IET Nanodielectrics

REVIEW

The Institution of Engineering and Technology WILEY

Review of machine learning-driven design of polymer-based dielectrics

Ming-Xiao Zhu[1] | Ting Deng[1] | Lei Dong[1] | Ji-Ming Chen[1] | Zhi-Min Dang[2]

[1]Department of Electrical Engineering, College of New Energy, China University of Petroleum (East China), Qingdao, China

[2]State Key Laboratory of Power System, Department of Electrical Engineering, Tsinghua University, Beijing, China

Correspondence

Ming-Xiao Zhu, College of New Energy, China University of Petroleum (East China), 66 Changjiang West Road, Qingdao, Shandong 266580, China.
Email: zhumx@upc.edu.cn

Zhi-Min Dang, State Key Laboratory of Power System, Department of Electrical Engineering, Tsinghua University, Beijing 100084, China.
Email: dangzm@tsinghua.edu.cn

Abstract

Polymer-based dielectrics are extensively applied in various electrical and electronic devices such as capacitors, power transmission cables and microchips, in which a variety of distinct performances such as the dielectric and thermal properties are desired. To fulfil these properties, the emerging machine learning (ML) technique has been used to establish a surrogate model for the structure–property linkage analysis, which provides an effective tool for the rational design of the chemical and morphological structure of polymers/nanocomposites. In this article, the authors reviewed the recent progress in the ML algorithms and their applications in the rational design of polymer-based dielectrics. The main routes for collecting training data including online libraries, experiments and high-throughput computations are first summarized. The fingerprints charactering the microstructures of polymers/nanocomposites are presented, followed by the illustration of ML models to establish a mapping between the fingerprinted input and the target properties. Further, inverse design methods such as evolution searching strategies and generative models are described, which are exploited to accelerate the discovery of new polymer-based dielectrics. Moreover, structure–property linkage analysis techniques such as Pearson correlation calculation, decision-tree-based methods and interpretable neural networks are summarized to identify the key features affecting the target properties. The future development prospects of the ML-driven design method for polymer-based dielectrics are also presented in this review.

KEYWORDS

fingerprinting, inverse design, machine learning, polymer-based dielectrics, structure-property linkage analysis

1 | INTRODUCTION

Polymer-based dielectrics are extensively used in various electrical and electronic devices such as capacitors [1, 2], power transmission cables [3, 4] and microchips [5, 6]. To adapt to different work scenarios, a variety of distinct properties such as dielectric and thermal parameters are desired for polymers, which can be tuned through tailoring of the chemical and morphological structure or adding nanofillers/additives to form nanocomposites [7–10]. Extensive efforts have been devoted to the design and development of polymer-based dielectrics to optimize their properties [11, 12]. However, due to the vastness of the chemical and structural space of polymers and nanofillers, the efficient development of polymers or nanocomposites are hampered by the high-cost and time-consuming experiments in the traditional scientific intuition or trial-and-error approach. As a result, new paradigms are expected to efficiently design polymer-based dielectrics with desired properties [13, 14].

The emerging machine learning (ML) technique trained on massive amounts of data establishes linkages between input fingerprints and output properties, which provides a powerful surrogate model for the structure–property linkage analysis [15–19]. Further, inverse design methods such as evolution searching (ES) strategies and generative models can be employed to explore the large space of potential materials, greatly accelerating the discovery and development of new polymers [20, 21]. For instance, Mannodi-Kanakkithodi et al. and Sharma et al. developed a ML-based genome approach to accelerate the discovery of on-demand polymers such as those

with desired dielectric constant, energy density and glass transition temperature, and several potential polymers have been identified [22, 23]. Based on the high-throughput phase-field calculations, Shen et al. presented a ML strategy to evaluate the energy storage capability of polymer nanocomposites [24]. We have also schemed ML models to accelerate the rational design of polymers with high thermal conductivity and polymer nanocomposites with desired dielectric properties [25, 26].

Figure 1 illustrates the workflow of ML methods for the rational design of polymer dielectrics. A dataset composed of a sufficient number of samples with known properties is prepared to train the ML models, in which the data can be derived from experiments or high-throughput computations [27–30]. The samples in a high-performance dataset are representative, which uniformly extracted from the whole data space. The second key component of ML methods is the representation of polymers or nanocomposites with a list of machine-readable characters, which are called 'fingerprints' or 'descriptors'. The selected features in fingerprints should be representative, that is, they must have some effects on the output properties, which can be chemical and structural features or characteristic parameters (e.g. band structure and density of states) depending on the target property [31]. Linear notations such as the simplified molecular-input line-entry system (SMILES) are common fingerprints describing the chemical information of polymers. Using the training dataset as input, the linkage between fingerprints and target properties can be learned by ML models such as kernel-based regression, decision tree and neural network, resulting in a surrogate model for predicting

dielectric properties [32, 33]. Further, inverse design methods such as ES strategies and generative models can be employed to screen the large space of potential materials, which greatly accelerates the discovery and development of new polymers [32, 33]. Moreover, the key features affecting the target properties can be identified by structure–property linkage analysis techniques such as Pearson correlation calculation, decision-tree-based methods and interpretable neural network.

This paper attempts to provide an overview of the application of ML-driven methods for the rational design of polymer or nanocomposite dielectrics. In Section 2, we discuss methods of generating training dataset. Next, the fingerprints of polymer-based dielectrics and the ML algorithms used in the property prediction are summarized in Section 3. We then review rational design protocols such as inverse design and structure–property linkage analysis methods in Section 4. Finally, the conclusions and future perspectives are summarized in Section 6.

2 | DATASET

The ML methods used in designing polymer-based dielectrics mostly involve supervised learning, in which the reliability of these models is largely determined by the amount and quality of samples in the training dataset. Online libraries, experiments and high-throughput computations are the main sources of training data [33, 34]. For instance, several online polymer databases such as PoLyInfo [35], CROW Polymer Property Database [36], Polymer Property Predictor and Database
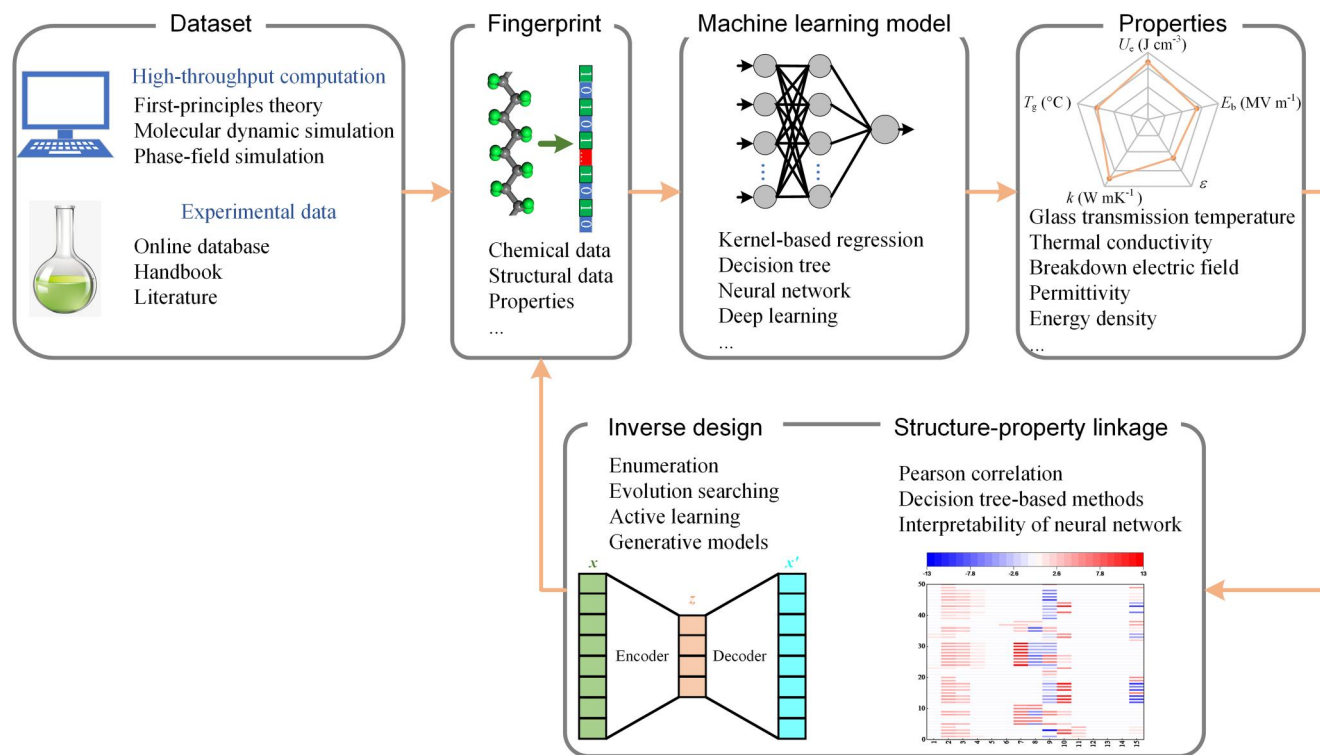

FIGURE 1 The schematic of machine learning methods for the rational design of polymer-based dielectrics

(NIST) [37] and Polymer Genome [38] have been built, allowing for the easy access of training data. By using the thermal parameters in PoLyInfo and QM9 as the training data, a ML framework referred to as 'transfer learning' was presented to predict the thermal conductivity of polymers [39]. In this method, given the limited supply of thermal conductivity parameters, prediction models of physically related proxy properties were pre-trained on large datasets from PoLyInfo and QM9, and then, the pre-trained models were fine-tuned using the limited data on the target property [39]. For polymer nanocomposites schema, only few online databases such as NanoMine have been built [40].

As an alternative approach, experimental data can be collected from scientific research studies including journal articles and handbooks. Timely retrieval of dielectrics data from research studies can be achieved using laborious manual excerption or ML-based natural language processing (NLP) techniques [33]. The manual excerption method searches for usable materials' data by manually screening the studies one by one. For example, 1210 experimentally measured permittivity values at different frequencies were manually collected from studies in the literature to train a ML model to instantly and accurately predict the frequency-dependent permittivity of polymers [34]. To aid accelerated discovery of polymer nanocomposites with the desired breakdown strength, permittivity and energy density, we have extracted hundreds of experimentally measured data from the literature to develop a ML model for predicting dielectric properties [26]. Owing to the difficulty in interpreting technical languages, which requires domain knowledge, the use of NLP in collecting material data is still in its infancy. Some initial studies have used the NLP method to capture materials' data from the literature and to predict new thermoelectric materials [41, 42].

Due to the synthetic difficulty of designed candidates and time-consuming experiments, the amount of dielectric property data is substantially limited. As an effective alternative approach, high-throughput computations [47] using first-principles theory [43, 44, 48–53], molecular dynamics (MD) [45, 54, 55], phase-filed model [46, 56, 57] and finite-element method [58, 59] have been employed to acquire property data, as shown in Figure 2a. Most dielectric properties of interest, for example, permittivity, conduction loss, breakdown strength, glass transmission temperature (T_g) and thermal conductivity, can be directly calculated or indirectly represented with some correlated parameters. For example, the breakdown strength and conduction loss of polymers are closely correlated with the carrier injection process at the electrode–dielectric interface as well as the carrier-transfer characteristics (known to be affected by the bandgap and trap states in polymers) [43, 49–53]. As a result, the charge injection barrier, bandgap and trap depth, which can be determined by the density functional theory (DFT), may serve as 'proxies' for breakdown strength and conduction loss [43, 49, 53]. And many research studies are devoted to investigating the effect of polymer molecular structures on these parameters [51, 52]. As an example, Figure 2b and c depict the DFT method to calculate the charge injection barrier from electrode to polymer and trap depth in polymer, respectively [43, 44].

Moreover, the electronic, ionic and total dielectric constant can be computed by the density functional perturbation theory as well [12, 22]. The thermodynamic properties of polymers or nanocomposites such as glass transmission temperature and thermal conductivity can be readily computed by MD simulations, for example, the non-equilibrium MD have been extensively adopted to calculate the thermal conductivity (Figure 2d) [25, 45]. Given the expensive computational cost of first-principle methods, small, length-scale models (<100 atoms) are generally built to characterize the dielectric properties. The classical MD with empirical force fields can model polymers in a nanoscale scale [33].

As for polymer nanocomposites, large-scale simulations (micrometre to millimetre) based on the phase-filed model or finite-element model (FEM) are required. Phase-field models have been developed to investigate the breakdown behaviour and effective permittivity of polymer nanocomposites (see Figure 2e), in which the effect of the microstructure of nanocomposites such as the shape and orientation of nanofillers can be considered [46, 56]. For example, based on high-throughput phase-field calculations of dielectric response, charge transport and the breakdown process, Shen et al. schemed a ML strategy to evaluate the energy storage capability of polymer nanocomposites [60]. Numerical computation techniques such as FEM are effective tools to understand the large-scale properties, for example, space charge transportation [58, 59] and the thermal conduction process [61]. As an example, a bipolar charge transport model is generally used to characterize the space charge distribution in dielectrics [58, 59], which is correlated with the breakdown strength [62]. Furthermore, a multi-scale modelling approach that applies ab initio, Monte Carlo, and continuum scales was proposed to estimate the breakdown strength of polymer nanocomposites based on the charge trapping effect of the nanofillers [63].

3 | ML STRATEGIES

ML-driven strategies are composed of two distinct steps: numerical representation of the materials in the dataset (fingerprinting) and establishment of a mapping between the fingerprinted input and the target property (learning).

3.1 | Fingerprinting

Fingerprinting encodes the polymers or nanocomposites with a list of machine-readable characters. The choice of the numerical descriptors should duly consider factors (e.g. microstructure and property values) that the target property may be correlated with, which require domain expertise or experience [31].

3.1.1 | Polymer fingerprints

Van Krevelen et al. found that polymer properties such as the glass transition temperature and solubility parameter are
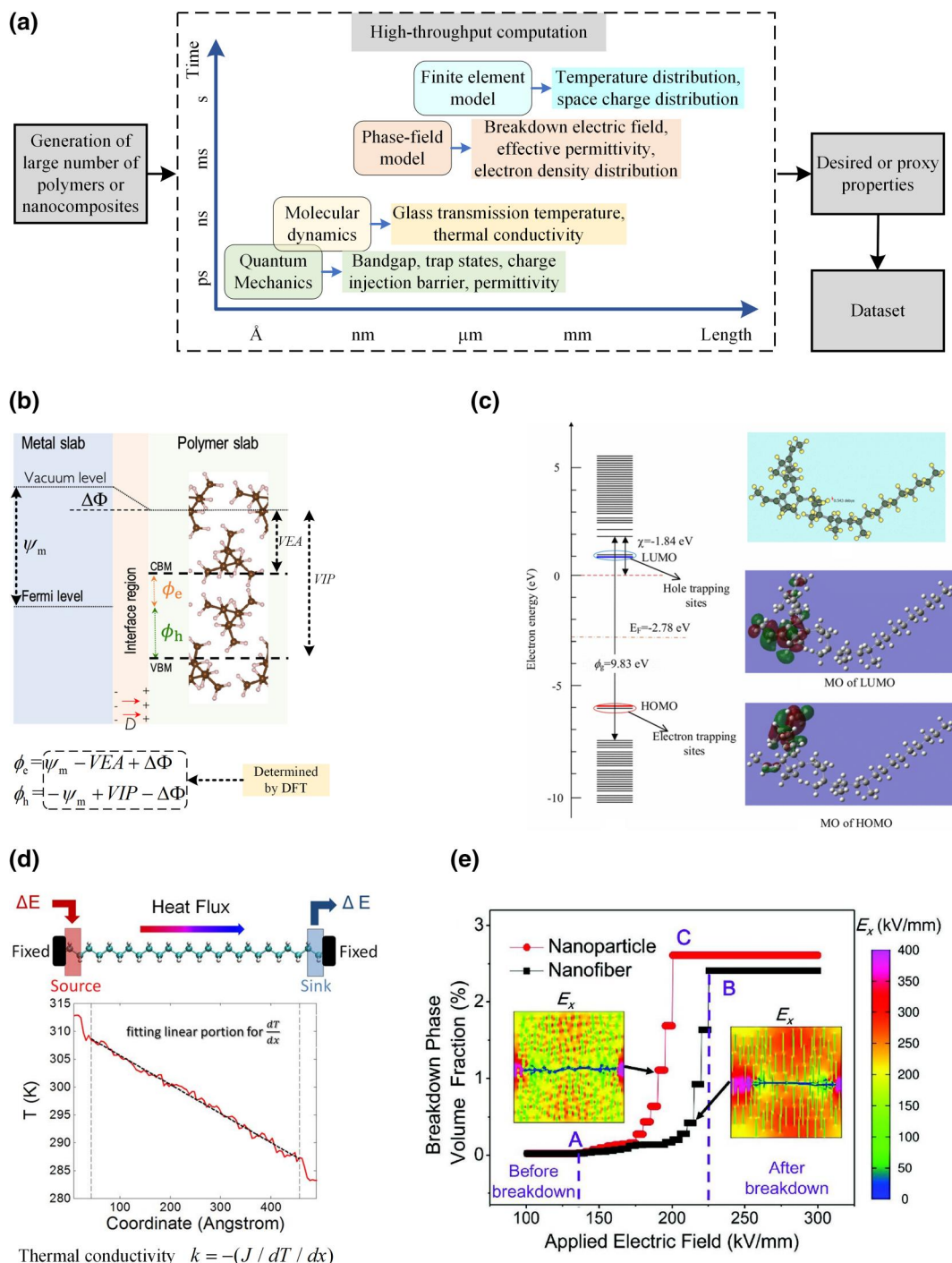
(a)

(b)

$$\phi_e = \psi_m - VEA + \Delta\Phi$$
$$\phi_h = -\psi_m + VIP - \Delta\Phi$$

Determined by DFT

(c)

(d)

Thermal conductivity $k = -(J / dT / dx)$

(e)

FIGURE 2 (a) The schematic of high-throughput computations and computation methods for various dielectric properties. (b) The schematic of density functional theory (DFT) to calculate the charge injection barrier from electrode to polymer [43]. Reprinted with permission from Kamal et al. [43]. Copyright 2020, American Chemical Society. (c) The schematic of DFT to calculate trap depth in polymer [44]. Reprinted with permission from Wang et al. [44]. Copyright 2019, IEEE. (d) Non-equilibrium molecular dynamics to calculate the thermal conductivity of polymer chain [45]. Reprinted with permission from Zhang et al. [45]. Copyright 2014, American Chemical Society. (e) Phase-filed model to simulate the breakdown process in polymer composites [46]. Reprinted with permission from Shen et al. [46]. Copyright 2017, John Wiley and Sons

correlated with the chemical structure (i.e. presence of chemical groups, end groups etc.) and the molecular weight distribution [64]. As a result, an 'atomic group contribution method' was schemed to estimate the polymer properties as a weighted sum of contributions from the constituting fragments [65].

Based on this, to accelerate the discovery of polymer dielectrics, Ramprasad et al. developed a more general and seamless pathway to represent all polymers [22, 29]. The polymers were fingerprinted by the occurrence of different types of building fragments (constituted by blocks such as
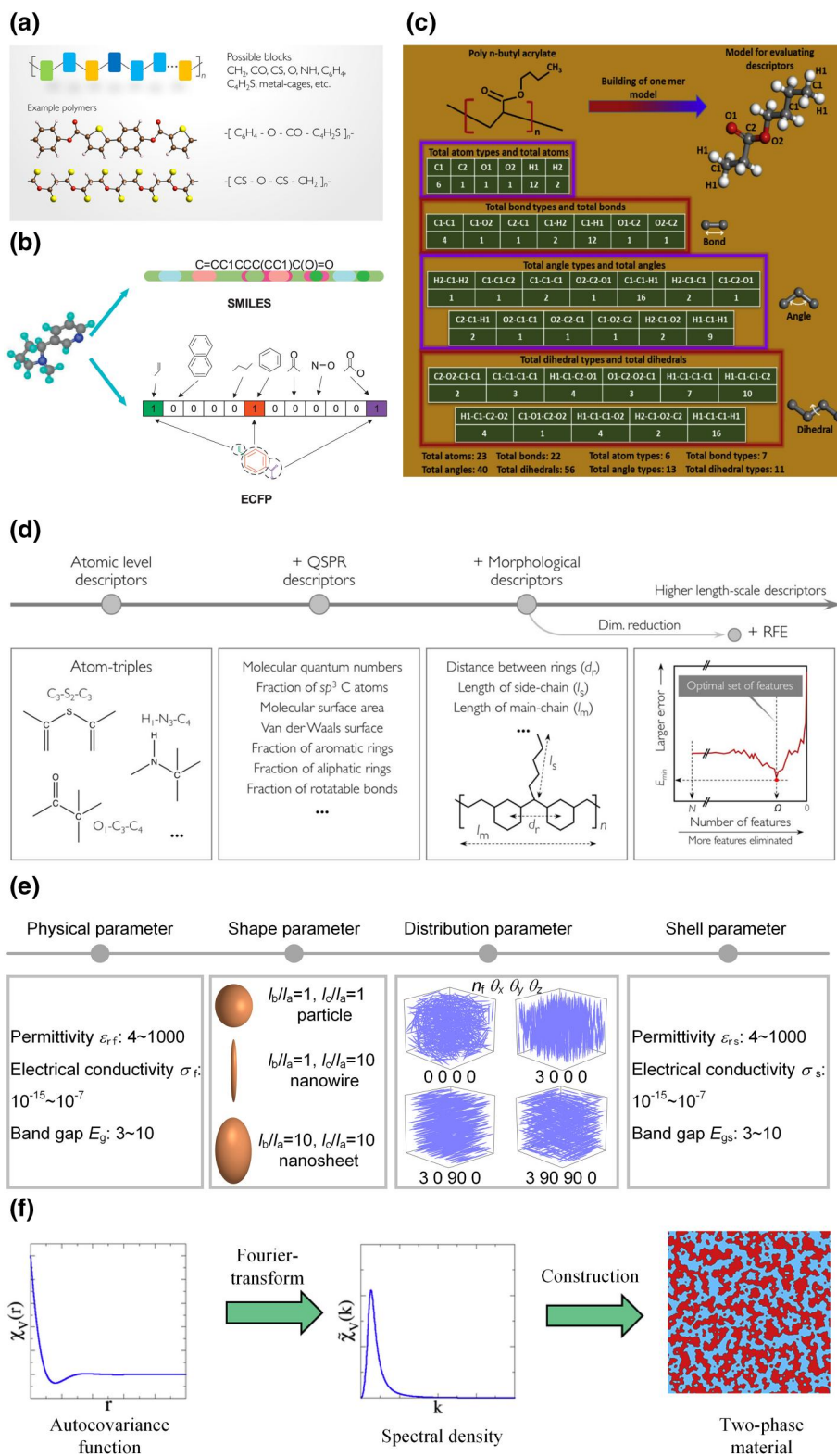
FIGURE 3 Different kinds of fingerprints for polymers and nanocomposites. (a) Fingerprints based on single, double, and triple components [29]. Reprinted with permission from Mannodi-Kanakkithodi et al. [29]. Copyright 2018, Elsevier. (b) Simplified Molecular-Input Line-Entry System (SMILES) and Extended-Connectivity Fingerprints (ECFPs) [25]. Reprinted with permission from Zhu et al. [25]. Copyright 2020, Elsevier. (c) and (d) Two kinds of hierarchical fingerprints [66, 67]. (c) Reprinted with permission from Bhowmik et al. [66]. Copyright 2021, Elsevier. (d) Reprinted with permission from Kim et al. [67]. Copyright 2018, American Chemical Society. (e) Fingerprints based on physical and geometrical parameters of nanocomposites [26]. Reprinted with permission from Zhu et al. [26]. Copyright 2021, American Chemical Society. (f) Fourier space form of spatial correlation functions to represent microstructure of composites [68]. Reprinted with permission from Chen and Torquato [68]. Copyright 2018, Elsevier

CH$_2$, CO, CS, O, NH, C$_6$H$_4$, and C$_4$H$_2$S) in terms of their number fractions [29], as shown in Figure 3a. A particular molecular fragment could be a singlet, doublet or triplet of contiguous blocks such as CH$_2$–NH pairs and –NH–CO–CH$_2$– triplets [22, 29]. After that, the linkage between fingerprints and properties (bandgap and dielectric constant) have been established using ML algorithms.

Line notations, encoding of the chemical structure into fixed-length strings or vectors, have been proved effective in describing molecules. Due to the human-readable and machine-friendly advantages, the SMILES has been extensively adopted to represent molecules [69–72]. To encode the linear notations into a machine-processable format, various molecular fingerprinting algorithms such as extended connectivity [73] have been presented to transform the SMILES into numerical vectors, which can be generated from the open-source RDKit software [74]. As depicted in Figure 3b, extended-connectivity fingerprints (ECFPs) use a list of binary identifiers to naturally and accurately represent the presence or absence of particular structures. We have used the SMILES and ECFP to build a structure–property relationship model, revealing key correlations between thermal conductivity and polymer structures [25]. It should be noted that the effect of polymer morphology on the target property is neglected in the SMILES. For this reason, hierarchical fingerprinting that captures atomistic to morphological structural information has been developed to improve the prediction accuracy. Figures 3c and d show the elements of two kinds of hierarchical fingerprinting [66, 67]. The descriptors in Figure 3c were constructed using the Materials Studio software of BIOVIA, which were divided into six groups, namely, atom type, molecular type, molecular attributes, atom name, bonding (bonds, angels and dihedral) and molecular weight [66]. The fingerprint in Figure 3d is composed of atom-level, block-level and chain-level descriptors [67].

3.1.2 | Nanocomposite fingerprints

Recently, by introducing inorganic nanofillers with specific characteristics, polymer nanocomposites have been extensively studied to obtain the desired target properties [9–11]. The properties of nanocomposites depend on various factors such as the physical parameters of nanofillers (permittivity, electrical conductivity, bandgap, thermal conductivity etc.), nanofiller–matrix interface characteristics (trap states, interfacial polarization, shell parameters in the core-shell structures etc.), and the geometric microstructure of nanofillers (filler shape, volume fraction, distribution, orientation etc.) [9–11, 26]. As a result, most of the existing fingerprints for nanocomposites are selected from these factors depending on the target property of interest [60, 75, 76], which is also called as physical-descriptor-based methodology [76]. The main advantage of physical descriptors is that they provide clear physical insights and offer meaningful mappings to processing parameters. For example, nanofiller features including the variables of morphology, permittivity, electrical conductivity and volume fraction are used to represent nanocomposites, then a ML strategy is schemed to evaluate the capability of energy storage [60]. To analyse the correlation between the filler doping scheme and the dielectric properties, we developed a fingerprint with a string of characters considering the physical parameters, shape, distribution of fillers, and shell properties in core–shell structures [26], as depicted in Figure 3e.

With the rapid development of convolutional neural networks (CNNs), which can directly extract the geometric features of composites, 2D cross-section images of 3D microstructures can be directly input into the CNN to predict various properties such as thermal conductivity [61, 77, 78]. In addition, some microstructure characterization and reconstruction (MCR) methods [79] such as spatial correlation functions (SCFs) [80, 81] and Fourier space representation of SCF (SDF) [68, 82] have received significant attention in representing the polymer composites, as shown in Figure 3f. The advantage of MCR is that the microstructure can be readily accomplished through a hierarchical reconstruction strategy [68, 80–82]. Although some techniques have been applied to characterize polymer nanocomposites, the underlying mechanisms behind the effects of nanofillers are not fully understood yet, for example, the effect of the nanofiller-matrix interface on the breakdown strength and dielectric constant. For this reason, more sophisticated fingerprints are expected to accurately predict specific properties in the future.

3.2 | ML algorithm

The ML algorithm aims to establish a mapping between the fingerprinted input and the target property, which provides an effective surrogate for estimating the target property. Currently, linear and non-linear regression algorithms have been applied to build the model, among which kernel-based regression and artificial neural networks (ANNs) are the most popular algorithms [83, 84]. A brief description of these algorithms is provided below.

The linear regression algorithm models the target property with a linear function of fingerprints, which is a simple method mainly used in some early works [64]. The prediction accuracy of these models is limited because they neglect the non-linear linkage between descriptors and properties. To solve that, the so-called kernel trick is often applied to realize complex regressions [32, 85]. Specifically, instead of using the original fingerprint x in linear regression, the data are first transformed into a higher-dimensional feature space by kernel functions such as radial basis function and polynomial kernels [32, 85]. This procedure results in a non-linear regression in the original feature space. Several kernel-based algorithms such as kernel ridge regression (KRR), support vector machine (SVM) and Gaussian process regression (GPR) have been employed [32, 85]. For instance, KRR has been applied to predict polymer properties (e.g. bandgap and dielectric constant) based on high-throughput DFT computations [22]. Yin et al. proposed an interval support vector regression with particle swarm optimization (PSO) to predict the dielectric constant for polymer

dielectrics at various frequencies and to accelerate the agile search for polymers with desirable properties [86]. In this work, the hyper-parameters in SVM are optimized by the PSO algorithm.

GPR is a well-established technique for building ML models for polymer dielectrics. It relies on the assumption that the unknown predicted function is sampled from multivariate Gaussian distribution and maintains a posterior distribution as observations are made [32, 33, 87]. The mean and variance for objective values can be well predicted based on the estimation of the covariance matrix, whose elements represent the covariance between two features [87], as shown in Figure 4a. The key advantage of GPR is that the uncertainty of the prediction can be provided, which is an essential ingredient of active learning for materials' design process. By using GPR to determine prediction value and uncertainty, an active-learning framework based on a balanced exploitation/exploration strategy was demonstrated to aid the discovery of polymers possessing high glass transition temperatures [88]. We have also adopted GPR to establish a linkage between the microstructure of nanocomposites and dielectric properties, and to discover doping schemes with desired properties [26]. A multi-fidelity information fusion approach based on GPR was proposed to

solve the problems where several datasets have varying levels of accuracy [89].

Decision-tree-based algorithms such as random forest (RF) are important non-linear ML methods. Random forest is an ensemble method of the decision tree, which combines multiple decision trees with a slightly randomized training process to improve performance [90–93]. The final regression or classification result is made by a weighted average or weighted vote of all forest predictions. More importantly, RF could provide an intrinsic metric to evaluate the importance of each descriptor, which is helpful for the rational design of polymers (see section 4.2). By using the RF as the classification method, Zhao et al. presented an automatic discovery method for polymers with low dielectric constants [92].

The ANN and deep neural network (CNN etc.) are well-developed ML algorithms that mimic the human brain to learn the linkages between certain descriptors and properties based on experience [94–101]. Figure 4b depicts the general architecture of the ANN, in which polymer fingerprints form the input layer [25]. The hidden layers build the relationships between the input and output layers, and the output layer represents the fitness of the candidate polymers. As for CNN, conventional and pool layers are added to extract key features from the input layer and
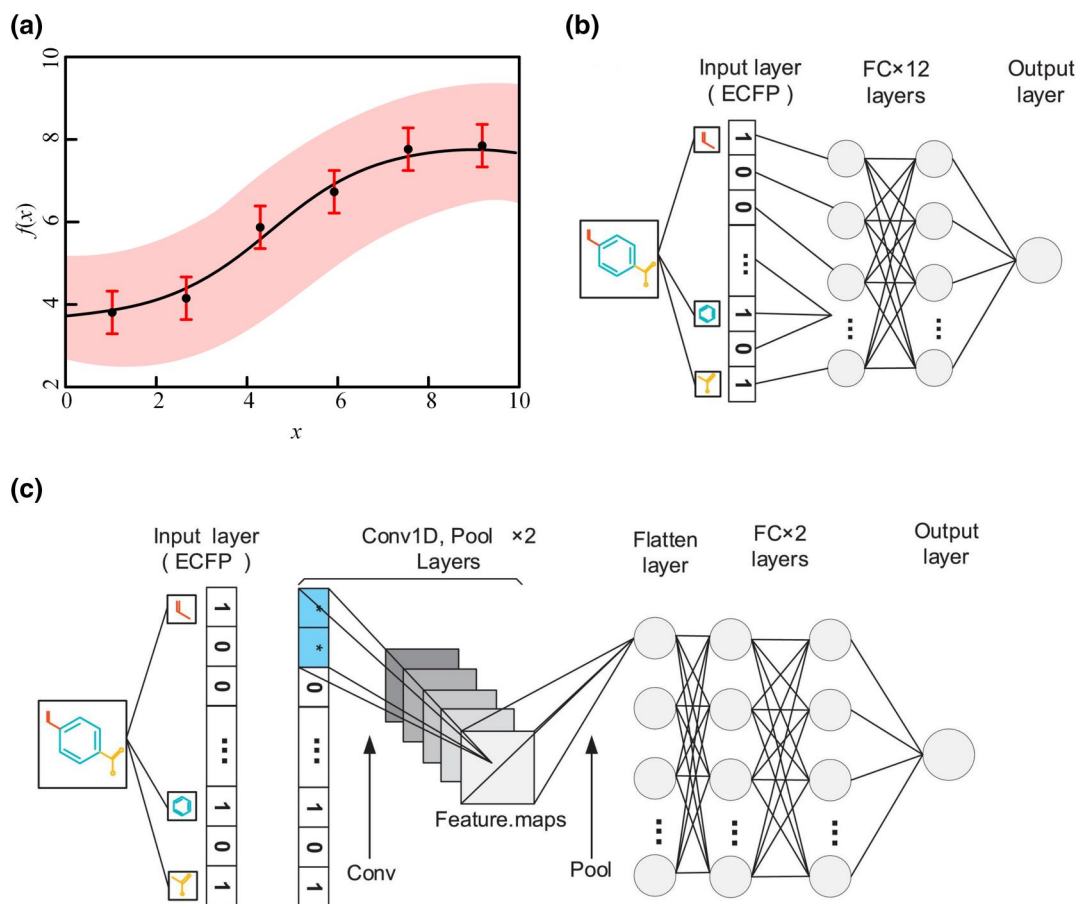

FIGURE 4 The schematic of (a) Gaussian process regression, (b) artificial neural networks and (c) convolutional neural networks [25]. Reprinted with permission from Zhu et al. [25]. Copyright 2020, Elsevier. ECFP, Extended-Connectivity Fingerprints

further reduce the dimension of the features [25, 97, 98], as shown in Figure 4c. Convolutional layers consist of a set of trainable filters, which are applied as discrete convolutions across the whole input, allowing the extraction of local features. Pooling layers further reduce the dimensionality of the features by combining subregions into a single output, among which max pooling is the most common method. Compared to the fully connected layers, the number of hidden neurons is drastically reduced by conventional and pool layers, thus allowing for far deeper networks. NNs have been applied to build ML models for the suitable selection of polymer–solvent pairs [99], glass transmission temperature [100] and thermal conductivity [25, 77, 95]. For the suitable selection of polymer–solvent pairs, a total of 11,958 polymer + good-solvent pairs and 8469 polymer + non-solvent pairs were employed to train a binary classification NN model to judge whether a solvent is good or insoluble for a specific polymer [99]. We used CNN, trained on thermal conductivities computed by classical MD simulations, to predict the thermal conductivity of polymers [25]. Wei et al. adopted 2D cross-section images of polymer composites as input to establish a CNN model, which have been demonstrated to predict effective thermal conductivity with high accuracy [77].

As described above, various ML algorithms are feasible for building a surrogate model between the fingerprinted input and the target property. These methods have their own advantages and disadvantages in terms of computational efficiency, the size of applicable dataset, and prediction capabilities. A brief comparison of these algorithms is provided in Table 1.

4 | INVERSE DESIGN OF POLYMER-BASED DIELECTRICS

Once the ML surrogate models are trained, they can be used to accelerate the discovery of polymer-based dielectrics by exploring the large space of potential materials, which can be called as inverse design process. Several methods such as enumeration, active learning, optimization algorithm and generative models have been utilized in the inverse design process. Moreover, by quantitatively analysing the structure–property relationships, variable importance measures could be determined to highlight the key features that play significant roles in target properties, which guide the rational design of polymer dielectrics.

4.1 | Inverse design methods

In the enumeration method, the target properties of a large pool of candidate materials are predicted with the previously trained ML model, followed by the selection and statistics of samples satisfying a certain screening criterion. Following this procedure, Mannodi-Kanakkithodi et al. screened a class of organic polymers involving several building blocks and discovered promising polymer dielectrics with desired bandgap and dielectric constant [12]. Additionally, we used a GPR-based ML model to screen promising polymer nanocomposites with desired permittivity, breakdown strength and energy density, resulting in several kinds of nanocomposites with desired properties [26], as shown in Figure 5a. These studies suggest the success of the enumeration method in searching for promising materials, which guarantees the diversity of the candidate materials. However, it is time-consuming to identify optimal polymer dielectrics by the enumeration method.

Active learning algorithms that iteratively fill the selected optimal point into training dataset were proven effective in materials design. As depicted in Figure 5b, active learning algorithms consist of three interwoven steps: (1) training the ML-based surrogate model for property prediction, (2) selecting the optimal sample based on the prediction results including values and uncertainties, and (3) supplementing the optimal sample into training dataset [88, 105]. The challenge in active learning is to

TABLE 1 Comparison of different ML algorithms

ML algorithm	Advantages	Disadvantages
Linear regression	Simplest method	Neglect of non-linear linkage between descriptors and properties
KRR, SVM	Low computational cost	Unfeasible for large datasets as the size of the kernel matrix scales quadratically with the number of features
GPR	The uncertainty for objective values can be well predicted	Requires a manageable dataset size and does not have the capability to train multiple properties in one single model
RF	Feasible for large datasets and provides an intrinsic metric to evaluate the importance of each descriptor	Might create over-complex trees and cause overfitting
ANN	Exhibits strong ability to capture non-linear complex relations from large-scale datasets	Requires much more training data, is time-consuming, and lacks interpretability; also called 'black boxes'.
Deep neural network	Feasible for graphical representations of materials and learns representations with different abstraction levels	Requires much more training data, is time-consuming, and lacks interpretability

Abbreviations: ANN, artificial neural network; GPR, Gaussian process regression; KRR, kernel ridge regression; ML, machine learning; RF, random forest; SVM, support vector machine.
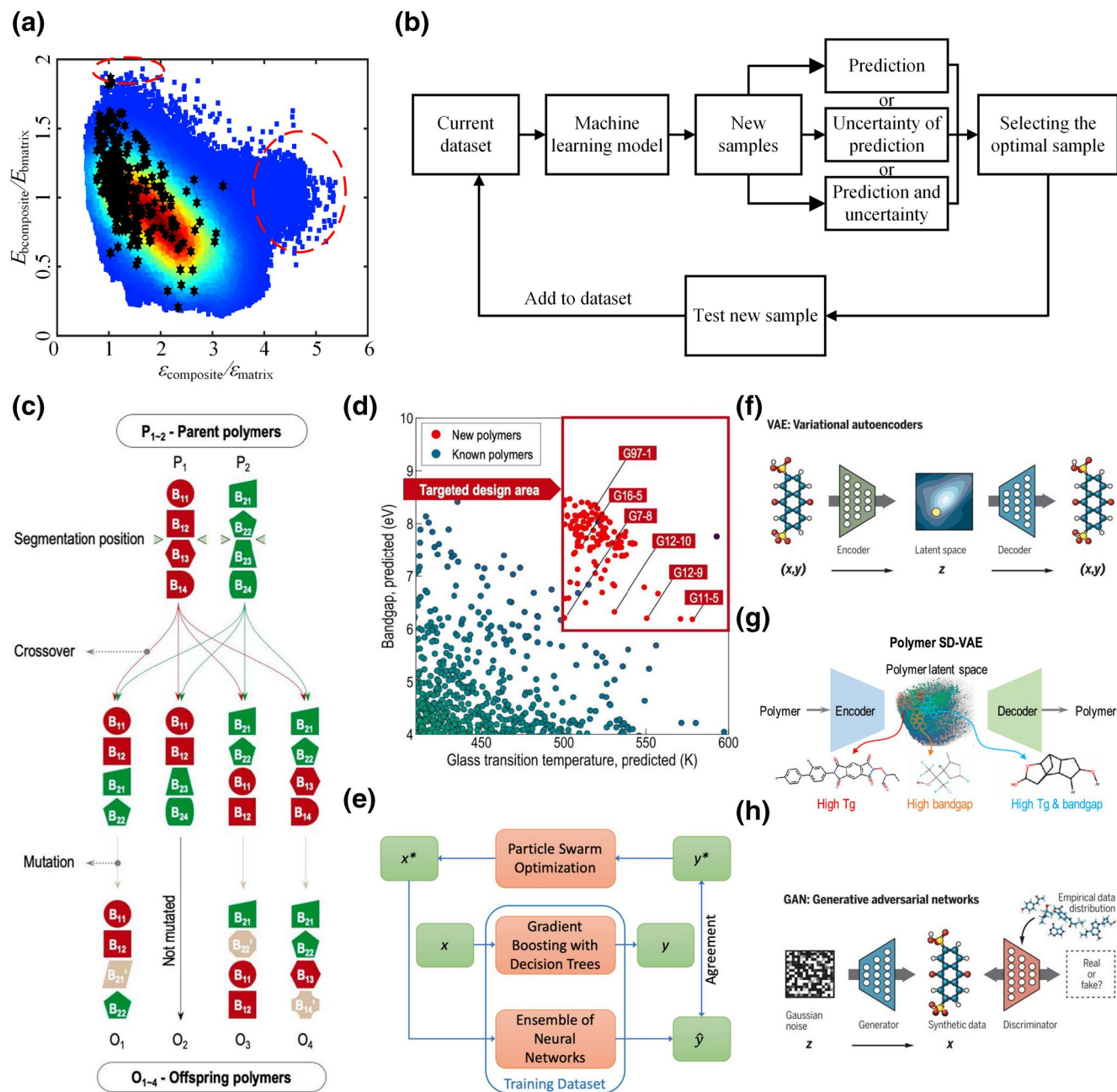
FIGURE 5 Inverse design methods for polymer-based dielectrics. (a) Enumeration method [26]. Reprinted with permission from Zhu et al. [26]. Copyright 2021, American Chemical Society. (b) Active learning algorithm. (c) and (d) Genetic algorithm method used to design polymers with high glass transition temperature and large bandgap [102]. Reprinted with permission from Kim et al. [102]. Copyright 2021, Elsevier. (e) Inverse design method based on particle swarm optimization [103]. (f) Variational autoencoder (VAE) [21]. Reprinted with permission from Sanchez-Lengeling and Aspuru-Guzik [21]. Copyright 2018, The American Association for the Advancement of Science. (g) VAE used to discover polymers with high T_g and bandgap [104]. Reprinted with permission from Batra et al. [104]. Copyright 2020, American Chemical Society. (h) Generative adversarial networks [21]. Reprinted with permission from Sanchez-Lengeling and Aspuru-Guzik [21]. Copyright 2018, The American Association for the Advancement of Science

balance the rapid convergence to optimal material (exploitation) with the need to traverse the material space to improve the model (exploration). Choosing the optimal sample requires ML models to provide both prediction and uncertainty values of the target property. As a result, the GPR algorithm and a combination of bootstrapping methods with standard ML algorithms (decision tree, SVM etc.), which can estimate the uncertainty of predictions, are common ML methods in active learning. Pure

exploitation, pure exploration and balanced exploitation and exploration principles have been utilized to select the optimal sample, as shown in Figure 5b. Kim et al. used active learning to discover polymers with $T_g > 450$ K, and they found that the balanced exploitation and exploration method had the best performance among the three principles [88].

Compared to the enumeration method, an optimization approach is preferable because it tests a smaller number of
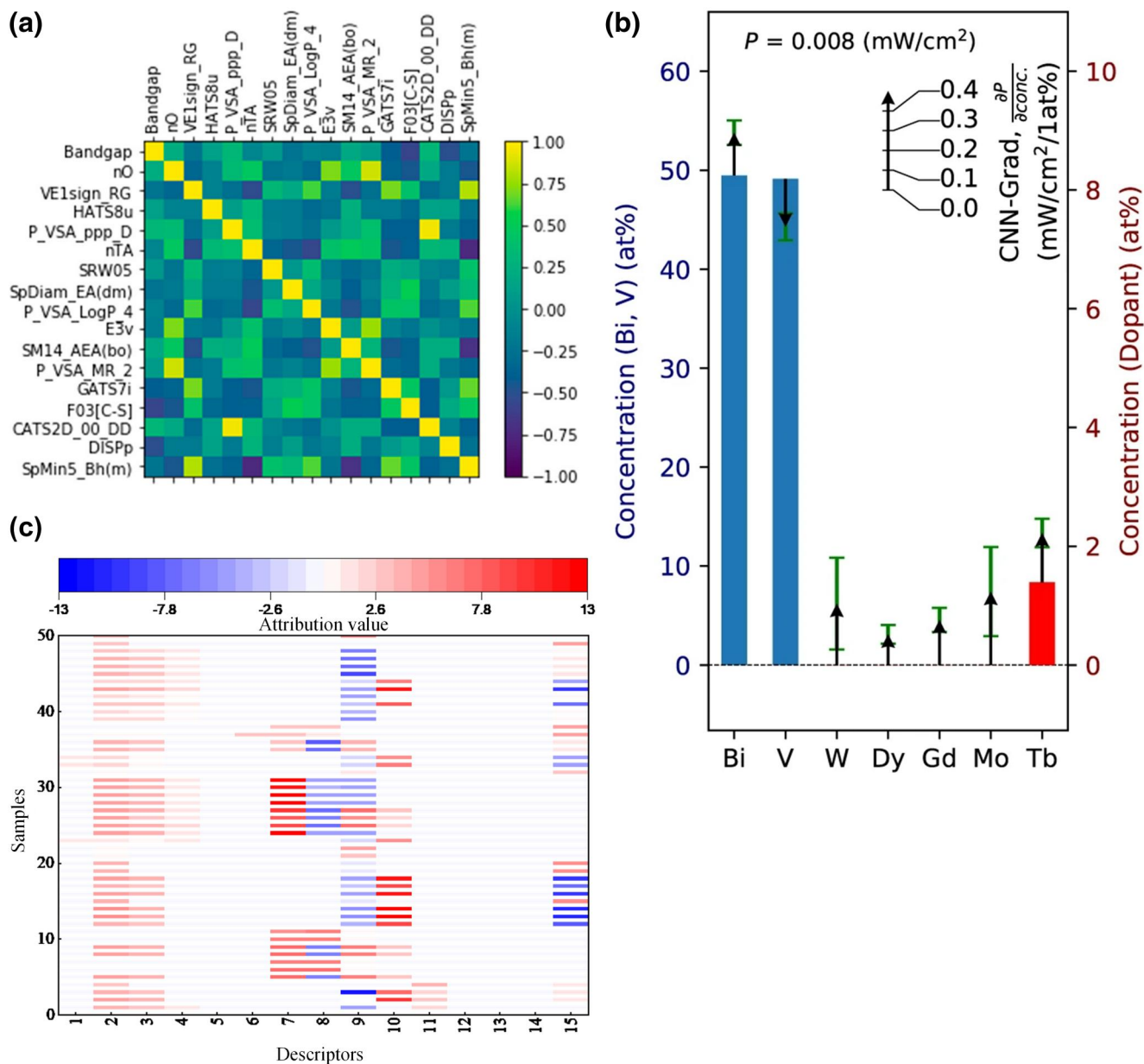
(a)

(b)

(c)

FIGURE 6   Methods for variable importance analysis. (a) Pearson correlation coefficients [118]. Reprinted with permission from Xu et al. [118]. Copyright 2021, American Chemical Society. (b) Gradients of convolutional neural networks (CNNs) model [119]. (c) Deep Learning Important FeaTures to investigate the contribution of different molecular structures to thermal conductivity of polymer chains

candidates when exploring the material space. ES algorithms such as the genetic algorithm (GA) and PSO are common choices for global optimization and have been used to search polymer space [102, 103, 106, 107]. ES completes a structured search through procedures inspired by natural evolution. At each iteration, parameter vectors ('genotypes', fingerprints in the ML) in a population are updated (selection, crossover and mutation in GA; movement of particle in PSO) to generate an offspring, followed by an evaluation of the objection function value. Kim et al. employed the GA method in tandem with ML models to design polymers with high glass transition temperature and large bandgap [102], as shown in Figure 5c, d. Additionally, Wei et al. demonstrated the application of GA to

discover unexpected thermal conductivity enhancement in disordered nanoporous graphene [107]. In another work, an inverse design by PSO and trained ML algorithm was demonstrated, predicting 17 polymer structures from user-defined cloud points [103], as shown in Figure 5e.

The generative model, stemming from the field of ML, is another approach for the inverse design of materials [21, 33]. Generative models learn to reproduce realistic samples from the distribution of data they are trained on. Several methods such as variational autoencoders (VAEs), generative adversarial networks (GANs) and recurrent neural networks are capable of building the generative models [104, 108–112]. In an autoencoder, the encoder learns to map the polymers to a lower-

T A B L E 2 Some examples of the ML-driven approach applied in designing polymers and nanocomposites

Target property	Data source	Fingerprint	ML model	Inverse design method	Reference
Polymers: Bandgap of the polymer and electron injection barrier (proxies for breakdown strength)	DFT computation	Hierarchical fingerprint in [53]	GPR	Enumeration	[53]
		SMILES in [43]			[43]
Polymers: Bandgap and dielectric constant (proxies for energy density)	DFT computation	Fingerprints based on singles, doubles and triples components	KRR	Enumeration	[22]
Polymers: Frequency-dependent dielectric constant	Experimental data in studies	Hierarchical fingerprint	GPR	Enumeration	[34]
Polymers: Dielectric constant	Experimental data in studies	Hierarchical fingerprint	Interval support vector regression	-	[86]
Polymers: Bandgap, glass transition temperature	Experimental data in studies	SMILES	GPR	GA in [102]	[102]
				VAE in [104]	[104]
Polymers: Glass transition temperature	Experimental data in studies	SMILES	GPR	Active learning	[88]
Polymers: Specific heat of polymers	Experimental data	Hierarchical fingerprint constructed using the Materials Studio software	Decision tree	-	[66]
Polymers: Thermal conductivity	MD simulations	SMILES	CNN	-	[25]
Polymers: Thermal conductivity	Online database	SMILES	Bayesian method	Enumeration	[39]
Nanocomposites: Breakdown strength, permittivity and energy density	Experimental data in studies	Descriptor-based fingerprint	GPR	Enumeration	[26]
Nanocomposites: Breakdown strength	Monte Carlo multi-scale simulation	MCR methods	GPR	GA	[79]
Nanocomposites: Energy density	Phase-field simulations	Descriptor-based fingerprint	NN	Enumeration	[60]
Nanocomposites: Thermal conductivity	FEM simulation	2D cross-sectional images	CNN	-	[61]

Abbreviations: CNN, convolutional neural network; DFT, density functional theory; FEM, finite-element model; GA, genetic algorithm; GPR, Gaussian process regression; KRR, kernel ridge regression; MCR, microstructure characterization and reconstruction MD, molecular dynamic; ML, machine learning; NN, neural network; SMILES, Simplified Molecular-Input Line-Entry System; VAE, variational autoencoder.

dimension space known as the latent space, while the decoder attempts to regain the original representation from the latent space [33], as shown in Figure 5f,g. The VAE achieves better generalizability (extended to uncovered spaces) by learning an approximation of the distribution of the input. For GAN in Figure 5h, the generator attempts to generate samples from a distribution, whereas the discriminator predicts whether the probability of a data is synthetic or real [33]. Although generative models have found successful application in molecule or drug discovery, they are just beginning to be utilized to find polymer dielectrics with desired properties. As an example, Batra et al. utilized a syntax-directed VAE in tandem with GPR models to discover polymers with high T_g and bandgap [104].

4.2 | Variable importance measures

Through a variable importance measure, the relevance of features with target properties can be evaluated, allowing a selection of the most important features and understanding of the ML model. Feature selection and dimensionality reduction algorithms, such as least absolute shrinkage and selection operator [113–115], sure independence screening and sparsifying operator (SISSO) [116], principal component analysis [117], are capable of selecting the best features. Additionally, Pearson correlation coefficients (PCCs) between various features and target properties can be calculated to represent the variable importance, in which PCC with −1 represents strong negative correlations and 1 represents strong positive correlations [118], as depicted in Figure 6a. Random forest and other decision-tree-based methods also possess the ability to evaluate the relevance of features through a variable importance measure [120].

Post-hoc interpretability of deep NN considers the possibility to identify composition–structure–property relationships, which aids human understanding of the data and provides guidance for rational design of polymer dielectrics [121]. For instance, the gradients of the CNN model quantify how the

predicted property varies with respect to each feature, can serve as a measure of the importance of input features and can be further analysed to interpret the data model [119]. Umehara et al. have used CNN gradient analysis to identify the composition–structure–property relationships for a series of $BiVO_4$ alloys [119], as shown in Figure 6b. Other methods or software packages such as Deep Learning Important FeaTures (DeepLIFT) [122], SHapley Additive exPlanations [123] and LIME [124] are available for NN interpretation. As an example, we have applied DeepLIFT to investigate the contribution of different molecular structures to thermal conductivity of polymer chains, as shown in Figure 6c.

5 | APPLICATIONS OF ML-DRIVEN DESIGN APPROACH

Polymer-based dielectrics have drawn a range of applications in various electrical and electronic devices. The polymer dielectrics should meet distinct properties such as the thermal and dielectric parameters in various applications, which can be improved by tailoring the chemical and morphological structure or adding nanofillers to form nanocomposites. The ML-driven approach has been proved an effective route to the rational design of polymer-based dielectrics. Table 2 summarizes some examples of the ML-driven approach applied in designing polymers and nanocomposites, in which the target properties, data sources, fingerprints, ML models and inverse design techniques are provided.

6 | CONCLUSIONS AND FUTURE PERSPECTIVES

The ML-driven approach trained on massive amounts of data has been proved to be a powerful technique for structure–property linkage analysis and the accelerating design of polymer-based dielectrics. The training data were mainly collected from online polymer libraries, experiments in the literature and high-throughput computations. Several kinds of fingerprints such as molecular weight distribution, occurrence of different types of building fragments, SMILES and hierarchical descriptors have been utilized to represent polymers, while the nanocomposites are generally described by physical descriptors and MCR methodology. Afterwards, different ML algorithms including linear-, kernel-based regression, non-linear regression and ANN have been exploited to establish a mapping between the fingerprinted input and the target property. Further, inverse design methods such as ES strategies and generative models can be employed to screen the large space of potential materials, which may greatly accelerate the discovery and development of new polymers dielectrics.

Although the material informatics techniques have achieved some success in discovering novel polymer dielectrics, there are some areas that should be addressed in the near future. At present, most ML-driven approaches are employed to design homopolymers. Further application to co-polymers, polymers with additives/nanocomposites has great practical significance. The prediction accuracy and generalization of ML models are strongly correlated with the quantity and quality of samples in the dataset, whereas these data are still limited for polymers and nanocomposites. This problem is expected to be solved by extracting scientific data untapped in numerous scientific journals with the ML-based NLP technique or developing advanced simulation methods such as the multi-scale modelling approach. With growing knowledge of the relationship between microstructures of polymers/nanocomposites and desired properties, other important descriptors such as the trap state (effects of chemical structures, additives, polymer-fillers interface etc.), morphologies (linear, cross-link, free volume etc.) and processing conditions should be incorporated into fingerprints to more accurately predict dielectric properties. Moreover, more advanced neural network algorithms (transfer learning, CNN etc.) and inverse design methods (VAE, GAN, and interpretable NN) can be applied for structure–property analysis, property prediction and polymer dielectrics' discovery.

CONFLICT OF INTEREST
None.

DATA AVAILABILITY STATEMENT
Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID
Ming-Xiao Zhu https://orcid.org/0000-0001-6154-7147

REFERENCES
1. Dang, Z.-M., et al.: Flexible nanodielectric materials with high permittivity for power energy storage. Adv. Mater. 25, 6334–6365 (2013)
2. Li, H., et al.: Dielectric polymers for high-temperature capacitive energy storage. Chem. Soc. Rev. 50, 6369–6400 (2021)
3. Li, Z., Du, B.: Polymeric insulation for high-voltage DC extruded cables: challenges and development directions. IEEE Electr. Insul. Mag. 34(6), 30–43 (2018)
4. Montanari, G.C., et al.: Next generation polymeric high voltage direct current cables—a quantum leap needed? IEEE Electr. Insul. Mag. 34(2), 24–31 (2018)
5. Burger, N., et al.: Review of thermal conductivity in composites: mechanisms, parameters and theory. Prog. Polym. Sci. 61, 1–28 (2016)
6. Huang, C., Qian, X., Yang, R.: Thermal conductivity of polymers and polymer nanocomposites. Mat. Sci. Eng. R Rep. 132, 1–22 (2018)
7. Yuan, C., et al.: Polymer/molecular semiconductor all-organic composites for high-temperature dielectric energy storage. Nat. Commun. 11, 3919 (2020)
8. Zhang, C., et al.: Grafting of antioxidant onto polyethylene to improve DC dielectric and thermal aging properties. IEEE Trans. Dielectr. Electr. Insul. 28(2), 541–549 (2021)

9. Tanaka, T., Montanari, G.C., Mulhaupt, R.: Polymer nanocomposites as dielectrics and electrical insulation-perspectives for processing technologies, material characterization and future applications. IEEE Trans. Dielectr. Electr. Insul. 11(5), 763–784 (2004)

10. Huang, X., et al.: High-*k* polymer nanocomposites with 1D filler for dielectric and energy storage applications. Prog. Polym. Sci. 100, 187–225 (2019)

11. Jiang, J., et al.: Polymer nanocomposites with interpenetrating gradient structure exhibiting ultrahigh discharge efficiency and energy density. Adv. Energy Mater. 9(15), 1803411 (2019)

12. Mannodi-Kanakkithodi, A., et al.: Rational co-design of polymer dielectrics for energy storage. Adv. Mater. 28(30), 6277–6291 (2016)

13. Le, T., et al.: Quantitative structure–property relationship modeling of diverse materials properties. Chem. Rev. 112, 2889–2919 (2012)

14. Liu, Y., et al.: Machine learning assisted materials design and discovery for rechargeable batteries. Energy Storage Mater. 31, 434–450 (2020)

15. Doan Tran, H., et al.: Machine-learning predictions of polymer properties with Polymer Genome. J. Appl. Phys. 128, 171104 (2020)

16. 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. Engineering. 5(6), 981–1192 (2019)

17. Butler, K.T., et al.: Machine learning for molecular and materials science. Nature. 559(7715), 547–555 (2018)

18. Kumar, J.N., Li, Q., Jun, Y.: Challenges and opportunities of polymer design with machine learning and high throughput experimentation. MRS Commun. 9(2), 537–544 (2019)

19. Sha, W., et al.: Machine learning in polymer informatics. InfoMat. 3(4), 353–361 (2021)

20. Jackson, N.E., Webb, M.A., Pablo, J.J.: Recent advances in machine learning towards multiscale soft materials design. Curr. Opin. Chem. Eng. 23, 106–114 (2019)

21. Sanchez-Lengeling, B., Aspuru-Guzik, A.: Inverse molecular design using machine learning: generative models for matter engineering. Science. 361, 360–365 (2018)

22. Mannodi-Kanakkithodi, A., et al.: Machine learning strategy for accelerated design of polymer dielectrics. Sci. Rep. 6, 20952 (2016)

23. Sharma, V., et al.: Rational design of all organic polymer dielectrics. Nat. Commun. 5, 4845 (2014)

24. Shen, Z.-H., et al.: Phase-field modeling and machine learning of electric-thermal-mechanical breakdown of polymer-based dielectrics. Nat. Commun. 10, 1843 (2019)

25. Zhu, M.-X., et al.: Machine-learning-driven discovery of polymers molecular structures with high thermal conductivity. Int. J. Heat Mass Tran. 162, 120381 (2020)

26. Zhu, M.-X., et al.: Rational design of high-energy-density polymer composites by machine learning approach. ACS Appl. Energy Mater. 4(2), 1449–1458 (2021)

27. Wan, X., et al.: Materials discovery and properties prediction in thermal transport via materials informatics: a mini-review. Nano Lett. 19(6), 3387–3395 (2019)

28. Balachandran, P.V.: Machine learning guided design of functional materials with targeted properties. Comput. Mater. Sci. 164(15), 82–90 (2019)

29. Mannodi-Kanakkithodi, A., et al.: Scoping the polymer genome: a roadmap for rational polymer dielectrics design and beyond. Mater. Today. 21(7), 785–796 (2018)

30. Meredig, B., et al.: Combinatorial screening for new materials in unconstrained composition space with machine learning. Phys. Rev. B. 89, 094104 (2014)

31. Ramprasad, R., et al.: Machine learning in materials informatics: recent applications and prospects. npj Comput. Mater. 3, 54 (2017)

32. Schmidt, J., et al.: Recent advances and applications of machine learning in solid-state materials science. npj Comput. Mater. 5, 83 (2019)

33. Chen, L., et al.: Polymer informatics: current status and critical next steps. Mat. Sci. Eng. R Rep. 144, 100595 (2021)

34. Chen, L., et al.: Frequency-dependent dielectric constant prediction of polymers using machine learning. npj Comput. Mater. 6, 61 (2020)

35. Otsuka, S., et al.: Polyinfo: polymer database for polymeric materials design. In: Fatos, X., Leonard, B., Nik, B. (eds.). International Conference on Emerging Intelligent Data and Web Technologies, Tirana, Albania, September 2011. pp. 22–29. IEEE, Los Alamitos (2011)

36. CROW. Polymer Properties Database [Internet]. http://www.polymerdatabase.com (2015). Accessed 10 Oct 2021

37. The Center for Hierarchical Materials Design (CHiMaD). Polymer Property Predictor and Database [Internet]. https://pppdb.uchicago.edu (2015). Accessed 10 Oct 2021

38. Huan, T.D., et al.: A polymer dataset for accelerated property prediction and design. Sci. Data. 3, 160012 (2016)

39. Wu, S., et al.: Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm. npj Comput. Mater. 5, 66 (2019)

40. Zhao, H., et al.: NanoMine schema: an extensible data representation for polymer nanocomposites. APL Mater. 6, 111108 (2018)

41. Tshitoyan, V., et al.: Unsupervised word embeddings capture latent knowledge from materials science literature. Nature. 571, 95–98 (2019)

42. Kim, E., et al.: Materials synthesis insights from scientific literature via text extraction and machine learning. Chem. Mater. 29, 9436–9444 (2017)

43. Kamal, D., et al.: Computable bulk and interfacial electronic structure features as proxies for dielectric breakdown of polymers. ACS Appl. Mater. Interfaces. 12(33), 37182–37187 (2020)

44. Wang, W., et al.: Interfacial charge dynamics of cross-linked polyethylene/ethylene-propylene-diene dual dielectric polymer as revealed by energy band structure. IEEE Trans. Dielectr. Electr. Insul. 26(6), 1755–1762 (2019)

45. Zhang, T., Wu, X., Luo, T.: Polymer nanofibers with outstanding thermal conductivity and thermal stability: fundamental linkage between molecular characteristics and macroscopic thermal properties. J. Phys. Chem. C. 118(36), 21148–21159 (2014)

46. Shen, Z.-H., et al.: High-throughput phase-field design of high-energy-density polymer nanocomposites. Adv. Mater. 30(2), 1704380 (2017)

47. Wang, C.C., et al.: Computational strategies for polymer dielectrics design. Polymer. 55, 979–988 (2014)

48. Zhu, J., et al.: Rational design of heat-resistant polymers with low curing energies by a materials genome approach. Chem. Mater. 32(11), 4527–4535 (2020)

49. Chen, X., et al.: First-principle investigation of the charge injection barriers of polyethylene and polytetrafluoroethylene oligomers. J. Appl. Phys. 126, 035101 (2019)

50. Sato, M., Kumada, A., Hidaka, K.: First-principles studies of carrier injection in polyethylene (PE) and ethylene-vinyl acetate copolymer (EVA) oligomers. IEEE Trans. Dielectr. Electr. Insul. 24(1), 574–582 (2017)

51. Meunier, M., Quirke, N., Aslanides, A.: Molecular modeling of electron traps in polymer insulators: chemical defects and impurities. J. Chem. Phys. 115, 2876–2881 (2001)

52. Takada, T., et al.: Determination of charge-trapping sites in saturated and aromatic polymers by quantum chemical calculation. IEEE Trans. Dielectr. Electr. Insul. 22(2), 1240–1249 (2015)

53. Kamal, D., et al.: Novel high voltage polymer insulators using computational and data-driven techniques. J. Chem. Phys. 154, 174906 (2021)

54. Wen, C., et al.: Determination of glass transition temperature of polyimides from atomistic molecular dynamics simulations and machine-learning algorithms. J. Polym. Sci. 58(11), 1521–1534 (2020)

55. Qin, H., et al.: Thermal transport in graphene-based layered materials: an analytical model validated with extensive molecular dynamics simulations. Carbon. 155, 114–121 (2019)

56. Shen, Z.-H., et al.: Phase-field model of electrothermal breakdown in flexible high-temperature nanocomposites under extreme conditions. Adv. Energy Mater. 8(20), 1800509 (2018)

57. Zhu, M.-X., et al.: Phase-field modeling of electric-thermal breakdown in polymers under alternating voltage. IEEE Trans. Dielectr. Electr. Insul. 27(4), 1128–1135 (2020)

58. Roy, S.L., et al.: Description of charge transport in polyethylene using a fluid model with a constant mobility: fitting model and experiments. J. Phys. D: Appl. Phys. 39, 1427–1436 (2006)

59. Chen, J., et al.: Space charge dynamics in double-layered insulation cable under polarity reversal voltage. IEEE Trans. Dielectr. Electr. Insul. 27(2), 622–630 (2020)

60. Shen, Z.-H., et al.: Designing polymer nanocomposites with high energy density using machine learning. npj Comput. Mater. 7, 110 (2021)

61. Rong, Q., et al.: Predicting the effective thermal conductivity of composites from cross sections images using deep learning methods. Compos. Sci. Technol. 184(10), 107861 (2019)

62. Li, S., Xie, D., Lei, Q.: Understanding insulation failure of nanodielectrics: tailoring carrier energy. High Volt. 5(6), 643–649 (2019)

63. Huang, Y., et al.: Predicting the breakdown strength and lifetime of nanocomposites using a multi-scale modeling approach. J. Appl. Phys. 122, 065101 (2017)

64. Van Krevelen, D., te Nijenhuis, K.: Properties of Polymers: Their Correlation With Chemical Structure; Their Numerical Estimation and Prediction From Additive Group Contributions. Elsevier Science (2009)

65. Bicerano, J.: Prediction of Polymer Properties. CRC Press (2002)

66. Bhowmik, R., et al.: Prediction of the specific heat of polymers from experimental data and machine learning methods. Polymer. 220, 123558 (2021)

67. Kim, C., et al.: Polymer genome: a data-powered polymer informatics platform for property predictions. J. Phys. Chem. C. 122, 17575–17585 (2018)

68. Chen, D., Torquato, S.: Designing disordered hyperuniform two-phase materials with novel physical properties. Acta Mater. 142, 152–161 (2018)

69. Weininger, D.: SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J. Chem. Inf. Comput. Sci. 28(1), 31–36 (1988)

70. Weininger, D., Weininger, A., Weininger, J.L.: SMILES. 2. Algorithm for generation of unique SMILES notation. J. Chem. Inf. Comput. Sci. 29, 97–101 (1989)

71. Lin, T.S., et al.: BigSMILES: a structurally-based line notation for describing macromolecules. ACS Cent. Sci. 5, 1523–1531 (2019)

72. Yan, C., et al.: Machine learning assisted discovery of new thermoset shape memory polymers based on a small training dataset. Polymer. 214, 123351 (2020)

73. Rogers, D., Hahn, M.: Extended-connectivity fingerprints. J. Chem. Inf. Model. 50(5), 742–754 (2010)

74. Rdkit: Open source toolkit for cheminformatics. https://www.rdkit.org/ (2012). Accessed 10 Oct 2021

75. Fetanat, M., et al.: Machine learning for advanced design of nanocomposite ultrafiltration membranes. Ind. Eng. Chem. Res. 60, 5236–5250 (2021)

76. Xu, H., et al.: A descriptor-based design methodology for developing heterogeneous microstructural materials system. J. Mech. Des. 136, 051007 (2014)

77. Wei, H., et al.: Predicting the effective thermal conductivities of composite materials and porous media by machine learning methods. Int. J. Heat Mass Tran. 127, 908–916 (2018)

78. Yang, Z., et al.: Deep learning approaches for mining structure-property linkages in high contrast composites from simulation datasets. Comput. Mater. Sci. 151, 278–287 (2018)

79. Schadler, L.S., et al.: A perspective on the data-driven design of polymer nanodielectrics. J. Phys. D: Appl. Phys. 53, 333001 (2020)

80. Jiao, Y., Stillinger, F.H., Torquato, S.: A superior descriptor of random textures and its predictive capacity. Proc. Natl. Acad. Sci. U S A. 106(42), 17634–17639 (2009)

81. Torquato, S., Beasley, J.D., Chiew, Y.C.: Two-point cluster function for continuum percolation. J. Chern. Phys. 88(10), 6540–6547 (1988)

82. Ghumman, U.F., et al.: A spectral density function approach for active layer design of organic photovoltaic cells. J. Mech. Des. 140, 111408 (2018)

83. Chen, A., Zhang, X., Zhou, Z.: Machine learning: accelerating materials development for energy storage and conversion. InfoMat. 2, 553–576 (2020)

84. Zhou, Z.: Journal of Materials Chemistry A and Materials Advances Editor's choice web collection: Machine learning for materials innovation. J. Mater. Chem. A. 9, 1295–1296 (2021)

85. Boser, B.E., Guyon, I.M., Vapnik, V.N.: A training algorithm for optimal margin classifiers. In: Haussler, D. (ed.) Proceedings of the Fifth Annual Workshop on Computational Learning Theory (COLT'92), Pittsburgh, July 1992. pp. 144–152. ACM Press, New York (1992)

86. Yi, Y., Wang, L.M., Yin, F.H.: Interval support vector regression enables high-throughput machine learning predictions for dielectric constant of polymer dielectrics. Appl. Phys. Lett. 118, 223901 (2021)

87. Deisenroth, M.P., Ng, J.W.: Distributed Gaussian processes. In: Bach, F., Blei, D. (eds.) ICML'15 Proceedings of the 32nd International Conference on Machine Learning Lille, July 2015 Vol. 37, pp. 1481–1490. ICML, Li (2015)

88. Kim, C., et al.: Active-learning and materials design: the example of high glass transition temperature polymers. MRS Commun. 9(3), 860–866 (2019)

89. Patra, A., et al.: A multi-fidelity information-fusion approach to machine learn and predict polymer bandgap. Comput. Mater. Sci. 171, 109286 (2020)

90. Quinlan, J.: Simplifying decision trees. Int. J. Man. Mach. Stud. 27, 221–234 (1987)

91. Breiman, L.: Random forests. Mach. Learn. 45, 5–32 (2001)

92. Liang, J., et al.: Machine-learning-assisted low dielectric constant polymer discovery. Mater. Chem. Front. 5, 3823–3829 (2021)

93. Mannodi-Kanakkithodi, A., Pilania, C., Ramprasad, R.: Critical assessment of regression-based machine learning methods for polymer dielectrics. Comput. Mater. Sci. 125, 123–135 (2016)

94. Goh, G.B., Hodas, N.O., Vishnu, A.: Deep learning for computational chemistry. J. Comput. Chem. 38(16), 1291–1307 (2017)

95. Yang, H., et al.: Machine learning and artificial neural network prediction of interfacial thermal resistance between graphene and hexagonal boron nitride. Nanoscale. 10, 19092–19099 (2018)

96. Nazarova, A.L., et al.: Dielectric polymer property prediction using recurrent neural networks with optimizations. J. Chem. Inf. Model. 61, 2175–2186 (2021)

97. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature. 521, 436–444 (2015)

98. Agrawal, A., Choudhary, A.: Deep materials informatics: applications of deep learning in materials science. MRS Commun. 9, 779–792 (2019)

99. Chandrasekaran, A., et al.: A deep learning solvent-selection paradigm powered by a massive solvent/nonsolvent database for polymers. Macromolecules. 53(12), 4764–4769 (2020)

100. Chen, G., Tao, L., Li, Y.: Predicting polymers' glass transition temperature by a chemical language processing model. Polymers. 13, 1898 (2021)

101. Dong, Y., et al.: Bandgap prediction by deep learning in configurationally hybridized graphene and boron nitride. npj Comput. Mater. 5, 26 (2019)

102. Kim, C., et al.: Polymer design using genetic algorithm and machine learning. Comput. Mater. Sci. 186, 110067 (2021)

103. Kumar, J.N., et al.: Machine learning enables polymer cloud-point engineering via inverse design. npj Comput. Mater. 5, 73 (2019)

104. Batra, R., et al.: Polymers for extreme conditions designed using syntax-directed variational autoencoders. Chem. Mater. 32, 10489–10500 (2020)

105. Balachandran, P.V., et al.: Adaptive strategies for materials design using uncertainties. Sci. Rep. 6, 19660 (2016)

106. Jennings, P.C., et al.: Genetic algorithms for computational materials discovery accelerated by machine learning. npj Comput. Mater. 5, 46 (2019)

107. Wei, H., Bao, H., Ruan, X.: Genetic algorithm-driven discovery of unexpected thermal conductivity enhancement by disorder. Nano Energy. 71, 104619 (2020)

108. Kingma, D.P., Welling, M.: Auto-encoding variational Bayes. arXiv:1312.6114v10 [stat.ML] (2013)

109. Kusner, M.J., Paige, B., Hernández-Lobato, J.M.: Grammar variational autoencoder. arXiv:1703.01925 [stat.ML] (2017)

110. Dai, H., et al.: Syntax-directed variational autoencoder for structured data. arXiv:1802.08786 [cs.LG] (2018)

111. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein GAN. arXiv:1701.07875 [stat.ML] (2017)

112. Hjelm, R.D., et al.: Boundary-seeking generative adversarial networks. arXiv:1702.08431 [stat.ML] (2017)

113. Santosa, F., Symes, W.W.: Linear inversion of band-limited reflection seismograms. SIAM J. Sci. Stat. Comp. 7, 1307–1330 (1986)

114. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B. 58, 267–288 (1996)

115. Kim, C., Pilania, G., Ramprasad, R.: From organized high-throughput data to phenomenological theory using machine learning: the example of dielectric breakdown. Chem. Mater. 28, 1304–1311 (2016)

116. Ouyang, R., et al.: SISSO: a compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. Phys. Rev. Mater. 2, 083802 (2018)

117. Matos, C.R.S., et al.: Principal component analysis of X-ray diffraction patterns to yield morphological classification of brucite particles. Anal. Chem. 79, 2091–2095 (2007)

118. Xu, P., et al.: Machine learning aided design of polymer with targeted band gap based on DFT computation. J. Phys. Chem. B. 125, 601–611 (2021)

119. Umehara, M., et al.: Analyzing machine learning models to accelerate generation of fundamental materials insights. npj Comput. Mater. 5, 34 (2019)

120. Stanev, V., et al.: Machine learning modeling of superconducting critical temperature. npj Comput. Mater. 4, 29 (2018)

121. Guidotti, R., et al.: A survey of methods for explaining black box models. arXiv:1802.01933v3 [cs.CY] (2018)

122. Shrikumar, A., Greenside, P., Kundaje, A.: Learning important features through propagating activation differences. arXiv:1704.02685v2 [cs.CV] (2019)

123. Lundberg, S.M., Lee, S.: A unified approach to interpreting model predictions. arXiv:1705.07874v2 [cs.AI] (2017)

124. Ribeiro, M.T., Singh, S., Guestrin, C.: Why should I trust you? Explaining the predictions of any classifier. arXiv:1602.04938v3 [cs.LG] (2016)