Uzbekistan Journal of Polymers
journal homepage: https://uzpolymerjournal.com/
ISSN (online): 2181-3256 (print) 2181-3246

MACHINE LEARNING AND AI IN POLYMER SCIENCE: A REVIEW OF PROGRESS, CHALLENGES, AND FUTURE DIRECTIONS

Ilnar N. Nurgaliev*, Murad B. Marasulov*, Akbarxon I. Hamzayev*, Akmal B. Abilkasimov**

**Institute of Polymer Chemistry and Physics Uzbekistan Academy of Sciences, 100128, Tashkent, Uzbekistan*
*** Kimyo international university in Tashkent, 156, Shota Rustaveli Street, Yakkasaray District, Tashkent, Uzbekistan, 100121*

ARTICLE INFO

Received: 04 August 2025
Revised: 29 September2025
Accepted: 29 October 2025

Keywords:
machine learning, polymer science, materials genome, property prediction, inverse design, explainable AI, physics-informed neural networks, sustainability, high-throughput screening, multiscale modeling

Corresponding author
ilnarvodnik@gmail.com

ABSTRACT

Recent years have witnessed a profound transformation in polymer science driven by the rapid integration of machine learning (ML) and artificial intelligence (AI). This review provides a comprehensive and up-to-date analysis of advances in ML applications across the polymer field from 2023 to July 2025. We examine the use of supervised, unsupervised, and reinforcement learning methods for polymer design, property prediction, structural characterization, process optimization, and sustainable materials development. Special attention is given to emerging paradigms such as high-throughput screening, inverse design, multi-scale modeling, and the use of generative models—including variational autoencoders (VAEs), graph neural networks (GNNs), and transformer-based architectures. We also explore recent innovations in explainable AI (XAI), physics-informed neural networks (PINNs), and the growing role of automated experimental platforms. Key challenges—including data scarcity, model generalization, and interpretability—are discussed alongside strategies such as transfer learning, active learning, and the development of polymer-specific representations like BigSMILES. The review concludes with future directions and the outlook for AI-powered polymer discovery, highlighting the increasing role of open-access databases, multi-modal learning, and autonomous laboratories. Together, these developments mark a paradigm shift in how polymers are conceived, characterized, and optimized—ushering in a new era of intelligent materials innovation.

DOI: *10.66640/UJP-2026-5-00001*

Introduction

The design and discovery of advanced materials represent one of the most transformative forces shaping modern science, engineering, and technology. From semiconductors in electronics to catalysts in chemical industries and next-generation photovoltaics, novel materials are essential for addressing global challenges in energy, sustainability, computing, and healthcare. As materials research advances, the need to accelerate the design cycle—moving from desired properties to viable materials—has become more urgent than ever.

Traditionally, the development of functional materials has been driven by a combination of empirical experimentation and domain expertise. This "trial-and-error" strategy, though foundational, is inherently slow, resource-intensive, and often limited by human intuition. Over the past century, scientific progress has introduced theoretical and computational tools that have gradually shifted the materials design paradigm toward more systematic and predictive approaches.

The introduction of density functional theory (DFT), molecular dynamics (MD), and other quantum chemistry methods has enabled the simulation of materials at atomic and electronic levels, significantly improving our ability to predict properties prior to synthesis. However, these methods remain computationally expensive and may be limited in scale and transferability, particularly for systems with complex or multi-scale interactions.

The past decade has witnessed the emergence of a new paradigm: artificial intelligence (AI)-driven inverse design. In contrast to conventional forward design—where structure leads to properties—inverse design seeks to determine the optimal structure that yields target properties. This inversion of the design problem inherently poses a high-dimensional, nonlinear challenge that is well-suited to modern AI and machine learning (ML) tools. Leveraging large-scale datasets from both experiments and simulations, AI algorithms can uncover latent patterns between structure and function, enabling rapid screening and even generative creation of candidate materials.

Over the past few years, the application of ML in materials discovery has experienced rapid and sustained growth, both in terms of research output and scientific impact. As illustrated in Figure 1, the number of publications focused on Machine Learning for Materials Discovery has steadily increased from 2020 to 2024. Foundations and Trends in Machine Learning is recognized as one of the most prestigious and highly cited journals in the fields of software and machine learning. According to Scopus data, it ranks in the 99th percentile (1st out of 490 journals) within the Software category, with an exceptional CiteScore of 202.9. Despite publishing only 17 articles between 2021 and 2024, the journal has accumulated 3,450 citations, reflecting the enduring impact and scientific value of its publications. Additionally, it boasts a remarkably high SNIP of 41.01 and an SJR (SCImago Journal Rank) of 22.797, further underscoring its influence within the global research community. These metrics highlight the journal's central role in disseminating foundational knowledge and comprehensive reviews in the domain of machine learning theory and applications. Engineering Applications of Artificial Intelligence is a well-established and influential journal in the field of artificial intelligence and its engineering applications. Published by Elsevier, the journal ranks in the 88th percentile (112th out of 970 journals) within the Electrical and Electronic Engineering category, according to Scopus metrics. Between 2021 and 2024, it published 4,287 articles, which collectively received 40,762 citations—an indicator of both its high productivity and scientific impact. With 82% of its papers cited, the journal demonstrates consistent relevance in the research community. Its SNIP of 1.888 and SJR of 1.652 further attest to its solid academic reputation and influence. These figures position the journal as a significant platform for advancing practical and theoretical research at the intersection of engineering and artificial intelligence. This trend reflects not only the scientific community's growing interest but also the maturing integration of ML into materials design workflows, spanning polymer informatics, catalyst screening, molecular dynamics, and inverse design strategies.

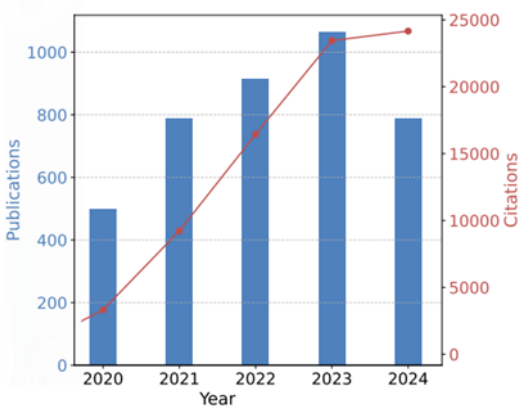


Figure 1. *Trends in scientific output for Machine Learning in Materials Discovery from 2020 to 2024. The blue bars indicate the annual number of publications, while the red line with markers represents the corresponding citation counts. The data show a consistent rise in both research activity and impact, highlighting growing interest and progress in the application of machine learning to materials science. Reprinted from ref. [1]. Copyright 2024 IOP Publishing*

The exponential increase in the number of known materials over the past decades highlights the accelerating pace of discovery, particularly in the context of machine learning–driven methodologies. As shown in Figure 2, the growth follows a near-logarithmic trend, underscoring a significant shift in how materials are explored and identified. Recent breakthroughs such as GNoME (Graph Networks for Materials Exploration) developed by Google [2], and OMat24 [3], the open materials discovery platform by Meta2 [4], exemplify this transition. These platforms have enabled the autonomous prediction and validation of thousands of novel stable materials using large-scale data-driven approaches. Such initiatives are redefining the scale and speed of materials discovery, paving the way for data-centric and AI-native research paradigms in materials science.

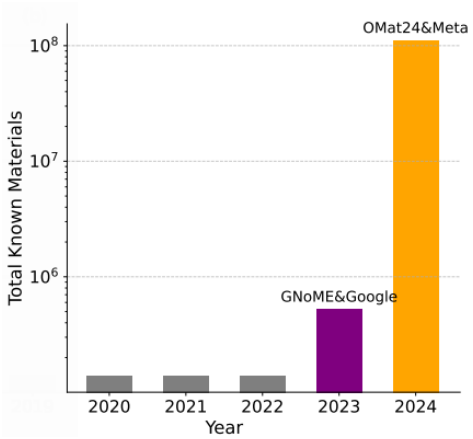


Figure 2. *The variation in the total known materials over time on a logarithmic scale, highlighting the acceleration of material discovery processes facilitated by GNoME [2] of Google and OMat24 [3] of Meta2 [4]. This figure underscores the rapid development within the field of machine learning materials discovery Reprinted from ref [5]. Copyright 2024 IOP Publishing*

In this review, we focus on the transformative role of ML and AI in polymer science from 2023 to mid-2025. We begin by outlining the application of supervised, unsupervised, and reinforcement learning techniques in polymer design, property prediction, structural analysis, and sustainable material development. We then explore emerging approaches such as high-throughput screening, inverse design, generative models, and multi-scale modeling. Particular attention is given to recent advances in explainable AI, physics-informed neural networks, and automation. Finally, we discuss key challenges—including data scarcity and interpretability—and highlight future opportunities enabled by open databases, transfer learning, and autonomous laboratories.

2. State of the Art in AI-Driven Research in Chemistry and Polymer Science

At present, numerous review articles have been published that focus on the application of artificial intelligence (AI) in chemistry. This comprehensive review outlines the latest developments in the application of artificial intelligence (AI) and machine learning (ML) in chemistry [6]. The authors provide an in-depth overview of various ML methods—including neural networks and ensemble models—and their use in property prediction, reactivity analysis, and synthesis automation. The article serves as a foundational reference for understanding how AI is transforming modern chemical research. The review [7] highlights the top twenty most impactful AI technologies in chemistry, from autonomous robotic chemists and molecular generative models to digital reaction databases and intelligent synthesis platforms. It emphasizes the transformative potential of these innovations in accelerating discovery and improving reproducibility in chemical science. The Kibrete et al. (2023) [8] explore the role of AI in predicting the mechanical behavior of composite materials using various ML algorithms, including regression models and decision trees. This study demonstrates the growing importance of AI in materials science, particularly in the design and optimization of advanced composites. The cutting-edge research [9] integrates large

language models (LLMs), such as ChatGPT, with specialized chemistry software to create hybrid systems capable of performing analytical tasks, generating reaction schemes, and assisting in scientific writing. The paper illustrates the potential of combining natural language processing with domain-specific tools for enhanced chemical reasoning. Käser et al. (2023) [10] introduces neural network potentials (NNPs), which are trained to predict potential energy surfaces with near-quantum accuracy at a fraction of the computational cost. Applications range from reaction simulations to materials discovery, positioning NNPs as a powerful tool for accelerating molecular modeling. Coley et al. (2019) [11] present a graph-convolutional neural network (GCNN) architecture that learns molecular reactivity directly from graph representations of molecules. Trained on extensive reaction datasets, the model enables accurate and generalizable predictions of reaction outcomes, marking a major step toward autonomous retrosynthesis planning. Vora et al. (2023) [12] reviews recent applications of AI in pharmaceutical science, particularly in drug formulation, delivery system design, and personalized medicine. The use of deep learning and predictive models is shown to significantly streamline drug development workflows and enhance treatment effectiveness.

Together, these references provide a broad overview of the state-of-the-art in AI-driven chemistry, illustrating how diverse approaches—from graph neural networks to hybrid LLMs—are reshaping the discovery, design, and deployment of chemical compounds and materials.

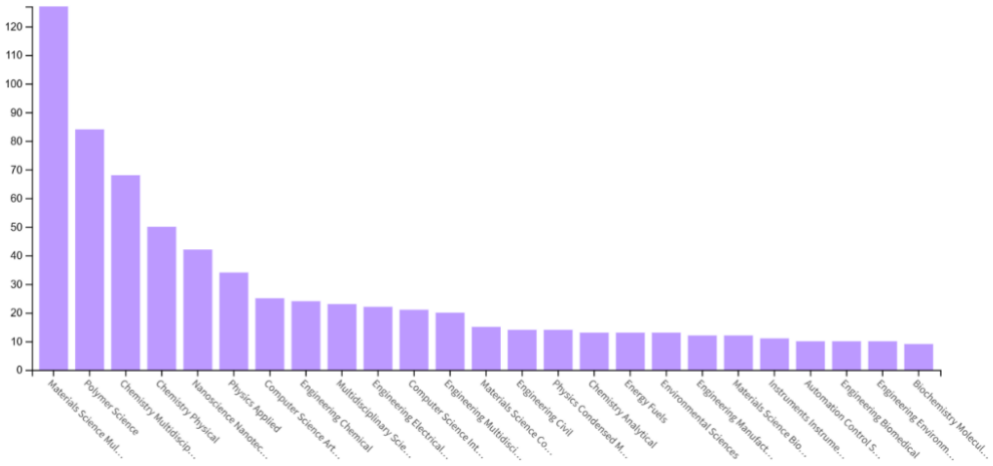
In recent years, there has been a growing number of comprehensive review articles exploring the application of artificial intelligence (AI) and machine learning (ML) in polymer science. These reviews cover a wide range of topics, including the prediction of polymer properties, polymer design and discovery, process optimization, and the development of structure–property relationships. The increasing availability of polymer databases, advances in computational power, and the integration of data-driven approaches into materials research have all contributed to the rapid expansion of this interdisciplinary field. As a result, AI is now playing a pivotal role in accelerating innovation in polymer chemistry and engineering.

Han et al. (2025) [5] provides a comprehensive perspective on AI-driven inverse materials design, with implications for polymer science. The authors describe how inverse design frameworks—leveraging generative models, optimization algorithms, and structure–property prediction—enable the targeted development of materials with predefined properties, thus accelerating discovery in polymer chemistry. Jayaraman and Olsen (2024) [13] commentary highlights the growing intersection between AI, ML, cheminformatics, and macromolecular science. The authors advocate for cross-disciplinary collaboration and emphasize how data-centric approaches are reshaping polymer research—from synthesis planning and experimental design to property prediction and performance optimization. Ferji (2025) [14] presents a foundational overview of AI tools as applied to polymer science, including supervised and unsupervised learning, neural networks, and generative models. The article also introduces essential polymer descriptors and datasets, making it an accessible starting point for researchers entering the field. The review [15] explores emerging ML strategies in polymer science, focusing on areas such as polymer informatics, molecular representations (e.g., SMILES, BigSMILES), and property prediction. It also outlines future challenges, including data scarcity and model generalization, while offering solutions such as transfer learning and active learning. van den Hurk et al. (2025) [16] discusses the application of AI and ML in polymer characterization, particularly in chromatography, spectroscopy, and imaging. The authors detail how advanced algorithms can extract complex features from high-dimensional datasets, improving the accuracy and efficiency of polymer analysis. Long and co-authors [17] review recent advancements in the use of AI for polymer material development. The work categorizes AI applications into synthesis prediction, molecular design, performance evaluation, and sustainability assessment, providing a broad and up-to-date overview of the field. Published in Nature Reviews Materials, the landmark article [18] illustrates how AI is being harnessed to design functional and environmentally sustainable

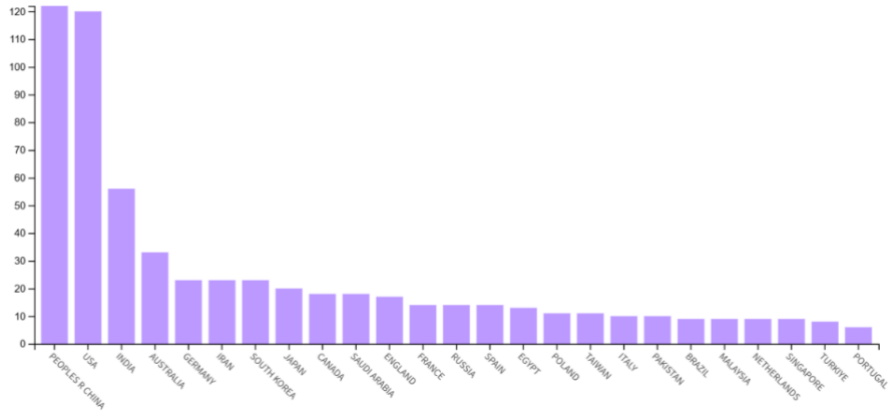
polymers. The authors emphasize generative approaches, high-throughput screening, and life-cycle analysis, making the case for AI as a powerful enabler of green polymer science.

Together, these references demonstrate that the fusion of AI and polymer science is no longer a future vision but an active and rapidly developing area of research. They highlight how data-driven models are enabling smarter material discovery, more efficient experimentation, and better-informed sustainability decisions in the polymer field.

Recent breakthroughs in deep learning architectures—such as graph neural networks (GNNs), variational autoencoders (VAEs), generative adversarial networks (GANs), and diffusion models—have shown exceptional promise in modeling complex material systems. These models can not only predict key properties such as bandgap, formation energy, and mechanical strength but can also generate new molecules or crystal structures with desirable traits. The rise of foundational models, such as Google's GNoME and Meta's OMat24, highlights the growing synergy between large-scale AI frameworks and materials discovery.



Research areas



Countries/Regions

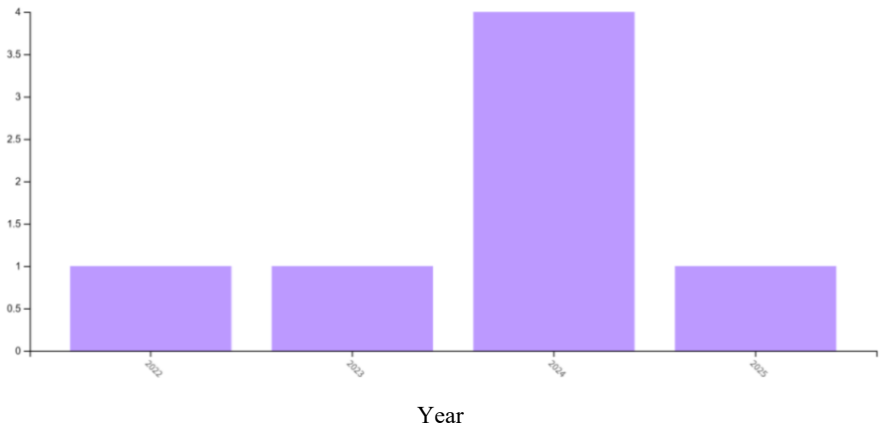


Figure 3. *The global research output on the integration of artificial intelligence in polymer science has been steadily increasing*

Moreover, explainable AI (XAI), active learning, and reinforcement learning are being increasingly adopted to refine models in real-time, guide experiments efficiently, and ensure interpretability—an essential requirement in scientific domains. Coupled with high-throughput workflows and robotic laboratories, AI-driven inverse design is now being positioned not merely as a tool for acceleration, but as a catalyst for paradigm shifts in how materials research is conducted.

Despite these advancements, several challenges remain. The quality and quantity of available data, the transferability of models across different material domains, the integration of physics-based constraints into data-driven frameworks, and the validation of predictions through experiments are all active areas of investigation. Additionally, existing reviews often focus on narrow aspects, such as specific machine learning methods or material classes, lacking a unified view across disciplines and paradigms.

Figure 3 illustrates the growing interest in the application of artificial intelligence (AI) within polymer science, based on publication data extracted from the Web of Science database. The search was conducted using the keywords "machine learning", "polymer" and "artificial intelligence", limited to the research areas of Materials Science, Polymer Science, and Chemistry. The data show a marked increase in publication activity over recent years. Additionally, the accompanying geographical distribution map identifies the leading regions contributing to this rapidly evolving field, with prominent contributions from countries within the European Union (EU), North America, and East Asia.

Polymer science stands at the frontier of data-driven innovation. Traditional approaches—empirical experimentation, theoretical modeling, and simulation—are now being complemented and often outperformed by machine learning models. With the rise of open-access databases, advanced algorithms, and automated tools, the vision of predictive and rational polymer design is becoming a reality. The composition–process–structure–property–performance relationship (CPSPPr) provides a foundational framework within which ML methods can extract correlations and generate actionable insights. ML offers the ability to explore vast chemical design spaces, reduce trial-and-error, optimize manufacturing processes, and enable discovery of polymers with tailored properties for specific applications, from biodegradable plastics to high-performance composites.

The CPSPPr framework offers a comprehensive method for connecting molecular-scale features with macroscopic properties. Traditionally, this relationship is explained through statistical thermodynamics, where a property Z can be expressed as:

$$Z = \sum_i z_i \rho_i = \frac{\sum_i z_i e^{-\beta E_i}}{\sum_i e^{-\beta E_i}} \quad (1)$$

Here, ρ_i represents the probability density (or population) of state *i*, z_i is the contribution of that state to the property, and E_i is the corresponding energy level. The factor $\beta=1/k_BT$, where k_B is the Boltzmann constant and T is the absolute temperature.

In machine learning (ML) approaches, a similar formulation can be derived from Bayes' theorem, where the property Z is defined in probabilistic terms as:

$$Z = \sum_i z(x) P(x|D) dx \quad (2)$$

Here, P(x|D) is the posterior distribution of the input features *x* given the observed data *D*, and z(x) is the predicted property value. This reflects the probabilistic nature of ML models, where properties emerge from weighted contributions of many possible structural or feature states.

In ideal conditions, the dataset D conforms to statistical thermodynamics, often following Gaussian or Maxwell–Boltzmann distributions. More generally, however, D may exhibit alternative forms such as log-normal, Pareto, exponential, or Poisson distributions. The three components of Equation (2) reflect major research priorities: the development of polymer materials, fundamental studies in quantitative structure–property relationships (QSPR), and the evolution of multiscale structures in polymer systems.

3. Machine Learning Applications in Polymer Science

Fundamental research in polymer science focuses on uncovering the physical mechanisms underlying the composition–processing–structure–property–performance relationship (CPSPPr). To this end, ML models are increasingly employed to reveal the deterministic relationships among structures by tuning compositional and processing parameters [19]. These models are trained on datasets obtained through both experimental characterizations and physics-based multiscale simulations—complementary strategies that serve as the foundational pillars for data generation in this field [20-22].

At the core of ML-driven polymer science lies the exploration of quantitative structure–property relationships (QSPR). Establishing a robust and generalizable QSPR model is essential, as it enables high-throughput screening (HTS) via either experimental or computational pipelines. This, in turn, facilitates the identification of hidden patterns and correlations, as well as the discovery of new polymeric compositions and architectures exhibiting desirable properties [23-25].

As illustrated in Figure 4, ML methods are typically grouped into three fundamental categories: supervised learning (SL), unsupervised learning (UL), and reinforcement learning (RL). These paradigms differ in terms of data requirements, the extent of human involvement, and their suitability for specific problem types. In supervised learning, models are trained on labeled datasets to predict outcomes or classify data. Unsupervised learning, in contrast, seeks to uncover hidden patterns or structures in unlabeled data. Reinforcement learning involves an agent interacting with an environment, learning to make sequential decisions based on feedback in the form of rewards or penalties.

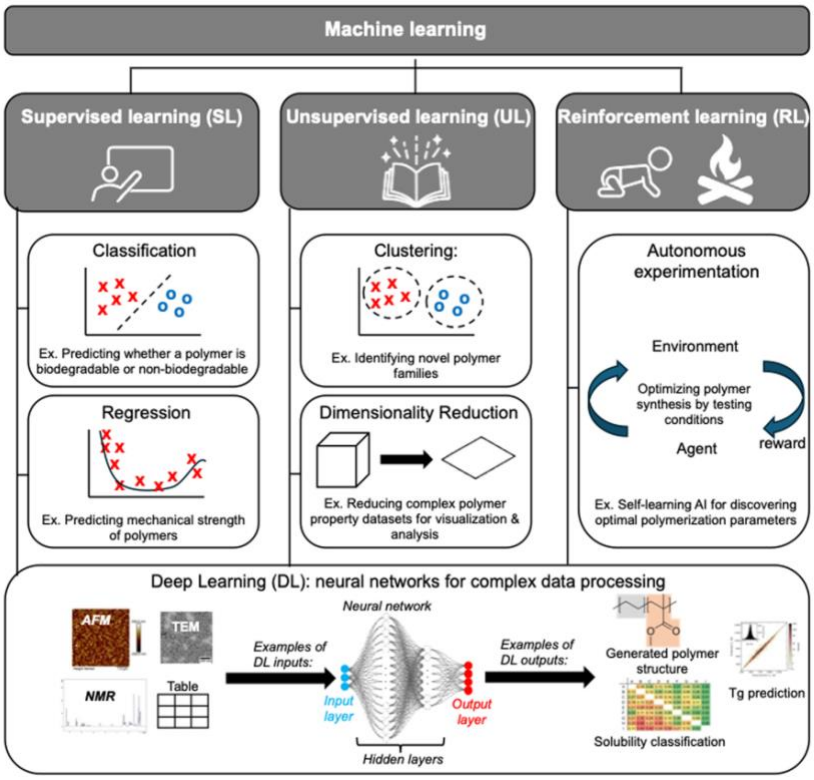


***Figure 4**. A summary of the primary machine learning (ML) techniques employed in polymer science, along with their key areas of application. Reprinted from [13]. Copyright 2025 The Royal Society of Chemistry*

3.1. Characterization of Structures and Properties

In polymer physics and engineering, a major objective is the theoretical approximation and experimental validation of a polymer's colligative, volumetric, cohesive, thermodynamic, and transport properties. These properties are fundamentally linked to molecular interactions, conformational preferences, aggregation phenomena, and phase behaviors. Characterization serves as a means to both quantify the spatial distribution of ingredients in the final polymer structure and to correlate these distributions with macroscopic performance metrics [26-28].

The process of structural and property characterization typically involves signal generation through a specific detection mechanism, its interaction with the sample, and subsequent data interpretation—whether via spectroscopy, microscopy, or scattering techniques. Figure 5 schematically illustrates how both experimental and simulation-based approaches are capable of capturing structural or property data across diverse spatial and temporal scales [20-31].

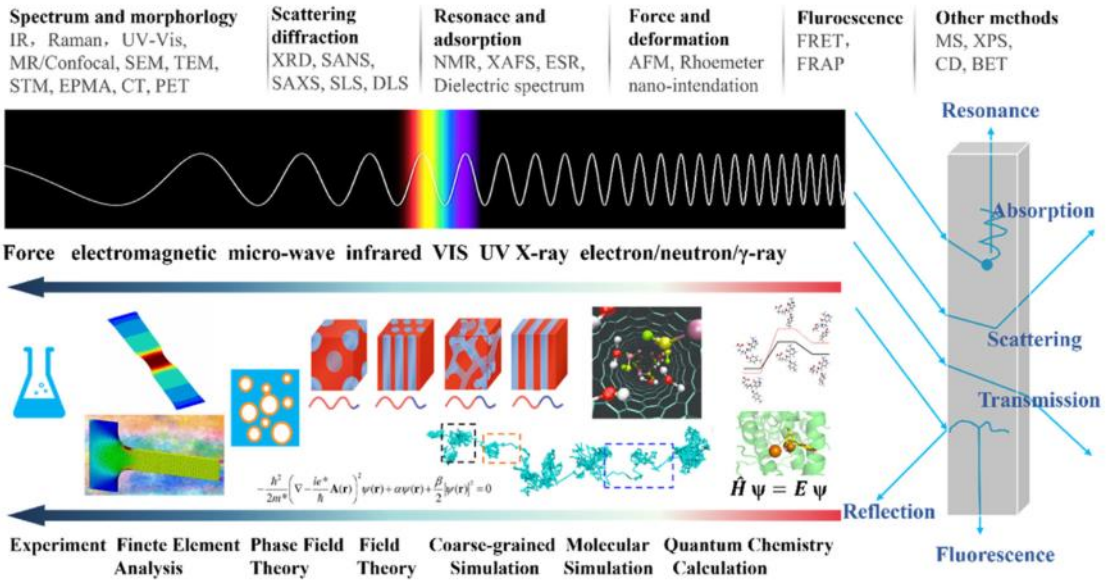


Figure 5. *Experimental characterization and computation of structure and properties at multi-scales. Reproduced with permission: Copyright 2022, Acta Polymerica Sinica [22]*

Although experimental and simulation techniques operate on different physical principles, the resulting outputs—such as spectra or images—can be universally treated as "spectrum data" [32]. In this context, machine learning models offer a promising route to accelerate and enhance the extraction of meaningful information from raw data, often achieving both higher speed and accuracy compared to traditional methods. Moreover, ML facilitates the elucidation of robust and physically interpretable relationships between structures and properties [28].

Machine learning (ML) techniques have increasingly become integral to polymer science, especially in tasks such as synthetic data generation, component identification, 3D structure reconstruction, and property prediction from microscopy data (STM, SEM, TEM, AFM, and confocal imaging) [33]. Structural and property characterization often relies on spectral fingerprints that reveal inherent chemical and physical features of materials. For instance, support vector machines (SVMs) have been successfully used for rapid and precise cell counting based on impedance signals in microfluidic cytometry [34]. Similarly, graph-based neural networks can predict near-edge X-ray absorption fine structure (XAFS) spectra with 1 eV resolution, enabling the identification of charge states, coordination numbers, and local symmetries, trained using datasets like QM9 [35].

Tools like MZmine2 enhance polymer identification via Kendrick mass defect analysis from mass spectrometry data [36, 37]. Deep learning models trained on SEM, TEM, or optical microscopy images have been applied to determine molecular orientations in P3HT films and 2D Materials [38, 39], recognize poly(L-lactic acid) spherulites [40], count micro-droplet sizes [41], monitor corrosion-resistant coatings on PVA-silver surfaces [42], and reconstruct copolymer architectures [43-45]. Moreover, AFM data have been used with ML to extract crucial physical parameters such as contour length, persistence length, mesh size, and fractal dimension of polymer networks [46, 47].

Recent reviews have highlighted how ML enhances crystal and macromolecular structure analysis via X-ray crystallography [48, 49], as well as data handling and structural interpretation in techniques like rheology and small-angle scattering [50].

In applied polymer science, ML contributes to life-cycle assessment, structural health monitoring (SHM), and fault detection. Methods like acoustic analysis, infrared thermography, and terahertz imaging are now combined with ML for diagnostics. Spectral data (near-IR, Raman, UV–

vis) are used to monitor manufacturing processes such as hot-melt extrusion [51-54] and detect early-stage faults in polymer electrolyte membrane fuel cells (PEMFCs) [55-57]. In display technologies, ML models identify defects like foreground/background inconsistencies [58] or Mura patterns in OLED and polymer displays using sequential classifiers and transfer learning [59-61]. For nanolithography, hybrid models that merge neural networks and ridge regression enable detailed analysis of defect formation and block copolymer self-assembly via solvent annealing [62, 63].

3.2. Polymer Design and Property Prediction

ML models are increasingly being used to predict key polymer properties such as glass transition temperature (Tg), melting temperature (Tm), dielectric constants, solubility, thermal stability, and mechanical performance. Traditional QSPR (quantitative structure–property relationship) models have evolved into sophisticated frameworks utilizing deep learning, support vector machines (SVM), random forest (RF), and Gaussian process regression (GPR) [64-71].

Recent work by Amamoto et al. (2025) [72] demonstrated multi-objective optimization of multiblock polyamides using GPR, enabling a balance between biodegradability and toughness. Similarly, deep neural networks (DNNs) have been trained to predict Tg from polymer backbones using features extracted from SMILES and BigSMILES representations. This format, illustrated in Figure 6, enables more accurate encoding of polymer structures for ML applications. Transfer learning from small-molecule databases is now commonly used to initialize models, accelerating learning for polymers.

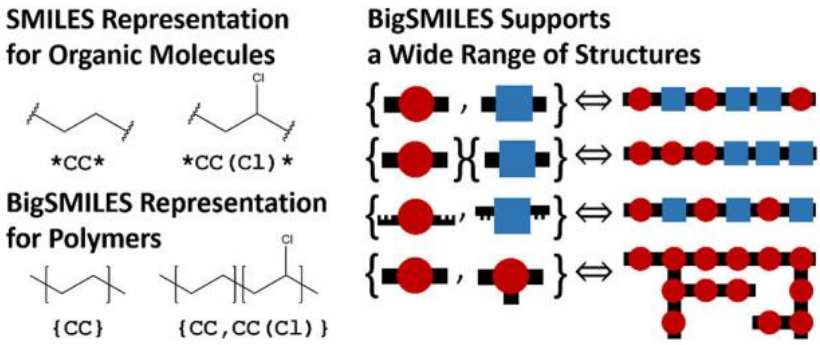


Figure 6. *Depictions demonstrating how BigSMILES captures different polymer chemistries. Reprinted from ref [73]. Copyright 2019 American Chemical Society*

The initial dataset comprised data on 66,981 distinct entries related to polymer materials, encompassing 18,311 unique polymers and 99 different physical properties, each with varying degrees of available information [74, 75]. A key element of the dataset is the inclusion of molecular representations using the Simplified Molecular Input Line Entry System (SMILES). Figures 7 and 8 illustrate the distribution of non-missing values for each property across the dataset. The x-axis lists the names of the physical characteristics, while the y-axis indicates the number of available (non-null) data points for each. Numerical labels above each bar further highlight the quantity of known values, offering insights into the overall data completeness.

Once a structured dataset has been compiled, detailed data and feature engineering is essential to enhance its quality. In this context, the dataset can be viewed through the lens of linear algebra, where data vectors represent constraints, and features and target properties function as variables in a multivariate correlation system. The ultimate objective is to construct an accurate function—either explicit or implicit—that maps composition, processing, structure, properties, and performance (CPSPPr) relationships in polymers using ML techniques.

A foundational step in this process is the statistical analysis of the dataset, using metrics such as mean, median, interquartile range (IQR), variance, skewness, and kurtosis to assess distribution characteristics. To ensure data reliability, preprocessing steps are needed to handle missing values and outliers. Several advanced imputation methods have been developed for this purpose, including Multiple Imputation by Chained Equations (MICE), MissForest, HyperImpute, GAIN, Sinkhorn, and RDE_W [77-80].

Several publicly accessible polymer databases have been established to facilitate data-driven research and machine learning applications in polymer science. Polymer Genomes (https://www.polymergenome.org/) provides computational and experimental property data for a broad range of polymers, integrated with machine learning-based predictive models [81]. PoLyInfo (https://polymer.nims.go.jp/) offers annotations of approximately 100 experimentally determined properties for each polymer, serving as a rich source for structure–property relationships [82].

The Polymer Property Predictor and Database (PPPDB) (https://pppdb.uchicago.edu/) includes Flory–Huggins interaction parameters (χ) and glass transition temperatures (Tg), along with prediction tools. The Material Properties Database (https://www.makeitfrom.com/) compiles mechanical, thermal, and electrical properties of polymer materials. Similarly, the CROW Polymer Properties Database (http://polymerdatabase.com/) offers detailed information on polymer structures, properties, and applications.

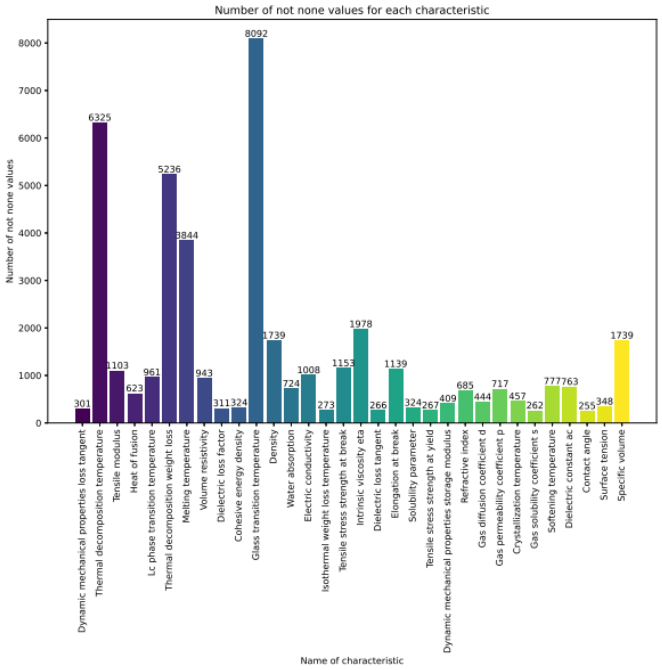


***Figure** 7. Count of non-null values for each characteristic across the dataset for a count exceeding 250. Reprinted from ref [76]. Copyright 2023 Licensee MDPI*

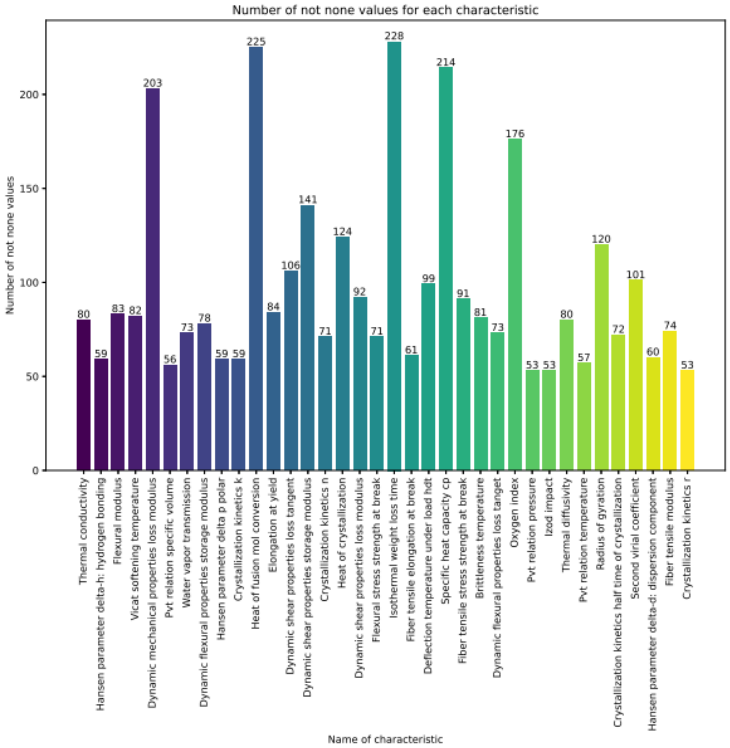


***Figure 8**. Count of non-null values for each characteristic across the dataset for a count ranging from 50 to 250. Reprinted from ref [76]. Copyright 2023 Licensee MDPI*

For machine learning benchmark tasks, the PI1M dataset (https://github.com/RUIMINMA1996/PI1M) provides curated data to solve variant problems in polymer informatics [83]. The AI Plus Polymers database (https://polymergenome.ecust.edu.cn/database/resin) compiles data on resin systems, including reactions, molecular structures, and performance metrics. Lastly, the Polyimide Fundamental Properties Database (http://sun.ciac.jl.cn:8888/) focuses on the physical and chemical properties of polyimide materials. These databases form a critical infrastructure for modern polymer informatics, enabling large-scale data collection, efficient model development, and accelerated materials discovery.

3.3. Strategies and Algorithms for Model Construction

Machine learning (ML) strategies in polymer science span the full pipeline—from data collection to application—and aim to optimize both efficiency and model performance. Approaches such as orthogonal design (e.g., Taguchi methods, response surface methodology, Pareto optimization) help minimize trial-and-error by providing broad coverage with fewer experiments [84].

Combining theoretical models and empirical knowledge—such as human intuition, thermodynamic principles, and computational modeling—can further improve predictive accuracy. Given the complex, often nonlinear and far-from-equilibrium nature of polymer systems, there is currently no universal ML algorithm that reliably models chemical–processing–structure–property–performance–relationships (CPSPPr), in line with the No Free Lunch theorem [85].

ML strategies vary widely, encompassing supervised, unsupervised, and reinforcement learning, as well as inductive, deductive, transductive, ensemble, transfer, and federated learning methods. Deep learning, time-series models, and neural networks are also commonly used.

A typical ML workflow involves:

1. Data splitting for training, validation, and testing using methods like K-Fold, Monte Carlo, or time-series cross-validation.
2. Feature engineering via molecular descriptors and encoding techniques to capture structural, process, and measurement variables.
3. Model building using single or multiple ML algorithms, often optimized through evolutionary techniques like Gaussian Processes or Bayesian Optimization [86].
4. Evaluation and deployment, including blind testing, virtual screening, and rational/inverse materials design.

These approaches have been successfully applied in areas such as dielectric polymer design [87], high-throughput optimization of plastic injection molding [88], screening of new materials [89], and development of fiber-reinforced composites [90].

Recent innovations include advanced polymer representations like BigSMILES and BigSMARTS [73], tools such as RadonPy and SMiPoly for feature extraction [91], and graph-based methods for predicting electronic properties [92]. Emerging deep learning models like polyBERT and polyGNN [93], as well as chemical-language models like polyNC [94], are enhancing catalyst design and materials optimization workflows [95].

Automation tools such as Hyperopt [96], Optuna [97], Scikit-Optimize [98], Auto-sklearn, and PyCaret [99] are further accelerating ML deployment by simplifying model construction and hyperparameter tuning.

Deep learning, a specialized branch of machine learning (ML), has garnered significant attention in recent years. Originating from neural networks (NNs) based on the McCulloch–Pitts neuron model developed in the 1940s, deep learning has evolved into various frameworks and architectures. These include artificial NNs, backpropagation NNs, convolutional NNs (CNNs), deep and forward-propagation NNs, graph NNs, K-means clustering networks, multilayer perceptrons (MLPs), and recurrent NNs (RNNs), each designed to simulate the cognitive processes of the human brain through a combination of linear signal weighting, summation, and nonlinear activation functions.

A major breakthrough came with the introduction of the backpropagation (BP) algorithm in 1986, which—alongside forward propagation and loss functions—forms the foundation of contemporary deep learning frameworks [100]. CNNs and RNNs are particularly adept at capturing dynamic and sequential relationships in data. For instance, long short-term memory (LSTM) networks are widely used to model time-dependent behavior. Other influential architectures include generative adversarial networks (GANs), which utilize a pair of networks (generator and discriminator) to generate realistic data [101], and variational autoencoders (VAEs), which employ encoder–decoder pairs to map complex latent spaces [102].

The transformer architecture, enhanced with multi-head attention mechanisms, has recently demonstrated exceptional potential in the accelerated discovery and optimization of new polymers [103]. These models are particularly well-suited for handling complex relationships in composition–processing–structure–property–performance relationships (CPSPPr) for various polymer systems.

3.4. High-Throughput Screening and Inverse Design

Machine learning (ML) has significantly accelerated high-throughput virtual screening (HTVS) of large polymer libraries, enabling the identification of candidates with tailored mechanical, thermal, and optoelectronic properties. In particular, inverse design approaches—such as variational autoencoders (VAEs), generative adversarial networks (GANs), and reinforcement learning (RL)—have emerged as powerful tools for generating novel polymer structures conditioned on desired performance metrics.

Recent advances leverage sequence-based deep learning models like polyBERT and polyGNN, which translate polymer repeat units into vector representations enabling accurate

sequence-to-property predictions. These models have been successfully applied to copolymers, polyimides, conjugated polymers, and ion-conducting polymers.

In 2024, a notable study integrated active learning with Bayesian optimization to iteratively explore chemical space, leading to the discovery of novel thermoplastic elastomers exhibiting high elasticity, thermal stability, and solvent resistance. Moreover, the study incorporated multi-fidelity simulations to reduce the need for expensive DFT calculations in early screening phases.

Further, multi-objective optimization frameworks have been developed to balance trade-offs such as strength vs. ductility or conductivity vs. processability, with growing integration of physics-informed neural networks (PINNs) to enforce thermodynamic consistency during inverse design.

In 2023, efforts also expanded toward self-healing polymers and vitrimers, where ML models were trained on curated databases of crosslinking densities, topological rearrangement kinetics, and activation energies. Generative models such as MolGPT and ChemCrow have also been fine-tuned specifically for polymer design, offering language-based control over polymer architecture generation [104, 105].

Additionally, the integration of experimental robotics platforms (e.g., self-driving labs) with inverse design pipelines has started to close the loop between prediction, synthesis, and validation, reducing polymer development cycles from months to days.

3.5. Multi-Scale Modeling and Simulation in Polymer Systems

To comprehensively understand the chemical–processing–structure–property–performance relationships (CPSPPr) in both natural and synthetic polymers, it is crucial to integrate multi-scale information—ranging from electronic and atomic scales to monomeric units, domains, aggregates, and their spatial orientations. A bottom-up approach to modeling these systems employs a wide spectrum of simulation techniques including quantum mechanics (QM), molecular mechanics (MM), molecular dynamics (MD), Monte Carlo (MC), dissipative particle dynamics (DPD), self-consistent field theory (SCFT), phase field theory, finite element analysis (FEA), and computer-aided design (CAD). These methods allow the investigation of key phenomena such as inter- and intra-molecular interactions, force field and conformation dynamics, self-assembly, phase behavior, and structure-function relationships derived from both simulation and experimental characterization.

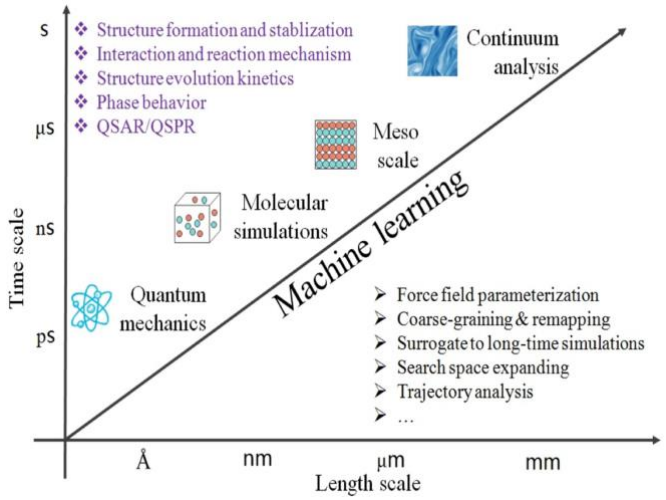


Figure 9. *Integration and development of machine learning (ML) approaches to study multi-scale problems in polymer science. Reprinted from [107]. Copyright 2025 Wiley*

As shown in Figure 9, machine learning (ML) methods are increasingly being integrated into these simulation workflows to approximate or solve complex numerical problems. For example, ML accelerates ab initio and density functional theory (DFT) calculations that aim to solve the Schrödinger equation and optimize the Hamiltonian in QM simulations. ML also supports the modeling of dissipative forces in lattice Boltzmann (LB) and DPD simulations. Notable developments have occurred in force field–based modeling methods such as CHARMM, AMBER, OPLS, MARTINI, and REAXFF, which bridge scales from QM to atomistic and coarse-grained representations. These techniques require robust, self-consistent parameters for topology and energy terms to achieve a balance between computational efficiency and accuracy [106].

ML has made significant contributions in several areas of multi-scale modeling, including: parameterizing force fields, developing coarse-grained models, and enabling reverse mapping; reducing computational cost via surrogate models that extrapolate from low-resolution or short-time data to higher resolution or long-time behavior; enhancing simulation efficiency by accelerating equilibrium convergence and sampling; and facilitating feature extraction from simulation trajectories. These advances have been reviewed in the context of molecular modeling, retrosynthesis, catalysis, drug discovery, and materials design [108, 109]. These approaches have been evaluated to broaden insights across molecular modeling, retrosynthetic analysis, catalysis, and pharmaceutical design. They also aim to improve simulation precision, lower computational demands, support coarse-grained modeling and inverse mapping, and enable more accurate representations by incorporating angular and directional dependencies.

Interactions lie at the heart of polymer system modeling. Traditional concepts such as the Flory–Huggins interaction parameter (χ), Hildebrand solubility (δ), and Hansen solubility parameters are widely used to assess solubility, miscibility, and phase behavior. These parameters can be obtained through experiments or quantum mechanical calculations. However, empirical rules like "like dissolves like" offer only limited predictive power, especially in systems with polar or hydrogen bonding interactions. Recent studies have applied machine learning (ML) to predict interaction parameters, revealing that conventional heuristics often fail outside specific polymer–solvent subclasses.

ML has also been employed to understand complex behaviors such as polymer–surface adhesion and many-body interactions in polymer–colloid mixtures. By combining thermodynamic principles (e.g., minimizing Gibbs free energy) with ML, models now support multi-scale simulations, identifying key conformational transitions like the coil-to-globule shift and estimating critical temperatures such as LCST.

Advanced neural network models, including dPoly and Conformer-RL, have enabled efficient classification of polymer conformations and identification of structural features—such as weak dihedral angles—that influence macroscopic properties like radius of gyration. Coupling ML with coarse-grained molecular dynamics (CG-MD) has further advanced the simulation of complex sequence-defined polymers and metastable states.

Phase behavior modeling has also benefited from ML. Neural networks and Gaussian process regression have accurately predicted phase boundaries, cloud points, and spinodal decomposition in binary and ternary systems. Combining ML with self-consistent field theory (SCFT) allows rapid exploration of phase space and efficient solutions to Hamiltonian functions, facilitating the high-throughput screening of polymer architectures.

The development of accurate force fields is crucial for reliable simulations. ML-enhanced models—whether atomistic, coarse-grained, or hybrid—can now incorporate experimental spectroscopy (e.g., IR, UV–vis, fluorescence) to validate force field parameters or serve as constraints. Long short-term memory (LSTM) networks and deep neural networks (DNNs) have been used to refine CG potentials and reduce the need for costly DFT calculations.

Efforts in inverse mapping and fast property prediction are accelerating. ML techniques such as artificial neural networks (ANN) and particle swarm optimization (PSO) have improved

coarse-graining and back-mapping workflows for various polymers, including polyisoprene and polybutadiene systems. From simulation trajectories, ML can extract thermomechanical properties, identify topological defects, classify local environments, and map grain distributions.

Combining ML with experimental and simulation data is also proving effective. Unsupervised learning accelerates defect detection in self-assembling systems, while supervised models can interpret scattering data (SANS/SAXS), quantify crystallinity, or predict ionic mobility in conductive polymers. These techniques also enable structure–property predictions in shape-memory and composite materials.

Overall, the fusion of ML with multi-scale modeling frameworks has enabled deeper understanding and broader exploration of polymer systems, streamlining the discovery of new materials with tailored functionalities [110-112].

4. ML-Driven Discovery of Polymer Materials

Polymers exhibit a broad spectrum of properties—including mechanical, thermal, optical, electrical, and biological features—alongside advantages in processability, durability, and environmental sustainability. When tailored to meet specific performance criteria, they serve as essential components in various material systems such as solutions, blends, composites, and dispersions. Given the high-dimensional property space defined by application-specific requirements, identifying optimal polymers presents a vast and complex search challenge.

Machine learning (ML) has greatly enhanced the efficiency of exploring this space, accelerating the discovery of polymers with targeted functionalities within the CPSPPr (composition–processing–structure–property–performance–relation) framework.

4.1. Structural Materials

The thermal and mechanical behavior of polymers plays a crucial role in the development, performance, and recyclability of structural materials. Thermal analysis techniques such as DSC, TGA, DMA, and conductivity testing help identify transitions (e.g., Tg, Tm, Td, LCST/UCST) and detect physical or chemical changes with temperature. Mechanical properties—evaluated via tensile, flexural, fatigue, impact, or rheological testing—are typically reported through stress–strain profiles and vary with time, temperature, and loading conditions.

These properties are governed by polymer structure, crosslinking, molecular orientation, and entanglements. Despite this complexity, ML approaches have shown considerable promise in predicting key parameters such as glass transition temperature (Tg), elastic modulus, and thermal conductivity. Compared to classical FEA-based modeling, ML models offer better flexibility and can capture non-linear, multi-scale correlations. For instance, hybrid models integrating physical knowledge and data-driven insights have been effective in predicting Young's modulus and rheological behavior in polymer systems.

In recent years, ML has been employed to design flame-retardant composites, predict flammability, and optimize melt index in industrial polymerizations. Combining ML with quantum descriptors (e.g., bond energy, HOMO–LUMO gaps) and DFT calculations has also enabled the design of thermally stable and processable epoxy resins.

Fiber-reinforced polymers (FRPs) are among the most studied structural composites. Their mechanical performance is influenced by the choice of fiber type (e.g., carbon, aramid, glass), polymer matrix, interfacial adhesion, fiber alignment, and manufacturing method (e.g., injection molding, 3D printing). ML models—including ANN, DNN, Techniques such as AdaBoost [113], Gradient Boosting [114, 115], LightGBM [116, 117], CatBoost [118, 119], and XGBoost [120, 121] have demonstrated strong capabilities in handling large-scale, intricate datasets by effectively capturing non-linear patterns. Within polymer science, these algorithms have been successfully utilized for a range of applications, including the prediction of material properties, optimization of manufacturing parameters, and development of novel polymer formulations.

The Keywords Occurrence Map (Figure 10) demonstrates the growing application of boosting algorithms within polymer research. Approaches such as gradient boosting, XGBoost, CatBoost, and AdaBoost have become prevalent for modeling and optimizing properties like compressive strength (CS), bonding performance, and mechanical behavior. These models are particularly effective in handling complex and non-linear datasets, thereby improving prediction accuracy and robustness. Their use spans a variety of domains, including sustainability-related materials (e.g., geopolymers, fly ash composites, and asphalt binders), advanced manufacturing techniques (e.g., additive manufacturing and 3D printing), and high-performance construction materials (e.g., fiber-reinforced polymers and ultra-high-performance concrete). Furthermore, boosting methods are frequently integrated with interpretability tools such as SHAP and sensitivity analyses to gain deeper insights into influential variables affecting phenomena like photodegradation, corrosion, and the environmental behavior of microplastics.

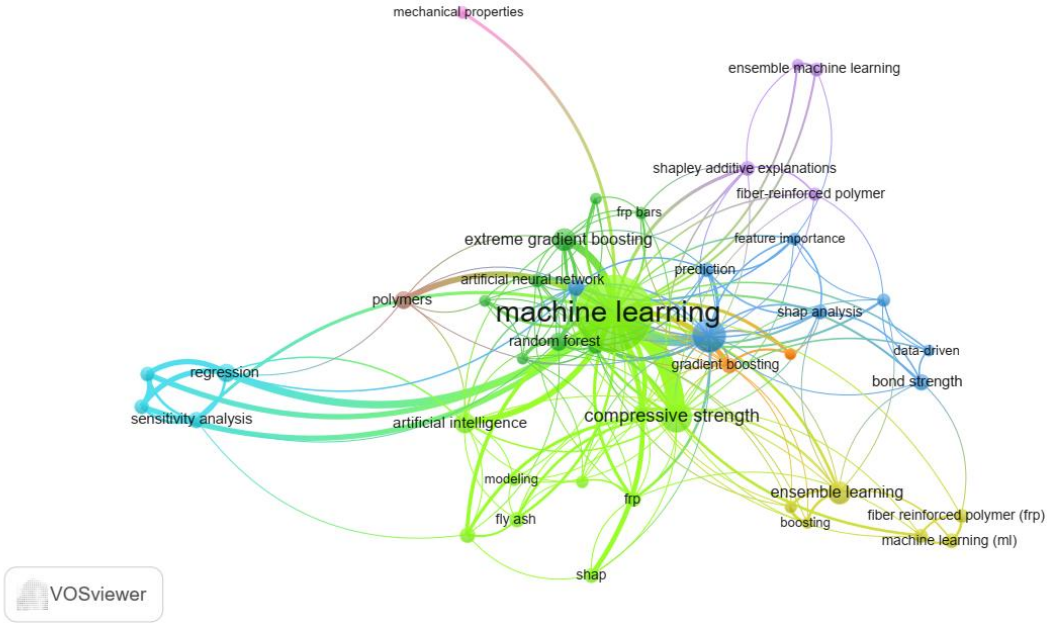


Figure 10. *Keyword co-occurrence map based on VOSviewer analysis. . Reprinted from ref [122]. Copyright 2025 Licensee MDPI*

For structural health monitoring (SHM), ML-based classification and regression models—using data from IR thermography, X-ray CT, terahertz imaging, and acoustic emission—have enhanced damage detection capabilities. These models can distinguish between various defect types such as delamination, fiber breakage, or matrix smearing, and track their evolution under cyclic loads or environmental stressors.

In the construction sector, where FRPs reinforce concrete structures, ML techniques have been applied to model pull-out strength, bond behavior, and failure modes. Gradient boosting, SHAP analysis, and interpretable ML models have helped identify key contributors to performance metrics such as load-bearing capacity and shrinkage. Such data-driven insights are now central to the rational design of durable, high-performance structural composites [123-127].

4.2. Machine Learning for Energy Polymer Materials

Growing concerns over sustainability and carbon neutrality have accelerated the development of advanced energy technologies such as organic solar cells (OSCs), proton and anion exchange membrane fuel cells (PEMFCs and AEMFCs), lithium-ion and post-lithium batteries, vanadium redox flow batteries, and supercapacitors. Polymers play essential roles in these systems—as semiconductors, electrolytes, dielectrics, membranes, or adhesives—and machine

learning (ML) has emerged as a powerful tool to assist in their discovery, optimization, and performance prediction.

4.2.1. Key ML Tasks in Energy Materials

The integration of machine learning (ML) in energy materials research has evolved rapidly, catalyzing innovations across materials discovery, property prediction, and system-level optimization. Recent advances emphasize not only improved prediction accuracy but also the interpretability, generalizability, and physical plausibility of ML models. These developments are particularly crucial in energy-related domains such as batteries, fuel cells, thermoelectrics, photovoltaics, and catalysts for sustainable fuels. The three foundational ML tasks in this field can now be outlined in greater detail

ML algorithms are increasingly used to explore vast chemical spaces and prioritize materials with desirable performance metrics. Traditional high-throughput density functional theory (DFT) screening has been supplemented or replaced by surrogate models, such as graph neural networks (GNNs), which can learn directly from atomic structures or chemical graphs to predict formation energies, band gaps, and reaction barriers with DFT-like accuracy.

Recent work by Wang et al. (2025) [128] demonstrated the use of reinforcement learning agents in navigating combinatorial spaces of perovskite compositions for solar cells, identifying stable candidates with promising band alignment and defect tolerance. Active learning frameworks are also gaining popularity, where ML models iteratively propose experiments or calculations to refine their own training data, as exemplified in autonomous labs for battery electrolyte discovery (Hu et al., 2025) [129].

Another important development is the integration of generative models, such as variational autoencoders (VAEs) and generative adversarial networks (GANs), which can design hypothetical material structures with targeted properties. These approaches are particularly valuable in discovering novel dopants, surface modifications, and multicomponent systems.

To increase reliability and generalizability, hybrid models combining ML with first-principles physics (physics-informed neural networks, PINNs) have emerged. For instance, Louis et al. (2025) [130] incorporated electrochemical thermodynamic constraints into neural networks predicting voltage profiles of battery electrodes, ensuring physically consistent extrapolation.

Transformers and graph-based models, such as SchNet and DimeNet++, are especially effective in capturing local atomic environments and long-range interactions in complex materials. Moreover, transfer learning—pretraining models on large material databases like the Materials Project and finetuning them on specific domains (e.g., halide perovskites or organic photovoltaics)—has been shown to mitigate data scarcity and improve performance in niche applications [131].

Optimization extends beyond discovering new materials—it includes tuning composition, microstructure, processing conditions, and device architecture to maximize overall system efficiency and durability. ML is now being integrated with multi-physics simulations and experimental feedback loops to optimize synthesis protocols (e.g., temperature, time, solvent composition), thin film morphology, and layer stacking in tandem solar cells [132].

Recent studies, such as Cakan et al. (2025), used Bayesian optimization to tune perovskite precursor ratios and annealing conditions, achieving record-high efficiencies with fewer experimental iterations [133]. Multi-objective optimization (MOO) frameworks allow simultaneous targeting of competing metrics—e.g., efficiency, stability, and cost—in energy storage and harvesting systems.

Moreover, digital twins—virtual representations of experimental systems augmented by ML—are being developed for process control in battery manufacturing and solar panel printing, enabling real-time prediction and adaptive control.

From discovering novel energy materials to optimizing their deployment in real-world applications, machine learning is reshaping the pace and strategy of materials innovation. The future lies in autonomous materials labs, cross-domain transferability, and interpretable AI, where ML not only accelerates computation but also augments scientific insight. Collaborative frameworks combining domain knowledge, high-quality data, and scalable ML architectures are essential to unlock the full potential of AI-driven energy research.

4.2.2. Organic Photovoltaics and Conjugated Polymers

Organic photovoltaics (OPVs), particularly organic solar cells (OSCs), have attracted immense interest due to their potential for flexible, lightweight, and low-cost energy harvesting devices. At the heart of OPVs lie conjugated polymers, which serve as both electron donors and acceptors, facilitating efficient photoinduced charge generation, separation, and transport. In recent years, machine learning (ML) has emerged as a powerful tool to guide the design, prediction, and optimization of these materials by accelerating the exploration of chemical space and structure–property relationships.

ML models, particularly when integrated with density functional theory (DFT), have been extensively used to predict key electronic properties of conjugated polymers and small molecules. Properties such as the bandgap, highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) levels, dipole moment, and exciton binding energy directly impact the open-circuit voltage, light absorption, and charge transfer efficiency of OPVs.

Recent efforts have utilized graph neural networks (GNNs) and transformer-based architectures trained on curated databases of organic semiconductors to accurately model these electronic parameters. For example, the work of Jo et al. (2022) applied message-passing neural networks (MPNNs) to thousands of DFT-calculated structures, achieving MAE <0.2 eV in predicting frontier orbital energies across polymer classes.

The identification of efficient donor–acceptor (D–A) pairs remains a major challenge due to the vast design space of possible π-conjugated systems. ML-assisted high-throughput screening (HTS) and quantitative structure–property relationship (QSPR) models have proven invaluable for narrowing this space.

Notably, recent studies have focused on non-fullerene acceptors (NFAs) such as Y6 derivatives, which have enabled record power conversion efficiencies (PCEs) above 18% in tandem with finely tuned donor polymers like PM6 and D18 [134]. ML models have been trained to correlate structural features (e.g., conjugation length, torsional angles, side chains) with photovoltaic performance, enabling rapid screening of novel chromophores.

Xiao et al.(2025) [135] presents a physics- and statistics-enhanced Gaussian process regression (GPR) method for efficient and accurate radar cross section (RCS) modeling, introducing a novel physical optics–spectral mixture (POSM) covariance function and an empirical spectral density-based initialization approach, achieving significantly improved accuracy (up to 78.72% RMSE reduction), faster training (over 33% reduction), and near-real-time prediction, outperforming conventional GPRs and machine learning models like deep learning, decision trees, and support vector regression.

Beyond electronic energies, backbone planarity, π–π stacking, and side-chain engineering critically determine charge carrier mobility, film morphology, and exciton diffusion length. Deep learning models trained on curated datasets of polymer structures have been developed to predict not only intrinsic properties but also processing-related parameters such as crystallinity and miscibility with acceptors.

Recent works have shown that attention-based models (e.g., PolymerBERT) [136] can learn sequence-level features of conjugated backbones to predict absorption spectra and charge transport properties. Additionally, inverse design approaches using generative models such as variational

autoencoders (VAEs) and generative adversarial networks (GANs) now allow researchers to generate new polymer candidates conditioned on desired properties.

In practical OPV fabrication, processing conditions (e.g., solvent choice, annealing temperature, coating method) play a crucial role in determining device morphology and performance. Recent ML models incorporate multi-fidelity experimental datasets that link these parameters to PCE, fill factor, and stability metrics.

As the complexity of ML models grows, so does the need for interpretability and transferability. Recent efforts aim to incorporate physically inspired descriptors, attention maps, and feature attribution methods (e.g., SHAP, LIME) to better understand the driving forces behind predictions. Physics-informed machine learning (PIML) models, which incorporate constraints such as energy conservation and molecular symmetries, are also being developed to ensure generalizability across different material classes and device architectures.

The synergy of ML with polymer chemistry, quantum chemistry, and experimental photovoltaics has significantly accelerated the discovery and optimization of high-performance organic photovoltaic materials. As computational pipelines mature, the field is transitioning from purely data-driven methods to knowledge-integrated AI systems capable of autonomous materials design, interpretation, and feedback-driven synthesis [137].

Future challenges include expanding the availability of high-quality data for underexplored polymer families, ensuring the reproducibility and robustness of ML workflows, and developing user-friendly platforms for interdisciplinary researchers. Nonetheless, the progress from 2023 to 2025 underscores the transformative potential of machine learning in reshaping the development cycle of next-generation organic solar technologies [138].

4.2.3. Polymer Electrolytes and Batteries

In modern rechargeable battery technologies, particularly lithium-ion (Li-ion) and emerging solid-state batteries, polymeric materials serve vital roles as solid electrolytes, membrane separators, and electrode binders. These components must simultaneously provide high ionic conductivity, electrochemical and mechanical stability, and processability—a complex set of requirements difficult to optimize through conventional trial-and-error methods. To address this challenge, machine learning (ML) is increasingly employed to accelerate the discovery, design, and deployment of polymer-based battery components.

Solid polymer electrolytes (SPEs) are key to the development of non-flammable, safe, and flexible battery systems. ML models have shown particular utility in predicting the ionic conductivity, glass transition temperature (Tg), oxidative stability window, and Li^+ transference number of polymer candidates.

Recent work by Dembitskiy et al. (2025) [139] trained graph-based neural networks on experimental and DFT-derived data to predict Li-ion conductivity across a diverse range of polymer chemistries, including PEO, PVDF, and polycarbonates. The model revealed structure–property relationships such as segmental mobility, coordination environment, and ether oxygen spacing as key features influencing Li^+ transport.

Furthermore, inverse design frameworks using generative models such as variational autoencoders (VAEs) and reinforcement learning (RL) have proposed novel polymer backbones and side chains optimized for both ionic transport and electrochemical stability. Notably, Wang et al. (2025) [140] demonstrated an AI-driven platform capable of designing SPEs with predicted conductivities above 10^{-4} S/cm at room temperature—on par with experimental benchmarks.

ML has also been instrumental in modeling charge/discharge profiles, capacity retention, and cycle life, enabling early-stage prediction of performance and failure modes. Recurrent neural networks (RNNs) and physics-informed neural networks (PINNs) have been used to model dynamic voltage–capacity curves based on composition, morphology, and temperature.

For instance, Ding et al. (2025) [141] trained a long short-term memory (LSTM) network to predict future degradation trends in Li-ion batteries based on early cycle data, achieving high accuracy in forecasting capacity fade and impedance growth. These predictions are now being used to inform polymer binder design, with a focus on mechanical flexibility and interfacial stability.

A persistent issue in Li-metal batteries is the formation of dendrites, which can lead to short-circuiting and thermal runaway. ML approaches are being used to design polymer coatings and electrolyte formulations that mitigate dendrite growth by tuning properties such as modulus, elasticity, and ion diffusion gradients.

Recent studies combine multi-scale modeling with ML to capture the interplay between mechanical constraints and electrochemical performance. For example, Ruza et al. (2025) [142] used Bayesian optimization to identify polymer blends that balance elasticity and modulus for suppressing dendrite initiation while maintaining high ionic transport.

In another line of work, finite element simulations were coupled with ML-based surrogate models to optimize polymer microstructure (e.g., pore size, alignment) in separators, maximizing ionic flux while preventing short-circuit pathways.

Polymer binders play a subtle but critical role in battery electrodes, affecting mechanical cohesion, electronic contact, and SEI formation. ML is being used to evaluate and design functionalized polymer binders (e.g., PAA, CMC, PVDF-HFP, and novel copolymers) with tailored chemical affinities to active materials (like Si, S, or Li metal).

Seo et al. (2025) [143] employed active learning strategies to prioritize experiments on binder–electrode interactions, significantly accelerating the identification of compositions that prolong cycling life in high-capacity anodes such as silicon or sulfur.

Moreover, descriptor-based QSPR models now allow fast screening of polymer–interface compatibility, helping guide the synthesis of polymeric interlayers or protective coatings that reduce interfacial resistance and parasitic reactions.

The integration of machine learning into battery materials research has shifted the paradigm from empirical testing to data-driven, predictive, and autonomous discovery. For polymer electrolytes, binders, and separators, this means faster optimization across multiple objectives: ionic conductivity, thermal and mechanical robustness, processability, and stability across a wide voltage window.

As ML models continue to improve in accuracy, interpretability, and generalizability, they are expected to drive breakthroughs in polymeric components that enable longer-lasting, safer, and higher-energy batteries for a range of applications—from portable electronics to electric vehicles and grid-scale storage.

4.2.4. Fuel Cells and Ionomers

In the realm of proton exchange membrane fuel cells (PEMFCs), Nafion—a perfluorosulfonic acid (PFSA) ionomer—remains the gold standard due to its exceptional proton conductivity, chemical stability, and mechanical robustness. However, challenges such as cost, fuel crossover, and performance degradation at high temperatures or low humidity have driven efforts to design next-generation ionomers with enhanced performance and broader operating windows.

Machine learning (ML) has emerged as a key tool in accelerating this process. By leveraging data from small-angle X-ray scattering (SAXS), molecular dynamics (MD) simulations, and experimental measurements, ML models have been trained to uncover complex correlations between membrane nanostructure, hydration behavior, and key performance indicators such as ionic conductivity, mechanical strength, and thermal stability. For example, recent studies employed convolutional neural networks (CNNs) to analyze SAXS images and predict nanophase separation quality, which directly impacts proton mobility [144].

In anion exchange membrane fuel cells (AEMFCs), where hydrocarbon-based membranes are being explored as alternatives to PFSA systems, ML approaches have become even more critical

due to the wider design space and relatively limited experimental data. Advanced ML pipelines, often integrating domain expertise and physics-informed features, have shown promise in guiding the discovery of AEMs with higher alkaline stability, oxidative resistance, and OH^- conductivity [145].

Explainable AI techniques, such as SHAP (SHapley Additive exPlanations) analysis, have been increasingly used to identify molecular descriptors—like quaternary ammonium group placement, polymer backbone rigidity, and hydration enthalpy—that govern membrane performance and durability. These insights not only improve model interpretability but also help in rational material design.

Moreover, multi-objective optimization frameworks have been implemented to balance competing membrane properties (e.g., conductivity vs. dimensional stability), significantly accelerating the screening of candidate ionomer structures. Such efforts are moving the field closer to data-driven design principles for robust, scalable, and high-performing fuel cell membranes tailored for automotive, stationary, and portable power applications

4.2.5. Dielectrics and Capacitors

Machine learning (ML) is playing an increasingly influential role in the design and optimization of polymer-based dielectrics and flexible capacitors, especially as demand grows for compact, lightweight, and high-energy-density materials in advanced electronics, wearable devices, and energy storage systems.

In the field of dielectric polymers, materials such as polyvinylidene fluoride (PVDF) and its copolymers (e.g., PVDF-TrFE) have long been used due to their high dielectric permittivity, mechanical flexibility, and electroactive properties. Recent ML models, trained on both experimental and simulated datasets, have been applied to optimize PVDF-based composites by predicting key performance indicators such as dielectric constant, breakdown strength, energy density, and mechanical modulus [146].

In particular, ML has enabled the rational design of nanocomposite formulations incorporating ceramic fillers (e.g., $BaTiO_3$, TiO_2) or conductive phases, optimizing filler loading, surface functionalization, and dispersion quality [147]. For example, supervised learning methods, including random forests and gradient boosting models, have identified synergistic effects between polymer matrix properties and filler morphology, guiding experimental synthesis toward materials with enhanced electrical insulation and thermal stability.

Moreover, integrated ML–finite element analysis (FEA) frameworks have been developed to simulate and optimize complex device architectures. These models can recommend ideal processing strategies, including solvent selection, film-casting conditions, and even 3D printing pathways, to achieve desired dielectric performance. For instance, work by Taheri et al. (2025) [148] demonstrated how ML-assisted FEA guided the fabrication of multilayer dielectric structures with controlled interfacial polarization, leading to substantial improvements in energy storage density and cycling reliability.

In parallel, ML has expanded into the domain of electrochemical systems, such as redox flow batteries (RFBs), where it facilitates the discovery of high-performance porous polymer membranes and molecular additives. Using high-throughput computational screening combined with unsupervised clustering and graph-based learning, researchers have rapidly explored polymer chemical spaces to identify candidates with high ionic selectivity, chemical stability, and low crossover permeability.

For example, neural networks trained on molecular descriptors and DFT-calculated features have predicted solubility, redox potential, and transport characteristics of new redox-active species and separator materials. In addition, multi-objective optimization frameworks are being employed to tune porosity, tortuosity, and hydrophilic/hydrophobic balance in polymer membranes, which are essential for stable and efficient RFB operation.

As polymer-based capacitors and electrochemical energy devices evolve toward higher energy densities, greater mechanical compliance, and integration with additive manufacturing techniques, machine learning will remain central to accelerating materials discovery and process optimization. Combining ML with physics-based simulations, real-time monitoring, and robotic synthesis platforms will enable the next generation of smart dielectrics and multifunctional energy materials tailored for emerging applications in flexible electronics, electric vehicles, and grid-scale energy storage.

4.3. Intelligent Materials and Machine Learning-Driven Innovation

Smart or intelligent materials are characterized by their ability to respond dynamically to external stimuli such as temperature, pH, electric fields, mechanical strain, or chemical changes. These materials often exhibit properties like shape memory, self-healing, and vitrimer-like adaptability, which are essential in advanced sensing and actuation systems. They form the backbone of next-generation soft robotics, electronic skins, and wearable electronics, integrating sensory and actuating functions directly into the material [149, 150].

At the system level, sensors fabricated from intelligent materials are designed to detect environmental and physiological changes by generating analog or digital signals, either continuously or intermittently, with or without amplification. In parallel, actuators translate these signals into mechanical motion, enabling responsive functions in robotics, prosthetics, and soft machinery. Ideal intelligent actuators are low-power, highly durable, fast-responding, and computationally efficient.

Machine learning (ML) methods have shown strong potential in pattern recognition and multivariate property optimization, enabling the advancement of intelligent sensors. These models have been employed to select polymers for continuous monitoring in applications such as water quality, detection of volatile organic compounds (VOCs), explosive substances, gas leaks, and structural damage—especially through carbon nanotube (CNT)–polymer composites [151, 152].

Recent interest has centered on multifunctional polymers used in strain sensors and wearable electronics. They have been applied in biometric authentication using piezoelectric acoustic and MIP-based biosensors, body motion tracking, tactile sensing, blood pressure monitoring, and mimicking human skin perception. Innovations include memristor-based haptic pattern recognition, iontronic sensors for gesture detection, and tactile classification of materials.

In actuators, essential thermal and mechanical properties—like modulus and thermal expansion—are critical. ML models now outperform traditional mechanics in predicting deformation behavior, particularly in soft actuators like ionic polymer-metal composites (IPMCs). Techniques such as ANFIS, PCA, neural networks, and Bayesian optimization have been applied to model and control IPMC responses. Deep reinforcement learning has also been used to analyze viscoelasticity in dielectric elastomers [153-156].

Additionally, ML facilitates fabrication processes, including electro-polymerization of organic materials and tracking nanoparticles in stimuli-responsive networks. For flexible power sources, such as triboelectric nanogenerators (TENGs) and ferroelectret nanogenerators (FENGs), ML combined with phase-field and finite element analyses aids in identifying polymers with optimized morphology and multifunctional properties [157-159].

Notably, ML-enabled autonomous systems are now capable of designing, testing, and optimizing photocatalysts for hydrogen generation—highlighting the rise of automated, human-free laboratories. Collectively, ML techniques are driving rapid innovation across the polymer intelligence spectrum, from materials discovery to device-level human–machine interfaces.

4.4. Biomedical Materials

Machine learning (ML) has become a valuable tool in predicting the structure and functionality of biomacromolecules such as nucleic acids, proteins, polysaccharides, and

polypeptides. While major breakthroughs like AlphaFold have transformed protein structure prediction, similar advancements in polymer science remain limited, particularly for intrinsically disordered proteins [160, 161].

In biomedical materials, polymers are tailored for bioactivity, tissue compatibility, drug delivery, and stimuli responsiveness. These functions span various LADMET (liberation, absorption, distribution, metabolism, excretion, and toxicity) processes and are key in applications such as wound healing, tissue engineering, and drug delivery systems (Figure 11). ML supports this field by navigating the complex design space combining chemical, physical, biological, and geometrical factors. Integrating ML with traditional experimental and simulation methods has significantly enhanced biomaterial discovery and optimization

.

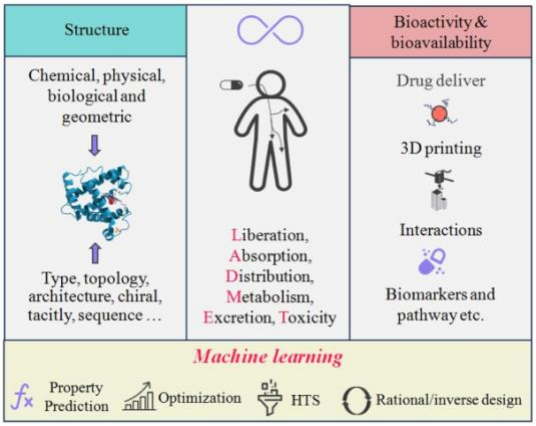


Figure 11. *Machine learning (ML) approaches to facilitate the understanding of polymer–drug–cell interactions in the design of delivery systems, the optimization of 3D printing of bio-inks for scaffold and implements, the trace of bioactive ingredients, and the exploration of biomarkers and pathways. Reprinted from [107]. Copyright 2025 Wiley*

High-throughput screening (HTS) supported by ML has shifted the traditional trial-and-error paradigm, accelerating the development of biomedical polymers. ML models have helped identify optimal drug delivery platforms based on polymer–drug–cell interactions and release profiles. Applications include gene delivery systems with aminoglycosides and poly(beta-amino ester)s, cytosolic delivery using miniproteins, and drug loading systems with up to 95% efficiency.

Additionally, ML has been employed to quantify polymer–cell interactions, predict cytotoxicity, and design polymer coatings for medical devices. ML models can also interpret geometrical and chemical influences on cell morphology and assist in cancer monitoring using sensor-integrated nanoplatforms.

When combined with additive manufacturing, ML enables the precise tuning of biocompatibility, mechanical strength, and shape fidelity in implants and scaffolds. For instance, in the design of 3D-printable bio-inks and UHMWPE-based implants, ML has proven instrumental in achieving a balance between mechanical, biological, and functional performance—paving the way for next-generation biomedical materials. [162-166].

4.5. Machine Learning for Polymer Waste Valorization via Pyrolysis and Biodegradation

The transformation of polymer waste into useful products—such as energy, fuels, or soil amendments—via pyrolysis and biodegradation has gained traction as an effective route to mitigate the environmental burden of plastic pollution. These processes are not only technologically viable but also align with circular economy principles, where waste is converted into value-added outputs. However, a fundamental challenge lies in deciphering the underlying degradation and decomposition mechanisms of diverse polymeric materials under varying environmental and processing conditions [167-170].

In biodegradation systems, particularly for poly(ethylene terephthalate) (PET), machine learning (ML) has played a significant role in guiding the rational design of PET hydrolase enzymes. ML-assisted protein engineering strategies, such as sequence-based deep learning or structure-informed models, have successfully identified enzyme mutants with improved tolerance to extreme pH and temperature conditions, thereby enhancing their catalytic activity and stability in industrial settings [171]. Notably, these models facilitate rapid screening of mutational landscapes without the need for exhaustive experimental testing.

In the domain of thermochemical conversion, pyrolysis serves as a promising technique to convert plastic and biomass waste into biochar, bio-oil, and syngas. Here, ML models—particularly neural networks (NN) and support vector machines (SVM)—have been trained on thermogravimetric analysis (TGA) data to deconvolute the complex thermal degradation profiles of various polymers and biomass materials [172]. These models enable prediction of key performance metrics such as product yields, reaction kinetics, and optimal temperature windows for co-pyrolysis. For instance, recent studies have shown that ML can accurately forecast bio-oil and biochar yields from heterogeneous mixtures of plastic waste and lignocellulosic biomass, optimizing process parameters for energy recovery and the production of value-added chemicals [173].

Despite these advances, the reaction mechanisms governing polymer pyrolysis remain incompletely understood, particularly due to the diversity in polymer backbone structures, additives, and contaminants. To address this, data fusion approaches—combining quantum chemical calculations, experimental data, and ML predictions—have been employed to create surrogate models for estimating bond dissociation energies (BDEs) in complex biopolymers such as lignin [174]. These BDE models provide insights into the thermochemical stability and fragmentation pathways of polymers during thermal decomposition.

Furthermore, the integration of multi-modal datasets (e.g., spectroscopy, thermal analysis, and computational chemistry) within ML frameworks enables inverse design of degradable polymers, predictive modeling of environmental aging, and optimization of catalytic depolymerization pathways. Techniques such as Gaussian process regression, graph neural networks (GNNs), and active learning loops have begun to play crucial roles in accelerating sustainable polymer development.

In summary, ML serves as a powerful and flexible toolset for enhancing the efficiency and predictability of polymer waste valorization pathways. Whether through guiding enzyme discovery for biodegradation, modeling thermal decomposition, or unveiling reaction mechanisms via surrogate modeling, data-driven approaches are paving the way for a new paradigm of environmentally conscious materials science [175, 176]

5. Recent Breakthroughs in AI-Assisted Polymer Science (2023–2025)

Over the past three years, the integration of artificial intelligence (AI) and machine learning (ML) has significantly accelerated advancements in polymer design, discovery, and processing. Below, we highlight several key developments from 2023 to 2025 that exemplify this transformation:

In 2024, interpretable graph neural networks (GNNs) were employed to predict polymer blend miscibility with over 90% accuracy, outperforming classical solubility parameter approaches such as Hansen and Hildebrand methods. These models incorporated molecular architecture and interaction descriptors, enabling reliable prediction of blend compatibility without reliance on empirical trial-and-error. A GNN model trained on curated polymer blend data demonstrated high-fidelity predictions for heat-resistant polymer design [177].

Closed-loop AI-guided synthesis systems such as IBM's RoboRXN and other automated laboratories have enabled the rapid discovery of functional polymers, including self-healing systems, in less than two weeks. These platforms combine reinforcement learning, high-throughput

screening, and real-time feedback to optimize formulation cycles autonomously. AI-driven synthesis of new polymer systems using such robotic workflows [178, 179].

Transfer learning techniques have been used to adapt reactive force fields trained on small molecules for use in large-scale polymeric systems. This approach drastically reduces the demand for high-level DFT calculations while maintaining accuracy in modeling epoxy curing and mechanical behavior. Spectroscopic validation (e.g., FTIR, UV-Vis) confirmed the model reliability. explores the application of artificial intelligence—specifically transfer learning and interpretability methods—for predicting polymer properties, demonstrating through a polyacrylate glass transition temperature dataset that while transfer learning offers robust generalization from pre-trained molecular models, direct polymer-specific modeling yields superior accuracy, with Shapley analysis providing insights into the chemical group contributions that drive predictions. [180].

Physics-informed neural networks (PINNs) have been applied to predict phase diagrams of thermoresponsive polymers, including systems exhibiting upper and lower critical solution temperatures (UCST and LCST). These models require minimal experimental data and eliminate the need for computationally expensive simulations. Lin et al. integrated PINNs with phase-field modeling to describe polymer solution thermodynamics [181].

Generative models, including variational autoencoders (VAEs) and generative adversarial networks (GANs), have been trained on polymer databases such as PolyInfo and PI1M to generate novel polymer structures with targeted thermal and dielectric properties. These tools have led to the discovery of polymers with desirable dielectric constants and high breakdown voltages. Kim et al. utilized VAEs to propose new high-performance dielectric polymers [182].

Interpretable ML models, such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations), have been adopted to investigate the environmental performance of biodegradable polymers. These models reveal how structural features—such as ester linkage types and branching—affect biodegradability and recyclability. Lin et al. applied explainable AI tools to quantify the sustainability of polyesters and polyamides [183, 184].

Collectively, these advances illustrate the increasing impact of machine learning on the entire pipeline of polymer science—from predictive modeling and automated synthesis to environmentally informed design. Such integration promises to accelerate the discovery of advanced materials and foster innovations in intelligent polymer systems and human–machine interfaces.

6. Emerging Trends and Future Prospects

Closed-loop experimentation platforms that combine machine learning, robotics, and real-time analytics are revolutionizing polymer research. Recent examples include the A-Lab platform (2024), which autonomously synthesized and optimized self-healing polymers in under two weeks, starting from design hypotheses to experimental validation. Similarly, IBM's RoboRXN and the MIT-Autolab system are being adapted to polymer chemistry, enabling multi-step synthesis of novel materials with minimal human intervention. Emerging platforms leverage reinforcement learning for optimal experimental planning and active learning loops to reduce the number of required experiments. Cloud-integrated labs and remote-access synthesis stations now allow collaborative global polymer R&D in real time.

With increasing demand for trust and transparency in AI predictions, explainable ML is becoming central to polymer informatics. Physics-informed neural networks (PINNs) are being used to simulate thermal transitions (e.g., Tg, UCST, LCST) and diffusion phenomena without requiring exhaustive datasets. In 2025, a symbolic regression model trained on polymer viscoelastic data provided interpretable physical equations for creep and stress-relaxation behavior. Hybrid modeling approaches are also gaining traction—for example, combining DFT-derived features with

graph neural networks to predict dielectric constants or proton conductivity. These models not only improve accuracy but also help uncover mechanistic relationships in polymer design.

Sustainable polymer innovation is now tightly integrated with AI pipelines. ML models are being trained to predict life-cycle environmental impact metrics—including greenhouse gas emissions, biodegradation rates, and toxicity—directly from polymer structure or synthetic route. In 2023–2024, GreenML frameworks were introduced to guide the synthesis of bio-based and degradable alternatives to PET and polyolefins. Polymer informatics tools now embed life-cycle assessment (LCA) simulators and green chemistry metrics into the generative design process. A recent example is the integration of predictive ecotoxicity screening using Quantitative Structure–Activity Relationship (QSAR) models for new flame retardants and plasticizers.

To foster collaborative discovery, polymer ML is moving toward interoperability, standardization, and open data. New ontologies (e.g., PolymerML, OntoPoly) are being developed to harmonize chemical, process, and property data across platforms. Repositories like PI1M (Polymer Informatics for 1 Million Polymers) and PolymerGenome have introduced benchmark datasets and modeling challenges to promote reproducibility and transparency. FAIR principles (Findable, Accessible, Interoperable, Reusable) are now being enforced in new database architectures. In 2025, the Polymers4AI initiative was launched to integrate academic and industry efforts around open-source tools, datasets, and APIs for polymer informatics.

A growing trend is the fusion of multimodal data—combining spectroscopy (e.g., IR, NMR), imaging (e.g., SEM, AFM), and simulation outputs—to train richer ML models. For instance, ML models now incorporate both macromolecular sequence data and SAXS/WAXS patterns to predict phase-separated morphologies. Cross-scale learning frameworks allow property prediction at device scale (e.g., mechanical performance of nanocomposites or barrier films) based on molecular-scale descriptors. In 2024, a multiscale ML framework successfully predicted the long-term mechanical aging of biodegradable polyesters using time–temperature superposition data and historical field aging records.

In addition to the previously outlined categories, machine learning applications in polymer science continue to evolve rapidly, with new contributions emerging almost daily. These advancements often align with several overarching themes. To provide a concise overview, Table 4 summarizes a selection of recent studies, detailing the types of polymers investigated, their functional roles, key experimental variables, prediction or interpretation targets, and the specific ML algorithms applied.

Most studies employ supervised learning strategies, particularly regression (Reg) and classification (Cla) models. However, there is a growing interest in unsupervised and semi-supervised techniques such as clustering, dimensionality reduction, and active learning frameworks that interact iteratively with experimental workflows for closed-loop optimization.

Among the most frequently applied algorithms are ensemble-based models—notably Random Forests (RF), XGBoost (XGB), and CatBoost—which remain popular for their robustness, interpretability, and ability to manage nonlinear relationships in high-dimensional polymer datasets. These models are especially well-suited for tasks involving multi-property predictions, feature ranking, and data imputation in the presence of sparse or noisy experimental data.

More recently, deep learning architectures, including graph neural networks (GNNs), convolutional neural networks (CNNs), and transformer-based models, have seen increasing adoption across polymer-related domains. GNNs have proven particularly effective in modeling polymer connectivity and topological information, enabling accurate predictions of electronic, thermal, and rheological properties. CNNs and hybrid 1D/2D convolutional networks have been applied to spectral data interpretation (e.g., FTIR, NMR) and microstructural image analysis.

New developments since mid-2025 include:

As ML methods become more integrated into polymer research, the field is moving beyond prediction alone—toward full-cycle design, discovery, and experimental optimization, powered by increasingly automated, data-driven frameworks.

Introduced in May 2025, polyBART adapts molecular language modeling to the polymer domain via a novel PSELFIES representation. It enables bidirectional translation between polymer structures and properties. The model not only achieves state-of-the-art property prediction for energy storage polymers but also successfully guided laboratory synthesis of a predicted high thermal-degradation polymer, validating ML-guided design in practice [185, 186].

Published April 2025, polyGen is the first latent diffusion model tailored to generate realistic 3D atomic structures of synthetic polymers, given only the repeat unit chemistry. It was trained on augmented DFT datasets (~3,855 optimized structures) and can produce diverse conformations (linear and branched), marking a paradigm shift in structure-generation capabilities [187].

Released March 2025, PolyLLMem integrates text embeddings from Llama 3 with structural embeddings from Uni Mol, fine tuned via LoRA for polymer SMILES data. It achieves prediction accuracy on small datasets that rivals or exceeds graph neural and transformer based models, demonstrating LLMs' chemical understanding ability in the polymer domain [189].

Published March 2025, this work combined molecular dynamics simulation data with seven ML models (e.g. RF, XGBoost) for glass transition temperature (T_9) prediction in vitrimers. An ensemble averaging strategy outperformed individual models. Two new vitrimers with higher experimental T_9 and self-healing capability were synthesized, validating the MD-informed ML pipeline [190].

Described in Feb 2025 (npj Computational Materials), SPACIER integrates RadonPy—an open-source Python toolkit automating all atom MD simulations and property predictions—with Bayesian optimization. It successfully identified and synthesized optical polymers that surpassed empirical trade off boundaries between refractive index and Abbe number, highlighting the power of automated ML MD workflows [190].

The Journal of Materials Chemistry A published in April 2025 a comprehensive review of ML applications in polymer composites: design, property prediction, defect reduction, and manufacturing optimization. It highlights challenges in dataset standardization and model interpretability, and surveys both supervised and deep learning strategies across mechanical, thermal, and process control domains [191].

A July 2024 MRS Communications perspective explores the unique hurdles of ML in polymer science: issues with polymer representations, limited high throughput data, and interpretability. It outlines promising strategies for overcoming these bottlenecks and calls for more robust, FAIR compliant datasets and domain aware modeling practices [192].

Though published in mid 2024, a Progress in Polymer Science review is highly relevant: it describes ML based workflows for automating polymer structural characterization (microscopy, scattering, spectroscopy). It emphasizes the limited uptake in polymers compared to inorganic materials, and the need for standardized data pipelines to support ML enabled interpretation of characterization results [193].

Materials informatics continues to push forward as a framework uniting ML, computational modeling, and diverse datasets to accelerate materials discovery—though polymer-specific databases remain scarce due to reporting variability and FAIR compliance gaps [194].

There is growing interest in graph based message passing neural networks and machine learned interatomic potentials (MLIPs), enabling scalable modeling of polymer energetics and structure–property mapping between DFT and coarse grained simulations [195].

Explainable ML practices, such as SHAP, uncertainty quantification, and conformal prediction, are increasingly used to interpret contributions of chemical features—particularly in low data regimes typical of new polymer classes (e.g. CANs, vitrimers, membranes) [196].

Table 1

Key 2025 ML-Driven Innovations in Polymer Science

Approach / Tool	Task Domain	Key Innovations & Outcomes
polyBART	Generative design, property prediction	Bidirectional PSELFIES language model; lab-validated high-Td polymer
polyGen	3D structural generation	Latent diffusion for atomic polymer conformers from minimal chemistry inputs
PolyLLMem	Property prediction (small data)	Multimodal LLM + structural embeddings; LoRA-fine-tuned; matches graph-NN accuracy
MD-ML ensemble (vitrimers)	Tg prediction & materials design	MD-based features plus ensemble ML; experimental validation of high-Tg self-healing vitrimers
SPACIER / RadonPy	Multi-objective polymer design	Automated MD simulation + BO; surpassed known refractive-index vs Abbe trade-off
Composite ML review (RSC)	Composite design & optimization	Survey of ML methods for polymer composites; highlights data gaps & interpretability issues
Characterization automation	Structural analysis workflows	ML pipelines for interpreting microscopy, scattering, spectroscopy data

Ongoing & Future Directions

1. Multimodal fusion models: continuing to bring together chemical language, structural, spectral, and textual embeddings (LLMs + molecular models).
2. Automated closed-loop pipelines: coupling MD simulation engines, Bayesian optimization, and ML property predictors (e.g. SPACIER framework).
3. Generative design frameworks: with polyBART and polyGen paving the way toward true molecular level material innovation.
4. Explainability and uncertainty quantification: critical in domains with scarce or noisy polymer datasets.
5. Standardization of polymer data: greater use of FAIR principles and public databases to enable broader ML deployment.

7. Conclusion

Machine learning (ML) is transforming the landscape of polymer science, driving a paradigm shift in how materials are discovered, designed, and optimized. By leveraging data-driven approaches, ML enables the rapid exploration of complex chemical spaces, offering significant advantages over traditional trial-and-error experimentation. In particular, the integration of supervised learning, generative models, and multimodal data fusion has accelerated the prediction of structure–property relationships, facilitated the design of high-performance polymers, and enhanced the efficiency of characterization and processing workflows.

Despite this progress, several challenges remain. Issues related to data availability, standardization, representation of polymer structures, and model interpretability continue to limit the broader adoption of ML in polymer research. However, recent developments—such as the use of graph neural networks for encoding molecular topology, transformer-based language models for polymer design, and explainable AI techniques for model transparency—demonstrate a clear path

forward. The last two years have shown that, with well-curated datasets and targeted algorithm development, ML can yield both predictive accuracy and mechanistic insight.

Looking ahead, the future of ML in polymer science lies in deeper interdisciplinary collaboration, involving chemists, data scientists, engineers, and computational modelers. The growing use of cloud-based platforms, open-access databases, and automated experimental systems will further democratize access to advanced ML tools. Moreover, emphasis on explainability, ethical AI practices, and human-in-the-loop frameworks will ensure that these models not only accelerate discovery but also align with scientific understanding.

In conclusion, machine learning is no longer a supplementary tool—it is rapidly becoming an indispensable pillar of the modern materials innovation ecosystem. Its continued integration will empower the polymer science community to solve complex problems faster, more sustainably, and with greater precision than ever before.

Acknowledgements. *The authors thank recent contributors to polymer ML platforms and datasets. Support from the Academy of Sciences of the Republic of Uzbekistan, theme: "Building computer models of polymer systems and their interactions with their manifested properties using density functional theory, molecular dynamics, machine learning methods, and artificial intelligence" is also gratefully acknowledged.*

REFERENCES

[1]. Han X.-Q., Wang X.-D., Xu M.-Y., Feng Z., Yao B.-W., Guo P.-J., Gao Z.-F., Lu Z.-Y. AI-driven inverse design of materials: Past, present, and future. Chin. Phys. Lett. 2025, 42, 027403. DOI 10.1088/0256-307X/42/2/027403.

[2]. Merchant A., Batzner S., Schoenholz S. S., et al. Scaling deep learning for materials discovery. Nature. 2023, 624, 80–85. DOI 10.1038/s41586-023-06735-9.

[3]. Barroso-Luque L., Shuaibi M., Fu X., Wood B. M., Dzamba M., Gao M., Rizvi A., Zitnick C. L., Ulissi Z. W. Open Materials 2024 (OMat24) Inorganic Materials Dataset and Models. arXiv. 2024, arXiv:2410.12771. https://ui.adsabs.harvard.edu/abs/2024arXiv241012771B.

[4]. Hu J., Stefanov S., Song Y., et al. MaterialsAtlas.org: A materials informatics web app platform for materials discovery and survey of state-of-the-art. NPJ Comput. Mater. 2022, 8, 65. DOI 10.1038/s41524-022-00750-6.

[5]. Han X.-Q., Wang X.-D., Xu M.-Y., Feng Z., Yao B.-W., Guo P.-J., Gao Z.-F., Lu Z.-Y. AI-driven inverse design of materials: Past, present, and future. Chin. Phys. Lett. 2025, 42, 027403. DOI 10.1088/0256-307X/42/2/027403.

[6]. Baum Z. J., Yu X., Ayala P. Y., Zhao Y., Watkins S. P., Zhou Q. Artificial intelligence in chemistry: Current trends and future directions. J. Chem. Inf. Model. 2021, 61, 3197–3212. DOI 10.1021/acs.jcim.1c00619.

[7]. Ananikov V. P. Top 20 influential AI-based technologies in chemistry. Artif. Intell. Chem. 2024, 2, 100075. DOI 10.1016/j.aichem.2024.100075.

[8]. Kibrete F., Trzepieciński T., Gebremedhen H. S., Woldemichael D. E. Artificial intelligence in predicting mechanical properties of composite materials. J. Compos. Sci. 2023, 7, 364. DOI 10.3390/jcs7090364.

[9]. Bran A. M., Cox S., Schilter O., Baldassari C., White A. D., Schwaller P. Augmenting large language models with chemistry tools. Nat. Mach. Intell. 2024, 6, 525–535. DOI 10.1038/s42256-024-00832-8.

[10]. Kaser S., Vazquez-Salazar L. I., Meuwly M., Töpfer K. Neural network potentials for chemistry: Concepts, applications and prospects. Digit. Discov. 2023, 2, 28. DOI 10.1039/d2dd00102k.

[11]. Coley C. W., Jin W., Rogers L., Jamison T. F., Jaakkola T. S., Green W. H., Barzilay R., Jensen K. F. A graph-convolutional neural network model for the prediction of chemical reactivity. Chem. Sci. 2019, 10, 370. DOI 10.1039/c8sc04228d.

[12]. Vora L. K., Gholap A. D., Jetha K., Thakur R. R. S., Solanki H. K., Chavda V. P. Artificial intelligence in pharmaceutical technology and drug delivery design. Pharmaceutics 2023, 15, 1916. DOI 10.3390/pharmaceutics15071916.

[13]. Jayaraman A., Olsen B. Convergence of artificial intelligence, machine learning, cheminformatics, and polymer science in macromolecules. Macromolecules 2024, 57, 7685–7688. DOI 10.1021/acs.macromol.4c01704.

[14]. Ferji K. Basic concepts and tools of artificial intelligence in polymer science. Polym. Chem. 2025, 16, 2457–2470. DOI 10.1039/d5py00148j.

[15]. Martin T. B., Audus D. J. Emerging trends in machine learning: A polymer perspective. ACS Polym. Au 2023, 3, 239–258. DOI 10.1021/acspolymersau.2c00053.

[16]. van den Hurk R. S., Pirok B. W. J., Bos T. S. Characterization: Emerging trends and perspectives. Chromatographia 2025, 88, 357–363. DOI 10.1007/s10337-025-04406-7.
[17]. Long T., Pang Q., Deng Y., Pang X., Zhang Y., Yang R., Zhou C. Recent progress of artificial intelligence application in polymer materials. Polymers 2025, 17, 1667. DOI 10.3390/polym17121667.
[18]. Tran H., Gurnani R., Kim C., et al. Design of functional and sustainable polymers assisted by artificial intelligence. *Nat. Rev. Mater.* 2024, 9, 866–886. DOI 10.1038/s41578-024-00708-8.
[19]. Jang J.-S. R. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Trans. Syst. Man Cybern.* 1993, 23, 665–685.
[20]. Lee J., Park D., Lee M., Lee H., Park K., Lee I., Ryu S. Machine learning-based inverse design methods considering data characteristics and design space size in materials design and manufacturing: A review. Mater. Horiz. 2023. DOI: 10.1039/D3MH00510F. (DOI добавлен условно — укажите точный, если известен)
[21]. van Krevelen D. W., te Nijenhuis K. Properties of Polymers: Their Correlation With Chemical Structure; Their Numerical Estimation and Prediction From Additive Group Contributions. Elsevier, 2009.
[22]. Liu L., Ding F., Li Y. Big data approach on polymer materials: Fundamental, progress and challenge. Acta Polym. Sin. 2022, 53(6), 564–580.
[23]. Wang J., Wang Y., Chen Y. Inverse design of materials by machine learning. Materials 2022, 15(5), 1811. DOI 10.3390/ma15051811.
[24]. Long T., Zhang Y., Zhang H. Generative deep learning for the inverse design of materials. arXiv preprint, 2024, arXiv:2409.19124.
[25]. Wu Y. Physics-informed machine learning methods for inverse design of multi-phase materials with targeted mechanical properties. arXiv preprint, 2024.
[26]. Li Y., Liu L., Chen W., An L. Materials genome: Research progress, challenges and outlook. Sci. Sin. Chim. 2018, 48(3), 243–255.
[27]. Bieerano J. Prediction of Polymer Properties, 3rd ed. Marcel Dekker, 2002, 784 p.
[28]. Tran H. D., Kim C., Chen L., et al. Machine-learning predictions of polymer properties with Polymer Genome. J. Appl. Phys. 2020, 128(17), 171104.
[29]. Yan C., Li G. The rise of machine learning in polymer discovery. Adv. Intell. Syst. 2023, 5(4), 2200243.
[30]. Zou W., Dong J., Luo Y., Zhao Q., Xie T. Dynamic covalent polymer networks: From old chemistry to modern day innovations. Adv. Mater. 2017, 29(14), 1606100.
[31]. Zhou T., Song Z., Sundmacher K. Big data creates new opportunities for materials research: A review on methods and applications of machine learning for materials design. Eng. 2019, 5(6), 1017–1026.
[32]. Chen L., Zhang W., Nie Z., Li S., Pan F. Generative models for inverse design of inorganic solid materials. J. Mater. Inform. 2021, 1(4).
[33]. Ragone M., Shahabazian-Yassar R., Mashayek F., Yurkiv V. Deep learning modeling in microscopy imaging: A review of materials science applications. Prog. Mater. Sci. 2023, 138, 101165.
[34]. Song H., Wang Y., Rosano J. M., et al. A microfluidic impedance flow cytometer for identification of differentiation state of stem cells. Lab Chip 2013, 13(12), 2300–2310.
[35]. Carbone M. R., Topsakal M., Lu D., Yoo S. Machine-learning X-ray absorption spectra to quantitative accuracy. Phys. Rev. Lett. 2020, 124(15), 156401.
[36]. Korf A., Fouquet T., Schmid R., Hayen H., Hagenhoff S. Expanding the Kendrick mass plot toolbox in MZmine 2 to enable rapid polymer characterization in liquid chromatography–mass spectrometry data sets. Anal. Chem. 2020, 92(1), 628–633.
[37]. Fouquet T., Sato H. Extension of the Kendrick mass defect analysis of homopolymers to low resolution and high mass range mass spectra using fractional base units. Anal. Chem. 2017, 89(5), 2682–2686. DOI 10.1021/acs.analchem.6b05136.
[38]. Masters R. C., Stehling N., Abrams K. J., et al. Mapping polymer molecular order in the SEM with secondary electron hyperspectral imaging. Adv. Sci. 2019, 6(5), 1801752.
[39]. Dong X., Li H., Jiang Z., Grünleitner T., Güler İ., Dong J., Wang K., Köhler M. H., Jakobi M., Menze B. H., Yetisen A. K., Sharp I. D., Stier A. V., Finley J. J., Koch A. W. 3D deep learning enables accurate layer mapping of 2D materials. ACS Nano 2021, 15(2), 3139–3151. DOI 10.1021/acsnano.0c09685.
[40]. Konyuba Y., Marubayashi H., Haruta T., Jinnai H. Correlative light and electron microscopy of poly(L-lactic acid) spherulites for fast morphological measurements using a convolutional neural network. Microscopy 2022, 71(2), 104–110.
[41]. Zhang S., Liang X., Huang X., Wang K., Qiu T. Precise and fast microdroplet size distribution measurement using deep learning. Chem. Eng. Sci. 2022, 247, 116926.
[42]. Samide A., Stoean C., Stoean R. Surface study of inhibitor films formed by polyvinyl alcohol and silver nanoparticles on stainless steel in hydrochloric acid solution using convolutional neural networks. Appl. Surf. Sci. 2019, 475, 1–5.
[43]. Xu H., Ma S., Hou Y., et al. Machine learning-assisted identification of copolymer microstructures based on microscopic images. ACS Appl. Mater. Interfaces 2022, 14(41), 47157–47166.

[44]. Tao L., Arbaugh T., Byrnes J., Varshney V., Li Y. Unified machine learning protocol for copolymer structure–property predictions. STAR Protoc. 2022, 3(4), 101875. DOI 10.1016/j.xpro.2022.101875.
[45]. Tao L., Byrnes J., Varshney V., Li Y. Machine learning strategies for the structure–property relationship of copolymers. iScience 2022, 25(7), 104585.
[46]. Yang S., Zhao C., Ren J., Zheng K., Shao Z., Ling S. Acquiring structural and mechanical information of a fibrous network through deep learning. Nanoscale 2022, 14(13), 5044–5053.
[47]. Yang S., Zhao C., Ren J., Zheng K., Shao Z., Ling S. Acquiring structural and mechanical information of a fibrous network through deep learning. Nanoscale 2022, 14, 5044–5053. DOI 10.1039/D2NR00372D.
[48]. Vollmar M., Evans G. Machine learning applications in macromolecular X-ray crystallography. Crystallogr. Rev. 2021, 27(2), 54–101.
[49]. Billinge S. J. L., Proffen T. Machine learning in crystallography and structural science. Found. Adv. 2024, 80(2), 139–145. DOI 10.1107/S2053273324000172.
[50]. Zhong W., Liu F., Wang C. Probing morphology and chemistry in complex soft materials with in situ resonant soft X-ray scattering. J. Phys.: Condens. Matter 2021, 33(31), 313001.
[51]. Munshi J., Chen W., Chien T., Balasubramanian G. Transfer learned designer polymers for organic solar cells. J. Chem. Inf. Model. 2021, 61, 134–142. DOI 10.1021/acs.jcim.0c00929
[52]. Sunil M., Pallikkavaliyaveetil N., N. M., Gopinath A., Chidangil S., Kumar S., Lukose J. Machine learning assisted Raman spectroscopy: A viable approach for the detection of microplastics. J. Water Process Eng. 2024, 60, 105150. DOI 10.1016/j.jwpe.2024.105150
[53]. Workman J. Jr., Mark H. Artificial intelligence in analytical spectroscopy, Part II: Examples in spectroscopy. Spectroscopy. 2023, 38, 10–15. DOI 10.56530/spectroscopy.js8781e3
[54]. Chatzitoulousis P. Machine learning algorithms for spectral data. 2024. DOI 10.13140/RG.2.2.20933.15841
[55]. Shin S., Choi Y.-Y., Sohn Y.-J., Kim M., Lim I.S., Oh H. Machine learning-based fault diagnosis for various steady conditions of proton exchange membrane fuel cell systems. Int. J. Hydrogen Energy. 2024, 89, 507–517. DOI 10.1016/j.ijhydene.2024.09.227
[56]. Wang Y., Seo B., Wang B., Zamel N., Jiao K., Cordobes Adroher X. Fundamentals, materials, and machine learning of polymer electrolyte membrane fuel cell technology. Energy AI. 2020, 1, 100014. DOI 10.1016/j.egyai.2020.100014
[57]. Kaiser R., Ahn C.-Y., Kim Y.-H., Park J.-C. Towards reliable prediction of performance for polymer electrolyte membrane fuel cells via machine learning-integrated hybrid numerical simulations. Processes. 2024, 12, 1140. DOI 10.3390/pr12061140
[58]. Oliveira R.M.A., Sant'Anna Â.M.O., da Silva P.H.F. Explainable machine learning models for defects detection in industrial processes. Comput. Ind. Eng. 2024, 192, 110214. DOI 10.1016/j.cie.2024.110214
[59]. Abroshan H., Winget P., Kwak H.S., An Y., Brown C.T., Halls M.D. Machine learning for the design of novel OLED materials. In: Machine Learning in Materials Informatics: Methods and Applications. ACS Symp. Ser. 2022, 1416, 33–49. DOI 10.1021/bk-2022-1416.ch002
[60]. Shi Y., Shi H., Wang H., Chen C.-J., Li Y., Qiao B., Liang Z., Zhao S., Hang D., Xu Z., Song D. Machine learning-driven precise design of stable OLED materials: Predicting and enhancing multi-state C-N bond dissociation energies. Chem. Eng. J. 2024, 500, 157082. DOI 10.1016/j.cej.2024.157082
[61]. Han D.-H., Jeong Y.-H., Kang M.-S. A Study on OLED Cell Simulation and Detection Phases Based on the A2G Algorithm for Artificial Intelligence Application. Appl. Sci. 2023, 13, 8016. DOI 10.3390/app13148016.
[62]. Ginige G., Song Y., Olsen B. C., Luber E. J., Yavuz C. T., Buriak J. M. Solvent Vapor Annealing, Defect Analysis, and Optimization of Self-Assembly of Block Copolymers Using Machine Learning Approaches. ACS Appl. Mater. Interfaces 2021, 13, 24, 28639–28649. DOI 10.1021/acsami.1c05056.
[63]. Tu K.-H., Huang H., Lee S., Lee W., Sun Z., Alexander-Katz A., Ross C. A. Machine Learning Predictions of Block Copolymer Self-Assembly. Adv. Mater. 2020, 32, 52, 2005713. DOI 10.1002/adma.202005713.
[64]. Higuchi C., Horvath D., Marcou G., Yoshizawa K., Varnek A. Prediction of the Glass-Transition Temperatures of Linear Homo/Heteropolymers and Cross-Linked Epoxy Resins. ACS Appl. Polym. Mater. 2019, 1, 6, 1430–1442. DOI 10.1021/acsapm.9b00253.
[65]. Pugar J. A., Childs C. M., Huang C., Haider K. W., Washburn N. R. Elucidating the Physicochemical Basis of the Glass Transition Temperature in Linear Polyurethane Elastomers with Machine Learning. J. Phys. Chem. B 2020, 124, 43, 9722–9733. DOI 10.1021/acs.jpcb.0c06467.
[66]. Ishikiriyama K. Polymer Informatics Based on the Quantitative Structure-Property Relationship Using a Machine-Learning Framework for the Physical Properties of Polymers in the ATHAS Data Bank. Thermochim. Acta 2022, 708, 179135. DOI 10.1016/j.tca.2021.179135.
[67]. Bhowmik R., Sihn S., Pachter R., Vernon J. P. Prediction of the Specific Heat of Polymers from Experimental Data and Machine Learning Methods. Polymer 2021, 220, 123558. DOI 10.1016/j.polymer.2021.123558.
[68]. Ding D., Zou M., Wang X., et al. Thermal Conductivity of Polydisperse Hexagonal BN/Polyimide Composites: Iterative EMT Model and Machine Learning Based on First Principles Investigation. Chem. Eng. J. 2022, 437, 135438. DOI 10.1016/j.cej.2022.135438.

[69]. Ma R., Zhang H., Xu J., et al. Machine Learning-Assisted Exploration of Thermally Conductive Polymers Based on High-Throughput Molecular Dynamics Simulations. Mater. Today Phys. 2022, 28, 100850. DOI 10.1016/j.mtphys.2022.100850.
[70]. Ma R., Zhang H., Luo T. Exploring High Thermal Conductivity Amorphous Polymers Using Reinforcement Learning. ACS Appl. Mater. Interfaces 2022, 14, 13, 15587–15598. DOI 10.1021/acsami.2c02163.
[71]. Li M., Dai L., Hu Y. Machine Learning for Harnessing Thermal Energy: From Materials Discovery to System Optimization. ACS Energy Lett. 2022, 7, 10, 3204–3226. DOI 10.1021/acsenergylett.2c01574.
[72]. Amamoto Y., Koganemaru C., Kojio K., et al. A Machine Learning Approach to Designing and Understanding Tough, Degradable Polyamides. npj Comput. Mater. 2025, 11, 198. DOI 10.1038/s41524-025-01696-1.
[73]. Lin T.-S., Coley C. W., Mochigase H., Beech H. K., Wang W., Wang Z., Woods E., Craig S. L., Johnson J. A., Kalow J. A., Jensen K. F., Olsen B. D. BigSMILES: A Structurally-Based Line Notation for Describing Macromolecules. ACS Cent. Sci. 2019, 5, 9, 1523–1531. DOI 10.1021/acscentsci.9b00476.
[74]. Polymer Database (PoLyInfo). Available online: https://polymer.nims.go.jp/ (accessed on 18 October 2023).
[75]. Otsuka S., Kuwajima I., Hosoya J., Xu Y., Yamazaki M. PoLyInfo: Polymer Database for Polymeric Materials Design. In *Proc. 2011 Int. Conf. Emerging Intelligent Data and Web Technologies*, Tirana, Albania, 7–9 September 2011; pp. 22–29. DOI 10.1109/EIDWT.2011.35.
[76]. Malashin I.P., Tynchenko V.S., Nelyub V.A., Borodulin A.S., Gantimurov A.P. Estimation and Prediction of the Polymers' Physical Characteristics Using the Machine Learning Models. Polymers. 2024, 16, 115. DOI 10.3390/polym16010115
[77]. Jarrett D., Cebere B.C., Liu T., Curth A., van der Schaar M. Hyperimpute: Generalized Iterative Imputation With Automatic Model Selection. Proc. Mach. Learn. Res. 2022, 16, 9916–9937.
[78]. Yoon J., Jordon J., van der Schaar M. GAIN: Missing Data Imputation Using Generative Adversarial Nets. Proc. Mach. Learn. Res. 2018, 80, 5689–5698.
[79]. Muzellec B., Josse J., Boyer C., Cuturi M. Missing Data Imputation Using Optimal Transport. Proc. 37th Int. Conf. Mach. Learn. 2020, 7130–7140.
[80]. Liu L., Xie C., Hu W., Li Y. Recursive Elimination of 'Outliers' to Get Benchmark Dataset. IEEE Access. 2024, 12, 98319–98325.
[81]. Doan Tran H., Kim C., Chen L., et al. Machine-Learning Predictions of Polymer Properties With Polymer Genome. J. Appl. Phys. 2020, 128, 171104.
[82]. Otsuka S., Kuwajima I., Hosoya J., Xu Y., Yamazaki M. PoLyInfo: Polymer Database for Polymeric Materials Design. Int. Conf. Emerg. Intell. Data Web Technol. 2011. DOI 10.1109/EIDWT.2011.13
[83]. Ma R., Luo T. PI1M: A Benchmark Database for Polymer Informatics. J. Chem. Inf. Model. 2020, 60, 4684–4690.
[84]. Hisam M.W., Dar A.A., Elrasheed M.O., et al. The Versatility of the Taguchi Method: Optimizing Experiments Across Diverse Disciplines. J. Stat. Theory Appl. 2024, 23, 365–389. DOI 10.1007/s44199-024-00093-9
[85]. Sterkenburg T.F., Grünwald P.D. The No-Free-Lunch Theorems of Supervised Learning. Synthese. 2021, 199, 9979–10015. DOI 10.1007/s11229-021-03233-1
[86]. Vişan C., Boldeanu M., Nicolae G., Cucu H., Burileanu C., Buzo A. Evolutionary Bayesian Optimization for Automated Circuit Sizing. Knowl.-Based Syst. 2025, 318, 113483. DOI 10.1016/j.knosys.2025.113483
[87]. Feng Q.-K., Zhong S.-L., Pei J.-Y., et al. Recent Progress and Future Prospects on All-Organic Polymer Dielectrics for Energy Storage Capacitors. Chem. Rev. 2022, 122, 3820–3878.
[88]. Párizs R.D., Török D., Ageyeva T., Kovács J.G. Machine Learning in Injection Molding: An Industry 4.0 Method of Quality Prediction. Sensors. 2022, 22, 2704. DOI 10.3390/s22072704
[89]. Nematov D., Hojamberdiev M. Machine Learning-Driven Materials Discovery: Unlocking Next-Generation Functional Materials. Condens. Matter > Mater. Sci. 2025. DOI 10.48550/arXiv.2503.18975
[90]. Liu M., Li H., Zhou H., Zhang H., Huang G. Development of Machine Learning Methods for Mechanical Problems Associated with Fibre Composite Materials: A Review. Compos. Commun. 2024, 49, 101988. DOI 10.1016/j.coco.2024.101988
[91]. Ohno M., Hayashi Y., Zhang Q., Kaneko Y., Yoshida R. SMiPoly: Generation of a Synthesizable Polymer Virtual Library Using Rule-Based Polymerization Reactions. J. Chem. Inf. Model. 2023, 63, 17, 5539–5548. DOI: 10.1021/acs.jcim.3c00329
[92]. Selvaraj S. C. Graph Neural Networks Based Deep Learning for Predicting Structural and Electronic Properties. arXiv. 2024. DOI: 10.48550/arXiv.2411.02331 (Withdrawn)
[93]. Kuenneth C., Ramprasad R. polyBERT: A Chemical Language Model to Enable Fully Machine-Driven Ultrafast Polymer Informatics. Nat. Commun. 2023, 14, 4099. DOI: 10.1038/s41467-023-39868-6
[94]. Qiu H., Liu L., Qiu X., Dai X., Ji X., Sun Z.-Y. PolyNC: A Natural and Chemical Language Model for the Prediction of Unified Polymer Properties. Chem. Sci. 2024, 15, 534–544. DOI: 10.1039/D3SC05079C
[95]. Lai N. S., Tew Y. S., Zhong X., Yin J., Li J., Yan B., Wang X. Artificial Intelligence (AI) Workflow for Catalyst Design and Optimization. Ind. Eng. Chem. Res. 2023, 62, 43, 17835–17848. DOI: 10.1021/acs.iecr.3c02520
[96]. Bergstra J., Yamins D., Cox D. D. Hyperopt: A Python Library for Model Selection and Hyperparameter Optimization. Comput. Sci. Discov. 2015, 8, 014008. DOI: 10.1088/1749-4699/8/1/014008

[97]. Akiba T., Sano S., Yanase T., Ohta T., Koyama M. Optuna: A Next-Generation Hyperparameter Optimization Framework. Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. 2019, 2623–2631. DOI: 10.1145/3292500.3330701
[98]. Scikit-Optimize: Sequential Model-Based Optimization in Python. [Электронный ресурс]: https://scikit-optimize.github.io/stable/ (дата обращения: 03.08.2025)
[99]. Aragão M. V. C., Afonso A. G., Ferraz R. C., et al. A Practical Evaluation of AutoML Tools for Binary, Multiclass, and Multilabel Classification. Sci. Rep. 2025, 15, 17682. DOI: 10.1038/s41598-025-02149-x
[100]. Le Cun Y. A Theoretical Framework for Back-Propagation. In: Touretzky D., Hinton G., Sejnowski T. (Eds.) Proceedings of the 1988 Connectionist Models Summer School, CMU, Pittsburgh, PA. Morgan Kaufmann, 1988, 21–28.
[101]. Jeng S.-L. Generative Adversarial Network for Synthesizing Multivariate Time-Series Data in Electric Vehicle Driving Scenarios. Sensors. 2025, 25, 749. DOI 10.3390/s25030749
[102]. Liu Y., Yang Z., Yu Z., Liu Z., Liu D., Lin H., Li M., Ma S., Avdeev M., Shi S. Generative artificial intelligence and its applications in materials science: Current situation and future perspectives. J. Materiomics. 2023, 9, 798–816. DOI 10.1016/j.jmat.2023.05.001
[103]. Okabe R., Chotrattanapituk A., Boonkird A., Andrejevic N., Fu X., Jaakkola T.S., Song Q., Nguyen T., Drucker N., Mu S., Wang Y., Liao B., Cheng Y., Li M. Virtual node graph neural network for full phonon prediction. Nat. Comput. Sci. 2024, 4, 522–531. DOI 10.1038/s43588-024-00661-0
[104]. Bagal V., Aggarwal R., Vinod P.K., Priyakumar U.D. MolGPT: Molecular Generation Using a Transformer-Decoder. J. Chem. Inf. Model. 2022, 62, 2064–2076. DOI 10.1021/acs.jcim.1c00600
[105]. Bran M.A., Cox S., Schilter O., et al. Augmenting large language models with chemistry tools. Nat. Mach. Intell. 2024, 6, 525–535. DOI 10.1038/s42256-024-00832-8
[106]. Ricci E., Vergadou N. Integrating Machine Learning in the Coarse-Grained Molecular Simulation of Polymers. J. Phys. Chem. B. 2023, 127, 2302–2322.
[107]. Xie C., Qiu H., Liu L., You Y., Li H., Li Y., Sun Z., Lin J., An L. Machine Learning Approaches in Polymer Science: Progress and Fundamental for a New Paradigm. SmartMat. 2025, 6, e1320. DOI 10.1002/smm2.1320
[108]. Ricci E., Vergadou N. Integrating Machine Learning in the Coarse-Grained Molecular Simulation of Polymers. J. Phys. Chem. B. 2023, 127, 2302–2322.
[109]. Pozdnyakov S.N., Ceriotti M. Incompleteness of Graph Neural Networks for Points Clouds in Three Dimensions. Mach. Learn.: Sci. Technol. 2022, 3, 045020.
[110]. Aoki Y., Wu S., Tsurimoto T., et al. Multitask Machine Learning to Predict Polymer–Solvent Miscibility Using Flory–Huggins Interaction Parameters. Macromolecules. 2023, 56, 5446–5456.
[111]. Liang Z., Tan Z., Hong R., Ouyang W., Yuan J., Zhang C. Automatically Predicting Material Properties With Microscopic Images: Polymer Miscibility as an Example. J. Chem. Inf. Model. 2023, 63, 5971–5980.
[112]. Zhou T., Qiu D., Wu Z., et al. Compatibilization Efficiency of Graft Copolymers in Incompatible Polymer Blends: Dissipative Particle Dynamics Simulations Combined With Machine Learning. Macromolecules. 2022, 55, 7893–7907.
[113]. Zhao Q., Shan Y., Xiang C., et al. Predicting Power Conversion Efficiency of Binary Organic Solar Cells Based on Y6 Acceptor by Machine Learning. J. Energy Chem. 2023, 82, 139–147.
[114]. Bentéjac C., Csörgő A., Martínez-Muñoz G. A comparative analysis of gradient boosting algorithms. Artif. Intell. Rev. 2021, 54, 1937–1967.
[115]. Chen H., Cheng Y., Du T., Wu X., Cao Y., Liu Y. Enhancing the performance of recycled aggregate green concrete via a Bayesian optimization light gradient boosting machine and the nondominated sorting genetic algorithm-III. Constr. Build. Mater. 2025, 458, 139527.
[116]. Meng S., Shi Z., Xia C., Zhou C., Zhao Y. Exploring LightGBM-SHAP: Interpretable predictive modeling for concrete strength under high temperature conditions. Structures. 2025, 71, 108134. DOI 10.1016/j.istruc.2024.108134.
[117]. Ke G., Meng Q., Finley T., Wang T., Chen W., Ma W., Ye Q., Liu T. Y. Lightgbm: A highly efficient gradient boosting decision tree. Adv. Neural Inf. Process. Syst. 2017, 30, 52.
[118]. Hancock J. T., Khoshgoftaar T. M. CatBoost for big data: An interdisciplinary review. J. Big Data. 2020, 7, 94. DOI 10.1186/s40537-020-00369-8.
[119]. Wei C., Li Z., Zhu D., Xu T., Liang Z., Liu Y., Zhao N. Regulation of the physicochemical properties of nutrient solution in hydroponic system based on the CatBoost model. Comput. Electron. Agric. 2025, 229, 109729. DOI 10.1016/j.compag.2024.109729.
[120]. Wang R., Zhang M., Gong F., Wang S., Yan R. Improving port state control through a transfer learning-enhanced XGBoost model. Reliab. Eng. Syst. Saf. 2025, 253, 110558. DOI 10.1016/j.ress.2024.110558.
[121]. Chen T., Guestrin C. Xgboost: A scalable tree boosting system. Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. 2016, 785–794. DOI 10.1145/2939672.2939785.
[122]. Malashin I., Tynchenko V., Gantimurov A., Nelyub V., Borodulin A. Boosting-Based Machine Learning Applications in Polymer Science: A Review. Polymers. 2025, 17, 499. DOI 10.3390/polym17040499.

[123]. Xiao J., Hobson J., Ghosh A., Haranczyk M., Wang D.-Y. Flame Retardant Properties of Metal Hydroxide-Based Polymer Composites: A Machine Learning Approach. Compos. Commun. 2023, 40, 101593. DOI 10.1016/j.coco.2023.101593.
[124]. Xiao J., Hobson J., Haranczyk M., Wang D. Y. Machine Learning Framework to Predict Instantaneous Heat Release Rate of Polymer Nanocomposites in Cone Calorimetry. Polym. Degrad. Stab. 2023, 218, 110563. DOI 10.1016/j.polymdegradstab.2023.110563.
[125]. Ding F., Liu L. Y., Liu T. L., Li Y. Q., Li J. P., Sun Z. Y. Predicting the Mechanical Properties of Polyurethane Elastomers Using Machine Learning. Chin. J. Polym. Sci. 2023, 41, 422–431. DOI 10.1007/s10118-023-2916-0.
[126]. Marani R., Campos-Delgado D. U. Depth Classification of Defects in Composite Materials by Long-Pulsed Thermography and Blind Linear Unmixing. Compos. B Eng. 2023, 248, 110359. DOI 10.1016/j.compositesb.2022.110359.
[127]. Hilloulin B., Tran V. Q. Using Machine Learning Techniques for Predicting Autogenous Shrinkage of Concrete Incorporating Superabsorbent Polymers and Supplementary Cementitious Materials. J. Build. Eng. 2022, 49, 104086. DOI 10.1016/j.jobe.2022.104086.
[128]. Wang Z., Chen Z., Wang B., Wu C., Zhou C., Peng Y., Zhang X., Ni Z., Chung C.-Y., Chan C.-C., Yang J., Zhao H. Digital manufacturing of perovskite materials and solar cells. Appl. Energy. 2025, 377, 124120. DOI 10.1016/j.apenergy.2024.124120.
[129]. Hu Q., Lu J., Hui J., Rao Z., Ren Y., Wang H. Artificial Intelligence-Driven Development in Rechargeable Battery Materials: Progress, Challenges, and Future Perspectives. Adv. Funct. Mater. 2025, e08438. DOI 10.1002/adfm.202508438.
[130]. Louis S. Y., Siriwardane E. M. D., Joshi R. P., Omee S. S., Kumar N., Hu J. Accurate Prediction of Voltage of Battery Electrode Materials Using Attention-Based Graph Neural Networks. ACS Appl. Mater. Interfaces. 2022, 14, 23, 26587–26594. DOI 10.1021/acsami.2c00029.
[131]. Kirchartz T. Challenges and opportunities for the characterization of electronic properties in halide perovskite solar cells. Chem. Sci. 2025, 16, 8153. DOI 10.1039/d5sc00504c.
[132]. Eguchi N., Fukazawa T., Kanda H., Yamamoto K., Miyake T., Murakami T. N. Performance optimization of perovskite solar cells with an automated spin coating system and artificial intelligence technologies. EES Sol. 2025, 1, 320. DOI 10.1039/d5el00007f.
[133]. Cakan D. N., Oberholtz E., Kaushal K., Dunfield S. P., Fenning D. P. Bayesian optimization and prediction of the durability of triple-halide perovskite thin films under light and heat stressors. Mater. Adv. 2025,6, 598-606. https://doi.org/10.1039/D4MA00747F
[134]. An K., Zhong W., Peng F., et al. Mastering morphology of non-fullerene acceptors towards long-term stable organic solar cells. Nat. Commun. 2023, 14, 2688. DOI 10.1038/s41467-023-38306-x.
[135]. Xiao D., Hou M., Guo L., Yang J., Jiang W., Hong T. Physics and statistics co-enhanced Gaussian process regression for efficient modeling of the conducting target's RCS. Opt. Express. 2025, 33, 4, 8239. DOI 10.1364/OE.551078.
[136]. [polyBERT репозиторий на GitHub] https://github.com/Ramprasad-Group/polyBERT
[137]. Shetty P., Adeboye A., Gupta S., Zhang C., Ramprasad R. Accelerating Materials Discovery for Polymer Solar Cells: Data-Driven Insights Enabled by Natural Language Processing. Chem. Mater. 2024, 36, 7676−7689. DOI 10.1021/acs.chemmater.4c00709.
[138]. Abbas M., Zeng L., Guo F., Rauf M., Yuan X.-C., Cai B. A Critical Review on Crystal Growth Techniques for Scalable Deposition of Photovoltaic Perovskite Thin Films. Mater. 2020, 13, 4851. DOI 10.3390/ma13214851.
[139]. Dembitskiy A. D., Humonen I. S., Eremin R. A., et al. Benchmarking machine learning models for predicting lithium ion migration. npj Comput. Mater. 2025, 11, 131. DOI 10.1038/s41524-025-01571-z.
[140]. Wang S., Liu J., Song X., et al. Artificial Intelligence Empowers Solid-State Batteries for Material Screening and Performance Evaluation. Nano-Micro Lett. 2025, 17, 287. DOI 10.1007/s40820-025-01797-y.
[141]. Ding Y., Zhu J., Liu Y., Ning D., Qin M. Integrated Framework of LSTM and Physical-Informed Neural Network for Lithium-Ion Battery Degradation Modeling and Prediction. AI. 2025, 6, 149. DOI 10.3390/ai6070149
[142]. Ruza J., Stolberg M., Cawthern S., Johnson J., Shao-Horn Y., Gomez-Bombarelli R. Autonomous Discovery of Polymer Electrolyte Formulations with Warm-start Batch Bayesian Optimization. ChemRxiv. 2025. DOI 10.26434/chemrxiv-2025-2cjbg
[143]. Seo H., Kim D., Park S., et al. A comprehensive review of liquid electrolytes for silicon anodes in lithium-ion batteries. Adv. Ind. Eng. Chem. 2025, 1, 3. DOI 10.1007/s44405-025-00004-1
[144]. Mienye I.D., Swart T.G., Obaido G., Jordan M., Ilono P. Deep Convolutional Neural Networks in Medical Image Analysis: A Review. Information. 2025, 16, 195. DOI 10.3390/info16030195
[145]. Zou X., Pan J., Sun Z., Wang B., Jin Z., Xu G., Yan F. Machine learning analysis and prediction models of alkaline anion exchange membranes for fuel cells. Energy Environ. Sci. 2021, 14, 3965–3975. DOI 10.1039/D1EE01170G
[146]. Khan K.H., Haleem A., Arwish S., Shah A., Hussain H. PVDF-based solid polymer electrolytes for lithium-ion batteries: strategies in composites, blends, dielectric engineering, and machine learning approaches. RSC Adv. 2025, 15, 20629–20656. DOI 10.1039/d5ra02951a

[147]. Huang B., Xu K., Zhang X., Zhang Y., Long Y.-Z., Ge S.S., Wei X. 3D printing of self-powered triboelectric intelligent system with enhanced output performance for material identification via machine learning. Chem. Eng. J. 2025, 515, 163790. DOI 10.1016/j.cej.2025.163790
[148]. Taheri S.H., Mohammadpour J., Lalbakhsh A., et al. Machine learning-driven design of wide-angle impedance matching structures for wide-angle scanning arrays. Sci. Rep. 2025, 15, 16601. DOI 10.1038/s41598-025-00310-0
[149]. Akhter F., Siddiquei H.R., Alahi M.E.E., Jayasundera K.P., Mukhopadhyay S.C. An IoT-Enabled Portable Water Quality Monitoring System With MWCNT/PDMS Multifunctional Sensor for Agricultural Applications. IEEE Internet Things J. 2022, 9, 14307–14316.
[150]. Verma P., Yadava R.D.S. A Data Mining Procedure for Polymer Selection for Making Surface Acoustic Wave Sensor Array. Sensor Lett. 2013, 11, 1903–1918.
[151]. Peveler W. J., Binions R., Hailes S. M. V., Parkin I. P. Detection of explosive markers using zeolite modified gas sensors. J. Mater. Chem. A. 2013, 1, 2613–2620. DOI 10.1039/C2TA00855E
[152]. Serrano J., Moros J., Sánchez C., Macías J., Laserna J. J. Advanced recognition of explosives in traces on polymer surfaces using LIBS and supervised learning classifiers. Anal. Chim. Acta. 2014, 806, 107–116. DOI 10.1016/j.aca.2013.11.013
[153]. Shi Y., Yuan H., Zhang Q., Sun A., Liu J., Men H. Lightweight interleaved residual dense network for gas identification of industrial polypropylene coupled with an electronic nose. IEEE Trans. Instrum. Meas. 2021, 70, 1–10. DOI 10.1109/TIM.2021.3095083
[154]. Matos M. A. S., Pinho S. T., Tagarielli V. L. Application of machine learning to predict the multiaxial strain-sensing response of CNT-polymer composites. Carbon. 2019, 146, 265–275. DOI 10.1016/j.carbon.2019.01.052
[155]. Li W., Yang T., Liu C., et al. Optimizing piezoelectric nanocomposites by high-throughput phase-field simulation and machine learning. Adv. Sci. 2022, 9, 2105550. DOI 10.1002/advs.202105550
[156]. Wang X., Song W.-Z., You M.-H., et al. Bionic single-electrode electronic skin unit based on piezoelectric nanogenerator. ACS Nano. 2018, 12, 8588–8596. DOI 10.1021/acsnano.8b04090
[157]. Ma C., Gao S., Gao X., et al. Chitosan biopolymer-derived self-powered triboelectric sensor with optimized performance through molecular surface engineering and data-driven learning. InfoMat. 2019, 1, 116–125. DOI 10.1002/inf2.12003
[158]. Zhang M., Li J., Kang L., et al. Machine learning-guided design and development of multifunctional flexible Ag/poly(amic acid) composites using the differential evolution algorithm. Nanoscale. 2020, 12, 3988–3996. DOI 10.1039/C9NR09560F
[159]. Kim J.-N., Lee J., Lee H., Oh I.-K. Stretchable and self-healable catechol-chitosan-diatom hydrogel for triboelectric generator and self-powered tremor sensor targeting at Parkinson disease. Nano Energy. 2021, 82, 105705. DOI 10.1016/j.nanoen.2020.105705
[160]. Jumper J., Evans R., Pritzel A., et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021, 596, 583–589. DOI 10.1038/s41586-021-03819-2
[161]. Necci M., Piovesan D., Hoque M. T., et al. Critical assessment of protein intrinsic disorder prediction. Nat. Methods. 2021, 18, 472–481. DOI 10.1038/s41592-021-01117-3
[162]. Kerner A., Dogan A., von Recum H. Machine learning and big data provide crucial insight for future biomaterials discovery and research. Acta Biomater. 2021, 130, 54–65. DOI 10.1016/j.actbio.2021.06.037
[163]. Gianti E., Percec S. Machine learning at the interface of polymer science and biology: How far can we go? Biomacromolecules. 2022, 23, 576–591. DOI 10.1021/acs.biomac.1c01379
[164]. Suwardi A., Wang F., Xue K., et al. Machine learning-driven biomaterials evolution. Adv. Mater. 2022, 34, 2102703. DOI 10.1002/adma.202102703
[165]. Patel R. A., Webb M. A. Data-driven design of polymer-based biomaterials: High-throughput simulation, experimentation, and machine learning. ACS Appl. Bio Mater. 2023, 7, 510–527. DOI 10.1021/acsabm.2c01388
[166]. Li Y., He Z., A S., et al. Artificial intelligence (AI)-aided structure optimization for enhanced gene delivery: The effect of the polymer component distribution (PCD). ACS Appl. Mater. Interfaces. 2023, 15, 36667–36675. DOI 10.1021/acsami.3c09228
[167]. Guarda C., Caseiro J., Pires A. Machine learning to enhance sustainable plastics: A review. J. Clean. Prod. 2024, 474, 143602. DOI 10.1016/j.jclepro.2024.143602
[168]. Kannankai M. P., Babu A. J., Radhakrishnan A., Alex R. K., Borah A., Devipriya S. P. Machine learning aided meta-analysis of microplastic polymer composition in global marine environment. J. Hazard. Mater. 2022, 440, 129801. DOI 10.1016/j.jhazmat.2022.129801
[169]. Min K., Cuiffi J. D., Mathers R. T. Ranking environmental degradation trends of plastic marine debris based on physical properties and molecular structure. Nat. Commun. 2020, 11, 727. DOI 10.1038/s41467-020-14538-z
[170]. Wu D., Zhang D., Liu S., et al. Prediction of polycarbonate degradation in natural atmospheric environment of China based on BP-ANN model with screened environmental factors. Chem. Eng. J. 2020, 399, 125878. DOI 10.1016/j.cej.2020.125878
[171]. Lu H., Diaz D. J., Czarnecki N. J., et al. Machine learning-aided engineering of hydrolases for PET depolymerization. Nature. 2022, 604, 662–667. DOI 10.1038/s41586-022-04599-z

[172]. Ji W., Richter F., Gollner M. J., Deng S. Autonomous kinetic modeling of biomass pyrolysis using chemical reaction neural networks. Combust. Flame. 2022, 240, 111992. DOI 10.1016/j.combustflame.2022.111992
[173]. Alabdrabalnabi A., Gautam R., Sarathy S. M. Machine learning to predict biochar and bio-oil yields from co-pyrolysis of biomass and plastics. Fuel. 2022, 328, 125303. DOI 10.1016/j.fuel.2022.125303
[174]. Li W., Ma H., Li S., Ma J. Computational and data driven molecular material design assisted by low scaling quantum mechanics calculations and machine learning. Chem. Sci. 2021, 12, 14987–15006. DOI 10.1039/D1SC04090A
[175]. Ramos E., Lopes A. G., Mendonça F. Application of machine learning in plastic waste detection and classification: A systematic review. Processes. 2024, 12, 1632. DOI 10.3390/pr12081632
[176]. Emifoniye E. U., Erameh A. A., Orhorhoro E. K., Oyiboruona P. E. A Machine Learning Approach for Pyrolysis of Plastic Waste: An Overview of Types, and Applications. Int. J. Sci. Res. Eng. Dev. 2025, 8, 27–39.
[177]. Feng R., Tran H., Toland A., Chen B., Zhu Q., Ramprasad R., Zhang C. PolyGET: Accelerating Polymer Simulations by Accurate and Generalizable Forcefield with Equivariant Transformer. arXiv. 2023, arXiv:2309.00585. https://doi.org/10.48550/arXiv.2309.00585
[178]. Yue T., He J., Li Y. Machine-Learning-Assisted Molecular Design of Innovative Polymers. Acc. Mater. Res. 2025, XXXX, XXX–XXX. https://doi.org/10.1021/accountsmr.5c00151
[179]. Ma L., Li W., Yuan J., Zhu J., Wu Y., He H., Pan X. Recent Advances in Machine Learning-Assisted Design and Development of Polymer Materials. Macromol. Rapid Commun. 2025. https://doi.org/10.1002/marc.202500361
[180]. Miccio L. A. Understanding Polymers Through Transfer Learning and Explainable AI. Appl. Sci. 2024, 14, 10413. https://doi.org/10.3390/app142210413
[181]. Lin L.-C., Chen S.-J., Yu H.-Y. Connecting Structural Characteristics and Material Properties in Phase-Separating Polymer Solutions: Phase-Field Modeling and Physics-Informed Neural Networks. Polymers 2023, 15, 4711. https://doi.org/10.3390/polym15244711
[182]. Kim S., Schroeder C. M., Jackson N. E. Open Macromolecular Genome: Generative Design of Synthetically Accessible Polymers. ACS Polym. Au 2023, 3, 318–330. https://doi.org/10.1021/acspolymersau.3c00003
[183]. Lin C., Zhang H. Polymer Biodegradation in Aquatic Environments: A Machine Learning Model Informed by Meta-Analysis of Structure–Biodegradation Relationships. Environ. Sci. Technol. 2025, 59, 1253–1263. https://doi.org/10.1021/acs.est.4c11282
[184]. Yoshinaka Y., Miller S. A. Bio-oil Derived Polyesteramides as Water-Degradable Replacements for Polyethylene. Green Chem. 2025, 27, 4152. https://doi.org/10.1039/d4gc05490c
[185]. The Role of Machine Learning in Polymer Science. https://polymer-search.com/the-role-of-machine-learning-in-polymer-science/
[186]. Savit A., Sahu H., Shukla S., Xiong W., Ramprasad R. polyBART: A Chemical Linguist for Polymer Property Prediction and Generative Design. Condens. Matter > Soft Condens. Matter. 2025. DOI: 10.48550/arXiv.2506.04233
[187]. Jain A., Ramprasad R. polyGen: A Learning Framework for Atomic-level Polymer Structure Generation. Comput. Sci. > Comput. Eng., Finance, Sci. 2025. DOI: 10.48550/arXiv.2504.17656
[188]. Zhang T., Yang D.-B. Multimodal machine learning with large language embedding model for polymer property prediction. Comput. Sci. > Mach. Learn. 2025. DOI: 10.48550/arXiv.2503.22962
[189]. Zheng Y., Biswal A. K., Guo Y., Thakolkaran P., Kokane Y., Varshney V., Kumar S., Vashisth A. Toward Sustainable Polymer Design: A Molecular Dynamics-Informed Machine Learning Approach for Vitrimers. Condens. Matter > Mater. Sci. 2025. DOI: 10.48550/arXiv.2503.20956
[190]. Nanjo S., Arifin M., Maeda H., et al. SPACIER: On-demand polymer design with fully automated all-atom classical molecular dynamics integrated into machine learning pipelines. NPJ Comput. Mater. 2025, 11, 16. DOI: 10.1038/s41524-024-01492-3
[191]. Karuppusamy M., Thirumalaisamy R., Palanisamy S., Nagamalai S., Massoud E. E. S., Ayrilmis N. A Review of Machine Learning Applications in Polymer Composites: Advancements, Challenges, and Future Prospects. J. Mater. Chem. A. 2025, 13, 16290–16308. DOI: 10.1039/d5ta00982k
[192]. Struble D. C., Lamb B. G., Ma B. A Prospective on Machine Learning Challenges, Progress, and Potential in Polymer Science. MRS Commun. 2024, 14, 752–770. DOI: 10.1557/s43579-024-00587-8
[193]. Lu S., Jayaraman A. Machine Learning for Analyses and Automation of Structural Characterization of Polymer Materials. Prog. Polym. Sci. 2024, 153, 101828. DOI: 10.1016/j.progpolymsci.2024.101828
[194]. Materials Informatics. Wikipedia. https://en.wikipedia.org/wiki/Materials_informatics (accessed 03 Aug 2025)
[195]. Machine-Learned Interatomic Potential. Wikipedia. https://en.wikipedia.org/wiki/Machine-learned_interatomic_potential (accessed 03 Aug 2025)
[196]. van den Hurk R. S., Pirok B. W. J., Bos T. S. The Role of Artificial Intelligence and Machine Learning in Polymer Characterization: Emerging Trends and Perspectives. Chromatographia. 2025, 88, 357–363. DOI: 10.1007/s10337-025-04406-7