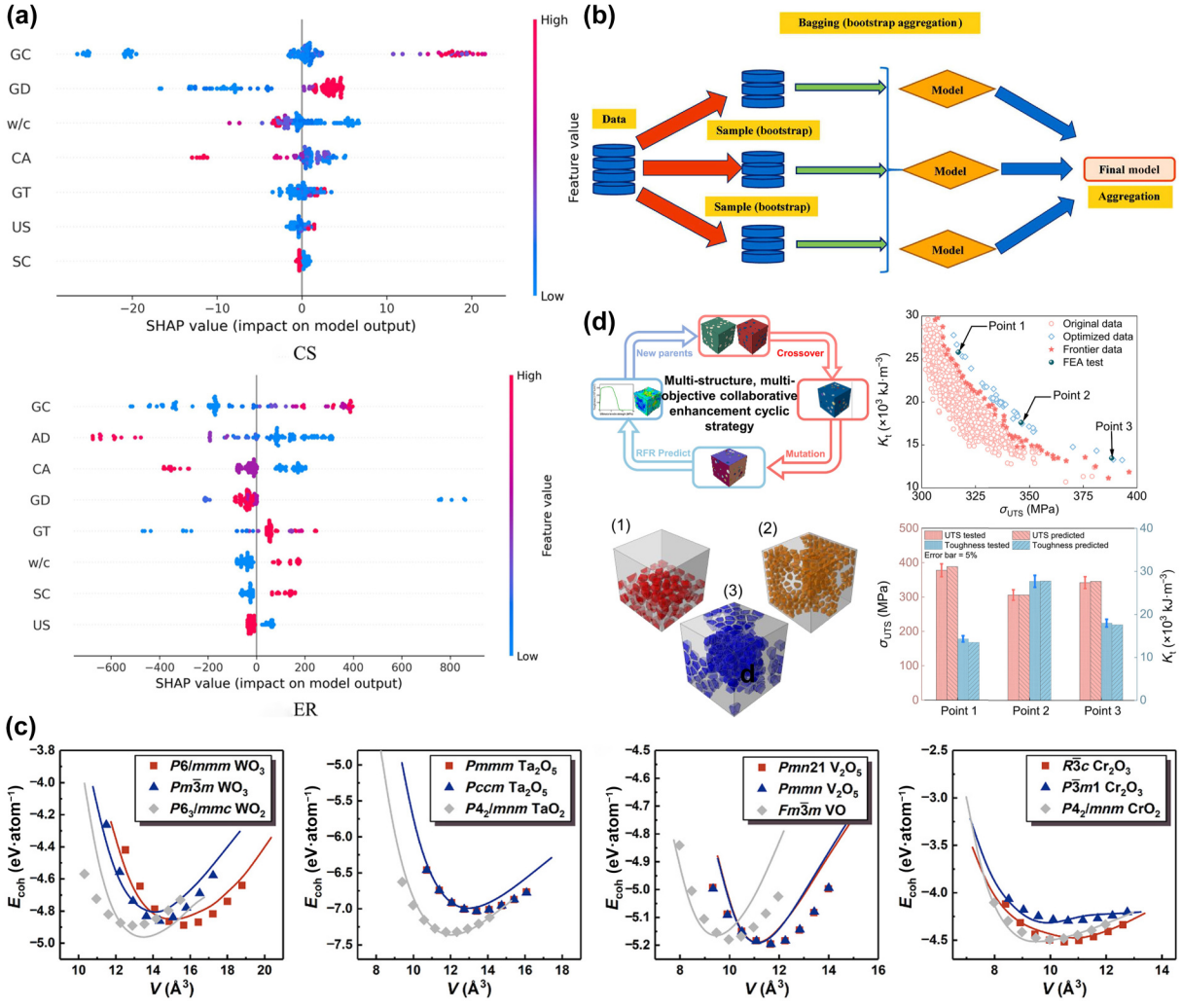

Figure 18 (a) SHAP values indicating the impact of parameters on the GEP model. Reproduced with permission from Ref. [191], © Alabduljabbar, H. et al. 2023. (b) Schematic diagram of Bagging algorithm. Reproduced with permission from Ref. [195], © Liu, G. L. et al. 2023. (c) Energy versus volume (EOS) of monometallic oxide. The line represents CTIP prediction, and the point represents DFT data. Reproduced with permission from Ref. [199], © Wu, Y. H. et al. 2024. (d) By using a multi-structure and multi-objective collaborative enhancement cycle strategy for reverse exploration, optimizing metal matrix material (MMC) and obtaining corresponding mechanical properties. Reproduced with permission from Ref. [200], © Elsevier Ltd. 2024.

system. Zhang et al. [193] utilized ANN optimization to modify ultra-high-performance concrete (UHPC) with eggshell powder nano silica, accurately capturing the interaction effects of multiple components (R^2 = 0.929). They revealed that a 10% premix increased strength by 11% through nucleation effect and pore filling, while simultaneously reducing the carbon footprint by 15%. Liu et al. [195] developed an AdaBoost integrated model (R^2 = 0.91) to quantify the dominant effects of curing time (32.7%) and nano silicon content (28.1%) on strength, and used graphical user interface (GUI) tools to enable rapid iteration of mix proportions (Fig. 18(b)).

Regarding the challenge of composite materials interface regulation, ML can be explained as the key to decoding the "nano matrix" synergistic mechanism. Meddage et al. [196] used the XGBoost model (R^2 = 0.981) combined with SHAP analysis to reveal for the first time the synergistic effect of GO particle size (optimal 43–60 μm) and polycarboxylate dispersants, and used explainable machine learning (XML) tools to visualized the component effects. Liu et al. [197] used genetic algorithm-backpropagation (GA-BP) neural network to optimize the ratio of nano-SiO_2 grafted rice straw fibers (R^2 = 0.9949), and quantified the structure–activity relationship between a 37.17% reduction in $Ca(OH)_2$ at the interface and a 40.1% improvement of wet dry cycle durability. SEM confirmed that 5–10 μm of nano coating inhibited interfacial transition zone (ITZ) weakening. This type of research collectively breaks through the limitations of traditional black box models and provides quantitative design criteria for interface engineering.

6.4.2 *Lightweight design*

By decoding the technical chain of micro phase transition mechanism, mesoscopic topology optimization, and macroscopic performance leap, the disruptive value of ML in the full chain design of "composition structure performance" of lightweight materials is systematically explained. The research evolution from single performance prediction (such as charge transfer ion potential (CTIP) strength simulation) to multi-objective collaboration (NSGA-II Pareto solution) and then to functional integration (negative Poisson's heat mechanical stability) highlights the core driving role of the intelligent driving paradigm in lightweight innovation in high-end fields such as aerospace and new energy equipment.

ML is reshaping the "atomic mesoscopic macroscopic" cross scale collaborative design paradigm of lightweight structural materials. Yao et al. [198] accurately analyzed the chemical short program evolution of Mo-Nb-Ta-W alloy using spectral neighbor analysis potential (SNAP) (error < 5%), revealing that an Nb > 25 at.% triggers the formation of B2 nano precipitates that pin edge dislocations while promoting screw dislocation cross slip. This dual mechanism reduces the slip resistance ratio by 40%, providing atomic scale design criteria for high-temperature lightweight alloys. The ML potential function strategy is deepened in complex systems. Wu et al. [199] used supervised learning to optimize CTIP for WTaVCr high entropy alloy and improved the distributed genetic algorithm (DBGA) to enhance parameter search efficiency. For the first time, they quantified the inhibitory effect of short-range chemical ordering on dislocation motion, demonstrating a 25% enhancement in strength. This work addresses the problem in parameterization of reaction potentials for multi-component systems (Fig. 18(c)).

Intelligent algorithms are innovating microstructure topology to address the bottleneck of strong resilient collaborative optimization. Zhong et al. [200] used random forest regression (R^2 > 0.92) to enhance the correlation between phase distribution and mechanical properties, and combined it with NSGA-II multi-objective optimization to obtain a Pareto solution for the "core–shell" network structure (ultimate tensile strength (UTS) > 500 MPa) (Fig. 18(d)). Finite element verification confirmed that the crack propagation path was extended by 200% and the energy absorption was tripled. Similarly, Zhao et al. [201] reconstructed the Halpin Tsai model through genetic programming (GP) and established an explicit structure–activity equation (R^2 > 0.93) for graphene origami/Cu metamaterials. This model predicted a 30% increase in modulus and demonstrated a 28,000 fold increase in computational efficiency compared to molecular dynamics at a negative Poisson's ratio of −0.59800 K with 100% H coverage, enabling a lightweight functional integrated design.

7 Breakthrough in enabling technologies

The collaborative evolution of data-driven and algorithmic innovation is reconstructing the paradigm of material research and development[139, 202–207]. This section focuses on the core technology group that supports the full-chain intelligent research and development. The data infrastructure realizes the structured integration and cross domain verification loop of multi-scale material data, whereas the algorithm innovation breaks through the bottleneck of data scarcity through physical embedding and small sample optimization, and uses explainable AI (XAI) to transform black box prediction into operable structure performance rules. Together, the two promote the transition of material design from experience trial and error to the "data model experiment" ternary driving paradigm. Their breakthrough progress lays the technical foundation for subsequent material genetic engineering and autonomous experimental systems.

7.1 Data infrastructure

The deep integration of multimodal databases with intelligent feature engineering systematically addresses the discontinuity in the "generation analysis value mining" chain of material data.

7.1.1 *Multimodal database*

Multimodal databases are emerging as core infrastructure for ML-driven material development, representing a paradigm shift from "structured data integration–machine-readable annotation–cross scale prediction closed loop". Yan et al. [208] pioneered the construction of the first open nanodatabase, PubVINAS, which transformed physical and chemical properties (1365 entries) and biological activity (2386 entries) data for 705 nanomaterials into standardized, machine-readable PDB structure files. They also defined 2142 geometric/surface/component descriptors, breaking through the unstructured limitations of traditional databases (Fig. 19(a)). This work was the first to achieve cross scale correlation modeling of "atomic structure (PDB annotation)–physical properties (zeta potential)–biological effects (cellular uptake)" (dual channel neural network R^2 > 0.5), laying a reliable data foundation for rational design of nanomaterials. Building on this paradigm, Wang et al. [15] further developed the ViNAS Pro platform, integrating PDB annotations for 14 types of nanomaterials with a full chain toolkit (including SMILES ligand

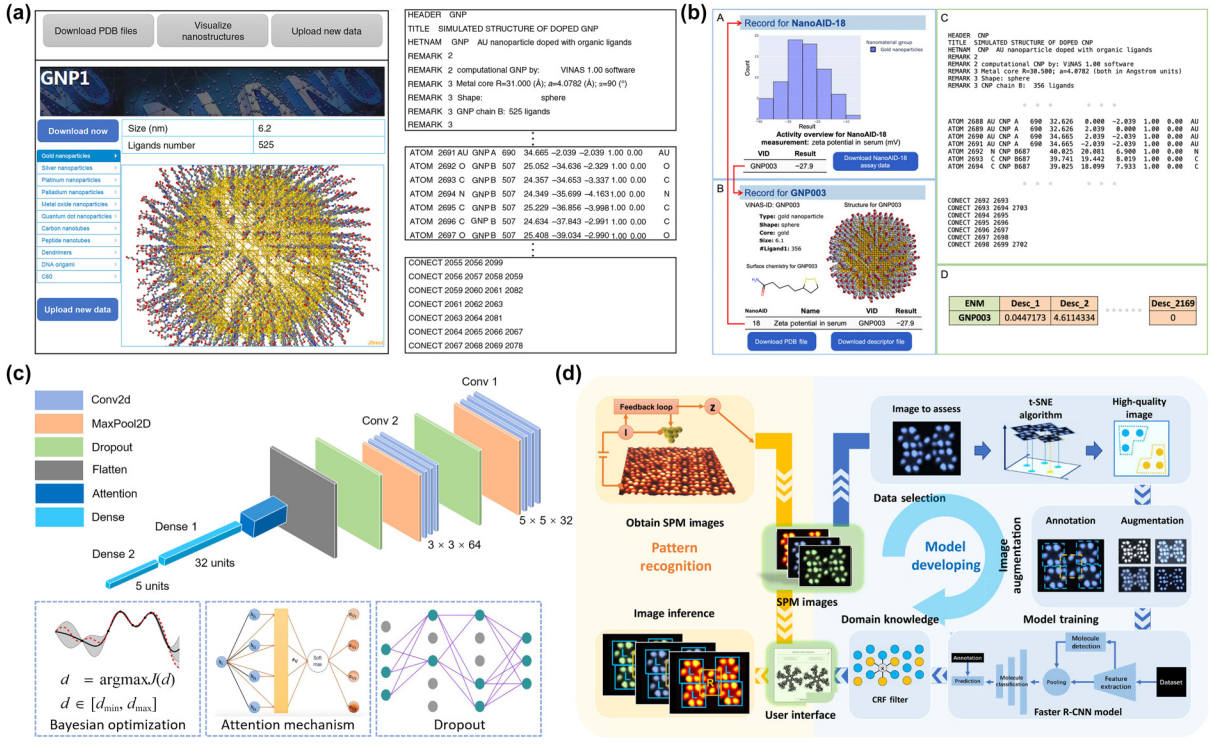

Figure 19 (a) PubVINAS online portal. Reproduced with permission from Ref. [208], © Yan, X. L. et al. 2020. (b) Integrate analysis databases and structural databases for data analysis on ViNAS Pro. Reproduced with permission from Ref. [15], © Wang, T. et al. 2024. (c) The architecture and training method of the constructed CNN model. Reproduced with permission from Ref. [164], © Liu, X. H. et al. 2024. (d) Overview of the workflow for automatic chiral molecule detection and identification. Reproduced with permission from Ref. [210], © American Chemical Society 2021.

annotation/VMD visualization/AutoNanoML modeling). This integration supports molecular dynamics simulations and virtual library generation, driving the transition of data from static storage to dynamic design (Fig. 19(b)). The standardized annotation reduces the surface energy prediction error of gold nanoparticles by 37% and improves the efficiency of calculating the Young's modulus of carbon nanotubes by 100 times.

A dedicated database forward prediction framework is accelerating the material screening revolution to meet specific domain requirements. Liu et al. [164] developed a universal framework for energy materials based on the Materials Project database, which innovatively uses self-coded Coulomb matrix descriptors instead of traditional DFT inputs (Fig. 19(c)). Combined with CNN models, it achieves the high-throughput prediction of key performance (voltage/capacity/energy density) of positive electrode materials for lithium/zinc/aluminum ion batteries (MAE ≤ 47.34 Wh·kg^{-1}). This framework screened 26 high-performance candidates out of 4300 materials, shortening the experimental validation cycle to 1/100 of that required by traditional methods, and highlighting the synergistic power of multimodal data and ML. The Coulomb matrix accurately captures the structural and performance distinctions between layered oxides (such as $LiCoO_2$) and polyanionic compounds (such as $LiFePO_4$) through the topological encoding of atomic coordinates and charges (feature importance weight > 0.8).

7.1.2 *Feature engineering framework*

At the data infrastructure level of intelligence material R&D, the innovation of feature engineering frameworks is systematically addressing the key challenges of generating, parsing, and mining value from multi-scale material data. For example, the ChIMES ML potential function developed by Lindsey et al. [209] breaks through the scale limitations of traditional DFT and constructs a feature engineering framework that can simulate > 100,000 atomic systems through high-precision fitting of quantum mechanics data. For the first time, this enables the generation of an atomic level dynamic dataset of carbon condensation processes under extreme conditions. This framework can not only capture the full pathway of non-equilibrium nanocarbon formation (such as the evolution mechanism of polymer seed clusters), but also analyzes the new maturation kinetics of "atomic free riding" through chemical bond recombination characteristics, revealing the quantitative relationship between system composition (oxygen concentration) and cluster morphology. The leap in data generation capability has laid the foundation for subsequent feature analysis, but the simulation data it relies on still need experimentally verification. In response to this, Li et al. [210] pioneered an intelligent parsing framework for experimental characterization data (Fig. 19(d)). By integrating STM imaging physical constraints with customized data augmentation techniques, a deep learning model was constructed to achieve molecular localization and chiral classification with only one single SPM image. This framework embeds domain knowledge into the feature extraction process (such as noise models and contrast perturbations), significantly improving the robustness of the model to non-ideal experimental images, solving the domain gap problem between simulated data and real scenes, and providing an *in-situ* validation tool for the nanocarbon morphology parameters predicted in Lindsey's research. Furthermore, Barnard et al.'s Shapley value analysis framework realizes a closed-loop quantification of feature importance [211]. Based on the graphene oxide energy prediction model, the contribution of structural features (O/H concentration) to the formation energy was

systematically analyzed, and it was found that extreme oxygen concentration samples dominated the model training (contribution > 30%), providing a transferable quantitative methodology for the conclusion of "composition structure sensitivity" in Lindsey's research. This framework not only confirms the physical law that every 1% change in oxygen concentration triggers a 0.8 eV energy fluctuation, but also guides future data sampling strategies through feature importance ranking, optimizing the generation efficiency of high-value feature data.

7.2 Algorithm innovation

7.2.1 Small sample learning

Small sample learning algorithms are becoming a core enabling technology for solving the scarcity dilemma of material data, with their innovation paths focusing on three directions: embedding physical prior, transferring cross domain knowledge, and optimizing active sampling. Bets et al. [212] tackled the challenge of learning periodic physical laws related to the bandgap of carbon nanotubes (Fig. 20(a)). They pioneered the use of a parameter periodic Snake activation function to accurately capture the discrete stepwise characteristics of chirality index bandgap, and achieved efficient data utilization through a two-stage transfer learning framework. First, they pre-trained the network on low precision density functional tight binding (DFTB) computation data (scale > 104) to extract periodic essence, then froze the Snake layer and fine-tuned only the ReLU layers to adapt to 137 high-precision experimental samples, maintaining excellent accuracy even when the prediction range far exceeds the training set. This work demonstrates that architecture design driven by physical mechanisms can significantly reduce the model's dependence on the target domain data volume. Further expanding to cross material systems, Hundi et al. [213] developed a microstructure feature transfer paradigm based on a pre-trained CNN model on hexagonal boron nitride (h-BN) radiation damage voxel images, and they achieved 98.7% of strength prediction accuracy with only 10% of the available graphite data. This revealed that low dimensional statistical features (including vacancy rate/atomic displacement) possess transferable predictive power across material systems. However, the presence of annotation noise in actual scenarios still constrains the robustness of small sample models. To address this, Groschner et al. [214] proposed a symmetry enhancement and noise immunity framework that utilizes the mirror symmetry of chiral nanoparticles to generate synthetic data, while adopting a co-teaching strategy for dual network mutual error correction. This approach maintains high prediction accuracy even under ≤ 40% label error rate, solving the feature space interference caused by morphological diversity. Ultimately, Kim et al. [215] advanced small sample optimization towards a closed loop using Bayesian active learning (Fig. 20(b)). By performing uncertainty quantification iteration screening on 93 high-value ligand experiments (occupying only 0.3% of the candidate space), it was found that phosphonate ligands increased the PLQY of perovskite nanocrystals by 40%, and established a volcano-shaped curve between the lowest unoccupied molecular orbital (LUMO) energy levels of ligand and defect passivation. This framework

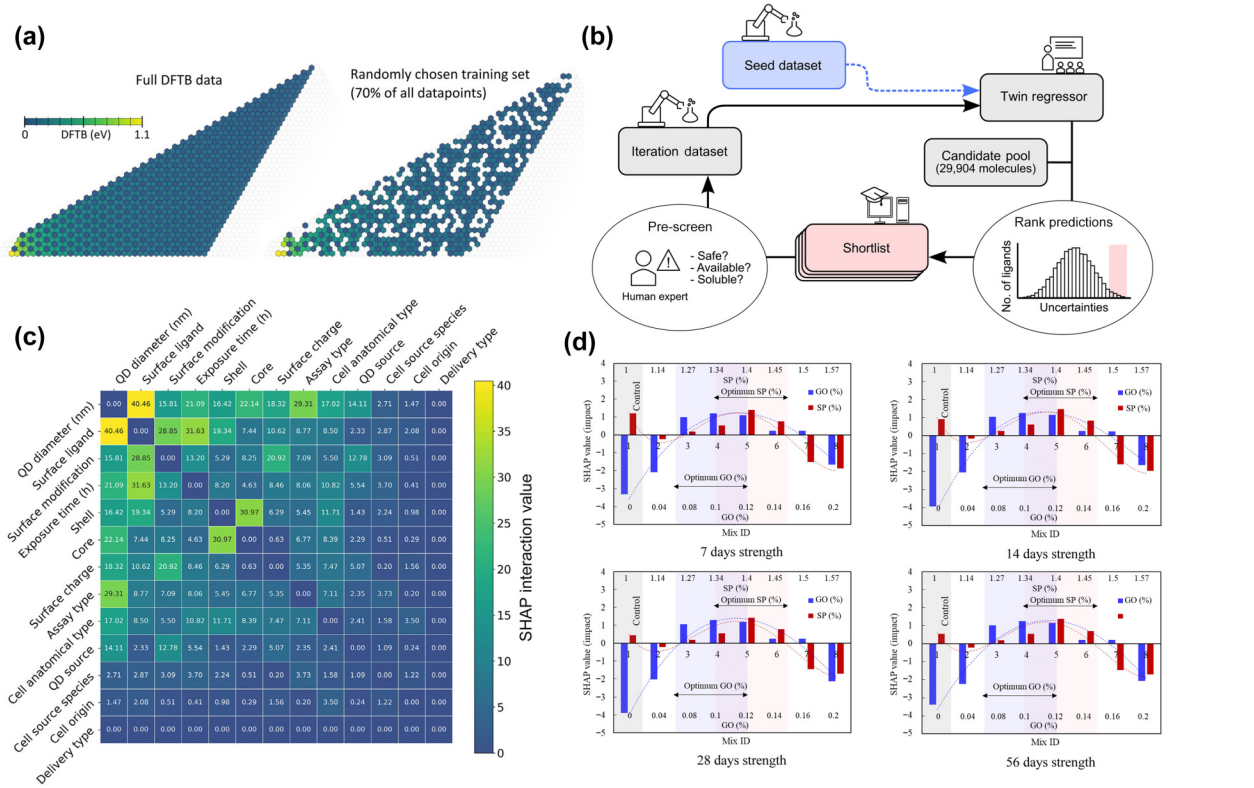

Figure 20 (a) DFTB data preparation. Reproduced with permission from Ref. [212], © Bets, K. V. et al. 2024. (b) Overview of the AI workflow for candidate selection. Reproduced with permission from Ref. [215], © American Chemical Society 2024. (c) The sum of SHAP interaction values for all instances of every two features. The feature sequence is arranged according to the sum of SHAP interaction values with other features. Reproduced with permission from Ref. [216], © Elsevier Ltd. 2021. (d) The influence and optimal dosage of graphene oxide and high-efficiency water reducing agent on compressive strength. Reproduced with permission from Ref. [196], © Meddage, D. P. P. et al. 2024.

couples data generation with model optimization, improving the sample information density from the source.

7.2.2 Interpretable AI

In recent years, XAI technology has become a core tool for solving the black box problem of material structure–activity relationship. Yu et al. [216] pioneered the "model interpretation mutual validation" framework, integrating the SHAP, partial dependence plot (PDP), and individual conditional expectation (ICE) multi-method cross validation mechanisms, quantitatively revealed for the first time the dominant factors of CdSe quantum dot cytotoxicity (zeta potential contribution > 40%), and found that there is a non-linear synergistic effect between particle size and solubility (cadmium dissolution rate increased by 300% when the particle size < 10 nm) (Fig. 20(c)). This methodology was systematically expanded in subsequent research. The same team integrated the LightGBM SHAP interpretation system with the RuleFit rule extraction algorithm for complex systems and applied it to the field of nanoagriculture [217]. Research found that when the soil clay content is greater than 30%, the absorption of metal oxide nanoparticles (MONPs) by plants decreases by 50%, and a quantifiable decision boundary of "concentration > 50 mg·kg^{-1}, absorption reduction" was generated, which significantly improved the operability of agricultural nanosafety design. In the field of building materials, Meddage et al. [196] further verified the cross material universality of XAI. Based on the XGBoost HAP framework, the "composition performance" mapping relationship was accurately analyzed, the synergistic mechanism of GO nanoparticle size of 43–60 μm and dispersant was clarified (with the maximum strength gain in the 0.08%–0.12% content range), and a visual decision-making tool was developed to achieve the reverse design of material ratios (Fig. 20(d)). It is worth noting that Gomes et al. [218] broke through the limitations of traditional statistics through symbolic regression algorithm and established a cross scale correlation model of "electronic structure biological effects". They found that Fe-doped TiO_2 inhibits the generation of reactive oxygen species under UV irradiation by regulating the bandgap (> 3.2 eV), providing a physical basis for explaining its environmental toxicity mechanism.

In response to challenges such as data heterogeneity and small sample bias, the research team developed innovative validation strategies, such as two-step cross validation (model interpretation and experimental validation), Bootstrap domain knowledge fusion validation, and feature stability analysis, which significantly improved the reliability of conclusions. The current bottleneck lies in modeling complex environmental medium effects and correcting industrial scale effects, which requires the next generation of XAI technology to deeply integrate multi-physics field mechanism models and transfer learning frameworks.

8 Challenges and limitations

8.1 Core challenges in machine learning for materials science

Although ML has shown great potential in accelerating the intelligent development of materials, covering various aspects from performance prediction to synthesis optimization, its inherent limitations must be recognized and addressed. Exploring these challenges in depth is crucial for responsible and sustainable development in this field.

8.1.1 Scarcity and quality of data

The performance of ML models largely depends on the availability of large, high-quality, and well labeled datasets. In the field of materials science, obtaining such data remains a major bottleneck. Experimental data, especially for new materials or complex synthesis processes, are often scarce, expensive, and time-consuming to generate. In addition, data from different sources may have inconsistencies due to differences in experimental protocols, equipment, and conditions. This scarcity and diversity may lead to insufficient model training, bias, or inability to capture the complex physical laws behind it. Technologies such as active learning, transfer learning, and advanced data augmentation methods specifically developed for material data are promising ways to alleviate this problem.

8.1.2 Transferability and generalization ability of the model

Models trained on specific datasets (such as perovskite solar cells) often perform poorly when applied to different but related material systems (such as organic photovoltaic devices). The reason for this lack of transferability is that ML models can only learn superficial statistical correlations from training data, which do not represent universal physical principles. Therefore, they may not be able to make inferences beyond the chemical or structural space represented by their training set. To improve generalization ability, it is necessary to incorporate domain knowledge and physical constraints into the model architecture (e.g., using physically constrained neural networks). Developing robust and interpretable models that can elucidate their reasoning process is crucial for building trust and ensuring that predictions have physical significance.

8.1.3 High computational costs

The development of state-of-the-art machine learning models, especially deep learning architectures with millions of parameters, requires a significant amount of computational resources. The process of hyperparameter tuning, training, and validation typically requires days or weeks of processing on a high-performance computing (HPC) cluster equipped with graphics processing units (GPU)/sensor processing units (TPU). This high computational cost may pose significant entry barriers for research teams with limited resources and slow down the iterative research cycle. Future work must focus on developing more efficient computational algorithms and model compression techniques to make ML tools more accessible to a wider range of materials fields.

8.1.4 The mystery of energy footprints in green synthesis and ML

In the interdisciplinary field of ML and green material synthesis, an interesting and challenging problem has emerged. The original intention of developing energy-saving or environmentally friendly materials, which is to use less energy to manufacture these materials, may be undermined by the enormous energy consumption of training large-scale ML models. The carbon footprint of training a large model may be equivalent to the lifetime emissions of several cars. This creates a paradox: We use an energy intensive tool to design low-energy materials. To address this issue, the field must prioritize model efficiency, hardware and infrastructure, algorithm optimization, and cost-benefit analysis.

8.1.5 Ethical issues and reproducibility

Although machine learning has brought transformative potential to

materials research, its responsible application requires addressing key ethical issues such as data privacy and algorithmic bias, while ensuring reproducibility through transparent code and data sharing. To reduce privacy risks and bias, strategies such as federated learning, differential privacy, diverse data management, and interpretable artificial intelligence technologies can be adopted. In addition, following findability, accessibility, interoperability, and reuse (FAIR) data principles, publicly releasing code, and benchmarking with the community are crucial for improving reproducibility, enhancing trust, and promoting fair and widespread application.

To address these challenges, the conclusion of this section requires the joint efforts of computer scientists, material scientists, and ethicists. In the field of materials science, the future of ML does not lie in replacing specialized knowledge, but in establishing a symbiotic relationship between data-driven discovery and fundamental physical principles. By publicly acknowledging and striving to overcome these limitations, we can drive the field towards a stronger, more replicable, and environmentally sustainable future.

8.2 Data-centric challenges and quality enhancement strategies

Although machine learning has shown great potential in accelerating the intelligent development of materials, its performance is highly dependent on the quality, quantity, and representativeness of the underlying data. The main challenges include data bias, annotation noise, and sample imbalance, which can seriously compromise the reliability and generality of the model. Below, we will discuss these issues and propose corresponding mitigation strategies.

8.2.1 Data bias

When the training data cannot represent the complete chemical or structural space, data bias occurs, which is usually caused by preferences in experimental conditions, material systems, or synthesis methods. For example, datasets may lean towards materials that have been thoroughly studied, such as perovskites or precious metal nanoparticles, whereas systems that have not been thoroughly studied, such as multi-component oxides or soft materials, may lead to overfitting of results. Possible solutions include activity-based learning, transfer learning, and domain adaptation.

8.2.2 Annotate noise

Labeling noise refers to errors in the labeling process, which may arise from subjective interpretation, instrument limitations, or inconsistent experimental procedures. For example, size measurement or spectral interpretation based on TEM may vary depending on the operator. Robust learning frameworks, data augmentation, and quantification of prediction uncertainty can be adopted as solutions.

8.2.3 Sample imbalance

When the frequency of certain categories or attributes in the dataset is too high or too low, sample imbalance may occur. For example, high-performance materials or rare failure modes may have a limited number of examples, leading to deviations in model performance. Possible solutions include resampling techniques, cost sensitive learning, and synthetic data generation.

8.2.4 Overall data quality improvement

To achieve reliable and scalable nanostructure data analysis, we propose the following key methods: Firstly, adopting a standardized data protocol is unified across the entire community, using ViNAS Pro tool to implement PDB like nanostructure encoding and annotation, and ensuring consistency in data format and semantics. Secondly, XAI technology is introduced to identify and correct biases in feature contributions using methods such as SHAP, local interpretable model agnostic explanations (LIME), or rule extraction, thereby improving model transparency and reliability. Finally, promoting collaborative review mechanisms, integrating high-quality and diverse datasets while protecting privacy through open data initiatives and federated learning frameworks, and facilitating cross institutional collaboration and data sharing.

9 Conclusions and future outlook

ML-driven intelligent material research and development has formed a full chain technology system that runs through "design synthesis characterization application", and its core breakthroughs are reflected in four dimensions.

(1) Innovation of the intelligent design paradigm

The emergence of advance dreverse design system constructed through reinforcement learning (such as protein conformation optimization) and evolutionary algorithms (such as reverse screening of multi-enzyme solid solutions) has enabled precise mapping from target performance goals to synthesis parameters. Generative models (such as ViNAS Pro) have constructed machine-readable virtual material libraries when coupled with cross scale modeling (such as PLIP potential function coupled PXRDnet structure reconstruction), successfully breaking the bottleneck of correlating "atomic nucleation, cluster evolution, and mesoscopic performance". However, the controllable generation of complex heterogeneous interfaces, such as core–shell structures, still relies on costly experimental verification.

(2) Autonomous closed-loop synthesis process

Microfluidics Bayesian optimization platforms (such as "artificial chemist" system) have achieved multi-objective collaborative regulation (such as quantum dot bandgap/quantum yield), and the single experimental cycle has been shortened to 2 min. Green preparation technology relies on multi-objective optimization (such as Pareto frontier equilibrium yield toxicity) and small sample strategies (such as transfer learning reusing MOF databases) to achieve a spatiotemporal yield of 62.4 $kg \cdot m^{-3} \cdot day^{-1}$ for antibacterial ZnO nanoparticles. However, the integration of high-temperature and high-pressure reactions and the prediction of metastable phase kinetics remain key challenges.

(3) Significant improvement in representational and analytical abilities

In-situ ML technology combinations (such as U-Net analysis of liquid phase TEM video) reveal dynamic mechanisms (such as activation energy of 12 kT for nanocubes). Intelligent morphology analysis (such as AuNR SMA spectral inversion) extracts geometric parameters directly from the optical response (size error < 5%), supporting synthesis closed-loop optimization. The current bottleneck lies in the limited ability of cross scale monitoring non-equilibrium processes, such as the real-time evolution of protein coronations.

(4) Application scenario closed-loop implementation

In the field of energy, ML potential functions (such as SNAP) have guided the design of high-temperature lightweight alloys

(Nb > 25 at.% triggers B2 precipitation phase), achieving a threefold increase in crack energy absorption through multi-objective optimization of strong toughness synergy. In biomedical applications, the GraphBNC framework predicts dynamic nano protein binding sites, whereas the LungVis platform quantifies alveolar targeting efficiency (B/A ratio of 0.07). In environmental governance, XGBoost-optimized GO@ZIF-7 adsorbents achieve a Pb^{2+} removal rate of 98%, and the ANN model guides a 15% reduction in carbon footprint of eggshell powder UHPC. For structural materials, genetic programming has reconstructed structure performance equations of graphene origami/Cu metamaterials, achieving synergy between negative Poisson's ratio (−0.59) and thermal mechanical stability.

Future technological breakthrough directions are:

(1) Dynamic cross-scale modeling

Future research should focus on developing physical information dynamic graph neural networks coupled with *in-situ* characterization data such as liquid-phase atomic force microscopy, to capture real-time interface evolution (such as catalytic bond cleavage and protein crown reconstruction). This requires constructing multi-physics coupled ML framework that quantifies cross effects of light, electricity, and heat force, with target performance such as perovskite LED junction temperature prediction error < 1.5 ℃.

(2) Algorithm hardware collaborative innovation

Next-generation breakthroughs demand the development of adaptive federated learning to address privacy issues in medical/industrial data, combined with small sample strategies such as symmetry enhanced synthetic data to improve noise robustness. Parallelly, it is advantageous to promote high-precision potential function training, which can accelerate non-equilibrium state simulation of ten thousand atom systems using quantum computing ML integration.

(3) Full chain intelligent system integration

It is necessary to build a "digital twin" platform that integrates ViNAS Pro structure generation, closed-loop autonomous synthesis robots, and *in-situ* characterization feedback to achieve design validation manufacturing integration. In parallel, developing generative AI-driven reverse design (such as diffusion models for generating novel materials with negative thermal expansion coefficients < -10^{-5}·K^{-1}) will further accelerate the discovery of advanced substances.

(4) Ethical and standardization challenges

The establishment of a nano safety prediction framework (such as quantitative cross reading algorithm to evaluate biological toxicity) can promote the extension of PDB annotations to international standards for multiphase soft materials. Furthermore, developing an online monitoring and self-calibration system, together with reinforcement learning based optimization of manufacturing parameters (such as nano dispersion uniformity control), will help address batch differences during industrial scaling up.

Data availability

Not applicable.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Nos. NSFC-22509178 and 52371240), the Changjiang Scholars Program of the Ministry of Education (No. Q2018270), China Postdoctoral Science Foundation (No. 2025M770225), and Yangzhou Innovation Capability Enhancement Program (No. YZ2022170).

Declaration of competing interest

All the contributing authors report no conflict of interests in this work.

Author contribution statement

H. J. Z., H. P., and J. N. D.: Conceiving the idea. J. M. X. and X. Y. Q.: Collecting the data and drafting the manuscript. H. J. Z., J. Z., J. X., and L. Z. L.: Revising the manuscript. X. Y. Q., M. S., and W. J. Z.: Revising the grammar and typesetting the images. All authors read and commented on the manuscript.

Use of AI statement

None.

References

[1] Zhang, Y. F.; Li, Q.; Feng, W. C.; Gao, S. J.; Yue, H. T.; Su, Y. C.; Zhou, H. J.; Huang, J. F.; Han, L. F.; Shakouri, M. et al. Regulating electron transfer in vanadium-based metal-organic frameworks via the synergy of linker engineering and machine learning for efficient and reversible aqueous zinc ion batteries. *Adv. Mater.* **2025**, *37*, 2507609.

[2] Dhoble, S.; Wu, T. H.; Kenry. Decoding nanomaterial-biosystem interactions through machine learning. *Angew. Chem., Int. Ed.* **2024**, *63*, e202318380.

[3] Li, Q.; Zhang, Y. F.; Feng, W. C.; Huang, J. F.; Wei, S. X.; Chen, G.; Liu, Y. W.; Du, M.; Yin, C. H.; Yang, Z. B. et al. Manganese-based metal-organic coordination for aqueous zinc-ion batteries with varying mechanical adaptability and machine learning-assisted performance decoding. *Adv. Mater.* **2025**, *37*, 2507951.

[4] Winkler, D. A. Role of artificial intelligence and machine learning in nanosafety. *Small* **2020**, *16*, 2001883.

[5] Wang, M.; Wang, T.; Cai, P. Q.; Chen, X. D. Nanomaterials discovery and design through machine learning. *Small Methods* **2019**, *3*, 1900025.

[6] Wang, Z. Z.; Ni, J. F.; Li, L.; Lu, J. Theoretical simulation and modeling of three-dimensional batteries. *Cell Rep. Phys. Sci.* **2020**, *1*, 100078.

[7] Zhu, S.; Ni, J. F.; Li, Y. Carbon nanotube-based electrodes for flexible supercapacitors. *Nano Res.* **2020**, *13*, 1825–1841.

[8] Chu, T. S.; Zhou, L.; Zhang, B. W.; Xuan, F. Z. Accurate atomic scanning transmission electron microscopy analysis enabled by deep learning. *Nano Res.* **2024**, *17*, 2971–2980.

[9] Yang, L.; Wang, H.; Leng, D. Y.; Fang, S. P.; Yang, Y. N.; Du, Y. R. Machine learning applications in nanomaterials: Recent advances and future perspectives. *Chem. Eng. J.* **2024**, *500*, 156687.

[10] Li, Q.; Zhang, Y. F.; Guo, X. T.; Yang, Z. B.; Wang, Y. X.; Chen, Y. M.; Liu, Y. W.; Yue, H. T.; Gao, S. J.; Zhou, H. J. et al. Nucleation and growth mechanisms of micro/nano structural manganese-trimesic acid coordinations for aqueous zinc-ion batteries. *Angew. Chem., Int. Ed.* **2025**, *64*, e202509741.

[11] Bi, P.; Zhang, M. C.; Li, S.; Lu, H. J.; Wang, H. M.; Liang, X. P.; Liang, H. R.; Zhang, Y. Y. Ultra-sensitive and wide applicable strain sensor enabled by carbon nanofibers with dual alignment for human machine interfaces. *Nano Res.* **2023**, *16*, 4093–4099.

[12] Chen, L.; Chen, S. Y.; Wu, J. C.; Chen, L. H.; Yang, S.; Chu, J.; Jiang, C. M.; Bi, S.; Song, J. H. A two-dimensional MoS_2 array based on artificial neural network learning for high-quality imaging. *Nano Res.* **2023**, *16*, 10139–10147.

[13] Li, W. W.; Duan, C. B.; Wei, Y.; Xu, H. Advancements in flexible memristors for neuromorphic computing: Materials, mechanisms, and applications in synaptic emulation. *FlexMat.* **2025**, *2*, 390–419.

[14] Lutz, I. D.; Wang, S. Z.; Norn, C.; Courbet, A.; Borst, A. J.; Zhao, Y. T.; Dosey, A.; Cao, L. X.; Xu, J. W.; Leaf, E. M. et al. Top-down design of protein architectures with reinforcement learning. *Science* **2023**, *380*, 266–273.

[15] Wang, T.; Russo, D. P.; Demokritou, P.; Jia, X. L.; Huang, H.; Yang, X. Y.; Zhu, H. An online nanoinformatics platform empowering computational modeling of nanomaterials by nanostructure annotations and machine learning toolkits. *Nano Lett.* **2024**, *24*, 10228–10236.

[16] Zhang, G. X.; Li, Y.; Du, G. Y.; Lu, J. Q.; Wang, Q. J.; Wu, K.; Zhang, S. T.; Chen, H. Y.; Zhang, Y. Z.; Xue, H. G. et al. Spiral-concave Prussian blue crystals with rich steps: Growth mechanism and coordination regulation. *Angew. Chem., Int. Ed.* **2025**, *64*, e202414650.

[17] Salazar, C. R.; Ammothum Kandy, A. K.; Furstoss, J.; Gromoff, Q.; Goniakowski, J.; Lam, J. Competing nucleation pathways in nanocrystal formation. *npj Comput. Mater.* **2024**, *10*, 199.

[18] Diao, S. H.; Wu, Q.; Li, S. M.; Xu, G. C.; Ren, X. L.; Tan, L. F.; Jiang, G. H.; Song, P.; Meng, X. W. From synthesis to properties: Expanding the horizons of machine learning in nanomaterials research. *Mater. Horiz.* **2025**, *12*, 4133–4164.

[19] Zhang, G. X.; Feng, W. C.; Du, G. Y.; Zhang, Y.; Yang, Y.; Xu, D.; Wang, T. Y.; Chen, H. Y.; Xue, H. G.; Shakouri, M. et al. Thermodynamically-driven phase engineering and reconstruction deduction of medium-entropy Prussian blue analogue nanocrystals. *Adv. Mater.* **2025**, *37*, 2503814.

[20] Duan, S. S.; Wang, J. Y.; Lin, Y.; Hong, J. L.; Lin, Y. C.; Xia, Y. E.; Li, Y. H.; Zhu, D.; Lei, W.; Su, W. M. et al. Highly durable machine-learned waterproof electronic glove based on low-cost thermal transfer printing for amphibious wearable applications. *Nano Res.* **2023**, *16*, 5480–5489.

[21] Feng, F.; Ou, Z. P.; Zhang, F. D.; Chen, J. X.; Huang, J. K.; Wang, J. X.; Zuo, H. Q.; Zeng, J. B. Artificial intelligence-assisted colorimetry for urine glucose detection towards enhanced sensitivity, accuracy, resolution, and anti-illuminating capability. *Nano Res.* **2023**, *16*, 12084–12091.

[22] Epps, R. W.; Bowen, M. S.; Volk, A. A.; Abdel-Latif, K.; Han, S. Y.; Reyes, K. G.; Amassian, A.; Abolhasani, M. Artificial chemist: An autonomous quantum dot synthesis bot. *Adv. Mater.* **2020**, *32*, 2001626.

[23] Park, Y. R.; Wang, G. Learning-effective mixed-dimensional halide perovskite QD synaptic array for self-rectifying and luminous artificial neural networks. *Adv. Funct. Mater.* **2024**, *34*, 2307971.

[24] Huang, F. Y.; Chen, L.; Zeng, Y.; Dai, W.; Wu, F.; Hu, Q.; Zhou, Y.; Shi, S. M.; Fang, L. C. Unveiling influences of metal-based nanomaterials on wheat growth and physiology: From benefits to detriments. *Chemosphere* **2024**, *364*, 143212.

[25] Allara, L.; Bertolotti, F.; Guagliardi, A. A deep learning approach for quantum dots sizing from wide-angle X-ray scattering data. *npj Comput. Mater.* **2024**, *10*, 54.

[26] Hu, H. H.; Jiang, W.; Han, X.; Wu, G.; Wang, H. R.; Shi, Y.; He, D. Y.; Ma, X. H.; Hong, X. Programmable self-assembly from two-dimensional nanosheets to spiral, twisted and branched nanostructures. *Angew. Chem., Int. Ed.* **2025**, *64*, e202416624.

[27] Rangel DaCosta, L.; Sytwu, K.; Groschner, C. K.; Scott, M. C. A robust synthetic data generation framework for machine learning in high-resolution transmission electron microscopy (HRTEM). *npj Comput. Mater.* **2024**, *10*, 165.

[28] Kløve, M.; Sommer, S.; Iversen, B. B.; Hammer, B.; Dononelli, W. A machine-learning-based approach for solving atomic structures of nanomaterials combining pair distribution functions with density functional theory. *Adv. Mater.* **2023**, *35*, 2208220.

[29] Dahl, J. C.; Niblett, S.; Cho, Y.; Wang, X. Z.; Zhang, Y.; Chan, E. M.; Alivisatos, A. P. Scientific machine learning of 2D perovskite nanosheet formation. *J. Am. Chem. Soc.* **2023**, *145*, 23076–23087.

[30] Yang, Y.; Shi, C. Q.; Feijóo, J.; Jin, J. B.; Chen, C. B.; Han, Y. M.; Yang, P. D. Dynamic evolution of copper nanowires during CO_2 reduction probed by *operando* electrochemical 4D-STEM and X-ray spectroscopy. *J. Am. Chem. Soc.* **2024**, *146*, 23398–23405.

[31] Ahmad, F.; Mahmood, A.; Muhmood, T. Machine learning-integrated omics for the risk and safety assessment of nanomaterials. *Biomater. Sci.* **2021**, *9*, 1598–1608.

[32] Kuznetsova, V.; Coogan, Á.; Botov, D.; Gromova, Y.; Ushakova, E. V.; Gun'ko, Y. K. Expanding the horizons of machine learning in nanomaterials to chiral nanostructures. *Adv. Mater.* **2024**, *36*, 2308912.

[33] Yang, H. T.; Li, J. L.; Lim, K. Z.; Pan, C. J.; Van Truong, T.; Wang, Q.; Li, K. R.; Li, S.; Xiao, X.; Ding, M. et al. Automatic strain sensor design via active learning and data augmentation for soft machines. *Nat. Mach. Intell.* **2022**, *4*, 84–94.

[34] Fu, L. L.; Hu, H. X.; Zhu, Q.; Zheng, L. F.; Gu, Y. M.; Wen, Y. P.; Ma, H. B.; Yin, H.; Ma, J. Machine learning assisted prediction of charge transfer properties in organic solar cells by using morphology-related descriptors. *Nano Res.* **2023**, *16*, 3588–3596.

[35] Kouchehbaghi, N. H.; Yousefzadeh, M.; Gharehaghaji, A.; Khosravi, S.; Khorsandi, D.; Haghniaz, R.; Cao, K.; Dokmeci, M. R.; Rostami, M.; Khademhosseini, A. et al. A machine learning-guided design and manufacturing of wearable nanofibrous acoustic energy harvesters. *Nano Res.* **2024**, *17*, 9181–9192.

[36] Wang, L. G.; Wu, J. B.; Wang, S. W.; Liu, H.; Wang, Y.; Wang, D. S. The reformation of catalyst: From a trial-and-error synthesis to rational design. *Nano Res.* **2024**, *17*, 3261–3301.

[37] Bang, K.; Hong, D.; Park, Y.; Kim, D.; Han, S. S.; Lee, H. M. Machine learning-enabled exploration of the electrochemical stability of real-scale metallic nanoparticles. *Nat. Commun.* **2023**, *14*, 3004.

[38] Isac, D. L.; Jalba, R. I.; Soriga, S. G.; Zhao, Y. H.; Tielens, F.; Tranca, I.; Isabela-Costinela, M. Computational investigation of the oxygen reduction reaction on the edges of differently-sized, shaped and terminated graphene nanoclusters. *Carbon* **2024**, *222*, 118942.

[39] Ferdosi, S.; Stukalov, A.; Hasan, M.; Tangeysh, B.; Brown, T. R.; Wang, T. Y.; Elgierari, E. M.; Zhao, X. Y.; Huang, Y. X.; Alavi, A. et al. Enhanced competition at the nano-bio interface enables comprehensive characterization of protein corona dynamics and deep coverage of proteomes. *Adv. Mater.* **2022**, *34*, 2206008.

[40] Wan, K. W.; He, J. X.; Shi, X. H. Construction of high accuracy machine learning interatomic potential for surface/interface of nanomaterials-a review. *Adv. Mater.* **2024**, *36*, 2305758.

[41] Miao, R. P.; Bissoli, M.; Basagni, A.; Marotta, E.; Corni, S.; Amendola, V. Data-driven predetermination of Cu oxidation state in copper nanoparticles: Application to the synthesis by laser ablation in liquid. *J. Am. Chem. Soc.* **2023**, *145*, 25737–25752.

[42] Suwardi, A.; Wang, F. K.; Xue, K.; Han, M. Y.; Teo, P.; Wang, P.; Wang, S. J.; Liu, Y.; Ye, E. Y.; Li, Z. B. et al. Machine learning-driven biomaterials evolution. *Adv. Mater.* **2022**, *34*, 2102703.

[43] Zhang, H. Y.; Yang, M.; Wu, Q. Y.; Xue, J. J.; Liu, H. Y. Engineering two-dimensional nanomaterials for photothermal therapy. *Angew. Chem., Int. Ed.* **2025**, *64*, e202424768.

[44] Yu, H. J.; Zhao, Z. L.; Liu, D.; Cheng, F. Integrating machine learning interpretation methods for investigating nanoparticle uptake during seed priming and its biological effects. *Nanoscale* **2022**, *14*, 15305–15315.

[45] Ji, Y. Z.; Ma, S.; Lv, S. Q.; Wang, Y. J.; Lü, S. Y.; Liu, M. Z. Nanomaterials for targeted delivery of agrochemicals by an all-in-

one combination strategy and deep learning. *ACS Appl. Mater. Interfaces* **2021**, *13*, 43374–43386.
[46] He, Y.; Liu, F.; Min, W. C.; Liu, G. H.; Wu, Y. B.; Wang, Y.; Yan, X. L.; Yan, B. De novo design of biocompatible nanomaterials using quasi-SMILES and recurrent neural networks. *ACS Appl. Mater. Interfaces* **2024**, *16*, 66367–66376.
[47] Yang, X. Z.; Liu, X.; Zhang, A. Q.; Lu, D. W.; Li, G.; Zhang, Q. H.; Liu, Q.; Jiang, G. B. Distinguishing the sources of silica nanoparticles by dual isotopic fingerprinting and machine learning. *Nat. Commun.* **2019**, *10*, 1620.
[48] Yue, X. Y; Li, Q. T.; Wang, Z. Q.; Duan, L. M. H.; Yang, W. K.; Pan, D.; Liu, H.; Liu, C. T.; Shen, C. Y. Machine learning facilitated gesture recognition using structural optimized wearable yarn-based strain sensor. *Nano Res.*, in press, https://doi.org/10.26599/NR.2025.94908084.
[49] Kuo, Y. A.; Jung, C.; Chen, Y. A.; Kuo, H. C.; Zhao, O. S.; Nguyen, T. D.; Rybarski, J. R.; Hong, S.; Chen, Y. I.; Wylie, D. C. et al. Massively parallel selection of nanocluster beacons. *Adv. Mater.* **2022**, *34*, 2204957.
[50] Rao, L. M.; Lu, X. Y.; Xu, L. L.; Zhu, Y. F.; Xue, T.; Ge, Y.; Duan, Z. S.; Duan, X. M.; Wen, Y. P.; Xu, J. K. Green synthesis of kudzu vine biochar decorated graphene-like $MoSe_2$ with the oxidase-like activity as intelligent nanozyme sensing platform for hesperetin. *Chemosphere* **2022**, *289*, 133116.
[51] Ceballos-Sanchez, O.; Navarro-López, D. E.; Mejía-Méndez, J. L.; Sanchez-Ante, G.; Rodríguez-González, V.; Sánchez-López, A. L.; Sanchez-Martinez, A.; Duron-Torres, S. M.; Juarez-Moreno, K.; Tiwari, N. et al. Enhancing antioxidant properties of CeO_2 nanoparticles with Nd^{3+} doping: Structural, biological, and machine learning insights. *Biomater. Sci.* **2024**, *12*, 2108–2120.
[52] Hejazi, D.; Liu, S. J.; Farnoosh, A.; Ostadabbas, S.; Kar, S. Development of use-specific high-performance cyber-nanomaterial optical detectors by effective choice of machine learning algorithms. *Mach. Learn.: Sci. Technol.* **2020**, *1*, 025007.
[53] Jin, K.; Wang, W. T.; Qi, G. P.; Peng, X. H.; Gao, H. N.; Zhu, H. J.; He, X.; Zou, H. X.; Yang, L.; Yuan, J. J. et al. An explainable machine-learning approach for revealing the complex synthesis path–property relationships of nanomaterials. *Nanoscale* **2023**, *15*, 15358–15367.
[54] Jintoku, H.; Futaba, D. N. Machine learning-assisted exploration and identification of aqueous dispersants in the vast diversity of organic chemicals. *ACS Appl. Mater. Interfaces* **2024**, *16*, 11800–11808.
[55] Kenry. Machine-learning-guided quantitative delineation of cell morphological features and responses to nanomaterials. *Nanoscale* **2024**, *16*, 19656–19668.
[56] Hassanien, A. E.; Abouelmagd, L. M.; Mahmoud, A. S.; Darwish, A. An optimized backpropagation neural network models for the prediction of nanomaterials concentration for purification industrial wastewater. *Eng. Appl. Artif. Intell.* **2023**, *126*, 107010.
[57] Chen, T. K.; Li, J. L.; Cai, P. F.; Yao, Q. F.; Ren, Z. K.; Zhu, Y. X.; Khan, S.; Xie, J. P.; Wang, X. N. Identification of chemical compositions from "featureless" optical absorption spectra: Machine learning predictions and experimental validations. *Nano Res.* **2023**, *16*, 4188–4196.
[58] Dang, Y.; Reddy, Y. V. M.; Cheffena, M. Facile E-nose based on single antenna and graphene oxide for sensing volatile organic compound gases with ultrahigh selectivity and accuracy. *Sens. Actuators B: Chem.* **2024**, *419*, 136409.
[59] Zhang, H. Z.; Fichthorn, K. A. Structural classification of Ag and Cu nanocrystals with machine learning. *Nanoscale* **2024**, *16*, 17154–17164.
[60] Zhang, M. Y.; He, H. J.; Huang, Y. Q.; Huang, R. F.; Wu, Z.; Liu, X. M.; Deng, H. Machine learning integrated high quantum yield blue light carbon dots for real-time and on-site detection of Cr(VI) in groundwater and drinking water. *Sci. Total Environ.* **2023**, *904*, 166822.
[61] Zhang, J.; Perrin, M. L.; Barba, L.; Overbeck, J.; Jung, S.; Grassy, B.; Agal, A.; Muff, R.; Brönnimann, R.; Haluska, M. et al. High-speed identification of suspended carbon nanotubes using Raman spectroscopy and deep learning. *Microsyst. Nanoeng.* **2022**, *8*, 19.
[62] Dai, Y. K.; Zhang, Z. W.; Wang, D.; Li, T. L.; Ren, Y. Z.; Chen, J. Q.; Feng, L. Y. Machine-learning-driven G-quartet-based circularly polarized luminescence materials. *Adv. Mater.* **2024**, *36*, 2310455.
[63] Jiang, Y. J.; Chen, Z. B.; Sui, N.; Zhu, Z. L. Data-driven evolutionary design of multienzyme-like nanozymes. *J. Am. Chem. Soc.* **2024**, *146*, 7565–7574.
[64] Li, S. C.; Barnard, A. S. Safety-by-design using forward and inverse multi-target machine learning. *Chemosphere* **2022**, *303*, 135033.
[65] Cui, L. Y.; Li, Q.; Zhang, Y. C.; Zhang, J.; Wang, Z.; Chen, J. K.; Zheng, B. Machine learning-assisted high-throughput screening of transparent organic light-emitting diode anode materials. *Chem. Sci.* **2024**, *15*, 19375–19389.
[66] Kioumourtzoglou, S.; Hof, S.; Kalk, C.; Toth, V.; Görlin, M.; Nováková, J.; Sá, J. Nanomaterials as a service (NaaS) concept: On-demand protocols for volume synthesis of nanomaterials. *Nanoscale Horiz.* **2024**, *9*, 1364–1371.
[67] Guo, G. B.; Saidi, T. L.; Terban, M. W.; Valsecchi, M.; Billinge, S. J. L.; Lipson, H. *Ab initio* structure solutions from nanocrystalline powder diffraction data via diffusion models. *Nat. Mater.*, in press, https://doi.org/10.1038/s41563-025-02220-y.
[68] Anker, A. S.; Kjær, E. T. S.; Juelsholt, M.; Christiansen, T. L.; Skjærvø, S. L.; Jørgensen, M. R. V.; Kantor, I.; Sørensen, D. R.; Billinge, S. J. L.; Selvan, R. et al. Extracting structural motifs from pair distribution function data of nanostructures using explainable machine learning. *npj Comput. Mater.* **2022**, *8*, 213.
[69] Chen, Q.; Chatzigeorgiou, G.; Meraghni, F.; Chen, X. F.; Yang, Z. B. Physics-informed deep homogenization approach for random nanoporous composites with energetic interfaces. *Eng. Appl. Artif. Intell.* **2025**, *140*, 109735.
[70] Gupta, N.; Jayaraman, A. Computational approach for structure generation of anisotropic particles (CASGAP) with targeted distributions of particle design and orientational order. *Nanoscale* **2023**, *15*, 14958–14970.
[71] Choi, J. H.; Jang, W.; Lim, Y. J.; Mun, S. J.; Bong, K. W. Highly flexible deep-learning-based automatic analysis for graphically encoded hydrogel microparticles. *ACS Sens.* **2023**, *8*, 3158–3166.
[72] Xi, Y.; Cheng, S. J.; Chao, S. Y.; Hu, Y. R.; Cai, M. S.; Zou, Y.; Liu, Z.; Hua, W.; Tan, P. C.; Fan, Y. B. et al. Piezoelectric wearable atrial fibrillation prediction wristband enabled by machine learning and hydrogel affinity. *Nano Res.* **2023**, *16*, 11674–11681.
[73] Zhou, N. N.; Ao, H. R.; Chen, X. M.; Gao, S.; Jiang, H. Y. A vector hybrid triboelectric sensor (HTS) for motion identification via machine learning. *Nano Res.* **2023**, *16*, 10120–10130.
[74] Jiang, T. Z.; Gai, S. L.; Yin, Y. Q.; Sun, Z. W.; Zhou, B. C.; Zhao, Y. B.; Ding, H.; Ahmad Ansari, A.; Yang, P. P. A light/thermal cascaded-driven equipment for machine recognition inspired by water lilies using as multifunctional soft actuator. *Chem. Eng. J.* **2024**, *495*, 153348.
[75] Abay, Z. M.; Wei, Y. X.; Tang, Z. G.; Liu, Y. G.; Chen, K.; Li, H.; Wang, Q. C.; Yuan, J.; Hu, P. P.; Lu, D. et al. Versatile MXene/(GO-AgNWs) electronic textile enabled by mixed-scale assembly strategy. *Nano Energy* **2025**, *139*, 110963.
[76] Jeindl, A.; Domke, J.; Hörmann, L.; Sojka, F.; Forker, R.; Fritz, T.; Hofmann, O. T. Nonintuitive surface self-assembly of functionalized molecules on Ag (111). *ACS Nano* **2021**, *15*, 6723–6734.
[77] Balraadjsing, S.; Peijnenburg, W. J. G. M.; Vijver, M. G. Building species trait-specific nano-QSARs: Model stacking, navigating model uncertainties and limitations, and the effect of dataset size. *Environ. Int.* **2024**, *188*, 108764.
[78] Bland, G. D.; Battifarano, M.; Pradas del Real, A. E.; Sarret, G.; Lowry, G. V. Distinguishing engineered TiO_2 nanomaterials from natural Ti nanomaterials in soil using spICP-TOFMS and machine learning. *Environ. Sci. Technol.* **2022**, *56*, 2990–3001.

[79] Chen, L. F.; Yang, S. W.; Li, Y. Q.; Liu, Z.; Wang, H.; Zhang, Y. Q.; Qi, K.; Wang, G.; He, P.; Ding, G. Q. Precursor symmetry triggered modulation of fluorescence quantum yield in graphene quantum dots. *Adv. Funct. Mater.* **2024**, *34*, 2401246.

[80] He, S.; Nader, K.; Abarrategi, J. S.; Bediaga, H.; Nocedo-Mena, D.; Ascencio, E.; Casanola-Martin, G. M.; Castellanos-Rubio, I.; Insausti, M.; Rasulev, B. et al. NANO. PTML model for read-across prediction of nanosystems in neurosciences. computational model and experimental case of study. *J. Nanobiotechnol.* **2024**, *22*, 435.

[81] Pinho, B.; Torrente-Murciano, L. Dial-a-particle: Precise manufacturing of plasmonic nanoparticles based on early growth information-redefining automation for slow material synthesis. *Adv. Energy Mater.* **2021**, *11*, 2100918.

[82] Chen, P. C.; Tang, Z. Y.; Zeng, Z. M.; Hu, X. F.; Xiao, L. P.; Liu, Y.; Qian, X. D.; Deng, C. Y.; Huang, R. Y.; Zhang, J. Z. et al. Machine-learning-guided morphology engineering of nanoscale metal-organic frameworks. *Matter* **2020**, *2*, 1651–1666.

[83] Wang, C.; Ko, B.; Najimu, M.; Sasmaz, E. Application of machine learning and laser-induced breakdown spectroscopy to flame spray pyrolysis for the prediction of catalyst properties. *Chem. Mater.* **2023**, *35*, 1926–1934.

[84] Braham, E. J.; Cho, J.; Forlano, K. M.; Watson, D. F.; Arròyave, R.; Banerjee, S. Machine learning-directed navigation of synthetic design space: A statistical learning approach to controlling the synthesis of perovskite halide nanoplatelets in the quantum-confined regime. *Chem. Mater.* **2019**, *31*, 3281–3292.

[85] Ferreira, L. F.; Giordano, G. F.; Gobbi, A. L.; Piazzetta, M. H. O.; Schleder, G. R.; Lima, R. S. Real-time and *in situ* monitoring of the synthesis of silica nanoparticles. *ACS Sens.* **2022**, *7*, 1045–1057.

[86] Sadeghi, S.; Bateni, F.; Kim, T.; Son, D. Y.; Bennett, J. A.; Orouji, N.; Punati, V. S.; Stark, C.; Cerra, T. D.; Awad, R. et al. Autonomous nanomanufacturing of lead-free metal halide perovskite nanocrystals using a self-driving fluidic lab. *Nanoscale* **2024**, *16*, 580–591.

[87] Fanourgakis, G. S.; Gkagkas, K.; Tylianakis, E.; Froudakis, G. E. A universal machine learning algorithm for large-scale screening of materials. *J. Am. Chem. Soc.* **2020**, *142*, 3814–3822.

[88] Jing, Y.; Low, A. K. Y.; Liu, Y.; Feng, M. J.; Lim, J. W. M.; Loh, S. M.; Rehman, Q.; Blundel, S. A.; Mathews, N.; Hippalgaonkar, K. et al. Stable and highly emissive infrared Yb-doped perovskite quantum cutters engineered by machine learning. *Adv. Mater.* **2024**, *36*, 2405973.

[89] Li, J. G.; Tu, Y. X.; Liu, R. L.; Lu, Y. H.; Zhu, X. Toward "on-demand" materials synthesis and scientific discovery through intelligent robots. *Adv. Sci.* **2020**, *7*, 1901957.

[90] Pan, B.; Madani, M. S.; Forsberg, A. P.; Brutchey, R. L.; Malmstadt, N. Solvent dependence of ionic liquid-based pt nanoparticle synthesis: Machine learning-aided in-line monitoring in a flow reactor. *ACS Nano* **2024**, *18*, 25542–25551.

[91] Ouassil, N.; Pinals, R. L.; Del Bonis-O'Donnell, J. T.; Wang, J. W.; Landry, M. P. Supervised learning model predicts protein adsorption to carbon nanotubes. *Sci. Adv.* **2022**, *8*, eabm0898.

[92] Jose, N. A.; Kovalev, M.; Bradford, E.; Schweidtmann, A. M.; Zeng, H. C.; Lapkin, A. A. Pushing nanomaterials up to the kilogram scale—An accelerated approach for synthesizing antimicrobial ZnO with high shear reactors, machine learning and high-throughput analysis. *Chem. Eng. J.* **2021**, *426*, 131345.

[93] Mekki-Berrada, F.; Ren, Z. K.; Huang, T.; Wong, W. K.; Zheng, F.; Xie, J. X.; Tian, I. P. S.; Jayavelu, S.; Mahfoud, Z.; Bash, D. et al. Two-step machine learning enables optimized nanoparticle synthesis. *npj Comput. Mater.* **2021**, *7*, 55.

[94] Wahl, C. B.; Aykol, M.; Swisher, J. H.; Montoya, J. H.; Suram, S. K.; Mirkin, C. A. Machine learning-accelerated design and synthesis of polyelemental heterostructures. *Sci. Adv.* **2021**, *7*, eabj5505.

[95] Al-Gheethi, A.; Sundram, N.; Crane, R.; Alburihi, A.; Mohamed, R. M. S. R.; Al-Shaibani, M. M.; Noman, E. A.; Ponnusamy, S. K.; Kamil, N. A. F. M. Metronidazole photocatalytic degradation by zinc oxide nanoparticles synthesized in watermelon peel extract; Advanced optimization, simulation and numerical models using machine learning applications. *Environ. Res.* **2022**, *212*, 113537.

[96] Shao, T. Q.; Zhang, H. R.; Wang, Z. J.; Zhang, T. H.; Xu, X.; Zhang, H. B.; Wang, X. F.; Guo, H. Y. *In-situ* grown carbon nanotubes on waste glass powder: Resource-efficient preparation and machine-learning based dispersion evaluation. *Case Stud. Constr. Mater.* **2024**, *21*, e03938.

[97] Mahmood, A. U.; Ghelardini, M. M.; Tracy, J. B.; Yingling, Y. G. Machine learning and small data-guided optimization of silica shell morphology on gold nanorods. *Chem. Mater.* **2024**, *36*, 9330–9340.

[98] Cai, Z. L.; Li, W.; Chung, Y. G.; Li, S.; Liang, T. G.; Wu, T. Transfer learning-assisted computational screening of metal-organic frameworks and covalent-organic frameworks for the separation of Xe/Kr noble gas. *Sep. Purif. Technol.* **2024**, *348*, 127752.

[99] Chatterjee, M.; Banerjee, A.; De, P.; Gajewicz-Skretna, A.; Roy, K. A novel quantitative read-across tool designed purposefully to fill the existing gaps in nanosafety data. *Environ. Sci.: Nano* **2022**, *9*, 189–203.

[100] Skorikov, A.; Heyvaert, W.; Albecht, W.; Pelt, D. M.; Bals, S. Deep learning-based denoising for improved dose efficiency in EDX tomography of nanoparticles. *Nanoscale* **2021**, *13*, 12242–12249.

[101] Zheng, B. W.; Gu, G. X. Machine learning-based detection of graphene defects with atomic precision. *Nanomicro Lett.* **2020**, *12*, 181.

[102] Gleason, S. P.; Dahl, J. C.; Elzouka, M.; Wang, X. Z.; Byrne, D. O.; Cho, H.; Gababa, M.; Prasher, R. S.; Lubner, S.; Chan, E. M. et al. Automated gold nanorod spectral morphology analysis pipeline. *ACS Nano* **2024**, *18*, 34646–34655.

[103] Kamšek, A. R.; Ruiz-Zepeda, F.; Bele, M.; Logar, A.; Dražić, G.; Hodnik, N. Structure–stability relationships in Pt-alloy nanoparticles using identical-location four-dimensional scanning transmission electron microscopy and unsupervised machine learning. *ACS Nano* **2025**, *19*, 2334–2344.

[104] Guo, G. B.; Goldfeder, J.; Lan, L.; Ray, A.; Yang, A. H.; Chen, B. Y.; Billinge, S. J. L.; Lipson, H. Towards end-to-end structure determination from x-ray diffraction data using deep learning. *npj Comput. Mater.* **2024**, *10*, 209.

[105] Serov, N.; Vinogradov, V. Inverse material search and synthesis verification by hand drawings via transfer learning and contour detection. *Small Methods* **2022**, *6*, 2101619.

[106] Mahjoubi, S.; Ye, F.; Bao, Y.; Meng, W. N.; Zhang, X. Identification and classification of exfoliated graphene flakes from microscopy images using a hierarchical deep convolutional neural network. *Eng. Appl. Artif. Intell.* **2023**, *119*, 105743.

[107] Sun, Z. J.; Shi, J.; Wang, J.; Jiang, M. Q.; Wang, Z.; Bai, X. P.; Wang, X. X. A deep learning-based framework for automatic analysis of the nanoparticle morphology in SEM/TEM images. *Nanoscale* **2022**, *14*, 10761–10772.

[108] Yao, L. H.; An, H.; Zhou, S.; Kim, A.; Luijten, E.; Chen, Q. Seeking regularity from irregularity: Unveiling the synthesis–nanomorphology relationships of heterogeneous nanomaterials using unsupervised machine learning. *Nanoscale* **2022**, *14*, 16479–16489.

[109] Duan, P. J.; Liu, J. Y.; Chen, L.; Li, M. X.; Pan, J. W.; Zhang, Z. Q.; Bai, C. W.; Chen, X. J.; Yu, H. Q.; Chen, F. Polymeric products deactivate carbon-based catalysts in catalytic oxidation reactions. *Nat. Water* **2025**, *3*, 178–190.

[110] Yao, L. H.; Ou, Z. H.; Luo, B. B.; Xu, C.; Chen, Q. Machine learning to reveal nanoparticle dynamics from liquid-phase TEM videos. *ACS Cent. Sci.* **2020**, *6*, 1421–1430.

[111] Sanin, A.; Flowers, J. K.; Piotrowiak, T. H.; Felsen, F.; Merker, L.; Ludwig, A.; Bresser, D.; Stein, H. S. Integrating automated electrochemistry and high-throughput characterization with machine

learning to explore Si-Ge-Sn thin-film lithium battery anodes. *Adv. Energy Mater.* **2025**, *15*, 2404961.
[112] Chen, J.; Zhang, H. R.; Wahl, C. B.; Liu, W.; Mirkin, C. A.; Dravid, V. P.; Apley, D. W.; Chen, W. Automated crystal system identification from electron diffraction patterns using multiview opinion fusion machine learning. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2309240120.
[113] Cho, M. G.; Sytwu, K.; Rangel DaCosta, L.; Groschner, C.; Oh, M. H.; Scott, M. C. Size-resolved shape evolution in inorganic nanocrystals captured via high-throughput deep learning-driven statistical characterization. *ACS Nano* **2024**, *18*, 29736–29747.
[114] Huang, Y.; Cao, J. Y.; Li, X. H.; Yang, Q.; Xie, Q. Q.; Liu, X.; Cai, X. M.; Chen, J. W.; Hong, H. X.; Li, R. B. Multimodal feature fusion machine learning for predicting chronic injury induced by engineered nanomaterials. *Nat. Commun.* **2025**, *16*, 2765.
[115] Huang, Y.; Wang, T. Q.; Li, Y.; Wang, Z.; Cai, X. M.; Chen, J. W.; Li, R. B.; Li, X. H. *In vitro*-to-*in vivo* extrapolation on lung toxicity induced by metal oxide nanoparticles via data-mining. *Environ. Sci. Technol.* **2025**, *59*, 1673–1682.
[116] Dang, Y.; Wang, G. T.; Su, G. H.; Lu, Z. W.; Wang, Y. Y.; Liu, T.; Pu, X.; Wang, X. X.; Wu, C.; Song, C. et al. Rational construction of a Ni/$CoMoO_4$ heterostructure with strong Ni–O–Co bonds for improving multifunctional nanozyme activity. *ACS Nano* **2022**, *16*, 4536–4550.
[117] Liao, Y. K.; Lai, Y. S.; Pan, F.; Su, Y. H. Hybrid-biotaxonomy-like machine learning enables an anticipated surface Plasmon resonance of Au/Ag nanoparticles assembled on ZnO nanorods. *J. Mater. Chem. A* **2023**, *11*, 11187–11201.
[118] Gabellini, C.; Şologan, M.; Pellizzoni, E.; Marson, D.; Daka, M.; Franchi, P.; Bignardi, L.; Franchi, S.; Posel, Z.; Baraldi, A. et al. Spotting local environments in self-assembled monolayer-protected gold nanoparticles. *ACS Nano* **2022**, *16*, 20902–20914.
[119] Tuchin, V. S.; Stepanidenko, E. A.; Vedernikova, A. A.; Cherevkov, S. A.; Li, D.; Li, L.; Döring, A.; Otyepka, M.; Ushakova, E. V.; Rogach, A. L. Optical properties prediction for red and near-infrared emitting carbon dots using machine learning. *Small* **2024**, *20*, 2310402.
[120] Gao, X. J.; Yan, J.; Zheng, J. J.; Zhong, S. L.; Gao, X. F. Clear-box machine learning for virtual screening of 2D nanozymes to target tumor hydrogen peroxide. *Adv. Healthc. Mater.* **2023**, *12*, 2202925.
[121] Xu, D. T.; Yin, W. Y.; Zhou, J.; Wu, L. Y.; Yao, H. D.; Sun, M. H.; Chen, P.; Deng, X. W.; Zhao, L. N. Rational design of MoS_2-supported Cu single-atom catalysts by machine learning potential for enhanced peroxidase-like activity. *Nanoscale* **2023**, *15*, 6686–6695.
[122] Zhang, C.; Yang, B. L.; Peng, Z. L.; Chen, S. H. Machine learning-based prediction of mechanical properties of N-doped γ-graphdiyne. *Sci. China Mater.* **2024**, *67*, 1129–1139.
[123] Halder, A. K.; Melo, A.; Cordeiro, M. N. D. S. A unified in silico model based on perturbation theory for assessing the genotoxicity of metal oxide nanoparticles. *Chemosphere* **2020**, *244*, 125489.
[124] Wei, Y. H.; Wu, J.; Wu, Y. X.; Liu, H. J.; Meng, F. Q.; Liu, Q. Q.; Midgley, A. C.; Zhang, X. Y.; Qi, T. Y.; Kang, H. L. et al. Prediction and design of nanozymes using explainable machine learning. *Adv. Mater.* **2022**, *34*, 2201736.
[125] Pihlajamäki, A.; Matus, M. F.; Malola, S.; Häkkinen, H. GraphBNC: Machine learning-aided prediction of interactions between metal nanoclusters and blood proteins. *Adv. Mater.* **2024**, *36*, 2407046.
[126] Lazarovits, J.; Sindhwani, S.; Tavares, A. J.; Zhang, Y. W.; Song, F. Y.; Audet, J.; Krieger, J. R.; Syed, A. M.; Stordy, B.; Chan, W. C. W. Supervised learning and mass spectrometry predicts the *in vivo* fate of nanomaterials. *ACS Nano* **2019**, *13*, 8023–8034.
[127] Sengottiyan, S.; Mikolajczyk, A.; Jagiełło, K.; Swirog, M.; Puzyn, T. Core, coating, or corona. The importance of considering protein coronas in nano-QSPR modeling of zeta potential. *ACS Nano* **2023**, *17*, 1989–1997.
[128] Mortazavi, B.; Novikov, I. S.; Shapeev, A. V. A machine-learning-based investigation on the mechanical/failure response and thermal conductivity of semiconducting BC_2N monolayers. *Carbon* **2022**, *188*, 431–441.
[129] Mortazavi, B.; Rajabpour, A.; Zhuang, X. Y.; Rabczuk, T.; Shapeev, A. V. Exploring thermal expansion of carbon-based nanosheets by machine-learning interatomic potentials. *Carbon* **2022**, *186*, 501–508.
[130] Mortazavi, B.; Shojaei, F.; Shahrokhi, M.; Azizi, M.; Rabczuk, T.; Shapeev, A. V.; Zhuang, X. Y. Nanoporous C_3N_4, C_3N_5 and C_3N_6 nanosheets; Novel strong semiconductors with low thermal conductivities and appealing optical/electronic properties. *Carbon* **2020**, *167*, 40–50.
[131] Mortazavi, B.; Shojaei, F.; Shapeev, A. V.; Zhuang, X. Y. A combined first-principles and machine-learning investigation on the stability, electronic, optical, and mechanical properties of novel C_6N_7-based nanoporous carbon nitrides. *Carbon* **2022**, *194*, 230–239.
[132] Mortazavi, B. Structural, electronic, thermal and mechanical properties of C_{60}-based fullerene two-dimensional networks explored by first-principles and machine learning. *Carbon* **2023**, *213*, 118293.
[133] Mortazavi, B.; Zhuang, X. Y. Ultrahigh strength and negative thermal expansion and low thermal conductivity in graphyne nanosheets confirmed by machine-learning interatomic potentials. *FlatChem* **2022**, *36*, 100446.
[134] Mortazavi, B.; Novikov, I. S.; Podryabinkin, E. V.; Roche, S.; Rabczuk, T.; Shapeev, A. V.; Zhuang, X. Y. Exploring phononic properties of two-dimensional materials using machine learning interatomic potentials. *Appl. Mater. Today* **2020**, *20*, 100685.
[135] Firestein, K. L.; von Treifeldt, J. E.; Kvashnin, D. G.; Fernando, J. F. S.; Zhang, C.; Kvashnin, A. G.; Podryabinkin, E. V.; Shapeev, A. V.; Siriwardena, D. P.; Sorokin, P. B. et al. Young's modulus and tensile strength of Ti_3C_2 MXene nanosheets as revealed by *in situ* TEM probing, AFM nanomechanical mapping, and theoretical calculations. *Nano Lett.* **2020**, *20*, 5900–5908.
[136] Labouta, H. I.; Asgarian, N.; Rinker, K.; Cramb, D. T. Meta-analysis of nanoparticle cytotoxicity via data-mining the literature. *ACS Nano* **2019**, *13*, 1583–1594.
[137] Yu, H. J.; Tang, S. Y.; Li, S. F. Y.; Cheng, F. Averaging strategy for interpretable machine learning on small datasets to understand element uptake after seed nanotreatment. *Environ. Sci. Technol.* **2023**, *57*, 12760–12770.
[138] Shirokii, N.; Din, Y.; Petrov, I.; Seregin, Y.; Sirotenko, S.; Razlivina, J.; Serov, N.; Vinogradov, V. Quantitative prediction of inorganic nanomaterial cellular toxicity via machine learning. *Small* **2023**, *19*, 2207106.
[139] Ma, Y.; Wang, J. L.; Wu, J. Y.; Tong, C. X.; Zhang, T. Meta-analysis of cellular toxicity for graphene via data-mining the literature and machine learning. *Sci. Total Environ.* **2021**, *793*, 148532.
[140] Zhang, H. N.; Barnard, A. S. Impact of atomistic or crystallographic descriptors for classification of gold nanoparticles. *Nanoscale* **2021**, *13*, 11887–11898.
[141] Scott-Fordsmand, J. J.; Gomes, S. I. L.; Pokhrel, S.; Mädler, L.; Fasano, M.; Asinari, P.; Tämm, K.; Jänes, J.; Amorim, M. J. B. Machine learning allowed interpreting toxicity of a Fe-doped CuO NM library large data set—An environmental *in vivo* case study. *ACS Appl. Mater. Interfaces* **2024**, *16*, 42862–42872.
[142] Yen, Y. T.; Lin, Y. S.; Chang, Y. J.; Li, M. T.; Chyueh, S. C.; Chang, H. T. Nanomaterial-based sensor arrays with deep learning for screening of illicit drugs. *Adv. Mater. Technol.* **2022**, *7*, 2200243.
[143] Liu, H. W.; Cheng, X.; Valanoor, N. Universal approach for predicting crystallography of heterogeneous epitaxial nanocrystals with multiple orientation relationships. *ACS Appl. Mater. Interfaces* **2016**, *8*, 34844–34853.
[144] Khot, A. C.; Dongale, T. D.; Nirmal, K. A.; Sung, J. H.; Lee, H. J.;

Nikam, R. D.; Kim, T. G. Amorphous boron nitride memristive device for high-density memory and neuromorphic computing applications. *ACS Appl. Mater. Interfaces* **2022**, *14*, 10546–10557.

[145] Praetorius, A.; Gundlach-Graham, A.; Goldberg, E.; Fabienke, W.; Navratilova, J.; Gondikas, A.; Kaegi, R.; Günther, D.; Hofmann, T.; Von Der Kammer, F. Single-particle multi-element fingerprinting (spMEF) using inductively-coupled plasma time-of-flight mass spectrometry (ICP-TOFMS) to identify engineered nanoparticles against the elevated natural background in soils. *Environ. Sci.: Nano* **2017**, *4*, 307–314.

[146] Parker, A. J.; Barnard, A. S. Unsupervised structure classes *vs*. supervised property classes of silicon quantum dots using neural networks. *Nanoscale Horiz.* **2021**, *6*, 277–282.

[147] Li, J.; Li, X. N.; Kah, M.; Yue, L.; Cheng, B. X.; Wang, C. X.; Wang, Z. Y.; Xing, B. S. Unlocking the potential of carbon dots in agriculture using data-driven approaches. *Sci. Total Environ.* **2024**, *944*, 173605.

[148] Ji, K. X.; Liu, P.; Wu, C. Y.; Li, Q.; Ge, Y.; Wen, Y. P.; Xiong, J. H.; Liu, X. X.; He, P. P.; Tang, K. J. et al. A deep learning strategy for discrimination and detection of multi-sulfonamides residues in aquatic environments using gold nanoparticles-decorated violet phosphorene SERS substrates. *Sens. Actuators B: Chem.* **2023**, *386*, 133736.

[149] Glaubitz, C.; Rothen-Rutishauser, B.; Lattuada, M.; Balog, S.; Petri-Fink, A. Designing the ultrasonic treatment of nanoparticle-dispersions via machine learning. *Nanoscale* **2022**, *14*, 12940–12950.

[150] Ge, Y.; Li, M. H.; Zhong, Y.; Xu, L. L.; Lu, X. Y.; Hu, J. Q.; Peng, Q. M.; Bai, L.; Wen, Y. P. Halloysite nanotube/black phosphorene nanohybrid modified screen-printed carbon electrode as an ultra-portable electrochemical sensing platform for smartphone-capable detection of maleic hydrazide with machine learning assistance. *Food Chem.* **2023**, *406*, 134967.

[151] Polat, C.; Kurban, M.; Kurban, H. Multimodal neural network-based predictive modeling of nanoparticle properties from pure compounds. *Mach. Learn.: Sci. Technol.* **2024**, *5*, 045062.

[152] Xie, Z. J.; Ou, H. R.; Xu, B. Y.; Zhan, H.; Wang, Z. P.; Yang, F.; Xu, J. S. Ion gel pressure sensor with high sensitivity and a wide linear range enabled by magnetically induced gradient microstructures. *ACS Appl. Mater. Interfaces* **2025**, *17*, 12720–12730.

[153] Wang, Z. S.; Xu, L. J.; Liu, W. L.; Chen, Y.; Yang, Q. N.; Tang, Z. M.; Tan, H. H.; Li, N.; Du, J. J.; Yu, M. L. et al. Tough, self-healing, adhesive double network conductive hydrogel based on gelatin-polyacrylamide covalently bridged by oxidized sodium alginate for durable wearable sensors. *Int. J. Biol. Macromol.* **2024**, *276*, 133802.

[154] Verma, G.; Gokarna, A.; Kadiri, H.; Lerondel, G.; Gupta, A. Flexible ZnO nanowire platform by metal-seeded chemical bath deposition: Parametric analysis and predictive modeling. *Appl. Mater. Today* **2024**, *40*, 102385.

[155] Simeone, F. C.; Costa, A. L. Assessment of cytotoxicity of metal oxide nanoparticles on the basis of fundamental physical-chemical parameters: A robust approach to grouping. *Environ. Sci.: Nano* **2019**, *6*, 3102–3112.

[156] Sizochenko, N.; Mikolajczyk, A.; Jagiello, K.; Puzyn, T.; Leszczynski, J.; Rasulev, B. How the toxicity of nanomaterials towards different species could Be simultaneously evaluated: A novel multi-nano-read-across approach. *Nanoscale* **2018**, *10*, 582–591.

[157] Fratalocchi, A.; Fleming, A.; Conti, C.; Di Falco, A. NIST-certified secure key generation via deep learning of physical unclonable functions in silica aerogels. *Nanophotonics* **2020**, *10*, 457–464.

[158] Findlay, M. R.; Freitas, D. N.; Mobed-Miremadi, M.; Wheeler, K. E. Machine learning provides predictive analysis into silver nanoparticle protein corona formation from physicochemical properties. *Environ. Sci.: Nano* **2018**, *5*, 64–71.

[159] Du, D.; Zhang, Y. W.; Li, X. W.; Xiao, N. M. First-principles calculations, machine learning and Monte Carlo simulations of the magnetic coercivity of Fe_xCo_{1-x} bulks and nanoclusters. *Nanomaterials* **2025**, *15*, 577.

[160] Diehl, H. P.; Brandon Sweeney, C.; Tran, T. Q.; Green, M. J. Computer vision and machine learning for assessing dispersion quality in carbon nanotube/resin systems. *Carbon* **2023**, *213*, 118230.

[161] Baro, B.; Shah, K.; Shah, K. H.; Roy, M.; Bayan, S. ZnO nanowall network-based tactile/gesture sensors and prediction with machine learning. *ACS Appl. Mater. Interfaces* **2024**, *16*, 61272–61283.

[162] Banerjee, P.; Filatov, A. S.; Zuo, X. B.; Diroll, B. T.; Shevchenko, E. V. Mechanistic insights into copper(I) and copper(II) cation exchange reactions in CdSe nanoplatelets. *Chem. Mater.* **2023**, *35*, 8872–8882.

[163] Zhao, L. N.; Chang, Y.; Qiu, S.; Liu, H. S.; Zhao, J. J.; Gao, J. F. High mechanical energy storage capacity of ultranarrow carbon nanowires bundles by machine learning driving predictions. *Adv. Energy Sustain. Res.* **2023**, *4*, 2300112.

[164] Liu, X. H.; Yang, K. Y.; Zhang, L. S.; Wang, W. T.; Zhou, S. D.; Wu, B.; Xiong, M. Y.; Yang, S. C.; Tan, R. A fast forward prediction framework for energy materials design based on machine learning methods. *Energy Mater. Adv.* **2024**, *5*, 0131.

[165] Long, Y.; Tao, Y.; Lv, W.; Yang, Q. H. Making 2D materials sparkle in energy storage via assembly. *Acc. Chem. Res.* **2024**, *57*, 2689–2699.

[166] Kiaghadi, M.; Sheikholeslami, M.; Alinia, A. M.; Boora, F. M. Predicting the performance of a photovoltaic unit via machine learning methods in the existence of finned thermal storage unit. *J. Energy Storage* **2024**, *90*, 111766.

[167] Ramesh Srenyvasan, V.; Muthuvel, S.; Kalil Rahiman, M.; Khan, S. S.; Rex, P. Zinc manganese oxide spinel phase nanomaterial layered on solar stills applications: An analytical study for improved performance compared with machine learning methods. *J. Water Process Eng.* **2025**, *71*, 107363.

[168] Lin, J. W.; Dong, H.; Cui, S. L.; Dong, W.; Sun, H. Fluid classification via the dual functionality of moisture-enabled electricity generation enhanced by deep learning. *ACS Appl. Mater. Interfaces* **2024**, *16*, 63723–63734.

[169] An, H.; Smith, J. W.; Ji, B. Q.; Cotty, S.; Zhou, S.; Yao, L. H.; Kalutantirige, F. C.; Chen, W. X.; Ou, Z. H.; Su, X. et al. Mechanism and performance relevance of nanomorphogenesis in polyamide films revealed by quantitative 3D imaging and machine learning. *Sci. Adv.* **2022**, *8*, eabk1888.

[170] Tan, E. X.; Chen, Y. C.; Lee, Y. H.; Leong, Y. X.; Leong, S. X.; Stanley, C. V.; Pun, C. S.; Ling, X. Y. Incorporating plasmonic featurization with machine learning to achieve accurate and bidirectional prediction of nanoparticle size and size distribution. *Nanoscale Horiz.* **2022**, *7*, 626–633.

[171] Gandhi, A. M.; Shanmugan, S.; Gorjian, S.; Pruncu, C. I.; Sivakumar, S.; Elsheikh, A. H.; Essa, F. A.; Omara, Z. M.; Panchal, H. Performance enhancement of stepped basin solar still based on OSELM with traversal tree for higher energy adaptive control. *Desalination* **2021**, *502*, 114926.

[172] Yang, L.; Liu, Q. L.; Kumar, P.; Sengupta, A.; Farnoud, A.; Shen, R. L.; Trofimova, D.; Ziegler, S.; Davoudi, N.; Doryab, A. et al. LungVis 1.0: An automatic AI-powered 3D imaging ecosystem unveils spatial profiling of nanoparticle delivery and acinar migration of lung macrophages. *Nat. Commun.* **2024**, *15*, 10138.

[173] Awere, C. O.; Sneha, A.; Rakkammal, K.; Muthui, M. M.; Kumari R, A.; Govindan, S.; Batur Çolak, A.; Bayrak, M.; Muthuramalingam, P.; Anadebe, V. C. et al. Carbon dot unravels accumulation of triterpenoid in *Evolvulus alsinoides* hairy roots culture by stimulating growth, redox reactions and ANN machine learning model prediction of metabolic stress response. *Plant Physiol. Biochem.* **2024**, *216*, 109142.

[174] Huang, Z. G.; Chen, S. N.; Ali, H. E.; Elkamchouchi, D. H.; Hu, J.;

Ali, E.; Zhang, J.; Huang, Y. D. Application of CNN and ANN in assessment the effect of chemical components of biological nanomaterials in treatment of infection of inner ear and environmental sustainability. *Chemosphere* **2023**, *331*, 138458.

[175] Pei, C. C.; Wang, Y.; Ding, Y. J.; Li, R. X.; Shu, W. K.; Zeng, Y.; Yin, X.; Wan, J. J. Designed concave octahedron heterostructures decode distinct metabolic patterns of epithelial ovarian tumors. *Adv. Mater.* **2023**, *35*, 2209083.

[176] Hao, T. T.; Zhou, H. Q.; Gai, P. P.; Wang, Z. L.; Guo, Y. X.; Lin, H.; Wei, W. T.; Guo, Z. Y. Deep learning-assisted single-atom detection of copper ions by combining click chemistry and fast scan voltammetry. *Nat. Commun.* **2024**, *15*, 10292.

[177] Yu, T.; Su, S. X.; Hu, J.; Zhang, J.; Xianyu, Y. L. A new strategy for microbial taxonomic identification through micro-biosynthetic gold nanoparticles and machine learning. *Adv. Mater.* **2022**, *34*, 2109365.

[178] Kwon, Y. T.; Kim, Y. S.; Kwon, S.; Mahmood, M.; Lim, H. R.; Park, S. W.; Kang, S. O.; Choi, J. J.; Herbert, R.; Jang, Y. C. et al. All-printed nanomembrane wireless bioelectronics using a biocompatible solderable graphene for multimodal human–machine interfaces. *Nat. Commun.* **2020**, *11*, 3450.

[179] Tran, A. T. T.; Hassan, K.; Tung, T. T.; Tripathy, A.; Mondal, A.; Losic, D. Graphene and metal-organic framework hybrids for high-performance sensors for lung cancer biomarker detection supported by machine learning augmentation. *Nanoscale* **2024**, *16*, 9084–9095.

[180] Liu, T.; He, J.; Lu, Z. W.; Sun, M. M.; Wu, M. J.; Wang, X. X.; Jiang, Y. Y.; Zou, P.; Rao, H. B.; Wang, Y. Y. A visual electrochemiluminescence molecularly imprinted sensor with Ag^{+}@UiO-66-NH_2 decorated $CsPbBr_3$ perovskite based on smartphone for point-of-care detection of nitrofurazone. *Chem. Eng. J.* **2022**, *429*, 132462.

[181] Wang, Y. D.; Sun, J. Z.; Yi, J. H.; Fu, R. J.; Liu, B.; Xianyu, Y. L. "Three-in-one" analysis of proteinuria for disease diagnosis through multifunctional nanoparticles and machine learning. *Adv. Sci.* **2025**, *12*, 2410751.

[182] Henriquez, D. V. D. O.; Kang, M. G.; Cho, I.; Choi, J.; Park, J.; Gul, O.; Ahn, J.; Lee, D. S.; Park, I. Low-power, multi-transduction nanosensor array for accurate sensing of flammable and toxic gases. *Small Methods* **2023**, *7*, 2201352.

[183] Shao, S. F.; Xie, C. Y.; Xia, Y. X.; Zhang, L.; Zhang, J.; Wei, S.; Kim, H. W.; Kim, S. S. Highly conjugated three-dimensional van der Waals heterostructure-based nanocomposite films for ultrahigh-responsive TEA gas sensors at room temperature. *J. Mater. Chem. A* **2022**, *10*, 2995–3008.

[184] Zhao, Y. L.; El-Sherif, D. M.; Cheng, J.; Zhang, H.; Hu, Z. Y.; Wu, Q.; Shen, H.; Yu, Z. L.; Xie, Y. F. Carboxymethyl cellulose-based colorimetric sensor array incorporating ionic liquid-tuned anthocyanin and nano-silica coating for pork freshness monitoring. *Food Hydrocolloids* **2025**, *158*, 110601.

[185] Abbas, W.; Zafar, F.; Abou Taleb, M. F.; Ameen, M.; Sami, A.; Mazhar, M. E.; Akhtar, N.; Fazal, M. W.; Ibrahim, M. M.; El-Bahy, Z. M. Machine learning trained poly(3,4-ethylenedioxythiophene) functionalized carbon matrix suspended Cu nanoparticles for precise monitoring of nitrite from pickled vegetables. *Food Chem.* **2024**, *460*, 140395.

[186] Al-Nowaiser, W. K.; Vohra, M. S.; Onaizi, S. A. Hybrid electrocoagulation/adsorption system using aluminum electrodes and novel GO@ZIF-7 nanocomposite for the effective removal of Pb(II) from wastewater. *Sep. Purif. Technol.* **2024**, *350*, 127828.

[187] Belcaid, A.; Beakou, B. H.; Bouhsina, S.; Anouar, A. Biosorption of cobalt and chromium from wastewater using manganese dioxide and iron oxide nanoparticles loaded on cellulose-based biochar: Modeling and optimization with machine learning (artificial neural network). *Int. J. Biol. Macromol.* **2024**, *282*, 136855.

[188] BinMakhashen, G. M.; Bahadi, S. A.; Al-Jamimi, H. A.; Onaizi, S. A. Ensemble meta machine learning for predicting the adsorption of anionic and cationic dyes from aqueous solutions using polymer/graphene/clay/MgFeAl-LTH nanocomposite. *Chemosphere* **2024**, *349*, 140861.

[189] Ebere Enyoh, C.; Wang, Q. Y. Box–Behnken design and machine learning optimization of PET fluorescent carbon quantum dots for removing fluoxetine and ciprofloxacin with molecular dynamics and docking studies as potential antidepressant and antibiotic. *Sep. Purif. Technol.* **2025**, *362*, 131975.

[190] Mahata, B.; Acharyya, S.; Banerji, P.; Guha, P. K. Assessment of fish adulteration using SnO_2 nanopetal-based gas sensor and machine learning. *Food Chem.* **2024**, *438*, 138039.

[191] Alabduljabbar, H.; Amin, M. N.; Eldin, S. M.; Javed, M. F.; Alyousef, R.; Mohamed, A. M. Forecasting compressive strength and electrical resistivity of graphite based nano-composites using novel artificial intelligence techniques. *Case Stud. Constr. Mater.* **2023**, *18*, e01848.

[192] Sun, H.; Amin, M. N.; Qadir, M. T.; Arifeen, S. U.; Iftikhar, B.; Althoey, F. Investigating the effectiveness of carbon nanotubes for the compressive strength of concrete using AI-aided tools. *Case Stud. Constr. Mater.* **2024**, *20*, e03083.

[193] Zhang, Y.; Zhang, Q. Z.; AlAteah, A. H.; Essam, A.; Mostafa, S. A. Predictive modeling for mechanical characteristics of ultra high-performance concrete blended with eggshell powder and nano silica utilizing traditional technique and machine learning algorithm. *Case Stud. Constr. Mater.* **2024**, *21*, e04025.

[194] Alyousef, R.; Nassar, R. U. D.; Fawad, M.; Farooq, F.; Gamil, Y.; Najeh, T. Predicting the properties of concrete incorporating graphene nano platelets by experimental and machine learning approaches. *Case Stud. Constr. Mater.* **2024**, *20*, e03018.

[195] Liu, G. L.; Zhao, H. Z.; Amin, M. N.; Zaman, A.; Hassan, A. M.; Ali, M.; Rehman, M. F. Strength predictive models of cementitious matrix by hybrid intrusion of nano and micro silica: Hyper-tuning with ensemble approaches. *J. Mater. Res. Technol.* **2023**, *26*, 1808–1832.

[196] Meddage, D. P. P.; Fonseka, I.; Mohotti, D.; Wijesooriya, K.; Lee, C. K. An explainable machine learning approach to predict the compressive strength of graphene oxide-based concrete. *Constr. Build. Mater.* **2024**, *449*, 138346.

[197] Liu, J. L.; Xie, X. F.; Li, L. F. Experimental study on mechanical properties and durability of grafted nano-SiO_2 modified rice straw fiber reinforced concrete. *Constr. Build. Mater.* **2022**, *347*, 128575.

[198] Yao, Y.; Cappola, J.; Zhang, Z. Y.; Zhu, Q.; Cai, W. J.; Yu, X. X.; Li, L. Nanostructure and dislocation interactions in refractory complex concentrated alloy: From chemical short-range order to nanoscale B2 precipitates. *Acta Mater.* **2024**, *281*, 120457.

[199] Wu, Y. H.; Yan, G. S.; Yu, W. S.; Shen, S. P. Investigating nanostructure–property relationship of WTaVCr high-entropy alloy via machine learning optimized reactive potential. *J. Mater. Res. Technol.* **2024**, *32*, 2624–2637.

[200] Zhong, Z. Y.; An, J.; Wu, D. A.; Gao, N.; Liu, L.; Wang, Z. Y.; Meng, F. C.; Zhou, X.; Fan, T. X. A machine learning strategy for enhancing the strength and toughness in metal matrix composites. *Int. J. Mech. Sci.* **2024**, *281*, 109550.

[201] Zhao, S. Y.; Zhang, Y. Y.; Zhang, Y. H.; Zhang, W.; Yang, J.; Kitipornchai, S. Genetic programming-assisted micromechanical models of graphene origami-enabled metal metamaterials. *Acta Mater.* **2022**, *228*, 117791.

[202] Lu, Z. Y.; Fan, Y.; Sun, Z. X.; He, X. D.; Yang, C. C.; Yin, H.; Zhang, J. Z.; Song, G. P.; Zheng, Y. T.; Bai, Y. L. A fast composition-stability machine learning model for screening MAX phases and guiding discovery of Ti_2SnN. *J. Adv. Ceram.* **2025**, *14*, 9221050.

[203] Wang, T.; Russo, D. P.; Bitounis, D.; Demokritou, P.; Jia, X. L.; Huang, H.; Zhu, H. Integrating structure annotation and machine learning approaches to develop graphene toxicity models. *Carbon* **2023**, *204*, 484–494.

[204] Mikolajczyk, A.; Sizochenko, N.; Mulkiewicz, E.; Malankowska, A.; Rasulev, B.; Puzyn, T. A chemoinformatics approach for the characterization of hybrid nanomaterials: Safer and efficient design perspective. *Nanoscale* **2019**, *11*, 11808–11818.

[205] Goldberg, E.; Scheringer, M.; Bucheli, T. D.; Hungerbühler, K. Prediction of nanoparticle transport behavior from physicochemical properties: Machine learning provides insights to guide the next generation of transport models. *Environ. Sci.: Nano* **2015**, *2*, 352–360.

[206] Gomez-Flores, A.; Bradford, S. A.; Cai, L.; Urík, M.; Kim, H. Prediction of attachment efficiency using machine learning on a comprehensive database and its validation. *Water Res.* **2023**, *229*, 119429.

[207] Gupta, A. K.; Zarkadoula, E.; Ziatdinov, M.; Kalinin, S. V.; Paduri, V. R.; Hachtel, J. A.; Zhang, Y. W.; Trautmann, C.; Weber, W. J.; Sachan, R. Nanoscale core–shell structure and recrystallization of swift heavy ion tracks in $SrTiO_3$. *Nanoscale* **2024**, *16*, 14366–14377.

[208] Yan, X. L.; Sedykh, A.; Wang, W. Y.; Yan, B.; Zhu, H. Construction of a web-based nanomaterial database by big data curation and modeling friendly nanostructure annotations. *Nat. Commun.* **2020**, *11*, 2519.

[209] Lindsey, R. K.; Goldman, N.; Fried, L. E.; Bastea, S. Chemistry-mediated ostwald ripening in carbon-rich C/O systems at extreme conditions. *Nat. Commun.* **2022**, *13*, 1424.

[210] Li, J. L.; Telychko, M.; Yin, J.; Zhu, Y. X.; Li, G. W.; Song, S. T.; Yang, H. T.; Li, J.; Wu, J. S.; Lu, J. et al. Machine vision automated chiral molecule detection and classification in molecular imaging. *J. Am. Chem. Soc.* **2021**, *143*, 10177–10188.

[211] Barnard, A. S.; Fox, B. L. Importance of structural features and the influence of individual structures of graphene oxide using Shapley value analysis. *Chem. Mater.* **2023**, *35*, 8840–8856.

[212] Bets, K. V.; O'Driscoll, P. C.; Yakobson, B. I. Physics-inspired transfer learning for ML-prediction of CNT band gaps from limited data. *npj Comput. Mater.* **2024**, *10*, 66.

[213] Hundi, P.; Shahsavari, R. Deep learning to speed up the development of structure–property relations for hexagonal boron nitride and graphene. *Small* **2019**, *15*, 1900656.

[214] Groschner, C. K.; Pattison, A. J.; Ben-Moshe, A.; Alivisatos, A. P.; Theis, W.; Scott, M. C. Classifying handedness in chiral nanomaterials using label error robust deep learning. *npj Comput. Mater.* **2022**, *8*, 149.

[215] Kim, M. A.; Ai, Q. X.; Norquist, A. J.; Schrier, J.; Chan, E. M. Active learning of ligands that enhance perovskite nanocrystal luminescence. *ACS Nano* **2024**, *18*, 14514–14522.

[216] Yu, H. J.; Zhao, Z. L.; Cheng, F. Predicting and investigating cytotoxicity of nanoparticles by translucent machine learning. *Chemosphere* **2021**, *276*, 130164.

[217] Yu, H. J.; Zhao, Z. L.; Luo, D.; Cheng, F. Interpretable machine learning for investigating complex nanomaterial–plant–soil interactions. *Environ. Sci.: Nano* **2022**, *9*, 4305–4316.

[218] Gomes, S. I. L.; Amorim, M. J. B.; Pokhrel, S.; Mädler, L.; Fasano, M.; Chiavazzo, E.; Asinari, P.; Jänes, J.; Tämm, K.; Burk, J. et al. Machine learning and materials modelling interpretation of *in vivo* toxicological response to TiO_2 nanoparticles library (UV and non-UV exposure). *Nanoscale* **2021**, *13*, 14666–14678.