Computational Materials Science 213 (2022) 111612

Contents lists available at ScienceDirect

Computational Materials Science

journal homepage: www.elsevier.com/locate/commatsci

Review article

Recent progress in generative adversarial networks applied to inversely designing inorganic materials: A brief review

Rahma Jabbar [a,*], Rateb Jabbar [b], Slaheddine Kamoun [a]

[a] Materials Engineering and Environment Laboratory (LR11ES46), National Engineering School of Sfax, Sfax University, P. Box 1173-3038, Sfax, Tunisia
[b] Department of Computer Science and Engineering, Qatar University, P.O. Box 2713 Doha, Qatar

ARTICLE INFO

Keywords:
Inorganic material
Generative adversarial networks
Deep learning
Inverse material design

ABSTRACT

Generative adversarial networks (GANs) are deep generative models (GMs) that have recently attracted attention owing to their impressive performance in generating completely novel images, text, music, and speech. Recently, GANs have made interesting progress in designing materials exhibiting desired functionalities, termed 'inverse materials design' (IMD). Because, discovering materials can lead to enormous technological progress, it is critical to provide a systematic review of new GAN applications to inversely designing inorganic materials. In this study, various aspects of GAN-based IMD were examined wherein IMD is a primary design process for discovering materials exhibiting desired features (physical properties, chemical formulae, etc.) by implementing constraints or conditions on input data or algorithms. We discussed fundamental materials databases and relevant machine-learning criteria. Furthermore, the comprehensive software tools currently available to materials scientists were presented. Descriptors including the criteria required for training GAN models were also discussed. Finally, we summarized both challenges and future direction for applying GANs to IMD research.

1. Introduction

Materials scientists aim to advance materials engineering and design because functional materials represent a basis of technological innovation [1]. In this field, it is critical to explore materials behavior and structure–property relationship to ensure adequate material applications [2] and to manage material constraints and limitations [3]. Nevertheless, due to the complexity of materials, experimental investigation and materials development are time-consuming and expensive [4]. Studies have applied computation approaches to calculating material properties and predicting and optimizing their structures [5–7].

Machine learning (ML) consists of sets of computational algorithms used for clustering [8], predicting [9], and classifying tasks [10,11]. Because ML has superhuman abilities, it has revolutionize all spheres of technology and science [12–15]. ML is becoming more ubiquitous across several areas, including healthcare [16,17], agriculture [18,19], geology [20,21], sustainability [22,23] and transportation [24–27]. In materials science field, ML and deep learning (DL) methods has been applied successfully in properties prediction and materials identification [28–30].

So far, studies have explored the relationship between material structures and properties by taking into account restrictive criteria for developing materials [31]. For example, generative models (GMs) have been used to generate materials with desired functionalities, which is termed "Inverse Materials Design" (IMD) [32].

Several studies have reviewed the impact of employing ML in materials science. Fundamental principles of ML models, material databases and tools [33–35] have been investigated. Schmidt et al. [36] discussed the ML fundamental principles, as well as the main tools used to boost materials research. They also investigated the application of solid-state materials design and predict their physical properties. Furthermore, considering materials science, Wei et al. [37] reviewed how ML is implemented. In their study, data processing, the overall prospects of ML models, and examples of applications predicting properties and generating different materials were discussed. Overall, current studies have primarily focused on reviewing how ML techniques are implemented in materials science.

For instance, Wang et al. [38] focused on ML-based IMD. According to the authors, GMs, global optimization (GO), and high-throughput virtual screening (HTVS) are important IMD strategies. Moreover, the authors classified and summarized the applications of ML to IMD according to the material type including porous and inorganic materials and polymers. Nevertheless, current research has not sufficiently investigated the challenges related to applying ML techniques to IMD.

* Corresponding author.
E-mail address: rahma.jabbar@enis.tn (R. Jabbar).

https://doi.org/10.1016/j.commatsci.2022.111612
Received 18 January 2022; Received in revised form 15 June 2022; Accepted 20 June 2022
Available online 1 July 2022

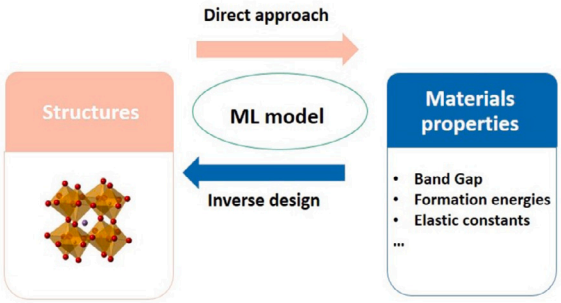

Fig. 1. Schematic representation of IMD process comparing to direct approach.

Chen et al. [39] selected the most well-known GMs, Generative adversarial networks (GANs), and Variational AutoEncoders (VAEs) and explored how they were applied to discovering inorganic systems. These models show promise for classifying unstable and stable materials and for discovering and predicting crystal structures exhibiting specific material properties.

Compared to VAEs, GANs can implicitly learn data distributions by iteratively verifying the data generated from the known prior latent spatial distributions. Moreover, GANs usually generate better photorealistic images than VAEs [40].

This review explores the advanced application of GMs for discovering and predicting inorganic materials. To the best of our knowledge, this paper is the first to provide a comprehensive survey that focuses precisely on implementing GAN models in IMD. First, the use of GAN models in different material-generating applications is discussed. Accordingly, research efforts concerning the challenges of applying an inverse design strategy requiring material properties to be inserted into the structures are reviewed. Subsequently, materials databases and representations requiring specific features to be trained using GAN models are discussed in depth. Finally, the inverse-design-strategy and GAN-model-related challenges that research communities face are presented and directions for further research are proposed.

The remainder of the paper is organized as follows. The first section introduces IMD and GAN models. Subsequently, the best GAN model applications for inversely designing inorganic materials are summarized. Then, the ML tools and materials databases available to researchers and the GAN descriptors are discussed in depth. Finally, the opportunities and challenges of GAN-based IMD are identified and discussed.

2. Inverse design and generative adversarial network backgrounds

2.1. Inverse design

Materials scientists have aimed to apply ML methods coupled with a computational approach to accelerate materials discoveries that have promising application potential [41,42]. Numerous studies have employed “structure” and “property” relationships to generate maps for discovering various molecules and materials. Inverse design, in contrast to the direct approach requiring material property computations, is an interesting process that shows promise for application to predicting and discovering organic and inorganic materials. Starting from the desired material functionality as the input, inverse design aims to identify new materials that exhibit this functionality (see Fig. 1). Noh et al. [43] categorized the inverse design strategies of inorganic crystals as high-throughput virtual screening, global optimization, and generative models.

High-throughput virtual screening [44], one of the earliest inverse design breakthroughs, is defined as the computational approximation of a large set of materials chosen based on intuition and density functional theory (DFT) or Hartree–Fock (HF) calculations to assess the material suitability for a particular function (property). High-throughput virtual screening computations usually involve training a predictor model to map a composition to a property, combinatorially screening different material compositions, and predicting hypothetical candidate materials.

Global optimization [45] is an algorithm that searches for superior bounds for optimizing objective functions and is more efficient than high-throughput virtual screening for navigating chemical spaces. Various stochastic global optimization methods including evolutionary algorithms (EAs), genetic algorithms (GAs), particle swarm optimization, and simulated annealing have been proposed.

Although all these methods have been used to discover materials, they are limited by the high computational costs of large DFT calculations and sample generation bias [43].

ML, or more precisely, generative models are an alternative [46] that could treat chemistry- or physics-based information distinctively compared to the first-principles methods to accelerate materials development.

2.2. Generative Adversarial Networks (GANs)

Generative models (GMs) are one of the inverse design strategies adopted to navigate chemical spaces and can be constructed using several ML algorithms such as GANs, VAEs, recurrent neural networks (RNNs), reinforcement learning (RL), and their hybrids [38].

GANs can implicitly learn data distributions [47]. Classical GANs, like the one shown in Fig. 2(a), combine two networks, a generator, and discriminator, regardless of their specific architectures. The GAN components cooperate in a zero-sum game or a minimax problem [48].

- The generator produces fake data, usually images, from a low-dimensional random noise vector as an input and can be trained to either transform images by feeding them with a real image as input or to generate counterfeit images by feeding them with a random input.
- The discriminator is updated to better distinguish between the fake images obtained from the generator and real images. When the discriminator cannot distinguish between the fake and real images, the model learns to generate images such as those obtained from the real data, and the generator terminates the training.

In a network context, “adversarial” means that networks compete against one other; that is, the generator attempts to fool the discriminator with generated data, whereas the discriminator attempts to expose the generator.

Several GAN extensions have been developed for specific tasks. The major innovations [49] include model structure improvement [50,51], theoretical extensions [52,53], and novel applications [54]. Furthermore, conditional GANs (CGANs) [53] have been identified by adding a condition vector input for both the generator and the discriminator, as shown in Fig. 2(b), to learn multimodal probability distributions. In addition, these approaches can be extended to cross-domain learning such as CycleGANs [55], which can learn how to translate latent spaces (e.g., transfer image styles).

Instead of using a discriminator to classify or predict the probability of whether generated images are real or fake, the Wasserstein GAN (WGAN) [56] changes or replaces the discriminator model with a critic that predicts a score for how real or fake an image appears. WGAN [50] is trained to minimize the Wasserstein distance, which measures the dissimilarity between the distribution difference of the real and fake materials, based on a loss function [see Fig. 2(C)].

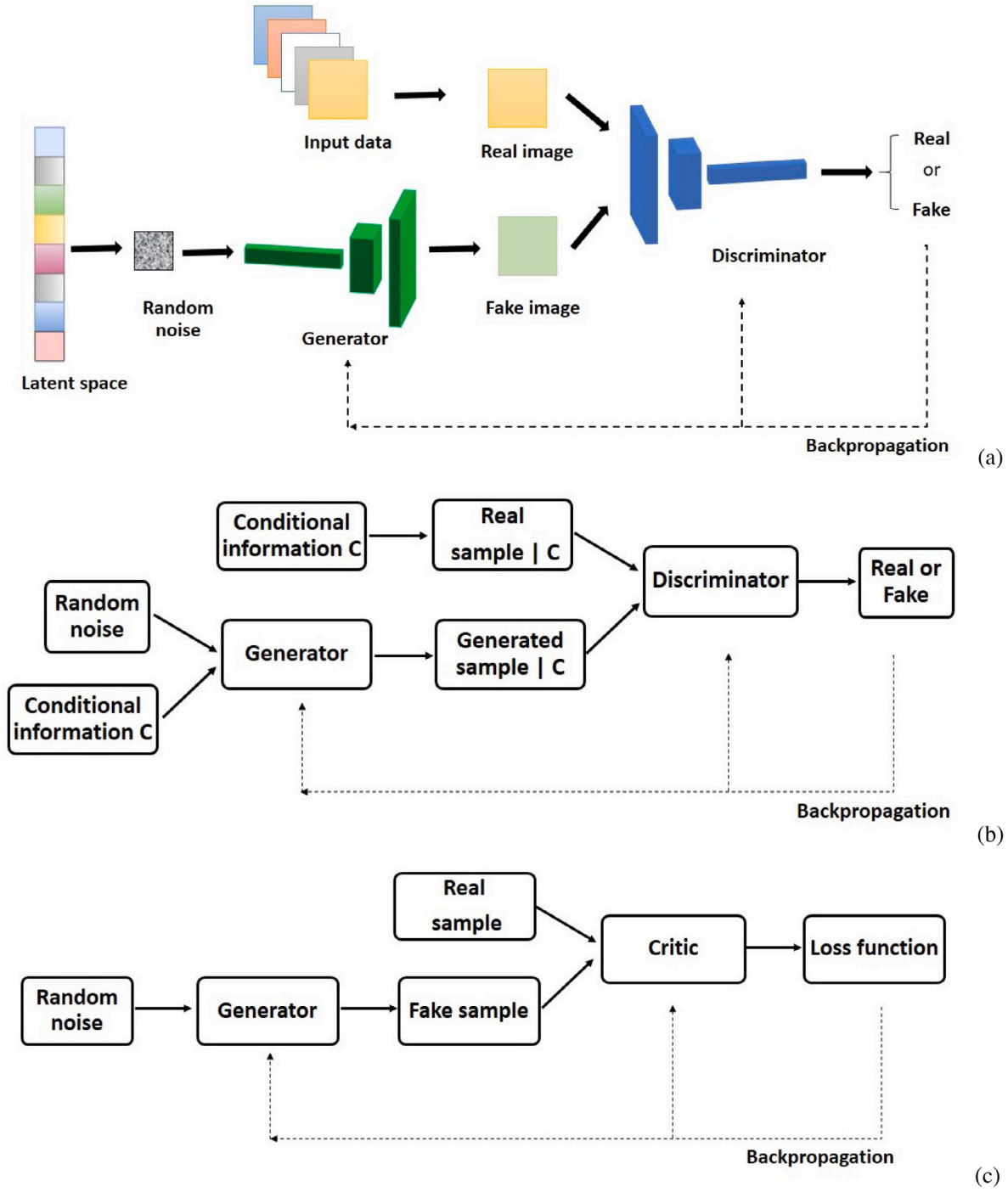

Fig. 2. The fundamental differences of computation procedures and structures between (a) the original GAN model and its variants: (b) CGAN and (c) WGAN.

3. Application of GANs to materials discovery

Although many ML models can be applied to determine structures (e.g., phase diagrams or crystal structures) or predict performance and fit (i.e., descriptors) [36], few materials science datasets are available for these models [35]. GANs can be used to solve the problem of insufficient training samples, which limit the learning ability of various types of ML. They can generate images in any category from scratch. In addition, combining GANs and other techniques (e.g., convolutional neural network (CNN) [57], autoencoder [58], etc.) can produce satisfying results [59].

The first approach to GAN applications in material science is CrystalGAN, proposed by Nouira et al. [60] wherein a CycleGAN-based strategy [55] is applied to generate a crystal structure. The main CrystalGAN goal is to generate novel ternary metal hybrids (i.e., A–B–H phases) from the observed A–H and B–H binary structures. The authors proposed using a matrix of lattice geometry and ionic positions of atoms as inputs. The model was inspired by a cross-domain learning strategy [61], which uses the CycleGAN schema to map one hydride system onto another to increase the number of data. The CrystalGAN model include three steps. The first-step GAN resembles a cross-domain GAN and generates pseudo binary samples in which the domains are mixed. The feature transfer procedure constructs more complex data

than that generated from the samples obtained in the first step. Finally, the second-step GAN predicts ternary stable chemical compounds while adhering to the geometric limitations. Although scientists have proven that novel ternary compounds can be predicted from known binaries, whether this approach can be applied to predict complex crystal structures is uncertain.

Kim et al. [62] used a 2D matrix representation of a set of atomic coordinates and cell parameters and employed a GAN to predict ternary compounds containing only Mg, Mn, and O atoms. Their model used a similar coordinate-based representation exhibiting invariant symmetry addressed with data augmentation technique [63] and an invariant permutation achieved through a symmetry operation. The GAN-based generative framework contained a generator to produce representations, critics, and classifiers to ensure that the predicted materials satisfied the composition conditions. The Wasserstein distance was computed using a critic model to improve the stability. Although the model discovered 23 crystal structures suitable for photoanode applications, the available training data were limited, and the model was restricted to predict Mg–Mn–O structures.

Although the previous models only considered general inverse design principles [43], scientists have developed models that can predict materials that are not included in the training data and subsequently analyze the material physical properties. However, the main inverse design goal is to directly discover materials by introducing desired properties to the input data. Materials were represented using only composition or implicit chemical properties while implementing physical properties (such as band gap and formation energy), as input conditions, is necessary to obtain a truly inverse design.

Accordingly, Dan et al. [64] developed a WGAN to predict chemically valid hypothetical materials using data inputs with specific properties. The discriminator and generator were modeled as a deep neural network (DNN) to develop the 'MatGAN model', which could predict two million materials, among which 92.53% were novel. The materials were described in a matrix indicating their chemical formulae. Although the material representation was restricted, the model performance was evaluated using various approaches (e.g., formation energy, holdout validation, and cross-validation). Of the predicted materials, 84.5% satisfied basic chemical rules such as charge neutrality and balanced electronegativity. Because the material band gap was selected to apply an inverse design, only wide-band gap materials were used as the training data. The predicted materials exhibited wide band gaps, which improved the MatGAN model efficiency. The model was combined with various ML models to classify materials based on the desired properties [65]. The application of this method has enabled the discovery of thermodynamically stable insulators and semiconductors exhibiting band gaps in specific ranges. The band gaps were verified by DFT calculations.

Various GMs have been trained to discover material structures, estimate the structural stability and material synthesizability [66], and predict the structural properties [65]. Including constraints or a loss function in GAN algorithms is useful to achieve a truly inverse design.

Long et al. [67] integrated a constraint network as a backpropagator into a constrained-crystal deep convolutional GAN (CCDCGAN) without embedding material properties in the input to automatically optimize the generator and predictor. The model, mainly contain convolution 3D layers [68], was applied to a binary Bi–Se system to predict distinct crystal structures based on the formation energy target property. The CCDCGAN was more efficient than the traditional GAN for generating stable structures.

Zhao et al. [69] developed a DNN-based GM to discover numerous crystal structures and used only cubic space group ($Fm\bar{3}m$, $F4\bar{3}m$ and $Pm\bar{3}m$) structures to train the model and simplify the model design (CubicGAN). Ternary and quaternary compounds were represented by the element chemical properties, coordinates, and space groups. The predicted structure effectiveness was investigated based on certain criteria including validity, diversity, and uniqueness. The material stability was also verified based on phonon dispersion calculations. A total of 506 hypothetical prototype material structures were discovered. Inspired by the WGAN model, the Wasserstein distance is the loss function.

Previously, a CubicGAN model was developed and then trained using various cubic and noncubic crystal structures [70]. To widen the crystal structure input range, space group symmetry operations were included in the training data. Unlike the original model, the optimized model predicted the crystal structures without constraining fractional coordinates. Both models consisted of a generator and a discriminator with some physical constraints. To predict more feasible crystal structures, additional physics-guided principles and constraints must be introduced to the model. Physical losses (e.g., distance and unit-cell size) were added to the Wasserstein loss. Although the authors proved that physical rules should be considered as constraints for predicting stable materials, the noncubic GAN model prediction performance was not evaluated.

Finally, GANs have been applied to IMD strategies to predict and design inorganic materials [71,72]. The progress in using GMs to design materials exhibiting specific properties increases the likelihood of designing more complex crystalline materials such as metal–organic frameworks (MOFs) and hybrid materials.

4. Materials databases

Traditionally, experiments have played a vital role in discovering and characterizing materials. Human intuition or coincidence is the key to the most important discoveries, and very few materials have been experimentally discovered owing to the rigorous equipment- and resource-related requirements [73]. Therefore, Computational techniques such as Monte Carlo simulations, molecular dynamics (MD) simulations, and DFT have enabled researchers to optimize structural geometries, calculate material properties, and elucidate structure–property relationships [74].

The development of computational methods has bought a computational revolution in materials science. Recently, scientists have combined empirical, theoretical, and computational data into "big" data or datasets such as those used by the Materials Genome Initiative (MGI) project [75]. By developing a materials innovation infrastructure, researchers at the MGI project aim to accelerate the design, discovery, deployment, and development of materials and a multitude of efficient ML applications.

To ensure the model prediction accuracy and effectiveness, the training data should be selected from an available databases or a published papers. Many available databases compile information about atomic coordinates and lattice parameters of crystal materials and their properties as well as phase diagrams for materials.

Table 1 summarizes the most well-known databases containing information about experimental materials and their computational characteristics. Notably, the Inorganic Crystal Structure database (ICSD) and the Cambridge Structural data (CSD) are restricted to crystal structures and their physico-chemical properties including several sets of experimental and computational data. The Materials Project and the Open Quantum Materials Database (OQMD) are defined as high-throughput DFT databases of computed materials structure and properties.

Some criteria are required for selecting data. Briefly, data should be findable, accessible, interoperable, and reusable (FAIR) to upgrade the data employment for new purposes [76]. The dataset size and quality also play a crucial role in training the model. Although similar materials science databases containing information about computational materials are publicly available, the data are limited, and the choice of the ML models is related to the dataset size. Fortunately, however, classical ML algorithms (e.g., support vector machines (SVMs), decision trees, and logistic regressions) only require a few data [77]. Because large datasets containing on the order of thousands of data points are recommended for DL [34], using limited data obeying certain restraint criteria to train ML models is challenging.

Table 1
Largest crystallographic databases.

Database	Description	URL
Inorganic Crystal Structure Database (ICSD)	Crystal structure database of inorganic compounds	https://icsd.fiz-karlsruhe.de
Cambridge Structural Database (CSD)	Repository for organic and metal–organic crystal structures	https://www.ccdc.cam.ac.uk
ChemSpider	Chemical structure	http://www.chemspider.com/
Pearson's Crystal Database	Crystal Structure Database for Inorganic Compounds	http://www.crystalimpact.com/pcd/
Open Quantum Materials Database (OQMD)	Database of DFT-calculated thermodynamic and structural properties of 1,022,603 materials.	http://oqmd.org/
Materials project	Database of materials and corresponding computed properties	https://materialsproject.org/
AFLOW	Database of materials and corresponding calculated mechanical, electronic and thermal properties.	http://aflowlib.org/

Table 2
Material- and ML-related software tools.

Tool	Description	Reference
Magpie descriptor	Materials-Agnostic Platform for Informatics and Exploration: This integrates material properties into one-dimensional vector features by calculating few statistics for each property of elements in compounds.	[81]
DScribe	Library of ML descriptors in materials science.	[82]
MEGNet	Materials graph network (MEGNet) is an implementation of Deep Mind's graph networks for unsupervised ML in materials science.	[83]
matminer	Python library for data mining material properties.	[84]
ChemML	ML and Informatics Program Suite for Chemical and Materials Sciences.	[85]
Pymatgen	Python Materials Genomics.	[86]
NOMAD	NOvel Materials Discovery (NOMAD) is collection of tools to explore correlations in material datasets.	[87]
PyXtal	Python Library for Crystal Structure Generation and Symmetry Analysis.	[88]
SchNet	DL architecture for quantum chemistry.	[89]
Amp	Package to facilitate ML for atomistic calculations.	[90]
PiNN	Python library for building molecule and material artificial neural networks (ANNs).	[91]

To accelerate the application of ML to materials science research, many authors [78,79] have compiled descriptors, ML tools, and DL models (e.g., SchNet, MEGNet, and Pymatgen) in open-source libraries. Table 2 presents some software tools and features available for non-professionals to train crystal material and are useful for predicting numerous materials and optical, electronic, and elastic properties for a large set of materials spaces [80].

5. Descriptors

Materials descriptors are a principal component for mapping input information for ML-application-related materials. In fact, chemical formulae, molecular structures, and many other data forms are not recognized by ML algorithms. Therefore, the data format must be unified and then the formatted data must be transformed into representations describing the input information [78]. These data representations are referred to as "descriptors", "features", or "input variables". Different data forms such as images, text or alphanumeric values that can be appropriately transformed into inputs can be analyzed using GAN algorithms.

Defining of an appropriate representation of a problem depend mainly on data type and algorithms. Therefore, descriptors are obtained by investigating the material aspects that may be correlated with the target property. Certain criteria are required for choosing good materials descriptors [35] which should uniquely and specifically characterize each crystal, correlate with the target property, be easily obtained, and adapt to the molecular size.

Many methods such as "Coulomb Matrices" (CMs), molecular graphs, "smooth-overlap-of-atomic-positions" (SOAP) atomic structures, and "bonds and angles ML" (BAML) have been used for obtaining descriptors for molecular and atomic structures. However, crystal representations must account for the lattice periodicity and additional space group symmetries, which complicates the process of obtaining descriptors. Many descriptors have been proposed for various crystal materials.

The descriptors have different advantages and solve problems from various perspectives. Because this review focused on descriptors trained using only a GAN and IMD, more criteria (e.g., symmetry, invertibility) are required for representing the crystal structure. Therefore, material descriptors have been converted from materials into representations or *vice versa*, which is important for GMs. Using symmetry descriptors, the model was extended to a continuous framework. However, simultaneously meeting all the criteria is exceptionally challenging.

Unlike the inverse design of molecules, that of 3D crystal structures has been rare owing to the challenges in obtaining a continuous representation in the latent space for DL [67]. 3D voxel space was used as a descriptor to predict porous zeolite materials and subsequently the autoencoder methods were provided by the atomic positions [71], and the silicon- and oxygen-atom and methane potential energy grids exhibited dimensions of $32 \times 32 \times 32$ points (see Fig. 3(a)). However,

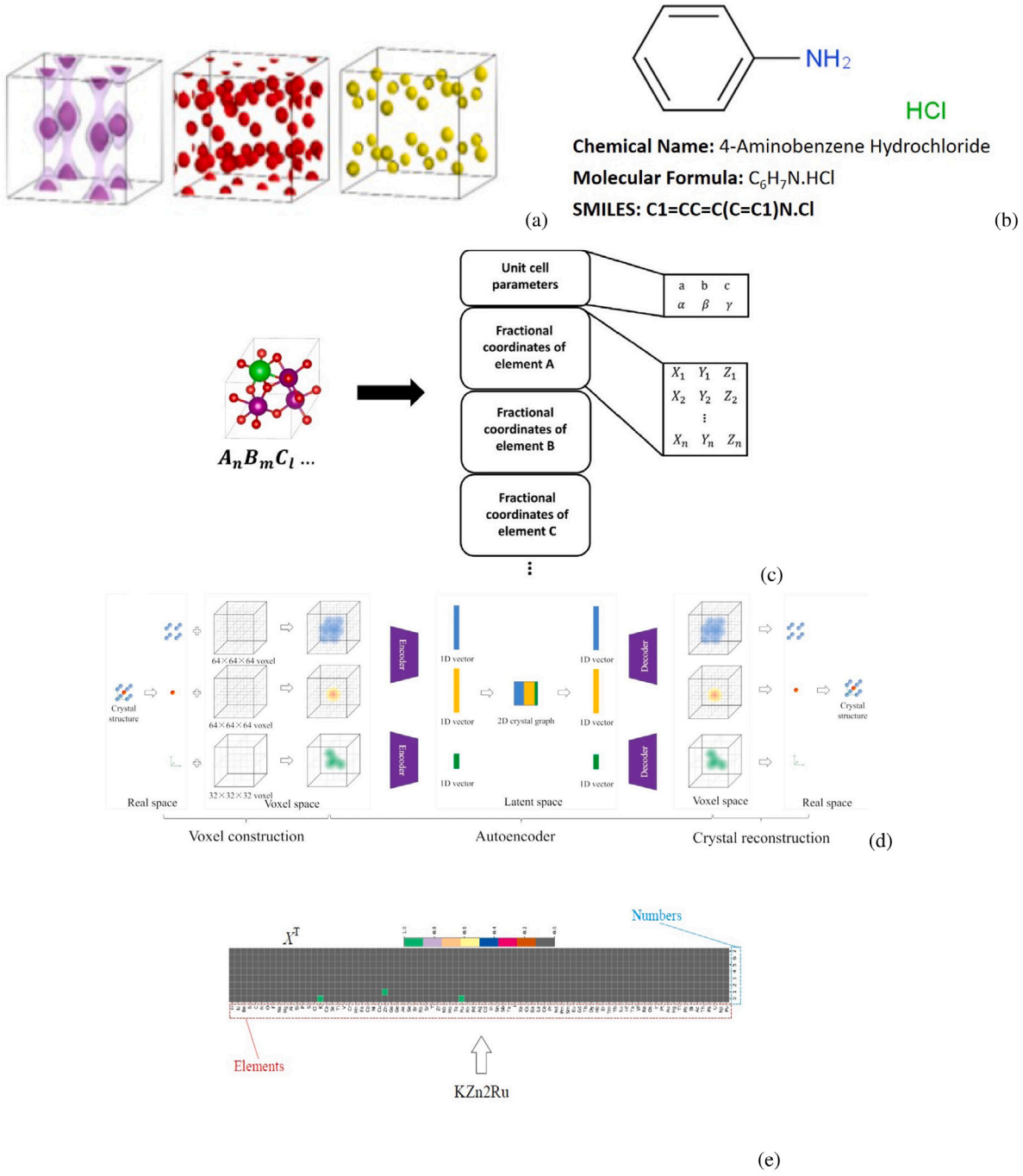

Fig. 3. Examples of various models of material descriptors used in IMD: (a) 3D grid of cubic voxels of zeolitic materials that indicate methane potential energy, silicon, and oxygen atoms in purple, red, and yellow colors respectively (figure reprinted from [71]), (b) SMILES string of the aniline hydrochloride crystal, (c) Matrix representation of ternary material (figure reprinted from [62]), (d) Image representation of 2D crystal graphs (figure reprinted from [67]), and (e) Matrix representation of material compositions of KZn_2Ru (figure reprinted from of [64]).

training 3D data models is computationally expensive because the inputs are complex [92].

Descriptors can convert physical 3D structures into a vector or a notation string to be fed into all types of ML models. A one-dimensional descriptor is the simplest possible molecular or atomic representation. For example, the Simplified Molecular-Input Line-Entry System (SMILES) comprises a sequence-based text representation, as shown in Fig. 3(b) and uses alphabet characters such as “c” and “C” to represent aromatic and aliphatic carbon atoms, respectively; “O” for oxygen atoms; and “−”, “=”, and “#” for single, double, and triple bonds, respectively, to describe molecules [93]. SMILES is a powerful method for obtaining descriptors because SMILES representations are unique; namely, a standard SMILES representation can ensure that each chemical molecule has only one SMILES string, and because SMILES uses a real language structure rather than just a computer data structure, which is naturally advantageous for using ML language models. SMILES descriptors have been widely employed as representations for predicting molecules, particularly for training GAN models to discover pharmaceutical drugs [46].

To avoid the complexity of space-filled volumetric representations, 3D data can be reduced to 2D descriptors. Accordingly, some researchers have used descriptors to represent the lattice parameters,

atomic positions, electron density maps [94,95], and atomic densities [96] of matrix crystals to describe materials.

Kim et al. [62], used a 2D matrix representation of the atomic coordinates and the cell information (see Fig. 3(c)). The descriptors are inspired by the "point cloud" approach in which objects are represented as a set of points and vectors exhibiting 3D coordinates. Long et al. [67] also aim to use 2D representations. They combine voxel grids of atomic positions and lattice constants with crystal graph [80] to predict a 2D crystal graph. The voxel space determining the crystal structure parameters was converted by encoding it into a 2D crystal graph through an autoencoder system [see Fig. 3(d)]. The lattice constants and atomic positions were translated into $32 \times 32 \times 32$ and $64 \times 64 \times 64$ voxel spaces, respectively. The entire process is reversible; that is, a random 2D crystal graph can be reconstructed into a crystal structure in real space, which is essential for GMs. Thus, a continuous latent space was constructed for both representations.

Dan et al. [64] used a sparse matrix consisting of material compositions or formulae as the training data. The rows and columns represented the number of atoms per element in the molecule and number of chemical elements used in the data, respectively. Fig. 3(e) shows the KZn_2Ru encoding matrix wherein each column represents one N atom, whereas the row is a one-hot encoding vector of the number of atoms of a specific element. Combining these representations with the GM (e.g., WGAN) enabled hypothetical materials [97] to be characterized more accurately than using the other molecular com position descriptors (Magpie, One-hot) to characterize the same materials. However, representing a material by its formula or composition has certain limitations. For example, each composition may correspond to multiple structural phases. Additionally, descriptors are limited to simple compounds.

Because previous studies have been limited by material chemical properties or formulae, more information including the elements chemical properties, structural information (e.g., elements, lattice parameters, and site fractional coordinates), and physical properties (e.g., formation energy and band gap) have to be added to the material representation.

6. Challenges and opportunities of GAN applications

ML models require numerous training data, and although engineers have recently made tremendous efforts to design inorganic materials, progress has been limited because discovering materials that meet diverse technical and economic constraints is challenging. Additionally, progress in characterizing material properties has been limited because such materials are complex and, therefore, the material properties are computationally expensive, very specific, and difficult to compute or measure. For example, elastic constants have been computed for less than 10% of the crystals cataloged in the Materials Project database because of the high computational effort required to predict these properties [83].

To tentatively solve the problem of using limited data to predict crystal structures, some researchers [62,63] have used data augmentation, which uses transformations to artificially increase the number of accessible training data [98]. This technique is useful for expanding the dataset size and preventing overfitting. Dan et al. [64] trained GAN models to characterize materials and predict hypothetical materials for input data. Introducing GM-generated databases was the key to predicting various material properties from small experimental material databases, even for databases with missing records, and the hypothetical data were used to train the various models.

Crystal structures representations have been defined by chemical compositions, atomic arrangement [97] and structural information. Although, physical properties and laws should be accounted to truly apply IMD. In fact, defining descriptors that translate the structure–property relationship remains challenging as discussed earlier in this paper. Material descriptors require the existence of various criteria. For the crystal structures representations, symmetry and periodicity have to be investigated. This would be an interesting and challenging research problem. Whereas, there are several options for selecting the cell to express a single crystal structure. Unfortunately, different choices might result in large or even unacceptable errors when encoding crystals exhibiting similar or identical structures.

GANs have interesting application prospects because they can be trained using annotation-free data and can generate realistic images and encourage high-frequency predictions. Nevertheless, training model is characterized by delicate parameters, difficult convergence, and instability. GANs are limited because they generate samples exhibiting little diversity, even when trained using multi-model data, which is called the "mode collapse problem" [99]. Although several recent GAN advances have attempted to solve this problem by using the Wasserstein distance as a loss function, the WGAN may be unstable for large loss-function gradients [100].

Another challenge in image generation is the lack of reliable and consistent metrics, making it difficult to evaluate the most outperforms GANs algorithm. Various metrics based on the predicted material physical properties (charge-neutral samples, formation energies, phonon dispersion calculations, etc.) have been calculated using DFT to evaluate GAN prediction accuracy. To overcome these challenges, researchers have proposed many GAN variants by redesigning the network architecture, changing the objective function form, and altering the optimization algorithms. Additionally, researchers have proposed other solutions such as using variational inference [101] and multiple discriminators to combine VAEs with GANs.

7. Conclusion

In this study, we reviewed promising GANs applications to IMD aimed at predicting functional materials exhibiting desired properties. GMs that are trained and able to discover stable structures can predict more materials, overcoming the limitation of a few experimental materials. Various discovered hypothetical materials have been manipulated as input datasets and have been effectively used to train various ML models. It is necessary to include material properties during IMD; thus, it is challenging to represent them with material compositions and structures. Some researchers have proposed using target properties to find input samples, while others have introduced constraint networks to GMs. Both models have shown impressive material prediction accuracy. However, it is still challenging to directly applied materials predicted by employing GMs, and, consequently, postprocessing is needed.

Finally, we have highlighted the potential of GM applications to accelerate the discovery and design of crystalline materials. Notably, the ML algorithm can effectively predict synthesis pathways based on possible chemical reactions [102]. Future studies should focus on synthesizing these predicted materials with particular properties drawing on the available experimental reaction databases and previous experimental results.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] N.J. Szymanski, Y. Zeng, H. Huo, C.J. Bartel, H. Kim, G. Ceder, Toward autonomous design and synthesis of novel inorganic materials, Mater. Horiz. 8 (2021) 2169–2198, http://dx.doi.org/10.1039/D1MH00495F.
[2] M. Kutz, Handbook of Materials Selection, John Wiley & Sons, 2002.
[3] J.-P. Zhang, Y.-B. Zhang, J.-B. Lin, X.-M. Chen, Metal azolate frameworks: From crystal engineering to functional materials, Chem. Rev. 112 (2) (2012) 1001–1033, http://dx.doi.org/10.1021/cr200139g, PMID: 21939178. arXiv:https://doi.org/10.1021/cr200139g.

[4] K. Rajan, Materials informatics, Mater. Today 8 (10) (2005) 38–45, http://dx.doi.org/10.1016/S1369-7021(05)71123-8.

[5] A.R. Oganov, C.W. Glass, Evolutionary crystal structure prediction as a tool in materials design, J. Phys.: Condens. Matter 20 (6) (2008) 064210, http://dx.doi.org/10.1088/0953-8984/20/6/064210.

[6] J.M. Rondinelli, E. Kioupakis, Predicting and designing optical properties of inorganic materials, Annu. Rev. Mater. Res. 45 (1) (2015) 491–518, http://dx.doi.org/10.1146/annurev-matsci-070214-021150, arXiv:https://doi.org/10.1146/annurev-matsci-070214-021150.

[7] X. Yin, C.E. Gounaris, Search methods for inorganic materials crystal structure prediction, Curr. Opin. Chem. Eng. 35 (2022) 100726, http://dx.doi.org/10.1016/j.coche.2021.100726, URL https://www.sciencedirect.com/science/article/pii/S2211339821000587.

[8] L. Cheng, N.B. Kovachki, M. Welborn, T.F. Miller, Regression clustering for improved accuracy and training costs with molecular-orbital-based machine learning, J. Chem. Theory Comput. 15 (12) (2019) 6668–6677, http://dx.doi.org/10.1021/acs.jctc.9b00884, PMID: 31638804. arXiv:https://doi.org/10.1021/acs.jctc.9b00884.

[9] T. Long, N.M. Fortunato, Y. Zhang, O. Gutfleisch, H. Zhang, An accelerating approach of designing ferromagnetic materials via machine learning modeling of magnetic ground state and curie temperature, Mater. Res. Lett. 9 (4) (2021) 169–174, http://dx.doi.org/10.1080/21663831.2020.1863876, arXiv:https://doi.org/10.1080/21663831.2020.1863876.

[10] S. Chittam, B. Gokaraju, Z. Xu, J. Sankar, K. Roy, Big data mining and classification of intelligent material science data using machine learning, Appl. Sci. 11 (18) (2021) 8596, http://dx.doi.org/10.3390/app11188596.

[11] W.S. Alaloul, A.H. Qureshi, Material classification via machine learning techniques: Construction projects progress monitoring, in: Deep Learning Applications, IntechOpen, 2021.

[12] W. Wang, K. Siau, Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity: A review and research agenda, J. Database Manage. (JDM) 30 (1) (2019) 61–79.

[13] I.H. Sarker, Machine learning: Algorithms, real-world applications and research directions, SN Comput. Sci. 2 (3) (2021) 160, http://dx.doi.org/10.1007/s42979-021-00592-x.

[14] A. Ben Said, A. Erradi, A probabilistic approach for maximizing travel journey WiFi coverage using mobile crowdsourced services, IEEE Access 7 (2019) 82297–82307, http://dx.doi.org/10.1109/ACCESS.2019.2924434.

[15] E. Baccour, A. Erbad, A. Mohamed, M. Hamdi, M. Guizani, RL-DistPrivacy: Privacy-aware distributed deep inference for low latency IoT systems, IEEE Trans. Netw. Sci. Eng. (2022) 1, http://dx.doi.org/10.1109/TNSE.2022.3165472.

[16] M.A. Ahmad, C. Eckert, A. Teredesai, Interpretable machine learning in healthcare, in: Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, in: BCB '18, Association for Computing Machinery, New York, NY, USA, 2018, pp. 559–560, http://dx.doi.org/10.1145/3233547.3233667.

[17] M.A. Elleuch, A.B. Hassena, M. Abdelhedi, F.S. Pinto, Real-time prediction of COVID-19 patients health situations using artificial neural networks and fuzzy interval mathematical modeling, 2021/06/24, Appl. Soft Comput. 110 (2021) 107643, http://dx.doi.org/10.1016/j.asoc.2021.107643, URL https://pubmed.ncbi.nlm.nih.gov/34188610.

[18] K.G. Liakos, P. Busato, D. Moshou, S. Pearson, D. Bochtis, Machine learning in agriculture: A review, Sensors 18 (8) (2018) http://dx.doi.org/10.3390/s18082674, URL https://www.mdpi.com/1424-8220/18/8/2674.

[19] S. Ayadi, A. Ben Said, R. Jabbar, C. Aloulou, A. Chabbouh, A.B. Achballah, Dairy cow rumination detection: A deep learning approach, in: I. Jemili, M. Mosbah (Eds.), Distributed Computing for Emerging Smart Networks, Springer International Publishing, Cham, 2020, pp. 123–139.

[20] M. Abdelhedi, R. Jabbar, T. Mnif, C. Abbes, Prediction of uniaxial compressive strength of carbonate rocks and cement mortar using artificial neural network and multiple linear regressions, Acta Geodyn. Geromater. 17 (3) (2020) 367–378.

[21] M. Abdelhedi, R. Jabbar, T. Mnif, C. Abbes, Ultrasonic velocity as a tool for geotechnical parameters prediction within carbonate rocks aggregates, Arab. J. Geosci. 13 (4) (2020) 180, http://dx.doi.org/10.1007/s12517-020-5070-0.

[22] R. Jabbar, E. Zaidan, A.B. Said, A. Ghofrani, Reshaping smart energy transition: An analysis of human-building interactions in Qatar using machine learning techniques, 2021, CoRR abs/2111.08333. arXiv:2111.08333. URL https://arxiv.org/abs/2111.08333.

[23] E. Zaidan, A. Abulibdeh, A. Alban, R. Jabbar, Motivation, preference, socioeconomic, and building features: New paradigm of analyzing electricity consumption in residential buildings, Build. Environ. 219 (2022) 109177, http://dx.doi.org/10.1016/j.buildenv.2022.109177, URL https://www.sciencedirect.com/science/article/pii/S0360132322004140.

[24] R. Jabbar, M. Shinoy, M. Kharbeche, K. Al-Khalifa, M. Krichen, K. Barkaoui, Driver drowsiness detection model using convolutional neural networks techniques for android application, in: 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), 2020, pp. 237–242, http://dx.doi.org/10.1109/ICIoT48696.2020.9089484.

[25] R. Jabbar, K. Al-Khalifa, M. Kharbeche, W. Alhajyaseen, M. Jafari, S. Jiang, Real-time driver drowsiness detection for android application using deep neural networks techniques, Procedia Comput. Sci. 130 (C) (2018) 400–407, http://dx.doi.org/10.1016/j.procs.2018.04.060.

[26] R. Jabbar, M. Shinoy, M. Kharbeche, K. Al-Khalifa, M. Krichen, K. Barkaoui, Urban traffic monitoring and modeling system: An IoT solution for enhancing road safety, in: 2019 International Conference on Internet of Things, Embedded Systems and Communications (IINTEC), 2019, pp. 13–18, http://dx.doi.org/10.1109/IINTEC48298.2019.9112118.

[27] R. Jabbar, K. Al-Khalifa, M. Kharbeche, W. Alhajyaseen, M. Jafari, S. Jiang, Applied internet of things IoT: Car monitoring system for modeling of road safety and traffic system in the state of Qatar, 2018 (3) (2018). http://dx.doi.org/10.5339/qfarc.2018.ICTPP1072. URL https://www.qscience.com/content/papers/10.5339/qfarc.2018.ICTPP1072.

[28] A.C. Mater, M.L. Coote, Deep learning in chemistry, J. Chem. Inf. Model. 59 (6) (2019) 2545–2559, http://dx.doi.org/10.1021/acs.jcim.9b00266, arXiv:https://doi.org/10.1021/acs.jcim.9b00266.

[29] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, L. Zdeborová, Machine learning and the physical sciences, Rev. Modern Phys. 91 (4) (2019) 045002.

[30] H.S. Stein, J.M. Gregoire, Progress and prospects for accelerating materials science with automated and autonomous workflows, Chem. Sci. 10 (42) (2019) 9640–9649.

[31] K.T. Butler, D.W. Davies, H. Cartwright, O. Isayev, A. Walsh, Machine learning for molecular and materials science, Nature 559 (7715) (2018) 547–555, http://dx.doi.org/10.1038/s41586-018-0337-2.

[32] N.C. Iovanac, R. MacKnight, B.M. Savoie, Actively searching: Inverse design of novel molecules with simultaneously optimized properties, J. Phys. Chem. A 126 (2) (2022) 333–340, http://dx.doi.org/10.1021/acs.jpca.1c08191, PMID: 34985908. arXiv:https://doi.org/10.1021/acs.jpca.1c08191.

[33] S. Lee, H. Byun, M. Cheon, J. Kim, J.H. Lee, Machine learning-based discovery of molecules, crystals, and composites: A perspective review, Korean J. Chem. Eng. 38 (10) (2021) 1971–1982, http://dx.doi.org/10.1007/s11814-021-0869-2.

[34] A.Y.-T. Wang, R.J. Murdock, S.K. Kauwe, A.O. Oliynyk, A. Gurlo, J. Brgoch, K.A. Persson, T.D. Sparks, Machine learning for materials scientists: An introductory guide toward best practices, Chem. Mater. 32 (12) (2020) 4954–4965, http://dx.doi.org/10.1021/acs.chemmater.0c01907, arXiv:https://doi.org/10.1021/acs.chemmater.0c01907.

[35] T. Zhou, Z. Song, K. Sundmacher, Big data creates new opportunities for materials research: A review on methods and applications of machine learning for materials design, Engineering 5 (6) (2019) 1017–1026, http://dx.doi.org/10.1016/j.eng.2019.02.011, URL https://www.sciencedirect.com/science/article/pii/S2095809918313559.

[36] J. Schmidt, M.R. Marques, S. Botti, M.A. Marques, Recent advances and applications of machine learning in solid-state materials science, Npj Comput. Mater. 5 (1) (2019) 1–36.

[37] J. Wei, X. Chu, X.-Y. Sun, K. Xu, H.-X. Deng, J. Chen, Z. Wei, M. Lei, Machine learning in materials science, InfoMat 1 (3) (2019) 338–358, http://dx.doi.org/10.1002/inf2.12028, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/inf2.12028. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/inf2.12028.

[38] J. Wang, Y. Wang, Y. Chen, Inverse design of materials by machine learning, Materials 15 (5) (2022) http://dx.doi.org/10.3390/ma15051811, URL https://www.mdpi.com/1996-1944/15/5/1811.

[39] L. Chen, W. Zhang, Z. Nie, S. Li, F. Pan, Generative models for inverse design of inorganic solid materials, J. Mater. Inform. 1 (1) (2021) 4, http://dx.doi.org/10.20517/jmi.2021.07.

[40] W. Lee, D. Kim, S. Hong, H. Lee, High-fidelity synthesis with disentangled representation, in: A. Vedaldi, H. Bischof, T. Brox, J.-M. Frahm (Eds.), Computer Vision – ECCV 2020, Springer International Publishing, Cham, 2020, pp. 157–174.

[41] Y. Hong, B. Hou, H. Jiang, J. Zhang, Machine learning and artificial neural network accelerated computational discoveries in materials science, WIREs Comput. Mol. Sci. 10 (3) (2020) e1450, http://dx.doi.org/10.1002/wcms.1450, arXiv:https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1450. URL https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcms.1450.

[42] L. Ward, C. Wolverton, Atomistic calculations and materials informatics: A review, Curr. Opin. Solid State Mater. Sci. 21 (3) (2017) 167–176, http://dx.doi.org/10.1016/j.cossms.2016.07.002, Materials Informatics: Insights, Infrastructure, and Methods. URL https://www.sciencedirect.com/science/article/pii/S1359028616301085.

[43] J. Noh, G.H. Gu, S. Kim, Y. Jung, Machine-enabled inverse design of inorganic solid materials: promises and challenges, Chem. Sci. 11 (2020) 4871–4881, http://dx.doi.org/10.1039/D0SC00594K.

[44] O.H. Omar, M. del Cueto, T. Nematiaram, A. Troisi, High-throughput virtual screening for organic electronics: a comparative study of alternative strategies, J. Mater. Chem. C 9 (2021) 13557–13583, http://dx.doi.org/10.1039/D1TC03256A.

[45] J.M.C. Marques, F.B. Pereira, J.L. Llanio-Trujillo, P.E. Abreu, M. Albertí, A. Aguilar, F. Pirani, M. Bartolomei, A global optimization perspective on molecular clusters, Phil. Trans. R. Soc. A 375 (2092) (2017) 20160198, http://dx.doi.org/10.1098/rsta.2016.0198, arXiv:https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.2016.0198. URL https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2016.0198.

[46] R. Gómez-Bombarelli, J.N. Wei, D. Duvenaud, J.M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T.D. Hirzel, R.P. Adams, A. Aspuru-Guzik, Automatic chemical design using a data-driven continuous representation of molecules, ACS Central Sci. 4 (2) (2018) 268–276, http://dx.doi.org/10.1021/acscentsci.7b00572, PMID: 29532027. arXiv:https://doi.org/10.1021/acscentsci.7b00572.

[47] D. Saxena, J. Cao, Generative adversarial networks (GANs): Challenges, solutions, and future directions, ACM Comput. Surv. 54 (3) (2021) http://dx.doi.org/10.1145/3446374.

[48] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, K. Weinberger (Eds.), Advances in Neural Information Processing Systems, Vol. 27, Curran Associates, Inc., 2014, URL https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf.

[49] K. Wang, C. Gou, Y. Duan, Y. Lin, X. Zheng, F.-Y. Wang, Generative adversarial networks: introduction and outlook, IEEE/CAA J. Autom. Sin. 4 (4) (2017) 588–598, http://dx.doi.org/10.1109/JAS.2017.7510583.

[50] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein generative adversarial networks, in: D. Precup, Y.W. Teh (Eds.), Proceedings of the 34th International Conference on Machine Learning, in: Proceedings of Machine Learning Research, vol. 70, PMLR, 2017, pp. 214–223, URL https://proceedings.mlr.press/v70/arjovsky17a.html.

[51] G.-J. Qi, Loss-sensitive generative adversarial networks on Lipschitz densities, Int. J. Comput. Vis. 128 (5) (2020) 1118–1140, http://dx.doi.org/10.1007/s11263-019-01265-2.

[52] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, P. Abbeel, InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets, in: Proceedings of the 30th International Conference on Neural Information Processing Systems, in: NIPS'16, Curran Associates Inc., Red Hook, NY, USA, 2016, pp. 2180–2188.

[53] M. Mirza, S. Osindero, Conditional generative adversarial nets, 2014, arXiv preprint arXiv:1411.1784.

[54] L. Yu, W. Zhang, J. Wang, Y. Yu, SeqGAN: Sequence generative adversarial nets with policy gradient, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 31, 2017, http://dx.doi.org/10.1609/aaai.v31i1.10804, (1). URL https://ojs.aaai.org/index.php/AAAI/article/view/10804.

[55] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2223–2232.

[56] L. Weng, From GAN to WGAN, 2019, http://dx.doi.org/10.48550/ARXIV.1904.08994, URL https://arxiv.org/abs/1904.08994.

[57] Y. Wang, L. Zhou, M. Wang, C. Shao, L. Shi, S. Yang, Z. Zhang, M. Feng, F. Shan, L. Liu, Combination of generative adversarial network and convolutional neural network for automatic subcentimeter pulmonary adenocarcinoma classification, Quant. Imaging Med. Surg. 10 (6) (2020) 1249–1264, URL https://pubmed.ncbi.nlm.nih.gov/32550134.

[58] K. Kim, H. Myung, Autoencoder-combined generative adversarial networks for synthetic image data generation and detection of jellyfish swarm, IEEE Access 6 (2018) 54207–54214, http://dx.doi.org/10.1109/ACCESS.2018.2872025.

[59] A. Mikołajczyk, M. Grochowski, Data augmentation for improving deep learning in image classification problem, in: 2018 International Interdisciplinary PhD Workshop (IIPhDW), 2018, pp. 117–122, http://dx.doi.org/10.1109/IIPHDW.2018.8388338.

[60] A. Nouira, N. Sokolovska, J.-C. Crivello, Crystalgan: learning to discover crystallographic structures with generative adversarial networks, 2018, arXiv preprint arXiv:1810.11203.

[61] T. Kim, M. Cha, H. Kim, J.K. Lee, J. Kim, Learning to discover cross-domain relations with generative adversarial networks, in: D. Precup, Y.W. Teh (Eds.), Proceedings of the 34th International Conference on Machine Learning, in: Proceedings of Machine Learning Research, vol. 70, PMLR, 2017, pp. 1857–1865, URL https://proceedings.mlr.press/v70/kim17a.html.

[62] S. Kim, J. Noh, G.H. Gu, A. Aspuru-Guzik, Y. Jung, Generative adversarial networks for crystal structure prediction, ACS Central Sci. 6 (8) (2020) 1412–1420, http://dx.doi.org/10.1021/acscentsci.0c00426, arXiv:https://doi.org/10.1021/acscentsci.0c00426.

[63] J. Wang, L. Perez, et al., The effectiveness of data augmentation in image classification using deep learning, Convolutional Neural Netw. Vis. Recognit. 11 (2017) 1–8.

[64] Y. Dan, Y. Zhao, X. Li, S. Li, M. Hu, J. Hu, Generative adversarial networks (GAN) based efficient sampling of chemical composition space for inverse design of inorganic materials, Npj Comput. Mater. 6 (1) (2020) 84, http://dx.doi.org/10.1038/s41524-020-00352-0.

[65] R. Xin, E.M.D. Siriwardane, Y. Song, Y. Zhao, S.-Y. Louis, A. Nasiri, J. Hu, Active-learning-based generative design for the discovery of wide-band-gap materials, J. Phys. Chem. C 125 (29) (2021) 16118–16128, http://dx.doi.org/10.1021/acs.jpcc.1c02438, arXiv:https://doi.org/10.1021/acs.jpcc.1c02438.

[66] Y. Song, E.M.D. Siriwardane, Y. Zhao, J. Hu, Computational discovery of new 2D materials using deep learning generative models, ACS Appl. Mater. Interfaces 13 (45) (2021) 53303–53313, http://dx.doi.org/10.1021/acsami.1c01044, PMID: 33985329. arXiv:https://doi.org/10.1021/acsami.1c01044.

[67] T. Long, N.M. Fortunato, I. Opahle, Y. Zhang, I. Samathrakis, C. Shen, O. Gutfleisch, H. Zhang, Constrained crystals deep convolutional generative adversarial network for the inverse design of crystal structures, Npj Comput. Mater. 7 (1) (2021) 66, http://dx.doi.org/10.1038/s41524-021-00526-4.

[68] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, 2015, http://dx.doi.org/10.48550/ARXIV.1511.06434, URL https://arxiv.org/abs/1511.06434.

[69] Y. Zhao, M. Al-Fahdi, M. Hu, E.M.D. Siriwardane, Y. Song, A. Nasiri, J. Hu, High-throughput discovery of novel cubic crystal materials using deep generative neural networks, Adv. Sci. 8 (20) (2021) 2100566, http://dx.doi.org/10.1002/advs.202100566, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/advs.202100566. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/advs.202100566.

[70] Y. Zhao, E. Siriwardane, J. Hu, Physics guided deep learning generative models for crystal materials discovery, 2021, arXiv preprint arXiv:2112.03528.

[71] B. Kim, S. Lee, J. Kim, Inverse design of porous materials using artificial neural networks, Sci. Adv. 6 (1) (2020) eaax9324, http://dx.doi.org/10.1126/sciadv.aax9324, arXiv:https://www.science.org/doi/pdf/10.1126/sciadv.aax9324. URL https://www.science.org/doi/abs/10.1126/sciadv.aax9324.

[72] B. Sanchez-Lengeling, C. Outeiral, G.L. Guimaraes, A. Aspuru-Guzik, Optimizing distributions over molecular space. An objective-reinforced generative adversarial network for inverse-design chemistry (ORGANIC), ChemRxiv (2017) http://dx.doi.org/10.26434/chemrxiv.5309668.v3.

[73] W. Yu, C. Zhu, Y. Tsunooka, W. Huang, Y. Dang, K. Kutsukake, S. Harada, M. Tagawa, T. Ujihara, Geometrical design of a crystal growth system guided by a machine learning algorithm, CrystEngComm 23 (2021) 2695–2702, http://dx.doi.org/10.1039/D1CE00106J.

[74] G. Chen, Z. Shen, A. Iyer, U.F. Ghumman, S. Tang, J. Bi, W. Chen, Y. Li, Machine-learning-assisted de novo design of organic molecules and polymers: Opportunities and challenges, Polymers 12 (1) (2020) http://dx.doi.org/10.3390/polym12010163, URL https://www.mdpi.com/2073-4360/12/1/163.

[75] J.J. de Pablo, N.E. Jackson, M.A. Webb, L.-Q. Chen, J.E. Moore, D. Morgan, R. Jacobs, T. Pollock, D.G. Schlom, E.S. Toberer, J. Analytis, I. Dabo, D.M. DeLongchamp, G.A. Fiete, G.M. Grason, G. Hautier, Y. Mo, K. Rajan, E.J. Reed, E. Rodriguez, V. Stevanovic, J. Suntivich, K. Thornton, J.-C. Zhao, New frontiers for the materials genome initiative, Npj Comput. Mater. 5 (1) (2019) 41, http://dx.doi.org/10.1038/s41524-019-0173-4.

[76] M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L.B. da Silva Santos, P.E. Bourne, J. Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, The FAIR Guiding Principles for scientific data management and stewardship, Sci. Data 3 (1) (2016) 160018, http://dx.doi.org/10.1038/sdata.2016.18.

[77] N. Artrith, K.T. Butler, F.-X. Coudert, S. Han, O. Isayev, A. Jain, A. Walsh, Best practices in machine learning for chemistry, Nature Chem. 13 (6) (2021) 505–508, http://dx.doi.org/10.1038/s41557-021-00716-z.

[78] A. Seko, A. Togo, I. Tanaka, Descriptors for machine learning of materials data, in: Nanoinformatics, Springer, Singapore, 2018, pp. 3–23.

[79] J. Lee, J. Shin, T.-W. Ko, S. Lee, H. Chang, Y. Hyon, Descriptors of atoms and structure information for predicting properties of crystalline materials, Mater. Res. Express 8 (2) (2021) 026302, http://dx.doi.org/10.1088/2053-1591/abe2d5.

[80] T. Xie, J.C. Grossman, Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties, Phys. Rev. Lett. 120 (2018) 145301, http://dx.doi.org/10.1103/PhysRevLett.120.145301, URL https://link.aps.org/doi/10.1103/PhysRevLett.120.145301.

[81] L. Ward, A. Agrawal, A. Choudhary, C. Wolverton, A general-purpose machine learning framework for predicting properties of inorganic materials, Npj Comput. Mater. 2 (1) (2016) 16028, http://dx.doi.org/10.1038/npjcompumats.2016.28.

[82] L. Himanen, M.O. Jäger, E.V. Morooka, F. Federici Canova, Y.S. Ranawat, D.Z. Gao, P. Rinke, A.S. Foster, DScribe: Library of descriptors for machine learning in materials science, Comput. Phys. Comm. 247 (2020) 106949, http://dx.doi.org/10.1016/j.cpc.2019.106949, URL https://www.sciencedirect.com/science/article/pii/S0010465519303042.

[83] C. Chen, W. Ye, Y. Zuo, C. Zheng, S.P. Ong, Graph networks as a universal machine learning framework for molecules and crystals, Chem. Mater. 31 (9) (2019) 3564–3572, http://dx.doi.org/10.1021/acs.chemmater.9b01294, arXiv:https://doi.org/10.1021/acs.chemmater.9b01294.

[84] L. Ward, A. Dunn, A. Faghaninia, N.E. Zimmermann, S. Bajaj, Q. Wang, J. Montoya, J. Chen, K. Bystrom, M. Dylla, K. Chard, M. Asta, K.A. Persson, G.J. Snyder, I. Foster, A. Jain, Matminer: An open source toolkit for materials data mining, Comput. Mater. Sci. 152 (2018) 60–69, http://dx.doi.org/10.1016/j.commatsci.2018.05.018, URL https://www.sciencedirect.com/science/article/pii/S0927025618303252.

[85] M. Haghighatlari, G. Vishwakarma, D. Altarawy, R. Subramanian, B.U. Kota, A. Sonpal, S. Setlur, J. Hachmann, ChemML: A machine learning and informatics program package for the analysis, mining, and modeling of chemical and materials data, WIREs Comput. Mol. Sci. 10 (4) (2020) e1458, http://dx.doi.org/10.1002/wcms.1458, arXiv:https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1458. URL https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcms.1458.

[86] S.P. Ong, W.D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V.L. Chevrier, K.A. Persson, G. Ceder, Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis, Comput. Mater. Sci. 68 (2013) 314–319, http://dx.doi.org/10.1016/j.commatsci.2012.10.028, URL https://www.sciencedirect.com/science/article/pii/S0927025612006295.

[87] C. Draxl, M. Scheffler, NOMAD: The FAIR concept for big data-driven materials science, MRS Bull. 43 (9) (2018) 676–682, http://dx.doi.org/10.1557/mrs.2018.208.

[88] S. Fredericks, K. Parrish, D. Sayre, Q. Zhu, PyXtal: A python library for crystal structure generation and symmetry analysis, Comput. Phys. Comm. 261 (2021) 107810, http://dx.doi.org/10.1016/j.cpc.2020.107810, URL https://www.sciencedirect.com/science/article/pii/S0010465520304057.

[89] K.T. Schütt, P. Kessel, M. Gastegger, K.A. Nicoli, A. Tkatchenko, K.-R. Müller, SchNetPack: A deep learning toolbox for atomistic systems, J. Chem. Theory Comput. 15 (1) (2019) 448–455, http://dx.doi.org/10.1021/acs.jctc.8b00908, arXiv:https://doi.org/10.1021/acs.jctc.8b00908.

[90] A. Khorshidi, A.A. Peterson, Amp: A modular approach to machine learning in atomistic simulations, Comput. Phys. Comm. 207 (2016) 310–324, http://dx.doi.org/10.1016/j.cpc.2016.05.010, URL https://www.sciencedirect.com/science/article/pii/S0010465516301266.

[91] Y. Shao, M. Hellström, P.D. Mitev, L. Knijff, C. Zhang, PiNN: A python library for building atomic neural networks of molecules and materials, J. Chem. Inf. Model. 60 (3) (2020) 1184–1193, http://dx.doi.org/10.1021/acs.jcim.9b00994, PMID: 31935100. arXiv:https://doi.org/10.1021/acs.jcim.9b00994.

[92] T. Hsu, W.K. Epting, H. Kim, H.W. Abernathy, G.A. Hackett, A.D. Rollett, P.A. Salvador, E.A. Holm, Microstructure generation via generative adversarial network for heterogeneous, topologically complex 3D materials, JOM 73 (1) (2021) 90–102, http://dx.doi.org/10.1007/s11837-020-04484-y.

[93] O. Prykhodko, S.V. Johansson, P.-C. Kotsias, J. Arús-Pous, E.J. Bjerrum, O. Engkvist, H. Chen, A de novo molecular generation method using latent vector based generative adversarial network, J. Cheminform. 11 (1) (2019) 74, http://dx.doi.org/10.1186/s13321-019-0397-9.

[94] C.J. Court, B. Yildirim, A. Jain, J.M. Cole, 3-D inorganic crystal structure generation and property prediction via representation learning, J. Chem. Inf. Model. 60 (10) (2020) 4518–4535, http://dx.doi.org/10.1021/acs.jcim.0c00464, PMID: 32866381. arXiv:https://doi.org/10.1021/acs.jcim.0c00464.

[95] J. Noh, J. Kim, H.S. Stein, B. Sanchez-Lengeling, J.M. Gregoire, A. Aspuru-Guzik, Y. Jung, Inverse design of solid-state materials via a continuous representation, Matter 1 (5) (2019) 1370–1384, http://dx.doi.org/10.1016/j.matt.2019.08.017, URL https://www.sciencedirect.com/science/article/pii/S2590238519301754.

[96] J. Hoffmann, L. Maestrati, Y. Sawada, J. Tang, J.M. Sellier, Y. Bengio, Data-driven approach to encoding and decoding 3-d crystal structures, 2019, arXiv preprint arXiv:1909.00949.

[97] T. Hu, H. Song, T. Jiang, S. Li, Learning representations of inorganic materials from generative adversarial networks, Symmetry 12 (11) (2020) http://dx.doi.org/10.3390/sym12111889, URL https://www.mdpi.com/2073-8994/12/11/1889.

[98] M. Bayer, M.-A. Kaufhold, C. Reuter, A survey on data augmentation for text classification, 2021, arXiv preprint arXiv:2107.03158.

[99] G. Cohen, R. Giryes, Generative adversarial networks, 2022, arXiv preprint arXiv:2203.00667.

[100] H. Chen, Challenges and corresponding solutions of generative adversarial networks (GANs): A survey study, J. Phys. Conf. Ser. 1827 (1) (2021) 012066, http://dx.doi.org/10.1088/1742-6596/1827/1/012066.

[101] Y. Sawada, K. Morikawa, M. Fujii, Study of deep generative models for inorganic chemical compositions, 2019, arXiv preprint arXiv:1910.11499.

[102] T.-S. Yao, C.-Y. Tang, M. Yang, K.-J. Zhu, D.-Y. Yan, C.-J. Yi, Z.-L. Feng, H.-C. Lei, C.-H. Li, L. Wang, L. Wang, Y.-G. Shi, Y.-J. Sun, H. Ding, Machine learning to instruct single crystal growth by flux method, Chin. Phys. Lett. 36 (6) (2019) 068101, http://dx.doi.org/10.1088/0256-307x/36/6/068101.