Please cite this article in press as: Yin et al., The data-intensive scientific revolution occurring where two-dimensional materials meet machine learning, Cell Reports Physical Science (2021), https://doi.org/10.1016/j.xcrp.2021.100482

Cell Reports
Physical Science

Review

The data-intensive scientific revolution occurring where two-dimensional materials meet machine learning

Hang Yin,[1,6] Zhehao Sun,[1,2,6] Zhuo Wang,[3,4,6] Dawei Tang,[2] Cheng Heng Pang,[3] Xuefeng Yu,[4] Amanda S. Barnard,[5,*] Haitao Zhao,[4,*] and Zongyou Yin[1,*]

SUMMARY

Machine learning (ML) has experienced rapid development in recent years and been widely applied to assist studies in various research areas. Two-dimensional (2D) materials, due to their unique chemical and physical properties, have been receiving increasing attention since the isolation of graphene. The combination of ML and 2D materials science has significantly accelerated the development of new functional 2D materials, and a timely review may inspire further ML-assisted 2D materials development. In this review, we provide a horizontal and vertical summary of the recent advances at the intersection of the fields of ML and 2D materials, discussing ML-assisted 2D materials preparation (design, discovery, and synthesis of 2D materials), atomistic structure analysis (structure identification and formation mechanism), and properties prediction (electronic properties, thermodynamic properties, mechanical properties, and other properties) and revealing their connections. Finally, we highlight current research challenges and provide insight into future research opportunities.

INTRODUCTION

Machine learning (ML) has been developing for decades. The fast development of new ML algorithms theories algorithms, data sources, and low-cost computation have accelerated the application of ML in various research areas.[1] Materials science, as a key subject that influences our lives, has also been influenced by development in ML.[2] The science of materials has experienced different stages, from the empirical paradigm, the model-based theoretical paradigm, to the computational third paradigm involving simulations. The data-driven fourth paradigm has brought new techniques and approaches capable of overcoming the limitations of its traditional predecessors.[3] Wider appreciation accompanied the announcement of the of materials genome initiative (MGI) in 2011,[4] a project aiming to integrate and share the data of materials for future manufacturing, supported by the rapid development of algorithms and the advancement of computational capability in the community. Although the quality of data still needs to be improved,[5] the integration of ML algorithms and big data have assisted scientists in revealing the hidden relationships among formation mechanisms, atomistic structures, and properties of materials.[3] In particular, two-dimensional (2D) materials, with unique physical and chemical properties, are widely studied for various reasons, each seeking a better understanding of a functional relationship.[6–8] The emerging superconductivity of 2D magic-angle superlattices[9] boosted the exploration of 2D materials and is an example of the relationship between a structural feature that can be controlled (the angle)

[1]Research School of Chemistry, Australian National University, Acton, ACT 2601, Australia

[2]School of Energy and Power Engineering, Dalian University of Technology, Dalian 116024, China

[3]Department of Chemical and Environmental Engineering, The University of Nottingham Ningbo China, Ningbo 315100, China

[4]Materials Interfaces Center, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, Guangdong, China

[5]School of Computing, Australian National University, Acton, ACT 2601, Australia

[6]These authors contributed equally

*Correspondence:
amanda.s.barnard@anu.edu.au (A.S.B.),
ht.zhao@siat.ac.cn (H.Z.),
zongyou.yin@anu.edu.au (Z.Y.)

https://doi.org/10.1016/j.xcrp.2021.100482

and a property that is highly desirable (the conductivity). However, the exploration of 2D materials is currently limited by the need for numerous tedious experiments, and researchers are looking to ML as a new strategy to accelerate their development.

ML contributes in many ways to materials exploration, from data collection (mining data from current references)[10] and data analytics (optimizing data for efficient analysis)[11,12] to analyzing results.[13] Several reviews have summarized the applications of ML in materials science, including the general progress of ML as applied to existing molecular and materials systems,[2] the design and discovery of new materials,[13,14] the development of new interatomic potentials,[15] and the use of these potentials in large-scale atomic simulations.[16] Recently Momeni et al.[17] summarized the calculation methods at various scales including atomistic scale, mesoscale, and macroscale, for the understanding and growth of 2D materials. This encompassed a detailed discussion of a variety of computational methods including density functional theory (DFT) and molecular dynamics (MD), mentioning the application of ML as one of the computational methods for 2D materials without elaboration. A more focused review of the applications of ML in 2D materials is therefore timely.

This review summarizes the recent advances of the applications of ML in 2D materials, including materials preparation, structure analysis, and property prediction. The review will begin with ML-guided 2D materials preparation, which includes the design and discovery of new 2D materials and ML-assisted 2D materials synthesis. ML-guided structure analysis will then be discussed based on structure identification and formation mechanisms. Subsequently, the advances of ML-assisted properties prediction and optimization will be summarized. Finally, a perspective on the current challenges and opportunities of this research area will be discussed.

ML-GUIDED MATERIALS PREPARATION

As mentioned above, ML can be a powerful and efficient tool to automatically extract information and knowledge from data and conduct complex and implicit tasks through the use of algorithms and sufficient computational power. ML has transformed research in some scientific areas such as biology, physics, medicine, and chemistry. In material science, a considerable amount of experimental and computational data has been generated, motivating researchers to explore data-driven science to capitalize on this investment. Since the single-layer graphene was first discovered by Novoselov, Geim, and their co-workers[18] in 2004, 2D materials have attracted tremendous interest from chemists, engineers, and material scientists, inspired by the potential applications in many fields ranging from semiconducting,[19] energy storage, sensing,[20] flexible electronics,[21] and catalysis.[22,23] Since there is a practical limitation to the fraction of all the possible 2D materials that a researcher can synthesize, the combination of the experimental and computational approaches, underpinned by the data handling ability of ML, could guide the design and preparation of 2D materials. This must be preceded by suitable knowledge extraction to identify meaningful patterns and relationships between chemical, mechanical, or topological features and functional properties as the target of materials design and discovery.

Design and discovery of new 2D materials with ML

2D materials have fascinating and unique chemical, physical, and electronic properties that are specific to their unique physicochemical structure and are promising to be applied in several different areas (as mentioned above). 2D materials include

graphene (including graphene oxides),[24,25] black phosphorus (BP),[26,27] hexagonal boron nitride (h-BN),[28] transition metal dichalcogenides (TMDs),[20,29] MXenes,[30] MBenes,[31] and silicene.[32] The time and cost involved in traditional trial-and-error methods are obstacles for large-scale design and discovery in each case, and are vulnerable to human biases that negatively impact ML models. Therefore, the ML-based data-driven workflow is promising to significantly promote efficiency and reduce the cost of design and discovery of 2D material. Reliable and sufficient data are the foundation of an ML process and 2D material databases. For example, the computational 2D materials database (C2DB)[33] and the joint automated repository for various integrated simulations using density functional theory (JARVIS-DFT),[34] which consists of $\sim 10^3$ 2D materials, have been constructed and are openly available, to facilitate more open research and collaboration. With the ability of ML to execute tasks and extract knowledge from available 2D material datasets such as these, it is anticipated that the design and discovery of 2D materials could be rationally guided by the revealed complex correlations and implicit data patterns. Under this subtitle, we focus on the process and application of ML in materials design and discovery, while particularly focusing on 2D materials, as summarized in Table 1.

Key elements to train ML for 2D materials design and discovery

Successful training of ML models is the most important part of ML-assisted design and discovery of 2D materials. Generally, the two most common types of problems addressed by ML in this domain are property prediction and pattern recognition. A simplified presentation of ML-participated case in material science is exhibited in Figure 1. As shown in Figure 1, given a target label, basic ML model training requires four key elements: the labeled data instances, the descriptor(s) of features, and the selection of algorithms and their hyper-parameters. The schematic to prepare and train an ML model in material science could be expressed as:[53,54]

$$Data(Target + Descriptors) + Algorithms + Hyperparameters = ML\ Model$$

where the target presents the labels that can be material properties or classes. Labels should be SMART (specific, measurable, attainable, relevant, and timely) targets.[53] Generally, a SMART target requires a comprehensive understanding of the phenomena and problems, which can guide the selection and transformation of data. For example, to study the stability of 2D materials, Siriwardane et al.[41] employed regression ML models where the formation energies were set as labels. The way a target is defined will determine the collection of data, the generation of descriptors, and the selection of ML algorithms, leading to different types of results.

ML is not a panacea, capable of solving all problems without expertise, and the form of input features and labels has an important influence on the outcome, as does the quality and quantity of data itself. Data must be cleaned and processed to be ML-ready.[5] The data collection could be automated by high-throughput computation or experiments, or fully autonomous, and can be in the form of tabular numerical data, images, signals, or text. Hundi and Shahsavari[39] used material microstructure presented as tabular numerical data to train deep-learning models (including convolutional neural networks [CNNs] and multilayer perceptrons [MLPs]) to accurately predict residual strengths of hexagonal boron nitride and graphene under different damage conditions of radiation and temperature. Lee et al.[37] trained a deep-learning model with the electron microscopy images of aberration-corrected scanning transmission to locate and classify point defects in a monolayer 2D TMD, $WSe_{2-2x}Te_{2x}$. Regardless of the source and form of data, the quantity of the data should be sufficiently large to cover the structure-property (or feature-label) space to be studied. The size of the dataset could influence the selection of ML models.

Table 1. The applications of data-intensive scientific revolution in 2D material discovery with machine learning

Target	Materials	Data source	Highly related descriptors	Predicted properties/ Model outputs	Algorithms	Achievements	Ref.
Materials discovery:	h-BN, AlN, GeS, MgI_2	Computational 2D Materials Database (C2DB)	*p* valence electron number	formation energy	deep neural network (DNN)	screened out 100 engineered point defects 2D material from 10,000 defect structures	35
Point defects in 2D materials			defect species chemical potential	band gap	transfer learning (TL)		
		Materials Project (MP)	electronegativity	Fermi energy	MEGNet (Graph Network)		
			homo energy		random forest (RF)		
			atomic weight				
			valence electron number				
			Van der Waals radius				
Properties prediction: Energy gap	graphene nanoflakes	self-generated	topological autocorrelation scores	energy gap	multiple linear regression (MLR) and binary decision tree (DT)	mapping energy gap and topological autocorrelation vectors for 622 graphene nanoflakes	36
					support vector machines (SVM)		
					artificial neural networks (ANN)		
					genetic algorithm (GA)		
Properties prediction: Strain fields	$WSe_{2-2x}Te_{2x}$ lattice	aberration-corrected Thermo Fisher Themis Z STEM, operated at 80 kV with probe current around 30 Pa	aberration-corrected STEM images	identification and classification of point defects in aberration-corrected annular dark-field (ADF) STEM images	fully convolutional networks (FCNs) with ResUNet architecture	generated high signal-to-noise class averages to measure 2D atomic spacings with up to 0.2 pm precision	37
Material discovery: Two-dimensional photovoltaic (2DPV) materials	Sb_2Se_2Te, Sb_2Te_3, and Bi_2Se_3	inorganic crystal structure database (ICSD)	packing factor (PF)	classification of PV/ non-PV materials	gradient boosting classifier (GBC)	26 2DPV candidates are successfully ruled out from 187,093 candidates	38
			Mulliken electronegativity		support vector machine (SVM)		
			Pauling electronegativity		random forest classifier (RFC)		
			cathode framework coordination		Ada boosting (Ada)		
			lattice parameter		stochastic gradient descent classifier (SGDC)		
			average number of valence electrons per atom count		decision tree classifier (DTC)		
			number of elements		logistic regression (LR)		
			sublattice neighbor count				
			bond ionicity of sublattice				
			average atomic volume				
			anion framework coordination				

(*Continued on next page*)

Table 1. *Continued*

Target	Materials	Data source	Highly related descriptors	Predicted properties/ Model outputs	Algorithms	Achievements	Ref.
Properties prediction: Residual strengths of various levels of radiation and temperature from the final atomic positions	h-BN, graphite	self-generated	voxelization hole score statistics atomic displacement statistic	residual strength of the processing condition	convolutional neural networks (CNNs) multilayer perceptrons (MLPs) transfer learning (TL)	trained two low computational cost models to predicted the structure and properties relations	39
Properties prediction: Thermodynamic stability	MoS_2, WS_2, MoSSe, WSSe, WsTe, GeTe, PbTe, Phosphorene, and $PtSe_2$	C2DB	prototype statistical functions SISSO generated features	classification of low, medium, or high thermostability convex hall energy formation energy	combination of stochastic gradient boosting (GB) decision tree classifier (DTC) sure independence screening and sparsifying operator (SISSO)	10 potential candidates were found of which two are yet unreported for photoelectrocatalytic (PEC) water splitting	40
Properties prediction: Formation energy	hexagonal MAB phases	MP	atomic volume number of P valence electrons ionization energies metallicity boiling temperature density geat molar capacity	formation energy	support vector machine (SVM) deep neural network (DNN) random forest regressor (RFR)	machine learning models are trained to predict the formation energies of the MAB phases to research their stability	41
Properties prediction: Semiconductor-metal transition	$MoS_{2-x}O_x$ bilayer	self-generated	average of the tangent of the Mo−S/O bond angle features based on the difference of oxygen fraction between two layers	formation energy Van der Waals (vdW) gap	random forest (RF)	discovery of two features that mainly characterize the SMT (concentration gradient of the chalcogen atom across the vdW gap and strain within the interface)	42

(*Continued on next page*)

Table 1. *Continued*

Target	Materials	Data source	Highly related descriptors	Predicted properties/ Model outputs	Algorithms	Achievements	Ref.
Properties prediction: Synthesizability of metal carbides and nitrides (MXenes) and their precursors	MXene	self-generated	formation energy	score of synthesizability	transductive bagging SVM	PU ML was adapted and applied to predict 2D materials synthesis	[43]
	MAX phases	C2DB	cohesive energy atom Bader charge mass per atom		k-Means Decision tree (DT) robust ensemble SVM (RESVM)		
Properties prediction: Exciton valley polarization landscape of 2D semiconductors	monolayer Tungsten diselenide (1L-WSe_2)	self-generated	position-dependent information extracted from its photoluminescence (PL) spectra at room temperature	low-temperature exciton valley polarization landscape	random forest (RF)	demonstration of representative empirical prediction of the low-temperature exciton valley polarization landscape of 1L-WSe_2	[44]
Properties prediction: Fracture strain, fracture strength, and Young's modulus	tungsten disulfide (WS_2)	self-generated	WS_2 type chirality temperature strain rate defect ratio	fracture strain, fracture strength, and Young's modulus	random forest (RF)	mechanical properties prediction of tungsten disulfide by application of random forest	[45]
Material discovery: Hierarchically structured allotropes of phosphorus	hierarchically structured allotropes of phosphorus	self-generated	smooth overlap of atomic positions (SOAP)	crystal structure	Gaussian approximation potential (GAP) driven random structure searching (GAP-RSS)	identification of a family of hierarchically structured allotropes based on a P8 cage as principal building unit	[46]
Property prediction: Critical superconducting temperature	25 unknown predicted superconductors were: 12 Cuprates, 7 iron-based crystals, and 6 others, with Tc ranging from ~32 K to ~138 K	inorganic crystal structure database (ICSD)	atomic mass average number of electrons average atomic magnetic moment atomic electronegativity	critical superconducting temperature	random forest regression (RFR), support vector regression (SVR), neural network regression	accuracy of critical superconducting temperature achieves over 92% confidence	[47]
Material discovery: Li-rich layered oxide cathode materials	Li-rich layered oxides compounds	self-generated	compositional eisorder for site synthesis data ionic radii electronegativity work function electron affinity	initial discharge capacity Coulombic efficiency capacity fade	support vector regression (SVR)	parameters for property enhancement and responsibility for initial discharge and Coulombic efficiency were proposed	[48]

(*Continued on next page*)

Table 1. *Continued*

Target	Materials	Data source	Highly related descriptors	Predicted properties/ Model outputs	Algorithms	Achievements	Ref.
Property prediction: Electron affinity, energy of the Fermi level, electronic band gap, ionization potential	graphene nanoflakes	self-generated	number of hydrogen atoms fraction of Zig-Zag edges (0 < FZZ < 1, where all other edges are armchair) surface area of graphene ratio between the two main structural axis average carbon atom coordination number average C-C bond length average C-C-C bond angle average C-H bond length average C-C-H bond angle	electron affinity energy of the Fermi level electronic band gap ionization potential	linear multiple regression (MLR) decision tree (DT) k-nearest neighbor (kNN) artificial neural networks (ANN) support vector machines (SVM)	electronic properties were predicted by using geometrical features with assistance of machine learning	49
Material discovery: Porous graphene with low thermal Conductivity	porous graphene	Molecular Dynamics (MD)	Hole density 2D distribution of holes	thermal conductivity	convolutional neural networks (CNN)	thermal conductivity were found efficiently and can inverse design optimal material by the revealed relation	50
Property prediction: Twist angle	twisted bilayer graphene (TBLG)	self-generated	Raman spectra	twist angle	kernel ridge regression (KRR) random forest regressor (RFR) artificial neural networks (ANN) principal-component analysis (PCA)	model of the one-to-one correspondence between given Raman spectrum and twist angle of TBLG is established	12
Property prediction: Band gap	functionalized MXene	self-generated	volume per atom standard deviations of the group number in periodic table standard deviation of melting point mean boiling point	band gap	kernel ridge regression (KRR) support vector regression (SVR) Gaussian process regression (GPR) bootstrap aggregating regression least absolute shrinkage and selection operator (LASSO)	trained ML model could bypass the band gap underestimation and avoid time-costly GW method	51
Material discovery: Efficient solar cell materials	CdTe, GaAs, $CuInSe_2$, $CuGaSe_2$, $ZnSnP_2$, $CdSnP_2$, $CH_3NH_3PbI_3$, etc.	JARVIS-DFT	classic force-field inspired descriptors (CFID)	spectroscopic limited maximum efficiency (SLME)	gradient boosting decision tree (GBDT) decision tree (DT) random forest (RF) k-nearest neighbor (kNN) multilayer perceptron (MLP)	general framework for the design of optimal solar cell materials with high efficiency was provided	52

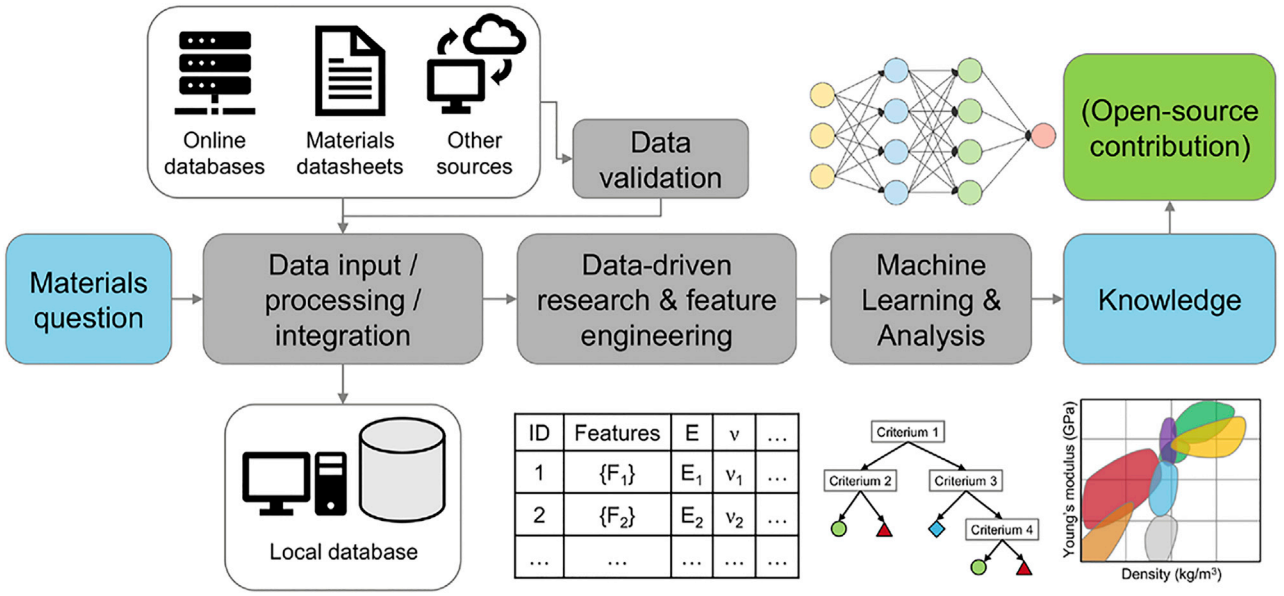

Figure 1. A general process to employ ML in materials science

A typical ML process includes data processing, feature engineering, ML models fitting, evaluation of model performance, comparison of performance across models, and visualization of results.[55] Copyright 2020, ACS Publications.

For instance, neural networks typically require larger dataset than classic statistical ML models to be feasible and useful.[55] There are techniques such as transfer learning and data augmentation that could be applied if the dataset size is small. For example, Frey et al.[35] used transfer learning to train a deep neural network to predict vital properties of promising defect/host systems in 2D materials. Compared with alternative studies, only 10% of the data typically required, around $\sim 10^3$ 2D structures, was needed to train the model. In terms of data quality, data biases and the type of distribution are critical issues that need to be addressed. Bias could come from experiments and computation errors[56] or from poor choices made by researchers.[57] An imbalanced dataset would lead to poor performance of ML models. To improve data quality, processing can be undertaken to reduce imbalance, change or normalize the distribution, and deal with systematic bias, but little can be done to expand the feature space once data collection and feature extraction has concluded.

The process of information gathering is called feature extraction, and the relevant characterization information is known as the feature set. Descriptors, which are often confused with features, are actually vectors of features with a common source. There may be dozens or even hundreds of features in a single descriptor, and multiple descriptors required to comprehensively represent a material. Features can either be measured characteristics of the molecules or materials (either numerical, images or spectra) or theoretical values calculated using formulas or algorithms.[58] There are some universal prerequisites for an ideal feature or descriptor suitable for ML. For example, descriptors should be unique and complete (in terms of features) to avoid evaluation bias. Descriptors should be capable of discrimination.[53] Features within descriptors should be differentiable and invariant to transformations. Once the features and descriptors are characterized or computed, sparsity and missing data (resulting in NaNs) must be corrected using feature selection or imputation. Care must be taken to ensure the number of features does not exceed the number of training instances. The feature space can be reduced using feature selection,

feature engineering, and dimension reduction methods if the set suffers from the curse of dimensionality.

Another critical fundamental stage of ML workflow is the algorithm selection, which is informed by the properties of the dataset (size, dimensionality, sparsity, imbalance) and the type of label (continuous or discrete). There is no one superior ML algorithm that is suitable for all problems, and every algorithm has its own advantages and disadvantages. Generally, there are three main categories of ML algorithms: supervised learning, unsupervised learning, and reinforcement learning. The most widely applied ML algorithm category in 2D material informatics is supervised learning, which uses labeled data to establish a mapping between chemical space and functional space. Supervised learning includes regression and classification, which predict continuous and discrete labels, respectively. In ML-assisted 2D material design and discovery processes, supervised learning is mainly employed for the property prediction and the exploration relationships between physicochemical structures and target properties.[12,35,36,38–45,47–49,51] Besides, supervised learning could also be used as a component in automated laboratories and can predict reaction conditions and process parameters.[59]

Most ML conducted outside of the scientific domain is unsupervised. Unsupervised learning is used for pattern recognition and does not require labeled data; the feature space will suffice. Two unsupervised learning techniques of interest to the 2D materials community are clustering and dimensionality reduction. Clustering groups the data based on similarity in the high dimensional feature space, and there are dozens of different clustering algorithms. The principal-component analysis (PCA) is a popular method for reducing the dimensionality of the feature space by transforming the features into principal components (that have no physical meaning). Manifold learning methods such as t-distributed stochastic neighbor embedding (t-SNE)[60] are applicable for visualizing high-dimensional data by providing each point corresponding coordination in a two or three-dimensional space.

Reinforcement learning (RL) is another subfield of ML, which mainly learns the relation between the action strategy and outcome to acquire the highest rewards to guide and enhance the next move. There are a few of studies that apply RL in material science. For example, Putin et al.[61] presented a framework that combines a generative adversarial network (GAN) and reinforcement learning trained on simplified molecular-input line-entry system (SMILES) string representation to generate the structure of small-molecule organic structures with target properties. Similarly, with the use of SMILE string, Popova et al.[62] employed RL to design molecules with desired properties. Furthermore, Zhou et al.[63] applied deep reinforcement learning trained on the descriptors selected by LASSO to optimize chemical reaction conditions.

Forward and inverse design of 2D materials

The conventional forward design of materials typically starts from a chemical space with features such as atomic identities, composition, and structure (ACS)[14] and ends at the functional space with the predictions of the material properties. The forward design has been widely successful but is restricted by prior knowledge. There could be many unreported compounds being omitted from the chemical space that might actually have the desired properties. Almost no datasets are truly exhaustive, and this introduces selection bias, the impact of which depends on how well (or how poorly) the configuration space has been sampled. Zunger[14] and Schleder et al.[53] have suggested that there are three types of forward design: (1) descriptive

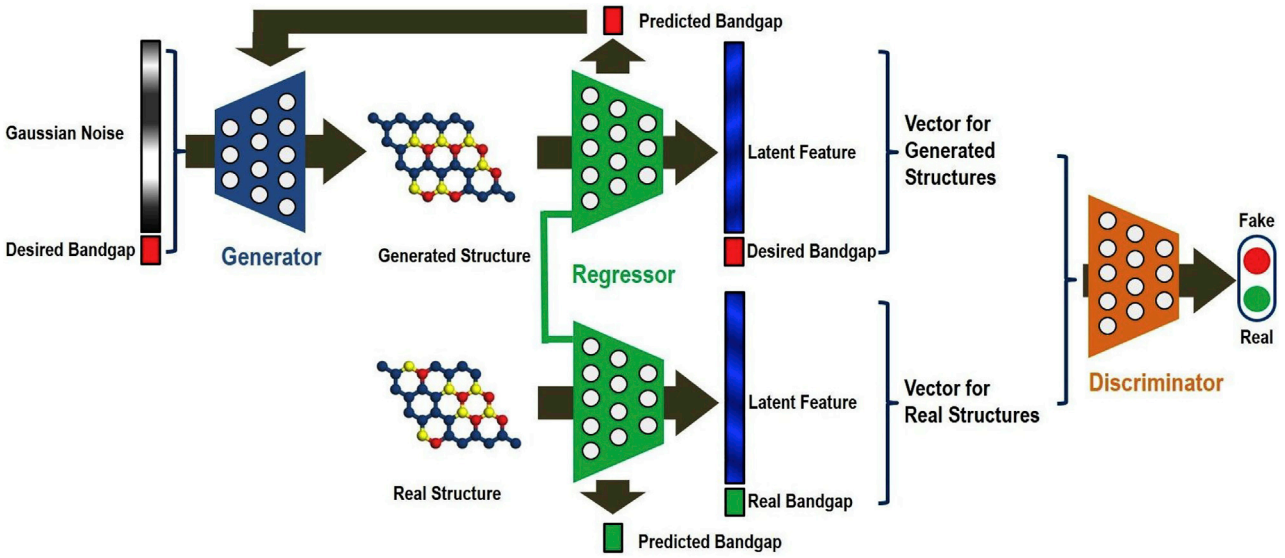

Figure 2. Schematic illustration of the application of GAN in 2D materials

The flow diagram demonstrates the architecture of applying GAN (RCGAN) for inverse structural design and discovery of 2D graphene/BN hybrids.[65] Copyright 2020, Elsevier.

methods, which are generally used for the interpretation and confirmation of experiment observations; (2) predictive methods, which are used for the prediction of new materials or properties; and (3) predictive methods for the design and discovery by predicting the function space for a class of materials.

In recent years, inverse design, which attempts to discover materials starting from a specific desired functionality, has been receiving attention. One approach is to comprehensively sampling the chemical space of material candidates with target properties with high-throughput computation or experiments and then select the compounds that have the best performance against some selection criteria. Zunger[14] has concluded that there are three modalities of inverse design: (1) with the implementation of optimization and search algorithms such as evolutionary algorithms and generative models to search the functional space, the ACS of material that has the optimal value of desired properties could be determined; (2) with the setting up of several filters based on some specific design principle and key properties, hierarchical virtual screening process could be applied to indicate the optimal material on a certain chemical space; (3) is slightly different from modality (2), in that modality three is to discover potential unknown compounds with optimal properties, so after the screening is completed, it is necessary to perform high-throughput calculations on the convex hull to identify the thermal stability of unknown materials. Although there are different approaches to inverse design, deep generative models have been most widely applied in this field,[64] using algorithms such as GANs, variational autoencoders (VAEs). and RL. For example, Dong et al.[65] present an inverse design framework (Figure 2) by using a regressional and conditional generative adversarial networks (RCGAN) to generate structures of 2D graphene/hexagonal boron nitride hybrids with specific bangap values.

Materials synthesis automation with ML

The possibility of ML-assisted chemical synthesis has drawn significant attention from the scientific community.[59] In particular, autonomous laboratories are becoming more widespread as next-generation facilities integrate an automated

Please cite this article in press as: Yin et al., The data-intensive scientific revolution occurring where two-dimensional materials meet machine learning, Cell Reports Physical Science (2021), https://doi.org/10.1016/j.xcrp.2021.100482

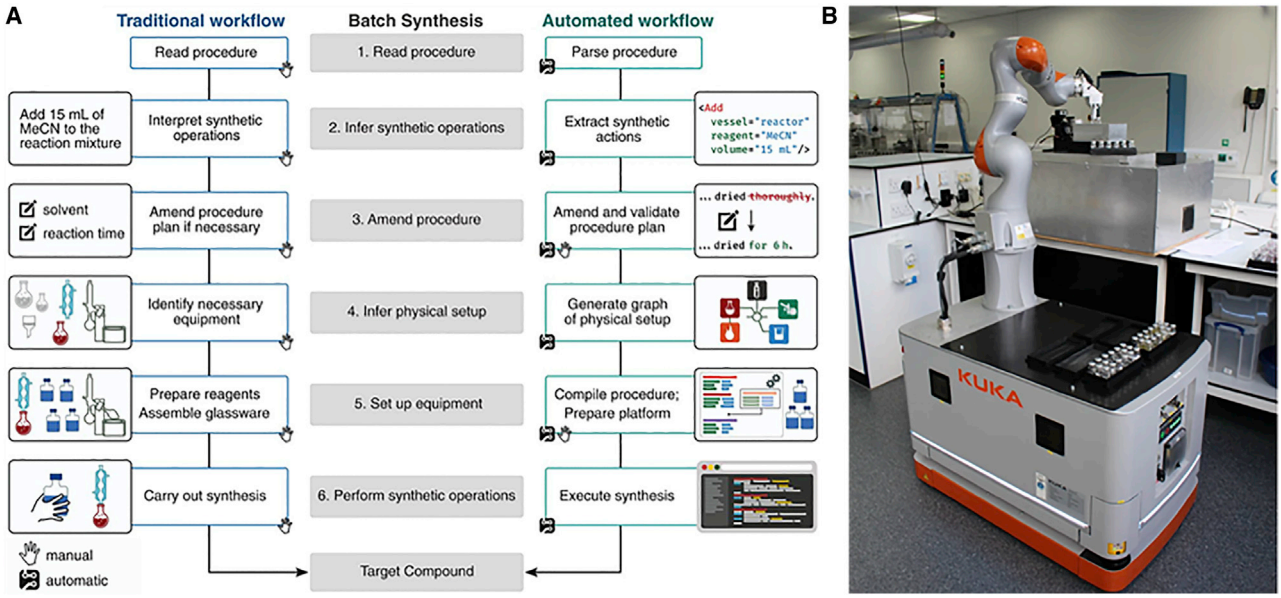

Figure 3. Demonstration of materials synthesis automation with ML

(A) Flowchart of both traditional and automated chemical synthesis execution process.[71] Copyright 2020, AAAS.

(B) The mobile chemist was transforming the samples to the photolysis station.[59] Copyright 2020, Springer Nature.

experimentation platform and artificial intelligence, leveraging ML methods to enable rapid self-guided experimentation.[66] Notably, ML models aimed at finding correlations between macroscopic variables such as the operating conditions, the reaction conditions, the plant operation parameter, and the yields of the products are enabling better prediction of process/structure relationships.[67,68] This has been demonstrated in synthesis planning[69,70] and reaction optimization.[63] Ding et al.[69] developed a data-intensive ML workflow to understand process dynamics and guide the atomic layer deposition of SiO_2 thin films. Based on the assumption that a well-trained ML model could conduct experiments without the supervision of humans, Mehr et al.[71] have proposed a universal system (Figure 3A) that could be used for the automatic synthesis of chemicals from the literature. Moreover, Burger et al.[59] employed a mobile robot (Figure 3B) performing 688 experiments autonomously for searching photocatalysts driven by a Bayesian search algorithm, which gave insights into using ML techniques in the automatic synthesis of 2D materials. It is expected that the integration of ML and automated experimentation platforms will substantially improve the rate of scientific discovery and provide more opportunities 2D material synthesis using more productive and innovative methods.

ML IN STRUCTURE ANALYSIS

Following preparation, materials characterization, including atomistic structure analysis, is the next step. The information on materials structure is fundamental to understanding the properties of the material and could be investigated by synergistic integration of experimental and theoretical approaches, e.g., transmission electron microscopy (TEM)[72] and Raman spectra[73] from the experiment, MD simulations,[74] and DFT[75] from theory, using appropriate data fusion methods. Combining ML with materials characterization has increased development and compensated for some of the shortcomings of these standalone approaches. ML-assisted structure identification and the understanding of the formation mechanisms will be discussed separately in this section.

Please cite this article in press as: Yin et al., The data-intensive scientific revolution occurring where two-dimensional materials meet machine learning, Cell Reports Physical Science (2021), https://doi.org/10.1016/j.xcrp.2021.100482

Table 2. The recent research on the integration of experimental techniques and ML for 2D materials characterization.

Techniques for characterization	Object	ML method	Image segmentation and classification	Ref.
Scanning transmission electron microscopy (STEM)	identify atomic species and type of defects	fully convolutional networks (FCNs)	pixel-wise classification	80
Scanning transmission electron microscopy (STEM)	locate and classify point defect	fully convolutional networks (FCNs)	pixel-wise classification	37
Scanning transmission electron microscopy (STEM)	identify defects	convolutional neural network (CNN)	pixel-wise classification	81
Transmission electron microscopy (TEM)	recognize the local atomic structure	convolutional neural network (CNN)	Meyer's algorithm	82
Atomic force microscopy (AFM) and Raman spectra	identify thicknesses, impurities, and stacking order	support vector machine (SVM)	red, green, and blue (RGB) channel intensities	83
Atomic force microscopy (AFM) and friction force microscopy (FFM)	identify the number of layers	spectral clustering (SClust)	spectral clustering algorithm	79

Structure identification

Structure identification, which is also called feature extraction (see above), is the basis for ML-assisted materials design. Structural information can be obtained from experiments or computer simulations and often requires manual analysis. However, with advances in computer vision, these repetitive tasks can be partially conducted automatically by an ML algorithm. Databases of images can directly provide quantitative predictions of complex correlations, especially for 2D materials at the nanoscale. Reducing the need for manual feature extraction also has the advantage of reducing evaluation bias, information bias in feature extraction, and confidence bias in labeling.

Among the existing techniques and tools, TEM[76] and electron diffraction, scanning transmission electron microscopy (STEM),[72] atomic and molecular resolved atomic force microscopy (AFM),[77] X-ray diffraction (XRD), field emission scanning electron microscopy (FESEM),[78] friction force microscopy (FFM),[79] and Raman spectroscopy[73] are typically used for characterization of the experimentally synthesized 2D structures, such as crystal size and morphology, phase, layer number, lattice distances, and facets from the micro scale to the atomic scale. However, these existing techniques still experience bottlenecks for 2D materials characterization and exploration, as well as introducing the risk of researcher errors. The studies in Table 2 combined ML and existing characterization techniques and tools to compensate for some of these bottlenecks. For instance, though the spatial resolution of STEM has reached down to the atomic level in 2D materials, the precision of atom-by-atom structural analysis has been limited to 8–20 pm, which is deficient in detecting local strain originated from substitutions and long-range strain fields from point defects.[72] For point defects, Ziatdinov et al.[80] developed a "weak supervision" method based on a deep neural network (DNN) that uses information about the coordinates of all atomic species in the image to identify multiple defects (Figure 4A). This method resembles the "human eye," which returns information about the atomic species and the location and type of defects on the surface of the sample in real time (Figure 4B). This work demonstrated an application of DNN to extract information from atomically resolved images including location of the atomic species and type of defects, which could be potentially refactored to overcome limitations in other 2D materials characterization.

The evaluation of defects of 2D materials also received attention from researchers. In 2020, Lee et al.[37] used ML to mine aberration-corrected STEM images to locate and classify point defects for 2D materials. The authors first applied fully convolutional networks (FCNs) to locate and classify each point defect in datasets of atomic-resolution images (Figures 5A and 5B) and then generated class-averaged images based

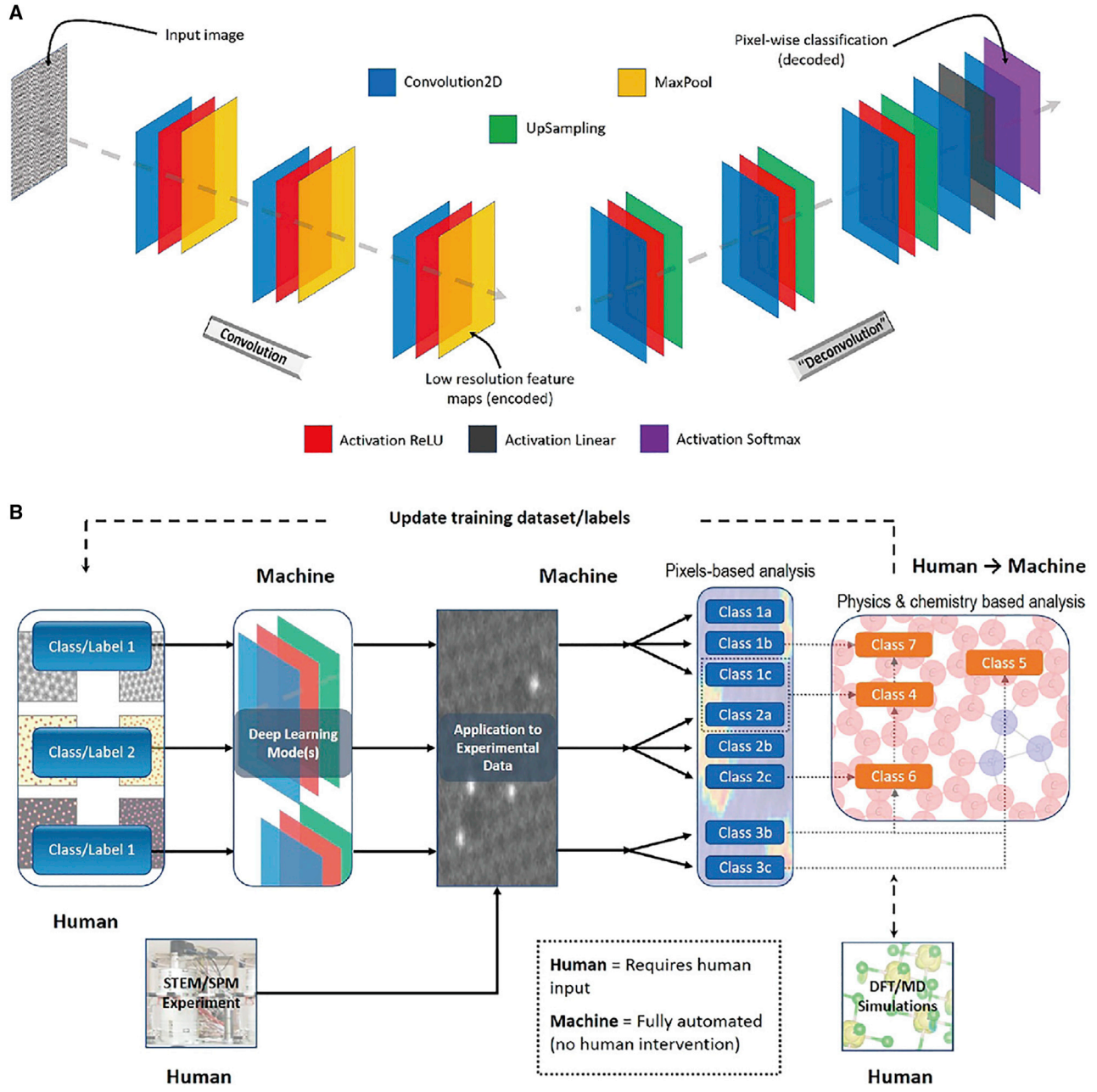

Figure 4. Applications of ML to a problem of finding atomic species and defects in a crystal lattice

(A) The schematic architecture of the fully convolutional networks has an encoder-decoder type of structure (or convolution-"deconvolution" structure). The final softmax layer outputs a pixel-wise classification for atomic species and/or defects.

(B) Schematics of a "weakly supervised" approach toward identifying lattice configurations and defects in the experimental data (see text for details). The parts of the process that require (do not require) human input is denoted as "Human" ("Machine").[80]

on the data to visualize the strain fields induced by single-atom defects in the 2D materials. The results show that Se vacancies introduce a complex, oscillating strain field in the $WSe_{2-2x}Te_{2x}$ lattice, which corresponds to the alternating ring of lattice expansion and contraction. At the same time, Maksov et al.[81] also introduced CNN into STEM to extract thousands of lattice defects from raw STEM data in a matter of seconds. Madsen et al.[82] performed CNN to recognize the local atomic

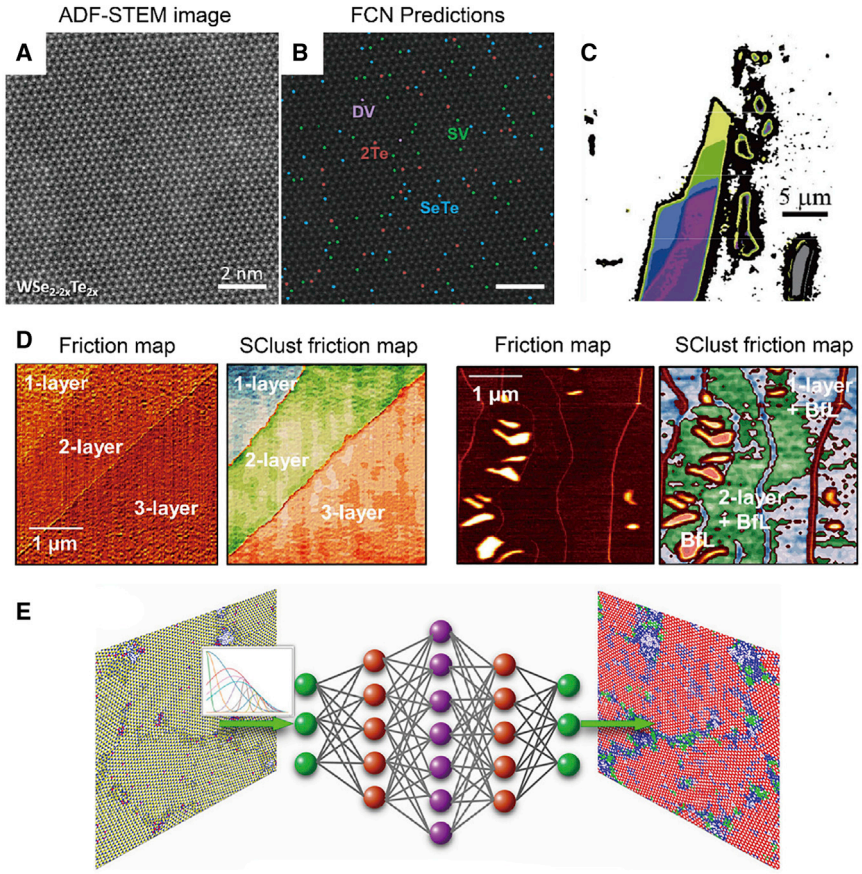

Figure 5. Applications of ML in structure identification of 2D materials

(A) Atomic-resolution ADF-STEM image of $WSe_{2-2x}Te_{2x}$. Scale bar = 2 nm.

(B) Chalcogen-site defects identified by fully convolutional networks (FCNs) overlaid on an image from (A). Scale bar = 2 nm.[37] Copyright 2020, ACS Publications.

(C) The test result of the sample, where regions of different layers have different colors. The as-identified regions of adhesive residues are blacked and the overexposed regions are grayed. Scale bar = 5 μm.[83] Copyright 2018, Springer Nature Publications.

(D) Friction (FFM) map and ML (SClust) friction map of epitaxial graphene film. Scale bar = 1 μm.[79] Copyright 2019, IOP Publications.

(E) Schematic of ML for defect identification in MoS_2 monolayer.[74] Copyright 2019, ACS Publications.

structure of single layers of defected graphene in TEM images. Since the assumption was that the sample will not have any strain, this situation is idealized and still needs to be validated for more 2D materials. These studies successfully took advantage of computer vision to mine atomic resolution datasets to assist in improving the measurement accuracy of STEM/TEM for electron beam sensitive materials.

The development of high-efficiency and large-area characterization technology has traditionally been a major obstacle in basic research and commercial applications of 2D nanostructures. When it comes to complex atomic construction, structural characterization faces an even greater challenge. Although 2D van der Waals (vdW) materials have abundant and unique functional properties,[84] characterizing the interface, surface doping, and counting the number of layers is still difficult. AFM with rational design and analysis can overcome this issue[85] but is still limited by an inability to observe the presence of surface adsorbates.[86] Lin et al.[83] introduced

the support vector machine (SVM) to optical identify 2D nanostructure, developing a method they referred to as ML optical identification (MOI). By identifying the characteristic color (red, green, and blue [RGB]) information in the optical photograph by using SVM, intelligent identification of individual 2D materials with different thicknesses, impurities, and stacking order can be realized (Figure 5C). The optical identification processes have also become more accurate. Using fast Fourier transform (FFT) on an images from electron and scanning probe microscopies is a typical example of image processing that can assist.[78] Since crystallographic domains have clearly different FFTs, Borodinov et al.[77] selected the absolute values of the FFT of the sliding window of AFM image as a feature for DNNs. After dimension reduction of the resulting dataset and extraction of abundance maps to populate the feature space, the DNNs-assigned the image pixels to relevant structure types and classified the pixels based on the values found in the maps. This work demonstrated the effective use of ML-assisted AFM to characterize and classify the crystallographic structure of 2D materials. Cellini et al.[79] performed detailed Å-indentation experiments and spectral clustering algorithm clustering to discover how the phase transition and interlayer elasticity of ultra-stiff diene-graphene affects the graphene-substrate interaction and the number of layers in epitaxial graphene grown on SiC and exfoliated graphene on SiO_2. With the help of spectral clustering algorithm, as shown in Figure 5D, this work realized layer number recognition and reconstruction of complex domains for graphene films, especially in areas with limited frictional contrast.

To move from successful synthesis of 2D materials to mass production, many synthesis methods have been improved and developed, including hydrothermal synthesis, mechanical exfoliation, physical vapor deposition (PVD), and chemical vapor deposition (CVD). However, optimal conditions for the synthesis of high-quality 2D materials are still a bottleneck. This is another area that ML can have an impact. For example, Hong et al.[74] performed multimillion-atom reactive molecular dynamics (RMD) simulations combined with a neural network to study a CVD-grown MoS_2 monolayer. By performing neural network-based analysis of the local atomic configuration at the origin of the grain growth, migration, and defect healing during the quenching-annealing steps, RMD simulations suggest that multiple heating-cooling treatments can contribute to the repair of grain boundaries for CVD-grown MoS_2 synthesis. The schematic for defect identification in MoS_2 monolayer is shown in Figure 5E. However, the approach may not be extensible. Low-symmetry configurations with complex features of 2D materials can *fatally* hamper DFT simulations of this type for some researchers due to the prohibitive computational cost involved, and ML could offer some promising strategies to overcome this challenge. Sheremetyeva et al.[12] compared kernel ridge regression (KRR), random forest regressor (RFR), and artificial neural networks (ANN)-based method to predict a continuous relationship between the twist angle and the stimulated Raman spectrum of twisted bilayer graphene. Different supervised regressors were used to predict the vibrational properties by extracting the structural information of the twist angle from Raman spectra supported by the semiempirical polarizability model. Unfortunately, the experimental Raman spectra of 2D materials are quite complicated, and some important features are missing in this work. A more reliable and accurate computational Raman spectra database needs to be established. In the case of experimental Raman spectra, Silva et al.[73] presented two complementary protocols using Raman spectroscopy developed by a neural network algorithm for the measurement of interlayer coupling and number of layers in graphene systems for statistical analysis, including the stacking order in mass-produced graphene. Iakovlev et al.[87] reported a neural network-based classification to identify the magnetic phase of a 2D ferromagnet

with Dzyaloshinskii-Moriya interaction within a certain magnetic field and temperature range. An advantage of neural network-based classification was presented by recognizing skyrmions of different types, and further introduced to experimental skyrmion magnetic configurations.

Structure-formation mechanism

Formation mechanisms that play a key role in synthesis involve physics and chemistry that are often poorly understood, creating obstacles for material structure prediction. ML methods can identify trends of the formation of structures based on historical data, but this is not the same as predicting synthesis routes. For the large-scale synthesis of uniform ultrathin metal sulfide nanocrystals, Du et al.[88] reported that the concentration of precursors, the composition of solvents, and the reaction temperature and time have coupled and complex relationship. This kind of relationship in the experiment is often difficult to merely rely on humans to judge the law and distinguish the relationship between the parameters and the structure formation. A better understanding of these complex relationships, with the help of experiments, simulation, and ML, will promote the development of high-efficiency and large-area 2D material synthesis technology, enabling mass production.

Structure-formation mechanisms are also a topic of theoretical investigation. Graphene oxide (GO), which is a hydrophilic 2D material based on graphene, contains O functional groups decorated in the sp^2 basal plane. Because the distribution of O functional groups largely determines the performance of GO, the complex relationship between formation conditions, concentration, and distribution of O present a challenge.[89] Motevalli et al.[75] used classification, regression, and causal inference to understand the relationship between the formation and concentration of defects and the concentration and distribution of oxygen groups. This work considered over 20,000 electronic structure simulations based on the density functional tight-binding (DFTB) method and extracted structural features in 220 dimensions. Primary causes, secondary causes, associatives and observations were used to assist in the layout and interpretation of the Bayesian network fitted to the 16 important features, as shown in the directed graph in Figure 6A and the self-organizing map (SOM), a type of unsupervised ANN also known as a Kohonen map, colored by defect concentrations, is shown in Figure 6B. The nanoflakes were classified using logistic regression to automatically separate the defective class, upon which ridge regression was used to identify the important features relating the defect concentration with the physicochemical structure. The possible causes of ruptured C−C bond defects in GO nanoflakes in this work were identified using Bayesian inference, and a probabilistic graph model was shown to provide to guide for the experimental synthesis of suitable GO nanoflakes by tuning the concentration of hydrogen.

Trends in the formation of different structures were identified by Siriwardane et al.[41] using SVM, DNN, and RFR models to predict formation energies of the MAB (M is the transition-metal element, A is one of group III-A or IV-A elements, and B is boron) phase. By using materials information on existing materials databases to train their model, they reported that the formation energy of MAB phases can be tuned by varying the A element, as opposed to M or B. This work inspired the full use of the existing materials database to mine the importance of available information and to reduce the estimated cost of experiments and simulations.

Phase transformations are closely related to the formation mechanisms. One of the TMDs, MoS_2, exhibits the unique physical, optical, and electrical properties correlated with its 2D ultrathin atomic layer structure.[29] 2H (semiconductor) and 1T (metallic)

Please cite this article in press as: Yin et al., The data-intensive scientific revolution occurring where two-dimensional materials meet machine learning, Cell Reports Physical Science (2021), https://doi.org/10.1016/j.xcrp.2021.100482

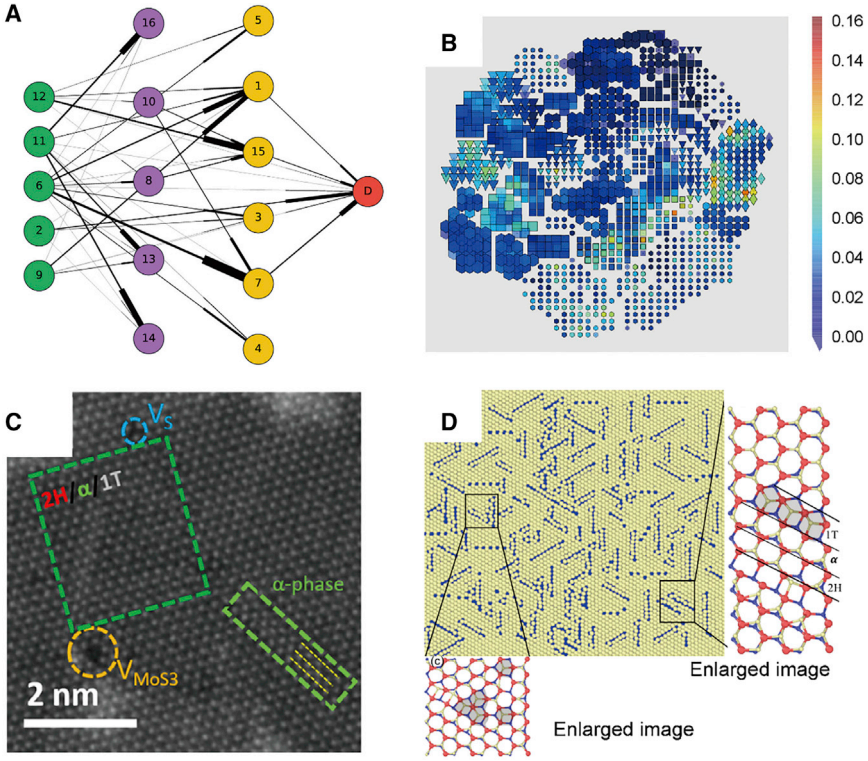

Figure 6. Applications of ML in structure-formation mechanism

(A) Directed graph of the Bayesian network causation model.

(B) Self-organizing map of the feature space colored by defect concentrations.[75] Copyright 2020, ACS Publications.

(C) Atomic-resolved high-angle annular dark-field scanning TEM image of monolayer MoS_2 shows various defect configurations. Scale bar = 2 nm.

(D) Atomic snapshot after 1 ns of MD simulation at 900 K. The top view of the sheet is shown where blue regions represent extended defects, and yellow beads are the S atoms in the top layer.[90] Copyright 2018, ACS Publications.

phases of MoS_2 show different properties, but the mechanisms underlying the 2H-to-1T transition during this phase transformation have not been fully investigated. This represents an ideal problem or simulation, but the main obstacle for using predictions from a computational simulation of the synthesis of 2D materials is the lack of connection between the simulation model and experimentally controllable parameters. In 2D material experiments and simulations, the spatial scale and timescale of the target system are often not fully accessible simultaneously, especially in the spatial distribution and dynamics of defects. Patra et al.[90] integrated the genetic algorithm (GA), *in situ* high-resolution transmission electron microscopy (HRTEM) and MD simulations to overcome this issue. A combination of GA and MD simulations was used to efficiently sample extended defect configurations (shown in Figures 6C and 6D) and HRTEM was used to verify the identification of stable structures throughout the evolutionary search. This work elucidates the defect aggregation and defect-driven phase transition mechanism in 2D TMD materials and also introduces a way to solve complex phenomena of 2D materials in multiple scales and across scales.

Numerous problems persist in understanding the formation mechanism for materials, even when utilizing ML. Bartel et al.[91] tested seven ML systems for formation energy on stability predictions, that is, ElFrac, Meredig, Magpie, AutoMat, ElemNet, Roost, and crystal graph convolutional neural network (CGCNN). It was reported that, even

Please cite this article in press as: Yin et al., The data-intensive scientific revolution occurring where two-dimensional materials meet machine learning, Cell Reports Physical Science (2021), https://doi.org/10.1016/j.xcrp.2021.100482

Table 3. Summary of ML-assisted study of 2D materials properties

Materials	Properties	Algorithms	Function of ML	Performance of ML	Ref.
Graphene	electronic properties	GA, SVM	predict the electronic properties from distribution of interatomic distances	accuracy >83% for energy of Fermi level and ionization potential; accuracy of ~80% for electron affinity; accuracy < 70% for energy of band gap	93
Graphene	electronic properties	MLR, DT, *k*NN, ANN, SVM	build QSPR model of each electronic properties with most relevant features	prediction of electronic properties: R^2 ~0.9	94
Graphene	electronic properties	MLR, DT, RF, SVM, ANN	predict the difference between the calculation results of DFTB and DFT.	accuracy of 94% for energy of Fermi level; accuracy of 88% for energy of band gap	49
Graphene	band gap	MLR, DT, SVM, ANN, GA	predict the band gap by using topological information.	prediction accuracy >80% with absolute error < 0.5 eV	36
Graphene-h-BN	band gap	ANN	predict the bandgap based on atomic system configuration	ANN method 1: R^2 = 0.950 ANN method 2: R^2 = 0.888	95
MXene	band gap	KRR, SVR, GPR, Bagging	predict the band gap	KRR: R^2 = 0.68, RMSE = 0.19 eV SVR: R^2 = 0.71, RMSE = 0.17 eV GPR: R^2 = 0.83, RMSE = 0.14 eV bagging: R^2 = 0.73, RMSE = 0.16 eV	51
MXene	band gap	GPR	predict the band edges	RMSE = 0.12 eV	96
$MoS_{2-x}O_x$ bilayer	band gap	RF	reveal the factors dominate SMT process	R^2 = 0.9	42
Stanene nanostructures	thermal conductivity	GA	assist the study of stanene nanostructures	enhance simulation efficiency	97
$MoS_{2(1-x)}Se_{2x}$ monolayer	thermal conductivity	SNAP	predict the parameters for EMD simulations	enhance simulation efficiency	98
2D materials	phononic properties	MTP	predict phonon dispersions, and phonons group velocity	accurately and efficiently replace DFPT method to evaluate phononic properties of complex structures	99
2D carbon nitrides	thermal conductivity	MTP	accelerate the MD simulation	enhance simulation efficiency	100
Diamanes	thermal conductivity	MTP	accelerate the MD simulation	enhance simulation efficiency	101
Porous graphene	thermal conductivity	CNN	search the structure with lowest thermal conductivity	efficient screening	50
2D materials	thermodynamic stability	GBDT, SISSO	evaluate thermodynamic stability of 2D materials	SISSO: RMSE = 0.165 eV/atom (low stability); RMSE = 0.245 eV/atom (medium stability); RMSE = 0.207 eV/atom (high stability)	40
$MoSe_2$ monolayer	mechanical properties	SVM	reveal the material type is the impact factor to influence the mechanical properties	–	102
WS_2 monolayer	mechanical properties	RF	reveal both the material type and defect ratio dominate the mechanical properties	–	45
Ti_3C_2 nanosheet	mechanical properties	MTP	reveal the correlation between strength and defects	–	103
WSe_2 monolayer	exciton valley polarization landscape	RF	predict the exciton valley polarization landscape at 15 K by using the data from 300 K	–	44
vdW layered $CuInP_2S_6$	ferroelectric properties	k-Means	assist to analyze the correlation between structure and property	–	104
$A_2B_2X_6$ monolayer	magnetic properties	KRR, ANN, SVM, extremely randomized trees (extra trees)	predict the magnetic properties and reveal the correlation between magnetic order and composition	extra forests: R^2 = 0.95, MAE = 0.30 for magnetic moment SVM: 82% success rate for ferromagnetic order; 80% success rate for antiferromagnetic order	105

(*Continued on next page*)

Table 3. *Continued*

Materials	Properties	Algorithms	Function of ML	Performance of ML	Ref.
2D graphene-based catalyst	binding energy	RF, SVR	predict binding energy between molecules and 2D catalysts	relaxed output structures: R^2 = 0.952, MAE = 0.166 unrelaxed input structures: R^2 = 0.865, MAE = 0.307	106
MXenes layered materials	toxicity	RF, SVM, extremely randomized trees (extra trees)	predict toxicity	–	107

though the formation energies can indeed be predicted well, all of the composition models perform poorly in predicting compound stability. In other words, formation energies do not completely determine the stability of the materials. The chemical formula proved to be far from an adequate descriptor to distinguish stable from unstable compounds within an arbitrary chemical space. Instead, the authors encourage the use of structural representations for materials discovery. This work emphasized the importance of assessing model performance and caution against the abuse of ML results that can be easily misunderstood.

ML-ASSISTED PROPERTY EXPLORATION

Since the properties of materials significantly influence their applications,[7] the rapid screening of materials properties may accelerate the development of various technologies. However, experimental and traditional theoretical screening requires significant investment of time and money, which may be reduced by integrating ML in the pipeline.[92] In the following sections, we will present the current advances of ML to predict specific properties (Table 3), including electronic properties, thermodynamic properties, mechanical properties, exciton valley polarization, ferroelectric properties, magnetic properties, binding energy, and toxicity.

Electronic properties

Electronic applications of 2D materials have been intensively studied,[85,108–110] and unsurprisingly these applications of are highly dependent on their electronic properties[111] such as the band gap, which is one of the fundamental properties that can be modulated to meet specific requirements. The discovery of new ways to control the electronic properties of 2D materials is essential for the advancement of electronic applications. Successful examples using ML to predict the electronic properties of 2D materials will be discussed in this section.

A dataset of electronic properties of graphene nanoflakes with various structures was generated via the DFTB method by Shi et al.,[112] and later ML was used to build models to predict the electronic properties of unknown structure that were not in the set. The 622 computational optimized graphene structures, including trigonal, hexagonal, and rectangular structures (Figures 7A–7C) from this dataset were used by Fernandez et al.[93] to study the structure-property relationship using the partial-least-square regression (PLSR) and radial distributions function (RDF) scores (Figure 7D). The results showed that the energy of the Fermi level and electron affinity correlates to the RDF scores at interatomic distances ranging from 2.0 to 10.0 Å, while band gap and ionization potential correlate to the corresponding value in the range of 3.0 to 30.0 Å. The as-built QSPR model can accurately predict Fermi level and ionization potential with more than 83% accuracy rate, where the prediction accuracy of ionization potential, and band gap reached ~80%, and 70%,

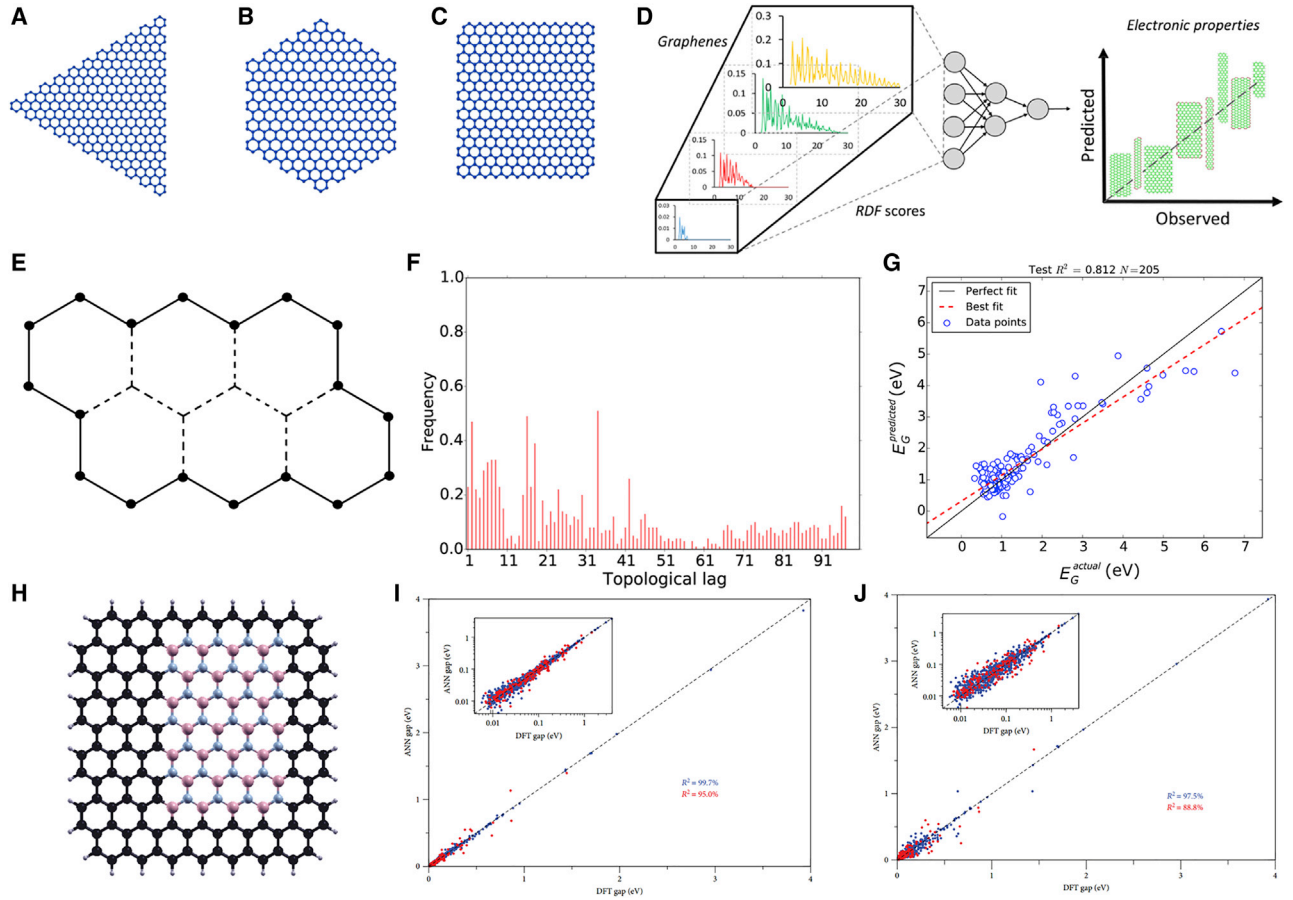

Figure 7. Applications of ML in the study of electronic properties of graphene and related materials

(A–C) Examples of (A) trigonal, (B) hexagonal, and (C) rectangular graphene structures in the dataset.

(D) Schematic illustration of electronic properties prediction via RDF scores.[93] Copyright 2015, ACS Publications.

(E) Molecular graph representation of a graphene nanoflake with edges connected by solid lines and all other interior nodes connected by dashed lines.

(F) Histograms of the ATS in the optimum SVM models of the energy gap of graphene nanoflakes in 200 independent GA runs.

(G) Scatterplot of the predictions for the energy gap of the graphene nanoflakes in the test set of 30% of the entire dataset, where "actual" refers to the values calculated by DFTB simulations and "predicted" corresponds to the SVM predictions.[36] Copyright 2016, ACS Publications.

(H) A typical graphene nanoflake with an embedded rectangular h-BN domain. The edges are passivated with hydrogen. Each system contains N = 166 C, B, or N atoms, colored in black, pink, and light blue, respectively, and N_H = 34 hydrogen atoms.

(I and J) Predicted ANN gap versus reference DFT gap, for typical fully connected networks with three layers: (I) method 1 (166/100/1 neurons) and (J) method 2 (20/100/1 neurons). The results corresponding to the training and test sets are represented in blue and red colors, respectively.[95] Copyright 2020, Hindawi.

respectively. Subsequently, they introduced more geometrical features to predict these electronic properties based on the same dataset of optimized graphene structures.[94] The nine inputs features (number of hydrogen atoms, fraction of zig-zag edges, surface area of graphene, ratio between the two main structural axis, average carbon atom coordination number, average C-C bond length, average C-C-C bond angle, average C-H bond length, and average C-C-H bond angle) yield a total of 511 input combinations. A collection of ML methods including multiple linear regression (MLR), decision tree (DT), *k*-nearest neighbor (*k*NN), ANNs, and SVM were applied to filter these input combinations and build rational models to predict the electronic properties of graphene including all four electron charge transfer properties.

As mentioned above, DFT is often not suitable for the problems involving a large number (>10^3) of atoms, which limits its applications, especially to 2D materials.

The DFTB method, an efficient quantum chemical approximation of DFT with self-consistent charges (SCC-DFTB), can be applied to calculate electronic properties of large 2D materials,[112] but with reduced accuracy than DFT. A fast and accurate correction term connecting the differences between the calculation results of DFT and DFTB can simultaneously improve accuracy and retain efficiency. Based on this goal, Fernandez et al.[49] built different ML models trained with structural features of graphene nanoflakes to predict the differences between DFT and DFTB calculations of two electronic properties (energy of the Fermi level and energy of the band gap), using the nine inputs features shown to be efficacious in their previous study,[94] and MLR, DT, random forest (RF), ANNs, and SVM. The RF model showed the best prediction accuracy of 94% in the prediction of the difference of the energy of the Fermi level, whereas the SVM model successfully predicted the difference of the energy of band gap with the highest prediction accuracy of 88%.

While these ML studies were highly successful, the requirement for complex features may be prohibitive and may not be suitable for rapid screening. Therefore, Fernandez et al. developed a method to accurately predict the band gap of graphene using the topology of its molecular graph, regardless of atom position information that can be difficult to assess experimentally.[36] The set of topological autocorrelation scores (ATSs) was introduced to conveniently describe the topology of graphene. As shown in Figure 7E, the structure of graphene can be represented as a molecular graph and be further transformed into a numerical ATS code. Different ML methods including MLR, DT, SVM, ANN, and GA were applied to optimize the model. As a result, the final model revealed that the topological distances in the range of 1 to 42 are the most informative ATS to predict band gaps (Figure 7F), with a high prediction accuracy of over 80% (Figure 7G) with a mean absolute error lower than 0.5 eV in the range of energy gap from 0 to 7 eV.

Nemnes et al. have also studied the band gap of graphene nanoflakes embedded with hexagonal boron nitride (Figure 7H) by using the combination of *ab initio* DFT and ANN.[95] As distinct from the single graphene that only contains C atoms, the hybrid structure of graphene-h-BN introduces additional B and N atoms, as well as the H atoms on the edges of hybrid material, which necessitate more complex structural information. It was previously known that both the size and position of the h-BN domains can influence the band gap of graphene-h-BN hybrid, so these were explicitly considered when using DFT to calculate band gaps as dataset. The authors designed two ANN models with different input neurons. The first model was a simple fully connected backpropagation ANN with the 200 inputs neurons for all the atoms of the graphene-h-BN hybrid, though is not necessary that the number of neurons matches the number of atoms. The second model simplified the number of inputs neurons from 200 to 20, with 4 of them represent the corresponding proportions of C, B, N, and H atoms, and 16 of them represent the normalized counts of atom quadruplets (X_i = C, B, N; Y_1, Y_2, Y_3 represent three nearest neighbors of X_i). Both ANN models showed good performance in band gap prediction, with accuracy over 90% (Figures 7I and 7J). With the simplified inputs the second model was suggested to be suitable for extension to other systems, though the improvement of this prediction was likely due to the reduction of the feature set, as opposed to the specific structure of the ANN.

MXene is a key 2D material since the different compositions exhibit various properties for a large range of applications. Singh et al. estimated the band gaps of MXenes by using different ML models using KRR, SVR, GPR, and bootstrap aggregating regression algorithms, with the input features including the boiling and melting

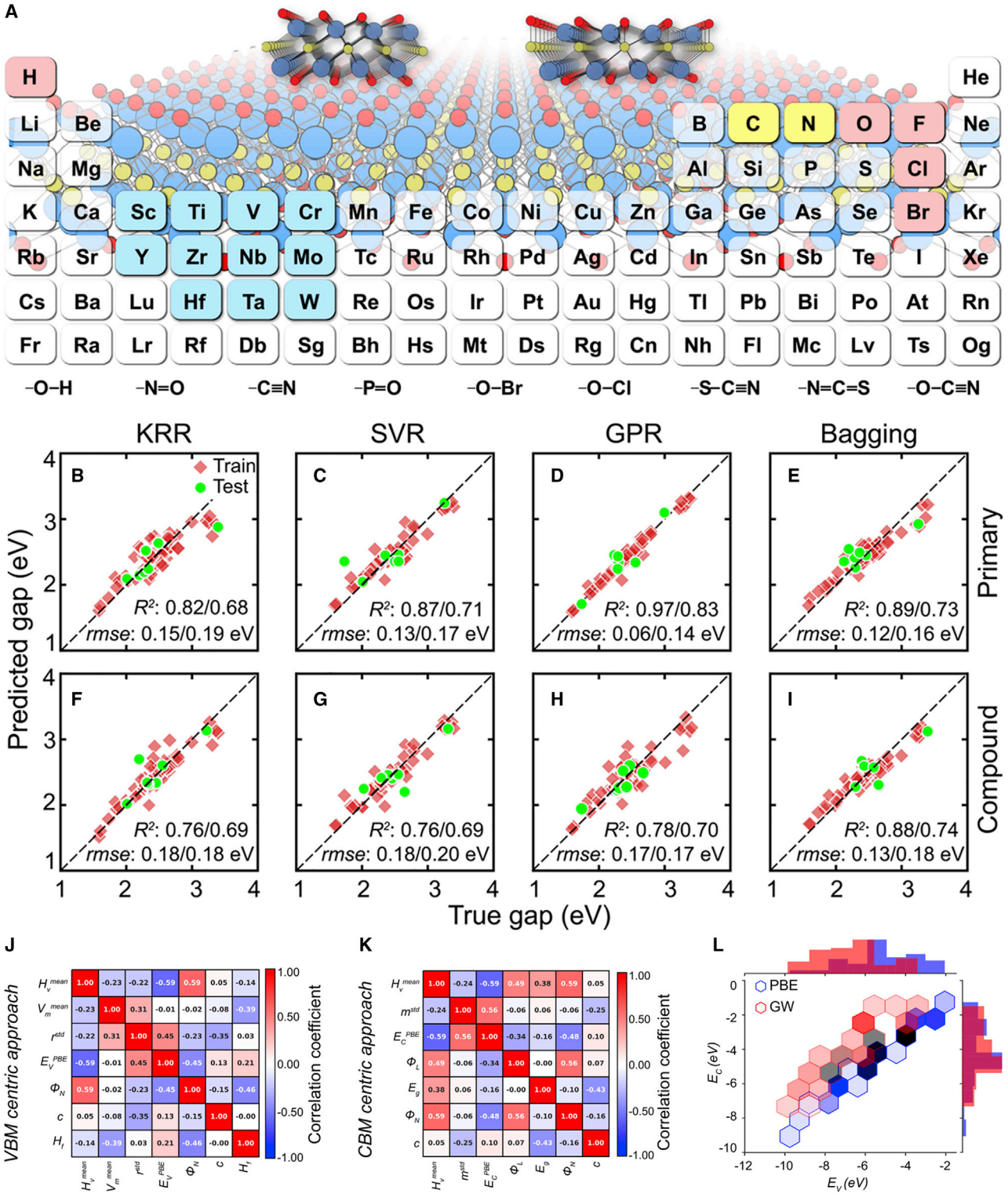

Figure 8. Prediction of band gaps of MXenes by using ML

(A) MXene composition. With 11 early transition metals M (blue) of IIIB to VIB groups; X, (C, N) (colored in yellow); and 14 surface functional groups T/T′ (red), a pool of 23 870 MXenes is generated. T/T′ consists of functionalization with elements (H, F, Cl, Br, O) and groups (CN, NO, PO, OH,

points, atomic radii, phases, bond lengths.[51] As shown in Figure 8A, a dataset of 23,870 MXenes with different compositions was developed, with 7,200 MXenes randomly selected for DFT structural optimization and band gap calculation. Thereafter, 76 semiconducting MXenes simulated with DFT (Perdew-Burke-Ernzerhof [PBE]) were selected to build the prediction models. After feature engineering, the number of input features was reduced from 47 to 8. As a result, the GPR model showed the best band gap prediction performance with the lowest root-mean-squared error (RMSE) of 0.14 eV (Figures 8B–8I). However, the result may not be convincing since the training dataset of 76 samples is very small, and convergence with respect to the training set size was not shown. For some of the applications of functionalized MXenes, such as photocatalysis, the position of the band edges is also important in addition to band gap. Hence, Singh et al. developed a GPR-based ML model to study the position of band edges of MXenes.[96] After a similar optimization process with previous study,[51] 76 MXenes having PBE band gap were selected from the dataset of MXenes, and 7 elemental and computed properties were chosen, on the basis of weighted importance that yielded by neighborhood component analysis (NCA), as input features (Figures 8J and 8K). In this work, GPR was selected as the basis for two approaches: (1) valence band maxima (VBM)-centric approach; and (2) conduction band minima (CBM)-centric approach. The CBM-centric approach showed better performance with a lower RMSE of 0.12 eV, which was attributed to better agreement between the calculated GW and PBE values of the CBM than the corresponding values of the VBM (Figure 8L).

Semiconductor-metal transition (SMT) is highly relevant to band gap modulation. By adjusting the band gap of a material it can switch between semiconductor and metallic states. Sun et al. used a cluster expansion method to determine that $MoS_{2-x}O_x$ bilayers can realize SMT.[42] In their study, a RF was applied to study the relationship between the calculated band gap and a set of structural features. In addition to the features generated by Matminer,[113] a few predefined features for the bilayer system were introduced. The results showed that the difference of the oxygen fraction in the two chalcogen-atom layers across the vdW gap (ΔF_{inter}) dominates the SMT process (Figure 9), with the assistance of average of the tangent of the Mo-S/O bond angle ($\overline{\tan\theta}$) as the secondary factor (Figure 9).

Thermodynamic properties

Thermodynamic properties of materials are essential for materials science since they directly reflect the influence of heat to the performance of materials. In this section, thermodynamic properties including thermodynamic stability and thermal conductivity and predictions of various ML algorithms will be compared and discussed.

Cherukara et al.[97] introduced a bond order potential (BOP) to study the thermal conductivity of stanine via MD simulations. They have found that stanene has a highly rippled structure, compared to other 2D materials such as graphene, and this softness and high anharmonic response resulted in significantly lower thermal

OCl, OBr, OCN, SCN, NCS) (shown at the bottom). Two prominent phases of MXene, namely, bb′ and cb, are shown at the top left and right, respectively.
(B–I) Band gap predictions of MXene. Scatterplots showing band gap predictions versus true (i.e., GW) gaps of important primary (B–E) and compound (F–I) feature-combinations.[51] Copyright 2018, ACS Publications.
(J and K) Heatmaps showing the correlation between 7 selected features via (J) the VBM centric approach and (K) the CBM centric approach.
(L) Hexabin distribution plot showing the distribution of PBE (blue) and GW (red) VBM and CBM at the absolute scale.[96] Copyright 2019, ACS Publications.

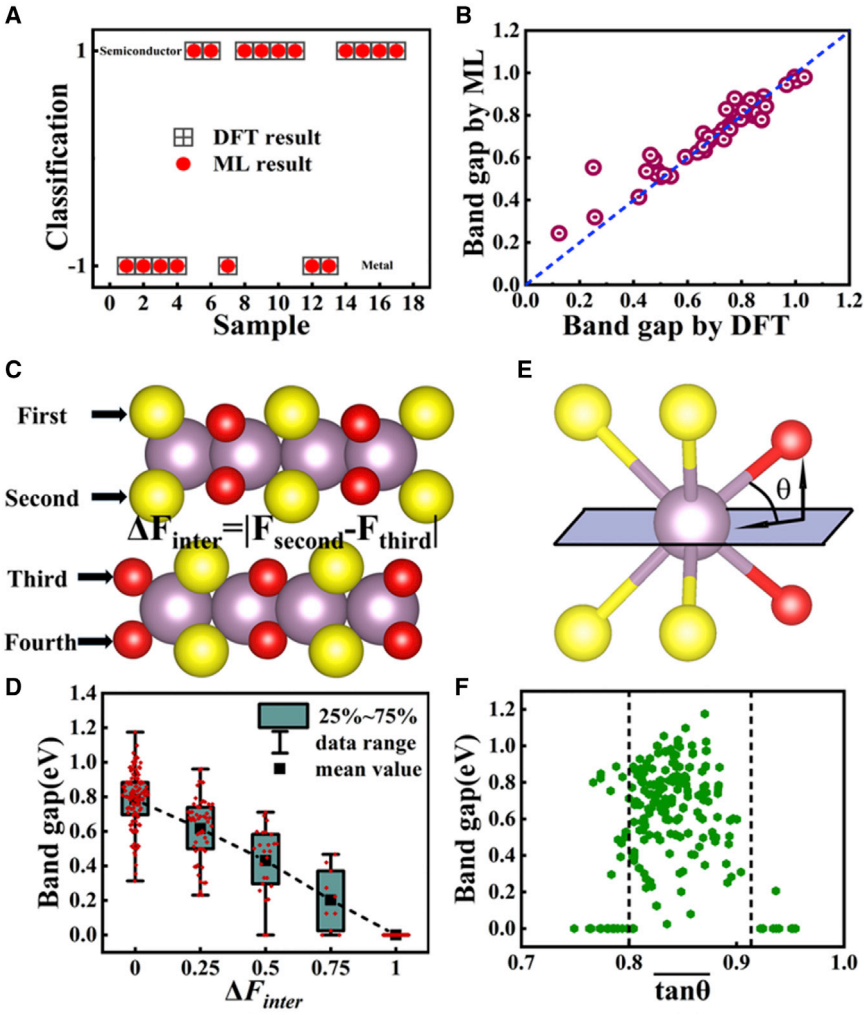

Figure 9. Study of SMT of $MoS_{2-x}O_x$ bilayers by ML

(A) Classification result of one typical test set, where y = −1 represents the metal structures; y = 1 represents semiconductor structures.

(B) Regression result of one typical test set using the random forest algorithm.

(C) Schematic diagram of the most important feature ΔF_{inter}.

(D) The relationship between the band gap data and ΔF_{inter} is shown below.

(E) Definition of the Mo-S/O bond angle $\overline{\tan\theta}$.

(F) The relationship between the band gap data and $\overline{\tan\theta}$ shown below.[42] Copyright 2020, ACS Publications.

conductivity. During the simulation the Tersoff parameters were optimized using GA, which indirectly contributed to the study of stanene nanostructures.

MD simulation is considered as an alternative method to Peierls-Boltzmann transport equation (PBTE) for the calculation of thermal conductivity; however, classic MD simulations can be influenced by inaccurate empirical interatomic potentials. For some thermodynamic studies, the availability and transferability of interatomic potentials has become a bottleneck for performing the simulations. Therefore, a new method named ML interatomic potential (MLIP) was proposed, to provide interatomic potential for further MD simulations, with both accuracy and efficiency.[114] Gu and Zhao[98] studied thermal conductivity of monolayer $MoS_{2(1-x)}Se_{2x}$ alloys based on

the combination of MD simulations and MLIP using a spectral neighbor analysis approach (SNAP). The introduction of SNAP, which demonstrated high accuracy in predicting phonon dispersion and anharmonic phonon properties, assisted the equilibrium MD (EMD) simulations to predict the thermal conductivity of monolayer $MoS_{2(1-x)}Se_{2x}$ alloys. Although the exclusion of long-range interaction in SNAP, and the absence of higher-order phonon scatterings (such as four-phonon scatterings) in PBTE, introduced slight differences between the results of EMD and PBTE, SNAP-assisted EMD still showed good accuracy in predicting thermal conductivity of monolayer $MoS_{2(1-x)}Se_{2x}$ alloys with potential for extension to other MoS_2, $MoSe_2$, and their mixtures-based nanostructures. Later, Mortazavi et al.[99] used the moment tensor potential (MTP), a fast and convenient method derived from MLIP, to study phononic properties of low-symmetry and porous 2D materials to replace traditional DFT simulations and improve computational efficiency. MTP was compared to the traditional density functional perturbation theory (DFPT) in the calculations of force constants, which can be used to evaluate photonic properties. The training dataset consists of a wide range of 2D materials, including monoelemental 2D lattices (such as graphene, and black phosphorene), binary 2D systems (such as CN, and BN), and ternary 2D lattices (such as BC_6N_6, and $BrCuTe_2$), allgenerated by using *ab initio* MD (AIMD) simulations. The MTP was trained with hundreds of parameters and yielded a competitive level of accuracy and reproducibility compared to the traditional DFPT in the prediction of phonon dispersions and phonons group velocity. It is expected that the combination of MD and MLIP can be extensively used to study thermal conductivity of many other 2D materials as well. For instance, Mortazavi et al.[100] used MTP to accelerate the non-equilibrium MD simulation to predict the thermal conductivity of several novel 2D carbon nitrides, including C_3N_4, C_3N_5, and C_3N_6. Subsequently, they used MTP to obtain the anharmonic force constants to further accelerate the study of thermal conductivity of semiconducting diamanes, including Janus C_4HF, C_4HCl, and C_4FCl and non-Janus C_2H, C_2F, and C_2Cl diamanes.[101]

The introduction of holes to 2D materials can efficiently tune their thermal conductivity, since the density and distribution of the holes dominate the thermal conductivity. Jiang et al. applied a CNN-based inverse design method to study the relationship between thermal conductivity reduction and hole distribution in monolayer graphene.[50] For the given structure shown in Figure 10A, there are 14 candidature sites for pores, yielding a total of 2^{14} = 16,384 possible porous graphene structures as training dataset. The thermal conductivity is related to the pore size and pore-size distribution (the number of holes out of the 14 possible hole sites), such that as the porosity increases the thermal conductivity decreased. However, each of the porosity levels covered a wide range of thermal conductivity (Figure 10B), indicating the spatial distribution of holes in significant. Using the trained CNN model as a first generation, the authors repeated the process through successive generations based on a larger dataset of porous graphene with 24 possible hole sites and porosity of 0.5, to search for porous graphene structures with the lowest thermal conductivity (Figure 10C). The top 5 porous graphene structures with the lowest thermal conductivity from different generations, along with 3 typical structures, are presented in Figure 10D, demonstrating the use of CNN for screening.

As pointed out by Zunger,[115] predicting the thermodynamic stability of materials in addition to desirable properties is necessary for practical translation. Schleder et al.[40] used ML to identify thermodynamically stable 2D materials using C2DB, to classify 2D materials as having low, medium, or high stability (Figure 11) according to the formation energy by using a combination of stochastic GBDT classifier and

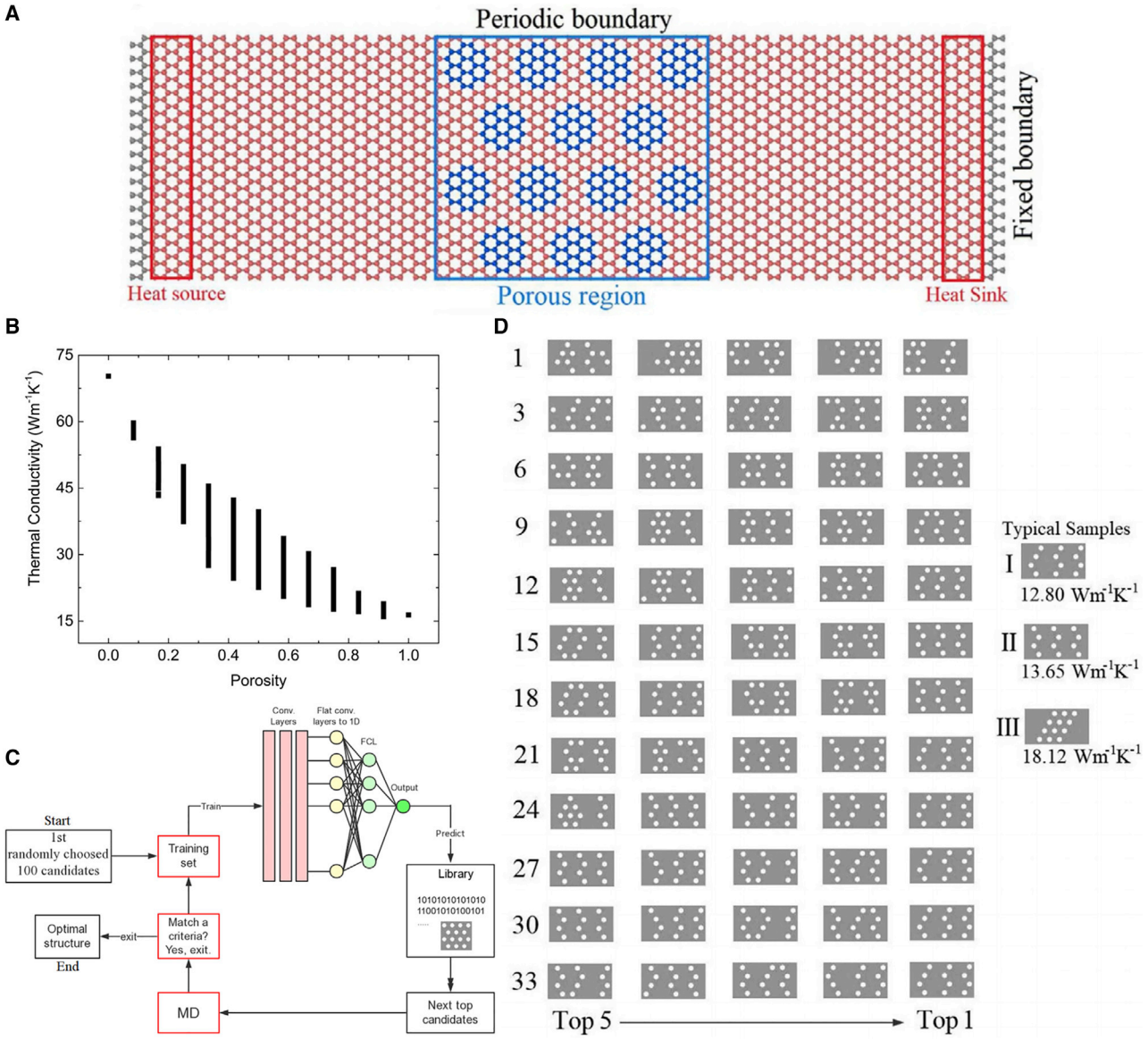

Figure 10. ML-assisted study of thermal conductivity of porous graphene

(A) The structure of porous graphene.

(B) Thermal conductivity versus the density of holes for porous graphene at room temperature.

(C) Schematic of the search algorithm based on the CNN model.

(D) The porous structure of the top 5 configurations from different generations (showed in left) with the lowest thermal conductivity searched by CNN model. I–III: three representative porous structures and their thermal conductivity value.[50] Copyright 2020, Elsevier.

the sure independence screening and sparsifying operator (SISSO) approach. The proposed approach can evaluate the stability of 2D materials using only the structural symmetry and compositions without information on the atomic positions. The approach was also used to predict the potential of these stable materials for photoelectrocatalytic water splitting.[40]

Mechanical properties

Mechanical properties, such as the Young's modulus and tensile strength, can directly influence the range of usefulness and durability of 2D materials.[116] The

Cell Reports
Physical Science
Review

CellPress
OPEN ACCESS

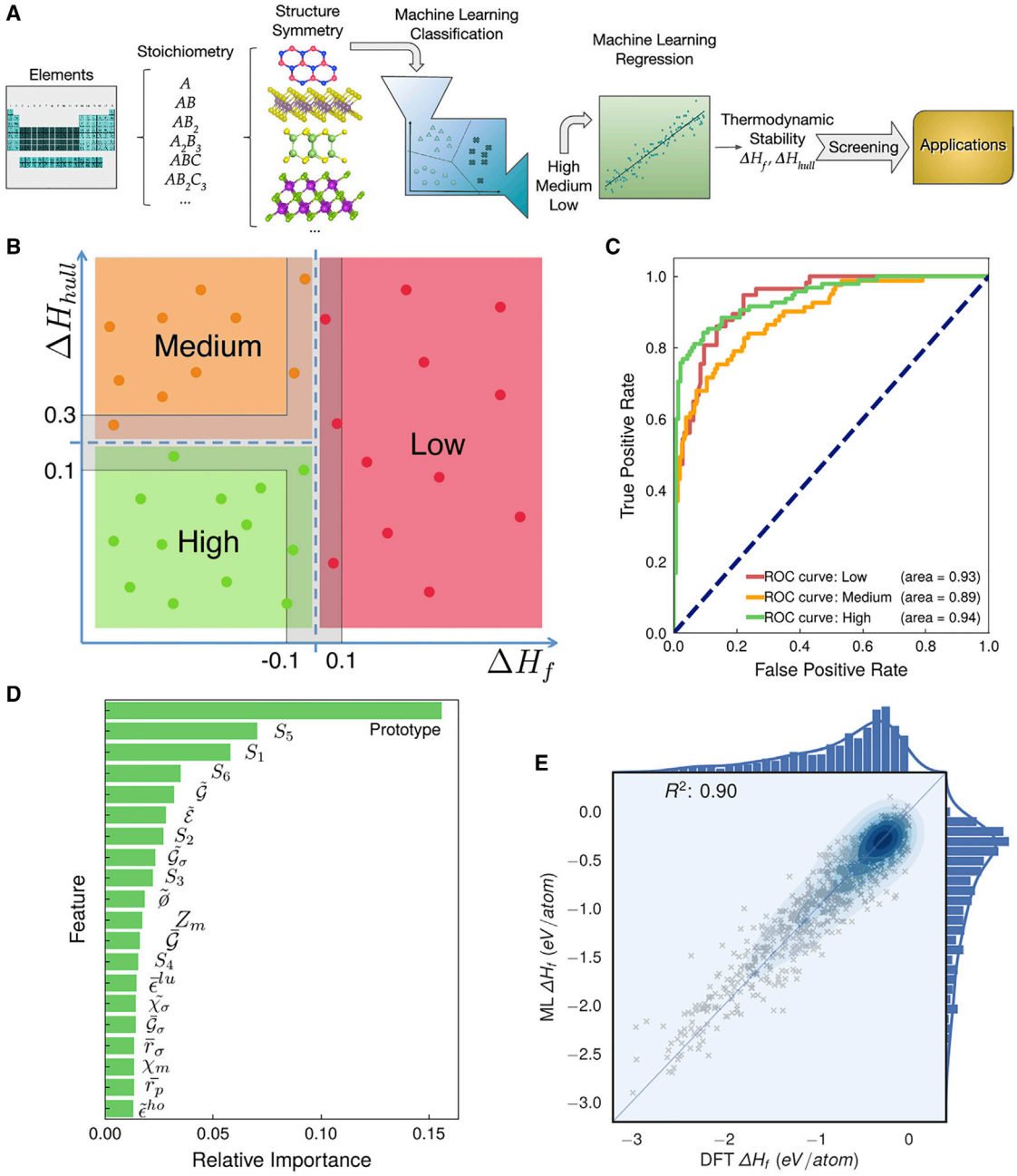

Figure 11. ML-assisted study of thermodynamic stability of 2D materials

(A) A schematic of ML-assisted process to explore 2D materials' thermodynamic stability.

(B) Classification of low, medium, or high stability based on the formation energy.

(C) Receiver operating characteristic (ROC) curve for the classification model and the corresponding area under curve (AUC) metric for each class: low (red), medium (orange), and high stability (green).

(D) Importance for the 20 most important features in the classification model.

(E) Correlation scatterplot of formation energies comparing the DFT calculated values with the regression model obtained via SISSO ML.[40] Copyright 2020, ACS Publications.

determination of mechanical properties usually needs complex and long-term experiments, whereas ML-assisted prediction may yield accurate results in a shorter period of time, as discussed in this section.

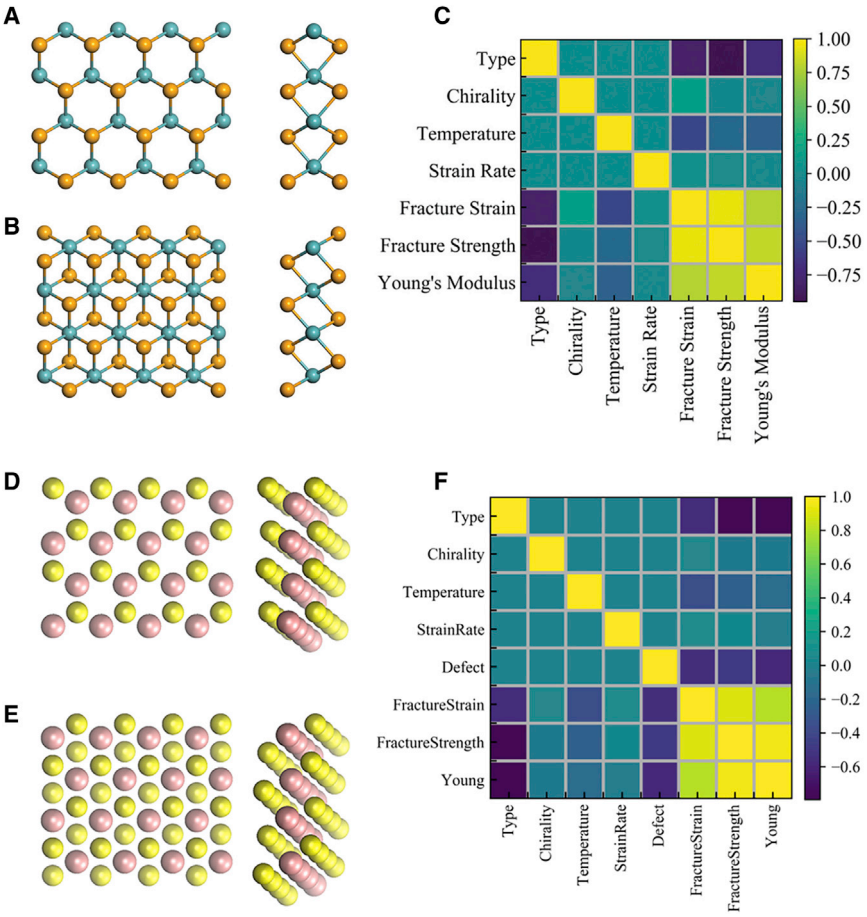

Figure 12. Study of mechanical properties of TMDs by ML

(A and B) Top and side views of (A) h-$MoSe_2$ and (B) t-$MoSe_2$.

(C) Standard correlation coefficients between the impact factors and the predicted mechanical properties of $MoSe_2$.[102] Copyright 2019, RSC Publishing.

(D and E) Top and side views of (D) h-WS_2 and (E) t-WS_2.

(F) Pearson's correlations among different features and target outputs.[45] Copyright 2019, ACS Publications.

Zhang et al.[102] studied the mechanical properties of h-$MoSe_2$ and t-$MoSe_2$ in both armchair and zigzag directions (Figures 12A and 12B) by using MD simulations and an SVM. The correlation between mechanical properties (fracture strain, fracture strength, and Young's modulus) of $MoSe_2$ and various impact factors ($MoSe_2$ type, chirality, temperature, and strain rate) were investigated. A dataset of 700 samples yielded from experiments and simulations were split into training and testing sets using a stratified 80/20 split. After the training with 4 input features ($MoSe_2$ type, chiral direction, temperature, and strain rate), the results showed that the type of $MoSe_2$ is the most important feature influencing the mechanical properties (Figure 12C). Subsequently, they have extended their study to the mechanical properties of both h-WS_2 and t-WS_2 monolayers (Figures 12D and 12E).[45] A larger dataset with a total of 3,600 MD simulations combining 2 types of WS_2, 2 directions, 6 temperatures, 5 strain rates, 6 defect ratios, and 5 initial conditions were generated for training and testing using RF. Five features including the type or material, chirality, temperature, strain rate, and defect ratio were considered as inputs, where fracture strain, fracture strength, and Young's modulus were used as the target labels. As

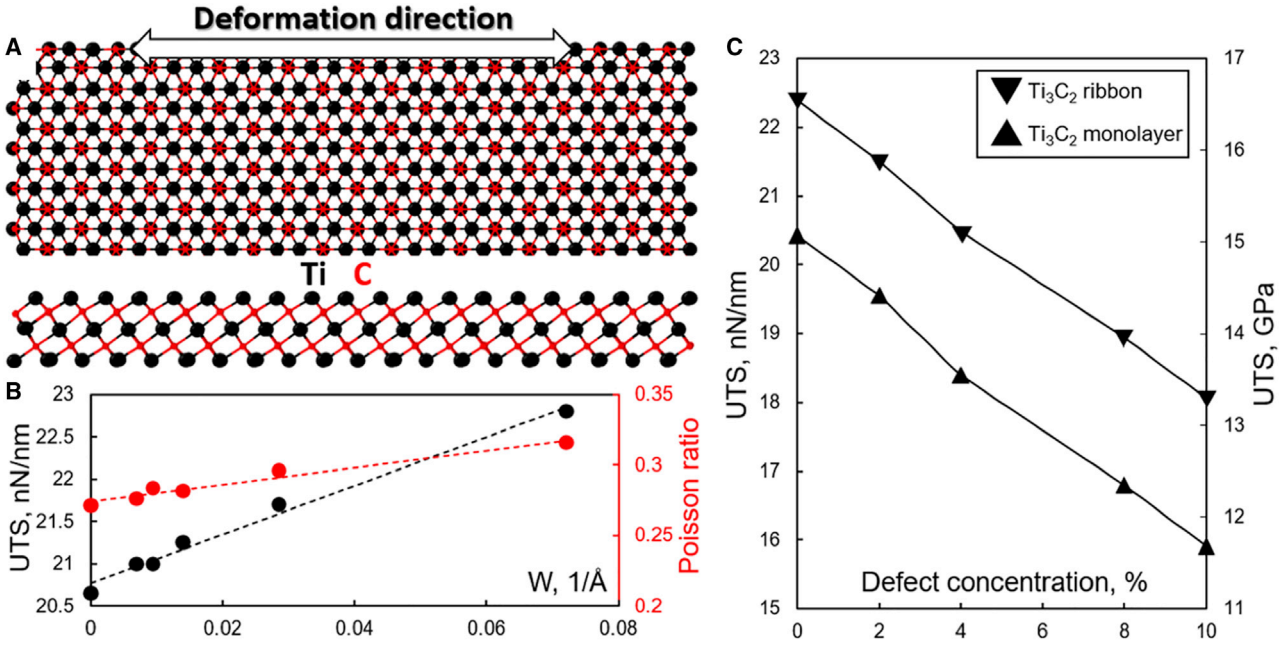

Figure 13. ML-assisted study of mechanical properties of Ti_3C_2 nanosheets

(A) Representation of a Ti_3C_2 nanoribbon (top and side views). Black and red spheres correspond to Ti and C atoms, respectively.

(B) Dependence of UTS (black) and Poisson ratio (red) on the inverse ribbon width *W*.

(C) Dependence of UTS in nN/nm (left axis) and GPa (right axis) units on monovacancy concentration for the Ti_3C_2 monolayer (triangles) and nanoribbon with a chosen width of 3.4 nm (upside down triangles).[103]

shown in Figure 12F, both WS_2 type and defect ratio dominate the mechanical properties of WS_2.

Firestein et al.[103] have fabricated Ti_3C_2 nanosheets and studied their Young's modulus and tensile strength by using a combination of experiments and simulations. In contrast with previous studies that are mostly based on computational datasets, this work has conducted a series of experimental characterizations prior to simulation. The Young's modulus of Ti_3C_2 was measured by AFM, where the tensile tests and the ultimate tensile strength (UTS) were conducted and measured by *in situ* TEM techniques, respectively. The experimental results showed the correlation of UTS and the thickness of nanosheets, which may be attributed to the presence of defects. To further understand the mechanical deformation of Ti_3C_2, the MLIP method was introduced to study the influence of geometrical parameters and structural defects on mechanical properties, revealing the negative correlation between the strength of Ti_3C_2 nanoribbons and vacancy concentration (Figure 13).

Other properties

Other properties and potential applications that have received less systemic study include the exciton valley polarization, ferroelectric properties, magnetic properties, binding energy, and toxicity of 2D materials. These properties are also important to the development of this domain and will be listed and discussed in this section.

The heterogeneity of exciton valley polarization is essential for both fundamental scientific research and future applications of 2D materials in valleytronic devices.

However, it is not simple to predict low-temperature heterogeneity of exciton valley polarization by using room-temperature measurements. Miyauchi et al.[44] used a RF to predict the low-temperature (15 K) exciton valley polarization landscape of a WSe_2 monolayer using the information extracted from photoluminescence (PL) spectra at room temperature (300 K). The decision trees were generated by input variables of PL characteristics extracted from mapping results at 300 K and output variables of position-dependent exciton valley polarization at 15 K, respectively. Bootstrap aggregation was applied in building decision trees to overcome the limited size of datasets. The RF showed good prediction capability, with 8 of 9 samples were successfully predicted, and discovered that PL intensity and trion-exciton intensity ratio at 300 K are the most important features in determining exciton valley polarization at low temperature.

Neumayer et al.[104] studied the ferroelectric properties of vdW layered $CuInP_2S_6$ using a combination of piezoresponse force microscopy (PFM) and *k*-means clustering. PFM was used to directly observe the correlation between strain and ferroelectric properties in vdW layered $CuInP_2S_6$, and output data were reduced by PCA prior to groups to reveal patterns in the experimental data. The combination of PFM and unsupervised learning indicated that the strain can be used to locally manipulate electromechanical behavior of $CuInP_2S_6$ and further enhance its functionality.

The magnetic properties, of $Cr_2Ge_2Te_6$, a monolayer $A_2B_2X_6$ material, such as magnetic order and magnetic moment, were studied using DFT calculations and support vector classification (SVC) by Rhone et al.[105] The DFT calculation that includes non-collinear spin and pain-orbit interactions was employed to generate a database of monolayer $Cr_2Ge_2Te_6$ containing atomic features. They achieved the accuracy rates of 82% and 80% for the prediction of ferromagnetic order and antiferromagnetic order, respectively, and revealed that the X site is important in determining the magnetic order of the structure and can significantly affect the magnetic coupling between adjacent A sites, which drives the magnetic ordering.

2D materials have also been widely employed as catalysts for various applications. The binding energy between molecules and metal atoms is a key factor to influence the attachment and detachment of small molecules to the catalysts and further influence the reaction efficiency and suitability. Barnard et al.[106] used RF and SVM to accurately predict the binding energies of 2D catalytic materials. A topological descriptor containing hundreds of features that can be divided into three groups (bond lengths/angles, statistical features, and partial radial distances) were used to train the models. The RF model predicted the binding energy between the molecules and metal atoms stabilized with doped, defective graphene (Figures 14A–14F), using a computational dataset containing approximate 1,700 samples. The method showed good accuracy in the prediction of binding energy, with R^2 = 0.952 (Figure 14G) for using the relaxed output structures (after quantum chemical simulations), and R^2 = 0.865 (Figure 14H) for unrelaxed input structures (without costly simulations), respectively.

As another key application of 2D materials is in biomedical applications, their cytotoxicity cannot be ignored. Birowska et al.[107] studied the cytotoxicity of MXenes by using RF involving two essential factors: the presence of transition metal oxides, and lithium atoms on surface. The model predicted toxicity for other 2D MXenes that have not been studied *in vitro* based on inputs features of surface characteristics,

Please cite this article in press as: Yin et al., The data-intensive scientific revolution occurring where two-dimensional materials meet machine learning, Cell Reports Physical Science (2021), https://doi.org/10.1016/j.xcrp.2021.100482

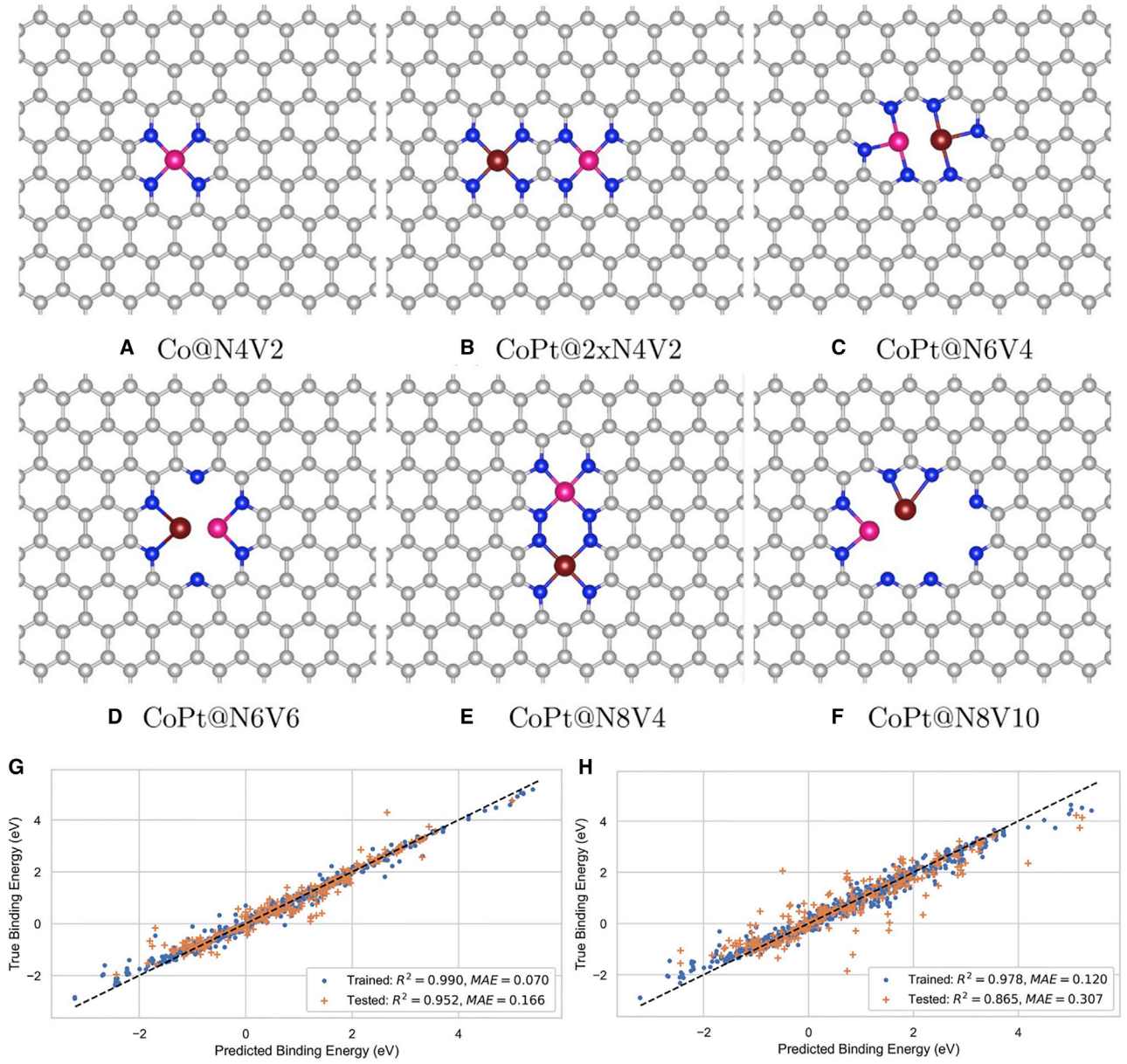

Figure 14. ML-assisted prediction of 2D catalytic materials

(A–F) Typical configurations of the six different pores supporting single metal atoms or paired single metal atoms that were considered in this study. Atoms are colored in gray (carbon), blue (nitrogen), pink (cobalt), and brown (platinum). Note: N and V refer to nitrogen and vacancy, respectively.

(G and H) Scatterplot predicting the binding energy on (G) N-doped porous graphene paired single atoms metal catalysts using random forest regression, and (H) porous graphene single atoms metal catalysts using random forest regression and only input features.[106] Copyright 2020, Wiley.

morphology, and physical structure. This work is inspiring since it showed that the undesirable properties such as the toxicity of 2D materials can be predicted as well as desirable electronic, thermodynamic, mechanical, or magnetic properties. However, it was found that geometrical descriptors alone are insufficient to train a ML model of toxicity, and inclusion of experimental information including material synthesis methods and characterizations would be beneficial. More information on materials themselves can assist in the interpretation of ML models, but the

prediction of toxicity requires information about the environment, interactions, and the biological availability. 2D materials with similar structural characteristics can exhibit different toxicity when exposed to different cell lines.

OUTLOOK

In conclusion, this review has discussed the recent advances of the application of ML in 2D materials, including materials preparation, structure analysis, and properties prediction. The current studies have demonstrated the advantages of the combination of ML and 2D materials data, either from simulations or experiments, pointing to a positive future. Both challenges and opportunities are still present, and will be highlighted in following paragraphs.

ML has a broad range of application scenarios and potential in every step of design, discovery, and developing synthesis processes of 2D materials. During the material design and discovery the difficulty and high cost of generating data could limit the performance of some ML methods, potentially resulting in under-fitting, orderfitting, low fidelity, or entirely misleading predictions due to selection bias. Some methods worth exploring to address this include transfer learning to train the ML model with the knowledge gained from related problems.[35] ML models can also be used as surrogates, replacing computationally intensive DFT,[40] or for data-driven feature selection to improve ML readiness of 2D material databases. ML could revolutionize 2D material synthesis since conventional experimental methods are slow and expensive and do not fulfill the growing demands in material manufacturing.[66] With the assistance of ML experiments could be automatic, and even autonomous, using guided high-throughput systems. For instance, a mobile robotic chemist guided by ML[59] was employed to perform 688 experiments within a ten-variable experimental space in 8 days, which is 1,000 times faster than manual approaches, and this autonomous system identified a six times more active photocatalyst mixtures. ML also has the potential to change the form of conventional experiments and increase the efficiency of 2D material discovery.

From a future perspective, we can envision two main strategies for developing and using ML in 2D material structure analysis, which will likely have long-term impact in this field. One is to focus on experimentally relevant features and use experimental designs informed by ML-assisted simulations of synthesis reactions, conditions, and dynamics. Another is improving the efficiency and effectiveness of ML models to accurately describe the underlying structural properties of 2D materials based on experimental information at scale, to enable real-time prediction and monitoring.

Compared to the large demand for 2D materials for a variety of applications, the number of studies related to the ML-assisted property prediction is still small. Even a basic strategy beginning with a simple 2D material, such as a graphene monolayer, and systematically introducing of more complexity (such as doping, multilayered materials, or heterostructures) will benefit the scientific community and accelerate 2D materials innovation. However, current studies are still worth improving for better performance, such as the prediction of band gaps of MXenes. The studies of different types of properties are still limited, and many properties including optical properties, superconductivity, toxicity are worth extensive investigation.

We believe that a more widespread use of ML in 2D materials design will help meet the increasing demands of industry and prepare 2D materials synthesis and characterization for development under industry 4.0.

Cell Reports
Physical Science
Review

CellPress
OPEN ACCESS

ACKNOWLEDGMENTS

This work was supported by the ANU Futures Scheme (Q4601024), the Australian Research Council (DP190100295, LE190100014), the National Natural Science Foundation of China (No. 51706114 and 51302166), Functional Materials Interfaces Genome (FIG) project, and Doctoral Fund of Ministry of Education of China (20133108120021).

AUTHOR CONTRIBUTIONS

H.Y., Z.S., and Z.W. prepared the manuscript. Z.Y., H.Z., and A.S.B. conceived the idea. All the authors contributed to revising the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Jordan, M.I., and Mitchell, T.M. (2015). Machine learning: Trends, perspectives, and prospects. Science *349*, 255–260.
2. Butler, K.T., Davies, D.W., Cartwright, H., Isayev, O., and Walsh, A. (2018). Machine learning for molecular and materials science. Nature *559*, 547–555.
3. Agrawal, A., and Choudhary, A. (2016). Perspective: Materials informatics and big data: Realization of the "fourth paradigm" of science in materials science. APL Mater. *4*, 053208.
4. de Pablo, J.J., Jackson, N.E., Webb, M.A., Chen, L.-Q., Moore, J.E., Morgan, D., Jacobs, R., Pollock, T., Schlom, D.G., Toberer, E.S., et al. (2019). New frontiers for the materials genome initiative. npj Comput. Mater. *5*, 41.
5. Barnard, A.S. (2020). Best Practice Leads to the Best Materials Informatics. Matter *3*, 22–23.
6. Yu, S., Wu, X., Wang, Y., Guo, X., and Tong, L. (2017). 2D Materials for Optical Modulation: Challenges and Opportunities. Adv. Mater. *29*, 1606128.
7. Lu, Z., Neupane, G.P., Jia, G., Zhao, H., Qi, D., Du, Y., Lu, Y., and Yin, Z. (2020). 2D Materials Based on Main Group Element Compounds: Phases, Synthesis, Characterization, and Applications. Adv. Funct. Mater. *30*, 2001127.
8. Yin, H., Xing, K., Zhang, Y., Dissanayake, D.M.A.S., Lu, Z., Zhao, H., Zeng, Z., Yun, J.-H., Qi, D.-C., et al. (2021). Periodic nanostructures: preparation, properties and applications. Chem. Soc. Rev. *50*, 6423–6482.
9. Cao, Y., Fatemi, V., Fang, S., Watanabe, K., Taniguchi, T., Kaxiras, E., and Jarillo-Herrero, P. (2018). Unconventional superconductivity in magic-angle graphene superlattices. Nature *556*, 43–50.
10. Tshitoyan, V., Dagdelen, J., Weston, L., Dunn, A., Rong, Z., Kononova, O., Persson, K.A., Ceder, G., and Jain, A. (2019). Unsupervised word embeddings capture latent knowledge from materials science literature. Nature *571*, 95–98.
11. Quesada-Cabrera, R., Weng, X., Hyett, G., Clark, R.J.H., Wang, X.Z., and Darr, J.A. (2013). High-throughput continuous hydrothermal synthesis of nanomaterials (part II): unveiling the as-prepared CexZryYzO2-δ phase diagram. ACS Comb. Sci. *15*, 458–463.
12. Sheremetyeva, N., Lamparski, M., Daniels, C., Van Troeye, B., and Meunier, V. (2020). Machine-learning models for Raman spectra analysis of twisted bilayer graphene. Carbon *169*, 455–464.
13. Moosavi, S.M., Jablonka, K.M., and Smit, B. (2020). The Role of Machine Learning in the Understanding and Design of Materials. J. Am. Chem. Soc. *142*, 20273–20287.
14. Zunger, A. (2018). Inverse design in search of materials with target functionalities. Nat. Rev. Chem. *2*, 0121.
15. Deringer, V.L., Caro, M.A., and Csányi, G. (2019). Machine Learning Interatomic Potentials as Emerging Tools for Materials Science. Adv. Mater. *31*, e1902765.
16. Kang, P.-L., Shang, C., and Liu, Z.-P. (2020). Large-Scale Atomic Simulation via Machine Learning Potentials Constructed by Global Potential Energy Surface Exploration. Acc. Chem. Res. *53*, 2119–2129.
17. Momeni, K., Ji, Y., Wang, Y., Paul, S., Neshani, S., Yilmaz, D.E., Shin, Y.K., Zhang, D., Jiang, J.-W., Park, H.S., et al. (2020). Multiscale computational understanding and growth of 2D materials: a review. npj Comput. Mater. *6*, 22.
18. Novoselov, K.S., Geim, A.K., Morozov, S.V., Jiang, D., Zhang, Y., Dubonos, S.V., Grigorieva, I.V., and Firsov, A.A. (2004). Electric field effect in atomically thin carbon films. Science *306*, 666–669.
19. Zeng, Z., Yin, Z., Huang, X., Li, H., He, Q., Lu, G., Boey, F., and Zhang, H. (2011). Single-layer semiconducting nanosheets: high-yield preparation and device fabrication. Angew. Chem. Int. Ed. Engl. *50*, 11093–11097.
20. Zhao, H., Ezeh, C.I., Ren, W., Li, W., Pang, C.H., Zheng, C., Gao, X., and Wu, T. (2019). Integration of machine learning approaches for accelerated discovery of transition-metal dichalcogenides as Hg0 sensing materials. Appl. Energy *254*, 113651.
21. Huang, S., Liu, Y., Zhao, Y., Ren, Z., and Guo, C.F. (2019). Flexible Electronics: Stretchable Electrodes and Their Future. Adv. Funct. Mater. *29*, 1805924.
22. Zhao, H., Ezeh, C.I., Yin, S., Xie, Z., Pang, C.H., Zheng, C., Gao, X., and Wu, T. (2020). MoO3-adjusted δ-MnO2 nanosheet for catalytic oxidation of Hg0 to Hg2+. Appl. Catal. B *263*, 117829.
23. Zhou, W., Yin, Z., Du, Y., Huang, X., Zeng, Z., Fan, Z., Liu, H., Wang, J., and Zhang, H. (2013). Synthesis of few-layer MoS2 nanosheet-coated TiO2 nanobelt heterostructures for enhanced photocatalytic activities. Small *9*, 140–147.
24. Yan, Y., Shin, W.I., Chen, H., Lee, S.-M., Manickam, S., Hanson, S., Zhao, H., Lester, E., Wu, T., and Pang, C.H. (2020). A recent trend: application of graphene in catalysis. Carbon Letters *31*, 177–199.
25. Huang, X., Yin, Z., Wu, S., Qi, X., He, Q., Zhang, Q., Yan, Q., Boey, F., and Zhang, H. (2011). Graphene-based materials: synthesis, characterization, properties, and applications. Small *7*, 1876–1902.
26. Li, L., Yu, Y., Ye, G.J., Ge, Q., Ou, X., Wu, H., Feng, D., Chen, X.H., and Zhang, Y. (2014). Black phosphorus field-effect transistors. Nat. Nanotechnol. *9*, 372–377.
27. Ling, X., Wang, H., Huang, S., Xia, F., and Dresselhaus, M.S. (2015). The renaissance of black phosphorus. Proc. Natl. Acad. Sci. USA *112*, 4523–4530.
28. Caldwell, J.D., Aharonovich, I., Cassabois, G., Edgar, J.H., Gil, B., and Basov, D.N. (2019). Photonics with hexagonal boron nitride. Nat. Rev. Mater. *4*, 552–567.
29. Yin, Z., Li, H., Li, H., Jiang, L., Shi, Y., Sun, Y., Lu, G., Zhang, Q., Chen, X., and Zhang, H. (2012). Single-layer MoS2 phototransistors. ACS Nano *6*, 74–80.
30. Gogotsi, Y., and Anasori, B. (2019). The Rise of MXenes. ACS Nano *13*, 8491–8494.
31. Bhaskar, G., Gvozdetskyi, V., Batuk, M., Wiaderek, K.M., Sun, Y., Wang, R., Zhang, C., Carnahan, S.L., Wu, X., Ribeiro, R.A., et al. (2021). Topochemical Deintercalation of Li

from Layered LiNiB: toward 2D MBene. J. Am. Chem. Soc. *143*, 4213–4223.

32. Vogt, P., De Padova, P., Quaresima, C., Avila, J., Frantzeskakis, E., Asensio, M.C., Resta, A., Ealet, B., and Le Lay, G. (2012). Silicene: compelling experimental evidence for graphenelike two-dimensional silicon. Phys. Rev. Lett. *108*, 155501.

33. Haastrup, S., Strange, M., Pandey, M., Deilmann, T., Schmidt, P.S., Hinsche, N.F., Gjerding, M.N., Torelli, D., Larsen, P.M., Riis-Jensen, A.C., et al. (2018). The Computational 2D Materials Database: high-throughput modeling and discovery of atomically thin crystals. 2D Materials *5*, 042002.

34. Choudhary, K., Garrity, K.F., Reid, A.C.E., DeCost, B., Biacchi, A.J., Hight Walker, A.R., Trautt, Z., Hattrick-Simpers, J., Kusne, A.G., Centrone, A., et al. (2020). The joint automated repository for various integrated simulations (JARVIS) for data-driven materials design. npj Comput. Mater. *6*, 173.

35. Frey, N.C., Akinwande, D., Jariwala, D., and Shenoy, V.B. (2020). Machine Learning-Enabled Design of Point Defects in 2D Materials for Quantum and Neuromorphic Information Processing. ACS Nano *14*, 13406–13417.

36. Fernandez, M., Abreu, J.I., Shi, H., and Barnard, A.S. (2016). Machine Learning Prediction of the Energy Gap of Graphene Nanoflakes Using Topological Autocorrelation Vectors. ACS Comb. Sci. *18*, 661–664.

37. Lee, C.-H., Khan, A., Luo, D., Santos, T.P., Shi, C., Janicek, B.E., Kang, S., Zhu, W., Sobh, N.A., Schleife, A., et al. (2020). Deep Learning Enabled Strain Mapping of Single-Atom Defects in Two-Dimensional Transition Metal Dichalcogenides with Sub-Picometer Precision. Nano Lett. *20*, 3369–3377.

38. Jin, H., Zhang, H., Li, J., Wang, T., Wan, L., Guo, H., and Wei, Y. (2020). Discovery of Novel Two-Dimensional Photovoltaic Materials Accelerated by Machine Learning. J. Phys. Chem. Lett. *11*, 3075–3081.

39. Hundi, P., and Shahsavari, R. (2019). Deep Learning to Speed up the Development of Structure-Property Relations For Hexagonal Boron Nitride and Graphene. Small *15*, e1900656.

40. Schleder, G.R., Acosta, C.M., and Fazzio, A. (2020). Exploring Two-Dimensional Materials Thermodynamic Stability via Machine Learning. ACS Appl. Mater. Interfaces *12*, 20149–20157.

41. Siriwardane, E.M.D., Joshi, R.P., Kumar, N., and Çakır, D. (2020). Revealing the Formation Energy-Exfoliation Energy-Structure Correlation of MAB Phases Using Machine Learning and DFT. ACS Appl. Mater. Interfaces *12*, 29424–29431.

42. Chen, Q., Chen, M., Zhu, L., Miao, N., Zhou, J., Ackland, G.J., and Sun, Z. (2020). Composition-Gradient-Mediated Semiconductor-Metal Transition in Ternary Transition-Metal-Dichalcogenide Bilayers. ACS Appl. Mater. Interfaces *12*, 45184–45191.

43. Frey, N.C., Wang, J., Vega Bellido, G.I., Anasori, B., Gogotsi, Y., and Shenoy, V.B. (2019). Prediction of Synthesis of 2D Metal Carbides and Nitrides (MXenes) and Their Precursors with Positive and Unlabeled Machine Learning. ACS Nano *13*, 3031–3041.

44. Tanaka, K., Hachiya, K., Zhang, W., Matsuda, K., and Miyauchi, Y. (2019). Machine-Learning Analysis to Predict the Exciton Valley Polarization Landscape of 2D Semiconductors. ACS Nano *13*, 12687–12693.

45. Wang, X., Han, D., Hong, Y., Sun, H., Zhang, J., and Zhang, J. (2019). Machine Learning Enabled Prediction of Mechanical Properties of Tungsten Disulfide Monolayer. ACS Omega *4*, 10121–10128.

46. Deringer, V.L., Pickard, C.J., and Proserpio, D.M. (2020). Hierarchically Structured Allotropes of Phosphorus from Data-Driven Exploration. Angew. Chem. Int. Ed. Engl. *59*, 15880–15885.

47. Liu, Z.-L., Kang, P., Zhu, Y., Liu, L., and Guo, H. (2020). Material informatics for layered high-TC superconductors. APL Mater. *8*, 061104.

48. Kireeva, N., and Pervov, V.S. (2020). Materials Informatics Screening of Li-Rich Layered Oxide Cathode Materials with Enhanced Characteristics Using Synthesis Data. Batter. Supercaps *3*, 427–438.

49. Fernandez, M., Bilić, A., and Barnard, A.S. (2017). Machine learning and genetic algorithm prediction of energy differences between electronic calculations of graphene nanoflakes. Nanotechnology *28*, 38LT03.

50. Wan, J., Jiang, J.-W., and Park, H.S. (2020). Machine learning-based design of porous graphene with low thermal conductivity. Carbon *157*, 262–269.

51. Rajan, A.C., Mishra, A., Satsangi, S., Vaish, R., Mizuseki, H., Lee, K.-R., and Singh, A.K. (2018). Machine-Learning-Assisted Accurate Band Gap Predictions of Functionalized MXene. Chem. Mater. *30*, 4031–4038.

52. Choudhary, K., Bercx, M., Jiang, J., Pachter, R., Lamoen, D., and Tavazza, F. (2019). Accelerated Discovery of Efficient Solar-cell Materials using Quantum and Machine-learning Methods. Chem. Mater. *31*, 5900–5908.

53. Schleder, G.R., Padilha, A.C.M., Acosta, C.M., Costa, M., and Fazzio, A. (2019). From DFT to machine learning: recent approaches to materials science–a review. J. Phys.: Mater. *2*, 032001.

54. Liu, Y., Zhao, T., Ju, W., and Shi, S. (2017). Materials discovery and design using machine learning. J. Materiomics *3*, 159–177.

55. Wang, A.Y.-T., Murdock, R.J., Kauwe, S.K., Oliynyk, A.O., Gurlo, A., Brgoch, J., Persson, K.A., and Sparks, T.D. (2020). Machine Learning for Materials Scientists: An Introductory Guide toward Best Practices. Chem. Mater. *32*, 4954–4965.

56. Chen, C., Zuo, Y., Ye, W., Li, X., Deng, Z., and Ong, S.P. (2020). A Critical Review of Machine Learning of Energy Materials. Adv. Energy Mater. *10*, 1903242.

57. Barnard, A.S., Motevalli, B., Parker, A.J., Fischer, J.M., Feigl, C.A., and Opletal, G. (2019). Nanoinformatics, and the big challenges for the science of small things. Nanoscale *11*, 19190–19201.

58. Todeschini, R., Consonni, V., and Gramatica, P. (2009). 4.05 - Chemometrics in QSAR. In Comprehensive Chemometrics, S.D. Brown, R. Tauler, and B. Walczak, eds. (Elsevier), pp. 129–172.

59. Burger, B., Maffettone, P.M., Gusev, V.V., Aitchison, C.M., Bai, Y., Wang, X., Li, X., Alston, B.M., Li, B., Clowes, R., et al. (2020). A mobile robotic chemist. Nature *583*, 237–241.

60. Zhong, M., Tran, K., Min, Y., Wang, C., Wang, Z., Dinh, C.-T., De Luna, P., Yu, Z., Rasouli, A.S., Brodersen, P., et al. (2020). Accelerated discovery of CO_2 electrocatalysts using active machine learning. Nature *581*, 178–183.

61. Putin, E., Asadulaev, A., Ivanenkov, Y., Aladinskiy, V., Sanchez-Lengeling, B., Aspuru-Guzik, A., and Zhavoronkov, A. (2018). Reinforced Adversarial Neural Computer for de Novo Molecular Design. J. Chem. Inf. Model. *58*, 1194–1204.

62. Popova, M., Isayev, O., and Tropsha, A. (2018). Deep reinforcement learning for de novo drug design. Sci. Adv. *4*, eaap7885.

63. Zhou, Z., Li, X., and Zare, R.N. (2017). Optimizing Chemical Reactions with Deep Reinforcement Learning. ACS Cent. Sci. *3*, 1337–1344.

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

65. Dong, Y., Li, D., Zhang, C., Wu, C., Wang, H., Xin, M., Cheng, J., and Lin, J. (2020). Inverse design of two-dimensional graphene/h-BN hybrids by a regressional and conditional GAN. Carbon *169*, 9–16.

66. Häse, F., Roch, L.M., and Aspuru-Guzik, A. (2019). Next-Generation Experimentation with Self-Driving Laboratories. Trends Chem. *1*, 282–291.

67. Sakiewicz, P., Piotrowski, K., Ober, J., and Karwot, J. (2020). Innovative artificial neural network approach for integrated biogas – wastewater treatment system modelling: Effect of plant operating parameters on process intensification. Renew. Sustain. Energy Rev. *124*, 109784.

68. Kumar, V., Kumar, A., Chhabra, D., and Shukla, P. (2019). Improved biobleaching of mixed hardwood pulp and process optimization using novel GA-ANN and GA-ANFIS hybrid statistical tools. Bioresour. Technol. *271*, 274–282.

69. Ding, Y., Zhang, Y., Ren, Y.M., Orkoulas, G., and Christofides, P.D. (2019). Machine learning-based modeling and operation for ALD of SiO2 thin-films using data from a multiscale CFD simulation. Chem. Eng. Res. Des. *151*, 131–145.

70. Lai, F., Sun, Z., Saji, S.E., He, Y., Yu, X., Zhao, H., Guo, H., and Yin, Z. (2021). Machine Learning-Aided Crystal Facet Rational Design with Ionic Liquid Controllable Synthesis. Small *17*, e2100024.

71. Mehr, S.H.M., Craven, M., Leonov, A.I., Keenan, G., and Cronin, L. (2020). A universal

system for digitization and automatic execution of the chemical synthesis literature. Science *370*, 101–108.

72. Huang, P.Y., Kurasch, S., Alden, J.S., Shekhawat, A., Alemi, A.A., McEuen, P.L., Sethna, J.P., Kaiser, U., and Muller, D.A. (2013). Imaging atomic rearrangements in two-dimensional silica glass: watching silica's dance. Science *342*, 224–227.

73. Silva, D.L., Campos, J.L.E., Fernandes, T.F.D., Rocha, J.N., Machado, L.R.P., Soares, E.M., Miquita, D.R., Miranda, H., Rabelo, C., Vilela Neto, O.P., et al. (2020). Raman spectroscopy analysis of number of layers in mass-produced graphene flakes. Carbon *161*, 181–189.

74. Hong, S., Nomura, K.I., Krishnamoorthy, A., Rajak, P., Sheng, C., Kalia, R.K., Nakano, A., and Vashishta, P. (2019). Defect Healing in Layered Materials: A Machine Learning-Assisted Characterization of MoS_2 Crystal Phases. J. Phys. Chem. Lett. *10*, 2739–2744.

75. Motevalli, B., Sun, B., and Barnard, A.S. (2020). Understanding and Predicting the Cause of Defects in Graphene Oxide Nanostructures Using Machine Learning. J. Phys. Chem. C *124*, 7404–7413.

76. Chen, X.-D., Xin, W., Jiang, W.-S., Liu, Z.-B., Chen, Y., and Tian, J.-G. (2016). High-Precision Twist-Controlled Bilayer and Trilayer Graphene. Adv. Mater. *28*, 2563–2570.

77. Borodinov, N., Tsai, W.-Y., Korolkov, V.V., Balke, N., Kalinin, S.V., and Ovchinnikova, O.S. (2020). Machine learning-based multidomain processing for texture-based image segmentation and analysis. Appl. Phys. Lett. *116*, 044103.

78. Zhu, J., Yin, Z., Yang, D., Sun, T., Yu, H., Hoster, H.E., Hng, H.H., Zhang, H., and Yan, Q. (2013). Hierarchical hollow spheres composed of ultrathin Fe2O3 nanosheets for lithium storage and photocatalytic water oxidation. Energy Environ. Sci. *6*, 987–993.

79. Cellini, F., Lavini, F., Berger, C., de Heer, W., and Riedo, E. (2019). Layer dependence of graphene-diamene phase transition in epitaxial and exfoliated few-layer graphene using machine learning. 2D Mater. *6*, 035043.

80. Ziatdinov, M., Dyck, O., Maksov, A., Li, X., Sang, X., Xiao, K., Unocic, R.R., Vasudevan, R., Jesse, S., and Kalinin, S.V. (2017). Deep Learning of Atomically Resolved Scanning Transmission Electron Microscopy Images: Chemical Identification and Tracking Local Transformations. ACS Nano *11*, 12742–12752.

81. Maksov, A., Dyck, O., Wang, K., Xiao, K., Geohegan, D.B., Sumpter, B.G., Vasudevan, R.K., Jesse, S., Kalinin, S.V., and Ziatdinov, M. (2019). Deep learning analysis of defect and phase evolution during electron beam-induced transformations in WS2. npj Comput. Mater. *5*, 12.

82. Madsen, J., Liu, P., Kling, J., Wagner, J.B., Hansen, T.W., Winther, O., and Schiøtz, J. (2018). A Deep Learning Approach to Identify Local Structures in Atomic-Resolution Transmission Electron Microscopy Images. Adv. Theory Simul. *1*, 1800037.

83. Lin, X., Si, Z., Fu, W., Yang, J., Guo, S., Cao, Y., Zhang, J., Wang, X., Liu, P., Jiang, K., and Zhao, W. (2018). Intelligent identification of two-dimensional nanostructures by machine-learning optical microscopy. Nano Res. *11*, 6316–6324.

84. Su, C., Yin, Z., Yan, Q.-B., Wang, Z., Lin, H., Sun, L., Xu, W., Yamada, T., Ji, X., Zettsu, N., et al. (2019). Waterproof molecular monolayers stabilize 2D materials. Proc. Natl. Acad. Sci. USA *116*, 20844–20849.

85. Yin, Z., Tordjman, M., Lee, Y., Vardi, A., Kalish, R., and Del Alamo, J.A. (2018). Enhanced transport in transistor by tuning transition-metal oxide electronic states interfaced with diamond. Sci. Adv. *4*, eaau0480.

86. Wastl, D.S., Weymouth, A.J., and Giessibl, F.J. (2014). Atomically resolved graphitic surfaces in air by atomic force microscopy. ACS Nano *8*, 5233–5239.

87. Iakovlev, I.A., Sotnikov, O.M., and Mazurenko, V.V. (2018). Supervised learning approach for recognizing magnetic skyrmion phases. Phys. Rev. B *98*, 174411.

88. Du, Y., Yin, Z., Zhu, J., Huang, X., Wu, X.-J., Zeng, Z., Yan, Q., and Zhang, H. (2012). A general method for the large-scale synthesis of uniform ultrathin metal sulphide nanocrystals. Nat. Commun. *3*, 1177.

89. Chua, C.K., and Pumera, M. (2014). Chemical reduction of graphene oxide: a synthetic chemistry viewpoint. Chem. Soc. Rev. *43*, 291–312.

90. Patra, T.K., Zhang, F., Schulman, D.S., Chan, H., Cherukara, M.J., Terrones, M., Das, S., Narayanan, B., and Sankaranarayanan, S.K.R.S. (2018). Defect Dynamics in 2-D MoS_2 Probed by Using Machine Learning, Atomistic Simulations, and High-Resolution Microscopy. ACS Nano *12*, 8006–8016.

91. Bartel, C.J., Trewartha, A., Wang, Q., Dunn, A., Jain, A., and Ceder, G. (2020). A critical examination of compound stability predictions from machine-learned formation energies. npj Comput. Mater. *6*, 97.

92. Barnard, A.S., and Opletal, G. (2019). Predicting structure/property relationships in multi-dimensional nanoparticle data using t-distributed stochastic neighbour embedding and machine learning. Nanoscale *11*, 23165–23172.

93. Fernandez, M., Shi, H., and Barnard, A.S. (2015). Quantitative Structure-Property Relationship Modeling of Electronic Properties of Graphene Using Atomic Radial Distribution Function Scores. J. Chem. Inf. Model. *55*, 2500–2506.

94. Fernandez, M., Shi, H., and Barnard, A.S. (2016). Geometrical features can predict electronic properties of graphene nanoflakes. Carbon *103*, 142–150.

95. Nemnes, G.A., Mitran, T.L., and Manolescu, A. (2019). Gap Prediction in Hybrid Graphene-Hexagonal Boron Nitride Nanoflakes Using Artificial Neural Networks. J. Nanomater. *2019*, 6960787.

96. Mishra, A., Satsangi, S., Rajan, A.C., Mizuseki, H., Lee, K.-R., and Singh, A.K. (2019). Accelerated Data-Driven Accurate Positioning of the Band Edges of MXenes. J. Phys. Chem. Lett. *10*, 780–785.

97. Cherukara, M.J., Narayanan, B., Kinaci, A., Sasikumar, K., Gray, S.K., Chan, M.K.Y., and Sankaranarayanan, S.K.R.S. (2016). Ab Initio-Based Bond Order Potential to Investigate Low Thermal Conductivity of Stanene Nanostructures. J. Phys. Chem. Lett. *7*, 3752–3759.

98. Gu, X., and Zhao, C.Y. (2019). Thermal conductivity of single-layer MoS2(1−x)Se2x alloys from molecular dynamics simulations with a machine-learning-based interatomic potential. Comput. Mater. Sci. *165*, 74–81.

99. Mortazavi, B., Novikov, I.S., Podryabinkin, E.V., Roche, S., Rabczuk, T., Shapeev, A.V., and Zhuang, X. (2020). Exploring phononic properties of two-dimensional materials using machine learning interatomic potentials. Appl. Mater. Today *20*, 100685.

100. Mortazavi, B., Shojaei, F., Shahrokhi, M., Azizi, M., Rabczuk, T., Shapeev, A.V., and Zhuang, X. (2020). Nanoporous C3N4, C3N5 and C3N6 nanosheets; novel strong semiconductors with low thermal conductivities and appealing optical/electronic properties. Carbon *167*, 40–50.

101. Raeisi, M., Mortazavi, B., Podryabinkin, E.V., Shojaei, F., Zhuang, X., and Shapeev, A.V. (2020). High thermal conductivity in semiconducting Janus and non-Janus diamanes. Carbon *167*, 51–61.

102. Wang, X., Hong, Y., Wang, M., Xin, G., Yue, Y., and Zhang, J. (2019). Mechanical properties of molybdenum diselenide revealed by molecular dynamics simulation and support vector machine. Phys. Chem. Chem. Phys. *21*, 9159–9167.

103. Firestein, K.L., von Treifeldt, J.E., Kvashnin, D.G., Fernando, J.F.S., Zhang, C., Kvashnin, A.G., Podryabinkin, E.V., Shapeev, A.V., Siriwardena, D.P., Sorokin, P.B., and Golberg, D. (2020). Young's Modulus and Tensile Strength of Ti_3C_2 MXene Nanosheets As Revealed by *In Situ* TEM Probing, AFM Nanomechanical Mapping, and Theoretical Calculations. Nano Lett. *20*, 5900–5908.

104. Neumayer, S.M., Susner, M.A., McGuire, M.A., Pantelides, S.T., Kalnaus, S., Maksymovych, P., and Balke, N. (2021). Lowering of Tc in Van Der Waals Layered Materials Under In-Plane Strain. IEEE Trans. Ultrason. Ferroelectr. Freq. Control *68*, 253–258.

105. Rhone, T.D., Chen, W., Desai, S., Torrisi, S.B., Larson, D.T., Yacoby, A., and Kaxiras, E. (2020). Data-driven studies of magnetic two-dimensional materials. Sci. Rep. *10*, 15795.

106. Melisande Fischer, J., Hunter, M., Hankel, M., Searles, D.J., Parker, A.J., and Barnard, A.S. (2020). Accurate prediction of binding energies for two-dimensional catalytic materials using machine learning. ChemCatChem *12*, 5109–5120.

107. Marchwiany, M.E., Birowska, M., Popielski, M., Majewski, J.A., and Jastrzębska, A.M. (2020). Surface-Related Features Responsible for Cytotoxic Behavior of MXenes Layered Materials Predicted with Machine Learning Approach. Materials (Basel) *13*, 3083.

108. Yin, Z., Zhang, X., Cai, Y., Chen, J., Wong, J.I., Tay, Y.-Y., Chai, J., Wu, J., Zeng, Z., Zheng, B., et al. (2014). Preparation of MoS2-MoO3 hybrid nanomaterials for light-emitting diodes. Angew. Chem. Int. Ed. Engl. *53*, 12560–12565.

109. Liu, J., Yin, Z., Cao, X., Zhao, F., Wang, L., Huang, W., and Zhang, H. (2013). Fabrication of flexible, all-reduced graphene oxide non-volatile memory devices. Adv. Mater. *25*, 233–238.

110. Yin, Z., Zeng, Z., Liu, J., He, Q., Chen, P., and Zhang, H. (2013). Memory devices using a mixture of MoS_2 and graphene oxide as the active layer. Small *9*, 727–731.

111. Tebyetekerwa, M., Zhang, J., Liang, K., Duong, T., Neupane, G.P., Zhang, L., Liu, B., Truong, T.N., Basnet, R., Qiao, X., et al. (2019). Quantifying Quasi-Fermi Level Splitting and Mapping its Heterogeneity in Atomically Thin Transition Metal Dichalcogenides. Adv. Mater. *31*, e1900522.

112. Shi, H., Barnard, A.S., and Snook, I.K. (2012). High throughput theory and simulation of nanomaterials: exploring the stability and electronic properties of nanographene. J. Mater. Chem. *22*, 18119–18123.

113. Ward, L., Dunn, A., Faghaninia, A., Zimmermann, N.E.R., Bajaj, S., Wang, Q., Montoya, J., Chen, J., Bystrom, K., Dylla, M., et al. (2018). Matminer: An open source toolkit for materials data mining. Comput. Mater. Sci. *152*, 60–69.

114. Behler, J. (2016). Perspective: Machine learning potentials for atomistic simulations. J. Chem. Phys. *145*, 170901.

115. Zunger, A. (2019). Beware of plausible predictions of fantasy materials. Nature *566*, 447–449.

116. Akinwande, D., Brennan, C.J., Bunch, J.S., Egberts, P., Felts, J.R., Gao, H., Huang, R., Kim, J.-S., Li, T., Li, Y., et al. (2017). A review on mechanics and mechanical properties of 2D materials—Graphene and beyond. Extreme Mech. Lett. *13*, 42–77.