npj | Computational Materials

www.nature.com/npjcompumats

REVIEW ARTICLE OPEN

Recent advances and applications of machine learning in solid-state materials science

Jonathan Schmidt[1], Mário R. G. Marques[1], Silvana Botti[2] and Miguel A. L. Marques[1]

One of the most exciting tools that have entered the material science toolbox in recent years is machine learning. This collection of statistical methods has already proved to be capable of considerably speeding up both fundamental and applied research. At present, we are witnessing an explosion of works that develop and apply machine learning to solid-state systems. We provide a comprehensive overview and analysis of the most recent research in this topic. As a starting point, we introduce machine learning principles, algorithms, descriptors, and databases in materials science. We continue with the description of different machine learning approaches for the discovery of stable materials and the prediction of their crystal structure. Then we discuss research in numerous quantitative structure–property relationships and various approaches for the replacement of first-principle methods by machine learning. We review how active learning and surrogate-based optimization can be applied to improve the rational design process and related examples of applications. Two major questions are always the interpretability of and the physical understanding gained from machine learning models. We consider therefore the different facets of interpretability and their importance in materials science. Finally, we propose solutions and future research paths for various challenges in computational materials science.

npj Computational Materials (2019)5:83 ; https://doi.org/10.1038/s41524-019-0221-0

INTRODUCTION

In recent years, the availability of large datasets combined with the improvement in algorithms and the exponential growth in computing power led to an unparalleled surge of interest in the topic of machine learning. Nowadays, machine learning algorithms are successfully employed for classification, regression, clustering, or dimensionality reduction tasks of large sets of especially high-dimensional input data.[1] In fact, machine learning has proved to have superhuman abilities in numerous fields (such as playing go,[2] self driving cars,[3] image classification,[4] etc). As a result, huge parts of our daily life, for example, image and speech recognition,[5,6] web-searches,[7] fraud detection,[8] email/spam filtering,[9] credit scores,[10] and many more are powered by machine learning algorithms.

While data-driven research, and more specifically machine learning, have already a long history in biology[11] or chemistry,[12] they only rose to prominence recently in the field of solid-state materials science.

Traditionally, experiments used to play the key role in finding and characterizing new materials. Experimental research must be conducted over a long time period for an extremely limited number of materials, as it imposes high requirements in terms of resources and equipment. Owing to these limitations, important discoveries happened mostly through human intuition or even serendipity.[13] A first computational revolution in materials science was fueled by the advent of computational methods,[14] especially density functional theory (DFT),[15,16] Monte Carlo simulations, and molecular dynamics, that allowed researchers to explore the phase and composition space far more efficiently. In fact, the combination of both experiments and computer simulations has allowed to cut substantially the time and cost of materials design.[17–20] The constant increase in computing power and the development of more efficient codes also allowed for computational high-throughput studies[21] of large material groups in order to screen for the ideal experimental candidates. These large-scale simulations and calculations together with experimental high-throughput studies[22–25] are producing an enormous amount of data making possible the use of machine learning methods to materials science.

As these algorithms start to find their place, they are heralding a second computational revolution. Because the number of possible materials is estimated to be as high as a googol (10^{100}),[26] this revolution is doubtlessly required. This paradigm change is further promoted by projects like the materials genome initiative (Materials genome initiative) that aim to bridge the gap between experiment and theory and promote a more data-intensive and systematic research approach. A multitude of already successful machine learning applications in materials science can be found, e.g., the prediction of new stable materials,[27–35] the calculation of numerous material properties,[36–51] and the speeding up of first-principle calculations.[52]

Machine learning algorithms have already revolutionized other fields, such as image recognition. However, the development from the first perceptron[53,54] up to modern deep convolutional neural networks was a long and tortuous process. In order to produce significant results in materials science, one necessarily has not only to play to the strength of machine learning techniques but also apply the lessons already learned in other fields.

[1]Institut für Physik, Martin-Luther-Universität, 06120 Halle-Wittenberg, Halle (Saale), Germany and [2]Institut für Festkörpertheorie und -optik, Friedrich-Schiller-Universität Jena, Max-Wien-Platz 1, 07743 Jena, Germany
Correspondence: Miguel A. L. Marques (miguel.marques@physik.uni-halle.de)

Received: 26 February 2019 Accepted: 17 July 2019
Published online: 08 August 2019

As the introduction of machine learning methods to materials science is still recent, a lot of published applications are quite basic in nature and complexity. Often they involve fitting models to extremely small training sets or even applying machine learning methods to composition spaces that could possibly be mapped out in hundreds of CPU hours. It is of course possible to use machine learning methods as a simple fitting procedure for small low-dimensional datasets. However, this does not play to their strength and will not allow us to replicate the success machine learning methods had in other fields.

Furthermore, and as always when entering a different field of science, nomenclature has to be applied correctly. One example is the expression "deep learning", which is responsible for a majority of the recent success of machine learning methods (e.g., in image recognition and natural language processing[55]). It is of course tempting to describe one's work as deep learning. However, denoting neural networks with one or two fully connected hidden layer as deep learning[56] is confusing for researchers new to the topic, and it misrepresents the purpose of deep-learning algorithms. The success of deep learning is rooted in the ability of deep neural networks to learn descriptors of data with different levels of abstraction without human intervention.[55,57] This is, of course, not the case in two-layer neural networks.

One of the major criticisms of machine learning algorithms in science is the lack of novel laws, understanding, and knowledge arising from their use. This comes from the fact that machine learning algorithms are often treated as black boxes, as machine-built models are too complex and alien for humans to understand. We will discuss the validity of the criticism and different approaches to this challenge.

Finally, there have already been a number of excellent reviews of materials informatics and machine learning in materials science in general,[13,58–62] as well as some other covering specifically machine learning in the chemical sciences,[63] in materials design of thermoelectrics and photovoltaics,[64] in the development of lithium-ion batteries,[65] and in atomistic simulations.[66] However, owing to the explosion in the number of works using machine learning, an enormous amount of research has already been published since the past reviews and the research landscape has quickly transformed.

Here we concentrate on the various applications of machine learning in solid-state materials science (especially the most recent ones) and discuss and analyze them in detail. As a starting point, we provide an introduction to machine learning, and in particular to machine learning principles, algorithms, descriptors, and databases in materials science. We then review numerous applications of machine learning in solid-state materials science: the discovery of new stable materials and the prediction of their structure, the machine learning calculation of material properties, the development of machine learning force fields for simulations in material science, the construction of DFT functionals by machine learning methods, the optimization of the adaptive design process by active learning, and the interpretability of, and the physical understanding gained from, machine learning models. Finally, we discuss the challenges and limitations machine learning faces in materials science and suggest a few research strategies to overcome or circumvent them.

BASIC PRINCIPLES OF MACHINE LEARNING

Machine learning algorithms aim to optimize the performance of a certain task by using examples and/or past experience.[67] Generally speaking, machine learning can be divided into three main categories, namely, supervised learning, unsupervised learning, and reinforcement learning.

Supervised machine learning is based on the same principles as a standard fitting procedure: it tries to find the unknown function that connects known inputs to unknown outputs. This desired result for unknown domains is estimated based on the extrapolation of patterns found in the labeled training data. Unsupervised learning is concerned with finding patterns in unlabeled data, as, e.g., in the clustering of samples. Finally, reinforcement learning treats the problem of finding optimal or sufficiently good actions for a situation in order to maximize a reward.[68] In other words, it learns from interactions.

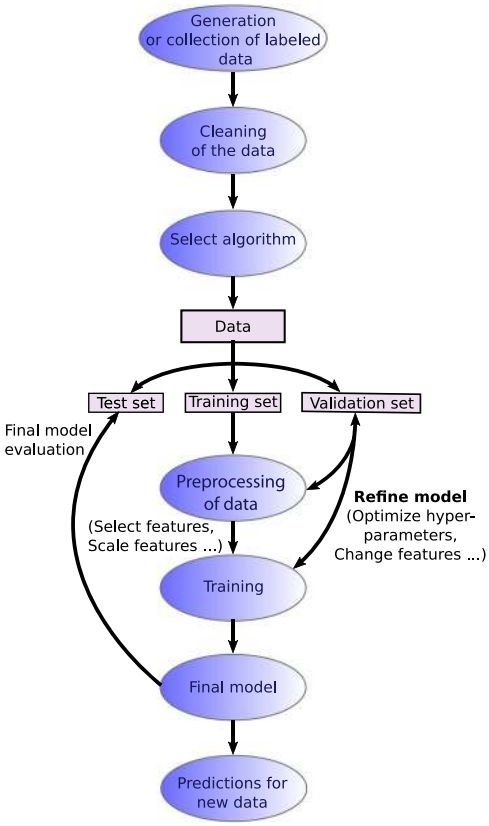

Fig. 1 Supervised learning workflow

Finally, halfway between supervised and unsupervised learning lies semi-supervised learning. In this case, the algorithm is provided with both unlabeled as well as labeled data. Techniques of this category are particularly useful when available data are incomplete and to learn representations.[69]

As supervised learning is by far the most widespread form of machine learning in materials science, we will concentrate on it in the following discussion. Figure 1 depicts the workflow applied in supervised learning. One generally chooses a subset of the relevant population for which values of the target property are known or creates the data if necessary. This process is accompanied by the selection of a machine learning algorithm that will be used to fit the desired target quantity. Most of the work consists in generating, finding, and cleaning the data to ensure that it is consistent, accurate, etc. Second, it is necessary to decide how to map the properties of the system, i.e., the input for the model, in a way that is suitable for the chosen algorithm. This implies to translate the raw information into certain features that will be used as inputs for the algorithm. Once this process is finished, the model is trained by optimizing its performance, usually measured through some kind of cost function. Usually this entails the adjustment of hyperparameters that control the training process, structure, and properties of the model. The data are split into various sets. Ideally, a validation dataset separate from the test and training sets is used for the optimization of the hyperparameters.

Every machine learning application has to consider the aspects of overfitting and underfitting. The reason for underfitting usually lies either in the model, which lacks the ability to express the complexity of the data, or in the features, which do not adequately describe the data. This inevitably leads to a high training error. On the other hand, an overfitted model interprets part of the noise in the training data as relevant information, therefore failing to reliably predict new data. Usually, an overfitted model contains more free parameters than the number required to capture the complexity of the training data. In order to avoid overfitting, it is essential to monitor during training not only the training error but also the error of the validation set. Once the validation error stops decreasing, a machine learning model can start to overfit. This problem is also discussed as the bias-variance trade off in machine learning.[70,71] In this context, the bias is an error based on wrong assumptions in the trained model, while high variance is the error resulting from too much sensitivity to noise in the training data. As such, underfitted models possess high bias while overfitted models have high variance.

Before the model is ready for applications, it has to be evaluated on previously unseen data, denoted as test set, to estimate its generalization and extrapolation ability.

Different methods ranging from a simple holdout, over *k*-fold cross-validation, leave-one-out cross-validation, Monte Carlo cross-validation,[72] up to leave-one-cluster-out cross-validation[73] can be used for the evaluation. All these methods rely on keeping some data hidden from the model during the training process. For a simple holdout, this is just performed once, while for *k*-fold cross-validation the dataset is separated into *k* equally sized sets. The algorithm is trained with all but one of these *k* subsets, which is used for testing. Finally, the process is repeated for every subset. For leave-one-out cross-validation, each sample is left out of the training set once and the model is evaluated for that sample. It has to be noted that research in chemistry has shown that this form of cross-validation is insufficient to evaluate adequately the predictive performance of quantitative structure–property relationship and should therefore be avoided.[74,75] Monte Carlo cross-validation is similar to *k*-fold cross-validation in the sense that the training and test set are randomly chosen. However, here the size of the training/test set is chosen independently from the number of folds. While this can be advantageous, it also means that a sample is not guaranteed to be in the test/training set. Leave-one-cluster-out cross-validation[73] was specifically developed for materials science and estimates the ability of the machine learning model to extrapolate to novel groups of materials that were not present in the training data. Depending on the target quantity, this allows for a more realistic evaluation and a better understanding of the limitations of the machine learning model. Leave-one-cluster-out cross-validation removes a cluster of materials and then considers the error for predictions of the materials belonging to the removed cluster. This is, for example, consistent with the finding in ref. [76] that models trained on superconductors with a specific superconducting mechanism do not have any predictive ability for superconductors with other mechanisms.

Before discussing various applications of machine learning in materials science, we will give an overview of the different descriptors, algorithms, and databases used in materials informatics.

Databases

Machine learning in materials science is mostly concerned with supervised learning. The success of such methods depends mainly on the amount and quality of data that is available, and this turns out to be one of the major challenges in material informatics.[77] This is especially problematic for target properties that can only be determined experimentally in a costly fashion (such as the critical temperature of superconductors—see section "Prediction of material properties—superconductivity"). For this reason, databases such as the materials project,[78] the inorganic crystal structure database,[79] and others (Materials genome initiative, The NOMAD archive, Supercon, National Institute of Materials Science 2011)[80–92] that contain information on numerous properties of known materials are essential for the success of materials informatics.

In order for these databases and for materials informatics to thrive, a FAIR treatment of data[93] is absolutely required. A FAIR treatment encompasses the four principles: findability, accessibility, interoperability, and repurposability.[94] In other words, researchers from different disciplines should be able to find and access data, as well as the corresponding metadata, in a commonly accepted format. This allows the application of the data for new purposes.

Traditionally, negative results are often discarded and left unpublished. However, as negative data are often just as important for machine learning algorithms as positive results,[28,95] a cultural adjustment toward the publication of unsuccessful research is necessary. In some disciplines with a longer tradition of data-based research (like chemistry), such databases already exist.[95] In a similar vein, data that emerges as a side product but are not essential for a publication are often left unpublished. This eventually results in a waste of resources as other researchers are then required to repeat the work. In the end, every single discarded calculation will be sorely missed in future machine learning applications.

Features

A pivotal ingredient of a machine learning algorithm is the representation of the data in a suitable form. Features in material science have to be able to capture all the relevant information, necessary to distinguish between different atomic or crystal environments.[96] The process itself, denoted as feature extraction or engineering, might be as simple as determining atomic numbers, might involve complex transformations such as an expansion of radial distribution functions (RDFs) in a certain basis, or might require aggregations based on statistics (e.g., average over features or the calculation of their maximum value). How much processing is required depends strongly on the algorithm. For some methods, such as deep learning, the feature extraction can be considered as part of the model.[97] Naturally, the best choice for the representation depends on the target quantity and the variety of the space of occurrences. For completeness, we have to mention that the cost of feature extraction and of target quantity evaluation must never be comparable.

Ideally, descriptors should be uncorrelated, as an abundant number of correlated features can hinder the efficiency and accuracy of the model. When this happens, further feature selection is necessary to circumvent the curse of dimensionality,[98] simplify models, and improve their interpretability as well as training efficiency. For example, several elemental properties such as the period and group in the periodic table, ionization potential, and covalent radius, can be used as features to model formation energies or distances to the convex hull of stability. However, it was shown that, to obtain acceptable accuracies, often only the period and the group are required.[99]

Having described the general properties of descriptors, we will proceed with a listing of the most used features in materials science. Without a doubt, the most studied type of features in this field are the ones related to the fitting of potential energy surfaces. In principle, the nuclear charges and the atomic positions are sufficient features, as the Hamiltonian of a system is usually fully determined by these quantities. In practice, however, while Cartesian coordinates might provide an unambiguous description of the atomic positions, they do not make a suitable descriptor, as

the list of coordinates of a structure are ordered arbitrarily and the number of such coordinates varies with the number of atoms. The latter is a problem, as most machine learning models require a fixed number of features as an input. Therefore, to describe solids and large clusters, the number of interacting neighbors has to be allowed to vary without changing the dimensionality of the descriptor. In addition, a lot of applications require that the features are continuous and differentiable with respect to atomic positions.

A comprehensive study on features for atomic potential energy surfaces can be found in the review of Bartó et al.[100]. Important points mentioned in their work are: (i) the performance of the model and its ability to differentiate between different structures do not depend directly on the descriptors but on the similarity measurement between them; (ii) the quality of the descriptors is related to the differentiability with respect to the movement of the atoms, completeness of the representation, and invariance to the basis symmetries of physics (rotation, reflection, translation, and permutation of atoms of the same species). For clarification, a set of invariant descriptors q_i, which uniquely determines an atomic environment up to symmetries, is defined as complete. An overcomplete set is then a set that includes more features than necessary.

Simple representations that show shortcomings as features are transformations of pairwise distances,[101–103] Weyl matrices,[104] and *Z*-matrices.[105] Pairwise distances (and also reciprocal or exponential transformations of these) only work for a fixed number of atoms and are not unique under permutation of atoms. The constrain on the number of atoms is also present for polynomials of pairwise distances. Histograms of pairwise atomic distances are non-unique: if no information on the angles between the atoms is given, of if the ordering of the atoms is unknown, it might be possible to construct at least two different structures with the same features. Weyl matrices are defined by the inner product between neighboring atoms positions, forming an overcomplete set, while permutations of the atoms change the order of the rows and columns. Finally, *Z*-matrices or internal coordinate representations are not invariant under permutations of atoms.

In 2012, Rupp et al.[106] introduced a representation for molecules based on the Coulomb repulsion between atoms I and J and a polynomial fit of atomic energies to the nuclear charge

$$M_{IJ} = \begin{cases} 0.5Z_I^{2.4} & \text{for } I = J \\ \frac{Z_I Z_J}{|\boldsymbol{R}_I - \boldsymbol{R}_J|} & \text{for } I \neq J \end{cases}. \tag{1}$$

The ordered eigenvalues (ε) of these "Coulomb matrices" are then used to measure the similarity between two molecules.

$$d(\varepsilon, \varepsilon\prime) = \sqrt{\sum_i \left|\varepsilon_i - \varepsilon_i'\right|^2}. \tag{2}$$

Here, if the number of atoms is not the same in both systems, ε is extended by zeros. In this representation, symmetrically equivalent atoms contribute equally to the feature function, the diagonalized matrices are invariant with respect to permutations and rotations, and the distance d is continuous under small variations of charge or interatomic distances. Unfortunately, this representation is not complete and does not uniquely describe every system. The incompleteness derives from the fact that not all degrees of freedom are taken into account when comparing two systems. The non-uniqueness can be demonstrated using as an example acetylene (C_2H_2).[107] In brief, distortions of this molecule can lead to several geometries that are described by the same Coulomb matrix.

Faber et al.[108] presented three distinct ways to extend the Coulomb matrix representation to periodic systems. The first of these features consists of a matrix where each element represents the full Coulomb interaction between two atoms and all their infinite repetitions in the lattice. For example:

$$X_{ij} = \frac{1}{N} Z_i Z_j \sum_{k,l} \varphi(|\boldsymbol{R}_k - \boldsymbol{R}_l|), \tag{3}$$

where the sum over k (l) is taken over the atom i (j) in the unit cell and its N closest equivalent atoms. However, as this double sum has convergence issues, one has to resort to the Ewald trick: X_{ij} is divided into a constant and two rapidly converging sums, one for the long-range interaction and another for the short-range interaction. Another extension by Faber et al. considers electrostatic interactions between the atoms in the unit cell and the atoms in the N closest unit cells. In addition, the long-range interaction is replaced by rapidly decaying interaction. In their final extension, the Coulomb interaction in the usual matrix is replaced by a potential that is symmetric with respect to the lattice vectors.

In the same line of work, Schütt et al.[109] extended the Coulomb matrix representation by combining it with the Bravais matrix. Unfortunately, this representation is plagued by a degeneracy problem that comes from the arbitrary choice of the coordinate system in which the Bravais matrix is written. Another representation proposed by Schütt et al. is the so called partial radial distribution function, which considers the density of atoms β in a shell of width dr and radius r centered around atom α (see Fig. 2):

$$g_{\alpha\beta}(r) = \frac{1}{N_\alpha V_r} \sum_i^{N_\alpha} \sum_j^{N_\beta} \theta(d_{\alpha_i\beta_j} - r)\theta(r + dr - d_{\alpha_i\beta_j}). \tag{4}$$

Here N_α and N_β are the number of atom of types α and β, V_r is the volume of the shell, and $d_{\alpha\beta}$ are the pairwise distances between two atom types.

Another form for representing the local structural environment was proposed by Behler and Parrinelo.[110] Their descriptors[111] involve an invariant set of atom-centered radial

$$G_i^{\mathrm{r}}(\{\boldsymbol{R}_i\}) = \sum_{j \neq i}^{\text{neighbors}} g^{\mathrm{r}}(R_{ij}), \tag{5}$$

and angular symmetry functions

$$G_i^{\mathrm{a}}(\{\boldsymbol{R}_i\}) = \sum_{j \neq i}^{\text{neighbors}} g^{\mathrm{a}}(\theta_{ijk}), \tag{6}$$

where θ_{ijk} is the angle between $\boldsymbol{R_j} - \boldsymbol{R}_i$ and $\boldsymbol{R_k} - \boldsymbol{R}_i$. While the radial functions G_i^{r} contain information on the interaction between pairs of atoms within a certain radius, the angular functions G_i^{a} contains additional information on the distribution of the bond angles θ_{ijk}. Examples for atom-centered symmetry functions are

$$G_i^{\mathrm{r}} = \sum_{j \neq i}^{\text{neighbors}} f_{\mathrm{c}}(R_{ij}) \mathrm{e}^{-\eta(R_{ij} - R_{\mathrm{s}})^2} \tag{7}$$

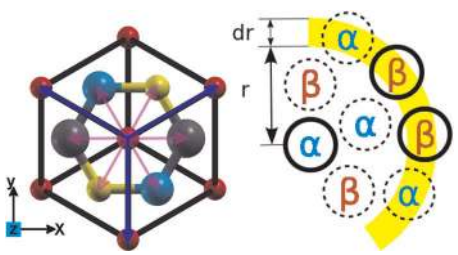

Fig. 2 Two crystal structure representations. (Left) A unit cell with the Bravais vectors (blue) and base (pink) represented. (Right) Depiction of a shell of the discrete partial radial distribution function $g_{\alpha\beta}(r)$ with width dr. (Reprinted with permission from ref. [109]. Copyright 2014 American Physical Society

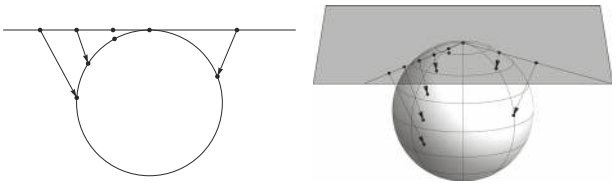

Fig. 3 Mapping of a flat space in one and two dimensions onto the surface of a sphere in one higher dimension. (Reprinted with permission from ref. [100]. Copyright 2013 American Physical Society.)

and

$$G_i^{\mathrm{a}} = 2^{1-\zeta} \sum_{\substack{jk \\ i\neq j\neq k}}^{\text{neighbors}} (1+\lambda\cos\theta_{ijk})^{\zeta} \mathrm{e}^{-\eta(R_{ij}^2+R_{ik}^2+R_{jk}^2)} \times f_{\mathrm{c}}(R_{ij}) f_{\mathrm{c}}(R_{ik}) f_{\mathrm{c}}(R_{jk}). \tag{8}$$

Here f_c is a cutoff function, leading to the neglect of interactions between atoms beyond a certain radius R_c. Furthermore, η controls the width of the Gaussians, R_s is just a parameter that shifts the Gaussians, λ determines the positions of the extrema of the cosine, and ζ controls the angular resolution. The sum over neighbors enforces the permutation invariance of these symmetry functions. Usually, 20–100 symmetry functions are used per atom, constructed by varying the parameters above. Beside atom centered, these functions can also be pair centered.[112]

A generalization of the atom-centered pairwise descriptor of Behler was proposed by Seko et al.[113] It consists of simple basis functions constructed from the multinomial expansion of the product between a cutoff function (f_c) and an analytical pairwise function (f_n) (for example, Gaussian, cosine, Bessel, Neumann, polynomial, or Gaussian-type orbital functions)

$$b_{n,p}^{i,j} = \left[\sum_k f_n(R_{jk}^i)\cdot f_{\mathrm{c}}(R_{jk}^i)\right]^p, \tag{9}$$

where p is a positive integer, and R_{jk}^i indicates the distance between atoms j and k of structure i. The descriptor then uses the sum of these basis functions over all the atoms in the structure $\left(\sum_j b_{n,p}^{i,j}\right)$.

A similar type of descriptor is the angular Fourier series (AFS),[100] which consists of a collection of orthogonal polynomials, like the Chebyshev polynomials $T_l(\cos\theta) = \cos(l\theta)$, and radial functions

$$\mathrm{AFS}_{nl} = \sum_{i,j>i} g_n(r_i) g_n(r_j) \cos(l\theta_{ij}). \tag{10}$$

These radial functions are expansions of cubic or higher-order polynomials

$$g_n(r) = \sum_a W_{na}\phi_a(r), \tag{11}$$

where

$$\phi_a(r) = (r_{\mathrm{c}} - r)^{a+2}/N_a. \tag{12}$$

A different approach for atomic environment features was proposed by Bartok et al.[100,114] and leads to the power spectrum and the bispectrum. The approach starts with the generation of an atomic neighbor density function

$$\rho(\mathbf{r}) = \delta(\mathbf{r}_0) + \sum_i \delta(\mathbf{r}-\mathbf{r}_i), \tag{13}$$

which is projected onto the surface of a four-dimensional sphere with radius r_0. As an example, Fig. 3 depicts the projection for 1 and 2 dimensions. Then the hyperspherical harmonic functions $U_{m'm}^j$ can be used to represent any function ρ defined on the surface of a four-dimensional sphere[115,116]

$$\rho = \sum_{j=0}^{\infty} \sum_{m,m'=-j}^{j} c_{m'm}^j U_{m'm}^j. \tag{14}$$

Combining these with the rotation operator and the transformation of the expansion coefficients under rotation leads to the formula

$$P_j = \sum_{m',m=-j}^{j} c_{m'm}^{j*} c_{m'm}^j \tag{15}$$

for the SO(4) power spectrum. On the other hand, the bispectrum is given by

$$B_{j_1 j_2 j} = \sum_{m_1',m_1=-j_1}^{j_1} c_{m_1',m_1}^{j_1} \sum_{m_2',m_2=-j_2}^{j_2} c_{m_2',m_2}^{j_2} \times \sum_{m',m=-j}^{j} C_{m\,m_1\,m_2}^{j\,j_1\,j_2} C_{m'\,m_1'\,m_2'}^{j\,j_1\,j_2} c_{m'm}^{j*}, \tag{16}$$

where $C_{m\,m_1\,m_2}^{j\,j_1\,j_2}$ are the Clebsch–Gordon coefficients of SO(4). We note that the representations above are truncated, based on the band limit $j_{\max}$ in the expansion.

Finally, one of the most successful atomic environment features is the following similarity measurement

$$K(\rho,\rho') = \left[\frac{k(\rho,\rho')}{\sqrt{k(\rho,\rho)k(\rho',\rho')}}\right]^{\zeta} \tag{17}$$

also known as the smooth overlap of atomic positions (SOAP) kernel.[100] Here ζ is a positive integer that enhances the sensitivity of the kernel to changes on the atomic positions and ρ is the atomic neighbor density function, which is constructed from a sum of Gaussians, centered on each neighbor:

$$\rho(\mathbf{r}) = \sum_i \mathrm{e}^{-\alpha|\mathbf{r}-\mathbf{r}_i|^2}. \tag{18}$$

In practice, the function ρ is then expanded in terms of the spherical harmonics. In addition, $k(\rho, \rho')$ is a rotationally invariant kernel, defined as the overlap between an atomic environment and all rotated environments:

$$k(\rho,\rho') = \int d\hat{R} \int d\mathbf{r}\, \rho(\mathbf{r})\rho'(\hat{R}\mathbf{r}). \tag{19}$$

The normalization factor $\sqrt{k(\rho,\rho)k(\rho',\rho')}$ ensures that the overlap of an environment with itself is one.

The SOAP kernel can be perceived as a three-dimensional generalization of the radial atom-centered symmetry functions and is capable of characterizing the entire atomic environment at once. It was shown to be equivalent to using the power or bispectrum descriptor with a dot-product covariance kernel and Gaussian neighbor densities.[100]

A problem with the above descriptors is that their number increases quadratically with the number of chemical species. Inspired by the Behler symmetry functions and the SOAP method, Artrith et al.[117] devised a conceptually simple descriptor whose dimension is constant with respect to the number species. This is achieved by defining the descriptor as the union between two sets of invariant coordinates, one that maps the atomic positions (or structure) and another for the composition. Both of these mappings consist of the expansion coefficients of the RDFs

$$\mathrm{RDF}_i(r) = \sum_a c_a^{\mathrm{RDF}} \phi_a(r) \quad \text{for} \quad 0 \le r \le R_{\mathrm{c}} \tag{20}$$

and angular distribution functions (ADF)

$$\mathrm{ADF}_i(\theta) = \sum_a c_a^{\mathrm{ADF}} \phi_a(\theta) \quad \text{for} \quad 0 \le r \le R_{\mathrm{c}}. \tag{21}$$

in a complete basis set ϕ_a (like the Chebyshev 94% average cross-

validation error). The advantage of stability

$$c_a^{\mathrm{RDF}} = \sum_{R_j} \phi_a(R_{ij}) f_c(R_{ij}) w_{tj} \tag{22}$$

and

$$c_a^{\mathrm{ADF}} = \sum_{R_j, R_k} \phi_a(\theta_{ijk}) f_c(R_{ij}) f_c(R_{ij}) w_{tj} w_{tk}. \tag{23}$$

Here f_c is a cut-off function that limits the range of the interactions. The weights w_{tj} and w_{tk} take the value of one for the structure maps, while the weights for the compositional maps depend on the chemical species, according to the pseudo-spin convention of the Ising model. By limiting the descriptor to two and three body interactions, i.e., radial and angular contributions, this method maintains the simple analytic nature of the Behler–Parrinelo approach. Furthermore, it allows for an efficient implementation and differentiation, while systematic refinement is assured by the expansion in a complete basis set.

Sanville et al.[118] used a set of vectors, each of which describes a five-atom chain found in the system. This information includes distances between the five atoms, angles, torsion angles, and functions of the bond screening factors.[119]

The simplex representation of a molecular structure of Kuz'min et al.[120,121] consists in representing a molecule as a system of different simplex descriptors, i.e., a system of different tetratromic fragments. These descriptors become consecutively more detailed with the increase of the dimension of the molecule representation. The simplex descriptor at the one-dimensional (1D) level consists on the number of combinations of four atoms for a given composition. At the two-dimensional (2D) level, the topology is also taken into account, while at the 3D level, the descriptor is the number of simplexes of fixed composition, topology, chirality, and symmetry. The extension of this methodology to bulk materials was proposed by Isayev et al.[122] and counts bounded and unbounded simplexes (see Fig. 4). While a bonded simplex characterizes only a single component of the mixture, unbounded simplexes can describe up to four components of the unit cell.

Isayef et al.[41] also adapted property-labeled material fragments[123] to solids. The structure of the material is encoded in a graph that defines the connectivity within the material based on its Voronoi tessellation[124,125] (see Fig. 5). Only strong bonding interactions are considered. Two atoms are seen as connected only when they share a Voronoi face and the interatomic distance does not exceed the sum of the Cordero covalent bond lengths.[126] In the graph, the nodes correspond to the chemical elements, which are identified through a plethora of elemental properties, like Mendeleev group, period, thermal conductivity, covalent radius, etc. The full graph is divided into subgraphs that correspond to the different fragments. In addition, information

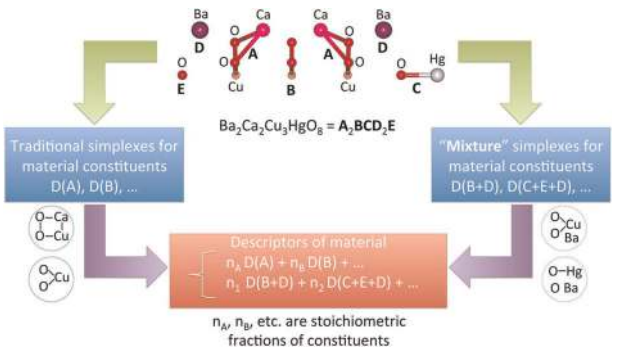

Fig. 4 Depiction of the generation of the simplex representation of molecular structure descriptors for materials. (Reprinted with permission from ref. [122]. Further permissions should be directed to the ACS.)

about the crystal structure (e.g., lattice constants) is added to the descriptor of the material, resulting in a feature vector of 2500 values in total. A characteristic of these graphs is their adjacency matrix, which consists of a square matrix of order n (number of atoms) filled with zeros except for the entries $a_{ij} = 1$ that occur when atom i and j are connected. Finally, for every property scheme q, the descriptors are calculated as

$$T = \sum_{i,j} |q_i - q_j| M_{ij}, \tag{24}$$

where the set of indices go over all pairs of atoms or over all pairs of bonded atoms, and M_{ij} are the elements of the product between the adjacency matrix of the graph and the reciprocal square distance matrix.

A different descriptor, named orbital-field matrix, was introduced by Pham et al.[127] Orbital-field matrices consist in the weighted product between one-hot vectors $\left(o_i^p\right)$, resembling those from the field of natural language processing. These vectors are filled with zeros with the exception of the elements that represent the electronic configuration of the valence of the atom. As an example, for the sodium atom with electronic configuration [Ne]$3s^1$, the one-hot vector is filled with zeros except for the first element, which is 1. The elements of the matrices are calculated from:

$$X_{ij}^p = \sum_{k=1}^{n_p} o_i^p \, o_j^k \, w_k(\theta_k^p, r_{pk}), \tag{25}$$

where the weight $w_k(\theta_k^p, r_{pk})$ represents the contribution of atom k to the coordination number of the center atom p and depends on the distance between the atoms and the solid angle θ_k^p determined by the face of the Voronoi polyhedron between the atoms. To represent crystal structures, the orbital-field matrices are averaged over the number of atoms N_p in the unit cell:

$$F_{ij} = \frac{1}{N_p} \sum_{p}^{N_p} X_{ij}^p. \tag{26}$$

Another way to construct features based on graphs is the crystal graph convolutional neural network (CGCNN) framework, proposed by Xie et al.[40] and shown schematically in Fig. 6. The atomic properties are represented by the nodes and encoded in the feature vectors v_i. Instead of using continuous values, each continuous property is divided into ten categories resulting in one-hot features. This is obviously not necessary for the discrete properties, which can be encoded as standard one-hot vectors without further transformations. The edges represent the bonding interactions and are constructed analogously to the property-labeled material fragments descriptor. Unlike most graphs, these crystal graphs allow for several edges between two nodes, due to periodicity. Therefore, the edges are encoded as one-hot feature vectors $u_{(i,j)_k}$, which translates into the kth bond between atom i and j. Crystal graphs do not form an optimal representation for predicting target properties by themselves; however, they can be improved by using convolution layers. After each convolution layer, the feature vectors gradually contain more information on the surrounding environment due to the concatenation between atom and bond feature vectors. The best convolution function of Xie et al. consisted of

$$\begin{aligned} \boldsymbol{v}_i^{(t+1)} = \boldsymbol{v}_i^{(t)} + \sum_{j,k} \sigma\left(\boldsymbol{z}_{(i,j)_k}^{(t)} \boldsymbol{W}_f^{(t)} + \boldsymbol{b}_f^{(t)}\right) \\ \odot g\left(\boldsymbol{z}_{(i,j)_k}^{(t)} \boldsymbol{W}_s^{(t)} + \boldsymbol{b}_s^{(t)}\right), \end{aligned} \tag{27}$$

where $\boldsymbol{W}_f^{(t)}$, $\boldsymbol{W}_s^{(t)}$, and $\boldsymbol{b}_i^{(t)}$ represent the convolution weight matrix, self-weight matrix, and the bias of the tth layer, respectively. In addition, e indicates element-wise multiplication, σ denotes the sigmoid function, and $z_{(i,j)_k}^{(t)}$ is the concatenation of

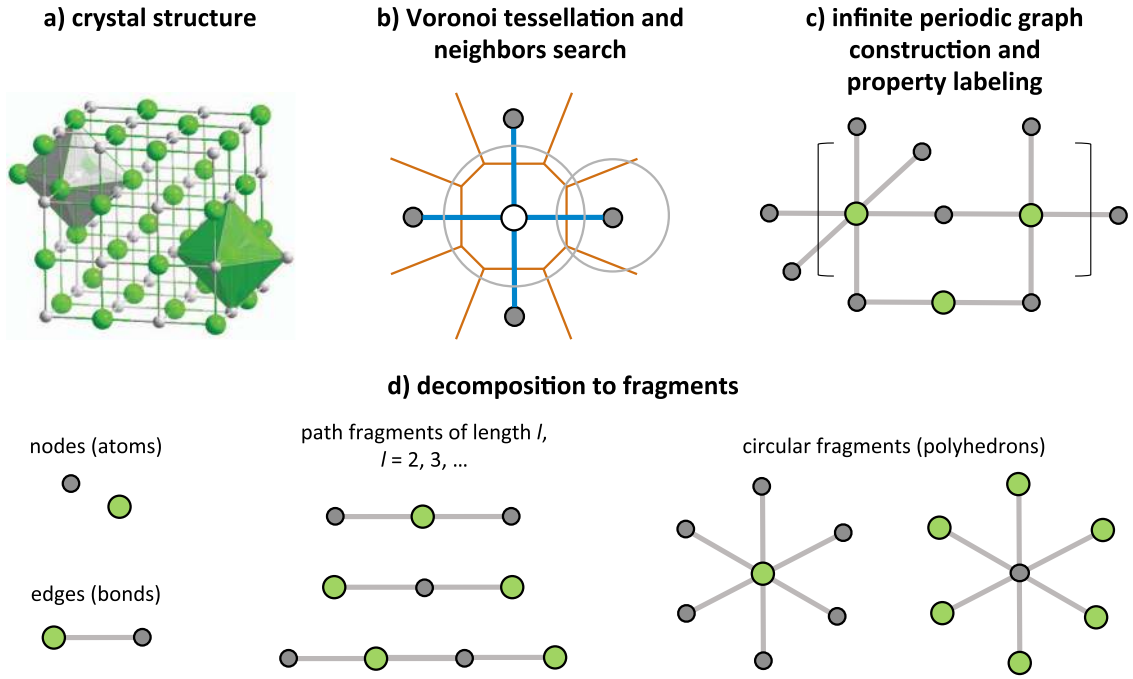

Fig. 5 Representation of the construction of property-labeled material fragments. The atomic neighbors of a crystal structure (**a**) are found via Voronoi tessellation (**b**). The full graph is constructed from the list of connections, labeled with a property (**c**) and decomposed into smaller subgraphs (**d**). (Reprinted with permission from ref. [41] licensed under the CC BY 4.0 [https://creativecommons.org/licenses/by/4.0/])

neighbor vectors:

$$z_{(ij)_k}^{(t)} = \boldsymbol{v}_i^{(t)} \oplus \boldsymbol{v}_j^{(t)} \oplus \boldsymbol{u}_{(ij)_k}, \tag{28}$$

Here $\oplus$ denotes concatenation of vectors.

After R convolutions, a pooling layer reduces the spatial dimensions of the convolution neural network. Using skip layer connections,[128] the pooling function operates not only on the last feature vector but also on all feature vectors (obtained after each convolution).

The idea of applying graph neural networks[129–131] to describe crystal structures stems from graph-based models for molecules, such as those proposed in refs. [131–140]. Moreover, all these models can be reorganized into a single common framework, known as message passing neural network[141] (MPNNs). The latter can be defined as a model operating on undirected graphs G, with edge features x_v and vertex features e_{vw}. In this context, the forward pass is divided into two phases: the message passing phase and the readout phase.

During the message passing phase, which lasts for T interaction steps, the hidden states h_v at each node in the graph are updated based on the messages m_v^{t+1}:

$$h_v^{t+1} = S_t\left(h_v^t, m_v^{t+1}\right), \tag{29}$$

where $S_t(\cdot)$ is the vertex update function. The messages are modified at each interaction by an update function $M_t(\cdot)$, which depends on all pairs of nodes (and their edges) in the neighborhood of v in the graph G:

$$m_v^{t+1} = \sum_{w \in N(v)} M_t\left(h_v^t, h_w^t, e_{vw}^t\right), \tag{30}$$

where $N(v)$ denotes the neighbors of v.

The readout phase occurs after T interaction steps. In this phase, a readout function $R(\cdot)$ computes a feature vector for the entire graph:

$$\hat{y} = R\left(\left\{h_v^T \in G\right\}\right). \tag{31}$$

Jørgensen et al.[142] extended MPNNs with an edge update network, which enforces the dependence of the information exchanged between atoms on the previous edge state and on the hidden states of the sending and receiving atom:

$$e_{vw}^{t+1} = E_t\left(h_v^{t+1}, h_w^{t+1}, e_{vw}^t\right), \tag{32}$$

where $E_t(.)$ is the edge update function. One example of MPNNs are causal generative neural networks. The message corresponds to $z_{(i,j)_k}^{(t)}$, defined in Eq. (27). Likewise, the hidden node update function corresponds to the convolution function of Eq. (27). In this case, we can clearly see that the hidden node update function depends on the message and on the hidden node state. The readout phase comes after R convolutions (or T iterations steps) and the readout function corresponds to the pooling layer function of the CGCNNs.

Up to now, we discussed very general features to describe both the crystal structure and chemical composition. However, should constrains be applied to the material space, the features necessary to study such systems can be vastly simplified. As mentioned above, elemental properties alone can be used as features, e.g., when a training set is restricted to only one kind of crystal structure and stoichiometry.[33,35,56,99,143] Consequently, the target property only depends on the chemical elements present in the composition. Another example can be found in ref. [144], where a polymer is represented by the number of building blocks (e.g., number of C_6H_4, CH_2, etc) or of pairs of blocks.

The crude estimations of properties can be an interesting supplement to standard features as discussed in ref. [77]. As its name implies, crude estimations of properties consist of the calculation of a target property (for example, the experimental band gap) utilizing crude estimators [for example, the DFT band gap calculated with the Perdew–Burke–Ernzerhof (PBE) approximation[145] to the exchange-correlation functional]. In principle, this approach can achieve successful results, as the machine learning algorithm no longer needs to predict a target property but rather an error or a difference between properties calculated with two well-defined methodologies.

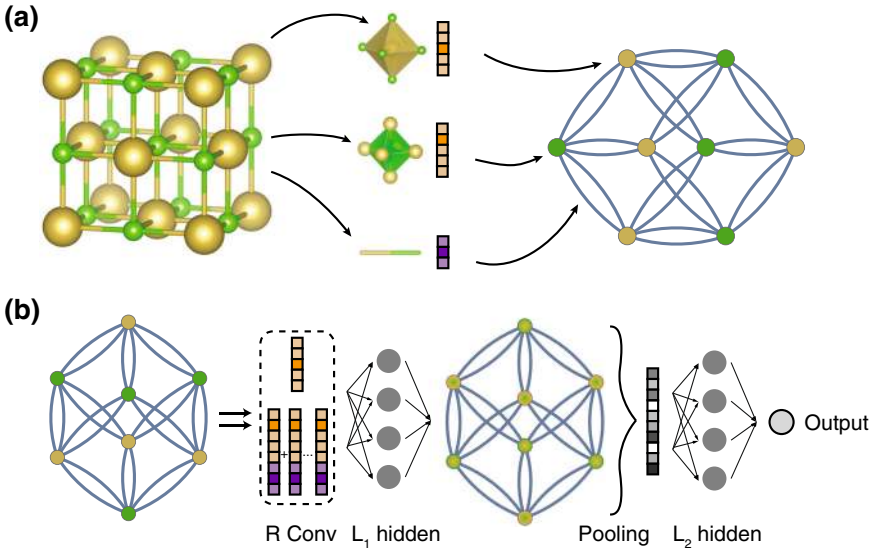

Fig. 6 Illustration of the crystal graph convolutional neural network. **a** Construction of the graph. **b** Structure of the convolutional neural network. (Reprinted with permission from ref. [40]. Copyright 2018 American Physical Society.)

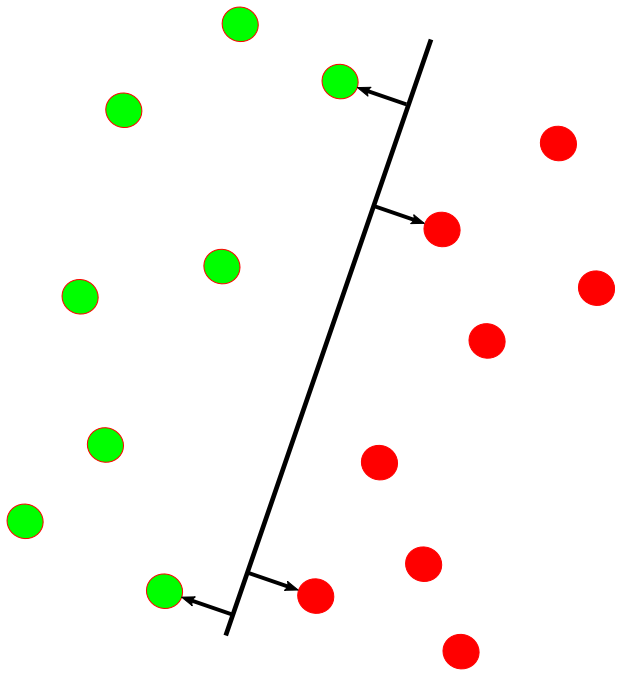

Fig. 7 Classification border of a support vector machine with the support vectors shown as arrows

Fischer et al.[146] took another route and used as features a vector that completely denotes the possible ground states of an alloy:

$$\mathbf{X} = (x_{c_1}, x_{c_2}, \ldots, x_{c_n}, x_{E_1}, x_{E_2}, \ldots, x_{E_c}), \tag{33}$$

where x_{c_i} denotes the all possible crystal structures present in the alloy at a given composition and x_{E_1} the elemental constituents of the system. In this way, the vector $\mathbf{X} = (\text{fcc, fcc, Au, Ag})$ would represent the gold–silver system. Furthermore, the probability density $p(\mathbf{X})$ denotes the probability that $\mathbf{X}$ is the set of ground states in a binary alloy. With these tools, one can find the most likely crystal structure for a given composition by sorting the probabilities and predict crystal structures by evaluating the conditional probability $p(\mathbf{X}|\mathbf{e})$, where $\mathbf{e}$ denotes unknown variables.

Having presented so many types of descriptors, the question that now remains concerns the selection of the best features for the problem at hand. In section "Basic principles of machine learning—Algorithms", we discuss some automatic feature selection algorithms, e.g., least absolute shrinkage and selection operator (LASSO), sure independence screening and sparsifying operator (SISSO), principal component analysis (PCA), or even decision trees. Yet these methods mainly work for linear models, and selecting a feature for, e.g., a neural network force field from the various features we described is not possible with any of these methods. A possible solution to this problem is to perform through benchmarks. Unfortunately, while there are many studies presenting their own distinct way to build features and applying them to some problem in materials science, fewer studies[96,100,147] actually present quantitative comparisons between descriptors. Moreover, some of the above features require a considerable amount of time and effort to be implemented efficiently and are not readily and easily available.

In view of the present situation, we believe that the materials science community would benefit greatly from a library containing efficient implementations of the above-mentioned descriptors and an assembly of benchmark datasets to compare the features in a standardized manner. Recent work by Himanen et al.[148] addresses part of the first problem by providing efficient implementations of common features. The library is, however, lacking the implementation of the derivatives. *SchNetPack* by Schütt et al.[149] also provides an environment for training deep neural network for energy surfaces and various material properties. Further useful tools and libraries can be found in refs. [150–152]

Algorithms

In this section, we briefly introduce and discuss the most prevalent algorithms used in materials science. We start with linear- and kernel-based regression and classification methods. We then introduce variable selection and extraction algorithms that are also largely based on linear methods. Concerning completely non-linear models, we discuss decision tree-based methods like random forests (RFs) and extremely randomized trees and neural networks. We start with simple fully connected feed-forward networks and convolutional networks and continue with more complex applications in the form of variational autoencoders (VAEs) and generative adversarial networks (GANs).

In ridge regression, a multi-dimensional least-squares linear-fit problem, including a L_2-regularization term, is solved:

$$\min_x |Ax - b|_2^2 + \lambda |x|_2^2. \tag{34}$$

The extra regularization term is included to favor specific solutions with smaller coefficients.

As complex regression problems can usually not be solved by a simple linear model, the so-called kernel trick is often applied to ridge regression.[153] Instead of using the original descriptor x, the data are first transformed into a higher-dimensional feature space $\phi(x)$. In this space, the kernel $k(x, y)$ is equal to the inner product $\langle \phi(x), \phi(y) \rangle$. In practice, only the kernel needs to be evaluated, avoiding an inefficient or even impossible explicit calculation of the features in the new space. Common kernels are, e.g.,[154], the radial basis function kernel

$$k_G(x, y) = e^{-\frac{|x-y|^2}{2\sigma^2}}, \tag{35}$$

or the polynomial kernel of degree d

$$k_P(x, y) = (x^T y + c)^d. \tag{36}$$

Solving the minimization problem given by Eq. (34) in the new feature space results in a non-linear regression in the original feature space. This is usually referred to as kernel ridge regression (KRR). KRR is generally simple to use, as for a successful application of KRR only very few hyperparameters have to be adjusted. Consequently, KRR is often used in materials science.

Support vector machines[155] (SVMs) search for the hyperplanes that divide a dataset into classes such that the margin around the hyperplane is maximized (see Fig. 7). The hyperplane is completely defined by the data points that lie the closest to the plane, i.e., the support vectors from which the algorithm derives its name.

Analogously to ridge regression, the kernel trick can be used to arrive at non-linear SVMs.[153] SVM regressors also create a linear model (non-linear in the kernel case) but use the so-called ε-insensitive loss function:

$$\text{Loss} = \begin{cases} 0 & \text{if } \varepsilon > |y - f(x)| \\ |y - f(x)| - \varepsilon & \text{otherwise} \end{cases} \tag{37}$$

where $f(x)$ is the linear model and ε a hyperparameter. In this way, errors smaller than the threshold defined by ε are neglected.

When comparing SVMs and KRR, no big performance differences are to be expected. Usually SVMs arrive at a sparser representation, which can be of advantage; however, their performance relies on a good setting of the hyperparameters. In most cases, SVMs will provide faster predictions and consume less memory, while KRR will take less time to fit for medium datasets. Nevertheless, owing to the generally low computational cost of both algorithms, these differences are seldom important for relatively small datasets. Unfortunately, neither method is feasible for large datasets as the size of the kernel matrix scales quadratically with the number of data points.

Gaussian process regression (GPR) relies on the assumption that the training data were generated by a Gaussian process and therefore consists of samples from a multivariate Gaussian distribution. The only other assumption that enter the regression are the forms of the covariance function $k(x, x')$ and the mean (which is often assumed to be zero). Based on the covariance matrix, whose elements represent the covariance between two features, the mean and the variance for every possible feature value can be predicted. The ability to estimate the variance is the main advantage of GPR, as the uncertainty of the prediction can be an essential ingredient of a materials design process (see section "Adaptive design process and active learning"). GPR also uses a kernel to define the covariance function. In contrast to KRR or SVMs where the hyperparameters of the kernel have be optimized with an external validation set, the hyperparameters in GPR can be optimized with gradient descent if the calculation of the covariance matrix and its inverse are computationally feasible. Although modern and fast implementations of Gaussian processes in materials science exist (e.g., COMBO[156]), their inherent scaling is quite limiting with respect to the data size and the descriptor dimension as a naive training requires an inversion of the covariance matrix of order $\mathcal{O}(N^3)$ and even the prediction scales with $\mathcal{O}(N^2)$ with respect to the size of the dataset.[157] Based on the principles of GPR, one can also produce a classifier. First, GPR is used to qualitatively evaluate the classification probability. Then a sigmoid function is applied to the latent function resulting in values in the interval [0, 1].

In the previous description of SVMs, KRR, and GPR, we assumed that a good feature choice is already known. However, as this choice can be quite challenging, methods for feature selection can be essential.

The LASSO[158,159] attempts to improve regression performance through the creation of sparse models through variable selection. It is mostly used in combination with least-squares linear regression, in which case it results in the following minimization problem[159]:

$$\min_{\beta,\beta_0} \sum_i (y_i - \beta_0 - \beta x_i)^2 \text{subject to} \sum_i |\beta_i| < t, \tag{38}$$

where y_i are the outcomes, x_i the features, and β the coefficients of the linear model that have to be determined. In contrast to ridge regression, where the L_2-norm of the regularization term is used, LASSO aims at translating most coefficients to zero. In order to actually find the model with the minimal number of non-zero components, one would have to use the so called L_0-norm of the coefficient vector, instead of the L_1-norm used in LASSO. (The L_0-norm of a vector is equal to its number of non-zero elements). However, this problem is non-convex and NP-hard and therefore infeasible from a computational perspective. Furthermore, it is proven[160] that the L_1-norm is a good approximation in many cases. The ability of LASSO to produce very sparse solutions makes it attractive for cases where a simple, maybe even simulatable model (see section "Discussion and conclusions—Interpretability"), is needed. The minimization problem from Eq. (38), under the constraint of the L_0-norm and the theory around it, is also known as compressed sensing.[161]

Ghiringhelli et al. described an extended methodology for feature selection in materials science based on LASSO and compressed sensing.[162] Starting with a number of primary features, the number of descriptors is exponentially increased by applying various algebraic/functional operators (such as the absolute value of differences, exponentiation, etc.) and constructing different combinations of the primary features. Necessarily, physical notions like the units of the primary features constrain the number of combinations. LASSO is then used to reduce the number of features to a point where a brute force combination approach to find the lowest error is possible. This approach is chosen in order to circumvent the problems pure LASSO faces when treating strongly correlated variables and to allow for non-linear models.

As LASSO is unfortunately still computationally infeasible for very high-dimensional feature spaces (>10^9), Ouyang et al. developed the SISSO[163] that combines sure independence screening,[164] other sparsifying operators, and the feature space generation from ref. [162]. Sure independence screening selects a subspace of features based on their correlation with the target variable and allows for extremely high-dimensional starting spaces. The selected subspace is than further reduced by applying the sparsifying operator (e.g., LASSO). Predicting the relative stability of octet binary materials as either rock-salt or zincblende was used as a benchmark. In this case, SISSO compared favorably with LASSO, orthogonal matching pursuit,[165,166] genetic programming,[167] and the previous algorithm from ref. [162] Bootstrapped-

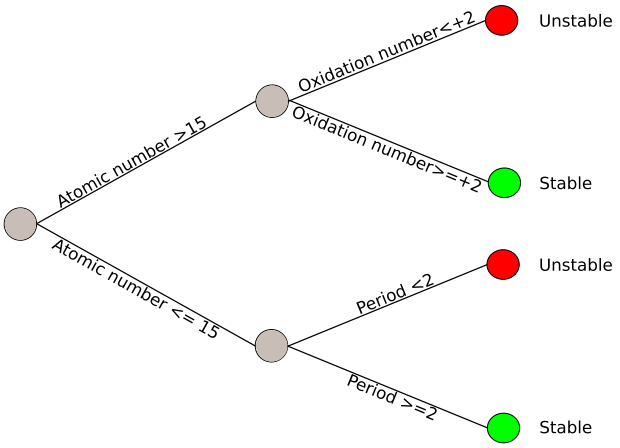

Fig. 8 Schema of a classification tree deciding whether a material is stable

projected gradient descent[168] is another variable selection method developed for materials science. The first step of bootstrapped-projected gradient descent consists in clustering the features in order to combat the problems other algorithms like LASSO face when encountering strongly correlated features. The features in every cluster are combined in a representative feature for every cluster. In the following, the sparse linear fit problem is approximated with projected gradient descent[169] for different levels of sparsity. This process is also repeated for various bootstrap samples in order to further reduce the noise. Finally, the intersection of the selected feature sets across the bootstrap samples is chosen as the final solution.

PCA[170,171] extracts the orthogonal directions with the greatest variance from a dataset, which can be used for feature selection and extraction. This is achieved by diagonalizing the covariance matrix. Sorting the eigenvectors by their eigenvalues (i.e., by their variance) results in the first principal component, second principal component, and so on. The broad idea behind this scheme is that, in contrast to the original features, the principal components will be uncorrelated. Furthermore, one expects that a small number of principal components will explain most of the variance and therefore provide an accurate representation of the dataset. Naturally, the direct application of PCA should be considered feature extraction, instead of feature selection, as new descriptors in the form of the principal components are constructed. On the other hand, feature selection based on PCA can follow various strategies. For example, one can select the variables with the highest projection coefficient from, respectively, the first n principal components when selecting n features. A more in-depth discussion of such strategies can be found in ref. [171].

The previous algorithms can be considered as linear models or linear models in a kernel space. An important family of non-linear machine learning algorithms is composed by decision trees. In general terms, decision trees are graphs in tree form,[172] where each node represents a logic condition aiming at dividing the input data into classes (see Fig. 8) or at assigning a value in the case of regressors. The optimal splitting conditions are determined by some metric, e.g., by minimizing the entropy after the split or by maximizing an information gain.[173]

In order to avoid the tendency of simple decision trees to overfit, ensembles such as RFs[174] or extremely randomized trees[175] are used in practice. Instead of training a single decision tree, multiple decision trees with a slightly randomized training process are built independently from each other. This randomization can include, for example, using only a random subset of the whole training set to construct the tree, using a random subset of the features, or a random splitting point when considering an optimal split. The final regression or classification result is usually obtained as an average over the ensemble. In this way, additional noise is introduced into the fitting process and overfitting is avoided.

In general, decision tree ensemble methods are fast and simple to train as they are less reliant on good hyperparameter settings than most other methods. Furthermore, they are also feasible for large datasets. A further advantage is their ability to evaluate the relevance of features through a variable importance measure, allowing a selection of the most relevant features and some basic understanding of the model. Broadly speaking, these are based on the difference in performance of the decision tree ensemble by including and excluding the feature. This can be measured, e.g., through the impurity reduction of splits using the specific feature.[176]

Extremely randomized trees are usually superior to RFs in higher variance cases as the randomization decreases the variance of the total model[175] and demonstrate at least equal performances in other cases. This proved true for several applications in materials science where both methods were compared.[99,177,178] However, as RFs are more widely known, they are still prevalent in materials science.

Boosting methods[179] generally combine a number of weak predictors to create a strong model. In contrast to, e.g., RFs where multiple strong learners are trained independently and combined through simple averaging to reduce the variance of the ensemble model, the weak learners in boosting are not trained independently and are combined to decrease the bias in comparison to a single weak learner. Commonly used methods, especially in combination with decision tree methods, are gradient boosting[180,181] and adaptive boosting.[182,183] In materials science, they were applied to the prediction of bulk moduli[184,185] and the prediction of distances to the convex hull, respectively.[99,186]

Ranging from feed-forward neural networks over self-organizing maps[187] up to Boltzmann machines[188] and recurrent neural networks,[189] there is a wide variety of neural network structures. However, until now only feed-forward networks have found applications in materials science (even if some Boltzmann machines are used in other areas of theoretical physics[190]). As such, in the following we will leave out "feed-forward" when referring to feed-forward neural networks. In brief, a neural network starts with an input layer, continues with a certain number of hidden layers, and ends with an output layer. The neurons of the nth layer, denoted as the vector x^n, are connected to the previous layer through the activation function $\phi(x)$ and the weight matrix A_{ij}^{n-1}:

$$x_i^n = \phi\left(\sum_j x_j^{n-1} A_{ij}^{n-1}\right). \tag{39}$$

The weight matrices are the parameters that have to be fitted during the learning process. Usually, they are trained with gradient descent style methods with respect to some loss function (usually L_2 loss with L_1 regularization), through a method known as back-propagation.

Inspired by biological neurons, sigmoidal functions were classically used as activation functions. However, as the gradient of the weight-matrix elements is calculated with the chain rule, deeper neural networks with sigmoidal activation functions quickly lead to a vanishing gradient,[191] hampering the training process. Modern activation functions such as rectified linear units[192,193]

$$\phi(x) = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases} \tag{40}$$

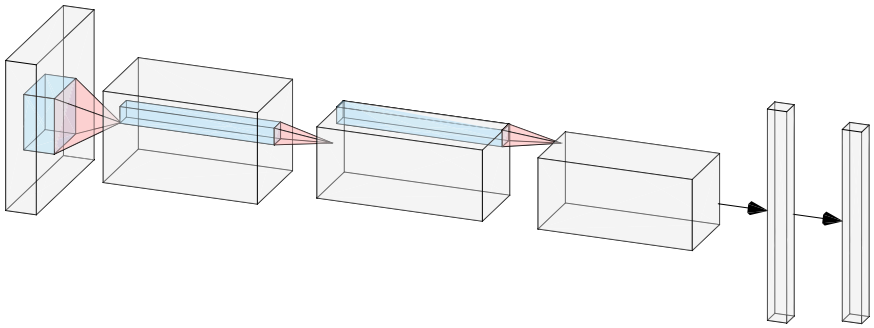

Fig. 9 Topology of a convolutional neural network starting with convolutional layers with multiple filters followed by pooling and two fully connected layers

or exponential linear units[194]

$$\phi(x) = \begin{cases} x & \text{if } x > 0 \\ a(e^x - 1) & \text{otherwise} \end{cases} \quad (41)$$

alleviate this problem and allow for the development of deeper neural networks.

The real success story of neural networks only started once convolutional neural networks were introduced to image recognition.[55,195] Instead of solely relying on fully connected layers, two additional layer variants known as convolutional and pooling layers were introduced (see Fig. 9).

Convolutional layers consist of a set of trainable filters, which usually have a receptive field that considers a small segment of the total input. The filters are applied as discrete convolutions across the whole input, allowing the extraction of local features, where each filter will learn to activate when recognizing a different feature. Multiple filters in one layer add an additional dimension to the data. As the same weights/filters are used across the whole input data, the number of hidden neurons is drastically reduced in comparison to fully connected layers, thus allowing for far deeper networks. Pooling layers further reduce the dimensionality of the representation by combining subregions into a single output. Most common is the max pooling layer that selects the maximum from each region. Furthermore, pooling also allows the network to ignore small translations or distortions. The concept of convolutional networks can also be extended to graph representations in material science,[139] in what can be considered MPNNs[141]. In general, neural networks with five or more layers are considered deep neural networks,[55] although no precise definition of this term in relation to the network topology exists. The advantage of deep neural networks is not only their ability to learn representations with different abstraction levels but also to reuse them.[97] Ideally, the invariance and differentiation ability of the representation should increase with increasing depth of the model.

Obviously, this saves resources that would otherwise be spent on feature engineering. However, some of these resources have now to be allocated to the development of the topology of the neural network. If we consider hard-coded layers (like pooling layers), one can once again understand them as feature extraction through human intervention. While some methods for the automatic development of neural network structures exist (e.g., the neuroevolution of augmenting topologies[196]), in practice the topologies of neural networks are still developed through trial and error. The extreme speedup in training time through graphics processing unit implementations and new methods that improve the training of deep neural networks, like dropout[197] and batch normalization,[198] also played a big role in the success story of neural networks. As these methods are included in open source libraries, like tensorflow[199] or pytorch,[200] they can easily be applied in materials science.

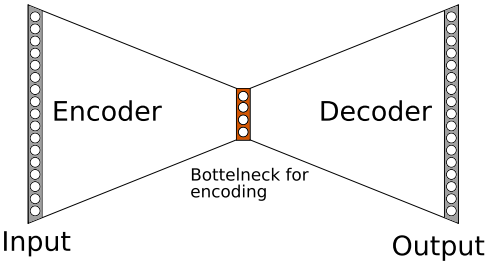

Fig. 10 Structure of an autoencoder

Neural networks can also be used in a purely generative manner, for example, in the form of autoencoders[201,202] or GANs.[203] Generative models learn to reproduce realistic samples from the distribution of data they are trained on. Naturally, one of the end goals of machine learning in materials science is the development of generative models, which can take into account material properties and therefore encompass most of the material design process.

Autoencoders are built with the purpose of learning a new efficient representation of the input data without supervision. Typically, the autoencoder is divided into two parts (see Fig. 10). The first half of the neural network is the encoder, which ends with a layer that is typically far smaller than the input layer in order to force the autoencoder to reduce the dimensionality of the data. The second half of the network, the decoder, attempts to regain the original input from the new encoded representation.

VAEs are based on a specific training algorithm, namely, stochastic gradient variational Bayes,[204] that assumes that the VAE learns an approximation of the distribution of the input. Naturally, VAEs can also be used as generative models by generating data in the form of the output of the encoder and subsequently decoding it.

GANs consist of two competing neural networks that are trained together (see Fig. 11): a generative model that attempts to produce samples from a distribution and a discriminative model that predicts the probability that an input belongs to the original distribution or was produced by the generative model. GANs have found great success in image processing[205,206] and have recently been introduced to other fields, such as astronomy,[207] particle physics,[208] genetics,[209] and also very recently to materials science.[29,210,211]

More information about these algorithms can be found in the references provided or in refs. [1,212–216]. We would like to note that the choice of the machine learning algorithm is completely problem dependent. It can be useful to establish a baseline for the quality of the model by using simple approaches (such as extremely randomized trees) before spending time optimizing hyperparameters of more complex models.

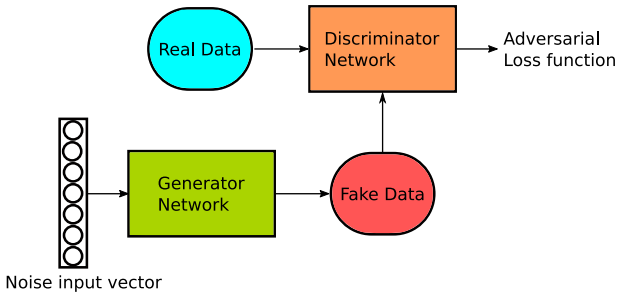

Fig. 11 Structure of a generative adversarial network

MATERIAL DISCOVERY

Nearly 30 years ago, the at the time editor of Nature, John Maddox wrote "One of the continuing scandals of physical science is that it remains in general impossible to predict the structure of even the simplest crystalline solids from a knowledge of their chemical composition."[217] While this is far from true nowadays, predicting the crystal structure based solely on the composition remains one of the most important (if not even the key) challenge in materials science, as any rational materials design has to be grounded in the knowledge of the crystal structure.

Unfortunately, the first-principle prediction of crystal or molecular structures is exceptionally difficult, because the combinatorial space is composed of all possible arrangements of the atoms in three-dimensional space and with an extremely complicated energy surface.[18] In recent years, advanced structure selection and generation algorithms such as random sampling,[218–221] simulated annealing,[222–224] metadynamics,[225] minima hopping,[226] and evolutionary algorithms,[19,227–233] as well as the progress in energy evaluation methods, expanded the scope of application of "classical" crystal structure prediction methods to a wider range of molecules and solid forms.[234] Nevertheless, these methods are still highly computationally expensive, as they require a substantial amount of energy and force evaluations. However, the search for new or better high-performance materials is not possible without searching through an enormous composition and structure space. As there are tremendous amounts of data involved, machine learning algorithms are some of the most promising candidates to take on this challenge.

Machine learning methods can tackle this problem from different directions. A first approach is to speed up the energy evaluation by replacing a first-principle method with machine learning models that are orders of magnitude faster (see section "Machine learning force fields"). However, the most prominent approach in inorganic solid-state physics is the so-called component prediction.[61] Instead of scanning the structure space for one composition, one chooses a prototype structure and scans the composition space for the stable materials. In this context, thermodynamic stability is the essential concept. By this we mean compounds that do not decompose (even in infinite time) into different phases or compounds. Clearly, metastable compounds like diamond are also synthesizable and advances in chemistry have made them more accessible.[235,236] Nevertheless, thermodynamically stable compounds are in general easier to produce and work with. The usual criterion for thermodynamic stability is based on the energetic distance to the convex hull, but in some cases the machine learning model will directly calculate the probability of a compound existing in a specific phase.

Component prediction

Clearly the formation energy of a new compound is not sufficient to predict its stability. Ideally, one would always want to use the distance to the convex hull of thermodynamic stability. In contrast to the formation energy, the distance to the convex hull considers the difference in free energy of all possible decomposition channels. De facto, this is not the case because our knowledge of the convex hull is of course incomplete. Fortunately, as our knowledge of the convex hull continuously improves with the discovery of new stable materials, this problem becomes less important over time. Lastly, most first-principle energy calculations are done at zero temperature and zero pressure, neglecting kinetic effects on the stability.

Faber et al.[35] applied KRR to calculate formation energies of two million elpasolites (with stoichiometry ABC_2D_6) crystals consisting of main group elements up to bismuth. Errors of around 0.1 eV/atom were reported for a training set of 10^4 compositions. Using energies and data from the materials project,[78] phase diagrams were constructed and 90 new stoichiometries were predicted to lie on the convex hull.

Schmidt et al.[99] first constructed a dataset of DFT calculations for approximately 250,000 cubic perovskites (with stoichiometry ABC_3) using all elements up to bismuth and neglecting rare gases and lanthanides. After testing different machine learning methods, extremely randomized trees[175] in combination with adaptive boosting[183] proved the most successful with an mean average error of 0.12 eV/atom. Curiously, the error in the prediction depends strongly on the chemical composition (see Fig. 12). Furthermore, an active learning approach based on pure exploitation was suggested (see section "Adaptive design process and active learning").

In ref. [186], the composition space for two ternary prototypes with stoichiometry AB_2C_2 (tI10-$CeAl_2Ga_2$ and the tP10-$FeMo_2B_2$ prototype structures) were explored for stable compounds using the approach developed in ref. [99]. In total, 1893 new compounds were found on the convex hull while saving around 75% of computation time and reporting false negative rates of only 0% for the tP10 and 9% for the tI10 compound.

Ward et al.[34] used standard RFs to predict formation energies based on features derived from Voronoi tessellations and atomic properties. Starting with a training set of around 30,000 materials, the descriptors showed better performance than Coulomb matrices[108] and partial RDFs[109] (see section "Basic principles of machine learning—Features" for the different descriptors). Surprisingly, the structural information from the Voronoi tessellation did not improve the results for the training set of 30,000 materials. This is based on the fact that very few materials with the same composition, but different structure, are present in the dataset. Changing the training set to an impressive 400,000 materials from the open quantum materials database[80] proved this point, as the error for the composition-only model was then 37% higher than for the model including the structural information.

A recent study by Kim et al.[237] used the same method for the discovery of quaternary Heusler compounds and identified 53 new stable structures. The model was trained for different datasets (complete open quantum materials database,[80] only the quaternary Heusler compounds, etc.). For the prediction of Heusler compounds, it was found that the accuracy of the model also benefited from the inclusion of other prototypes in the training set. It has to be noted that studies with such large datasets are not feasible with kernel-based methods (e.g. KRR, SVMs) due to their unfavorable computational scaling.

Li et al.[33] applied different regression and classification methods to a dataset of approximately 2150 $A_{1-x}A'_xB_{1-y}B'_yO_3$ perovskites, materials that can be used as cathodes in high-temperature solid oxide fuel cell.[238] Elemental properties were used as features for all methods. Extremely randomized trees proved to be the best classifiers (accuracy 0.93, F_1-score 0.88) while KRR and extremely randomized trees had the best performance for regression, with mean average errors of <17 meV/atom. The errors in this work are difficult to compare to others as the elemental composition space was very limited.

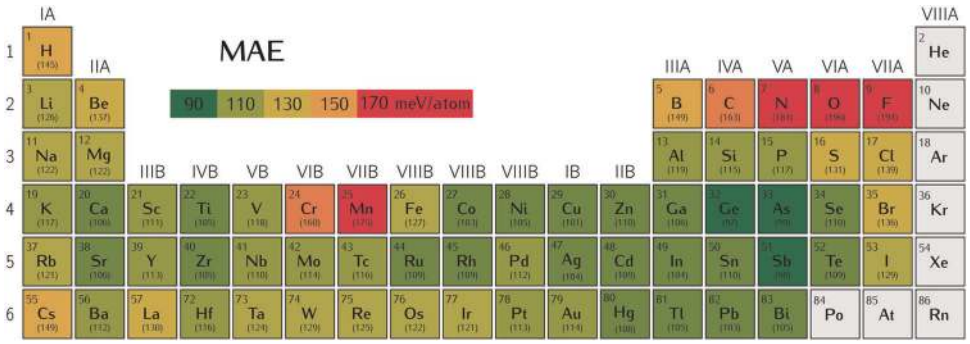

Fig. 12 Mean average error (in meV/atom) for adaptive boosting used with extremely random trees averaged over all perovskites containing the element. The numbers in parentheses are the actual averaged error for each element. (Reprinted with permission from ref. [99]. Copyright 2017 American Chemical Society.)

Another work treating the problem of oxide–perovskite stability is ref. [56]. Using neural networks based only on the elemental electronegativity and ionic radii, Ye et al. achieved a mean average error of 30 meV/atom for the prediction of the formation energy of unmixed perovskites. Unfortunately, their dataset contained only 240 compounds for training, cross-validation, and testing. Ye et al.[56] also achieved comparable errors for mixed perovskites, i.e. perovskites with two different elements on either the A- or B-site. Mean average errors of 9 and 26 meV/atom were then obtained, respectively, for unmixed and mixed garnets with the composition $C_3A_2D_3O_{12}$. By reducing the mixing to the C-site and including additional structural descriptors, Ye et al. were able to once again decrease the latter error to merely 12 meV/atom. If one compares this study to, e.g., refs. [1,35], the errors seem extremely small. This is easily explained once we notice that ref. [56] only considers a total compound space of around 600 compounds in comparison to around 250,000 compounds in ref. [1]. In other words, the complexity of the problem differs by more than two orders of magnitude.

The CGCNNs (see section "Basic principles of machine learning—Features") developed by Xie et al.,[40] the *MatErials Graph Networks*[132] by Chen et al., and the MPNNs by Jørgensen et al.[142] also allow for the prediction of formation energies and therefore can be used to speed up component prediction.

Up to this point, all component prediction methods presented here relied on first-principle calculations for training data. Owing to the prohibitive computational cost of finite temperature calculations, nearly all of this data correspond to zero temperature and pressure and therefore neglects kinetic effects on the stability. Furthermore, metastable compounds, such as diamond, which are stable for all practical purposes and essential for applications, risk to be overlooked. The following methods bypass this problem through the use of experimental training data.

The first structure prediction model that relies on experimental information can be traced back to the 1920s. One example that was still relevant until the past decade is the tolerance factor of Goldschmidt,[239] a criterion for the stability of perovskites. Only recently, modern methods like SISSO,[163] gradient tree boosting,[180] and RFs[174] improved upon these old models and allowed a rise in precision from 74% to >90%[143,240,241] for the stability prediction of perovskites. Balachandran et al.[241] also predicted whether the material would exist as a cubic or non-cubic perovskite, reaching a 94% average cross-validation error. The advantage of stability prediction based on experimental data is a higher precision and reliability, as the theoretical distance to the convex hull is a good but far from perfect indicator for stability. Taking the example of perovskites, one has to increase the distance to the convex hull up to 150 meV/atom just to find even 95% of the perovskites present in the inorganic crystal structure database[79] (see Fig. 13).

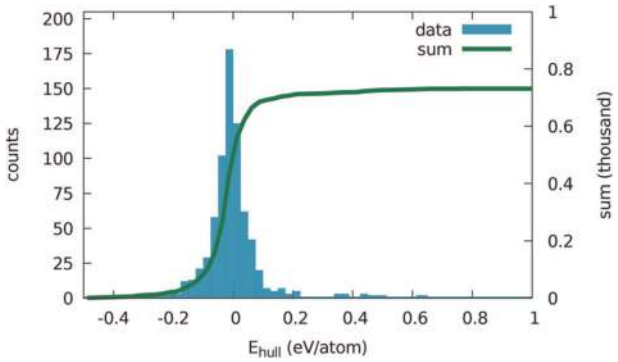

Fig. 13 Histogram of the distance to the convex hull for perovskites included in the inorganic crystal structure database.[79] Calculations were performed within density functional theory with the Perdew–Burke–Ernzerhof approximation. The bin size is 25 meV/atom. (Reprinted with permission from ref. [99]. Copyright 2017 American Chemical Society.)

Another system with a relatively high number of experimentally known structures are the AB_2C Heusler compounds. Oliynyk et al.[242] used RFs and experimental data for all compounds with AB_2C stoichiometry from Pearson's crystal data[243] and the alloy phase diagram database[88] to build a model to predict the probability to form a full-Heusler compound with a certain composition. Using basic elemental properties as features, Olynyk et al. were able to successfully predict and experimentally confirm the stability of several novel full-Heusler phases.

Legrain et al. extended the principle of this work to half-Heusler *ABC* compounds. While comparing the results of three ab initio high-throughput studies[37,244,245] to the machine learning model, they found that the predictions of the high-throughput studies were neither consistent with each other nor with the machine learning model. The inconsistency between the first-principle studies is due to different publication dates that led to different knowledge about the convex hulls and to slightly differing methodologies. The machine learning model performs well with 9% false negatives and 1% false positives (in this case, positive means stable as half-Heusler structure). In addition, the machine learning model was able to correctly label several structures for which the ab initio prediction failed. This demonstrates the possible advantages of experimental training data, when it is available.

Zheng et al.[36] applied convolutional neural networks and transfer learning[246] to the prediction of stable full-Heusler compounds AB_2C. Transfer learning considers training a model for one problem and then using parts of the model, or the knowledge gained during the first training process, for a second

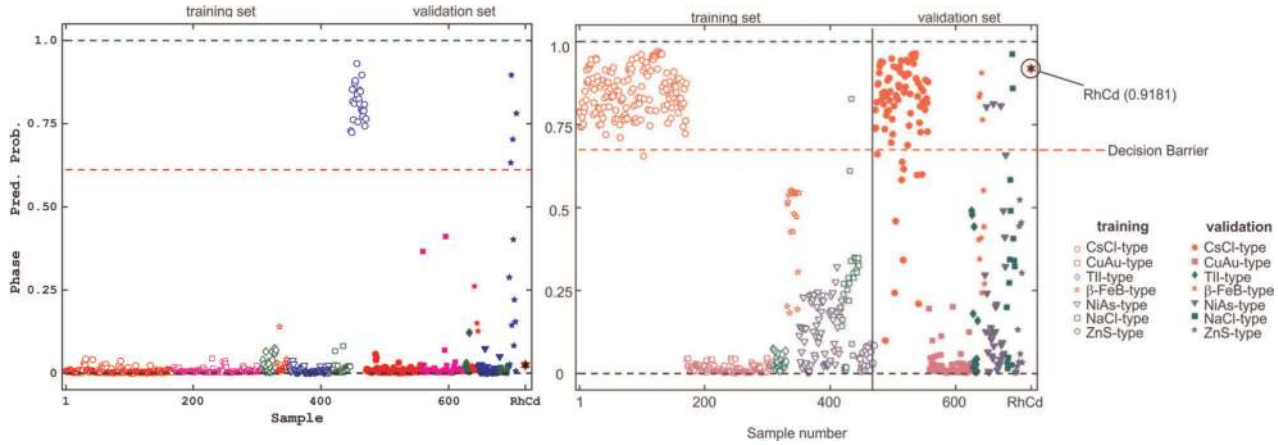

Fig. 14 Predicted probability for ZnS-type (left) and CsCl-type structures (right) for the support vector machine model with 31 features. (Reprinted with permission from ref. [32]. Further permissions should be directed to the ACS.)

training thereby reducing the data required. An image of the periodic table representation was used in order to take advantage of the great success of convolutional neural networks for image recognition. The network was first trained to predict the formation energy of around 65,000 full-Heusler compounds from the open quantum materials database,[80] resulting in a mean absolute error of 7 meV/atom (for a training set of 60,000 data points) and 14 meV/atom (for a training set of 5000 compositions). The weights of the neural network were then used as starting point for the training of a second convolutional neural network that classified the compositions as stable or unstable according to training data from the inorganic crystal structure database.[79] Unfortunately, no data concerning the accuracy of the second network was published.

Hautier et al.[247] combined the use of experimental and theoretical data by building a probabilistic model for the prediction of novel compositions and their most likely crystal structures. These predictions were then validated with ab initio computations. The machine learning part had the task to provide the probability density of different structures coexisting in a system based on the theory developed in ref. [146]. Using this approach, Hautier et al. searched through 2211 ABO compositions where no ternary oxide existed in the inorganic crystal structure database[79] and where the probability of forming a compound was larger than a threshold. This resulted in 1261 compositions and 5546 crystal structures, whose energy was calculated using DFT. To assess the stability, the energies of all decomposition channels that existed in the database were also calculated, resulting in 355 new compounds on the convex hull.

It is clear that component prediction via machine learning can greatly reduce the cost of high-throughput studies through a preselection of materials by at least a factor of ten.[1] Naturally, the limitations of stability prediction according to the distance to the convex hull have to be taken into consideration when working on the basis of DFT data. While studies based on experimental data can have some advantage in accuracy, this advantage is limited to crystal structures that are already thoroughly studied, e.g., perovskites, and consequently a high number of experimentally stable structures is already known. For a majority of crystal structures, the number of known experimentally stable systems is extremely small and consequently ab initio data-based studies will definitely prevail over experimental data-based studies. Once again, a major problem is the lack of any benchmark datasets, preventing a quantitative comparison between most approaches. This is even true for work on the same structural prototype. Considering, for example, perovskites, we notice that three groups predicted distances to the convex hull.[33,56,99] However, as the underlying composition spaces and datasets are completely different it is hardly possible to compare them.

Structure prediction

In contrast to the previous section, where the desired output of the models was a value quantifying the probability of compositions to condense in one specific structure, models in this chapter are concerned with differentiating multiple crystal structures. Usually this is a far more complex problem, as the theoretical complexity of the structural space dwarfs the complexity of the composition space. Nevertheless, it is possible to tackle this problem with machine learning methods.

Early attempts, which predate machine learning, include, e.g., Pettifor structural maps that use elementary properties to separate different binary or ternary structures from each other in a 2D plot, allowing the prediction of new stable structures.[248–251] In some sense, Pettifor maps are already closely related to recent work, such as ref. [163], where a structural map for binary structures based on chemical properties was developed with SISSO. Some of the first applications of modern machine learning crystal structure prediction can be found in ref. [146]. There, Fischer et al. developed an approach based on the cumulant expansion method, described in ref. [252], to predict the probability of an elemental composition forming a specific binary crystal structure. Their method estimates the correlation of the stability of two structures with respect to their composition. The model was trained with data from ref. [90] and evaluated with leave-one-out cross-validation. It was able to predict the correct structure in 90% of the cases during the first five guesses, in comparison to 62% when picking the structures according to their frequency in the dataset. It has to be mentioned here once again that leave-one-out cross validation is not a good method to evaluate the extrapolation ability of such models.[74,75]

Olynyk et al.[32] applied cluster resolution feature selection[253] to the classification of binary crystal structures. These features were then used as input for partial least-squares discriminant analysis (PLS-DA) and SVMs. In order to reduce the complexity of the problem, only the seven most common binary prototype structures were considered. A dataset of 706 compounds was divided into three sets, 235 for feature selection, 235 for optimization of the PLS-DA and SVMs, and 236 for validation. The SVMs performed better with an average false positive rate of 5.8%, a false negative rate of 7.3%, and an accuracy of 93.2%, compared to the PLS-DA with 3.5%, 34.0%, and an accuracy of 77.1%. It has to be noted that these values differ significantly depending on the crystal system that one tries to predict (see Fig. 14). This approach was adapted by Olynyk et al.[254] to equiatomic ternary compounds. SVMs were used on a dataset of ~1500

	$Ca(Ca_{0.5}Nd_{0.5})_2NbO_6$	$Ca_2Nb_2O_7$	$CaTiO_3$	$CeAl_2Ga_2$	Cu	CuZrSiAs	FeAs	$GdFeO_3$	K_2NiF_4	$LaAlO_3$	$MgAl_2O_4$	$MgCu_2$	NaCl	$NaFeO_2$	TiNiSi	Other	Recall
$Ca(Ca_{0.5}Nd_{0.5})_2NbO_6$	94	0	0	0	0	0	0	0	0	0	0	0	0	0	0	43	0.686
$Ca_2Nb_2O_7$	0	95	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0.655
$CaTiO_3$	0	0	133	0	0	0	0	12	0	5	0	0	0	0	1	105	0.522
$CeAl_2Ga_2$	0	0	0	161	0	0	0	0	0	0	0	0	0	0	0	28	0.847
Cu	0	0	0	0	56	0	0	0	0	0	0	0	0	0	0	70	0.444
CuZrSiAs	0	0	0	0	0	93	0	0	0	0	0	0	0	0	0	21	0.816
FeAs	0	0	0	0	0	0	88	0	0	0	0	0	0	0	0	15	0.854
$GdFeO_3$	0	0	9	0	0	0	0	454	0	19	0	0	1	0	0	120	0.753
K_2NiF_4	0	0	0	0	0	0	0	3	81	2	0	0	0	0	0	56	0.570
$LaAlO_3$	0	0	2	0	0	0	0	33	1	92	0	0	0	0	0	27	0.594
$MgAl_2O_4$	0	0	0	0	0	0	0	0	0	0	315	0	0	0	0	69	0.820
$MgCu_2$	0	0	0	0	0	0	0	0	0	0	0	58	0	0	0	53	0.523
NaCl	0	0	0	0	0	0	0	1	0	0	1	0	140	1	0	81	0.625
$NaFeO_2$	0	0	0	0	0	0	0	0	0	0	0	0	0	105	0	34	0.755
TiNiSi	0	0	0	1	0	0	3	0	0	0	0	0	0	0	45	65	0.395
Other	0	5	29	5	37	2	18	59	21	16	31	14	21	12	8	20984	0.986
Total	105	100	173	167	93	95	109	562	103	134	103	72	162	118	54	21821	0.950
Precision	0.895	0.950	0.769	0.964	0.602	0.979	0.807	0.808	0.786	0.687	0.908	0.806	0.864	0.890	0.833	0.962	

Fig. 15 Confusion matrix for cutoff size of 100 (a perfect confusion matrix is diagonal). (Reprinted with permission from ref. [30]. Copyright 2018 American Chemical Society.)

ternary compounds from Pearson's crystal data. Reducing the number of features via cluster resolution, from an initial 1000 features to 110, resulted in a sensitivity of 97.3%, accuracy of 96.9%, and specificity of 93.9%.

As crystal structure prediction is only the first step in the rational design process, combining stability determination with property design is necessary. Balachandran et al.[31] studied a set of 60,000 potential $x\mathrm{BiMe'}_y\mathrm{Me''}_{1-y}\mathrm{O}_3$–$(1-x)\mathrm{PbTiO}_3$ perovskites with several machine learning methods. First, SVMs were used to classify them into perovskites and non-perovskites, followed by a prediction of the Curie temperature of those classified as perovskites. Once a candidate was experimentally synthesized, it was added to the training set and the process was repeated. Of the ten synthesized compounds, six perovskites were found, whose highest Curie temperature was reported to be 898 K.

Graser et al.[30] applied RFs for crystal structure classification of 24,215 compounds from Pearson's crystal data[243] database. Naturally, a lot of prototypes only have very few representatives (<10) in the database. In order to circumvent this problem, prototypes with fewer instances than a certain cutoff number were put into a group denoted as "other." As the "other" class comprised between 92.51% and 64.1% of the dataset, depending on the choice of cutoff number, this greatly reduced the complexity of the dataset. Graser et al. then researched the change in predictive ability of the model with respect to the cutoff number. As expected, the recall improved with increasing cutoff number. The confusion table (see Fig. 15) demonstrates that, through the use of large datasets, even a simple method can achieve impressive results for an extremely challenging task like crystal structure prediction.

Park et al.[255] tackled the problem of crystal structure prediction from a slightly different perspective. All the others methods discussed up to now used the chemical composition, or data derived from the chemical composition, and structure as descriptors. In contrast, Park et al. used powder X-ray diffraction patterns to determine the crystal system, extinction group, and space group of inorganic compounds. Because the three-dimensional electron density is contracted into a 1D diffraction pattern, the symmetry of the crystal is often not fully determined from the diffraction pattern alone, especially for low-symmetry structures. While software for indexing and determining the space group exists, it requires substantial expertise and human input to obtain the correct results. Previous machine learning attempts in this field mostly considered the task of feature engineering (e.g., PCA[256–259] or manual featurization[260,261]) or considered smaller datasets and shallower neural networks. In contrast, in ref. [255] deep convolutional neural networks were developed, using X-ray patterns as input and giving as output the space group, extinction group, or crystal system. For the training data, structural data from the inorganic crystal structure database[79] was used to calculate randomly perturbed spectra, which simulated real spectra. During testing on a dataset that amounted to around 20% of the training set, the network reached accuracies of 81.14, 83.83, and 94.99% for space group, extinction group, and crystal system classifications, respectively. Furthermore, the model was able to correctly identify the structural system of two novel compounds[262,263] whose prototype structure did not appear in the database (and therefore neither in the training set). Albeit the model was not performing better than human experts using a software like TREOR,[264] it has the potential to be a useful tool to non-experts and in order to speed up the identification process of X-ray diffraction spectra in general. The success of the model is not surprising, as the use of convolutional neural networks in image classification[265–267] is well established in computer science.

A similar approach for crystal structure classification was followed in ref. [268]. Zilleti et al. used convolutional neural networks to classify crystal structures by a simulated 2D diffraction fingerprint. The approach was limited to the classification of crystal structures, because the 2D diffraction pattern is not unambiguous for all space groups, and consequently the neural network is not able to distinguish between rhombohedral and hexagonal structures, for example. They also used attentive response maps,[269–273] trying to achieve some interpretability

Table 1. Summary of material properties predicted with machine learning methods and corresponding references

Property	References
Curie temperature	31,283–287
Vibrational free energy and entropy	288
Band gap	40,41,132,159,283,289–300
Dielectric breakdown strength	38,44,45
Lattice parameter	300
Debye temperature and heat capacity	41–43
Glass transition temperature	301,302
Thermal expansion coefficient	41
Thermal boundary resistance	303
Thermal conductivity	37,46–51,304,305
Local magnetic moments	127,306
Melting temperature	39,48,307
Magnetocaloric effects	283
Grain boundaries	308
Grain boundary energy	309–312
Grain boundary mobility	312
Interface energy	300
Seebeck coefficient	46,313,314
Thermoelectric figure of merit	315
Bulk and shear moduli	40–42,132,184,185,316
Electrical resistivity	46
Density of states	109,317,318
Fermi energy and Poisson ratio	40
Dopant solution energy	319
Metal–insulator classification	65
Topological invariants	320–326
Superconducting critical temperature	73,76,122,327–329
Li-ion conductivity and battery state-of-charge	65,330,331

and visualization of the model. This will be further discussed in section "Discussion and conclusions—Interpretability".

Further interesting work concerning micro-structure and material characterization, using machine learning-based image processing, can be found in refs. [274–279].

If we consider structure prediction through machine learning, we also have to consider global structure prediction methods, where the whole energy surface has to be explored efficiently. This can also be considered as a surrogate-based optimization problem (see section "Adaptive design process and active learning"), where the expensive experiment is the local geometry optimization through DFT. Yamashita et al.[280] started with a large set of initial structures from which a random subset was locally optimized and used to train a Bayesian regressor to predict the energy. Using Thompson sampling,[281] structures from the initial set were sampled randomly according to their probability of minimizing the energy. The approach was tested on NaCl and Y_2Co_{17} and reduced the average number of trials until finding the optimal structure by, respectively, 31% and 39% when compared to random structure selection.

Lastly, we discuss two works that introduced modern neural network architectures to crystal structure prediction and generation. Both methods have also been used recently for micro-structures by Li et al.[210,282]

Ryan et al.[28] applied VAEs (see section "Basic principles of machine learning—Algorithms") to crystal structure prediction. The 42-layer VAEs develop a more efficient representation for the input (see section "Basic principles of machine learning—Features"). For training and testing, a dataset of around 50,000 crystal structures from the inorganic crystal structure database[79] and the crystallography open database[89] were used. The encoding of the original descriptors was used as input for a five-layer sigmoid classifier that predicts the most likely elements to form the topology represented by the atomic fingerprints. A third auxiliary neural network, in this case a five-layer softmax classifier, combined the non-normalized atomic fingerprints and the output of the sigmoid classifier that improves the prediction. To predict directly the crystal structure from this approach, one requires training data of negatives or, in other words, knowledge of crystal structures that do not exist. Unfortunately, no such database is available in physics. In order to circumvent this problem, Ryan et al. calculated the likelihood of the existence of a structure as the product of the probabilities of elements existing at the single atomic sites. Application to test data demonstrated a clear superiority of this approach in comparison with random choices.

Nouira et al.[29] introduced a GAN-based strategy (see section "Basic principles of machine learning—Algorithms") to crystal structure generation in the form of CrystalGAN. Specifically, they created a novel GAN structure to generate stable ternary structures on the basis of binary hydrides. Remarkably, the method generates structures of higher complexity and is able to include constraints based on domain knowledge. However, as no data about the stability of the generated structures were published, the evaluation of the usefulness of this approach is still pending. A second application of GANs in materials science, and in particular in chemistry, can be found in ref. [211].

PREDICTION OF MATERIAL PROPERTIES

Machine learning methods have proven to be successful in the prediction of a large number of material properties. An overview of different properties that were predicted can be found in Table 1. In the following, we discuss in depth a few properties, studied in various works, which provide good examples for current challenges in computational materials science, and possible strategies to overcome them.

Band gaps

Design of functional materials for applications like light-emitting diodes (LEDs), photovoltaics, scintillators, or transistors, always requires detailed knowledge of the band gap. Consequently, a lot of effort was invested in theoretical methods for high-throughput calculations of this electronic property. It is well known that standard exchange correlation functionals, like the PBE,[145] systematically underestimate band gaps in comparison to experimental results. More modern functionals like the modified Becke–Johnson by Tran and Blaha[332] or the strongly constrained and appropriately normed meta-GGA[333] by Jianwei et al. improve upon these results. However, the state-of-the-art higher-fidelity methods still remain the many-body *GW* approximation or hybrid functionals. Unfortunately, these usually come with a prohibitively high computational cost. Machine learning is one possibility to overcome this obstacle by either directly predicting band gaps based on experimental or theoretical training data or by using the results of low-fidelity methods to predict experimental or high-fidelity theoretical results.

Zhuo et al.[289] tried to circumvent the problems of the different theoretical methods by directly predicting experimental band gaps. Their approach started with a classification of the materials as either metal or non-metal using SVM classifiers and then progressed by predicting the band gap with SVM regressors. The performance of the resulting models in predicting experimental band gaps lies somewhere between basic functionals (like the PBE) and hybrid functionals. The error turns out to be comparable to, e.g., ref. [40,41]. However, Zhuo et al. improved upon those earlier

machine learning results, as the error is with respect to the experimental results instead of DFT calculations. While there have been earlier attempts at using experimental band gap training data (e.g., ref. [290]), the dataset used by Zhou et al. includes >6000 band gaps, dwarfing all previous datasets.

Lee et al.[291] approached the problem from a different perspective by using low-fidelity DFT gaps (modified Becke–Johnson and PBE), as well as basic crystalline and elemental properties as features for optimized link state routing, LASSO, and nonlinear SVR. They predicted gaps calculated with G_0W_0 starting from the ground-state obtained with the Heyd–Scuseria–Ernzerhof hybrid functional.[334] SVR using both the modified Becke–Johnson and the PBE gaps, as well as the other features, yielded the best results, with a root mean square error of 0.24 eV.

Pilania et al.[292] applied a co-kriging statistical learning framework to learn high-fidelity band gaps. Their approach differed from previous ones, as low-fidelity band gaps were not explicitly used as features and were therefore not necessarily needed as input for all compounds.

Another interesting attempt at the prediction of high-fidelity band gaps can be found in ref. [293]. Rajan et al. used KRR, SVR, GPR, and decision tree boosting methods to predict the G_0W_0 band gaps of MXenes. They started by generating and selecting features with LASSO[159] and then optimizing the feature space for each method. Counterintuitively, the PBE band gap was not included in the optimized feature space of any method. However, other researchers suggest to include this information[283,294] and stress the importance of so called crude estimations of property[77] (see section "Basic principles of machine learning—Features").

Weston et al.[295] investigated the band gaps of kesterite compounds and developed a logistic regression classifier for the prediction of the direct–indirect property of these band gaps. A total of 184 semiconducting materials were used for training, and the best model demonstrated an accuracy, recall, precision, and f_1 score of around 90%.

Bulk and shear moduli

Two other popular properties in solid-state machine learning are the bulk and shear moduli, which determine the stress–strain relations in the linear range. They are also correlated with other properties like the bonding strength, thermal conductivity,[184,335,336] charge carrier mobility,[337] and of course the hardness of the material.[338,339] As such, they are often used as a proxy in the search for superhard[340] (hardness >40 GPa) materials. In general, these properties are available as a result of DFT calculations; however, they are too computationally expensive for really large high-throughput studies. Two less computationally expensive alternatives exist, specifically force-field methods[341] and theoretical models for the direct calculation of bulk and shear moduli. However, force fields lack accuracy, and most theoretical models only span a highly restricted chemical and structural space.[342–345] This opens up the question whether machine learning algorithms can show better generalizability.

de Jong et al.[184] developed a new machine learning technique, called gradient boosting machine local polynomial regression, that extends the principles of gradient boosting frameworks[180] to the case of multivariate local polynomial regression.[346] They used this technique to predict the Voigt–Reuss–Hill averages[347] of the bulk and shear moduli on the basis of elemental properties. In this case, they used the volume per atom, row number, cohesive energy, and the electronegativity as features. The use of the cohesive energy as a feature is slightly problematic as it also requires DFT calculations. The training set consisted of around 2000 materials and a root mean square error of 0.075 log(GPa) and 0.138 log(GPa) were reached for the logarithm of the bulk and shear moduli. The logarithm was used to decrease the emphasis on large values. It has to be noted that the training set was biased toward metallic compounds and rather simple materials.

The previously discussed CGCNNs by Xie et al.[40] also allows for the prediction of bulk and shear moduli. Test set errors of 0.105 log(GPa) and 0.127 log(GPa) for these properties were reported for the dataset of ref. [184]. The network was also tested on 1585 materials that were recently added to the materials project database.[78] Once again the network demonstrated good generalizability for the new dataset with different crystal groups. Doubling the size of the training data, the MEGNet model of Chen et al.[132] obtained around 10% lower errors for bulk and shear moduli.

A similar study for siliceous zeolites was performed in ref. [185], where gradient boosting regressors were used to predict once again the logarithm of the bulk and shear moduli. They obtained an error of 0.102 ± 0.034 log(GPa) for the bulk and 0.0847 ± 0.022 log(GPa) for the shear moduli. Even if the training set only contained 121 zeolites, this method seems to compare favorably to the 5 conventional force field methods[348–352] reported in ref. [353]. In contrast to ref. [184], Evans et al. used structural and local descriptors as the challenge was to differentiate between the different siliceous zeolites and not between materials of different elemental composition.

Furmanchuk et al.[42] used RFs to predict the bulk modulus. A wide variety of 1428 compounds from the thermoelectric design lab database,[91] containing from unitary up to quinary combinations of 62 elements, was used for training. The notable fact about this study is that thermal effects, which are usually neglected in DFT calculations, were included through Birch and Murnaghan fits.[354,355] As features, properties of the element itself and experimentally measured properties of elemental substances were used. Another set of 356 theoretically calculated materials and 69 experimentally measured ones was kept for testing. A root mean square error of 18.75 GPa was reported for the first set, while no error was reported for the experimental set.

Isayef et al.[41] developed an extension of property-labeled material fragments to be used for solids. As this leads to a very general feature vector, one can apply it to the prediction of a variety of properties.[41] Using gradient boosting decision trees and a training set of around 3000 materials, they achieved errors of 14.25 and 18.43 GPa for the bulk and shear moduli, respectively. It has to be noted that this training set only considered unary to ternary compounds and neglected quaternary compounds. In contrast, these were also considered in, e.g., refs. [184,185,316].

Another interesting machine learning study of the bulk and shear moduli of solids is ref. [316]. Mansouri et al. combined elemental and structural properties as descriptors and used SVRs to screen a chemical space of around 120,000 materials for superhard incompressible materials. This was actually followed by the synthesis and characterization of two novel superhard materials. Once again, the cohesive energy was identified as one of the crucial features for both moduli.

Topological states

The discovery of topological insulators has sparked an extreme interest into the field of topological states in condensed matter.[356–358] It is therefore not surprising that in the past 2 years machine learning ansätze were introduced to the topic. In general, learning topological phases is a highly non-trivial task as topological invariants are inherently non-local. In the field of topological states, neural networks are by far the most relevant machine learning method used.[320–326] In refs. [321,323,324], this technique was used to predict the topological invariants of, respectively, 1D topological insulators of the A 3 class, the 2D XY-model, and 1D topological insulators of the A 3 class, as well as 2D insulators of the A class. In the later two works, analysis of the neural network confirmed that it learned both the winding

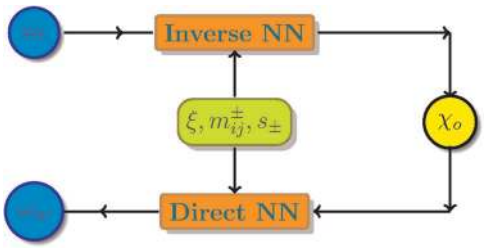

Fig. 16 Strategy to solve the inverse design problem for photonic devices,[325] where χ_0 is the structure parameter suggested by the inverse network and ξ, $m_{ij}^{\pm}$, and $s_{\pm}$ are extra inputs for both networks. (Reprinted with permission from ref. [325] licensed under the CC BY 4.0 cc license.)

number formula[321] and a formula for the Berry curvature in the case of the A class insulators[323] (see section "Discussion and conclusions—Interpretability" for a more extensive discussion). Another interesting application that takes advantage of the extreme success of neural networks for image classification is quantum loop topography.[320] In this method, an image representing the Hamiltonian or wave function is constructed and entered into a neural network that decides on the topological phase of the system with great accuracy. Although the input is the result of Monte Carlo simulations, it is rather efficient, as it only requires single steps and not Monte Carlo averages.

While the previous examples are still mostly concerned with theoretical models, more recent work is already concerned with designing topological photonic devices directly through machine learning methods.[325] Pilozzi et al. designed photonic devices described by the Aubry–André–Harper model[359,360] with neural networks (see Fig. 16). The desired property are edge states with specific frequencies ω_t, which are determined by a set of structure parameters. In order to solve the problem, direct and inverse models are combined. The process starts with the inverse neural network, determining the structure parameters required for an edge state with frequency ω_t^{ind}. In order to simplify the problem, some categorical features are not included in the neural network, but actually one neural network is trained for each categorical value. The obtained structure parameters from the inverse neural network are used as input for the direct neural network that produces a new frequency ω_t^{dir}. If the discrepancy between the two frequencies is smaller than a certain threshold $\omega_t^{\text{ind}} - \omega_t^{\text{dir}} < \delta$, the structure parameters from the indirect neural network are accepted. This self-consistent approach is used to filter out the unphysical structures from the results of the inverse neural network.

Superconductivity

Even 30 years after its discovery,[361] unconventional superconductivity remains one of the unsolved challenges of theoretical condensed matter physics. As machine learning methods do not require a complete theoretical understanding of the problem, determining the critical temperature T_c is an obvious challenge for these methods. In the case of critical temperatures, data accumulation is problematic, as there are few computational methods to calculate critical temperatures,[362–364] and these are limited to conventional superconductors. Moreover, they are far less widely available than, e.g., methods to calculate the band gap or bulk moduli. On the one hand, this is a drawback as it severely limits the acquisition of data, but on the other hand machine learning methods could prove even more important as no general working theoretical model exists.

There was some early work, akin to machine learning, on clustering superconductors based on quantum structure diagrams[365,366] and some more recent work concerning the filtering of materials for cuprate superconductors based on their electronic structure.[367] A discussion of similar design approaches can be found in ref. [368]. In refs. [327,328], the superconducting critical temperature is fitted to the lattice parameters with an SVM. Unfortunately, both studies clearly suffer from the difficulty of accumulating data. The former is concerned with iron-based superconductors and has a training set of 30 materials while the latter only treats doped MgB_2 with a training set of 40 materials. Even though these examples do not take advantage of the fortes of machine learning methods, they still reach an error 1.17 K and 1.10 K, admittedly for a very limited domain. The actual search for superconductors in a larger domain is far more challenging, because the Kohn–Luttinger theorem[369] suggests that fermionic systems with a Coulomb interaction are in general superconducting for $T \to 0$. This presents a difficulty, as leaving compounds with no reported critical temperature out of the dataset, or assuming that critical temperature is zero, would either lead to a misrepresentation or underrepresentation of data.[76] However, as we are often interested in high-temperature superconductors from a technological perspective, we can circumvent the problem by classifying potential superconductors as low or high T_C, instead of using a regressor to predict the critical temperature.

Isayev et al.[122] used RFs to divide superconductors into groups, one group with T_C below and one group with T_C >20 K and RFs and partial least squared regression to build a continuous model of the transition temperature. The training set size was nevertheless still very limited (464 classification, 295 regression).

A study by Stanev et al.[76] considered a larger training set of around 14,000 materials from the SuperCon database.[82] Superconductors were first classified into groups with T_c below and above 10 K, resulting in an accuracy and F_1 score of about 92%. The features were created using Magpie[370] and consisted of elemental properties and combinations of them. Interestingly enough, when reducing the number of descriptors to only the three used in refs. [365,366], specifically the average number of valence electrons, the metallic electronegativity differences, and orbital radii differences, the accuracy of the classifier only decreased by around 3%. This suggests that little progress was made in terms of such features in the meantime.

The regression model for $\log(T_C)$ was built for materials with transition temperatures >10 K to avoid the previously discussed problems and reached an R^2 score of around 0.88. By dividing the training set into different groups of superconductors, Stanev et al. could demonstrate that the model recovered physical knowledge, such as the isotope effect or other empirical relations.[371] Furthermore, it was clear that the model was not able to extrapolate from one group of superconductors to another, e.g., from conventional to cuprate superconductors. This is, of course, expected owing to the different superconducting mechanisms involved in the two families. This extrapolation problem of materials science machine learning models and methods to estimate it are also discussed in ref. [73]. Finally, Stanev et al. applied the classifier and regressor to the materials in the inorganic crystal structure database[79] and scanned it for new high-T_C superconductors. As a byproduct, a feature of the band structure that is known to increase the T_C was recovered even though no electronic structure data were used in the model.

Ling et al.[329] also used RFs in research concerning the design process of materials including superconductors, but as their research is mainly concerned with the optimization of the design process, we will discuss it later in the section "Adaptive design process and active learning". Another interesting study[372] of superconductors applied *k*-means clustering, PCA, and Bayesian linear unmixing to scanning tunneling microscopy data in order to extract meaningful data regarding electronic interactions in the spin-density wave regime. Note that these are expected to play a key role in the existence of unconventional superconductivity.

ADAPTIVE DESIGN PROCESS AND ACTIVE LEARNING

The previous chapters were concerned with the prediction of the stability, atomic structure, and physical properties. Necessarily, all of these methods have the end goal of minimizing the time until a new optimal material with tailored properties is found. This can either imply the minimization or maximization of a single property or the search for a material on the Pareto front in the case of multiple objectives. In order to reach this goal, we aim at reducing the number of "experiments" that have to be carried out, as these are the most time consuming and expensive segment of the discovery process. In our case, experiment may denote computationally expensive calculations, like the ones necessary to obtain the phonon and electron transport properties required for the design of thermoelectrics. A more general discussion of such optimization problems can be found in the literature under the name of surrogate-based optimization[373,374] and active learning.

The adaptive design process consists of two interwoven tasks: (i) A surrogate model has to be developed; (ii) Based on the prediction of the surrogate model, optimal infill points have to be chosen in order to retrain the surrogate model and finally find the optimum.

The challenge in this process is to balance the end goal of finding the best material (exploitation) with the need to explore the space of materials in order to improve the model.[375] The most naive strategy is naturally pure exploitation, in which case the design algorithm always chooses the material with the highest prediction of the target value (lowest in the case of minimization). From other fields, e.g., drug design[376,377] and quantitative structure–activity relationship in chemistry,[378,379] it is already known that such an unsophisticated approach is far from optimal. More sophisticated policies, such as maximum likelihood of improvement or maximum expected improvement, try to strike a balance between these strategies. However, choosing the optimal experiment according to one's strategy requires machine learning models that not only return predictions but also the uncertainty of a prediction.

Starting from this requirement, the most obvious algorithm choice are Bayesian prediction models like Gaussian processes[380] as they also provide the variance of the predicted function. Gaussian processes have been applied to a wide variety of structure optimization and design problems in materials science. A few examples are optimizing thermal conductance in nanostructures,[47] predicting interface[309] and crystal[280] structures, optimizing materials for thermoelectric[381,382] and optoelectric[382] devices, or optimizing GaN LEDs.[383] Furthermore, these studies already resulted in successfully synthesized materials.[384,385] We already discussed in the section "Basic principles of machine learning—Algorithms" that the inherent scaling of Gaussian processes both with respect to training set size as well as feature dimension is quite bad.[157] At the moment, a lot of adaptive design studies still treat extremely small datasets (see, e.g., ref. [384] with a training set size of 22), in which case this is irrelevant. However, a large number of the previously discussed models for stability or property prediction use high-dimensional descriptors and are therefore also unsuitable for Bayesian methods.[329,386]

One alternative to Bayesian predictors are standard machine learning algorithms, like SVRs or decision tree methods, in combination with bootstrapping methods to estimate the uncertainty. In ref. [375], Balachandran et al. compared different surrogate models and strategies on a set of M_2AX compounds for the optimization of elastic properties. From a pure prediction perspective, SVRs with radial basis function slightly outperformed Gaussian processes for training set sizes >120 materials. Different design strategies were then used in combination with the SVR. It turned out that efficient global optimization,[387] as well as knowledge gradient,[388] showed the best results. Xue et al.[384] obtained similar results concerning the choice of algorithms for the composition optimization of NiTi-based shape memory alloys. Starting with a set of 22 materials, Xue et al. successfully synthesized 14 materials (from a total of 36 synthesized in total during 9 feedback loops), which were superior to the original dataset.

Balachandran et al.[389] also applied SVRs in combination with efficient global optimization to the maximization of the band gap of $A_{10}(BO_4)_6X_2$ apatites. In this case, the performance for two feature sets, one containing the Shannon ionic radii and the other one the Pauling electronegativity differences was compared. Interestingly, the design based on the ionic radii performed better, finding the optimal material after 22 materials (13 materials in the initial training set, 9 chosen by the design algorithm) in comparison to 30 for the electronegativities, while having a far larger error in the machine learning model (0.54 eV compared to 0.19 eV for the electronegativities). The result is most likely due to the fact that, of the three atomic species considered for the B-site (P, V, As), P provides clearly higher band gaps than the other elements and has a different ionic radius while the electronegativity of P and As are nearly the same.[389] Using this information, the algorithm eliminated all compositions without P on the B-site. This example demonstrates that sometimes the algorithm with the highest predictive power will not necessarily lead to the best optimal design results. A combination of the two predictors leads to even better results with the optimal composition after one iteration; however, the mean absolute error of the model was still slightly worse (0.21 eV) than the one of the purely electronegativity-based model.

Ling et al.[329] treated a high-dimensional (with respect to the descriptor space) materials design problem with the RF framework FUELS.[390] By adding a bias term to the uncertainty, which accounts for noise and missing degrees of freedom, they expanded upon previous uncertainty estimates from refs. [391,392]. Tested on 4 datasets (magnetocaloric, thermoelectric, superconductors, and thermoelectric) with higher descriptor number (respectively, 54, 54, 56, 22), FUELS compared favorably with the Bayesian framework COMBO and random sampling, while being roughly an order of magnitude faster. In order to evaluate various selection strategies or model algorithms, different metrics were used. In materials science, a commonly used metric is the number of experiments until the optimal material is found. While this metric has some merit, in most cases opportunity cost (the distance of the current best from the overall best) or the number of experiments until the current best is within a specific distance (e.g., 1%) is superior and is also used more often in the literature.[393,394]

Monte Carlo tree searches[395] are a second algorithm with superior scaling that has recently been introduced to materials science. The application is inspired by its success in go,[2] where a combination of neural networks, reinforcement learning, and Monte Carlo tree search allowed for the first superhuman performance in this ancient strategy game. Dieb et al.[396] implemented a materials design version in the form of the open source library MDTS. Using the test case of the optimal design of thermoelectric Si-Ge alloys, they demonstrated that, although Bayesian optimization has advantages for small problems due to its advanced prediction abilities, Monte Carlo tree search design time stays close to constant (see Fig. 17) with increasing problem size. Furthermore, and in contrast to genetic algorithms, it does not require the determination of hyperparameters. Owing to the unfavorable scaling of Bayesian optimization, at some point the computational effort of the design becomes larger than the computational effort of the experiments, at which point Monte Carlo methods become superior. For the interface structure optimization in ref. [396], this is already the case for interfaces with >22 atoms. Further applications to the determination of grain boundary structures[397] and the structure of boron-doped graphene[398] also demonstrate the viability of the method for

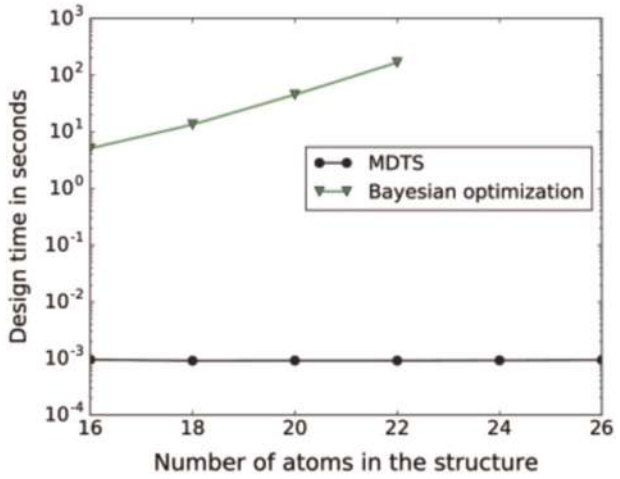

Fig. 17 Design time of Bayesian optimization and Monte–Carlo tree search for different numbers of atoms in the interface. (Reprinted with permission from ref. [396] licensed under the CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).)

structure design problems. A more in-depth review of Bayesian optimization and Monte Carlo tree search in materials design can be found in ref. [399].

Sawada et al.[400] also developed an algorithm for multi-component design based on game tree search. Optimizing the composition in a seven-component Heusler compound, the algorithm proved to be around nine times faster than expected improvement or upper confidence bound[401] strategies based on Gaussian processes.

Dehghannasiri et al.[402] proposed an experimental design framework based on the mean objective cost of uncertainty. This is defined as the expected difference in cost between the material, which minimizes the expected cost for a surrogate model and the optimal material.[403] Applying the framework to the minimization of the dissipation energy of shape memory alloys demonstrated the superiority of the algorithm to pure exploitation and random selection.

So far, none of the discussed algorithms considered nested decision problems or cases where it is more efficient to carry out experiments in batches of similar experiments instead of one at a time. The latter is, for example, true for the case of the photoactive device design considered by Wang et al.[404] The size of thiol-gold nanoparticles and their density on the surface determine the efficiency of the device. While one can easily explore different densities of nanoparticles in a batch of experiments, it is difficult to change the size of the nanoparticle due to the cost of their synthesis. Therefore, it is more efficient to consider a nested problem where the algorithm first chooses a size and then a batch of densities. Wang et al. extended the concept of knowledge gradient[388] to the case of nested decisions and batches of experiments. Applying it to the previously described design problem, the new algorithm proved to be superior to all naive strategies (pure exploitation/exploration, or ε-greedy which chooses either pure exploration or exploitation with probability ε) and also to sequential knowledge gradient (batch size 1) if one considers the number of batches. If one instead considers the total number of experiments, the performance of knowledge gradient was only slightly better.

If we consider typical design problems, one often has to consider multiple objectives. For example, for the design of a shape memory alloy, one desires a specific finish temperature, thermal hysteresis, and possibly a high maximum transformation strain. Naturally, this requires more sophisticated measures of improvement (see ref. [405] for a review) than single objective optimization methods. A typical measure is the expected hypervolume improvement[406] that measures the change in hypervolume of the space dominated by the best known materials. Solomou et al.[407] applied this metric to the optimization of shape memory alloys in combination with a Gaussian process model, once for two objectives (specific finish temperature and thermal hysteresis) and once for three objectives (adding the maximum transformation strain), and demonstrated that it is clearly superior to a random or purely exploitative strategy.

Talapatra et al.[408] also combined expected hyper-volume improvement with Gaussian processes in order to simultaneously maximize the bulk modulus while minimizing the shear modulus. Instead of using a single Gaussian regressor, they developed a method called Bayesian model averaging, which combines different models. This approach can prove useful in cases where the available data is too limited to choose good features or hyperparameters.

Gopakumar et al.[409] compared both SVRs and Gaussian processes on multiple datasets: optimal thermal hysteresis and transition temperature for shape memory alloys, optimal bulk and Young's modulus for M_2AX phases, and optimal piezoelectric modulus and band gap for piezoelectric materials. SVRs performed better as regressors and were consequently chosen as surrogate model. Several optimal design strategies were used, specifically random, exploitation, exploration, centroid, and maximin. For the smallest dataset, maximin surprisingly performed only slightly better for large experimental budgets and worse than pure exploitation for small budgets. However, for the larger dataset of elastic moduli both centroid and maximin proved to be clearly superior.

An additional popular choice of global optimization algorithms that can also be applied to adaptive design, especially to structure development, are genetic algorithms. Reviews of their application to materials design can be found in refs. [230,410].

It is difficult to compare the ability of the different optimal design algorithms and frameworks discussed in this section because no systematic study has ever been carried out. Nevertheless, it is quite clear that, given sufficient data, adaptive design algorithms produce superior results in comparison to naive strategies like pure exploration or exploitation, which are unfortunately still extremely common in materials science. Furthermore, several works demonstrated that experimental resources are used more efficiently if they are allocated to the suggestions of the design algorithm instead of a larger initial random training set. Machine learning models can be quite limited in their accuracy; however, the inclusion of knowledge of this uncertainty in the design process can alleviate these limitations. This allows for a feedback cycle between experimentalists and theoreticians, which increases trust and cooperation and reduces the number of expensive experiments.

MACHINE LEARNING FORCE FIELDS

As previously discussed, first-principle calculations can accurately describe most systems but at a high computational price. Usually this price is too high for use in molecular dynamics, Monte Carlo, global structural prediction, or other simulation techniques that require frequent evaluations of the energy and forces. Even DFT is limited to molecular dynamics runs of a few picoseconds and simulations with hardly more than thousands of atoms. For this reason, the research concerning empirical potentials and the development of models for the potential energy surfaces never faded away.

In fact, most molecular dynamics simulations are normally computed with classical force fields.[411–417] As these potentials often scale linearly with the number of atoms, they are computationally inexpensive and the loss in accuracy is overlooked in favor of the possibility to perform longer simulations or

simulations with hundreds of thousands or even millions of atoms. Another approach is DFT-based tight binding.[418–420] This quantum–mechanical technique scales with the cube of the number of electrons but has a much smaller prefactor than DFT. Certainly, calculations performed with this method are not as accurate as in DFT, but they are more reliable than classical force field calculations. In addition to the reduced precision, the construction of force fields and tight-binding parameters is unfortunately not straightforward.

Neural networks were the first machine learning method used in the construction of potential energy surfaces. As early as 1992, Sumper et al.[421] used a neural network to relate the vibration spectra of a polyethylene molecule with its potential energy surface. Unfortunately, the large amount of input data and architecture optimization required deemed this approach as too cumbersome and difficult to apply to other molecular systems. It was the work of Blank et al.[422] in 1995 that really showed the potential, and marked the birth, of machine learning force fields. Their work on the surface diffusion of CO/Ni(111) relied on neural network potentials, which mapped the energy of a system with its structure, mainly the lateral position of the center of mass, the angle of the molecular axis relative to the surface normal, and the position of the center of mass. The training set was obtained from electronic structure calculations and no further approximations were used. Their seminal study proved that neural networks could be used to make accurate and efficient predictions of the potential energy surface for systems with several degrees of freedom.

Since then, many machine learning potentials were reported. As several reviews on these potentials can be easily found in the literature,[112,423–425] here we discuss only the most prominent and recent approaches related to materials science.

One of the most successful applications of machine learning to the creation of a reliable representation of the potential energy surface is the Behler and Parrinelo approach.[110] Here the total energy of a system is represented as a sum of atomic contributions E_i. This became the standard for all later machine learning force fields, as it allows their application to very large systems. In the Behler–Parrinelo approach, a multilayer perceptron feedforward neural network is used to map each atom to its contribution to the energy. Every atom of a system is described by a set of symmetry functions, which serve as input to a neural network of that element. Every element in the periodic table is characterized by a different network. As the neural network function provides an energy, analytical differentiation with respect to the atomic positions or the strain delivers, respectively, forces and stresses. This approach was originally applied to bulk silicon, reproducing DFT energies up to an error of 5 meV/atom. Furthermore, molecular dynamic simulations using this potential were able to reproduce the RDF of a silicon melt at 3000 K. Many applications of this methodology to the field of materials science have appeared since then, for example, to carbon,[426] sodium,[427] zinc oxide,[428] titanium dioxide,[111] germanium telluride,[429] copper,[430] gold,[431] and Al-Mg-Si alloys.[432]

Since its publication in 2007, several improvements were made to the Behler and Parrinelo approach. In 2015, Ghasemi et al. proposed a charge equilibration technique via neural networks,[433] where an environment-dependent atomic electronegativity is obtained from the neural networks and the total energy is computed from a charge equilibration method. This technique successfully reproduced several bulk properties of CaF_2.[434] In 2011, the cost function was expanded to include force terms.[424,428] This extension was first proposed by Witkoskie et al.[435] and later extended and generalized by Pukrittayakamee et al.[436,437] These works show that the inclusion of the gradients in the training substantially improves the accuracy of the force fields, not only due to the increase of the size of the training set but also due to the additional restrictions in the training. Hajinazar et al.[438] devised a strategy to train hierarchical multicomponent systems, starting with elemental substances and going up to binaries, ternaries, etc. They then applied this technique to the calculations of defects and formation energies of Cu, Pd, and Ag systems and were able to obtain an excellent reproduction of phonon dispersions. Another improvement concerns the replacement of the original Behler–Parrinelo symmetry functions by descriptors that can be systematically improved. One such descriptor is given by Chebyshev polynomials,[117] which also allow for the creation of potentials for materials with several chemical elements, due to its constant complexity with respect to the number of species. Potentials constructed with this descriptor are the reported machine learning potentials that can describe more chemical species, with 11 so far.

Artrith et al.[439] proved the applicability of specialized neural network potentials in their study of amorphous Li–Si phases. They compared the results obtained with two different sampling methods. The first involved a delithiation algorithm, which coupled a genetic algorithm with a specialized potential trained with only 725 structures close to the crystalline Li_xSi_{1-x} phase. The second method consisted of an extensive molecular dynamics heat-quench sampling and a more general potential. Figure 18 shows the accuracy of the latter neural network potential.

We note that not only machine learning methods are changing the field of materials science but also machine learning methodologies. The spectral neighbor analysis potential[440] from Thompson et al. consists of a linear fit that associates an atomic environment, represented by the four-dimensional bispectrum components, with the energies of solids and liquids. The first application of these potential to tantalum showed promising results, as it was able to correctly reproduce the relative energy of different phases. Furthermore, in the application of this potential to molybdenum by Chen et al.,[441] PCA was used to examine the distribution of the features in the space. This technique increases the efficiency of the fitting, as it ensures a good coverage of the feature space and reduces the number of structures in the training set. Their potential achieved good accuracies for energies and stresses (9 meV and 0.9 GPa, respectively). Although the accuracy in the forces was considerably worse (0.30 eVÅ), they also managed to reproduce correctly several mechanical properties such as the bulk modulus, lattice constants, or phonon dispersions (see Fig. 19). Wood et al.[442] proposed an improvement of the

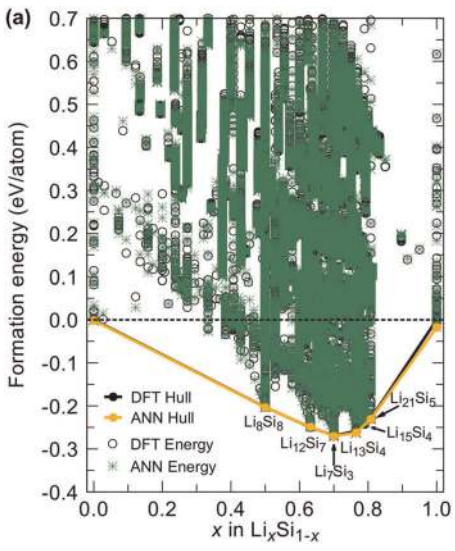

Fig. 18 Phase diagram of 45,000 Li_xSi_{1-x} structures depicting the formation energies predicted using the general neural network potential (green stars) and the density functional theory reference formation energies (black circles). (Reprinted from ref. [439], with the permission of AIP Publishing.)

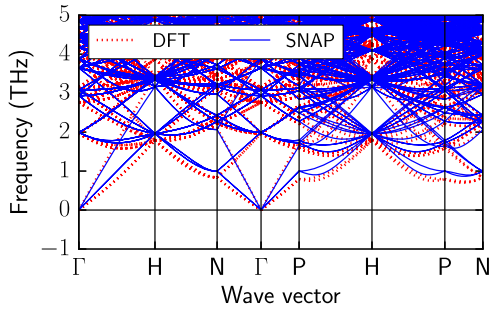

Fig. 19 Comparison between the phonon dispersion curves obtained with density functional theory and the spectral neighbor analysis potential model for a $5 \times 5 \times 5$ supercell of Mo. (Reprinted with permission from ref. [441]. Copyright 2017 American Physical Society.)

model that consisted in the introduction of quadratic terms in the bispectrum components and Li et al.[443] introduced a two-step model fitting work-flow for multi-component systems and applied it to the binary Ni–Mo alloy.

Other linear models include the work of Seko et al.,[113] who reproduced potential energy surfaces to Na and Mg using KRR and LASSO combined with the multinomial expansion descriptor (see section "Basic principles of machine learning—Features"). Phonon dispersion and specific heat curves calculated with the LASSO technique for hcp-Mg were in good agreement with the DFT results. Using a similar methodology, Seko et al. applied elastic net regression,[444,445] a generalization of the LASSO technique, to 10 other elemental metals[446] (Ag, Al, Au, Ca, Cu, Ga, In, K, Li, and Zn). The resulting potential yielded a good accuracy for energies, forces, and stresses, enabling the prediction of several physical properties, such as lattice constants and phonon spectra.

In a different approach, Li et al.[447] devised a molecular dynamics scheme that relies on forces obtained by either Bayesian inference using GPR or by on-the-fly quantum mechanical calculations (tight binding, DFT, or other). Certain simulations in materials science involve steps where complex, recurring, chemical bonding geometries are encountered. The principal idea behind this scheme is that an adaptive approach can handle the occurrence of unseen geometries while the recurring ones are trained for. This is achieved by the following predictor-corrector algorithm[448,449]: After n steps of the simulation with a force field, the latest configuration is selected for quantum mechanical treatment and the accuracy of the force field is tested. Should the accuracy fall below a certain threshold, the force field is refitted. This scheme might not be the most efficient for a singular molecular dynamics cycle but excels when the simulations involve monotonic cycles between two temperatures, for example. Applications to silicon,[447] aluminum, and uranium[450] (with linear regression) reveal accuracies for forces <100 meV/Å. The phonon density of states and melting temperature of aluminum obtained with this scheme are also in good agreement with ab initio calculations. In the same spirit, Glielmo et al.[451] employed vectorial Gaussian process[452,453] regression to predict forces using vector two-body kernels of covariant nature. Their results for nickel, silicon, and iron indicate that the inclusion of symmetries results in a more efficient learning and that it is not necessary to impose energy conservation to achieve force covariance. Additional improvements of this methodology include the replacement of the features by higher-order n-body-based kernels.[454]

Another family of highly successful machine learning potentials is the Gaussian approximation potentials (GAPs). First introduced by Bartók et al.,[114] these potentials interpolate the atomic energy in the bispectrum space using GPR. Tests for semiconductors and iron revealed a remarkable reproduction of the ab initio potential energy surface. Advances in this methodology include the replacement of the bispectrum descriptor by the SOAP descriptor and the training of not only energies but also forces and stresses,[455] the generalization of the approach for solids[456] by adding two- and three-body descriptors, and the possibility to compare structures with multiple chemical species.[457] The materials studied in these works were tungsten, carbon, and silicon, respectively. The application of the GAPs to bcc ferromagnetic iron by Dragoni et al.[458] proves the accuracy of these potentials for both DFT energetics and thermodynamical properties. In particular, bulk point defects, phonons, the Bain path, and Γ surfaces[459] are correctly reproduced. By combining single-point DFT calculations, GAPs, and random structure search,[220,221] Deringer et al. showed a procedure that simultaneously explores and fits a complex potential energy surface.[460] They used 500 random structures to train a GAP model, which was then used to perform the conjugate gradient steps of the random search. The minimum structures were added to the training set after being recalculated with single-point DFT calculations. The potential for boron resulting from this procedure was able to describe the energetics of multiple polymorphs, which included αB_{12} and βB_{106}.

The GAP methodology was also applied to graphene.[461] The potential constructed by Rowe et al. was able to reproduce DFT phonon dispersion curves at 0 K. In addition, the potential predicted quantitatively the lattice parameter, phonon spectra at finite temperature, and the in-plane thermal expansion. Other works concerning GPR include its application to formaldehyde and comparison of the results with neural networks[462] and the acceleration of geometry optimization for some molecules.[463]

Jacobsen et al. presented another structure optimization technique based on evolutionary algorithms and atomic potentials constructed using KRR.[464] To represent the atomic environment, they used the fingerprint function proposed by Oganov and Valle.[465] By using the atomic potentials to estimate the energy, they were able to reach a considerable speed-up of the search for the global minimum structure of SnO_2(110)-(4 × 1).

In an unconventional way to construct atomic potentials, Han et al.[466] presented a deep neural network that, for each atom in a structure, takes as input N_c functions of the distance between the atom and its neighbors, where N_c is the maximum number of neighbors considered. As a consequence, some of the inputs of the neural network have to be zero. Furthermore, the potential might have transferability problems if ever used on a structure with smaller inter-atomic distances than the ones considered in the training set. Nevertheless, their potential showed good accuracy in energy predictions for copper and zirconium. Zhang et al.[467] improved this methodology with the generalization of the loss function to include forces and stresses.

DFT FUNCTIONALS

The application of machine learning techniques also spread to the creation of exchange and correlation potential and energy functionals. The first application emerged from the work of Tozer et al.[468] in 1996, where they devised a one-layer feed-forward multiperceptron neural network to map the electronic density $\rho(r)$ to the exchange and correlation potential $v_{xc}(r)$ at the same points. Technically, this exchange and correlation functional belongs to the family of local-density approximations. Tozer et al. trained the neural network on two different datasets, first on the data of a single water molecule and afterwards on several molecules (namely, Ne, HF, N_2, H_2O, and H_2). Using 3768 data points calculated with a regular molecular numerical integration scheme,[469] the method achieved an accuracy of 2–3% in the exchange and correlation energy of the water molecule. When applied in a self-consistent Kohn–Sham calculation, the potential lead to eigenvalues and optimized geometries congruent with the local density approximation. On the other hand, for the set of

several molecules, Tozer et al. obtained an error of 7.6% using 1279 points. The points were obtained in the same manner as before but were constrained to avoid successive points with similar densities. This potential generated geometries close to the local density approximation and good eigenvalues for molecules sufficiently represented in the training set. Meanwhile, and as expected, the neural network potential failed for molecules not sufficiently represented in the training like LiH and Li_2.

In 2012, Snyder et al. tackled the problem of noninteracting spinless fermions confined to a 1D box.[470] They employed KRR to construct a machine learning approximation for the kinetic energy functional of the density. This is the idea behind orbital-free DFT and an attempt to bypass the need to solve a Shrödinger-like equation. The kinetic energy and density pairs of up to four electrons were obtained using Numerov's method[471] for several external potentials. These potentials were created using a linear combination of three Gaussian dips with random depths, widths, and centers. Furthermore, 1000 densities were taken for the test set while M were taken for the training set. For $M = 200$ chemical accuracy was achieved, as no error surpassed 1 kcal/mol. To obtain the correct behavior of the functional derivative of this energy, which is necessary for the self-consistent DFT procedure, PCA was used. Self-consistent calculations with this functional led to a range of similar densities instead of a unique density and to higher errors in the energy than when using the exact density. Nevertheless, the functional reached chemical accuracy.

This methodology was later improved during the study of the bond breaking for a 1D model of a diatomic molecule, subjected to a soft Coulomb interaction.[472] The training data consisted of Kohn–Sham energies and densities calculated with the local-density approximation for 1D H_2, H_2, Li_2, Be_2, and LiH with different nuclear separations. Choosing up to 20 densities for each molecule for the training set produced smaller errors in the kinetic energy functional than those due to the approximation to the exchange-correlation functional. This new functional was able to produce binding energy curves indistinguishable from the local-density approximation.

A different path was taken by Brockherde et al.[473] that, instead of solving the Kohn–Sham equations self-consistently as usually, used KRR to learn the Hohenberg–Kohn map between the potential $v(r)$ and the density $n(r)$. Among the machine learning community, this approach is normally designated as transductive inference. The energy is obtained from the density, also using KRR. When applied to the problem of noninteracting spinless fermions confined to a 1D box (same problem as in ref. [470]), this machine learning map reproduced the correct energy up to 0.042 kcal/mol (if calculated in a grid) or 0.017 kcal/mol (using other basis sets), for a training set of 200 samples. Comparison of this map with other machine learning maps that learn only the kinetic energy reveals that the Hohenberg–Kohn map approach is much more accurate. Furthermore, this map achieved similar results when applied to molecules, reaching accuracies of 0.0091 kcal/mol for water and 0.5 kcal/mol for benzene, ethane, and malinaldehyde. These values measure the difference to the PBE energy. The training sets consisted of 20 points for the water and 2000 points for the other molecules. To generate the training sets for the larger molecules, molecular dynamics simulations using the general amber force-field[474] were used to yield a large set of geometries. These were subsequently sampled using the k-means approach to obtain 2000 representative structures that were only then evaluated using the PBE functional. In addition, the precision of the density prediction for benzene was compared with the results for the local-density approximation and PBE. Not only did the Hohenberg–Kohn map produce densities with errors smaller than the difference between different functionals (when evaluated on a grid) but these errors were also smaller than the ones introduced by evaluating the PBE functional using a Fourier basis representation instead of the evaluation on the grid.

A distinct approach comes from Liu et al.,[475] who applied a neural network to determine the value of the range-separation parameter μ of the long-range corrected Becke–Lee–Yang–Parr functional.[476,477] They trained a neural network, characterized by one hidden layer, with 368 thermochemical and kinetic energies. These values came from experimental data and from highly accurate quantum chemistry calculations. When compared with the original functional ($\mu = 0.47$), the new functional improved the accuracy of heats of formation and atomization energies while performing slightly worse in the calculation for ionization potentials, reaction barriers, and electronic affinities.

Nagai et al.[478] trained a neural network with 2 hidden layers (300 nodes) to produce the projection from the charge density onto the Hartree-exchange-correlation potential (v_{Hxc}). For that, they solved a simple model of two interacting spinless fermions under the effect of a 1D Gaussian potential, using exact diagonalization. The ground state density was then used to calculate v_{Hxc} using an inverse Kohn–Sham method based on the Haydock–Foulkes variational principle.[479,480] When applied in the Kohn–Sham self-consistent cycle, this potential reproduced the exact densities and total energies, provided that a suitable training set was chosen (see Fig. 20). The system studied by the authors admits as solution either a bound and an unbound state or two bound states, depending on the Gaussian potential. Choosing points surrounding the boundary for the training set of the neural network leads to the most accurate results, with errors around 10^{-3} a.u. everywhere except at the boundary (where they can

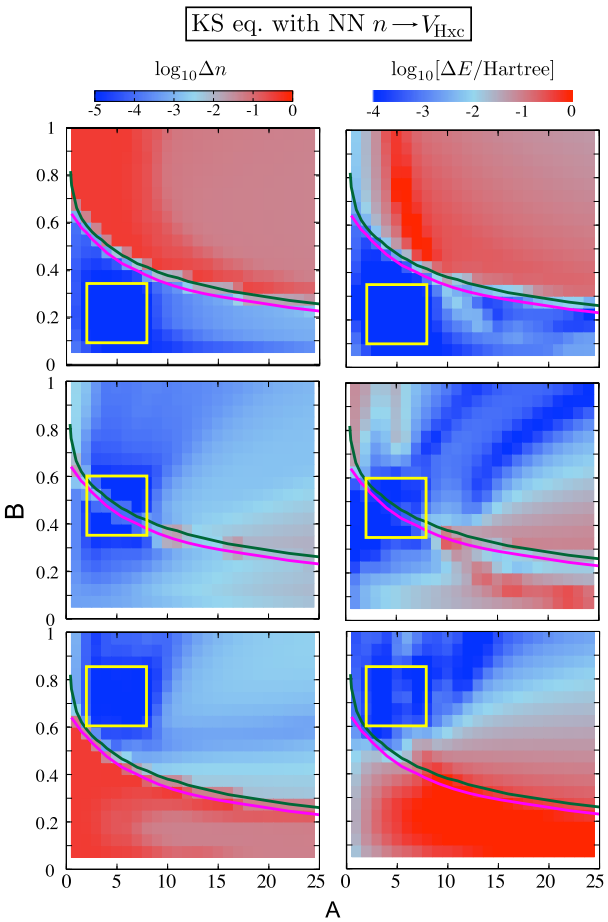

Fig. 20 Transferability of the neural network v_{Hxc}. The bold frames indicate the training set and the lines show the boundary between solutions with (green) and without (pink) the Coulomb interaction. The errors are plotted as color maps. (Reprinted from ref. [478] with the permission of AIP Publishing.)

almost reach 1 a.u.). On the other hand, choosing points in one of the regions results in a poor description of the other region.

DISCUSSION AND CONCLUSIONS

Interpretability

We already noted in the introduction that a major criticism of machine learning techniques is that their black-box algorithms do not provide us with new "physical laws" and that their inner workings remain outside our understanding.[481] For example, Ghiringhelli et al. argue that "a trustful prediction of new promising materials, identification of anomalies, and scientific advancement are doubtful," if the scientific connection between features and prediction is unknown.[96] Johnson writes in the context of quantitative structure–activity relationships: "By not following through with careful, designed, hypothesis testing we have allowed scientific thinking to be co-opted by statistics and arbitrarily defined fitness functions."[378] The main concern is that models not based on physical principles might fail in completely unexpected cases (that are trivial for humans) while providing a very good result on average. Such cases can only be predicted and prevented if one understands the causality between the inputs and outputs of the model. Furthermore, especially in applications where a single failure is extremely expensive or potentially deadly (as in medicine), the lack of trust in black-box machine learning models stops their widespread use even when they provide a superior performance.[273]

As there are different concepts of interpretability, we will define its various facets according to Lipton et al.[482] To start with, we can divide interpretability into transparency and post hoc explanations, which consist of additional information provided by or extracted from a model.

Transparency can once again be split into the concepts of simulatability, decomposability, and algorithmic transparency. Simulatability is a partially subjective notion and concerns the ability of humans to follow and retrace the calculations of the model. This is, e.g., the case for sparse linear models such as the ones resulting from LASSO,[159] SISSO,[163] or flat decision tree models. Decomposability is closely related to the intelligibility of a model and describes whether its various parts (input, parameters, calculations) allow for an intuitive interpretation. Algorithmic transparency considers our understanding of the error surface (e.g., whether the training will converge to a unique solution). This is clearly not the case for modern neural networks, for example.

Post hoc interpretability considers the possibility to extract additional information from the model. Examples for this are variable importance from a decision tree model or active response maps, which highlight regions of a picture that were particularly important for its classification by a convolutional neural network.

Starting from these concepts of interpretability, it is obvious that the notion of a complex model runs counter to the claim that it is simulatable by a human. Furthermore, models that are simulatable (e.g., low-dimensional linear models) and accurate often require unintuitive highly processed features that reduce the decomposability[483] (e.g., spectral neighbor analysis potential potentials) in order to reach a comparable performance to a more complex model. In contrast, a complex model like a deep convolutional neural network only requires relatively simple unengineered features and relies on its own ability to extract descriptors of different abstraction levels. In this sense, there is a definite conflict between the complexity and accuracy of a model, on one hand, and a simulatable decomposable model on the other hand.

The simplest examples of models that are simulatable are techniques based in dimensionality reduction or feature selection algorithms, like SISSO.[163] These are usually used in combination with linear fits and result in simple equations describing the problem. An example is the estimation of the probability of a material to exist as a perovskite (ABX_3), as given in ref. [143]:

$$\tau = \frac{r_X}{r_B} - n_A\left(n_A - \frac{r_A/r_B}{\ln(r_A/r_B)}\right), \tag{42}$$

where n_A is the oxidation state of A and r_i is the ionic radius of ion i. Another example is given by Kim et al.,[38] who used LASSO, as well as RF and KRR, to predict the dielectric breakdown field of elemental and binary insulators, on the basis of eight features obtained from first-principle calculations (e.g. band gap, phonon cutoff frequency, etc.). In the end, all three methods determined the same two features as optimal and demonstrated nearly the same error. However, Kim et al. favored LASSO,[159] because it provided a simple analytical formula, even if no further knowledge was gained from the formula. In any case, the knowledge of the analytical formula and therefore the simulatability seems to be far less relevant than the knowledge of the most relevant physical variables. In general, we can even argue that simulatability is not relevant for materials science as computational methods based on physical reasoning, like DFT or tight binding, are even further removed from simulatability than most machine learning models.

A second method that provides a variable importance measure (see section "Basic principles of machine learning—Features") are RFs or other decision tree-based methods. Stanev et al. demonstrate the usefulness of this method for post hoc interpretability in ref. [76], by recovering numerous known (e.g., isotope effect) and some unknown rules and limits for the superconducting critical temperature. This was done by first reducing the number of features via variable importance measure (Gini importance) and subsequently visualizing the correlation between the features and the critical temperature (see Fig. 21).

Pankajakshan et al.[168] developed bootstrapped-projected gradient descent as a feature selection method specifically for materials science. The motivation came from some consistency issues for correlated or linearly dependent variables (present, for example, in LASSO), which bootstrapped-projected gradient descent can alleviate through extra clustering and bootstrapping. In their work, Pankajakshan et al. used machine learning mostly to find and understand descriptors, in order to improve the *d*-band model of catalysts for CO_2 reduction,[484] instead of actually using the machine learning model for predictions. This can definitely be a reasonable approach in cases where datasets are too small and incomplete for any successful extrapolation. Notwithstanding, in most cases it is questionable if a classical (in the sense of "non-machine learning") model should be used directly when a machine learning model is superior, as in the case of the *d*-band model.[485] Of course, it is a bonus when a classical model exists, as it can be used to check for consistency issues or as a crude estimation of property. However, in our opinion, pragmatic applications of advanced materials design should always use the best model.

While RFs and linear fits are considered more accessible from a interpretability point of view, deep neural networks are one of the prime examples for algorithms that are traditionally considered a black box. While their complex nature often results in superior performance in comparison to simpler algorithms, an unwanted consequence is the lack of simulatability and algorithmic transparency. As the lack of interpretability is one of the main challenges for a wider adoption of neural networks in industry and experimental sciences, post hoc methods to visualize the response and understand the inner workings of neural networks were developed during the past years. One example are attentive response maps for image recognition networks that highlight regions of the picture according to their importance in the decision making process. Kumar et al.[271] demonstrated that, by combining the understanding gained from attentive response maps with domain knowledge and applying it to the design

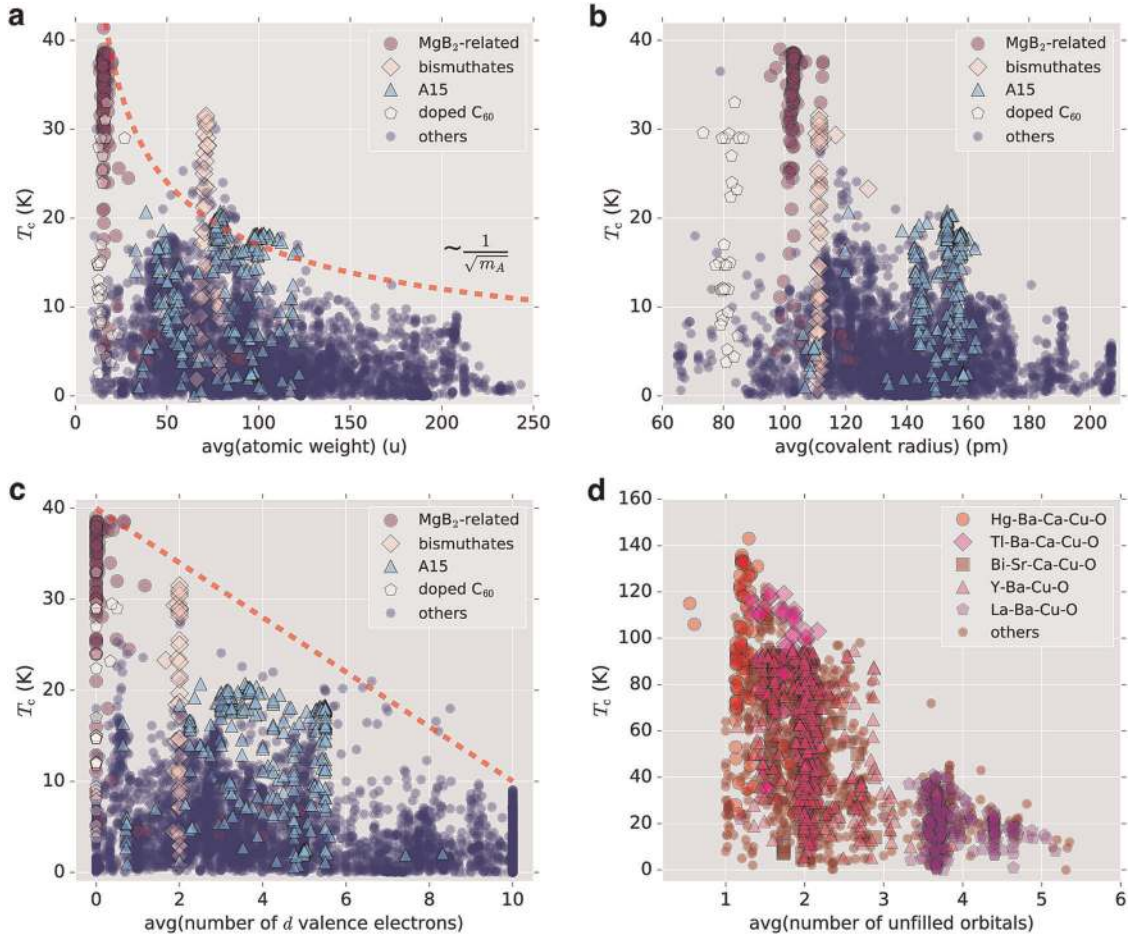

Fig. 21 Superconducting critical temperature T_C plotted versus the various features; **a** demonstrates the isotope effect and **b**–**d** show how the critical temperature is limited and influenced by various physical quantities of the materials. (Reprinted with permission from ref. [76] licensed under the CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).)

process of the neural network, one can not only achieve a better informed decision making process but also higher performance. An improvement of the performance through integration of domain knowledge is not completely surprising, but the result is nevertheless remarkable, as usually higher interpretability comes at the cost of a lower performance.

Zilleti et al.[268] introduced attentive response maps, as implemented in ref. [271], to materials science in order to visualize the ability of their convolutional neural networks to recognize crystal structures from diffraction patterns. The response maps of the different convolutional layers demonstrate that the neural networks recover the position of the diffraction peaks and their orientation as features (see Fig. 22).

A second example, which demonstrates the ability of neural networks to convey additional post hoc information, is described in ref. [40]. Xie et al. used a crystal graph convolutional neural network to learn the distance to the convex hull of perovskites ABX_3. By using the output of the pooling layers instead of the fully connected layers as a predictor, the energy can be split into contributions from the different crystal sites (see Fig. 23). This allowed Xie et al. to not only confirm the importance of the radii of the A- and B-atoms but also to gain new insights that were then used for an efficient combinatorial search of perovskites. In ref. [486], Xie et al. follow up with the interpretation of the features extracted from the convolutional neural networks and demonstrated how similarity patterns emerge for different material groups and at different scales.

Zhang et al.[321] also highlighted the ability of convolutional neural networks to extract physically meaningful features out of un-engineered descriptors. They built a convolutional neural network (two convolutional layers, one fully connected layer) to calculate the topological winding number of 1D band insulators with chiral symmetry based on their Hamiltonian as input data

$$\begin{bmatrix} h_x(0) & h_x\left(\frac{2\pi}{L}\right) & \cdots \\ h_y(0) & h_y\left(\frac{2\pi}{L}\right) & \cdots \end{bmatrix}^T = \begin{bmatrix} \cos(\Phi) & \cos(\Phi + \Delta\Phi) & \cdots \\ \sin(\Phi) & \sin(\Phi + \Delta\Phi) & \cdots \end{bmatrix}^T. \quad (43)$$

From the theoretical equation for the winding number,[487] one can derive that the second convolutional layer should produce an output linearly depending on $\Delta\Phi$ with the exception of a jump at $\Delta\Phi = \pi$. We can see in Fig. 24 that this is exactly the case, and consequently, the convolutional neural network actually learned the discrete formula for the winding number. Sun et al.[323] studied similar models of higher complexity with deep convolutional neural networks and were also able to demonstrate that their networks learned the known mathematical formulas for the winding and the Chern numbers.[488]

Naturally, neural networks will never reach the algorithmic transparency of linear models. However, representative datasets, a good knowledge of the training process, and a comprehensive validation of the model can usually overcome this obstacle. Furthermore, if we consider the possibilities for post hoc explanations or the decomposability of neural networks, they are actually far more interpretable than their reputation might suggest.

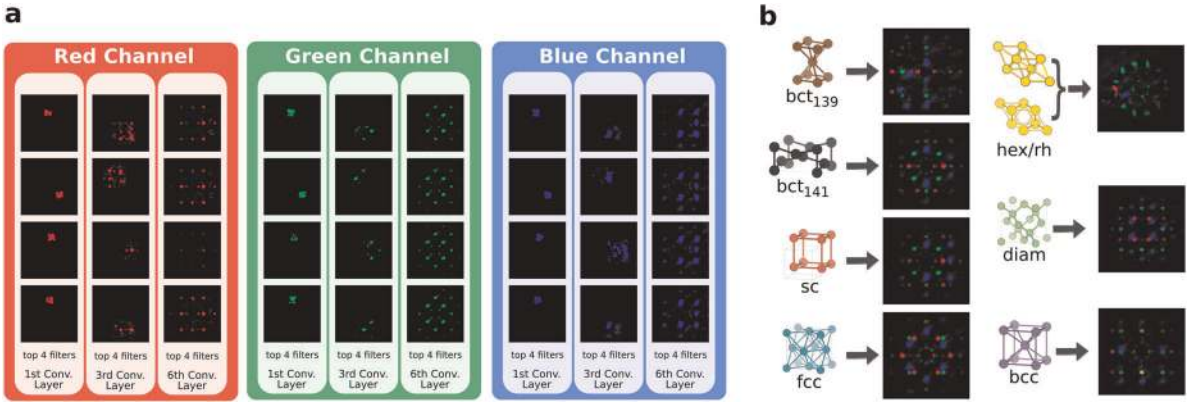

Fig. 22 **a** Attentive response maps for the four most activated filters of the first, third, and last convolutional layers for simple cubic lattices. The brightness of the pixel represents the importance of the location for classification. **b** Sum of the last convolutional layer filters for all seven crystal classes showing that the network learned crystal templates automatically from the data. (Reprinted with permission from ref. [268] licensed under the CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).)

To conclude this chapter, we would like to summarize a few points: (i) Interpretability is not a single algorithmic property but a multifaceted concept (simulatability, decomposability, algorithmic transparency, post hoc knowledge extraction) (ii) The various facets have different priorities depending on the dataset and the research goal. (iii) Simulatability is usually non-existent in materials science (e.g., in DFT or Monte Carlo simulations) regardless of whether one uses a machine learning or a classical algorithm. Therefore, it should probably not be a point of concern in materials informatics. (iv) Part of the progress of materials informatics has to include the increasing use of post hoc knowledge techniques, like attentive response maps, to improve the viability of, and the trust in, high-performing black-box models. Often this knowledge alleviates the fear that the model is operating on unphysical principles.[268,321,323]

Conclusions

Just like the industrial revolution, which consisted of the creation of machines that could perform mechanical tasks more efficiently than humans, in the field of machine learning machines are progressively trained to identify patterns and to find relations between properties and features more efficiently than us. In materials science, machine learning is mostly applied to classification and regression problems. In this context, we discussed a wide variety of quantitative structure–property relationships, which encompass a high number of properties essential for modern technology. It seems likely that further properties, should they be needed, can also be predicted with a similar level of accuracy.

If we consider the direction of future research, there will be a clear division between methodologies depending on the availability of data. For continuous properties, which can be calculated realistically for $\geq 10^5$ materials, we assume that universal models and especially deep neural networks, like Xie et al.'s crystal graph convolutional networks[40] or Chen et al.'s *MatErials Graph Networks*,[132] will be the future. They are able to predict a diverse set of properties, such as formation energies, band gaps, Fermi energies, bulk moduli, shear moduli, and Poisson ratios for a wide material space (87 elements, 7 lattice systems, and 216 space groups in the case of ref. [40]). At the same time, they reach an accuracy with respect to DFT calculations that is comparable with (or even smaller than) the DFT errors with respect to experiment. Such models have the potential to end the need for applications trained for only a single structural prototype and/or property, which can in turn drastically reduce the amount of resources spent by single researchers. Comparing to the state of the art of neural network architectures and training methods in fields like image recognition and natural language procession, we can also expect that the success of neural network models will only increase once modern topologies, training methods, and fast implementations reach a wider audience in materials science. To reach this goal, a closer interdisciplinary collaboration with computer scientists will be essential.

In other cases that are characterized by a lack of data, several strategies are very promising. First of all, one can take into consideration surrogate-based optimization (active learning), which allows researchers to optimize the results achieved with a limited experimental or computational budget. Surrogate-based optimization allows us to somewhat overlook the limited accuracy of the machine learning models while nevertheless arriving at sufficient design results. As the use of such optimal design algorithms is still confined to relatively few studies with small datasets, much future work can be foreseen in this direction. A second strategy to overcome the limited data available in materials science is transfer learning. While it has already been applied with success in chemistry,[489] wider applications in solid-state materials informatics are still missing. A last strategy to handle the small datasets that are so common in materials science was discussed by Zhang et al. in ref. [77]. Crude estimation of properties basically allows us to shift the problem of predicting a property to the problem of predicting the error of the crude model with respect to the higher-fidelity training data. Up to now, this strategy was mostly used for the prediction of band gap, as datasets of different fidelity are openly available (DFT, *GW*, or experimental). Moreover the use of crude estimators allows researchers to benefit from decades of work and expertise that went into classical (non-machine learning) models. If the lower-fidelity data are not available for all materials, it is also possible to use a co-kriging approach that still profits from the crude estimators but does not require it for every prediction.[292]

Component prediction is a highly effective way to speed up the material discovery process and we expect high-throughput searches of all common crystal structure prototypes that were not yet researched in the coming years. While the prediction of the energy can also be considered, a quantitative structure–property relationships, metastable materials, and an incomplete knowledge of the theoretical convex hull have to be taken into account. Several studies demonstrated that better accuracy can be achieved with experimental training data. However, as experimental data are seldom available and expensive to generate, the number of prototypes for which studies analog to ref. [143] are an option will quickly be exhausted. A second challenge is the lack of published data of failed experiments. In this case, a cultural shift toward the publication of all valid data, may it be positive or negative, is required.

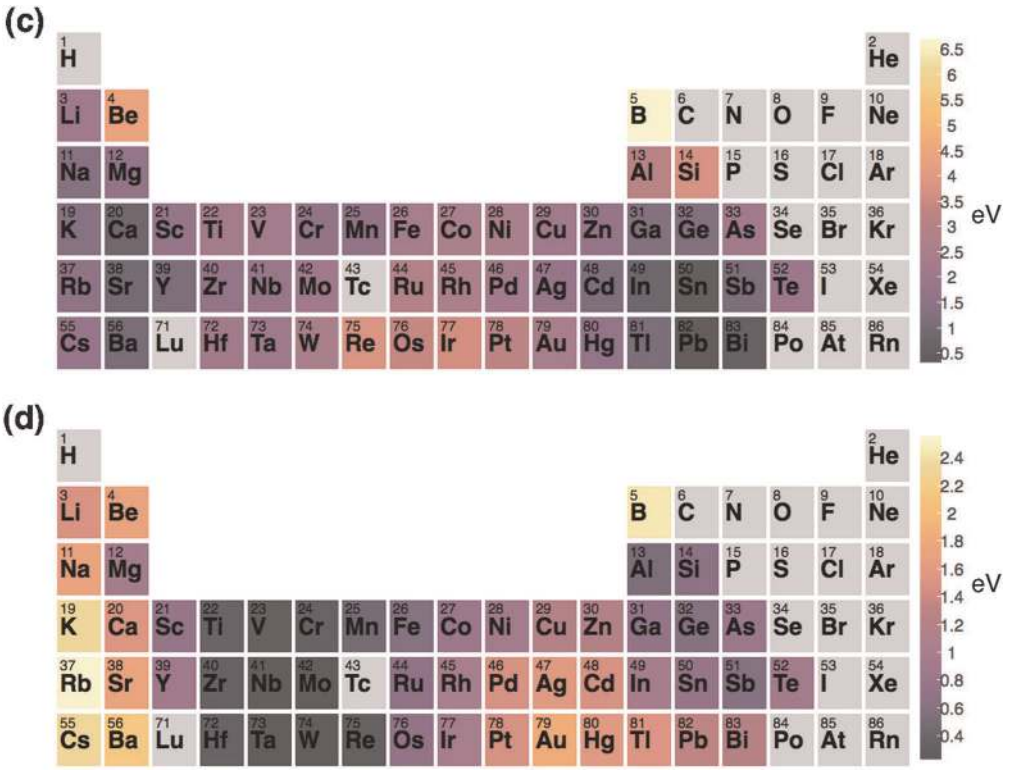

Fig. 23 Contributions to the distance to the convex hull per element, A site (**c**) and B site (**d**). (Reprinted with permission from ref. [40]. Copyright 2018 American Physical Society.)

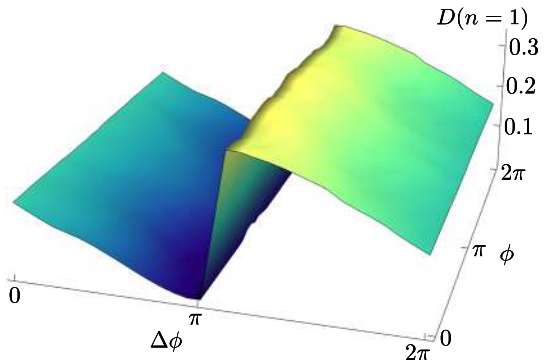

Fig. 24 Output of the second layer as a function of $\delta\Phi$ and Φ. (Reprinted with permission from ref. [321]. Copyright 2018 American Physical Society.)

The direct prediction or generation of a crystal structure is still an extremely challenging problem. While several studies demonstrate how to differentiate between a small number of prototypes for a certain composition, the difficulty quickly rises with an increasing number of possible crystal structures. This is amplified by the fact that the majority of available data belongs to only a small number of extensively researched prototypes. Recently, more complex modern neural network structures (e.g., VAEs, GANs, etc.) were introduced to the problem, with some interesting results. Moreover, the use of machine learning-based optimization algorithms, like Bayesian optimization for global structure prediction, is also a direction that should be further explored.

Machine learning was successfully integrated with other numerical techniques, such as molecular dynamics and global structural prediction. Force fields built with neural networks enjoy an efficiency that parallels that of classical force fields and an accuracy comparable to the reference method (usually DFT in solid state, although in chemistry some force fields already achieved coupled cluster accuracy[489]). Consequently, we expect them to completely replace classical force fields in the long term. Owing to their vastly superior numerical scaling, machine learning methods allow us to tackle challenging problems, which go far beyond the limitations of current electronic structure methods, and to investigate novel, emerging phenomena that stem from the complexity of the systems.

The majority of early machine learning applications to solid-state materials science employed straightforward and simple-to-use algorithms, like linear kernel models and decision trees. Now, that these proofs-of-concept exist for a variety of application, we expect that research will follow two different directions. The first will be the continuation of the present research, the development of more sophisticated machine learning methods, and their applications in materials science. Here one of the major problems is the lack of benchmarking datasets and standards. In chemistry, a number of such datasets already exists, such as the QM7 dataset,[490,491] QM8 dataset,[491,492] QM7b dataset,[493,494] etc. These are absolutely essential to measure the progress in features and algorithms. While we discussed countless machine learning studies in this review, definitive quantitative comparisons between the different works were mostly impossible, impeding the evaluation of progress and thereby progress itself. It has to be noted that there has been one recent competition for the prediction of formation energies and band gaps.[495] In our opinion, this is an very important step in the right direction. Unfortunately, the dataset used in this competition was extremely small and specific, putting the generalizability of the results to larger and more diverse datasets into doubt.

The second direction regards the usability of machine learning models. In the electronic structure community, both the models (e.g., new approximations to the exchange-correlation functional of DFT) and the computer codes are developed by a relatively

small group of experts and put at the disposal of the much larger community of materials scientists. Even though this is slowly starting to change, models from most publications are not publicly available. This results in most researchers spending resources on building their own models to solve very specific problems. We note that frameworks to disseminate models are now starting to emerge.[496]

In conclusion, we reviewed the latest applications of machine learning in the field of materials science. These applications have been mushrooming in the past couple of years, fueled by the unparalleled success that machine learning algorithms have found in several different fields of science and technology. It is our firm conviction that this collection of efficient statistical tools are indeed capable of speeding up considerably both fundamental and applied research. As such, they are clearly more than a temporary fashion and will certainly shape materials science for the years to come.

AUTHOR CONTRIBUTIONS

The authors contributed equally to the search and analysis of the literature and to the discussion and writing of the manuscript.

ADDITIONAL INFORMATION

Competing interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Marsland, S. *Machine Learning* (CRC Press, Taylor & Francis Inc., Boca Raton, FL, 2014).
2. Silver, D. et al. Mastering the game of go with deep neural networks and tree search. *Nature* **529**, 484–489 (2016).
3. Bojarski, M. et al. End to end learning for self-driving cars. Preprint at arXiv:1604.07316 (2016).
4. He, K., Zhang, X., Ren, S. & Sun, J. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In *2015 IEEE International Conference on Computer Vision (ICCV)* (eds Bajcsy, R. & Hager, G.) 1026–1034 (IEEE, Piscataway, NJ, 2015).
5. Liu, S.-S. & Tian, Y.-T. Facial expression recognition method based on gabor wavelet features and fractional power polynomial kernel PCA. In *Advances in Neural Networks - ISNN 2010* (eds Zhang, L., Lu, B.-L. & Kwok, J.) 144–151 (Springer, Berlin, Heidelberg, 2010).
6. Waibel, A. & Lee, K.-F. (eds) *Readings in Speech Recognition* (Morgan Kaufmann, Burlington, MA, 1990).
7. Pazzani, M. & Billsus, D. Learning and revising user profiles: the identification of interesting web sites. *Mach. Learn.* **27**, 313–331 (1997).
8. Chan, P. K. & Stolfo, S. J. Toward scalable learning with non-uniform class and cost distributions: a case study in credit card fraud detection. In *KDD'98 Proc. Fourth International Conference on Knowledge Discovery and Data Mining* (eds Agrawal, R., Stolorz, P. & Piatetsky, G.) 164–168 (AAAI Press, New York, NY, 1998).
9. Guzella, T. S. & Caminhas, W. M. A review of machine learning approaches to spam filtering. *Expert Syst. Appl.* **36**, 10206–10222 (2009).
10. Huang, C.-L., Chen, M.-C. & Wang, C.-J. Credit scoring with a data mining approach based on support vector machines. *Expert Syst. Appl.* **33**, 847–856 (2007).
11. Baldi, P. & Brunak, S. *Bioinformatics: The Machine Learning Approach* (The MIT Press, Cambridge, MA, 2001).
12. Noordik, J. H. *Cheminformatics Developments: History, Reviews and Current Research* (IOS Press, Amsterdam, 2004).
13. Rajan, K. Materials informatics. *Mater. Today* **8**, 38–45 (2005).
14. Martin, R. M. *Electronic Structure: Basic Theory and Practical Methods* (Cambridge University Press, Cambridge, 2008).
15. Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **136**, B864–B871 (1964).
16. Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, A1133–A1138 (1965).
17. Olson, G. B. Designing a new material world. *Science* **288**, 993–998 (2000).
18. Oganov, A. R. (ed.) *Modern Methods of Crystal Structure Prediction* (Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2010).
19. Oganov, A. R. & Glass, C. W. Crystal structure prediction using ab initio evolutionary techniques: principles and applications. *J. Chem. Phys.* **124**, 244704 (2006).
20. Newnham, R. E. *Properties of materials: anisotropy, symmetry, structure* (Oxford University Press, Oxford, 2005).
21. Curtarolo, S. et al. The high-throughput highway to computational materials design. *Nat. Mater.* **12**, 191–201 (2013).
22. Green, M. L. et al. Fulfilling the promise of the materials genome initiative with high-throughput experimental methodologies. *Appl. Phys. Rev.* **4**, 011105 (2017).
23. Koinuma, H. & Takeuchi, I. Combinatorial solid-state chemistry of inorganic materials. *Nat. Mater.* **3**, 429–438 (2004).
24. Suram, S. K., Haber, J. A., Jin, J. & Gregoire, J. M. Generating information-rich high-throughput experimental materials genomes using functional clustering via multitree genetic programming and information theory. *ACS Comb. Sci.* **17**, 224–233 (2015).
25. Potyrailo, R. et al. Combinatorial and high-throughput screening of materials libraries: review of state of the art. *ACS Comb. Sci.* **13**, 579–633 (2011).
26. Walsh, A. The quest for new functionality. *Nat. Chem.* **7**, 274–275 (2015).
27. Lookman, T., Eidenbenz, S., Alexander, F. & Barnes, C. (eds) *Materials Discovery and Design by Means of Data Science and Optimal Learning* (Springer International Publishing, Basel, 2018).
28. Ryan, K., Lengyel, J. & Shatruk, M. Crystal structure prediction via deep learning. *J. Am. Chem. Soc.* **140**, 10158–10168 (2018).
29. Nouira, A., Sokolovska, N. & Crivello, J.-C. Crystalgan: learning to discover crystallographic structures with generative adversarial networks. Preprint at arXiv:1810.11203 (2018).
30. Graser, J., Kauwe, S. K. & Sparks, T. D. Machine learning and energy minimization approaches for crystal structure predictions: a review and new horizons. *Chem. Mater.* **30**, 3601–3612 (2018).
31. Balachandran, P. V., Kowalski, B., Sehirlioglu, A. & Lookman, T. Experimental search for high-temperature ferroelectric perovskites guided by two-step machine learning. *Nat. Commun.* **9**, 1668 (2018).
32. Oliynyk, A. O., Adutwum, L. A., Harynuk, J. J. & Mar, A. Classifying crystal structures of binary compounds AB through cluster resolution feature selection and support vector machine analysis. *Chem. Mater.* **28**, 6672–6681 (2016).
33. Li, W., Jacobs, R. & Morgan, D. Predicting the thermodynamic stability of perovskite oxides using machine learning models. *Comput. Mater. Sci.* **150**, 454–463 (2018).
34. Ward, L. et al. Including crystal structure attributes in machine learning models of formation energies via Voronoi tessellations. *Phys. Rev. B* **96**, 024104 (2017).
35. Faber, F. A., Lindmaa, A., von Lilienfeld, O. A. & Armiento, R. Machine learning energies of 2 million elpasolite (ABC_2D_6) crystals. *Phys. Rev. Lett.* **117**, 135502 (2016).
36. Zheng, X., Zheng, P. & Zhang, R.-Z. Machine learning material properties from the periodic table using convolutional neural networks. *Chem. Sci.* **9**, 8426–8432 (2018).
37. Carrete, J., Li, W., Mingo, N., Wang, S. & Curtarolo, S. Finding unprecedentedly low-thermal-conductivity half-Heusler semiconductors via high-throughput materials modeling. *Phys. Rev. X* **4**, 011019 (2014).
38. Kim, C., Pilania, G. & Ramprasad, R. From organized high-throughput data to phenomenological theory using machine learning: the example of dielectric breakdown. *Chem. Mater.* **28**, 1304–1311 (2016).
39. Seko, A., Maekawa, T., Tsuda, K. & Tanaka, I. Machine learning with systematic density-functional theory calculations: application to melting temperatures of single- and binary-component solids. *Phys. Rev. B* **89**, 054303 (2014).
40. Xie, T. & Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **120**, 145301 (2018).
41. Isayev, O. et al. Universal fragment descriptors for predicting properties of inorganic crystals. *Nat. Commun.* **8**, 15679 (2017).
42. Furmanchuk, A., Agrawal, A. & Choudhary, A. Predictive analytics for crystalline materials: bulk modulus. *RSC Adv.* **6**, 95246–95251 (2016).
43. Kauwe, S. K., Graser, J., Vazquez, A. & Sparks, T. D. Machine learning prediction of heat capacity for solid inorganics. *Integr. Mater. Manuf. Innov.* **7**, 43–51 (2018).
44. Kim, C., Pilania, G. & Ramprasad, R. Machine learning assisted predictions of intrinsic dielectric breakdown strength of ABX_3 perovskites. *J. Phys. Chem. C* **120**, 14575–14580 (2016).
45. Yuan, F. & Mueller, T. Identifying models of dielectric breakdown strength from high-throughput data via genetic programming. *Sci. Rep.* **7**, 17594 (2017).
46. Gaultois, M. W. et al. Perspective: Web-based machine learning models for real-time screening of thermoelectric materials properties. *APL Mater.* **4**, 053213 (2016).

47. Ju, S. et al. Designing nanostructures for phonon transport via Bayesian optimization. *Phys. Rev. X* **7**, 021024 (2017).
48. Seko, A., Hayashi, H., Nakayama, K., Takahashi, A. & Tanaka, I. Representation of compounds for machine-learning prediction of physical properties. *Phys. Rev. B* **95**, 144110 (2017).
49. Sosso, G. C., Deringer, V. L., Elliott, S. R. & Csányi, G. Understanding the thermal properties of amorphous solids using machine-learning-based interatomic potentials. *Mol. Simul.* **44**, 866–880 (2018).
50. Wei, H., Zhao, S., Rong, Q. & Bao, H. Predicting the effective thermal conductivities of composite materials and porous media by machine learning methods. *Int. J. Heat. Mass Tran.* **127**, 908–916 (2018).
51. Wu, Y.-J., Sasaki, M., Goto, M., Fang, L. & Xu, Y. Electrically conductive thermally insulating Bi–Si nanocomposites by interface design for thermal management. *ACS Appl. Nano Mater.* **1**, 3355–3363 (2018).
52. Jalem, R. et al. Bayesian-driven first-principles calculations for accelerating exploration of fast ion conductors for rechargeable battery application. *Sci. Rep.* **8**, 5845 (2018).
53. Rosenblatt, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **65**, 386 (1958).
54. McCulloch, W. S. & Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **5**, 115–133 (1943).
55. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
56. Ye, W., Chen, C., Wang, Z., Chu, I.-H. & Ong, S. P. Deep neural networks for accurate predictions of crystal stability. *Nat. Commun.* **9**, 3800 (2018).
57. Ren, Z. & Lee, Y. J. Cross-domain self-supervised multi-task feature learning using synthetic imagery. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (eds Bai, Y., Zhang, Y., Ding, M. & Ghanem, B.) 762–771 (IEEE, Piscataway, NJ, 2018).
58. Rajan, K. Materials informatics: the materials gene and big data. *Annu. Rev. Mater. Res.* **45**, 153–169 (2015).
59. Mueller, T., Kusne, A. G. & Ramprasad, R. in *Reviews in Computational Chemistry* (eds Parrill, A. L. & Lipkowitz, K. B.) Ch. 4 (John Wiley & Sons, Inc., Hoboken, NJ, 2016).
60. Correa-Baena, J.-P. et al. Accelerating materials development via automation, machine learning, and high-performance computing. *Joule* **2**, 1410–1420 (2018).
61. Liu, Y., Zhao, T., Ju, W. & Shi, S. Materials discovery and design using machine learning. *J. Mater.* **3**, 159–177 (2017).
62. Ward, L. et al. Strategies for accelerating the adoption of materials informatics. *MRS Bull.* **43**, 683–689 (2018).
63. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559**, 547–555 (2018).
64. Butler, K. T., Frost, J. M., Skelton, J. M., Svane, K. L. & Walsh, A. Computational materials design of crystalline solids. *Chem. Soc. Rev.* **45**, 6138–6146 (2016).
65. Shi, S. et al. Multi-scale computation methods: Their applications in lithium-ion battery research and development. *Chin. Phys. B* **25**, 018212 (2016).
66. Ward, L. & Wolverton, C. Atomistic calculations and materials informatics: a review. *Curr. Opin. Solid State Mater. Sci.* **21**, 167–176 (2017).
67. Alpaydin, E. *Introduction to Machine Learning* (The MIT Press, Cambridge, MA, 2014).
68. Sutton, R. S. & Barto, A. G. *Reinforcement Learning* (The MIT Press, Cambridge, MA, 2018).
69. Nguyen, H., Maeda, S.-i. & Oono, K. Semi-supervised learning of hierarchical representations of molecules using neural message passing. Preprint at arXiv:1711.10168 (2017).
70. Geman, S., Bienenstock, E. & Doursat, R. Neural networks and the bias/variance dilemma. *Neural Comput.* **4**, 1–58 (1992).
71. Sammut, C. & Webb, G. I. *Encyclopedia of Machine Learning and Data Mining* (Springer Publishing Company, Incorporated, New York, NY, 2017).
72. Picard, R. R. & Cook, R. D. Cross-validation of regression models. *J. Am. Stat. Assoc.* **79**, 575–583 (1984).
73. Meredig, B. et al. Can machine learning identify the next high-temperature superconductor? Examining extrapolation performance for materials discovery. *Mol. Syst. Des. Eng.* **3**, 819–825 (2018).
74. Tropsha, A., Gramatica, P. & Gombar, V. K. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb. Sci.* **22**, 69–77 (2003).
75. Golbraikh, A. & Tropsha, A. Beware of q2. *J. Mol. Graph. Modell.* **20**, 269–276 (2002).
76. Stanev, V. et al. Machine learning modeling of superconducting critical temperature. *npj Comput. Mater.* **4**, 29 (2018).
77. Zhang, Y. & Ling, C. A strategy to apply machine learning to small datasets in materials science. *npj Comput. Mater.* **4**, 25 (2018).
78. Jain, A. et al. Commentary: The materials project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
79. F. H. Allen, G. G. & Sievers, R. (eds) *Crystallographic Databases* (International Union of Crystallography, Chester, 1987).
80. Saal, J. E., Kirklin, S., Aykol, M., Meredig, B. & Wolverton, C. Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD). *JOM* **65**, 1501–1509 (2013).
81. Kirklin, S. et al. The open quantum materials database (OQMD): assessing the accuracy of DFT formation energies. *npj Comput. Mater.* **1**, 15010 (2015).
82. Groom, C. R., Bruno, I. J., Lightfoot, M. P. & Ward, S. C. The Cambridge structural database. *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.* **72**, 171–179 (2016).
83. Hachmann, J. et al. The Harvard clean energy project: large-scale computational screening and design of organic photovoltaics on the world community grid. *J. Phys. Chem. Lett.* **2**, 2241–2251 (2011).
84. Puchala, B. et al. The materials commons: a collaboration platform and information repository for the global materials community. *JOM* **68**, 2035–2044 (2016).
85. Mullin, R. Citrine informatics. *CEN Glob. Enterp.* **95**, 34–34 (2017).
86. de Jong, M. et al. Charting the complete elastic properties of inorganic crystalline compounds. *Sci. Data* **2**, 150009 (2015).
87. Zakutayev, A. et al. An open experimental database for exploring inorganic materials. *Sci. Data* **5**, 180053 (2018).
88. Villars, P., Okamoto, H. & Cenzual, K. *ASM Alloy Phase Diagrams Database* (ASM International, Materials Park, OH, 2006).
89. Gražulis, S. et al. Crystallography open database (COD): an open-access collection of crystal structures and platform for world-wide collaboration. *Nucleic Acids Res.* **40**, D420–D427 (2011).
90. Villars, P. et al. The Pauling file, binaries edition. *J. Alloy. Comp.* **367**, 293–297 (2004).
91. Gorai, P. et al. TE design lab: a virtual laboratory for thermoelectric material design. *Comput. Mater. Sci.* **112**, 368–376 (2016).
92. Haastrup, S. et al. The Computational 2D Materials Database: high-throughput modeling and discovery of atomically thin crystals. *2D Mater.* **5**, 042002 (2018).
93. Wilkinson, M. D. et al. The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
94. Draxl, C. & Scheffler, M. NOMAD: the FAIR concept for big data-driven materials science. *MRS Bull.* **43**, 676–682 (2018).
95. Raccuglia, P. et al. Machine-learning-assisted materials discovery using failed experiments. *Nature* **533**, 73–76 (2016).
96. Ghiringhelli, L. M., Vybiral, J., Levchenko, S. V., Draxl, C. & Scheffler, M. Big data of materials science: critical role of the descriptor. *Phys. Rev. Lett.* **114**, 105503 (2015).
97. Bengio, Y., Courville, A. & Vincent, P. Representation learning: a review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 1798–1828 (2013).
98. Bellman, R. E. *Adaptive Control Processes: A Guided Tour* (Princeton University Press, Princeton, NJ, 2015).
99. Schmidt, J. et al. Predicting the thermodynamic stability of solids combining density functional theory and machine learning. *Chem. Mater.* **29**, 5090–5103 (2017).
100. Bartók, A. P., Kondor, R. & Csányi, G. On representing chemical environments. *Phys. Rev. B* **87**, 184115 (2013).
101. Raff, L., Komanduri, R. & Hagan, M. *Neural Networks in Chemical Reaction Dynamics* (Oxford University Press, Oxford, 2012).
102. Braams, B. J. & Bowman, J. M. Permutationally invariant potential energy surfaces in high dimensionality. *Int. Rev. Phys. Chem.* **28**, 577–606 (2009).
103. Swamidass, S. J. et al. Kernels for small molecules and the prediction of mutagenicity, toxicity and anti-cancer activity. *Bioinformatics* **21**, i359–i368 (2005).
104. Weyl, H. *The Classical Groups: Their Invariants and Representations* (Princeton University Press, Princeton, NJ, 1997).
105. Jensen, F. *Introduction to Computational Chemistry* (Wiley, New York, NY, 2013).
106. Rupp, M., Tkatchenko, A., Müller, K.-R. & von Lilienfeld, O. A. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.* **108**, 058301 (2012).
107. Moussa, J. E. Comment on «fast and accurate modeling of molecular atomization energies with machine learning». *Phys. Rev. Lett.* **109**, 059801 (2012).
108. Faber, F., Lindmaa, A., von Lilienfeld, O. A. & Armiento, R. Crystal structure representations for machine learning models of formation energies. *Int. J. Quantum Chem.* **115**, 1094–1101 (2015).
109. SchÃ¼tt, K. T. et al. How to represent crystal structures for machine learning: towards fast prediction of electronic properties. *Phys. Rev. B* **89**, 205118 (2014).
110. Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
111. Artrith, N. & Urban, A. An implementation of artificial neural-network potentials for atomistic materials simulations: performance for TiO_2. *Comput. Mater. Sci.* **114**, 135–150 (2016).

112. Behler, J. Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.* **145**, 170901 (2016).
113. Seko, A., Takahashi, A. & Tanaka, I. Sparse representation for a potential energy surface. *Phys. Rev. B* **90**, 024101 (2014).
114. Bartók, A. P., Payne, M. C., Kondor, R. & Csányi, G. Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* **104**, 136403 (2010).
115. Khersonskii, V. K., Moskalev, A. N. & Varshalovich, D. A. *Quantum Theory of Angular Momemtum* (World Scientific Publishing, Singapore, 1988).
116. Meremianin, A. V. Multipole expansions in four-dimensional hyperspherical harmonics. *J. Phys. A Math. Gen.* **39**, 3099–3112 (2006).
117. Artrith, N., Urban, A. & Ceder, G. Efficient and accurate machine-learning interpolation of atomic energies in compositions with many species. *Phys. Rev. B* **96**, 014112 (2017).
118. Sanville, E., Bholoa, A., Smith, R. & Kenny, S. D. Silicon potentials investigated using density functional theory fitted neural networks. *J. Phys. Condens. Matter* **20**, 285219 (2008).
119. Baskes, M. Determination of modified embedded atom method parameters for nickel. *Mater. Chem. Phys.* **50**, 152–158 (1997).
120. Kuz'min, V. E. et al. Hierarchic system of QSAR models (1D–4D) on the base of simplex representation of molecular structure. *J. Mol. Model.* **11**, 457–467 (2005).
121. Kuz'min, V. E., Artemenko, A. G. & Muratov, E. N. Hierarchical QSAR technology based on the simplex representation of molecular structure. *J. Comput. Aid. Mol. Des.* **22**, 403–421 (2008).
122. Isayev, O. et al. Materials cartography: representing and mining materials space using structural and electronic fingerprints. *Chem. Mater.* **27**, 735–743 (2015).
123. Ruggiu, F., Marcou, G., Varnek, A. & Horvath, D. ISIDA property-labelled fragment descriptors. *Mol. Inform.* **29**, 855–868 (2010).
124. Blatov, V. A. Voronoi-Dirichlet polyhedra in crystal chemistry: theory and applications. *Crystallogr. Rev.* **10**, 249–318 (2004).
125. Carlucci, L., Ciani, G., Proserpio, D. M., Mitina, T. G. & Blatov, V. A. Entangled two-dimensional coordination networks: a general survey. *Chem. Rev.* **114**, 7557–7580 (2014).
126. Cordero, B. et al. Covalent radii revisited. *Dalton Trans*. 2832–2838 (2008).
127. Pham, T. L. et al. Machine learning reveals orbital interaction in materials. *Sci. Technol. Adv. Mat.* **18**, 756–765 (2017).
128. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (eds Bajcsy, R., Li, F.-F., & Tuytelaars, T.) 770–778 (IEEE, Piscataway, NJ, 2016).
129. Gori, M., Monfardini, G. & Scarselli, F. A new model for learning in graph domains. *Proc. 2005 IEEE Int. Jt. Conf. Neural Netw. 2005.* **2**, 729–734 (2005).
130. Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M. & Monfardini, G. The graph neural network model. *IEEE Trans. Neural Netw.* **20**, 61–80 (2009).
131. Li, Y., Tarlow, D., Brockschmidt, M. & Zemel, R. Gated graph sequence neural networks. Preprint at arXiv:1511.05493 (2015).
132. Chen, C., Ye, W., Zuo, Y., Zheng, C. & Ong, S. P. Graph networks as a universal machine learning framework for molecules and crystals. *Chem. Mater.* **31**, 3564–3572 (2019).
133. Schütt, K. T., Arbabzadah, F., Chmiela, S., Müller, K. R. & Tkatchenko, A. Quantum-chemical insights from deep tensor neural networks. *Nat. Commun.* **8**, 13890 (2017).
134. Kipf, T. N. & Welling, M. Semi-supervised classification with graph convolutional networks. Preprint at arXiv:1609.02907 (2016).
135. Kearnes, S., McCloskey, K., Berndl, M., Pande, V. & Riley, P. Molecular graph convolutions: moving beyond fingerprints. *J. Comput. Aided Mol. Des.* **30**, 595–608 (2016).
136. Bruna, J., Zaremba, W., Szlam, A. & LeCun, Y. Spectral networks and locally connected networks on graphs. Preprint at arXiv:1312.6203 (2013).
137. Battaglia, P. W., Pascanu, R., Lai, M., Rezende, D. & Kavukcuoglu, K. Interaction networks for learning about objects, relations and physics. Preprint at arXiv:1612.00222 (2016).
138. Defferrard, M., Bresson, X. & Vandergheynst, P. Convolutional neural networks on graphs with fast localized spectral filtering. Preprint at arXiv:1606.09375 (2016).
139. Duvenaud, D. K. et al. Convolutional networks on graphs for learning molecular fingerprints. In *Advances in Neural Information Processing Systems 28* (eds. Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M. & Garnett, R.) 2224–2232 (Curran Associates, Inc., Red Hook, NY, 2015).
140. Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A. & Müller, K.-R. Schnet – a deep learning architecture for molecules and materials. *J. Chem. Phys.* **148**, 241722 (2018).
141. Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O. & Dahl, G. E. Neural message passing for quantum chemistry. In *Proc. 34th International Conference on Machine Learning, vol. 70 of Proc. Machine Learning Research* (eds. Precup, D. & Teh, Y. W.) 1263–1272 (PMLR, International Convention Centre, Sydney, 2017).
142. Jørgensen, P. B., Jacobsen, K. W. & Schmidt, M. N. Neural message passing with edge updates for predicting properties of molecules and materials. Preprint at arXiv:1806.03146 (2018).
143. Bartel, C. J. et al. New tolerance factor to predict the stability of perovskite oxides and halides. Preprint at arXiv:1801.07700 (2018).
144. Mannodi-Kanakkithodi, A., Pilania, G., Huan, T. D., Lookman, T. & Ramprasad, R. Machine learning strategy for accelerated design of polymer dielectrics. *Sci. Rep.* **6**, 20952 (2016).
145. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
146. Fischer, C. C., Tibbetts, K. J., Morgan, D. & Ceder, G. Predicting crystal structure by merging data mining with quantum mechanics. *Nat. Mater.* **5**, 641–646 (2006).
147. Jäger, M. O. J., Morooka, E. V., Canova, F. F., Himanen, L. & Foster, A. S. Machine learning hydrogen adsorption on nanoclusters through structural descriptors. *npj Comput. Mater.* **4**, 37 (2018).
148. Himanen, L. et al. Dscribe: library of descriptors for machine learning in materials science. Preprint at arXiv:1904.08875 (2019).
149. Schütt, K. T. et al. SchNetPack: a deep learning toolbox for atomistic systems. *J. Chem. Theory Comput.* **15**, 448–455 (2018).
150. Ward, L. et al. Matminer: an open source toolkit for materials data mining. *Comput. Mater. Sci.* **152**, 60–69 (2018).
151. Yao, K., Herr, J. E., Toth, D., Mckintyre, R. & Parkhill, J. The tensormol-0.1 model chemistry: a neural network augmented with long-range physics. *Chem. Sci.* **9**, 2261–2269 (2018).
152. Ong, S. P. et al. Python materials genomics (pymatgen): a robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **68**, 314–319 (2013).
153. Boser, B. E., Guyon, I. M. & Vapnik, V. N. A training algorithm for optimal margin classifiers. In *Proc. Fifth Annual Workshop on Computational learning theory - COLT'92* (ed Haussler, D.) 144–152 (ACM Press, New York City, NY, 1992).
154. Schölkopf, B., Tsuda, K. & Vert, J.-P. (eds.) *Kernel Methods in Computational Biology* (MIT Press, Cambridge, MA, 2004).
155. Devroye, L, Györfi, L. & Lugosi, G. in *A Probabilistic Theory of Pattern Recognition. Stochastic Modelling and Applied Probability* 187–213 (Springer, New York, NY, 1996).
156. Ueno, T., Rhone, T. D., Hou, Z., Mizoguchi, T. & Tsuda, K. COMBO: an efficient Bayesian optimization library for materials science. *Mater. Des.* **4**, 18–21 (2016).
157. Deisenroth, M. P. & Ng, J. W. Distributed Gaussian processes. In *ICML'15 Proc. 32nd International Conference on International Conference on Machine Learning - Volume 37* (eds Bach, F. & Blei, D.) 1481–1490 (ICML, Lille, 2015).
158. Santosa, F. & Symes, W. W. Linear inversion of band-limited reflection seismograms. *SIAM J. Sci. Stat. Comp.* **7**, 1307–1330 (1986).
159. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B* **58**, 267–288 (1996).
160. Candès, E. J., Romberg, J. K. & Tao, T. Stable signal recovery from incomplete and inaccurate measurements. *Commun. Pure Appl. Math.* **59**, 1207–1223 (2006).
161. Eldar, Y. C. & Kutyniok, G. (eds.) *Compressed Sensing: Theory and Applications* (Cambridge University Press, Cambridge, 2012).
162. Ghiringhelli, L. M. et al. Learning physical descriptors for materials science by compressed sensing. *New J. Phys.* **19**, 023017 (2017).
163. Ouyang, R., Curtarolo, S., Ahmetcik, E., Scheffler, M. & Ghiringhelli, L. M. SISSO: a compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. *Phys. Rev. Mater.* **2**, 083802 (2018).
164. Fan, J. & Lv, J. Sure independence screening for ultrahigh dimensional feature space. *J. R. Stat. Soc. Ser. B* **70**, 849–911 (2008).
165. Tropp, J. A. & Gilbert, A. C. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inf. Theory* **53**, 4655–4666 (2007).
166. Pati, Y., Rezaiifar, R. & Krishnaprasad, P. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In *Proc. 27th Asilomar Conference on Signals, Systems and Computers* (ed Singh, A.) 40–44 (IEEE Comput. Soc. Press, Los Alamitos, CA, 1993).
167. Schmidt, M. & Lipson, H. Distilling free-form natural laws from experimental data. *Science* **324**, 81–85 (2009).
168. Pankajakshan, P. et al. Machine learning and statistical analysis for materials science: stability and transferability of fingerprint descriptors and chemical insights. *Chem. Mater.* **29**, 4190–4201 (2017).
169. Jain, P., Tewari, A. & Kar, P. On iterative hard thresholding methods for high-dimensional m-estimation. *Adv. Neural Inf. Process. Syst.* **27**, 685–693 (2014).
170. Pearson, K. LIII. On lines and planes of closest fit to systems of points in space. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **2**, 559–572 (1901).
171. Jolliffe, I. *Principal Component Analysis* (Springer-Verlag, Berlin, 2002).
172. Quinlan, J. Simplifying decision trees. *Int. J. Man. Mach. Stud.* **27**, 221–234 (1987).
173. Quinlan, J. R. Induction of decision trees. *Mach. Learn.* **1**, 81–106 (1986).
174. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
175. Geurts, P., Ernst, D. & Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **63**, 3–42 (2006).

176. Strobl, C., Boulesteix, A.-L., Zeileis, A. & Hothorn, T. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics* **8**, 25 (2007).
177. Toyao, T. et al. Toward effective utilization of methane: machine learning prediction of adsorption energies on metal alloys. *J. Phys. Chem. C* **122**, 8315–8326 (2018).
178. Shandiz, M. A. & Gauvin, R. Application of machine learning methods for the prediction of crystal system of cathode materials in lithium-ion batteries. *Comput. Mater. Sci.* **117**, 270–278 (2016).
179. Schapire, R. E. The strength of weak learnability. *Mach. Learn.* **5**, 197–227 (1990).
180. Friedman, J. H. Greedy function approximation: a gradient boosting machine. *Ann. Stat.* **29**, 1189–1232 (2001).
181. Mason, L., Baxter, J., Bartlett, P. L. & Frean, M. R. in *Advances in Neural Information Processing Systems 12* (eds. Solla, S. A., Leen, T. K. & Müller, K.) 512–518 (MIT Press, Cambridge, MA, 2000).
182. Freund, Y. & Schapire, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **55**, 119–139 (1997).
183. Drucker, H. Improving regressors using boosting techniques. In *ICML '97 Proc. Fourteenth International Conference on Machine Learning* (ed Kaufmann, M.) 107–115 (ICML, Lille, 1997).
184. de Jong, M. et al. A statistical learning framework for materials science: application to elastic moduli of k-nary inorganic polycrystalline compounds. *Sci. Rep.* **6**, 34256 (2016).
185. Evans, J. D. & Coudert, F.-X. Predicting the mechanical properties of zeolite frameworks by machine learning. *Chem. Mater.* **29**, 7833–7839 (2017).
186. Schmidt, J., Chen, L., Botti, S. & Marques, M. A. L. Predicting the stability of ternary intermetallics with density functional theory and machine learning. *J. Chem. Phys.* **148**, 241728 (2018).
187. Kohonen, T. *Self-Organizing Maps* (Springer, Berlin, 2001).
188. Ackley, D. H., Hinton, G. E. & Sejnowski, T. J. A learning algorithm for Boltzmann machines. *Cogn. Sci.* **9**, 147–169 (1985).
189. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).
190. Carleo, G. & Troyer, M. Solving the quantum many-body problem with artificial neural networks. *Science* **355**, 602–606 (2017).
191. Kolen, J. F. & Kremer, S. C. (eds). in *A Field Guide to Dynamical Recurrent Networks*. Ch. 11 (Wiley-IEEE Press, Hoboken, NJ, 2001).
192. Nair, V. & Hinton, G. E. in *ICML'10 Proc. 27th International Conference on International Conference on Machine Learning* (eds Fürnkranz, J. & Joachims, T.) 807–814 (Omnipress, Athens, 2010).
193. Glorot, X., Bordes, A. & Bengio, Y. Deep sparse rectifier neural networks. In *Proc. Fourteenth International Conference on Artificial Intelligence and Statistics*, *vol. 15 of Proc. Machine Learning Research* (eds Gordon, G., Dunson, D. & Dudík, M.) 315–323 (PMLR, London, 2011).
194. Clevert, D.-A., Unterthiner, T. & Hochreiter, S. Fast and accurate deep network learning by exponential linear units (ELUs). Preprint at arXiv:1511.07289 (2015).
195. LeCun, Y. et al. Handwritten digit recognition with a back-propagation network. *Adv. Neural Inf. Process. Syst.* **2**, 396–404 (1990).
196. Stanley, K. O. & Miikkulainen, R. Evolving neural networks through augmenting topologies. *Evol. Comput.* **10**, 99–127 (2002).
197. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
198. Ioffe, S. & Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. Preprint at arXiv:1502.03167 (2015).
199. Abadi, M. et al. TensorFlow: large-scale machine learning on heterogeneous systems. *arXiv*:1603.04467. https://arxiv.org/abs/1603.04467 (2011).
200. Paszke, A. et al. Automatic differentiation in pytorch. In *NIPS 2017 Autodiff Workshop: The Future of Gradient-based Machine Learning Software and Techniques* (2017).
201. Plaut, D. C. & Hinton, G. E. Learning sets of filters using back-propagation. *Comput. Speech Lang.* **2**, 35–61 (1987).
202. Hinton, G. E. Reducing the dimensionality of data with neural networks. *Science* **313**, 504–507 (2006).
203. Goodfellow, I. et al. Generative adversarial nets. In *Advances in Neural Information Processing Systems 27* (eds Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D. & Weinberger, K. Q.) (Curran Associates, Inc., Red Hook, NJ, 2014).
204. Kingma, D. P. & Welling, M. Auto-encoding variational Bayes. Preprint at arXiv:1312.6114 (2013).
205. Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A. A. Image-to-image translation with conditional adversarial networks. Preprint at arXiv:1611.07004 (2017).
206. Ledig, C. et al. Photo-realistic single image super-resolution using a generative adversarial network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (eds Chellappa, R., Zhang, Z. & Hoogs, A.) 105–114 (IEEE, Piscataway, NJ, 2017).
207. Schawinski, K., Zhang, C., Zhang, H., Fowler, L. & Santhanam, G. K. Generative adversarial networks recover features in astrophysical images of galaxies beyond the deconvolution limit. *Mon. Not. R. Astron. Soc. Lett.* L110–L114 (2017).
208. Paganini, M., de Oliveira, L. & Nachman, B. CaloGAN: Simulating 3D high energy particle showers in multilayer electromagnetic calorimeters with generative adversarial networks. *Phys. Rev. D* **97**, 014021 (2018).
209. Ghahramani, A., Watt, F. M. & Luscombe, N. M. Generative adversarial networks uncover epidermal regulators and predict single cell perturbations. *bioRxiv*. https://doi.org/10.1101/262501 (2018).
210. Li, X. et al. A deep adversarial learning methodology for designing microstructural material systems. In *ASME. International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, *Volume 2B: 44th Design Automation Conference*, pp. V02BT03A008 (ASME, New York, NY, 2018).
211. Sanchez-Lengeling, B., Outeiral, C., Guimaraes, G. L. & Aspuru-Guzik, A. Optimizing distributions over molecular space. an objective-reinforced generative adversarial network for inverse-design chemistry (ORGANIC). ChemRxiv preprint 5309668/3 (2017).
212. Géron, A. *Hands-On Machine Learning with Scikit-Learn and TensorFlow* (O'Reilly UK Ltd., Farnham, 2017).
213. Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning* (The MIT Press, Cambridge, MA, 2017).
214. Murphy, K. P. *Machine Learning: A Probabilistic Perspective* (MIT Press Ltd, Cambridge, MA, 2012).
215. Bishop, C. M. *Pattern Recognition and Machine Learning* (Springer-Verlag New York Inc., New York, NY, 2006).
216. Kelleher, J. D., Mac Namee, B. & D'Arcy, A. *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms*, *Worked Examples*, *and Case Studies* (MIT Press Ltd, Cambridge, MA, 2015).
217. Maddox, J. Crystals from first principles. *Nature* **335**, 201–201 (1988).
218. Freeman, C. M. & Catlow, C. R. A. Structure predictions in inorganic solids. *J. Chem. Soc. Chem. Commun*. 89–91 (1992).
219. van Eijck, B. P. & Kroon, J. Structure predictions allowing more than one molecule in the asymmetric unit. *Acta Crystallogr. Sect. B* **56**, 535–542 (2000).
220. Pickard, C. J. & Needs, R. J. High-pressure phases of silane. *Phys. Rev. Lett.* **97**, 045504 (2006).
221. Pickard, C. J. & Needs, R. J. Ab initio random structure searching. *J. Phys. Condens. Matter* **23**, 053201 (2011).
222. Pannetier, J., Bassas-Alsina, J., Rodriguez-Carvajal, J. & Caignaert, V. Prediction of crystal structures from crystal chemistry rules by simulated annealing. *Nature* **346**, 343–345 (1990).
223. Schön, J. C. & Jansen, M. First step towards planning of syntheses in solid-state chemistry: determination of promising structure candidates by global optimization. *Angew. Chem. Int. Ed.* **35**, 1286–1304 (1996).
224. Doll, K., Schön, J. C. & Jansen, M. Structure prediction based on ab initio simulated annealing. *J. Phys. Conf. Ser.* **117**, 012014 (2008).
225. Martoňák, R., Laio, A. & Parrinello, M. Predicting crystal structures: the Parrinello-Rahman method revisited. *Phys. Rev. Lett.* **90**, 075503 (2003).
226. Goedecker, S. Minima hopping: an efficient search method for the global minimum of the potential energy surface of complex molecular systems. *J. Chem. Phys.* **120**, 9911–9917 (2004).
227. Bush, T. S., Catlow, C. R. A. & Battle, P. D. Evolutionary programming techniques for predicting inorganic crystal structures. *J. Mater. Chem.* **5**, 1269–1272 (1995).
228. Woodley, S. M. & Catlow, R. Crystal structure prediction from first principles. *Nat. Mater.* **7**, 937–946 (2008).
229. Gottwald, D., Kahl, G. & Likos, C. N. Predicting equilibrium structures in freezing processes. *J. Chem. Phys.* **122**, 204503 (2005).
230. Paszkowicz, W. Genetic algorithms, a nature-inspired tool: survey of applications in materials science and related fields. *Mater. Manuf. Process.* **24**, 174–197 (2009).
231. Glass, C. W., Oganov, A. R. & Hansen, N. USPEX–evolutionary crystal structure prediction. *Comput. Phys. Commun.* **175**, 713–720 (2006).
232. Wang, Y., Lv, J., Zhu, L. & Ma, Y. Crystal structure prediction via particle-swarm optimization. *Phys. Rev. B* **82**, 094116 (2010).
233. Wang, Y., Lv, J., Zhu, L. & Ma, Y. CALYPSO: a method for crystal structure prediction. *Comput. Phys. Commun.* **183**, 2063–2070 (2012).
234. Reilly, A. M. et al. Report on the sixth blind test of organic crystal structure prediction methods. *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.* **72**, 439–459 (2016).
235. Zakutayev, A. et al. Experimental synthesis and properties of metastable $CuNbN_2$ and theoretical extension to other ternary copper nitrides. *Chem. Mater.* **26**, 4970–4977 (2014).

236. Shoemaker, D. P. et al. In situ studies of a platform for metastable inorganic crystal growth and materials discovery. *Proc. Natl Acad. Sci. USA* **111**, 10922–10927 (2014).
237. Kim, K. et al. Machine-learning-accelerated high-throughput materials screening: discovery of novel quaternary Heusler compounds. *Phys. Rev. Mater.* **2**, 123801 (2018).
238. Jacobs, R., Mayeshiba, T., Booske, J. & Morgan, D. Material discovery and design principles for stable, high activity perovskite cathodes for solid oxide fuel cells. *Adv. Energy Mat.* **8**, 1702708 (2018).
239. Goldschmidt, V. M. Die gesetze der krystallochemie. *Die Nat.* **14**, 477–485 (1926).
240. Pilania, G., Balachandran, P. V., Gubernatis, J. E. & Lookman, T. Classification of ABO_3 perovskite solids: a machine learning study. *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.* **71**, 507–513 (2015).
241. Balachandran, P. V. et al. Predictions of new ABO_3 perovskite compounds by combining machine learning and density functional theory. *Phys. Rev. Mater.* **2**, 043802 (2018).
242. Oliynyk, A. O. et al. High-throughput machine-learning-driven synthesis of full-Heusler compounds. *Chem. Mater.* **28**, 7324–7331 (2016).
243. Villars, P. *Pearson's Crystal Data, Crystal Structure Database for Inorganic Compounds* (ASM International, Materials Park, OH, 2007).
244. Ma, J. et al. Computational investigation of half-Heusler compounds for spintronics applications. *Phys. Rev. B* **95**, 024411 (2017).
245. Zhang, X., Yu, L., Zakutayev, A. & Zunger, A. Sorting stable versus unstable hypothetical compounds: The case of multi-functional ABX half-Heusler filled tetrahedral structures. *Adv. Funct. Mater.* **22**, 1425–1435 (2012).
246. Weiss, K., Khoshgoftaar, T. M. & Wang, D. A survey of transfer learning. *J. Big Data* **3**, 9 (2016).
247. Hautier, G., Fischer, C. C., Jain, A., Mueller, T. & Ceder, G. Finding nature's missing ternary oxide compounds using machine learning and density functional theory. *Chem. Mater.* **22**, 3762–3767 (2010).
248. Pettifor, D. A chemical scale for crystal-structure maps. *Solid State Commun.* **51**, 31–34 (1984).
249. Pettifor, D. G. The structures of binary compounds. I. Phenomenological structure maps. *J. Phys. C Solid State Phys.* **19**, 285–313 (1986).
250. Pettifor, D. G. Structure maps for pseudobinary and ternary phases. *Mater. Sci. Tech.* **4**, 675–691 (1988).
251. Glawe, H., Sanna, A., Gross, E. K. U. & Marques, M. A. L. The optimal one dimensional periodic table: a modified pettifor chemical scale from data mining. *New J. Phys.* **18**, 093011 (2016).
252. Morita, T. Cluster variation method of cooperative phenomena and its generalization II. quantum statistics. *J. Phys. Soc. Jpn.* **12**, 1060–1063 (1957).
253. Sinkov, N. A. & Harynuk, J. J. Cluster resolution: a metric for automated, objective and optimized feature selection in chemometric modeling. *Talanta* **83**, 1079–1087 (2011).
254. Oliynyk, A. O. et al. Disentangling structural confusion through machine learning: Structure prediction and polymorphism of equiatomic ternary phases ABC. *J. Am. Chem. Soc.* **139**, 17870–17881 (2017).
255. Park, W. B. et al. Classification of crystal structure using a convolutional neural network. *IUCrJ* **4**, 486–494 (2017).
256. Obeidat, S. M., Al-Momani, I., Haddad, A. & Yasein, M. B. Combination of ICP-OES, XRF and XRD techniques for analysis of several dental ceramics and their identification using chemometrics. *Spectroscopy* **26**, 141–149 (2011).
257. MITSUI, T. & SATOH, M. Determination of ammonium nitrate in dynamite without separation by multivariate analysis using X-ray diffractometer. *J. Chem. Softw.* **4**, 33–40 (1998).
258. Chen, Z. P. et al. Enhancing the signal-to-noise ratio of X-ray diffraction profiles by smoothed principal component analysis. *Anal. Chem.* **77**, 6563–6570 (2005).
259. Matos, C. R. S., Xavier, M. J., Barreto, L. S., Costa, N. B. & Gimenez, I. F. Principal component analysis of X-ray diffraction patterns to yield morphological classification of brucite particles. *Anal. Chem.* **79**, 2091–2095 (2007).
260. Tatlier, M. Artificial neural network methods for the prediction of framework crystal structures of zeolites from XRD data. *Neural Comput. Appl.* **20**, 365–371 (2010).
261. Agatonovic-Kustrin, S., Wu, V., Rades, T., Saville, D. & Tucker, I. Ranitidine hydrochloride X-ray assay using a neural network. *J. Pharm. Biomed. Anal.* **22**, 985–992 (2000).
262. Park, W. B., Shin, N., Hong, K.-P., Pyo, M. & Sohn, K.-S. A new paradigm for materials discovery: heuristics-assisted combinatorial chemistry involving parameterization of material novelty. *Adv. Funct. Mater.* **22**, 2258–2266 (2012).
263. Park, W. B., Singh, S. P. & Sohn, K.-S. Discovery of a phosphor for light emitting diode applications and its structural determination, $Ba(Si,Al)_5(O,N)_8:Eu^{2+}$. *J. Am. Chem. Soc.* **136**, 2363–2373 (2014).
264. Werner, P. E., Eriksson, L. & Westdahl, M. TREOR, a semi-exhaustive trial-and-error powder indexing program for all symmetries. *J. Appl. Crystallogr.* **18**, 367–370 (1985).
265. LeCun, Y. et al. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* **1**, 541–551 (1989).
266. Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2324 (1998).
267. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25* (eds Pereira, F., Burges, C. J. C., Bottou, L. & Weinberger, K. Q.) 1097–1105 (Curran Associates, Inc., Red Hook, NY, 2012).
268. Ziletti, A., Kumar, D., Scheffler, M. & Ghiringhelli, L. M. Insightful classification of crystal structures using deep learning. *Nat. Commun.* **9**, 2775 (2018).
269. Zeiler, M. D. & Fergus, R. Visualizing and understanding convolutional networks. In *Computer Vision — ECCV 2014* (eds Fleet, D., Pajdla, T., Schiele, B. & Tuytelaars, T.) 818–833 (Springer International Publishing, Basel, 2014).
270. Bach, S. et al. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS ONE* **10**, e0130140 (2015).
271. Kumar, D., Menkovski, V., Taylor, G. W. & Wong, A. Understanding anatomy classification through attentive response maps. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (eds Wang, F. et al.) 1130–1133 (IEEE, Piscataway, NJ, 2018).
272. Montavon, G., Lapuschkin, S., Binder, A., Samek, W. & Müller, K.-R. Explaining nonlinear classification decisions with deep Taylor decomposition. *Pattern Recogn.* **65**, 211–222 (2017).
273. Kumar, D., Wong, A. & Taylor, G. W. Explaining the unexplained: a class-enhanced attentive response (CLEAR) approach to understanding deep neural networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* 1686–1694 (IEEE, Piscataway, NJ, 2017).
274. Dimiduk, D. M., Holm, E. A. & Niezgoda, S. R. Perspectives on the impact of machine learning, deep learning, and artificial intelligence on materials, processes, and structures engineering. *Integr. Mater. Manuf. Innov.* **7**, 157–172 (2018).
275. Kalinin, S. V., Sumpter, B. G. & Archibald, R. K. Big–deep–smart data in imaging for guiding materials design. *Nat. Mater.* **14**, 973–980 (2015).
276. Liu, Z. et al. Tomogan: low-dose X-ray tomography with generative adversarial networks. Preprint at arXiv:1902.07582 (2019).
277. Liu, R., Agrawal, A., Liao, W., Choudhary, A. & De Graef, M. Materials discovery: Understanding polycrystals from large-scale electron patterns. In *2016 IEEE International Conference on Big Data (Big Data)* (ed Joshi, J.) 2261–2269 (IEEE, Piscataway, NJ, 2016).
278. Wang, B. et al. Deep learning for analysing synchrotron data streams. In *2016 New York Scientific Data Summit (NYSDS)* 1–5 (IEEE, 2016).
279. DeCost, B. L., Jain, H., Rollett, A. D. & Holm, E. A. Computer vision and machine learning for autonomous characterization of am powder feedstocks. *JOM* **69**, 456–465 (2017).
280. Yamashita, T. et al. Crystal structure prediction accelerated by Bayesian optimization. *Phys. Rev. Mater.* **2**, 013803 (2018).
281. Chapelle, O. & Li, L. An empirical evaluation of Thompson sampling. In *Advances in Neural Information Processing Systems 24* (eds Shawe-Taylor, J., Zemel, R. S., Bartlett, P. L., Pereira, F. & Weinberger, K. Q.) 2249–2257 (Curran Associates, Inc., Red Hook, NY, 2011).
282. Li, X. et al. A transfer learning approach for microstructure reconstruction and structure-property predictions. *Sci. Rep.* **8**, 13461 (2018).
283. Zhang, B. et al. Machine learning technique for prediction of magnetocaloric effect in La(Fe,Si/Al)$_1$3-based materials. *Chin. Phys. B* **27**, 067503 (2018).
284. Balachandran, P. V., Xue, D. & Lookman, T. Structure–Curie temperature relationships in $BaTiO_3$-based ferroelectric perovskites: Anomalous behavior of (Ba, Cd)TiO_3 from DFT, statistical inference, and experiments. *Phys. Rev. B* **93**, 144111 (2016).
285. Sanvito, S. et al. Accelerated discovery of new magnets in the Heusler alloy family. *Sci. Adv.* **3**, e1602241 (2017).
286. Zhai, X., Chen, M. & Lu, W. Accelerated search for perovskite materials with higher Curie temperature based on the machine learning methods. *Comput. Mater. Sci.* **151**, 41–48 (2018).
287. Dam, H. C. et al. Important descriptors and descriptor groups of Curie temperatures of rare-earth transition-metal binary alloys. *J. Phys. Soc. Jpn.* **87**, 113801 (2018).
288. Legrain, F., Carrete, J., van Roekeghem, A., Curtarolo, S. & Mingo, N. How chemical composition alone can predict vibrational free energies and entropies of solids. *Chem. Mater.* **29**, 6220–6227 (2017).
289. Zhuo, Y., Tehrani, A. M. & Brgoch, J. Predicting the band gaps of inorganic solids by machine learning. *J. Phys. Chem. Lett.* **9**, 1668–1673 (2018).
290. Dey, P. et al. Informatics-aided bandgap engineering for solar materials. *Comput. Mater. Sci.* **83**, 185–195 (2014).
291. Lee, J., Seko, A., Shitara, K., Nakayama, K. & Tanaka, I. Prediction model of band gap for inorganic compounds by combination of density functional theory calculations and machine learning techniques. *Phys. Rev. B* **93**, 115104 (2016).

292. Pilania, G., Gubernatis, J. & Lookman, T. Multi-fidelity machine learning models for accurate bandgap predictions of solids. *Comput. Mater. Sci.* **129**, 156–163 (2017).
293. Rajan, A. C. et al. Machine-learning-assisted accurate band gap predictions of functionalized MXene. *Chem. Mater.* **30**, 4031–4038 (2018).
294. Sparks, T., Kauwe, S. & Welker, T. Extracting knowledge from DFT: experimental band gap predictions through ensemble learning. ChemRxiv preprint 7236029 (2018).
295. Weston, L. & Stampfl, C. Machine learning the band gap properties of kesterite $I_2-II-IV-V_4$ quaternary compounds for photovoltaics applications. *Phys. Rev. Mater.* **2**, 085407 (2018).
296. Gu, T., Lu, W., Bao, X. & Chen, N. Using support vector regression for the prediction of the band gap and melting point of binary and ternary compound semiconductors. *Solid State Sci.* **8**, 129–136 (2006).
297. Pilania, G. et al. Machine learning bandgaps of double perovskites. *Sci. Rep.* **6**, 19375 (2016).
298. Setyawan, W., Gaume, R. M., Lam, S., Feigelson, R. S. & Curtarolo, S. High-throughput combinatorial database of electronic band structures for inorganic scintillator materials. *ACS Comb. Sci.* **13**, 382–390 (2011).
299. Lu, S. et al. Accelerated discovery of stable lead-free hybrid organic-inorganic perovskites via machine learning. *Nat. Commun.* **9**, 3405 (2018).
300. Pilania, G. & Liu, X.-Y. Machine learning properties of binary wurtzite superlattices. *J. Mater. Sci.* **53**, 6652–6664 (2018).
301. Cassar, D. R., de Carvalho, A. C. & Zanotto, E. D. Predicting glass transition temperatures using neural networks. *Acta Mater.* **159**, 249–256 (2018).
302. Liu, Y., Zhao, T., Yang, G., Ju, W. & Shi, S. The onset temperature (Tg) of As Se1 glasses transition prediction: a comparison of topological and regression analysis methods. *Comput. Mater. Sci.* **140**, 315–321 (2017).
303. Zhan, T., Fang, L. & Xu, Y. Prediction of thermal boundary resistance by the machine learning method. *Sci. Rep.* **7**, 7109 (2017).
304. Seko, A. et al. Prediction of low-thermal-conductivity compounds with first-principles anharmonic lattice-dynamics calculations and Bayesian optimization. *Phys. Rev. Lett.* **115**, 205901 (2015).
305. van Roekeghem, A., Carrete, J., Oses, C., Curtarolo, S. & Mingo, N. High-throughput computation of thermal conductivity of high-temperature solid phases: the case of oxide and fluoride perovskites. *Phys. Rev. X* **6**, 041061 (2016).
306. Pham, T.-L. et al. Learning structure-property relationship in crystalline materials: a study of lanthanide–transition metal alloys. *J. Chem. Phys.* **148**, 204106 (2018).
307. Pilania, G., Gubernatis, J. E. & Lookman, T. Structure classification and melting temperature prediction in octet AB solids via machine learning. *Phys. Rev. B* **91**, 214302 (2015).
308. Kikuchi, S., Oda, H., Kiyohara, S. & Mizoguchi, T. Bayesian optimization for efficient determination of metal oxide grain boundary structures. *Phys. B* **532**, 24–28 (2018).
309. Kiyohara, S., Oda, H., Tsuda, K. & Mizoguchi, T. Acceleration of stable interface structure searching using a kriging approach. *Jpn. J. Appl. Phys.* **55**, 045502 (2016).
310. Kiyohara, S., Oda, H., Miyata, T. & Mizoguchi, T. Prediction of interface structures and energies via virtual screening. *Sci. Adv.* **2**, e1600746 (2016).
311. Zhu, Q., Samanta, A., Li, B., Rudd, R. E. & Frolov, T. Predicting phase behavior of grain boundaries with evolutionary search and machine learning. *Nat. Commun.* **9**, 467 (2018).
312. Rosenbrock, C. W., Homer, E. R., Csányi, G. & Hart, G. L. W. Discovering the building blocks of atomic systems using machine learning: application to grain boundaries. *npj Comput. Mater.* **3**, 29 (2017).
313. Furmanchuk, A. et al. Prediction of Seebeck coefficient for compounds without restriction to fixed stoichiometry: a machine learning approach. *J. Comput. Chem.* **39**, 191–202 (2017).
314. Abdellahi, M., Bahmanpour, M. & Bahmanpour, M. Modeling Seebeck coefficient of $Ca_{3-x}M_xCo_4O_9$ (M = Sr, Pr, Ga, Ca, Ba, La, Ag) thermoelectric ceramics. *Ceram. Int.* **41**, 345–352 (2015).
315. Carrete, J., Mingo, N., Wang, S. & Curtarolo, S. Nanograined half-Heusler semiconductors as advanced thermoelectrics: an ab initio high-throughput statistical study. *Adv. Funct. Mater.* **24**, 7427–7432 (2014).
316. Tehrani, A. M. et al. Machine learning directed search for ultraincompressible, superhard materials. *J. Am. Chem. Soc.* **140**, 9844–9853 (2018).
317. Yeo, B. C., Kim, D., Kim, C. & Han, S. S. Pattern learning electronic density of states. Preprint at arXiv:1808.03383 (2018).
318. Broderick, S. R., Aourag, H. & Rajan, K. Classification of oxide compounds through data-mining density of states spectra. *J. Am. Ceram. Soc.* **94**, 2974–2980 (2011).
319. Meredig, B. & Wolverton, C. Dissolving the periodic table in cubic zirconia: data mining to discover chemical trends. *Chem. Mater.* **26**, 1985–1991 (2014).
320. Zhang, Y. & Kim, E.-A. Quantum loop topography for machine learning. *Phys. Rev. Lett.* **118**, 216401 (2017).
321. Zhang, P., Shen, H. & Zhai, H. Machine learning topological invariants with neural networks. *Phys. Rev. Lett.* **120**, 066401 (2018).
322. Deng, D.-L., Li, X. & Sarma, S. D. Machine learning topological states. *Phys. Rev. B* **96**, 195145 (2017).
323. Sun, N., Yi, J., Zhang, P., Shen, H. & Zhai, H. Deep learning topological invariants of band insulators. *Phys. Rev. B* **98**, 085402 (2018).
324. Beach, M. J. S., Golubeva, A. & Melko, R. G. Machine learning vortices at the Kosterlitz-Thouless transition. *Phys. Rev. B* **97**, 045207 (2018).
325. Pilozzi, L., Farrelly, F. A., Marcucci, G. & Conti, C. Machine learning inverse problem for topological photonics. *Commun. Phys.* **1**, 57 (2018).
326. Carrasquilla, J. & Melko, R. G. Machine learning phases of matter. *Nat. Phys.* **13**, 431–434 (2017).
327. Owolabi, T. O., Akande, K. O. & Olatunji, S. O. Prediction of superconducting transition temperatures for Fe-based superconductors using support vector machine. *Adv. Phys. Theor. Appl.* **35**, 12–26 (2014).
328. Owolabi, T. O., Akande, K. O. & Olatunji, S. O. Estimation of superconducting transition temperature t_c for superconductors of the doped MgB_2 system from the crystal lattice parameters using support vector regression. *J. Supercond. Nov. Magn.* **28**, 75–81 (2014).
329. Ling, J., Hutchinson, M., Antono, E., Paradiso, S. & Meredig, B. High-dimensional materials and process optimization using data-driven experimental design with well-calibrated uncertainty estimates. *Integr. Mater. Manuf. Innov.* **6**, 207–217 (2017).
330. Sendek, A. D. et al. Machine learning-assisted discovery of solid Li-ion conducting materials. *Chem. Mater.* **31**, 342–352 (2019).
331. Waag, W., Fleischer, C. & Sauer, D. U. Critical review of the methods for monitoring of lithium-ion batteries in electric and hybrid vehicles. *J. Power Sources* **258**, 321–339 (2014).
332. Tran, F. & Blaha, P. Accurate band gaps of semiconductors and insulators with a semilocal exchange-correlation potential. *Phys. Rev. Lett.* **102**, 226401 (2009).
333. Sun, J., Ruzsinszky, A. & Perdew, J. P. Strongly constrained and appropriately normed semilocal density functional. *Phys. Rev. Lett.* **115**, 036402 (2015).
334. Heyd, J., Scuseria, G. E. & Ernzerhof, M. Hybrid functionals based on a screened Coulomb potential. *J. Chem. Phys.* **118**, 8207–8215 (2003).
335. Snyder, G. J. & Toberer, E. S. Complex thermoelectric materials. *Mater. Sustain. Energy* 101–110 (2010).
336. Cahill, D. G., Watson, S. K. & Pohl, R. O. Lower limit to the thermal conductivity of disordered crystals. *Phys. Rev. B* **46**, 6131–6140 (1992).
337. Yan, J. et al. Material descriptors for predicting thermoelectric performance. *Energy Environ. Sci.* **8**, 983–994 (2015).
338. Liu, A. Y. & Cohen, M. L. Prediction of new low compressibility solids. *Science* **245**, 841–842 (1989).
339. Gilman, J. J. *Electronic Basis of the Strength of Materials* (Cambridge University Press, Cambridge, 2001).
340. Kaner, R. B. Materials science: designing superhard materials. *Science* **308**, 1268–1269 (2005).
341. Kramer, G. J., Farragher, N. P., van Beest, B. W. H. & van Santen, R. A. Interatomic force fields for silicas, aluminophosphates, and zeolites: derivation based on ab initio calculations. *Phys. Rev. B* **43**, 5068–5080 (1991).
342. Cohen, M. L. Theory of bulk moduli of hard solids. *Mater. Sci. Eng. A* **105-106**, 11–18 (1988).
343. Xu, B., Wang, Q. & Tian, Y. Bulk modulus for polar covalent crystals. *Sci. Rep.* **3**, 3068 (2013).
344. Cohen, M. L. Calculation of bulk moduli of diamond and zinc-blende solids. *Phys. Rev. B* **32**, 7988–7991 (1985).
345. Lam, P. K., Cohen, M. L. & Martinez, G. Analytic relation between bulk moduli and lattice constants. *Phys. Rev. B* **35**, 9190–9194 (1987).
346. Loader, C. *Local Regression and Likelihood* (Springer-Verlag, Berlin, 1999).
347. Hill, R. The elastic behaviour of a crystalline aggregate. *Proc. Phys. Soc. Sect. A* **65**, 349–354 (1952).
348. Sastre, G. & Gale, J. D. Derivation of an interatomic potential for germanium- and silicon-containing zeolites and its application to the study of the structures of octadecasil, ASU-7, and ASU-9 materials. *Chem. Mater.* **15**, 1788–1796 (2003).
349. Tsuneyuki, S., Tsukada, M., Aoki, H. & Matsui, Y. First-principles interatomic potential of silica applied to molecular dynamics. *Phys. Rev. Lett.* **61**, 869–872 (1988).
350. van Beest, B. W. H., Kramer, G. J. & van Santen, R. A. Force fields for silicas and aluminophosphates based on ab initio calculations. *Phys. Rev. Lett.* **64**, 1955–1958 (1990).
351. Gale, J. D. Analytical free energy minimization of silica polymorphs. *J. Phys. Chem. B* **102**, 5423–5431 (1998).
352. Sanders, M. J., Leslie, M. & Catlow, C. R. A. Interatomic potentials for SiO_2. *J. Chem. Soc. Chem. Commun*. 1271–1273 (1984).
353. Siddorn, M., Coudert, F.-X., Evans, K. E. & Marmier, A. A systematic typology for negative Poisson's ratio materials and the prediction of complete auxeticity in pure silica zeolite JST. *Phys. Chem. Chem. Phys.* **17**, 17927–17933 (2015).

354. Birch, F. Finite elastic strain of cubic crystals. *Phys. Rev.* **71**, 809–824 (1947).
355. Murnaghan, F. D. The compressibility of media under extreme pressures. *Proc. Natl Acad. Sci. USA* **30**, 244–247 (1944).
356. Wang, J. & Zhang, S.-C. Topological states of condensed matter. *Nat. Mater.* **16**, 1062–1067 (2017).
357. Hasan, M. Z. & Kane, C. L. Colloquium: topological insulators. *Rev. Mod. Phys.* **82**, 3045–3067 (2010).
358. Moore, J. E. The birth of topological insulators. *Nature* **464**, 194–198 (2010).
359. Aubry, S. & André, G. Analyticity breaking and Anderson localization in incommensurate lattices. *Ann. Isr. Phys. Soc.* **3**, 18 (1980).
360. Harper, P. G. The general motion of conduction electrons in a uniform magnetic field, with application to the diamagnetism of metals. *Proc. Phys. Soc. Sect. A* **68**, 879–892 (1955).
361. Bednorz, J. G. & Müller, K. A. Possible high T_c superconductivity in the Ba–La–Cu–O system. *Z. Phys. B Condens. Matter* **64**, 189–193 (1986).
362. Eliashberg, G. M. Interactions between electrons and lattice vibrations in a superconductor. *Sov. Phys. JETP* **11:3**, 7354388 (1960).
363. Lüders, M. et al. Ab initio theory of superconductivity. I. Density functional formalism and approximate functionals. *Phys. Rev. B* **72**, 024545 (2005).
364. Marques, M. A. L. et al. Ab initio theory of superconductivity. II. application to elemental metals. *Phys. Rev. B* **72**, 024546 (2005).
365. Rabe, K. M., Phillips, J. C., Villars, P. & Brown, I. D. Global multinary structural chemistry of stable quasicrystals, high-T_c ferroelectrics, and high-T_c superconductors. *Phys. Rev. B* **45**, 7650–7676 (1992).
366. Villars, P. & Phillips, J. Quantum structural diagrams and high-T_c superconductivity. *Phys. Rev. B* **37**, 2345–2348 (1988).
367. Klintenberg, M. & Eriksson, O. Possible high-temperature superconductors predicted from electronic structure and data-filtering algorithms. *Comput. Mater. Sci.* **67**, 282–286 (2013).
368. Norman, M. R. Materials design for new superconductors. *Rep. Prog. Phys.* **79**, 074502 (2016).
369. Kohn, W. & Luttinger, J. M. New mechanism for superconductivity. *Phys. Rev. Lett.* **15**, 524–526 (1965).
370. Ward, L., Agrawal, A., Choudhary, A. & Wolverton, C. A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Comput. Mater.* **2**, 16028 (2016).
371. Matthias, B. T. Empirical relation between superconductivity and the number of valence electrons per atom. *Phys. Rev.* **97**, 74–76 (1955).
372. Ziatdinov, M. et al. Deep data mining in a real space: separation of intertwined electronic responses in a lightly doped $BaFe_2As_2$. *Nanotechnology* **27**, 475706 (2016).
373. Nguyen, A.-T., Reiter, S. & Rigo, P. A review on simulation-based optimization methods applied to building performance analysis. *Appl. Energ.* **113**, 1043–1058 (2014).
374. Forrester, A. I. & Keane, A. J. Recent advances in surrogate-based optimization. *Prog. Aerosp. Sci.* **45**, 50–79 (2009).
375. Balachandran, P. V., Xue, D., Theiler, J., Hogden, J. & Lookman, T. Adaptive strategies for materials design using uncertainties. *Sci. Rep.* **6**, 19660 (2016).
376. Schneider, G. et al. Voyages to the (un)known: adaptive design of bioactive compounds. *Trends Biotechnol.* **27**, 18–26 (2009).
377. Bajorath, J. et al. Navigating structure–activity landscapes. *Drug Discov. Today* **14**, 698–705 (2009).
378. Johnson, S. R. The trouble with QSAR (or how i learned to stop worrying and embrace fallacy). *J. Chem. Inf. Model.* **48**, 25–26 (2008).
379. Maggiora, G. M. On outliers and activity cliffs – why QSAR often disappoints. *J. Chem. Inf. Model.* **46**, 1535–1535 (2006).
380. Rasmussen, C. E. & Williams C. K. I. *Gaussian Processes for Machine Learning* (MIT Press Ltd, Cambridge, MA, 2005).
381. Yamawaki, M., Ohnishi, M., Ju, S. & Shiomi, J. Multifunctional structural design of graphene thermoelectrics by Bayesian optimization. *Sci. Adv.* **4**, eaar4192 (2018).
382. Bassman, L. et al. Active learning for accelerated design of layered materials. *npj Comput. Mater.* **4**, 74 (2018).
383. Rouet-Leduc, B., Barros, K., Lookman, T. & Humphreys, C. J. Optimisation of GaN LEDs and the reduction of efficiency droop using active machine learning. *Sci. Rep.* **6**, 24862 (2016).
384. Xue, D. et al. Accelerated search for materials with targeted properties by adaptive design. *Nat. Commun.* **7**, 11241 (2016).
385. Xue, D. et al. Accelerated search for $BaTiO_3$-based piezoelectrics with vertical morphotropic phase boundary using Bayesian learning. *Proc. Natl Acad. Sci. USA* **113**, 13301–13306 (2016).
386. Shan, S. & Wang, G. G. Survey of modeling and optimization strategies to solve high-dimensional design problems with computationally-expensive black-box functions. *Struct. Multidiscip. Optim.* **41**, 219–241 (2009).
387. Jones, D. R., Schonlau, M. & Welch, W. J. Efficient global optimization of expensive black-box functions. *J. Glob. Optim.* **13**, 455–492 (1998).
388. Frazier, P., Powell, W. & Dayanik, S. The knowledge-gradient policy for correlated normal beliefs. *INFORMS J. Comput.* **21**, 599–613 (2009).
389. Balachandran, P. V. et al. in *Materials Discovery and Design* 59–79 (Springer International Publishing, Basel, 2018).
390. Hutchinson, M., Paradiso, S. & Ward, L. *Citrine Informatics Lolo* https://citrine.io/ (2016).
391. Efron, B. *Model Selection Estimation and Bootstrap Smoothing* (Division of Biostatistics, Stanford University, Stanford, CA, 2012).
392. Wager, S., Hastie, T. & Efron, B. Confidence intervals for random forests: the jackknife and the infinitesimal jackknife. *J. Mach. Learn. Res.* **15**, 1625–1651 (2014).
393. Lindström, D. Evaluation of a surrogate based method for global optimization. *Int. J. Comput. Electr. Autom. Control Inf. Eng.* **9**, 1636–1642 (2015).
394. Powell, W. B. & Ryzhov, I. O. *Optimal Learning* (John Wiley & Sons, Inc., Hoboken, NJ, 2012).
395. Browne, C. B. et al. A survey of Monte Carlo tree search methods. *IEEE Trans. Comp. Intel. AI* **4**, 1–43 (2012).
396. Dieb, T. M. et al. MDTS: automatic complex materials design using Monte Carlo tree search. *Sci. Technol. Adv. Mat.* **18**, 498–503 (2017).
397. Kiyohara, S. & Mizoguchi, T. Searching the stable segregation configuration at the grain boundary by a Monte Carlo tree search. *J. Chem. Phys.* **148**, 241741 (2018).
398. Dieb, T. M., Hou, Z. & Tsuda, K. Structure prediction of boron-doped graphene by machine learning. *J. Chem. Phys.* **148**, 241716 (2018).
399. Dieb, T. M. & Tsuda, K. in *Nanoinformatics* 65–74 (Springer Singapore, Singapore, 2018).
400. Sawada, R., Iwasaki, Y. & Ishida, M. Boosting material modeling using game tree search. *Phys. Rev. Mat.* **2**, 103802 (2018).
401. Okamoto, Y. Applying Bayesian approach to combinatorial problem in chemistry. *J. Phys. Chem. A* **121**, 3299–3304 (2017).
402. Dehghannasiri, R. et al. Optimal experimental design for materials discovery. *Comput. Mater. Sci.* **129**, 311–322 (2017).
403. Yoon, B.-J., Qian, X. & Dougherty, E. R. Quantifying the objective cost of uncertainty in complex dynamical systems. *IEEE Trans. Signal Proces.* **61**, 2256–2266 (2013).
404. Wang, Y., Reyes, K. G., Brown, K. A., Mirkin, C. A. & Powell, W. B. Nested-batch-mode learning and stochastic optimization with an application to sequential multi-stage testing in materials science. *SIAM J. Sci. Comput.* **37**, B361–B381 (2015).
405. Wagner, T., Emmerich, M., Deutz, A. & Ponweiser, W. On expected-improvement criteria for model-based multi-objective optimization. In *Parallel Problem Solving from Nature*, (eds Schaefer, R., Cotta, C., Kolodziej, J. & Rudolph, G.) *PPSN XI* 718–727 (Springer, Berlin, Heidelberg, 2010).
406. Emmerich, M. T. M., Deutz, A. H. & Klinkenberg, J. W. Hypervolume-based expected improvement: monotonicity properties and exact computation. In *2011 IEEE Congress of Evolutionary Computation (CEC)* 2147–2154 (IEEE, Piscataway, NJ, 2011).
407. Solomou, A. et al. Multi-objective Bayesian materials discovery: Application on the discovery of precipitation strengthened NiTi shape memory alloys through micromechanical modeling. *Mater. Des.* **160**, 810–827 (2018).
408. Talapatra, A. et al. Autonomous efficient experiment design for materials discovery with Bayesian model averaging. *Phys. Rev. Mat.* **2**, 113803 (2018).
409. Gopakumar, A. M., Balachandran, P. V., Xue, D., Gubernatis, J. E. & Lookman, T. Multi-objective optimization for materials discovery via adaptive design. *Sci. Rep.* **8**, 3738 (2018).
410. Johnson, D. D. in *Informatics for Materials Science and Engineering* 349–364 (Elsevier, Amsterdam, 2013).
411. Tersoff, J. New empirical model for the structural properties of silicon. *Phys. Rev. Lett.* **56**, 632–635 (1986).
412. Stillinger, F. H. & Weber, T. A. Computer simulation of local order in condensed phases of silicon. *Phys. Rev. B* **31**, 5262–5271 (1985).
413. van Duin, A. C. T., Dasgupta, S., Lorant, F. & Goddard, W. A. ReaxFF: a reactive force field for hydrocarbons. *J. Phys. Chem. A* **105**, 9396–9409 (2001).
414. MacKerell, A. D. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
415. Daw, M. S. & Baskes, M. I. Embedded-atom method: derivation and application to impurities, surfaces, and other defects in metals. *Phys. Rev. B* **29**, 6443–6453 (1984).
416. Daw, M. S., Foiles, S. M. & Baskes, M. I. The embedded-atom method: a review of theory and applications. *Mater. Sci. Rep.* **9**, 251–310 (1993).
417. Becker, C. A., Tavazza, F., Trautt, Z. T. & de Macedo, R. A. B. Considerations for choosing and using force fields and interatomic potentials in materials science and engineering. *Curr. Opin. Solid State Mater. Sci.* **17**, 277–283 (2013).
418. Seifert, G. & Joswig, J.-O. Density-functional tight binding – an approximate density-functional theory method. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2**, 456–465 (2012).

419. Koskinen, P. & Mäkinen, V. Density-functional tight-binding for beginners. *Comput. Mater. Sci.* **47**, 237–253 (2009).
420. Porezag, D., Frauenheim, T., Köhler, T., Seifert, G. & Kaschner, R. Construction of tight-binding-like potentials on the basis of density-functional theory: application to carbon. *Phys. Rev. B* **51**, 12947–12957 (1995).
421. Sumpter, B. G. & Noid, D. W. Potential energy surfaces for macromolecules. a neural network technique. *Chem. Phys. Lett.* **192**, 455–462 (1992).
422. Blank, T. B., Brown, S. D., Calhoun, A. W. & Doren, D. J. Neural network models of potential energy surfaces. *J. Chem. Phys.* **103**, 4129–4137 (1995).
423. Handley, C. M. & Popelier, P. L. A. Potential energy surfaces fitted by artificial neural networks. *J. Phys. Chem. A* **114**, 3371–3383 (2010).
424. Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys.* **134**, 074106 (2011).
425. Behler, J. First principles neural network potentials for reactive simulations of large molecular and condensed systems. *Angew. Chem. Int. Ed.* **56**, 12828–12840 (2017).
426. Khaliullin, R. Z., Eshet, H., Kühne, T. D., Behler, J. & Parrinello, M. Nucleation mechanism for the direct graphite-to-diamond phase transition. *Nat. Mater.* **10**, 693–697 (2011).
427. Eshet, H., Khaliullin, R. Z., Kühne, T. D., Behler, J. & Parrinello, M. Ab initio quality neural-network potential for sodium. *Phys. Rev. B* **81**, 184107 (2010).
428. Artrith, N., Morawietz, T. & Behler, J. High-dimensional neural-network potentials for multicomponent systems: applications to zinc oxide. *Phys. Rev. B* **83**, 153101 (2011).
429. Sosso, G. C., Miceli, G., Caravati, S., Behler, J. & Bernasconi, M. Neural network interatomic potential for the phase change material GeTe. *Phys. Rev. B* **85**, 174103 (2012).
430. Artrith, N. & Behler, J. High-dimensional neural network potentials for metal surfaces: a prototype study for copper. *Phys. Rev. B* **85**, 045439 (2012).
431. Boes, J. R., Groenenboom, M. C., Keith, J. A. & Kitchin, J. R. Neural network and ReaxFF comparison for Au properties. *Int. J. Quantum Chem.* **116**, 979–987 (2016).
432. Kobayashi, R., Giofré, D., Junge, T., Ceriotti, M. & Curtin, W. A. Neural network potential for Al-Mg-Si alloys. *Phys. Rev. Mater.* **1**, 053604 (2017).
433. Ghasemi, S. A., Hofstetter, A., Saha, S. & Goedecker, S. Interatomic potentials for ionic systems with density functional accuracy based on charge densities obtained by a neural network. *Phys. Rev. B* **92**, 045131 (2015).
434. Faraji, S. et al. High accuracy and transferability of a neural network potential through charge equilibration for calcium fluoride. *Phys. Rev. B* **95**, 104105 (2017).
435. Witkoskie, J. B. & Doren, D. J. Neural network models of potential energy surfaces: prototypical examples. *J. Chem. Theory Comput.* **1**, 14–23 (2005).
436. Pukrittayakamee, A., Hagan, M., Raff, L., Bukkapatnam, S. & Komanduri, R. in *Intelligent Engineering Systems Through Artificial Neural Networks: Smart Systems Engineering Computational Intelligence in Architecting Complex Engineering Systems*, Vol. 17, 469–474 (ASME Press, New York, NY, 2007).
437. Pukrittayakamee, A. et al. Simultaneous fitting of a potential-energy surface and its corresponding force fields using feedforward neural networks. *J. Chem. Phys.* **130**, 134101 (2009).
438. Hajinazar, S., Shao, J. & Kolmogorov, A. N. Stratified construction of neural network based interatomic models for multicomponent materials. *Phys. Rev. B* **95**, 014114 (2017).
439. Artrith, N., Urban, A. & Ceder, G. Constructing first-principles phase diagrams of amorphous Li_xSi using machine-learning-assisted sampling with an evolutionary algorithm. *J. Chem. Phys.* **148**, 241711 (2018).
440. Thompson, A., Swiler, L., Trott, C., Foiles, S. & Tucker, G. Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials. *J. Comput. Phys.* **285**, 316–330 (2015).
441. Chen, C. et al. Accurate force field for molybdenum by machine learning large materials data. *Phys. Rev. Mater.* **1**, 043603 (2017).
442. Wood, M. A. & Thompson, A. P. Extending the accuracy of the SNAP interatomic potential form. *J. Chem. Phys.* **148**, 241721 (2018).
443. Li, X.-G. et al. Quantum-accurate spectral neighbor analysis potential models for Ni–Mo binary alloys and fcc metals. *Phys. Rev. B* **98**, 094104 (2018).
444. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. Roy. Stat. Soc. Ser. B* **67**, 301–320 (2005).
445. Hastie, T, Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning* (Springer, New York, NY, 2009).
446. Seko, A., Takahashi, A. & Tanaka, I. First-principles interatomic potentials for ten elemental metals via compressed sensing. *Phys. Rev. B* **92**, 054113 (2015).
447. Li, Z., Kermode, J. R. & Vita, A. D. Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces. *Phys. Rev. Lett.* **114**, 096405 (2015).
448. Vita, A. D. & Car, R. A novel scheme for accurate MD simulations of large systems. *MRS Proc.* **491**, 473 (1997).
449. Csányi, G., Albaret, T., Payne, M. C. & Vita, A. D. Learn on the fly: a hybrid classical and quantum-mechanical molecular dynamics simulation. *Phys. Rev. Lett.* **93**, 175503 (2004).
450. Kruglov, I., Sergeev, O., Yanilkin, A. & Oganov, A. R. Energy-free machine learning force field for aluminum. *Sci. Rep.* **7**, 8512 (2017).
451. Glielmo, A., Sollich, P. & Vita, A. D. Accurate interatomic force fields via machine learning with covariant kernels. *Phys. Rev. B* **95**, 214302 (2017).
452. Evgeniou, T., Micchelli, C. A. & Pontil, M. Learning multiple tasks with kernel methods. *J. Mach. Learn. Res.* **6**, 615–637 (2005).
453. Álvarez, M. A., Rosasco, L. & Lawrence, N. D. Kernels for vector-valued functions: a review. *Found. Trends Mach. Learn.* **4**, 195–266 (2012).
454. Glielmo, A., Zeni, C. & Vita, A. D. Efficient nonparametric n-body force fields from machine learning. *Phys. Rev. B* **97**, 184307 (2018).
455. Szlachta, W. J., Bartók, A. P. & Csányi, G. Accuracy and transferability of Gaussian approximation potential models for tungsten. *Phys. Rev. B* **90**, 104108 (2014).
456. Deringer, V. L. & Csányi, G. Machine learning based interatomic potential for amorphous carbon. *Phys. Rev. B* **95**, 094203 (2017).
457. De, S., Bartók, A. P., Csányi, G. & Ceriotti, M. Comparing molecules and solids across structural and alchemical space. *Phys. Chem. Chem. Phys.* **18**, 13754–13769 (2016).
458. Dragoni, D., Daff, T. D., Csányi, G. & Marzari, N. Achieving DFT accuracy with a machine-learning interatomic potential: thermomechanics and defects in bcc ferromagnetic iron. *Phys. Rev. Mater.* **2**, 013808 (2018).
459. Vtek, V. Intrinsic stacking faults in body-centred cubic crystals. *Philos. Mag.* **18**, 773–786 (1968).
460. Deringer, V. L., Pickard, C. J. & Csányi, G. Data-driven learning of total and local energies in elemental boron. *Phys. Rev. Lett.* **120**, 156001 (2018).
461. Rowe, P., Csányi, G., Alfè, D. & Michaelides, A. Development of a machine learning potential for graphene. *Phys. Rev. B* **97**, 054303 (2018).
462. Kamath, A., Vargas-Hernández, R. A., Krems, R. V., Carrington, T. & Manzhos, S. Neural networks vs Gaussian process regression for representing potential energy surfaces: a comparative study of fit quality and vibrational spectrum accuracy. *J. Chem. Phys.* **148**, 241702 (2018).
463. Schmitz, G. & Christiansen, O. Gaussian process regression to accelerate geometry optimizations relying on numerical differentiation. *J. Chem. Phys.* **148**, 241704 (2018).
464. Jacobsen, T., Jørgensen, M. & Hammer, B. On-the-fly machine learning of atomic potential in density functional theory structure optimization. *Phys. Rev. Lett.* **120**, 026102 (2018).
465. Oganov, A. R. & Valle, M. How to quantify energy landscapes of solids. *J. Chem. Phys.* **130**, 104504 (2009).
466. Han, J., Zhang, L., Car, R. & Weinan, E. Deep potential: a general representation of a many-body potential energy surface. *Commun. Comput. Phys.* **23**, 629 (2018).
467. Zhang, L., Han, J., Wang, H., Car, R. & E, W. Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics. *Phys. Rev. Lett.* **120**, 143001 (2018).
468. Tozer, D. J., Ingamells, V. E. & Handy, N. C. Exchange-correlation potentials. *J. Chem. Phys.* **105**, 9200–9213 (1996).
469. Murray, C. W., Handy, N. C. & Laming, G. J. Quadrature schemes for integrals of density functional theory. *Mol. Phys.* **78**, 997–1014 (1993).
470. Snyder, J. C., Rupp, M., Hansen, K., Müller, K.-R. & Burke, K. Finding density functionals with machine learning. *Phys. Rev. Lett.* **108**, 253002 (2012).
471. Hairer, E, Nørsett, S. P. & Wanner, G. *Solving Ordinary Differential Equations I: Nonsitff Problems* (Springer, Berlin, 1993).
472. Snyder, J. C. et al. Orbital-free bond breaking via machine learning. *J. Chem. Phys.* **139**, 224104 (2013).
473. Brockherde, F. et al. Bypassing the Kohn-Sham equations with machine learning. *Nat. Commun.* **8**, 872 (2017).
474. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).
475. Liu, Q. et al. Improving the performance of long-range-corrected exchange-correlation functional with an embedded neural network. *J. Phys. Chem. A* **121**, 7273–7281 (2017).
476. Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **38**, 3098–3100 (1988).
477. Leininger, T., Stoll, H., Werner, H.-J. & Savin, A. Combining long-range configuration interaction with short-range density functionals. *Chem. Phys. Lett.* **275**, 151–160 (1997).
478. Nagai, R., Akashi, R., Sasaki, S. & Tsuneyuki, S. Neural-network Kohn-Sham exchange-correlation potential and its out-of-training transferability. *J. Chem. Phys.* **148**, 241737 (2018).
479. Kadantsev, E. S. & Stott, M. J. Variational method for inverting the Kohn-Sham procedure. *Phys. Rev. A* **69**, 012502 (2004).
480. Foulkes, W. M. C. & Haydock, R. Tight-binding models and density-functional theory. *Phys. Rev. B* **39**, 12520–12536 (1989).

481. Vellido, A., Martn-Guerrero, J. D. & Lisboa, P. J. Making machine learning models interpretable. In *20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)* 163–172 (Bruges (Belgium), 2012). Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.
482. Lipton, Z. C. The mythos of model interpretability. *Queue* **16**, 30:31–30:57 (2018).
483. Lipton, Z. C., Kale, D. C. & Wetzel, R. Modeling missing data in clinical time series with RNNs. In *Proc. Machine Learning for Healthcare 2016* (eds Doshi-Velez, F., Fackler, J., Kale, D., Wallace, B. & Wiens, J.) 253–270 (Proceedings of Machine Learning Research, Children's Hospital LA, Los Angeles, CA, USA, 2016).
484. Kitchin, J. R., Nørskov, J. K., Barteau, M. A. & Chen, J. G. Role of strain and ligand effects in the modification of the electronic and chemical properties of bimetallic surfaces. *Phys. Rev. Lett.* **93**, 156801 (2004).
485. Ma, X., Li, Z., Achenie, L. E. K. & Xin, H. Machine-learning-augmented chemisorption model for CO_2 electroreduction catalyst screening. *J. Phys. Chem. Lett.* **6**, 3528–3533 (2015).
486. Xie, T. & Grossman, J. C. Hierarchical visualization of materials space with graph convolutional neural networks. *J. Chem. Phys.* **149**, 174111 (2018).
487. Alexander, J. W. Topological invariants of knots and links. *Trans. Am. Math. Soc.* **30**, 275–275 (1928).
488. Chern, S.-S. Characteristic classes of Hermitian manifolds. *Ann. Math.* **47**, 85 (1946).
489. Smith, J. S. et al. Outsmarting quantum chemistry through transfer learning. ChemRxiv preprint 6744440 (2018).
490. Ramakrishnan, R., Dral, P. O., Rupp, M. & von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data* **1**, 140022 (2014).
491. Ruddigkeit, L., van Deursen, R., Blum, L. C. & Reymond, J.-L. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *J. Chem. Inf. Model.* **52**, 2864–2875 (2012).
492. Ramakrishnan, R., Hartmann, M., Tapavicza, E. & von Lilienfeld, O. A. Electronic spectra from TDDFT and machine learning in chemical space. *J. Chem. Phys.* **143**, 084111 (2015).
493. Blum, L. C. & Reymond, J.-L. 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13. *J. Am. Chem. Soc.* **131**, 8732–8733 (2009).
494. Montavon, G. et al. Machine learning of molecular electronic properties in chemical compound space. *New J. Phys.* **15**, 095003 (2013).
495. Sutton, C. et al. Nomad 2018 kaggle competition: solving materials science challenges through crowd sourcing. Preprint at arXiv:1812.00085 (2018).
496. Chard, R. et al. DLHub: model and data serving for science. Preprint at arXiv:1811.11213 (2018).