Energy and AI 16 (2024) 100361

Contents lists available at ScienceDirect

Energy and AI

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

Review

Machine learning for CO_2 capture and conversion: A review

Sung Eun Jerng [a,1], Yang Jeong Park [b,c,d,1], Ju Li [b,c,d,*]

[a] Department of Environmental and Energy Engineering, The University of Suwon, 17, Wauan-gil, Bongdam-eup, Hwaseong-si, 18323, Gyeonggi-do, Republic of Korea

[b] Department of Materials Science and Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, 02139, MA, United States of America

[c] Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, 02139, MA, United States of America

[d] MIT-IBM Watson AI Lab, 75 Binney Street, Cambridge, 02142, MA, United States of America

ARTICLE INFO

Keywords:
Machine learning
CO_2 conversion
CO_2 capture
Amine
Ionic liquids
Single-atom alloys
High-entropy catalysts

ABSTRACT

Coupled electrochemical systems for the direct capture and conversion of CO_2 have garnered significant attention owing to their potential to enhance energy- and cost-efficiency by circumventing the amine regeneration step. However, optimizing the coupled system is more challenging than handling separated systems because of its complexity, caused by the incorporation of solvent and heterogeneous catalysts. Nevertheless, the deployment of machine learning can be immensely beneficial, reducing both time and cost owing to its ability to simulate and describe complex systems with numerous parameters involved. In this review, we summarized the machine learning techniques employed in the development of CO_2 capture solvents such as amine and ionic liquids, as well as electrochemical CO_2 conversion catalysts. To optimize a coupled electrochemical system, these two separately developed systems will need to be combined via machine learning techniques in the future.

1. Introduction

Since the industrial revolution, the atmospheric level of carbon dioxide (CO_2) has rapidly accumulated, resulting in climate change and global warming [1]. Indeed, the climate emergency has deteriorated year after year [2]. Recently, the United Nations announced the end of global warming and the start of global boiling [3]. Thousands of scientists are struggling to find a solution to mitigate this meteorological crisis [4]. Among various sustainable energy solutions, carbon capture, utilization, and storage (CCUS) has emerged as a much-needed technology to achieve negative emissions beyond the carbon-neutral energy cycle [5,6]. Solvent-based chemisorption is the basic principle for carbon capture and has been investigated primarily in amine solvents and ionic liquids (ILs) [7,8]. The amine solvent is already considered TRL-9 level and has been widely applied to power plants to capture emitted CO_2 for decades [7]. However, to achieve the goal of limiting the temperature increase to 1.5° from the Paris Climate Agreement, continuous research and development for improving CO_2 conversion efficiency is necessary [9].

The CO_2 capture technologies can be categorized by their reaction mechanism into absorption and adsorption reactions [10]. CO_2 capture by absorption mostly involves applying amines such as monoethanolamine (MEA) [11–13], ILs [14,15], and deep eutectic solvents [16,17], in which CO_2 reacts with the solvent molecules to form chemical bonds. Meanwhile, the adsorption of CO_2 on solid active sites typically occurs in porous materials such as metal–organic frameworks [18,19], zeolites [20,21], and activated carbon [22,23]. The selectivity of CO_2 from flue gas, absorption/adsorption capacity, regeneration rate, and physico-chemical changes of the captured state are important factors, particularly in the design of solvents [24].

There are various CO_2 conversion methods, such as electrochemical [25], photochemical [26], biochemical [27], and thermocatalytic reactions [28], that convert CO_2 into high-value-added products. Electrochemical CO_2 conversion is regarded as the holy grail because clean electricity generated from solar, wind or nuclear energy [29,30] may be utilized. To enhance the desired product selectivity while suppressing the side reactions, a catalyst design that considers the active sites, intermediates, and adsorption energy is imperative [31–33]. Designing a catalyst is difficult because of the cost and challenges associated with experimentally identifying and characterizing catalysts, particularly when dealing with ultra-fast complicated catalytic processes at the atomic level [34,35]. This limits the exploration of a wide range

* Corresponding author at: Department of Materials Science and Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, 02139, MA, United States of America.
E-mail address: liju@mit.edu (J. Li).

[1] The two authors contributed equally to this paper.

https://doi.org/10.1016/j.egyai.2024.100361
Received 7 January 2024; Received in revised form 18 March 2024; Accepted 21 March 2024
Available online 30 March 2024
2666-5468/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

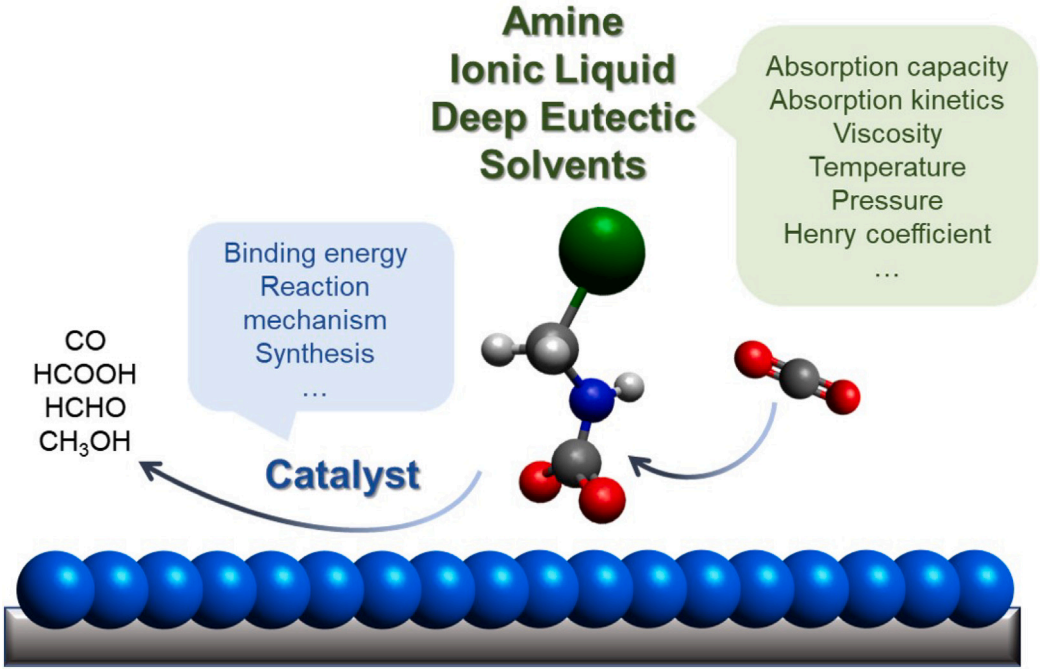

Fig. 1. Descriptors in CO_2 capture and electrochemical conversion reaction.

of catalyst materials and hinders the identification of the optimal catalyst [36]. In addition, to understand the catalysis mechanism in complex electrochemical systems, the detailed reaction steps and intermediate species involved in catalytic reactions require additional investigation [8,37]. In particular, high-entropy alloy (HEA) catalysts, which consist of multiple elements, have complex active centers that present catalytic properties and mechanisms that are challenging to understand [38]. While computational simulations such as density functional theory (DFT) [39] and molecular dynamics (MD) [40] have been implemented to aid in catalyst design, they also have limitations [41, 42]. DFT computations, which are commonly used to study catalyst surfaces, can be computationally expensive and difficult to scale up to simulate large surface areas and complex reactions [42].

Recently, the integration of CO_2 capture and conversion has received much attention owing to its advantage in bypassing the huge energy loss at the amine regeneration step, leading to the mitigation of energy loss and cost [43,44]. In the integrated CO_2 capture and electrochemical conversion reaction, solvents such as amines and ILs capture CO_2 in the reactive state and convert CO_2 to higher value-added products at the electrode. This process can be performed under aqueous [45,46], aqueous–nonaqueous [47], or nonaqueous systems [48, 49], resulting in the generation of different products at the electrode. Enhancing the performance of integrated CO_2 capture and conversion is challenging because of the need to optimize the CO_2 capturing solvent as well as the electrocatalysts. However, until recently, CO_2 capture and conversion have been treated as separate processes, resulting in a lack of studies on the combined systems. Moreover, owing to the complexity of the system, devising solvents and catalysts using traditional methods, such as first-principles calculations or trial-and-error, is time- and energy-consuming.

Machine learning (ML) has been adopted to discover and optimize high-performance CO_2 capture and conversion solvents and catalysts [39,50]. The importance of ML implementation in materials design lies in its ability to accelerate the discovery and optimization of materials [34,51]. ML algorithms can analyze large amounts of data and extract patterns/relationships that may not be readily apparent to humans [52] from descriptors designed to reflect human prior knowledge (Fig. 1). This can help investigators identify active and selective catalysts, optimize reaction conditions, and predict new catalyst formulations [53]. The utilization of ML can lead to a reduction in the time and cost of experimental or computational studies, making catalyst design more efficient [54]. Additionally, ML can assist in the search for high-performance materials with complex compositions and hierarchical structures, which can be challenging using traditional methods [55]. Fig. 2 represents the number of publications for CCUS and ML, indicating a growing research interest from 2001 to 2024.

ML can be used to predict the thermodynamic and electronic properties of materials such as amines [8,56,57], ILs [58–61], and catalysts [62], enabling faster and more efficient screening of materials with excellent performance. ML models can be trained using data obtained from computational methods such as DFT, ReaxFF [63], or kinetic Monte Carlo [50]. These models can then be utilized to perform computer-aided catalyst design and predict properties such as catalyst activity and selectivity [64]. Through the use of ML, researchers can quickly and economically predict the electronic and thermodynamic properties of catalysts, which can accelerate the discovery of innovative and superior catalysts for various applications, including CO_2 reduction reactions (CO_2RR) [65].

ML can also be applied to reveal the mechanisms of catalytic activity and selectivity in CO_2RR, helping to gain insight into the relevant reaction pathways and improve catalyst design [35,41]. For example, the selectivity of catalytic CO_2RR is affected by the binding strength of competitive intermediates, such as *COOH, *OCHO, *CO, and *H [38,66,67]. ML techniques can be employed to investigate the binding strength of these intermediates and tailor the surface conditions of catalysts, providing a promising and efficient approach for exploring CO_2RR selectivity [40,65]. Furthermore, ML can provide useful information for explaining the chemical bonding that binds relevant adsorbates with certain adsorption energies on model surface sites [38]. ML models serve as computationally efficient surrogates, minimizing expensive quantum chemical calculations while enabling the accelerated screening of the catalyst design space.

Herein, we reviewed ML techniques for CO_2 capture solvents (amine and ILs) and catalysts in the integrated CO_2 capture and conversion system. For the ideal case, it is more desirable to simultaneously consider the combined effects of solvents and catalysts. However, owing to the complexity of the system, the reported results have investigated the two different fields separately. First, we summarized the ML techniques for

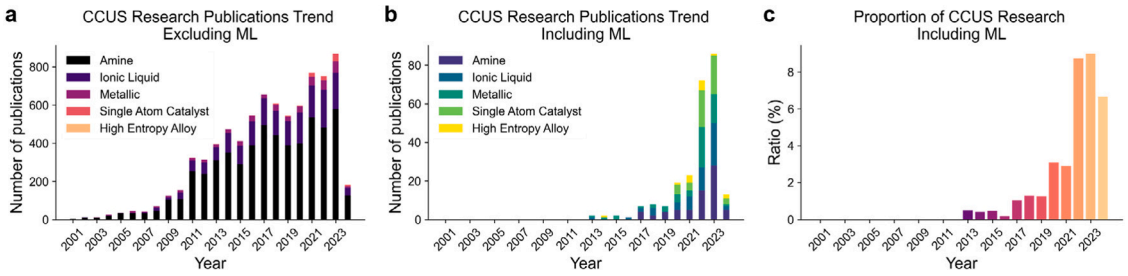

Fig. 2. Research trends of CCUS technology. (a) Number of research articles on traditional CCUS technology according to the keyword. (b) Number of research articles on CCUS technology with ML applications according to the keyword. (c) The ratio of CCUS research incorporated with ML applications.

CO_2 capture and conversion systems; then, we provided examples of ML implementation in the development of amines, ILs, and electrocatalysts. To model a practical integrated CO_2 capture and conversion system, ML will play an important role in uniting these two separate fields in the future.

2. Introduction of machine learning approaches

In this section, we provide an overview of data sources and ML techniques widely used for CO_2 capture and conversion.

2.1. Data sources and databases

The availability of reliable and extensive materials databases is critical to developing ML models to predict material properties and behaviors for enabling advanced materials design processes such as high-throughput screening. Most publicly accessible materials databases depend on DFT-calculated properties and have been developed over the past two decades. One of the earliest projects was the 'Materials Genome Project', a collection of data on lithium-ion battery materials [68], which later evolved into the famous 'Materials Project' [69]. Other notable material databases include Automatic FLOW for Materials Discovery [70], the Open Quantum Materials Database [71], and the Materials Cloud [72]. Several review papers [73–75] have extensively documented publicly available material databases and are highly recommended for perusal.

The other notable dataset is the open catalyst project (OCP) [76,77] released by Fundamental AI Research at Meta AI and Carnegie Mellon University to promote the discovery of new catalysts for renewable energy storage through AI research. It includes 1.3 million molecular relaxations with results from over 260 million DFT calculations. Its ability to perform relaxed (local minimum) energy prediction distinguishes it from other databases focusing on structure–property relationships.

It is noteworthy that computationally inexpensive tools such as xTB [78] and DeepDFT [79] continue to emerge as alternatives to DFT, which was previously widely used to calculate the electronic structure of materials. Additionally, the development of machine learning interatomic potential (MLIP) has enabled large-scale simulations cheaper than existing ab initio molecular dynamics (AIMD), so that the barrier of computational demand of quantum chemistry decreases continuously. Employing these tools can be of great help in building large datasets and training ML models.

Despite the advancement of such large datasets, most research to date in the catalyst design field has been performed based on experimental data collected from the literature. This is because computationally describing behavior is challenging. This complicates the comparison between prospective algorithms and limits the potential of large datasets. Nevertheless, the collection, organization, and standardization of experimental data have the potential to drive significant advances in this field. In particular, recent developments in the field of natural language processing (NLP) automatically extract information from unstructured academic literature [80,81], helping to organize shared knowledge of the community and discover new insights with relatively little effort.

2.2. Supervised learning

Supervised learning is the most common type of ML algorithm, in which a model is trained on labeled data, signifying that each input data point is paired with its corresponding output label. The goal is to learn a mapping function that predicts output labels for unseen test data.

2.2.1. Traditional machine learning

ML algorithms fit labels y_i from input $\mathbf{x}_i$ for datapoint i. Feature engineering is emphasized to improve the performance of ML algorithms and reduce computational costs. Detailed descriptions of these traditional ML algorithms, including multiple linear regression (MLR), random forest (RF), gradient boosting, extreme gradient boosting (XGBoost), support vector machine (SVM), and Gaussian process regression (GPR), are provided in several other reviews [82–84]. Thus, we focus on ML models frequently that have been used in CCUS applications.

(1) Radial basis function neural networks (RBFNNs). An RBFNN is an artificial neural network (ANN) that employs radial basis functions as activation functions. Girosi et al. [85] claim the superiority of RBFNNs by proving the existence and uniqueness of the best approximation. It comprises three layers in a feed-forward architecture. Stochastic gradient descent, in its various forms, is the most common technique used to train the weights and biases of RBF networks, which can be considered as special weights.

(2) Classification and regression tree (CART). The term CART, first introduced by Breiman [86], refers to decision trees (DTs) that are used for either classification or regression. These DTs exhibit some differences, such as the procedure used to determine where splitting occurs. Readers could refer to other materials that explain the details of CART [86,87].

2.2.2. Deep learning

Deep learning (DL) is a subfield of ML that particularly utilizes deep neural networks (DNNs) involving many layers, which are inspired by the human brain. Unlike traditional ML algorithms, the importance of feature engineering is less emphasized in DL, as DNNs can automatically extract meaningful features during the training process. The advancement of DL was led by the proof of the universal approximation theorem for various neural network architectures [88], the backpropagation algorithm [89], which works by propagating the errors backward through the network, and the development of specialized hardware, such as graphics processing units (GPUs).

Given the various applications of DL, it has become a useful tool in materials discovery. For example, convolutional neural networks (CNNs) have been adopted to several characterization tools like X-ray diffraction patterns [90–92], infrared spectra [93,94], Raman spectroscopy [95], and mass spectroscopy [96,97]. Graph neural network (GNN) models [98–101], which can extract semantic information from the relationship between nodes and edges, have also been used to

model structure–property relationship based on local quantum chemical interaction. GNNs have shown promising results in modeling complex interactions and also have gained popularity in other tasks such as chemical reaction prediction [102,103]. These advancements have proven valuable in reducing the computational cost of DFT calculations [79]. In consideration of these advantages, the use of GNNs is promising for catalyst design and discovery.

2.3. Feature selection and explainability

Feature selection is the method of finding an optimal subset of features to improve prediction performance. Wrapper methods follow a greedy algorithm to find optimal combinations of features against evaluation metrics [104]. Ultimately, less important features are removed, allowing researchers to focus on more impactful features. However, greedy algorithms do not guarantee finding the global optimal set of features by their nature. They have a risk of overlooking synergistic feature interactions.

As ML models are often considered black boxes, the need for explainable AI (XAI) has emerged. Analyzing feature importance, which is an example of XAI implementation, may provide additional chemical intuition from the data. Tree-based ML models such as DTs, RF, and XGBoost, can calculate feature importance based on how much each feature contributes to reducing the impurity (e.g., Gini impurity, entropy) when making decisions. The impurity represents the degree of mixing of a target label within a dataset present in a node when splitting the data. The SHAP (SHapley Additive exPlanations) value [105] treats the model as a cooperative game with the same number of players as features and reveals the individual contributions of each player (or feature) to the model's output for each example or observation. However, we must be careful not to ignore the importance of scientific domain knowledge if we focus too much on algorithmic solutions in feature importance analysis.

2.4. Active learning

Collecting and annotating training datasets for ML can be expensive. Active learning (AL) is an ML method that aims to improve model training efficiency while minimizing the number of labeled samples required. AL frameworks select the most informative samples from an unlabeled dataset and label them using an oracle such as a human annotator. There are three main strategies such as membership query synthesis, stream-based selective sampling, and pool-based sampling [106]. AL has drawn research interest [107,108] in materials science, given the computational cost and experimental cost.

As most AL approaches are pool-based, problems are typically defined as follows [109]. There is an initial labeled dataset and an unlabeled dataset. In each iteration, a batch of samples is selected from the unlabeled dataset based on a basic learned model and acquisition function, and the corresponding labels are queried from the oracle. The basic model is retrained on the updated labeled dataset. AL termination strategies can be defined by various pre-defined criteria, for example, when the budget is exhausted or a desired model performance is reached. The design of the acquisition function determines how the data space is explored, and there are several querying strategies. The uncertainty sampling strategy requests samples that have a high prediction uncertainty to improve its understanding of uncertain regions in the data space. However, uncertainty sampling can easily lead to a lack of query sample diversity [106]. We should be aware that DL models tend to be overconfident [110] when combining DL with AL.

2.5. Gaussian process Bayesian optimization

Bayesian optimization and AL are both methodologies that share interconnected concepts that emphasize enhancing model performance predicated on data. Both techniques can leverage concepts akin to uncertainty sampling for labeling or sampling. However, Bayesian optimization predominantly focuses on the optimization of intricate functions, whereas AL prioritizes the selection and labeling of the most informative data points from unlabeled datasets. Gaussian process Bayesian optimization (GPBO) employs Gaussian processes (GPs) to perform modeling and utilizes an acquisition function to determine the subsequent sampling location.

3. Machine learning applications

3.1. Machine learning for CO_2 capture using amine solvent

The technology most widely applied to separate CO_2 from emissions is amine-mediated CO_2 capture [43]. Indeed, amine exhibits a large CO_2 capture capability of approximately 4000 mmol/L, while that of water is approximately 1200 mmol/L [44]. However, the CO_2 absorption kinetics and capabilities vary with the amine molecular structure. Particularly, the number of carbon atoms bonded to the nitrogen atom affects the pKa of the amine site, resulting in greater CO_2 binding strength. In addition, the proportions of carbamate anion, ammonium cation, and (bi)carbonate species in the solvent vary according to various factors such as steric hindrance, temperature, and CO_2 pressure. Thus, the optimization of the amine structure and reaction environment is critical to enhancing the CO_2 capture reaction. However, developing precise algorithms to reflect these factors is challenging owing to their complex nonlinear relationships.

Recently, ML techniques have been extensively developed to predict CO_2 absorption and design amine molecules (see Table 1). Particularly, supervised learning methods such as ANN are widely applied to interpret the thermodynamic properties of CO_2 in amine solvents, such as CO_2 solubility. Representatively, Chen et al. [111] developed back-propagation neural networks (BPNN) and RBFNN, based on the CO_2 solubility values of 12 amine solvents such as MEA, diethanolamine (DEA), and methyldiethanolamine (MDEA). The proposed model demonstrated good agreement with the corresponding experimental values at various amine concentrations, temperatures, and CO_2 partial pressures. Both BPNN and RBFNN models exhibited acceptable average absolute relative errors in terms of the wide input parameter ranges. However, BPNN showed slightly better average absolute relative errors and root-mean-square errors than RBFNN. These results were also compared with eight other numerical models, which highlighted the accuracy of BPNN and RBFNN. Zhang et al. [112] developed the BPNN and general regression neural network (GRNN) to demonstrate that ML can predict the CO_2 solubility, as well as the viscosity and density of solvents. By incorporating 433 datasets from the literature, BPNN showed better predictions for CO_2 solubility and viscosity, while GRNN demonstrated more precise calculations for the density of the potassium lysinate and its mixture. Yarveicy et al. [113] adopted four ML models (ANN, adaptive neuro-fuzzy inference system (ANFIS), least-squares support vector machine (LSSVM), and adaptive boosting in conjunction with the classification and regression tree (AdaBoost-CART) methods) to predict the CO_2 solubility of piperazine in aqueous solvent. Literature values encompassing a wide range of concentrations, temperatures, and pressures were applied to train the models, with AdaBoost-CART showing the most accurate prediction. Ghiasi et al. [114] also used AdaBoostCART model to predict the CO_2 solubility of ethanolamine solvents such as MEA, DEA, and triethanolamine (TEA). The AdaBoost-CART model showed superior results over those of previous neural-based literature models. In addition, an analysis of the feature importance shows that the most influential factor for CO_2 solubility of MEA and TEA was temperature. However,

Table 1
Summary of ML studies in predicting desired properties of amine solvent.

Reference	ML algorithm	Target property	Data source
[111]	BPNN and RBFNN	CO_2 solubility	Literature experimental data
[112]	BPNN, general regression NN	CO_2 solubility, viscosity, density	Literature experimental data
[113]	ANN, ANFIS, LSSVM, and AdaBoost-CART	CO_2 solubility of piperazine	Literature experimental data
[114]	AdaBoost-CART, AdaBoost, NN	CO_2 solubility	Literature experimental data
[115]	BPNN and RBFNN	Density, viscosity, refractive index, heat capacity, thermal conductivity, and thermal diffusivity	in-house experimental data
[116]	BPNN-based model	mass transfer coefficient	Literature experimental data
[117]	XGBoost	adsorption rate	Experimentally verified
[118]	GPBO	regeneration energy	Experimentally verified

unlike MEA and TEA, DEA showed a large dependence on pressure. Pouryousefi et al. [115] demonstrated that ANN-based models could accurately predict a much wider range of physico-chemical properties than empirical models such as Weiland, Nissan–Grunberg, Gladston–Dale, and Redlich–Kister. To predict density, viscosity, refractive index, heat capacity, thermal conductivity, and thermal diffusivity in quaternary systems, such as 4-(diethyl amino)-2-buthanol and MDEA blended with MEA, experimental data were prepared in-house and models were developed based on the BPNN and RBFNN schemes.

The mass transfer coefficient of amines can also be calculated using ML. Dong et al. [116] developed a BPNN-based model to predict the mass transfer coefficient of CO_2 in amine-mediated absorption, using 3935 datasets with 23 amine systems obtained from the literature. Therefore, ML models have shown huge potential in the development of surrogate models to evaluate the physico-chemical properties of amines.

Furthermore, ML has been integrated into a more complex workflow to advance beyond the prediction of physico-chemical properties, ultimately contributing to the design of amines. Orlov et al. [117] developed an ML model framework by integrating kinetic experiments, MD simulations, ML modeling, and large-scale virtual screening to discover a tertiary amine with high CO_2 capture capability as shown in Fig. 4. The RF and XGBoost algorithms were applied to predict quantitative structure–property relationship (QSPR) based on the OPERA [119] and ISIDA Fragmentor [120,121]. Commercially available amines from public datasets were evaluated by the ML models, and some were tested experimentally. Results were verified by experiments, which showed that the CO_2 capture performance of MDEA and 1-ethyl-3-pyrrolidinol was enhanced dramatically when assisted by piperazine.

In addition, the amine regeneration energy can also be predicted using ML. Hwang et al. [118] used GPBO to optimize the operating conditions of a novel water-lean amine solvent for the post-combustion capture of CO_2 capture. A GPR model was adopted as a surrogate model, and the solvent regeneration energies of K_2Sol and MEA were predicted from the gas-to-liquid ratio and stripper pressure. The models helped avoid time-consuming first-principles-based model construction. The results showed that the regeneration energy of K_2Sol was only 2.8 GJ/tCO_2, while that of MEA was 4.3 GJ/tCO_2. K_2Sol also showed superior properties in terms of CO_2 loading, cycling capacity, regeneration temperature, and degradation. These studies demonstrate that the application of an advanced ML framework can expedite the exploration/optimization of amines.

3.2. Machine learning for CO_2 capture in ionic liquids

Another extensively investigated solvent used to capture CO_2 is ILs [50]. ILs are molten salts that are non-volatile liquids composed of organic cations and organic/inorganic anions. ILs have the advantages of high CO_2 absorption, thermal stability, and facile regeneration. The properties of ILs can be tuned depending on the structure and composition of cations and anions [122].

Similar to the application of ML to amines, early studies of ML applied to ILs focused on demonstrating the potential of predictive performance through comparison with traditional models. Zhao et al. [123] predicted IL viscosity using MLR and SVM algorithms with 1502 experimental viscosity datasets from the IL Thermo Database [124], using temperature, pressure, and quantum-chemical descriptors as features (Fig. 5A). By comparing the average absolute relative deviation, the SVM model (Fig. 5B) showed better predictive results than MLR. Venkatraman et al. [125] developed a DT and RF model to characterize the relationship between the structure of ILs and their CO_2 absorption. Over 10,000 CO_2 solubility datasets of 185 ILs from the literature were employed to train the model. Their prediction was more accurate than those from the quantum-chemistry-based COSMOtherm model. Ghiasi et al. [126] applied 5330 experimental CO_2 solubility datasets of 66 ILs to develop a model for predicting the CO_2 solubility of ILs at various temperatures and pressures. By using the CART method, which is considered a white-box model unlike other ML models, the model prediction was more accurate than other pre-published black-box ML methods such as LSSVM, ANFIS, and ANN. Aghaie et al. [127] categorized literature datasets into two categories: thermodynamic variables and molecular structural variables. By utilizing thermodynamic and QSPR ML models, such as LSSVM, DT, RF, and MLR, the most influential input parameters in CO_2 solubility were determined. From their results, the pressure, temperature, and highest-occupied/lowest-unoccupied molecular orbital fraction were critical factors in determining the CO_2 capture capability of ILs. Song et al. [128] used ANN, SVM, and a group contribution (GC) model to predict the CO_2 solubility of ILs. The GC model is a technique used in chemical thermodynamics and chemical engineering to estimate the thermodynamic properties of complex molecules based on the contributions of smaller, well-defined chemical groups or fragments. Through training with 10,116 CO_2 solubility data, the ANNGC model exhibited the most accurate CO_2 solubility prediction (Fig. 5 C and D). Henry's law constants have also been utilized to predict the CO_2 solubility in ILs. Wu et al. [129] applied 160 experimental Henry's law constants of CO_2 in 32 ILs to train the ML models. Among MLP, RF, and MLR, MLP showed the best performance in CO_2 solubility prediction. In addition, the ionic fragment contributions (IFC) based on GC, SVM, and ANN were combined to develop a QSPR model [130]. By training with 13,055 and 415 records of CO_2 and N_2 data, respectively, the IFC-SVM and IFC-ANN models successfully predicted the CO_2 and N_2 solubility of ILs. These approaches could provide guidance on determining promising ILs.

Conductor-like Screening Model for Real Solvents (COSMO-RS) [131,132] is a thermodynamics method based on quantum chemistry to predict chemical potentials in liquids. Taheri et al. [133] developed a screening method using ANN and COSMO-RS to predict the solubility and selectivity of ILs, particularly in CO_2 and H_2S. Henry's law constant can also be calculated by using COSMO-RS theory through DFT calculations of the geometric and electronic properties. Kuroki et al. [134] used the obtained Henry's law constant dataset to train a GPR model.

From the wrapper method, the electronic feature of the ions was found to be the most influential factor in characterizing the CO_2 capture capacity. Based on the model, superior ILs for CO_2 absorption were selected among 402,114 IL candidates. The findings were validated by synthesizing and analyzing [P66614][A] IL candidates.

It is worth noting that the latest deep learning technique utilizing GNN architectures outperformed traditional ML models in CO_2 absorption prediction. Jian et al. [135] compared traditional ML models with several GNN architectures such as graph convolution networks, graph attention networks, and graph isomorphism networks (Fig. 5 E). GNNs directly process molecules as inputs, eliminating the need for costly first-principles calculations to obtain meaningful features. This provides an advantage for large-scale applications and demonstrates the potential for further research.

Furthermore, for the expeditious discovery of high-performing ILs, ML has been applied to model complex workflows. Zhang et al. [136] designed an ML framework that combined multi-player Monte Carlo tree search (MP-MCTS) and recurrent neural network (RNN) to generate and test multiple ILs within a parallel scheme. The MP-MCTS algorithm is used to search the chemical space of ILs by iterative generation and testing of new structures. A tree node corresponds to one symbol of SMILES, and the RNN was trained on 414,972 SMILES to predict the relevant properties. In the selection step, a path from the root to an edge node at the current level was constructed by selecting the children nodes that maximize the upper confidence bound. CO_2 solubility, selectivity, and desorption were considered in the effort to discover novel ILs with high selectivity for CO_2 from flue gas (CO_2/N_2) and syngas (CO_2/H_2). The application of the ML framework, along with real experiments, has proven to be effective in discovering potential IL candidates in actual applications.

3.3. Machine learning for CO_2 capture in deep eutectic solvents

Deep eutectic solvents(DESs) are comprised of hydrogen bond acceptors (HBAs) and hydrogen bond donors (HBDs). DESs have the advantages of low cost, facile preparation, non-volatility, and non-flammability. While the dominant force in the ionic liquids is the ionic interactions between anions and cations, hydrogen bonding is the most influential interaction in the DESs. Comprehensive studies have been conducted on CO_2 capture and separation using DESs. Particularly, theoretical calculations such as the QSPR method and COSMO-RS method [137,138] have been performed to predict the solubility and Henry's constant of CO_2 in DESs. To develop DESs with high CO_2 absorption capacity, high absorption rate, selectivity, and low energy demand, modulating the gas solubility, viscosity, and absorption enthalpy of DESs is imperative. For the physical CO_2 absorption, Henry's constant and the structure of cations/anions affect the absorption capacity. For the chemical CO_2 absorption, the functional groups of the DESs determine the reaction with CO_2. However, due to the strong covalent binding of CO_2, the regeneration energy demand increases as a result.

The goal of most researchers is generally to predict CO_2 solubility by applying ML models. Sagar et al. [139] developed a CO_2 solubility prediction model based on BPNN. The three hydrogen bond acceptors and the lactic acid hydrogen bond donor were analyzed and the structural properties, experimental conditions, and thermodynamic behavior were applied as the input data. The alky chain length and pressure demonstrated a linear relation with the CO_2 solubility while the temperature showed a contrast tendency.

Instead of directly inputting experimental features, COSMO-RS [131,132] is one of the most popular ways to derive a feature. Wang et al. [140] utilized 1011 CO_2 solubility data at various temperatures and pressures to train the model and calculated COSMO-RS derived descriptors via structural parameters generated from Gaussian 09. The QSPR model was developed using Random Forest. The Random Forest combined with COSMO-RS-derived parameters showed R^2 value of 0.9758 and an average absolute relative deviation (AARD) of 7.76%. Lemaoui et al. [141] developed an advanced descriptor to improve performance on the dataset of Wang's group. The molecular descriptors were extracted from COSMO-RS. 94 DES mixtures at 150 compositions, pressures, and temperatures were applied to create 2327 data sets. The developed MLP model showed 0.986 ± 0.002 R^2 value and 4.504 ± 0.507 AARD. Mohan et al. [142] combined COSMO-RS with an artificial neural network for CO_2 solubility prediction. They applied 1973 CO_2 solubility data points in 132 DESs at various DES molar ratios, temperatures, and pressures. The model based on the artificial neural network with COSMO-RS-derived features demonstrated a superior AARD of 2.72% while that of COSMO-RS model was 23.4%. Thus, the machine learning model showed promising tools to design effective DESs for CO_2 capture.

ML models can be combined with various methods to discover promising candidates beyond prediction. Makarov et al. [143] proposed a two-step screening method to discover promising CO_2 absorption DESs. At first, ML models were developed to predict the CO_2 absorption capacity of DESs. The CO_2 absorption capacity data of 162 DESs and 232 ILs were applied for training the model. The transformer convolutional neural fingerprint and the Random Forest with extended connectivity fingerprint showed the best performance. Using SHAP method, the 30 descriptors were determined and found that the temperature and pressure affected the CO_2 absorption capacity most. In the second step, they combined the ML technique with the Redlich–Kister thermodynamic model to calculate the melting temperature. As a result, 1447 DESs were suggested as a promising candidate for high CO_2 absorption capacity at room temperature.

Overall, research to design carbon capture solvent using ML follows a four-stage sequence shown in Fig. 3. Initially, the model takes into account factors like the molecular structure of solvents and environmental parameters (Stage 1). An ML model is constructed to encapsulate the nonlinear interdependencies among these influential factors (Stage 2). The ML model is trained to predict properties of amines and ILs such as solubility, density, viscosity, refractive index, heat capacity, and thermal conductivity (Stage 3). Then optimal amine/IL solvent is identified from candidate pools (Stage 4). By employing ML algorithms with feature engineering, the methodology circumvents the challenging task of first-principles-based model construction.

3.4. Machine learning for CO_2 conversion catalysts

The electrochemical conversion of CO_2 has been widely investigated to generate high value-added products. Some of the reactions are outlined below:

$$CO_2 + 2H^+ + 2e^- \rightarrow CO + H_2O \quad E0 = -0.103 \text{ V } vs \text{ RHE} \tag{1}$$

$$CO_2 + H^+ + 2e^- \rightarrow HCOO^- \quad E0 = -0.225 \text{ V } vs \ RHE \tag{2}$$

$$CO_2 + 8H^+ + 8e^- \rightarrow CH_4 + 2H_2O \quad E0 = +0.169 \text{ V } vs \ RHE \tag{3}$$

$$2CO_2 + 12H^+ + 12e^- \rightarrow C_2H_4 + 4H_2O \quad E0 = +0.079 \text{ V } vs \ RHE \tag{4}$$

$$CO_2 + 6H^+ + 6e^- \rightarrow CH_3OH + 2H_2O \quad E0 = +0.02 \text{ V } vs \ RHE \tag{5}$$

$$2CO_2 + 12H^+ + 12e^- \rightarrow C_2H_5OH + 3H_2O \quad E0 = +0.09 \text{ V } vs \ RHE \tag{6}$$

The reaction mechanism of CO_2RR is intricate owing to the multiple protons and electrons associated with the process. Moreover, in the aqueous system, the hydrogen evolution reaction accompanying CO_2RR leads to a low faradaic efficiency for the desired product. In addition, to devise high-performance catalysts, various parameters such as adsorption/desorption energy, intermediates, CO_2 pressure, and CO_2 temperature need to be determined to reveal the products with high selectivity and faradaic efficiency. Thus, various ML techniques such as neural networks have been utilized to determine effective descriptors and predict catalyst performance.

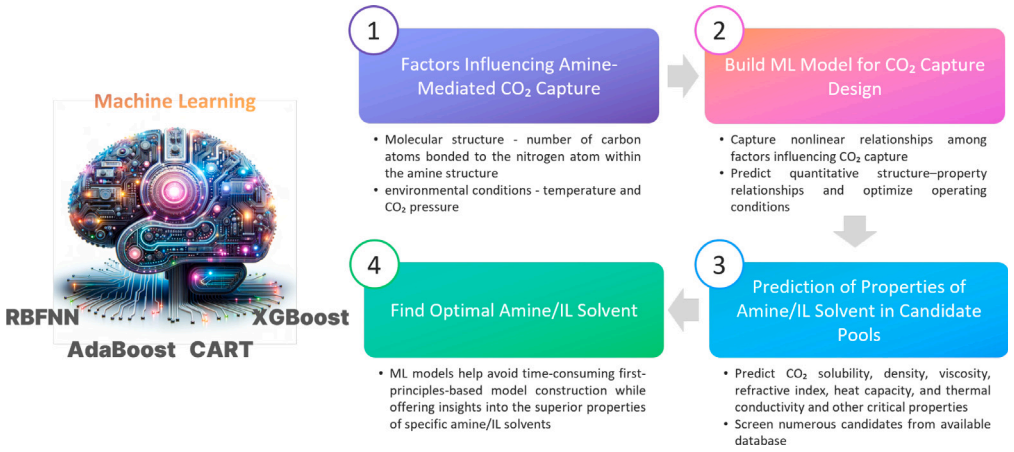

Fig. 3. Schematic overview of machine learning-assisted framework for CO_2 capture with amine solvents and ionic liquids.

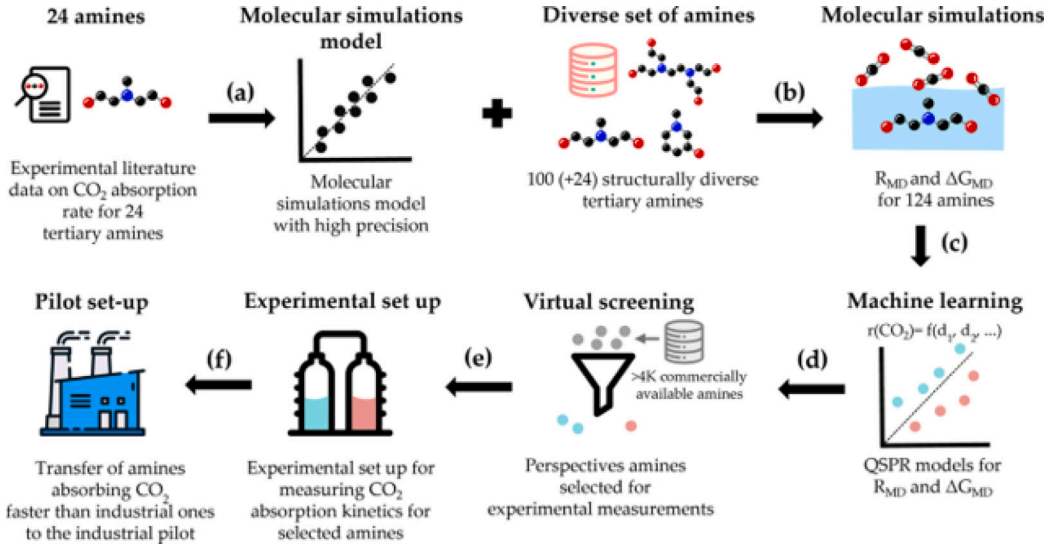

Fig. 4. The workflow of the methodology to predict the amine properties that combines ML and MD simulations.
Source: Reproduced with permission from Ref. [117].
© 2022 Springer Nature Publication.

3.4.1. *Bi-metallic CO_2 conversion catalyst*

The optimization of high-performance bimetallic/multi-metallic materials with atomically precise structures and composition requires time-consuming high-throughput experiments and/or quantum-chemical calculations [36]. However, the implementation of ML models can dramatically enhance time- and cost-efficiency in the prediction of the electrochemical CO_2 conversion performance of the catalyst [64]. Thus, ML models have been developed to identify CO_2 conversion descriptors such as adsorption energy, temperature, and electronegativity [35]. Among the proposed descriptors, the adsorption energy is the most widely investigated in the design of catalysts. In 2015, Ma et al. [146] used an ANN to calculate the CO adsorption energy onto multi-metallic alloy catalysts and proposed 100 high-performance terminated Cu complex catalysts. In addition, Ulissi et al. [144] incorporated an iterative ML model refinement framework with DFT to identify the active sites of Ni/Ga intermetallics (Fig. 6 a). They calculated the CO adsorption energy at various Ni sites and discovered that Ni atoms surrounded by Ga atoms demonstrated the best CO_2 conversion. Noh et al. [147] proposed an AL framework using the d-band width from muffin-tin orbital theory and the electronegativity as descriptors to predict chemisorption energies. Their model achieved high accuracy in predicting the energies of CO adsorption onto alloy systems and identified Cu_3Y@Cu* as a promising electrochemical CO_2 reduction catalyst with a lower overpotential than that of Au catalyst. Tran et al. [145] introduced an AL framework that enables theoretical screening to guide DFT calculations for the discovery of effective electrocatalysts for CO_2 reduction and H_2 evolution (Fig. 6 b). In this study, 131 candidate surfaces across various alloys of 31 different elements were evaluated to highlight potential candidates with high performance.

In addition, ML, multiscale simulations, and quantum mechanics (QM) were applied to identify the active sites of Au nanoparticles and dealloyed Au_3Fe core–shell nanoparticles [148]. The CO adsorption energy and HOCO formation energy at the Surface Defect, Step111, StepUnder111, Step110, Step311, Step TB (twin boundary), and StepUnder TB were calculated. The predicted performance (a-value) was visualized by mapping, illustrating that the Surface Defect and StepUnder 111 sites were the most critical sites in catalysis. In addition, Cu–Al alloy electrocatalyst was discovered as a promising CO_2 reducing catalyst by calculating the CO and H adsorption energies at the catalytic active sites [149]. From the experimental verification, the favorable Cu coordination resulted in a high faradaic efficiency of > 80% in ethylene production. Xing et al. [150] developed a GBR-based algorithm, named WLEDZ, to explore alloy catalysts for CO_2RR. The model reduced the dimensionality of the feature space from 13 to 5 features using the Pearson correlation matrix, yielding the following features: work function, local electronegativity, electronegativity,

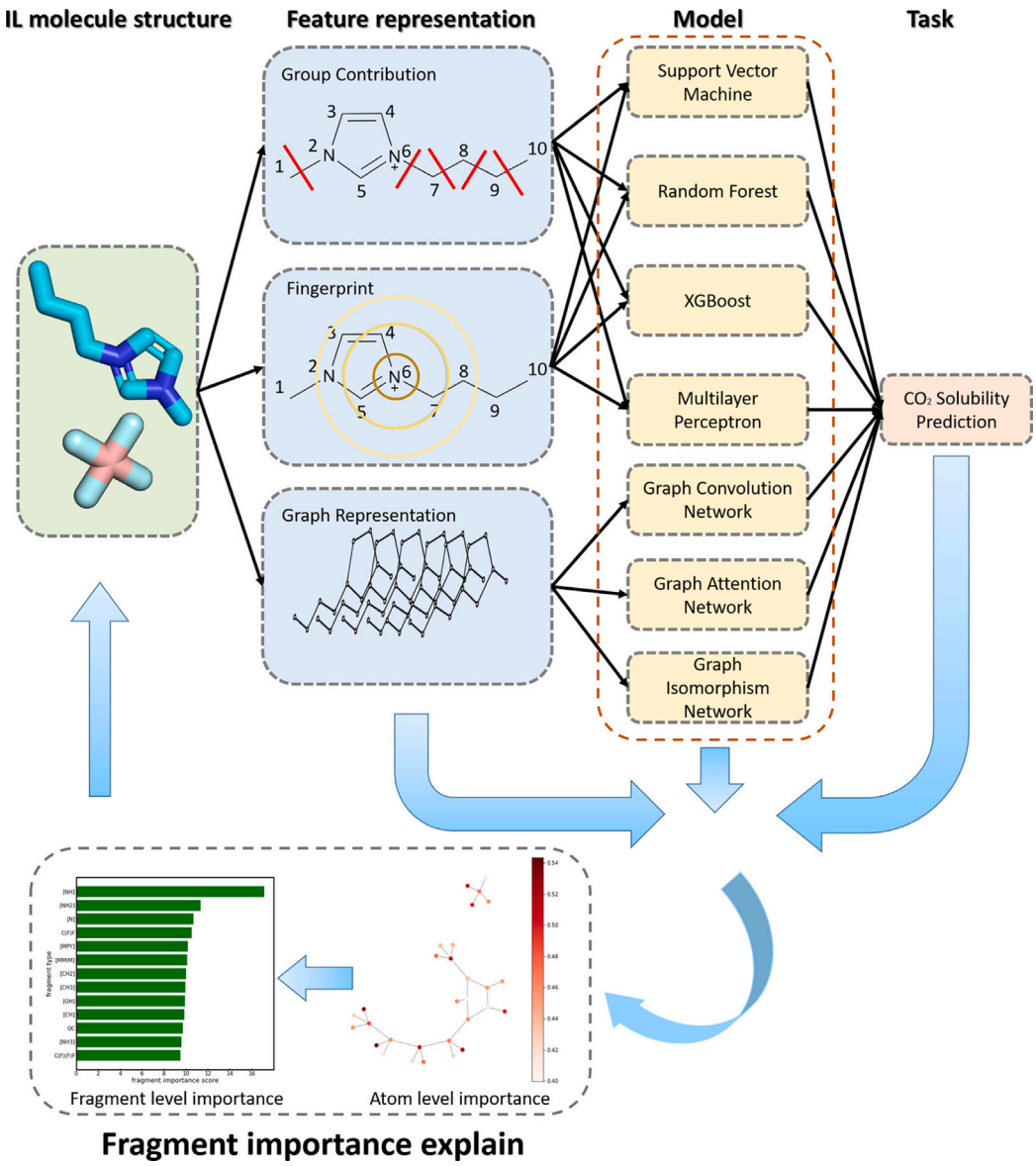

Fig. 5. A framework that employs both shallow ML models and deep learning models based on descriptors to predict CO_2 solubility. An explainable GNN is used for atom-level and fragment-level importance analysis in IL molecule pairs.
Source: Reproduced with permission from Ref. [135].
© 2022 American Chemical Society Publications.

interplanar spacing, and atomic number. The reduced-feature model performed just as well as models with a greater number of features. In addition, the model successfully predicted the adsorption energy of key intermediates in the CO_2RR process and identified 8 highly promising Cu-based bimetallic catalysts. The predictions made by the ML model were further supported by DFT calculations. Feng et al. [151] utilized a gradient boosting regression (GBR) model working with 17 features to investigate the reaction activity and selectivity of 1218 dual-metal-site catalysts. Their ML model identified Mn–Ru, Mn–Os, Zn–Ru, and Co–Au–N6-Gra-model 3 as promising candidates. In addition, the limiting potential of transition metal phthalocyanine dual-metal-site catalysts was successfully predicted via a DFT-based ML-accelerated (DFT-ML) method [152]. From the GBR algorithm, Ag–MoPc was predicted as a promising electrocatalyst with a limiting potential of −0.33 V. As evidenced by the studies highlighted above, ML models have successfully identified novel catalysts, revealed complex reaction mechanisms, and attained high predictive performance. The synergy between ML and experimental validation holds promise for the accelerated discovery of next-generation CO_2 reduction catalysts. As the field continues to evolve, ML is anticipated to play an even more central role in developing and optimizing electrochemical CO_2 conversion processes.

3.4.2. CO_2RR single-atom catalyst

To design single-atom catalysts (SACs), the employment of QM calculations to describe microscopic atomic behavior is imperative. In contrast to the preceding sections, ML applications are centered on mitigating the computational expenses associated with QM calculations, which are typically resource-intensive, rather than addressing the explosion of chemical compound combinations.

Zhu et al. [154] predicted reaction intermediates and products in CO_2RR using SACs in zeolites. XGBoost, GBR, and ExtraTrees models demonstrated satisfactory prediction performance of the changes in free energy (ΔG) for all possible intermediates and product probabilities (P) of CO_2 reduction to methanol, methane, and formaldehyde on 26 SACs in zeolites, using descriptors related to reaction pathways, metals, charge transfer, hydrogen-bond interactions, and geometry. The CO adsorption energy of atomically dispersed metal–nonmetal

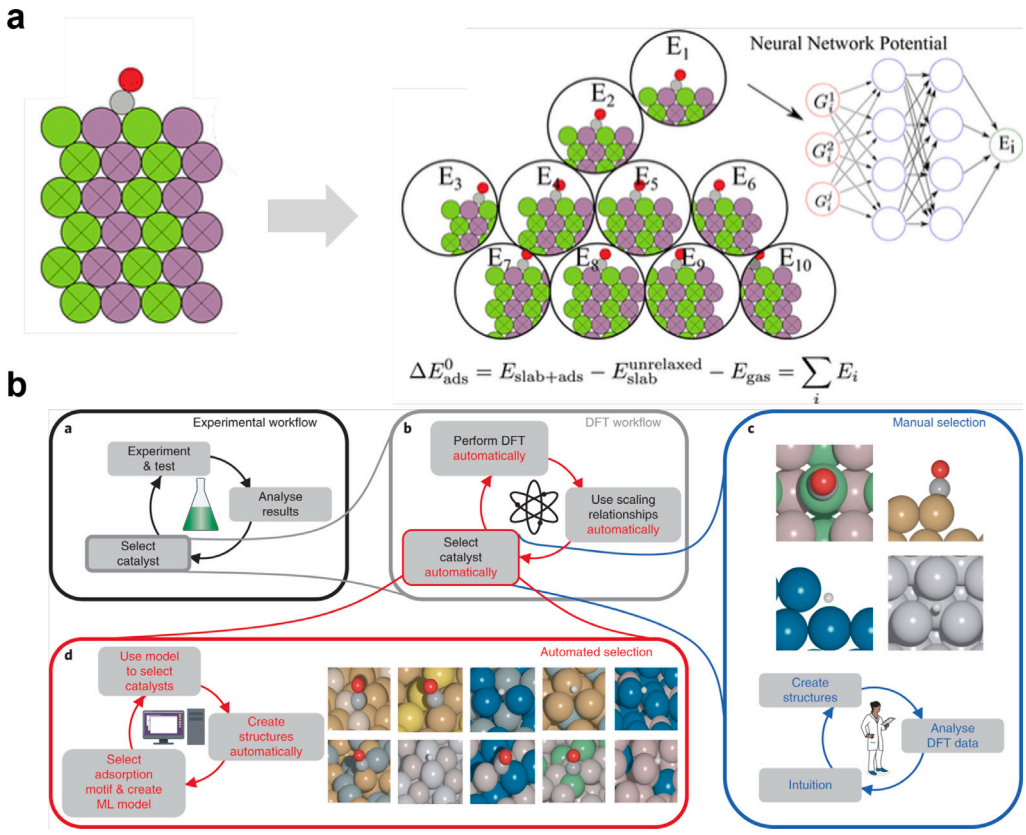

Fig. 6. (a) An ordinary method employing a neural network to predict adsorption energy on a typical bimetallic surface with a CO adsorbate. Reproduced with permission from Ref. [144], © 2017, American Chemical Society Publications. (b) The active learning workflow that combines ML and DFT with experimental catalyst discovery. Unlike traditional workflows (depicted in blue) that rely on scientific intuition for candidate selection, this proposed workflow operates automatically. Reproduced with permission from Ref. [145], © 2018, Springer Nature Publication. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

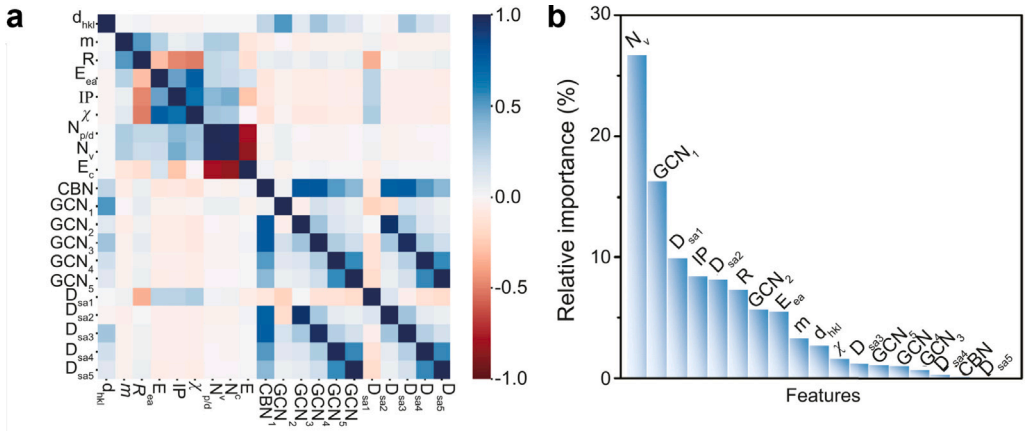

Fig. 7. (a) Heat map displaying feature correlations using Pearson correlation coefficient analysis. (b) Ranking of feature importance based on mean impact factor analysis. *Source:* Reproduced with permission from Ref. [153].
© 2023 Elsevier.

co-doped graphene was calculated using extreme gradient boosting regression [155]. Among the 1060 single-atom catalysts, with 20 transition metal atoms and 53 nonmetal bonded environments, 94 potential single-atom catalysts were suggested as promising CO_2RR electrocatalysts. Wang et al. [153] used high-throughput DFT calculations and ML techniques to screen the optimal active sites of Cu-based single-atom alloy catalysts among 2669 configurations for CO_2RR by predicting the energies of CO adsorption onto Cu-based single-atom alloy catalysts (Fig. 7). They identified ZnCu, AgCu, GaCu, GeCu, and PCu single-atom alloy catalysts as promising CO_2RR catalysts.

3.4.3. CO_2RR high entropy catalysts

In addition, ML techniques have been widely applied to complex systems such as HEAs, which are composed of five or more elements. Chen et al. [157] predicted the electrocatalytic performance of the FeCoNiCuMo HEA catalyst by calculating the adsorption energies of COOH*, CO*, and CHO* from DFT and NN models. Pedersen et al. [158] presented a probabilistic and unbiased approach by combining DFT and a supervised GPR model to predict the CO and H adsorption energies on the (111) surfaces of the disordered CoCuGaNiZn and AgAuCuPdPt alloys. By optimizing the compositions of these alloys to increase the likelihood of sites with weak H adsorption

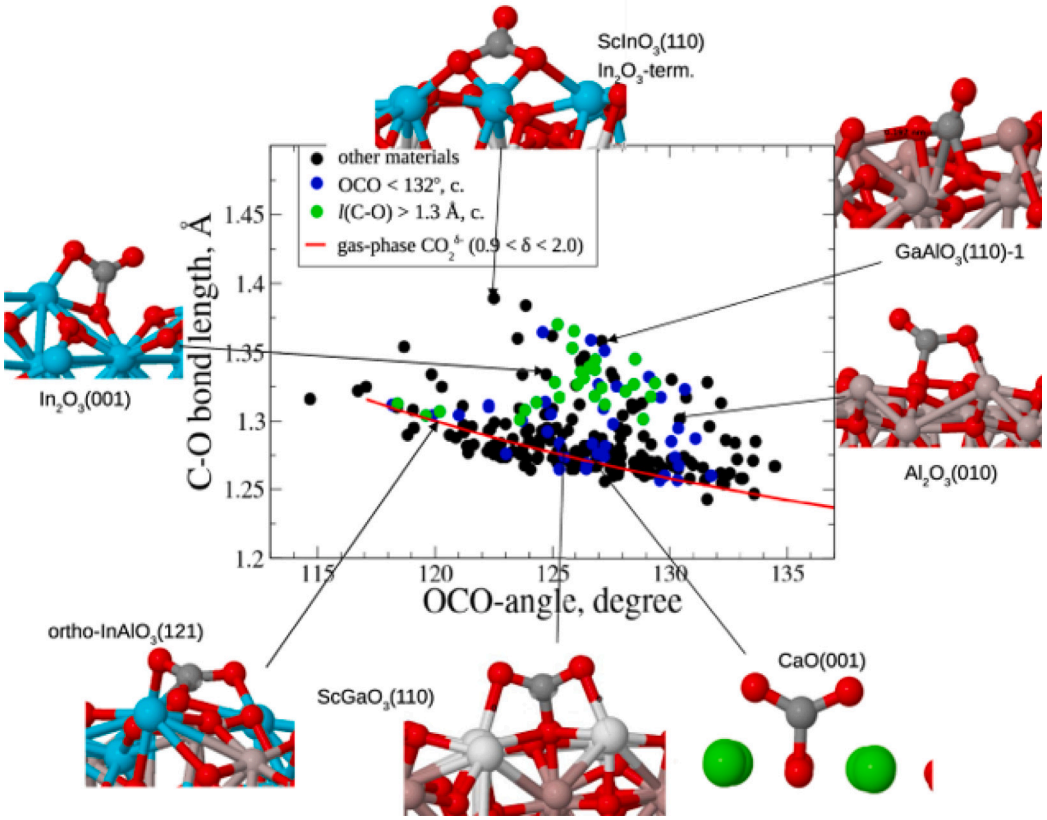

Fig. 8. The OCO-angle in charged gas-phase CO_2 is indicated with the red line, and adsorbed CO_2 structures are depicted with colored dots representing different categories: blue for adsorption sites from the unconstrained subgroup with OCO < 132°, green for the subgroup of sites with l(C–O) > 1.30 Å, and black for the remaining samples. Subgroups obtained with the Sabatier principle constraint are marked with "c". (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
Source: Reproduced with permission from Ref. [156].
© 2022 Springer Nature Publications.

and strong CO adsorption, the aim was to enhance the selectivity for CO reduction and H suppression. Their approach focused on increasing the probability of desired surface sites, rather than constructing specific arrangements of atoms. They identified several promising catalysts that meet the criteria of weak H adsorption and strong CO adsorption, some of which were validated in existing literature. Additionally, they demonstrated the model's ability to suggest GaNi as a locally optimal catalyst candidate for CO_2 and CO reduction, even without prior knowledge of its catalytic properties. This study was supported by an independent experimental realization of the CO_2 reduction reaction [159]. Mazheika et al. [156] identified specific catalyst genes (features) that correlate with the activation of CO_2 for chemical conversion. By combining the subgroup discovery model and first-principles data for various oxides, they found that the surfaces of known effective catalysts consistently exhibit combinations of genes that result in the elongation of the C–O bond, indicating CO_2 activation. They proposed a new indicator for CO_2 activation, a large C–O bond distance, which automatically satisfied the Sabatier principle for low- and medium-temperature CO_2 conversion. The subgroup discovery approach identified the surface properties that strongly correlate with CO_2 activation, including the charges of oxygen atoms and surface cations, surface geometric features, electrostatic potential, polarizability, and other descriptors. They also showed that previous indicators, such as a decrease in the OCO angle, are not suitable for CO_2 activation and can lead to carbonation poisoning. Based on their findings, they also proposed several promising oxide-based catalysts for the CO_2 conversion reaction (Fig. 8) .

3.4.4. *Metal–organic framework*

Metal–organic frameworks (MOFs) are porous materials in which metal ions are linked via organic linkers to form multi-dimensional structures. Owing to their large surface area, MOFs are widely applied for CO_2 capture and utilization. Tons of computational simulations are required to discover innovative organic linkers and structures that can modulate their performances and properties. ML offers a circumvention that can reduce computational costs.

ML using feature engineering can shorten the time to predict performance by integrating chemical domain knowledge. Dureckova et al. [161] developed a model based on gradient boosted trees regression method using 358,400 MOFs and 1166 network topologies data sets. The QSPR models and atomic property weighted radial distribution functions (AP-RDFs) were applied to discover descriptors. Their model showed R^2 of CO_2 working capacities as 0.944 and that of CO_2/H_2 selectivities as 0.872. Their model could ensure 1000 MOFs with high performance. Orhan et al. [162] suggested Effective Point Charge (EPoCh) descriptors that quantify the effect of MOFs' partial charges for predicting CO_2 adsorption. Though the Henry coefficient was the most influential descriptor, the EPoCh descriptors were demonstrated as the second most important. EPoCh descriptors calculations were hundreds of thousands of times faster than that of the Henry coefficient. Given that the number of hypothetical MOFs grows intensively, this computation reduction is a meaningful factor.

The application of DL automatically extracts features, eliminating the need for feature engineering that requires expert domain knowledge. Yang et al. [163] reported a DL model to predict the CO_2, H_2O, and CO adsorption in MOFs as well as Henry coefficients. As a result, the Monte Carlo simulations could be accelerated by one

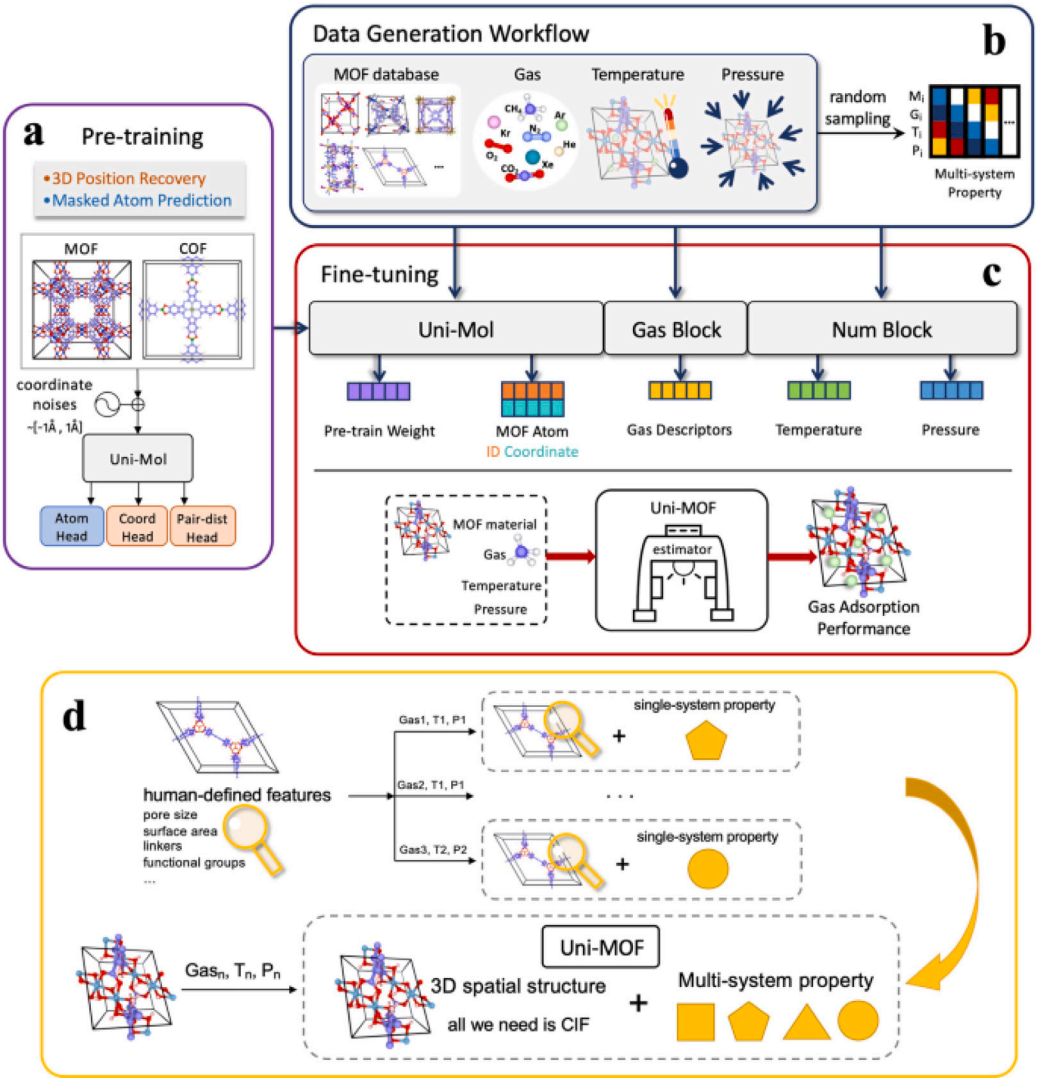

Fig. 9. The proposed Uni-MOF framework does not require additional analytical calculations for materials. It predicts properties solely based on the crystallographic information file (CIF) of MOF materials. (a) Pre-training workflow to learn 3D spatial representation by predicting the types of masked atoms and denoising 3D coordinates. (b) Generating cross-system performance datasets using random sampling of operating conditions. (c) Uni-MOF for unified gas adsorption prediction was fine-tuned from the embedding of the pre-trained model. (d) Overall workflow of Uni-MOF.
Source: Reproduced with permission from Ref. [160].
© 2024 Springer Nature Publications.

hundred times. Wang et al. [160] developed Uni-MOF which is aimed at predicting various gas adsorption performances of MOFs (Fig. 9). Over 631,000 three-dimensional MOF and COF sets were applied to train the model so that Uni-MOF can present structural demonstrations as well as gas adsorption capacity predictions at various ambiances. By leveraging unlabeled data using pretraining, this transformer-based approach can remove bias caused by feature engineering.

Simulation using MLIP provides physical and chemical explanations to humans instead of black box ML. Zheng et al. [164] predicted the CO_2 chemisorption of MOFs by developing quantum-informed machine learning force fields (QMLFFs). QMLFFs showed over a thousand times faster calculations than first-principle simulations. Using QMLFF-based molecular dynamics, the diffusion coefficient and the binding free energy landscape of CO_2 chemisorption in Mg-MOF-74 well matched with that of experimental values.

4. Outlook

In this section, we suggest a few recently developed AI technologies such as self-supervised learning, deep generative models, and large language models which have not been widely dealt with in the CCUS domain yet.

4.1. Self-supervised learning

Self-supervised learning (SSL) is proposed to overcome the scarcity of labeled data and to extract the full potential of large unlabeled data. The basic idea is that ML models are trained on target labels which can be automatically generated inherently to extract useful features. Afterward, extracted features can be applied to various downstream tasks by using fine-tuning or transfer learning. For example, contrastive learning is one of the most successful SSL approaches. The core idea is to pull together representations of positive pairs, which means similar samples while pushing apart representations of negative pairs, which means dissimilar samples in an embedding space.

Several studies about SSL were conducted in the materials science domain. Wang et al. [165] disclosed a model called MolCLR, which is trained by applying contrastive learning on 2D graph representations of 10 M molecules. Fang et al. [166] proposed an SSL method to predict

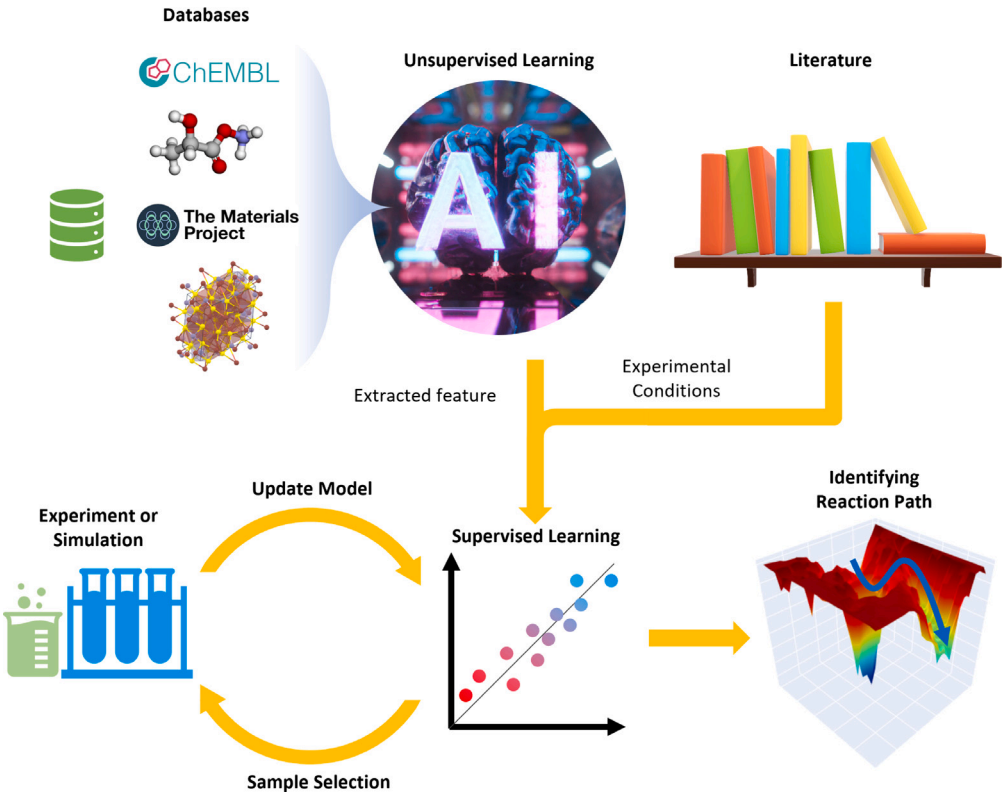

Fig. 10. The proposed framework leveraging automatically extracted features using self-supervised learning. By combining general knowledge about molecules and crystals with experimental conditions, supervised ML models can be developed effectively. This model can guide exploration in the real world using active learning. Accelerated simulation using machine learning interatomic potential can offer chemical and physical insight by identifying reaction paths.

local spatial structure and global spatial structure of 3D graph representations of 20 M molecules obtained from Merck molecular force field [167]. Ahmad et al. [168] adopted RoBERTa [169] pretraining strategy to understand SMILES representation of 77 M molecules. Koker et al. [170] applied contrastive learning to crystalline materials to introduce CrystalCLR. Magar et al. [171] extended the Barlow Twins [172] approach which is popular in the computer vision domain to crystalline materials domain. These models mentioned are publicly available. Using extracted knowledge in these pretraining processes has the potential to be more efficient than developing new features based on traditional chemical domain knowledge or quantum mechanical calculations, but this approach has not been sufficiently explored (Fig. 10). In particular, it can bring out the full potential of ML models in the CCUS field, which relies on a limited number of training data.

4.2. Deep generative models

Generative models focus on learning the probability distribution of the data to generate new samples that resemble the original data. Once the model has learned the data distribution, it can generate new samples by sampling from the learned distribution. Several generative models were explored, including generative adversarial networks (GANs), variational autoencoders (VAEs), normalizing flows, energy-based models, and autoregressive models [173]. These generative models have been widely applied to molecules [174–178] and crystals [179,180]. Generative models offer potential opportunities as they may discover completely new materials that have never been explored before at a much lower computational cost than traditional high throughput computation. However, synthesizing these new materials, applying them to the real process, and optimizing the process remains constrained by experimental costs. Therefore, further research is needed on sophisticated sampling and filtering strategies to select promising candidates.

4.3. Large language models

Large language models (LLMs) have recently attracted attention for their diverse capabilities. Text representations such as IUPAC names or SMILES can be spontaneously integrated with natural language text, so there are many possibilities for integration with text-written chemical domain knowledge. Guo et al. [181] benchmarked LLM's abilities such as understanding, reasoning, and explaining in eight chemistry tasks. In particular, Jablonka et al. [182] demonstrated the possibility of an inverse design of a molecule with a desired bandgap by fine-tuning ChatGPT with a limited number of data. By aligning knowledge of molecular and crystal structures with academic literature, new materials can be discovered and experimental processes can be optimized. However, LLMs are known to suffer from the hallucination problem, which can result in incorrect or fabricated information [183]. Research is required on various tasks that integrate chemical domain knowledge to predict, explain, or suggest experimental details while alleviating hallucinations.

5. Conclusion

The pressing urgency of climate change, driven by anthropogenic CO_2 emissions, has led to a global effort toward the development of advanced technologies, including ML applications to efficiently capture emitted carbon. ML techniques contributed CCUS community by combining with quantum mechanical simulations to pave the way for insights into catalyst behavior and dynamics to be generated that were previously unachievable or highly time-consuming. They have also enabled the rapid screening of thousands of candidate materials, drastically reducing the time from conceptualization to realization. However, challenges still remain to maximize the full capabilities of AI.

The first issue is the establishment of benchmarks for AI applications in CCUS to promote progress in this field by standardizing future research. Despite the appearance of extensive databases of molecules and solids, a method for harnessing their potential has not been examined in this field. Extracting meaningful features from huge databases and standardizing fragmented literature datasets can incorporate methods from other fields, resulting in more sophisticated ML prediction models.

The second issue is the development of strategies for exploring undiscovered chemical space. The combined framework between conditional generative models and MLIP simulations may play predominant roles in *in silico* design of carbon capture and conversion materials.

The third issue is the enhancement of the interpretability of ML models. In particular, previous research has focused dominantly on demonstrating the power of ML models as black-box models. In contrast, uncovering reaction pathways can augment the explanatory power of ML models and provide insights into chemical mechanisms. Further study on simulation-combined ML frameworks can facilitate an understanding of the underlying dynamics, leading to advancement toward the goal of the research community.

CRediT authorship contribution statement

Sung Eun Jerng: Writing – original draft, Visualization, Validation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Yang Jeong Park:** Writing – original draft, Visualization, Project administration, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Ju Li:** Writing – original draft, Validation, Supervision.

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.

Data availability

No data was used for the research described in the article.

Acknowledgments

This work was supported by a grant from the National Research Foundation of Korea (NRF) funded by the Korean government, Ministry of Science and ICT (MSIT) (No. 2021R1A6A3A01086766, 2021R1A6A3A03044878). This research was also supported by a grant (code 2023-006) from Gyeonggi Technology Development Program funded by Gyeonggi Province, Republic of Korea. This study was supported by the research grant of The University of Suwon, Republic of Korea in 2023 (2023-0166).

References

[1] Mac Dowell N, Fennell PS, Shah N, Maitland GC. The role of CO_2 capture and utilization in mitigating climate change. Nat Climate Change 2017;7(4):243–9.
[2] Peters GP, Andrew RM, Canadell JG, Friedlingstein P, Jackson RB, Korsbakken JI, et al. Carbon dioxide emissions continue to grow amidst slowly emerging climate policies. Nature Clim Change 2020;10(1):3–6.
[3] Türk V. World view. Nature 2023;623:9.
[4] Chen H, Zheng Y, Li J, Li L, Wang X. AI for nanomaterials development in clean energy and carbon capture, utilization and storage (CCUS). ACS Nano 2023.
[5] Navarro-Jaén S, Virginie M, Bonin J, Robert M, Wojcieszak R, Khodakov AY. Highlights and challenges in the selective reduction of carbon dioxide to methanol. Nat Rev Chem 2021;5(8):564–79.
[6] Sharifian R, Wagterveld R, Digdaya I, Xiang C, Vermaas D. Electrochemical carbon dioxide capture to close the carbon cycle. Energy Environ Sci 2021;14(2):781–814.
[7] Bui M, Adjiman CS, Bardow A, Anthony EJ, Boston A, Brown S, et al. Carbon capture and storage (CCS): the way forward. Energy Environ Sci 2018;11(5):1062–176.
[8] Gupta S, Li L. The potential of machine learning for enhancing CO_2 sequestration, storage, transportation, and utilization-based processes: a brief perspective. JOM 2022;74(2):414–28.
[9] Segets D, Andronescu C, Apfel U-P. Accelerating CO_2 electrochemical conversion towards industrial implementation. Nat Commun 2023;14(1):7950.
[10] Yu C-H, Huang C-H, Tan C-S, et al. A review of CO_2 capture by absorption and adsorption. Aerosol Air Qual Res 2012;12(5):745–69.
[11] Chai SYW, Ngu LH, How BS. Review of carbon capture absorbents for CO_2 utilization. Greenh Gases: Sci Technol 2022;12(3):394–427.
[12] Meng F, Meng Y, Ju T, Han S, Lin L, Jiang J. Research progress of aqueous amine solution for CO_2 capture: A review. Renew Sustain Energy Rev 2022;168:112902.
[13] Aghel B, Janati S, Wongwises S, Shadloo MS. Review on CO_2 capture by blended amine solutions. Int J Greenh Gas Control 2022;119:103715.
[14] Lian S, Song C, Liu Q, Duan E, Ren H, Kitamura Y. Recent advances in ionic liquids-based hybrid processes for CO_2 capture and utilization. J Environ Sci 2021;99:281–95.
[15] Shukla SK, Khokarale SG, Bui TQ, Mikkola J-PT. Ionic liquids: Potential materials for carbon dioxide capture and utilization. Front Mater 2019;6:42.
[16] Krishnan A, Gopinath KP, Vo D-VN, Malolan R, Nagarajan VM, Arun J. Ionic liquids, deep eutectic solvents and liquid polymers as green solvents in carbon capture technologies: a review. Environ Chem Lett 2020;18:2031–54.
[17] Liu Y, Dai Z, Zhang Z, Zeng S, Li F, Zhang X, et al. Ionic liquids/deep eutectic solvents for CO_2 capture: Reviewing and evaluating. Green Energy Environ 2021;6(3):314–28.
[18] Ding M, Flaig RW, Jiang H-L, Yaghi OM. Carbon capture and conversion using metal–organic frameworks and MOF-based materials. Chem Soc Rev 2019;48(10):2783–828.
[19] Hu Z, Wang Y, Shah BB, Zhao D. CO_2 capture in metal–organic framework adsorbents: an engineering perspective. Adv Sustain Syst 2019;3(1):1800080.
[20] Kumar S, Srivastava R, Koh J. Utilization of zeolites as CO_2 capturing agents: Advances and future perspectives. J CO_2 Utilization 2020;41:101251.
[21] Fu D, Davis ME. Carbon dioxide capture with zeotype materials. Chem Soc Rev 2022.
[22] Mukherjee A, Okolie JA, Abdelrasoul A, Niu C, Dalai AK. Review of post-combustion carbon dioxide capture technologies using activated carbon. J Environ Sci 2019;83:46–63.
[23] Azmi NZM, Buthiyappan A, Raman AAA, Patah MFA, Sufian S. Recent advances in biomass based activated carbon for carbon dioxide capture–a review. J Ind Eng Chem 2022;116:1–20.
[24] Borhani TN, Wang M. Role of solvents in CO_2 capture processes: The review of selection and design methods. Renew Sustain Energy Rev 2019;114:109299.
[25] Overa S, Ko BH, Zhao Y, Jiao F. Electrochemical approaches for CO_2 conversion to chemicals: a journey toward practical applications. Acc Chem Res 2022;55(5):638–48.
[26] Wang P, Dong R, Guo S, Zhao J, Zhang Z-M, Lu T-B. Improving photosensitization for photochemical CO_2-to-CO conversion. Natl Sci Rev 2020;7(9):1459–67.
[27] Yaashikaa P, Kumar PS, Varjani SJ, Saravanan A. A review on photochemical, biochemical and electrochemical transformation of CO_2 into value-added products. J CO_2 Utilization 2019;33:131–47.
[28] Alam MI, Cheula R, Moroni G, Nardi L, Maestri M. Mechanistic and multiscale aspects of thermo-catalytic CO_2 conversion to C_1 products. Catal Sci Technol 2021;11(20):6601–29.
[29] Kibria MG, Edwards JP, Gabardo CM, Dinh C-T, Seifitokaldani A, Sinton D, et al. Electrochemical CO_2 reduction into chemical feedstocks: from mechanistic electrocatalysis models to system design. Adv Mater 2019;31(31):1807166.
[30] Gao F-Y, Bao R-C, Gao M-R, Yu S-H. Electrochemical CO_2-to-CO conversion: electrocatalysts, electrolytes, and electrolyzers. J Mater Chem A 2020;8(31):15458–78.
[31] Xiao C, Zhang J. Architectural design for enhanced C_2 product selectivity in electrochemical CO_2 reduction using Cu-based catalysts: a review. ACS Nano 2021;15(5):7975–8000.
[32] Zhao Y, Zheng L, Jiang D, Xia W, Xu X, Yamauchi Y, et al. Nanoengineering metal organic framework-based materials for use in electrochemical CO_2 reduction reactions. Small 2021;17(16):2006590.
[33] Huang Y, Rehman F, Tamtaji M, Li X, Huang Y, Zhang T, et al. Mechanistic understanding and design of non-noble metal-based single-atom catalysts supported on two-dimensional materials for CO_2 electroreduction. J Mater Chem A 2022;10(11):5813–34.
[34] Schlexer Lamoureux P, Winther KT, Garrido Torres JA, Streibel V, Zhao M, Bajdich M, et al. Machine learning for computational heterogeneous catalysis. ChemCatChem 2019;11(16):3581–601.
[35] Liu J, Luo W, Wang L, Zhang J, Fu X-Z, Luo J-L. Toward excellence of electrocatalyst design by emerging descriptor-oriented machine learning. Adv Funct Mater 2022;32(17):2110748.
[36] Nandy A, Duan C, Taylor MG, Liu F, Steeves AH, Kulik HJ. Computational discovery of transition-metal complexes: from high-throughput screening to machine learning. Chem Rev 2021;121(16):9927–10000.

[37] Xu J, Cao X-M, Hu P. Perspective on computational reaction prediction using machine learning methods in heterogeneous catalysis. Phys Chem Chem Phys 2021;23(19):11155–79.
[38] Yang Z, Gao W. Applications of machine learning in alloy catalysts: rational selection and future development of descriptors. Adv Sci 2022;9(12):2106043.
[39] Zhang X, Zhou Z. Perspective on theoretical models for CO_2 electrochemical reduction. J Phys Chem C 2022;126(8):3820–9.
[40] Orupattur NV, Mushrif SH, Prasad V. Catalytic materials and chemistry development using a synergistic combination of machine learning and ab initio methods. Comput Mater Sci 2020;174:109474.
[41] Dattila F, Seemakurthi RR, Zhou Y, López N. Modeling operando electrochemical CO_2 reduction. Chem Rev 2022;122(12):11085–130.
[42] Wu Q, Pan M, Zhang S, Sun D, Yang Y, Chen D, et al. Research progress in high-throughput screening of CO_2 reduction catalysts. Energies 2022;15(18):6666.
[43] Sullivan I, Goryachev A, Digdaya IA, Li X, Atwater HA, Vermaas DA, et al. Coupling electrochemical CO_2 conversion with CO_2 capture. Nat Catal 2021;4(11):952–8.
[44] Jerng SE, Gallant BM. Electrochemical reduction of CO_2 in the captured state using aqueous or nonaqueous amines. Iscience 2022;25(7).
[45] Chen L, Li F, Zhang Y, Bentley CL, Horne M, Bond AM, et al. Electrochemical reduction of carbon dioxide in a monoethanolamine capture medium. ChemSusChem 2017;10(20):4109–18.
[46] Lee G, Li YC, Kim J-Y, Peng T, Nam D-H, Sedighian Rasouli A, et al. Electrochemical upgrade of CO_2 from amine capture solution. Nat Energy 2021;6(1):46–53.
[47] Pérez-Gallent E, Vankani C, Sánchez-Martínez C, Anastasopol A, Goetheer E. Integrating CO_2 capture with electrochemical conversion using amine-based capture solvents as electrolytes. Ind Eng Chem Res 2021;60(11):4269–78.
[48] Bhattacharya M, Sebghati S, Vercella YM, Saouma CT. Electrochemical reduction of carbamates and carbamic acids: implications for combined carbon capture and electrochemical CO_2 recycling. J Electrochem Soc 2020;167(8):086507.
[49] Khurram A, He M, Gallant BM. Tailoring the discharge reaction in Li-CO_2 batteries through incorporation of CO_2 capture chemistry. Joule 2018;2(12):2649–66.
[50] Yan Y, Borhani TN, Subraveti SG, Pai KN, Prasad V, Rajendran A, et al. Harnessing the power of machine learning for carbon capture, utilisation, and storage (CCUS)–a state-of-the-art review. Energy Environ Sci 2021;14(12):6122–57.
[51] Mai H, Le TC, Chen D, Winkler DA, Caruso RA. Machine learning for electrocatalyst and photocatalyst design and discovery. Chem Rev 2022;122(16):13478–515.
[52] Sulley GA, Montemore MM. Recent progress towards a universal machine learning model for reaction energetics in heterogeneous catalysis. Curr Opin Chem Eng 2022;36:100821.
[53] Guan Y, Chaffart D, Liu G, Tan Z, Zhang D, Wang Y, et al. Machine learning in solid heterogeneous catalysis: Recent developments, challenges and perspectives. Chem Eng Sci 2022;248:117224.
[54] Tamtaji M, Gao H, Hossain MD, Galligan PR, Wong H, Liu Z, et al. Machine learning for design principles for single atom catalysts towards electrochemical reactions. J Mater Chem A 2022;10(29):15309–31.
[55] Mou T, Pillai HS, Wang S, Wan M, Han X, Schweitzer NM, et al. Bridging the complexity gap in computational heterogeneous catalysis with machine learning. Nat Catal 2023;6(2):122–36.
[56] Rahimi M, Moosavi SM, Smit B, Hatton TA. Toward smart carbon capture with machine learning. Cell Rep Phys Sci 2021;2(4).
[57] Yang Z, Chen B, Chen H, Li H. A critical review on machine-learning-assisted screening and design of effective sorbents for carbon dioxide (CO_2) capture. Front Energy Res 2023;10:1043064.
[58] Koutsoukos S, Philippi F, Malaret F, Welton T. A review on machine learning algorithms for the ionic liquid chemical space. Chem Sci 2021;12(20):6820–43.
[59] Yusuf F, Olayiwola T, Afagwu C. Application of artificial intelligence-based predictive methods in ionic liquid studies: A review. Fluid Phase Equilib 2021;531:112898.
[60] Shama VM, Swami AR, Aniruddha R, Sreedhar I, Reddy BM. Process and engineering aspects of carbon capture by ionic liquids. J CO_2 Utilization 2021;48:101507.
[61] Sun J, Sato Y, Sakai Y, Kansha Y. A review of ionic liquids design and deep eutectic solvents for CO_2 capture with machine learning. J Clean Prod 2023;137695.
[62] Zhang Z, Cao X, Geng C, Sun Y, He Y, Qiao Z, et al. Machine learning aided high-throughput prediction of ionic liquid@ mof composites for membrane-based CO_2 capture. J Membr Sci 2022;650:120399.
[63] Van Duin AC, Dasgupta S, Lorant F, Goddard WA. ReaxFF: a reactive force field for hydrocarbons. J Phys Chem A 2001;105(41):9396–409.
[64] Toyao T, Maeno Z, Takakusagi S, Kamachi T, Takigawa I, Shimizu K-i. Machine learning for catalysis informatics: recent applications and prospects. Acs Catal 2019;10(3):2260–97.
[65] Daiyan R, Saputera WH, Masood H, Leverett J, Lu X, Amal R. A disquisition on the active sites of heterogeneous catalysts for electrochemical reduction of CO_2 to value-added chemicals and fuel. Adv Energy Mater 2020;10(11):1902106.
[66] Wang Y, Liu J, Zheng G. Designing copper-based catalysts for efficient carbon dioxide electroreduction. Adv Mater 2021;33(46):2005798.
[67] Zhu S, Delmo EP, Li T, Qin X, Tian J, Zhang L, et al. Recent advances in catalyst structure and composition engineering strategies for regulating CO_2 electrochemical reduction. Adv Mater 2021;33(50):2005484.
[68] Nosengo N, et al. The material code. Nature 2016;533(7601):22–5.
[69] Jain A, Ong SP, Hautier G, Chen W, Richards WD, Dacek S, et al. Commentary: The materials project: A materials genome approach to accelerating materials innovation. APL Mater 2013;1(1).
[70] Toher C, Oses C, Hicks D, Gossett E, Rose F, Nath P, et al. The AFLOW fleet for materials discovery. In: Handbook of materials modeling: methods: theory and modeling. Springer; 2020, p. 1785–812.
[71] Kirklin S, Saal JE, Meredig B, Thompson A, Doak JW, Aykol M, et al. The open quantum materials database (OQMD): assessing the accuracy of DFT formation energies. npj Comput Mater 2015;1(1):1–15.
[72] Talirz L, Kumbhar S, Passaro E, Yakutovich AV, Granata V, Gargiulo F, et al. Materials cloud, a platform for open computational science. Sci Data 2020;7(1):299.
[73] Himanen L, Geurts A, Foster AS, Rinke P. Data-driven materials science: status, challenges, and perspectives. Adv Sci 2019;6(21):1900808.
[74] Tao Q, Xu P, Li M, Lu W. Machine learning for perovskite materials design and discovery. npj Comput Mater 2021;7(1):23.
[75] Sorokina M, Steinbeck C. Review on natural products databases: where to find data in 2020. J Cheminform 2020;12(1):20.
[76] Chanussot L, Das A, Goyal S, Lavril T, Shuaibi M, Riviere M, et al. Open catalyst 2020 (OC20) dataset and community challenges. Acs Catal 2021;11(10):6059–72.
[77] Tran R, Lan J, Shuaibi M, Wood BM, Goyal S, Das A, et al. The open catalyst 2022 (OC22) dataset and challenges for oxide electrocatalysts. ACS Catal 2023;13(5):3066–84.
[78] Bannwarth C, Ehlert S, Grimme S. GFN2-xTB—An accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. J Chem Theory Comput 2019;15(3):1652–71.
[79] Jørgensen PB, Bhowmik A. Deepdft: Neural message passing network for accurate charge density prediction. 2020, arXiv preprint arXiv:2011.03346.
[80] Swain MC, Cole JM. ChemDataExtractor: a toolkit for automated extraction of chemical information from the scientific literature. J Chem Inf Model 2016;56(10):1894–904.
[81] Shetty P, Rajan AC, Kuenneth C, Gupta S, Panchumarti LP, Holm L, et al. A general-purpose material property data extraction pipeline from large polymer corpora using natural language processing. npj Comput Mater 2023;9(1):52.
[82] Liu Y, Esan OC, Pan Z, An L. Machine learning for advanced energy materials. Energy AI 2021;3:100049.
[83] Wei J, Chu X, Sun X-Y, Xu K, Deng H-X, Chen J, et al. Machine learning in materials science. InfoMat 2019;1(3):338–58.
[84] Schleder GR, Padilha AC, Acosta CM, Costa M, Fazzio A. From DFT to machine learning: recent approaches to materials science–a review. J Phys: Mater 2019;2(3):032001.
[85] Girosi F, Poggio T. Networks and the best approximation property. Biol Cybern 1990;63(3):169–76.
[86] Breiman L. Classification and regression trees. Routledge; 2017.
[87] Krzywinski M, Altman N. Classification and regression trees. Nature Methods 2017;14(8):757–8.
[88] Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. Neural Netw 1989;2(5):359–66.
[89] Werbos PJ. Applications of advances in nonlinear sensitivity analysis. In: System modeling and optimization: proceedings of the 10th IFIP conference new york city, USA, August 31–September 4, 1981. Springer; 2005, p. 762–70.
[90] Wang H, Xie Y, Li D, Deng H, Zhao Y, Xin M, et al. Rapid identification of X-ray diffraction patterns based on very limited data by interpretable convolutional neural networks. J Chem Inf Model 2020;60(4):2004–11.
[91] Lim B, Bellec E, Dupraz M, Leake S, Resta A, Coati A, et al. A convolutional neural network for defect classification in bragg coherent X-ray diffraction. npj Comput Mater 2021;7(1):115.
[92] Park WB, Chung J, Jung J, Sohn K, Singh SP, Pyo M, et al. Classification of crystal structure using a convolutional neural network. IUCrJ 2017;4(4):486–94.
[93] Jung G, Jung SG, Cole JM. Automatic materials characterization from infrared spectra using convolutional neural networks. Chem Sci 2023;14(13):3600–9.
[94] Lim B, Bellec E, Dupraz M, Leake S, Resta A, Coati A, et al. A convolutional neural network for defect classification in bragg coherent X-ray diffraction. npj Comput Mater 2021;7(1):115.
[95] Mozaffari MH, Tay L-L. A review of 1D convolutional neural networks toward unknown substance identification in portable Raman spectrometer. 2020, arXiv preprint arXiv:2006.10575.
[96] Wei J, Wang X, Wang Z, Cao J. Qualitative detection of pesticide residues using mass spectral data based on convolutional neural network. SN Appl Sci 2021;3:1–13.
[97] Seddiki K, Saudemont P, Precioso F, Ogrinc N, Wisztorski M, Salzet M, et al. Cumulative learning enables convolutional neural network representations for small mass spectrometry data classification. Nat Commun 2020;11(1):5595.

[98] Schütt KT, Sauceda HE, Kindermans P-J, Tkatchenko A, Müller K-R. Schnet–a deep learning architecture for molecules and materials. J Chem Phys 2018;148(24).

[99] Gasteiger J, Groß J, Günnemann S. Directional message passing for molecular graphs. 2020, arXiv preprint arXiv:2003.03123.

[100] Choudhary K, DeCost B. Atomistic line graph neural network for improved materials property predictions. npj Comput Mater 2021;7(1):185.

[101] Gasteiger J, Shuaibi M, Sriram A, Günnemann S, Ulissi Z, Zitnick CL, et al. Gemnet-oc: developing graph neural networks for large and diverse molecular simulation datasets. 2022, arXiv preprint arXiv:2204.02782.

[102] Do K, Tran T, Venkatesh S. Graph transformation policy network for chemical reaction prediction. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. 2019, p. 750–60.

[103] Chen S, Jung Y. A generalized-template-based graph neural network for accurate organic reactivity prediction. Nat Mach Intell 2022;4(9):772–80.

[104] Kuhn M, Johnson K. Feature engineering and selection: A practical approach for predictive models. Chapman and Hall/CRC; 2019.

[105] Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst 2017;30.

[106] Ren P, Xiao Y, Chang X, Huang P-Y, Li Z, Gupta BB, et al. A survey of deep active learning. ACM Comput Surv (CSUR) 2021;54(9):1–40.

[107] Kusne AG, Yu H, Wu C, Zhang H, Hattrick-Simpers J, DeCost B, et al. On-the-fly closed-loop materials discovery via Bayesian active learning. Nat Commun 2020;11(1):5966.

[108] Saal JE, Oliynyk AO, Meredig B. Machine learning in materials discovery: Confirmed predictions and their underlying approaches. Annu Rev Mater Res 2020;50:49–69.

[109] Zhan X, Wang Q, Huang K-h, Xiong H, Dou D, Chan AB. A comparative survey of deep active learning. 2022, arXiv preprint arXiv:2203.13450.

[110] Hein M, Andriushchenko M, Bitterwolf J. Why relu networks yield high-confidence predictions far away from the training data and how to mitigate the problem. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019, p. 41–50.

[111] Chen G, Luo X, Zhang H, Fu K, Liang Z, Rongwong W, et al. Artificial neural network models for the prediction of CO_2 solubility in aqueous amine solutions. Int J Greenh Gas Control 2015;39:174–84.

[112] Zhang Z, Li H, Chang H, Pan Z, Luo X. Machine learning predictive framework for CO_2 thermodynamic properties in solution. J CO_2 Utilization 2018;26:152–9.

[113] Yarveicy H, Ghiasi MM, Mohammadi AH. Performance evaluation of the machine learning approaches in modeling of CO_2 equilibrium absorption in piperazine aqueous solution. J Mol Liq 2018;255:375–83.

[114] Ghiasi MM, Abedi-Farizhendi S, Mohammadi AH. Modeling equilibrium systems of amine-based CO_2 capture by implementing machine learning approaches. Environ Prog Sustain Energy 2019;38(5):13160.

[115] Pouryousefi F, Idem R, Supap T, Tontiwachwuthikul P. Artificial neural networks for accurate prediction of physical properties of aqueous quaternary systems of carbon dioxide (CO_2)-loaded 4-(Diethylamino)-2-butanol and Methyldiethanolamine blended with Monoethanolamine. Ind Eng Chem Res 2016;55(44):11614–21.

[116] Dong S, Quan H, Zhao D, Li H, Geng J, Liu H. Generic AI models for mass transfer coefficient prediction in amine-based CO_2 absorber, part I: BPNN model. Chem Eng Sci 2022;264:118165.

[117] Orlov AA, Valtz A, Coquelet C, Rozanska X, Wimmer E, Marcou G, et al. Computational screening methodology identifies effective solvents for CO_2 capture. Commun Chem 2022;5(1):37.

[118] Hwang J, Kim J, Lee HW, Na J, Ahn BS, Lee SD, et al. An experimental based optimization of a novel water lean amine solvent for post combustion CO_2 capture process. Appl Energy 2019;248:174–84.

[119] Mansouri K, Grulke CM, Judson RS, Williams AJ. OPERA models for predicting physicochemical properties and environmental fate endpoints. J Cheminform 2018;10(1):1–19.

[120] Varnek A, Fourches D, Horvath D, Klimchuk O, Gaudin C, Vayer P, et al. ISIDA-platform for virtual screening based on fragment and pharmacophoric descriptors. Curr Comput-Aided Drug Des 2008;4(3):191.

[121] Ruggiu F, Marcou G, Varnek A, Horvath D. ISIDA property-labelled fragment descriptors. Mol Inform 2010;29(12):855–68.

[122] Wei D, Ivaska A. Applications of ionic liquids in electrochemical sensors. Anal Chim Acta 2008;607(2):126–35.

[123] Zhao Y, Huang Y, Zhang X, Zhang S. A quantitative prediction of the viscosity of ionic liquids using S σ-profile molecular descriptors. Phys Chem Chem Phys 2015;17(5):3761–7.

[124] Dong Q, Muzny CD, Kazakov A, Diky V, Magee JW, Widegren JA, et al. Ilthermo: A free-access web database for thermodynamic properties of ionic liquids. J Chem Eng Data 2007;52(4):1151–9.

[125] Venkatraman V, Alsberg BK. Predicting CO_2 capture of ionic liquids using machine learning. J CO_2 Utilization 2017;21:162–8.

[126] Ghiasi MM, Mohammadi AH. Application of decision tree learning in modelling CO_2 equilibrium absorption in ionic liquids. J Mol Liq 2017;242:594–605.

[127] Aghaie M, Zendehboudi S. Estimation of CO_2 solubility in ionic liquids using connectionist tools based on thermodynamic and structural characteristics. Fuel 2020;279:117984.

[128] Song Z, Shi H, Zhang X, Zhou T. Prediction of CO_2 solubility in ionic liquids using machine learning methods. Chem Eng Sci 2020;223:115752.

[129] Wu T, Li W-L, Chen M-Y, Zhou Y-M, Zhang Q-Y. Prediction of Henry's law constants of CO_2 in imidazole ionic liquids using machine learning methods based on empirical descriptors. Chem Pap 2021;75:1619–28.

[130] Tian Y, Wang X, Liu Y, Hu W. Prediction of CO_2 and n_2 solubility in ionic liquids using a combination of ionic fragments contribution and machine learning methods. J Mol Liq 2023;383:122066.

[131] Klamt A. Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena. J Phys Chem 1995;99(7):2224–35.

[132] Klamt A, Jonas V, Bürger T, Lohrenz JC. Refinement and parametrization of COSMO-RS. J Phys Chem A 1998;102(26):5074–85.

[133] Taheri M, Zhu R, Yu G, Lei Z. Ionic liquid screening for CO_2 capture and H_2S removal from gases: The syngas purification case. Chem Eng Sci 2021;230:116199.

[134] Kuroki N, Suzuki Y, Kodama D, Chowdhury FA, Yamada H, Mori H. Machine learning-boosted design of ionic liquids for CO_2 absorption and experimental verification. J Phys Chem B 2023;127(9):2022–7.

[135] Jian Y, Wang Y, Barati Farimani A. Predicting CO_2 absorption in ionic liquids with molecular descriptors and explainable graph neural networks. ACS Sustain Chem Eng 2022;10(50):16681–91.

[136] Zhang K, Wu J, Yoo H, Lee Y. Machine learning-based approach for tailor-made design of ionic liquids: Application to CO_2 capture. Sep Purif Technol 2021;275:119117.

[137] Liu Y, Yu H, Sun Y, Zeng S, Zhang X, Nie Y, et al. Screening deep eutectic solvents for CO_2 capture with COSMO-RS. Front Chem 2020;8:82.

[138] Song Z, Hu X, Wu H, Mei M, Linke S, Zhou T, et al. Systematic screening of deep eutectic solvents as sustainable separation media exemplified by the CO_2 capture process. ACS Sustain Chem Eng 2020;8(23):8741–51.

[139] Sagar A, Upadhyayula S. Implementation of artificial neural networks in the assessment of CO_2 solubility in deep eutectic and ionic liquid solvents–performance and cost comparison. Sustain Chem Climate Action 2022;1:100007.

[140] Wang J, Song Z, Chen L, Xu T, Deng L, Qi Z. Prediction of CO_2 solubility in deep eutectic solvents using random forest model based on COSMO-RS-derived descriptors. Green Chem Eng 2021;2(4):431–40.

[141] Lemaoui T, Boublia A, Lemaoui S, Darwish AS, Ernst B, Alam M, et al. Predicting the CO_2 capture capability of deep eutectic solvents and screening over 1000 of their combinations using machine learning. ACS Sustain Chem Eng 2023;11(26):9564–80.

[142] Mohan M, Demerdash O, Simmons BA, Smith JC, Kidder MK, Singh S. Accurate prediction of carbon dioxide capture by deep eutectic solvents using quantum chemistry and a neural network. Green Chem 2023;25(9):3475–92.

[143] Makarov DM, Fadeeva YA, Golubev VA, Kolker AM. Designing deep eutectic solvents for efficient CO_2 capture: A data-driven screening approach. Separation and Purification Technology 2023;325:124614.

[144] Ulissi ZW, Tang MT, Xiao J, Liu X, Torelli DA, Karamad M, et al. Machine-learning methods enable exhaustive searches for active bimetallic facets and reveal active site motifs for CO_2 reduction. Acs Catal 2017;7(10):6600–8.

[145] Tran K, Ulissi ZW. Active learning across intermetallics to guide discovery of electrocatalysts for CO_2 reduction and H_2 evolution. Nat Catal 2018;1(9):696–703.

[146] Ma X, Li Z, Achenie LE, Xin H. Machine-learning-augmented chemisorption model for CO_2 electroreduction catalyst screening. J Phys Chem Lett 2015;6(18):3528–33.

[147] Noh J, Back S, Kim J, Jung Y. Active learning with non-ab initio input features toward efficient CO_2 reduction catalysts. Chemical science 2018;9(23):5152–9.

[148] Chen Y, Huang Y, Cheng T, Goddard III WA. Identifying active sites for CO_2 reduction on dealloyed gold surfaces by combining machine learning with multiscale simulations. J Am Chem Soc 2019;141(29):11651–7.

[149] Zhong M, Tran K, Min Y, Wang C, Wang Z, Dinh C-T, et al. Accelerated discovery of CO_2 electrocatalysts using active machine learning. Nature 2020;581(7807):178–83.

[150] Xing M, Zhang Y, Li S, He H, Sun S. Prediction of carbon dioxide reduction catalyst using machine learning with a few-feature model: WLEDZ. J Phys Chem C 2022;126(40):17025–35.

[151] Feng H, Ding H, He P, Wang S, Li Z, Zheng Z, et al. Data-driven design of dual-metal-site catalysts for the electrochemical carbon dioxide reduction reaction. J Mater Chem A 2022;10(36):18803–11.

[152] Wan X, Zhang Z, Niu H, Yin Y, Kuai C, Wang J, et al. Machine-learning-accelerated catalytic activity predictions of transition metal phthalocyanine dual-metal-site catalysts for CO_2 reduction. J Phys Chem Lett 2021;12(26):6111–8.

[153] Wang D, Cao R, Hao S, Liang C, Chen G, Chen P, et al. Accelerated prediction of cu-based single-atom alloy catalysts for CO_2 reduction by machine learning. Green Energy Environ 2023;8:820–30.

[154] Zhu Q, Gu Y, Liang X, Wang X, Ma J. A machine learning model to predict CO_2 reduction reactivity and products transferred from metal-zeolites. ACS Catal 2022;12(19):12336–48.

[155] Chen A, Zhang X, Chen L, Yao S, Zhou Z. A machine learning model on simple features for CO_2 reduction electrocatalysts. J Phys Chem C 2020;124(41):22471–8.
[156] Mazheika A, Wang Y-G, Valero R, Viñes F, Illas F, Ghiringhelli LM, et al. Artificial-intelligence-driven discovery of catalyst genes with application to CO_2 activation on semiconductor oxides. Nature Commun 2022;13(1):419.
[157] Chen ZW, Gariepy Z, Chen L, Yao X, Anand A, Liu S-J, et al. Machine-learning-driven high-entropy alloy catalyst discovery to circumvent the scaling relation for CO_2 reduction reaction. ACS Catal 2022;12(24):14864–71.
[158] Pedersen JK, Batchelor TA, Bagger A, Rossmeisl J. High-entropy alloys as catalysts for the CO_2 and co reduction reactions. Acs Catal 2020;10(3):2169–76.
[159] Nellaiappan S, Katiyar NK, Kumar R, Parui A, Malviya KD, Pradeep K, et al. High-entropy alloys as catalysts for the CO_2 and CO reduction reactions: Experimental realization. Acs Catal 2020;10(6):3658–63.
[160] Wang J, Liu J, Wang H, Zhou M, Ke G, Zhang L, et al. A comprehensive transformer-based approach for high-accuracy gas adsorption predictions in metal-organic frameworks. Nature Commun 2024;15(1):1904.
[161] Dureckova H, Krykunov M, Aghaji MZ, Woo TK. Robust machine learning models for predicting high CO_2 working capacity and CO_2/H_2 selectivity of gas adsorption in metal organic frameworks for precombustion carbon capture. J Phys Chem C 2019;123(7):4133–9.
[162] Orhan IB, Le TC, Babarao R, Thornton AW. Accelerating the prediction of CO_2 capture at low partial pressures in metal-organic frameworks using new machine learning descriptors. Commun Chem 2023;6(1):214.
[163] Yang C-T, Pandey I, Trinh D, Chen C-C, Howe JD, Lin L-C. Deep learning neural network potential for simulating gaseous adsorption in metal–organic frameworks. Mater Adv 2022;3(13):5299–303.
[164] Zheng B, Oliveira FL, Neumann Barros Ferreira R, Steiner M, Hamann H, Gu GX, et al. Quantum informed machine-learning potentials for molecular dynamics simulations of CO_2's chemisorption and diffusion in Mg-MOF-74. ACS Nano 2023;17(6):5579–87.
[165] Wang Y, Wang J, Cao Z, Barati Farimani A. Molecular contrastive learning of representations via graph neural networks. Nat Mach Intell 2022;4(3):279–87.
[166] Fang X, Liu L, Lei J, He D, Zhang S, Zhou J, et al. Geometry-enhanced molecular representation learning for property prediction. Nat Mach Intell 2022;4(2):127–34.
[167] Halgren TA. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. J Comput Chem 1996;17(5–6):490–519.
[168] Ahmad W, Simon E, Chithrananda S, Grand G, Ramsundar B. Chemberta-2: Towards chemical foundation models. 2022, arXiv preprint arXiv:2209.01712.
[169] Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. Roberta: A robustly optimized bert pretraining approach. 2019, arXiv preprint arXiv:1907.11692.
[170] Koker T, Quigley K, Spaeth W, Frey NC, Li L. Graph contrastive learning for materials. 2022, arXiv preprint arXiv:2211.13408.
[171] Magar R, Wang Y, Barati Farimani A. Crystal twins: self-supervised learning for crystalline material property prediction. npj Comput Mater 2022;8(1):231.
[172] Zbontar J, Jing L, Misra I, LeCun Y, Deny S. Barlow twins: Self-supervised learning via redundancy reduction. In: International conference on machine learning. PMLR; 2021, p. 12310–20.
[173] Bond-Taylor S, Leach A, Long Y, Willcocks CG. Deep generative modelling: A comparative review of vaes, gans, normalizing flows, energy-based and autoregressive models. IEEE Trans Pattern Anal Mach Intell 2021;44(11):7327–47.
[174] De Cao N, Kipf T. MolGAN: An implicit generative model for small molecular graphs. 2018, arXiv preprint arXiv:1805.11973.
[175] Jin W, Barzilay R, Jaakkola T. Junction tree variational autoencoder for molecular graph generation. In: International conference on machine learning. PMLR; 2018, p. 2323–32.
[176] Zang C, Wang F. Moflow: an invertible flow model for generating molecular graphs. In: Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining. 2020, p. 617–26.
[177] Vignac C, Krawczuk I, Siraudin A, Wang B, Cevher V, Frossard P. Digress: Discrete denoising diffusion for graph generation. 2022, arXiv preprint arXiv:2209.14734.
[178] Bagal V, Aggarwal R, Vinod P, Priyakumar UD. Molgpt: molecular generation using a transformer-decoder model. J Chem Inf Model 2021;62(9):2064–76.
[179] Xie T, Fu X, Ganea O-E, Barzilay R, Jaakkola T. Crystal diffusion variational autoencoder for periodic material generation. 2021, arXiv preprint arXiv:2110.06197.
[180] Pakornchote T, Choomphon-Anomakhun N, Arrerut S, Atthapak C, Khamkaeo S, Chotibut T, et al. Diffusion probabilistic models enhance variational autoencoder for crystal structure generative modeling. Sci Rep 2024;14(1):1275.
[181] Guo T, Nan B, Liang Z, Guo Z, Chawla N, Wiest O, et al. What can large language models do in chemistry? A comprehensive benchmark on eight tasks. Adv Neural Inf Process Syst 2024;36.
[182] Jablonka KM, Schwaller P, Ortega-Guerrero A, Smit B. Leveraging large language models for predictive chemistry. Nat Mach Intell 2024;1–9.
[183] Tonmoy S, Zaman S, Jain V, Rani A, Rawte V, Chadha A, et al. A comprehensive survey of hallucination mitigation techniques in large language models. 2024, arXiv preprint arXiv:2401.01313.