ACS Sustainable Chemistry & Engineering

pubs.acs.org/journal/ascecg

Research Article

Predicting CO_2 Absorption in Ionic Liquids with Molecular Descriptors and Explainable Graph Neural Networks

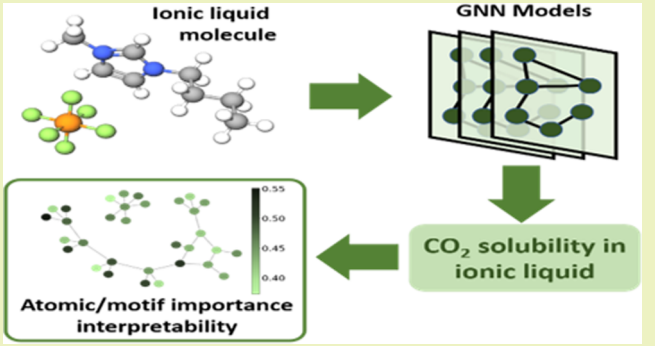
Yue Jian, Yuyang Wang, and Amir Barati Farimani*

Cite This: *ACS Sustainable Chem. Eng.* 2022, 10, 16681−16691

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Ionic liquids (ILs) provide a promising solution for CO_2 capture and storage to mitigate global warming. However, identifying and designing the high-capacity IL from the giant chemical space require expensive and exhaustive simulations and experiments. Machine learning (ML) can accelerate the process of searching for desirable ionic molecules through accurate and efficient property predictions in a data-driven manner. However, existing descriptors and ML models for the ionic molecule suffer from the inefficient adaptation of molecular graph structure. Besides, few works have investigated the explainability of ML models to help understand the learned features that can guide the design of efficient ionic molecules. In this work, we develop both fingerprint-based ML models and graph neural networks (GNNs) to predict the CO_2 absorption in ILs. Fingerprint works on graph structure at the feature extraction stage, while GNNs directly handle molecule structure in both the feature extraction and model prediction stage. We show that our method outperforms previous ML models by reaching a high accuracy (MAE of 0.0137, R^2 of 0.9884). Furthermore, we take the advantage of GNN representation and develop a substructure-based explanation method that provides insight into how each chemical fragment within IL molecules contributes to the CO_2 absorption prediction of ML models. We also show that our result agrees with some ground truth on functional group importance from the theoretical understanding of CO_2 absorption in ILs, which can advise on the design of novel and efficient functional ILs in the future.



KEYWORDS: *explainable deep learning, graph neural networks, machine learning, CO_2 absorption, ionic liquid, global warming*

INTRODUCTION

Global warming is a major environmental problem in our world. Based on the prediction of the Intergovernmental Panel on Climate Change (IPCC), the average temperature of our world will increase by about 1.9 °C if we do not take any action by 2100.[1] Among all of the greenhouse gases, CO_2 makes the most contribution to global warming to an extent of about 78.6%.[2] How to effectively capture and store CO_2 is crucial for solving the global warming problem. Existing methods, including physisorption/chemisorption,[3,4] membrane separation[5] or molecular sieves,[6] carbamation, amine physical absorption,[7] amine dry scrubbing,[8] and mineral carbonation,[9,10] have been introduced to absorb CO_2. However, the reagents used in these methods suffer from insufficient carbon dioxide storage capacity, high energy demand in absorption process, and low thermal stability.[11,12] The evaporation and degradation of reagents may lead the storage process to become costly.[13]

Ionic liquids (ILs) are families of molten salt that remains in liquid state at room temperature. Over the past decades, they have received significant attention and have been an intensive research area due to their unique physical and chemical properties, such as nonvolatility, high chemical stability, high CO_2 solubility, and easy operation at liquid state. These properties make ILs an ideal candidate for CO_2 storage.[14−22] Usually, IL comprises a pair of ions with different charges, and the combination of ions largely determines the properties of ILs. However, such combinations of cations and anions as well as the various selections of cations and anions themselves make it challenging to exhaust the design space of IL for efficient CO_2 storage through experiments. To efficiently estimate the CO_2 absorption of ILs, researchers have investigated the quantitative structure−property relationship (QSPR). QSPR methods aim at building mathematical models for the prediction of numerical properties based on the structural information of chemical compounds.[23,24] However, traditional methods used in QSPR such as molecular dynamic (MD) and

Received: October 5, 2022
Revised: November 15, 2022
Published: December 2, 2022

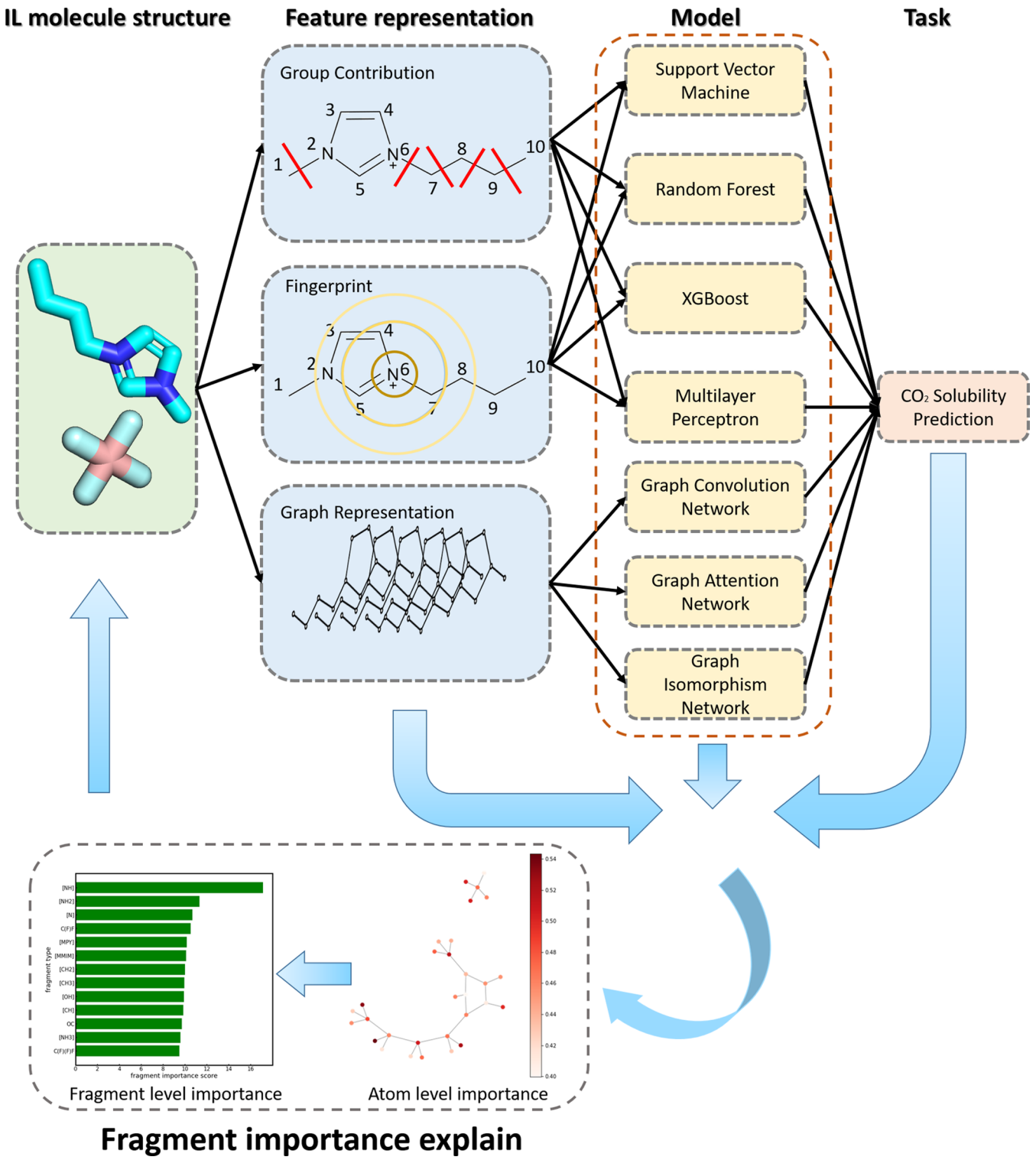


Figure 1. Overview of methods in the paper. We first use GC and FP as descriptors. Shallow machine learning models (SVM, RF, XGBoost) and deep learning models (MLP) are built on top of the descriptors for CO_2 solubility prediction. Besides descriptor-based machine learning, GNN-based models (GCN, GAT, GIN) are employed to perform the solubility prediction. Furthermore, we develop an explanation method (IL explainer) for ionic molecules that can take in a fine-tuned GNN model, and an IL data point, and return the fragment importance of the molecule. We finally make the importance explanation for the functional groups within cations across the whole data set.

density functional theory (DFT) can be computationally challenging for ILs due to the complexity of inter- and intramolecular interactions.[25−27]

The recent development of machine learning (ML) methods bears promise for QSPR modeling through accurate and efficient property predictions of chemical compounds. Compared with conventional simulation methods like MD or DFT,[28] ML methods have demonstrated similar accuracy but with less computational cost in various chemical applications.[29−31] Especially, several works have explored applying

https://doi.org/10.1021/acssuschemeng.2c05985
ACS Sustainable Chem. Eng. 2022, 10, 16681−16691

ML models to solving ionic liquid problems via various descriptors of IL molecules. Group contribution theory (GC) is one of the earliest descriptors for IL molecules.[32−35] GC manually breaks down a molecule into different characteristic functional groups and counts the existence frequency of each group. However, this descriptor is highly human experience-dependent and may lead to the loss of information for substructures within the group. Another way of finding molecule descriptors is the quantum chemical descriptor (QC), which utilizes the calculated properties from DFT to provide submolecular-level representations for IL molecules.[36−38] However, to gain QC descriptors, one needs to perform expensive and time-consuming QC calculations like DFT to acquire the properties. ML models like support vector machine (SVM), random forest (RF), and deep learning models such as multilayer perceptron (MLP), convolutional neural network (CNN), and recurrent neural network (RNN) have been applied on top of the descriptors to perform various property prediction tasks.[39,40] But, both GC and QC descriptors can lack the modeling of the structural information of molecules, which confines the performance of ML models. Other molecular descriptors like extended-connectivity fingerprints (ECFPs) create a feature vector by iteratively aggregating the neighbor information of each atom and hashing that into a vector.[41,42] ECFPs directly encode the structural information of molecules and can be more expressive. Such molecular fingerprints (FPs) and how different ML models built upon them perform have not been well studied for CO_2 absorption in ILs.

Recently, GNNs have been shown to be a powerful tool for molecule feature representation and property prediction, and have received a significant amount of attention.[43−49] At the feature representation stage, GNNs directly work on the molecular structure. It treats the molecule as a graph and utilizes an adjacent matrix to encode the bond edge and connectivity, as well as a node feature matrix to encode the atom and related properties. This representation is more generalizable, stable, and less computationally expensive compared with GC and QC descriptors. At the model training and prediction stage, GNN aggregates the node message through edge during the forward process,[46,50,51] and it outperforms other neural-network-based models on unstructured graphical data.[46] GNNs have been involved in many areas related to molecules, such as drug discovery, quantum chemistry, and structural biology.[52] However, existing works using GNNs on ILs tend to focus solely on one family of anions or cations, which still needs to be expanded and generalized.[53,54]

Besides building ML models to obtain accurate and efficient predictions, how to explain the output from ML models given certain input data is also an active research area.[25,55] In traditional experimental and computational chemistry, researchers heavily rely on their knowledge and experience in designing new compounds. On the other hand, as a black box, the intermediate decision process of the ML models is hard to unveil. Understanding how ML models make decisions can provide us with extra insights into how the structure of the input molecule affects the property of IL and new IL design from a data-driven perspective. Explainable algorithms have been developed on GNNs to analyze the importance of each input edge.[56] However, in the IL research area, researchers usually focus on the prediction performance of GNN on various properties but ignore the explainability of the GNN model. Benefiting from graph representation, GNNs have the potential to provide an explanation of the molecular structure importance that reaches atom and bond levels.

In this paper, we introduce two categories of methods for CO_2 solubility prediction, namely, descriptor-based machine learning models and GNNs. Besides, we also develop an explanation method for IL molecule substructure importance analysis. For descriptor-based machine learning models, GC and FP are included as descriptors. We then compare the expressiveness of FP with GC on different machine learning models. For the GNN part, we include graph convolutional networks (GCNs), graph attention networks (GATs), and graph isomorphism networks (GINs) to perform the CO_2 solubility prediction. Moreover, we make two improvements to data representation and the GNN framework to build an IL explainer. First, instead of treating cation and anion as two separate graphs, we treat them as a single undirected graph and feed the whole graph into one GNN network. Second, we substitute the final pooling layer of GNN with a global node that connects with every atom within one data point. Based on that, we develop an IL molecule explainer by combining the improved GNN framework with the subgraph-based GNN explaining methods.[56] Benefiting from the two improvements we make, the IL molecule explainer can provide an importance score insight into a single atom level within the IL molecule. We also find that our explanation method can provide a reasonable fragment importance ranking for the IL molecule in the prediction task and can be a useful tool to guide the design of new IL molecules. To the best of our knowledge, this is one of the first works in applying GNNs and the fragment importance explanation study that reaches the single atom level for CO_2 absorption in ILs.

METHODS

Figure 1 shows the overview of the whole work. IL molecule pairs are represented in three ways, which are GC, FP, and graph representation. GC and FP are combined with various machine learning models such as SVM, RF, XGBoost, and MLP to perform solubility prediction tasks. For GNNs, we utilize three GNN frameworks, which are GCN, GAT, and GIN for property prediction. Furthermore, an explanation method is developed and can provide both atom-level and fragment-level importance analysis for IL molecule pair by taking in a fine-tuned GNN model and a target IL molecule pair.

IL Data Set. The data set contains 10 117 data points of CO_2 solubility in various ionic liquid (IL) systems at different temperatures and pressure,[57] which is initially collected and published by Lei et al.[14] Each data point includes the SMILES of cation and anion, temperature, pressure, and the solubility of CO_2 (expressed as the mole fraction of CO_2 in ionic liquid). Solubilities of ILs range from 0.0000648 to 0.9516 in mole fraction, the temperatures range from 243.2 to 453.15 K, and the pressures range from 0.00798 to 499.9 bar. Cations of ILs include imidazolium, pyrrolidinium, pyridinium, piperidinium, ammonium, phosphonium, and sulfonium. Anions of ILs contain tetrafluoroborate [BF_4], chloride [Cl], dicyanamide [DCA], nitrate [NO_3], hexafluorophosphate [PF_6], thiocyanate [SCN], tricyanomethanide [$C(CN)_3$], hydrogen sulfate [HSO_4], bis (trifluoro methylsulfonyl) amide [Tf_2N], methylsulfate [$MeSO_4$], etc. Following previous works, the data set is randomly split into training and test sets by the ratio of 80 and 20%, respectively.

Descriptor-Based Machine Learning Models. *Descriptor Engineering.* We utilize the Morgan fingerprint as the descriptor for the IL molecule pairs. Morgan FP is a method for generalizing molecular signatures with molecule structure information.[41] We use RDKit to generate FP.[58] Specifically, for the generation hyperparameters, the radius is set to be 3, and the number of bits is 2048

https://doi.org/10.1021/acssuschemeng.2c05985
ACS Sustainable Chem. Eng. 2022, 10, 16681−16691

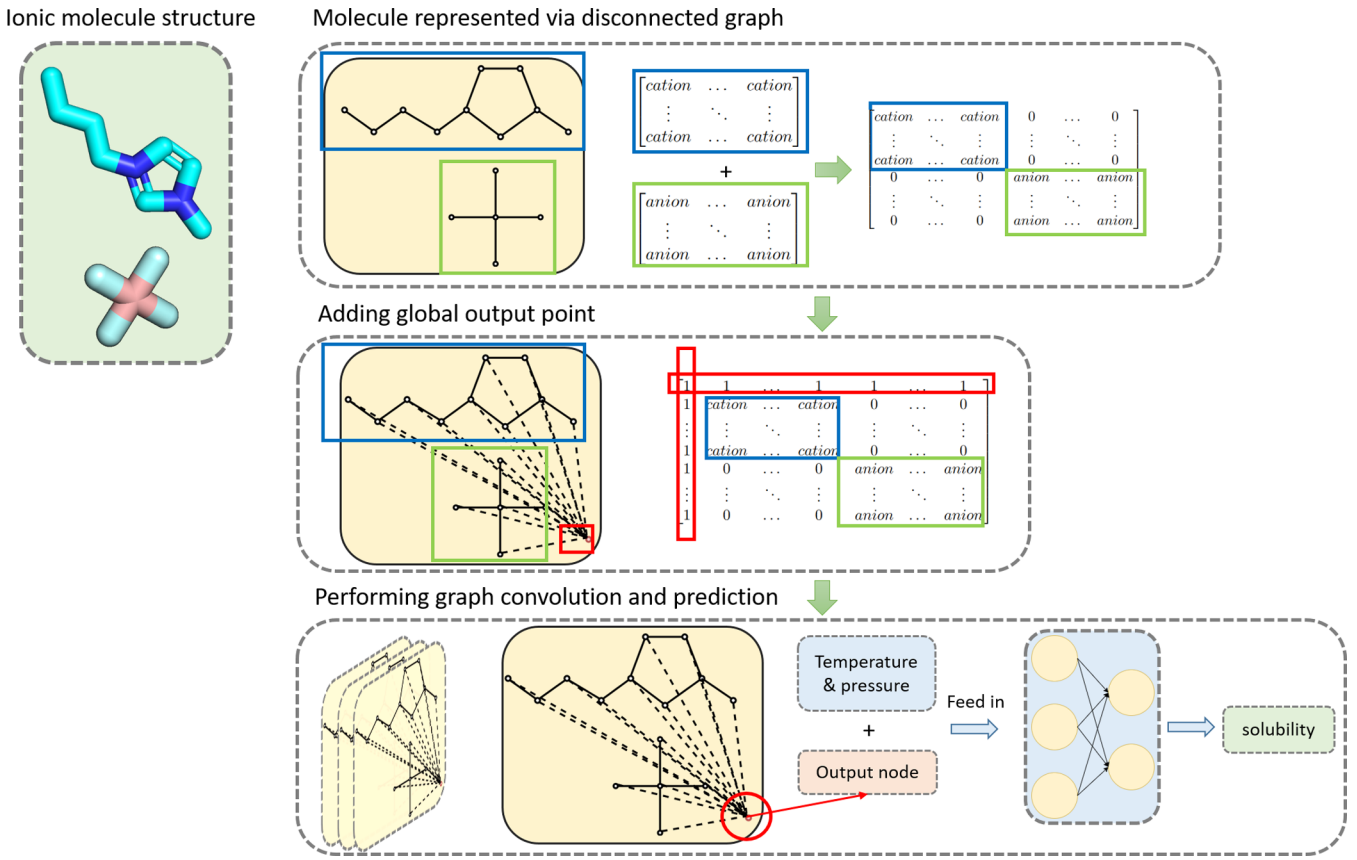


Figure 2. Illustration of how GNN predicts solubility for ionic molecule pairs. First, the cation and anion are treated as two separate graphs with an adjacent matrix. Then, we concatenate two molecule graphs diagonally. Further, a global node is added to the graph. Finally, two molecule graphs with a global node are merged into a single undirected graph. The final graph will be fed into GNN after several layers of message passing; the message of the global node is extracted and put into the final classification layer to get the solubility result.

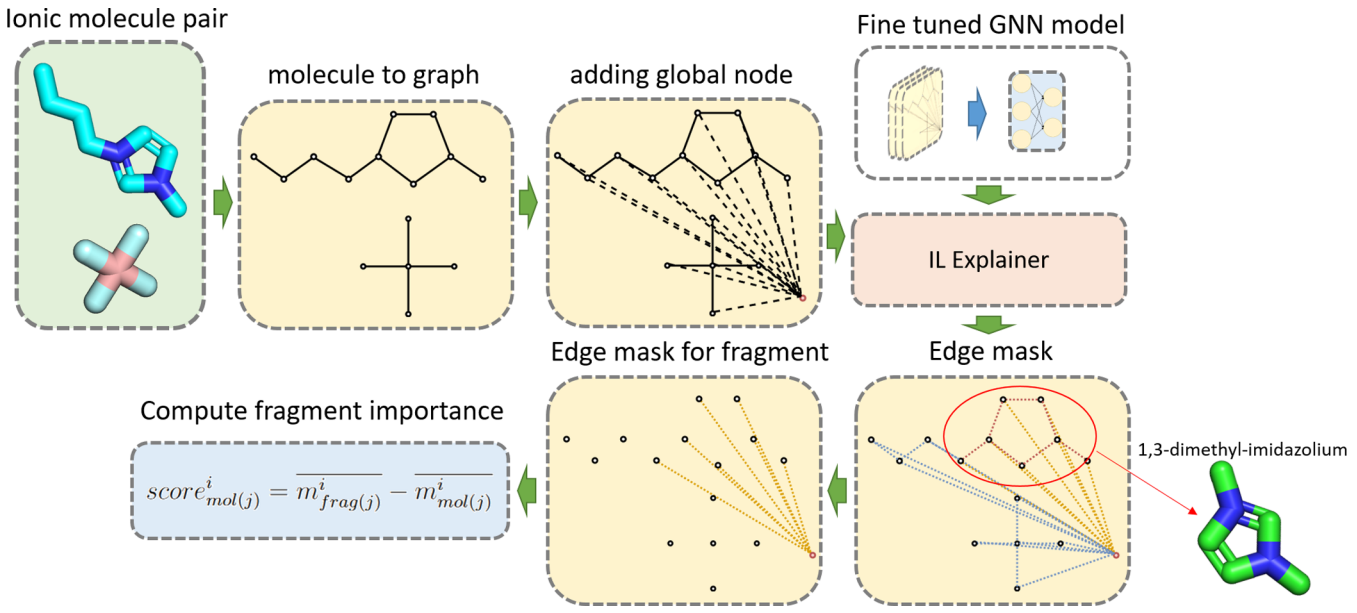


Figure 3. Making an explanation of an ionic molecule pair and extracting the fragment importance of 1,3-dimethyl-imidazolium from the edge mask. To perform the explanation, we feed the molecule graph and a trained GNN model (i.e., GCN, GAT, GIN) into the IL explainer. Then, we run the explainer to perform the optimization on the mask matrix. Finally, we get an edge importance mask for the input graph. The importance of the edge between the atom node and the global node is seen as the importance of the atom. After that, we normalize and compute the fragment importance based on the importance of atoms that the fragment includes.

for each molecule. For a certain ionic molecule pair, we first obtain the FP for each molecule and then concatenate them into a vector with a length of 4096. By adding the temperature (K) and pressure (bar), the length of the final vector is 4098. We also included GC

https://doi.org/10.1021/acssuschemeng.2c05985
ACS Sustainable Chem. Eng. 2022, 10, 16681−16691

descriptor as a baseline method to compare with FP.[57] For GC, we slice all of the ionic molecules into 51 fragments; each molecule pair is then mapped to a vector with a length of 51, and each element in the vector counts the existing frequency of each fragment in a certain molecule pair. After that, temperature and pressure are added to the vector. The length of the final vector would be 53.

Descriptor-Based Machine Learning Models. We use four ML models including SVM,[59] RF,[60] XGBoost,[61] and MLP[62] to comprehensively investigate the molecular descriptors. Those models take input as the descriptors as well as the conditions (i.e., temperature, pressure) and predict CO_2 solubility. The details and hyperparameters of the ML models can be found in the Model Details section in the Supporting Information.

Explainable GNN. In this section, we first introduce molecular graphs and GNN models for IL CO_2 solubility prediction, which is shown in Figure 2. We then conduct a post hoc interpretability study on the GNN model by generalizing an explanation for the input molecule through the GNN model shown in Figure 3.

IL Molecular Graph. Graph representation treats a molecule as a graph G with a set of nodes V and edges E.[46] In our case, given an ionic molecule, each atom is treated as a node, while bonds are edges. Further, a node matrix $H \in \mathbb{R}^{N\times M}$, where N denotes the number of nodes and M denotes the embedding dimension of node features, and an adjacent matrix $A \in \mathbb{R}^{N\times N}$ is used to represent the bond connectivity of the molecule graph. Atomic number, hybridization, aromatic, atom degree, and charge are included in atom features. Bond type, whether the bond belongs in the ring, and whether it is an aromatic bond are included in edge features. Moreover, each data point contains a cation and an anion; instead of treating both molecules separately with two graphs, we treat them as one single unconnected graph. Mathematically, we concatenate two adjacent matrices diagonally and two node matrices vertically. The way we concatenate the matrix is shown in Figure 2. Figure 2 also describes the whole process of using GNN to predict CO_2 solubility. Furthermore, to accommodate the need for IL explainer development, we substitute the final pooling layer of the model with a global node. This node takes the responsibility of outputting the final solubility and is connected to each node within the molecule graph. The details and the intuition of adding this global node will be introduced in the following section.

Graph Neural Network. Graph neural network (GNN) is a deep learning method that directly works on graph-structured data. Let G denote a graph, V denotes the set of nodes belonging to graph G, and E denote a set of edges of the graph G. In the kth layer of GNN, the model takes a set of node representation $\{h_v^{(k-1)} | v \in \mathcal{N}_{(v)}\}$ as input, then performs a feature aggregation as eq 1, and updates every node feature in the set $\{h_v^{(k-1)} | v \in \mathcal{N}_{(v)}\}$ to $\{h_v^{(k)} | v \in \mathcal{N}_{(v)}\}$. Here in eq 1, $\mathcal{N}_v$ denotes the set of neighbor nodes of node v and f_θ denotes an update function for all of the aggregated features in the layer[46,50,51,63,64]

$$h_v^{(k)} = f_\theta(h_v^{(k-1)}, \text{aggregate}(\{h | j \in \mathcal{N}_{(v)}\})) \tag{1}$$

In this work, three widely used and powerful GNN models are applied to perform CO_2 solubility prediction, which develops different aggregation and combination operations to learn from molecular graphs. GCN[46] aggregates and updates the node features by summing up all of the normalized node features of the neighbors for a single node and multiplying the sum with a learnable weight matrix. The update function is shown in eq 2

$$h_v^{(k)} = \mathbf{W}^{\mathrm{T}} \sum_{j\in\mathcal{N}_{(v)}\cup\{v\}} \left(\frac{e_{j,v}}{\sqrt{\hat{d}_j \hat{d}_v}} h_v^{(k-1)} \right) \tag{2}$$

GIN[51,63] is introduced to retain the information during the aggregation process by using an MLP as an update function, as shown in eq 3. Here, $\epsilon^{(k)}$ is a learnable parameter. Different from GCN, which uses mean pooling as the update function, GIN substitutes mean pooling with MLP. This change makes the update function from surjective to injective and thus enhances the expressiveness of the model

$$h_v^{(k)} = \text{MLP}^{(k)}\left((1 - \epsilon^{(k)}) \cdot h_v^{(k-1)} + \sum_{j\in\mathcal{N}(v)} h_j^{(k-1)} \right) \tag{3}$$

GAT[50,64] improves the expressiveness of GNNs by deploying the attention mechanism. Graph attention assigns a learnable weight for each edge when performing feature aggregation on nodes. In this way, the model can learn the weight on each edge during the training process and thus weigh each neighbor node with the difference in importance during the forward process of the model. The update function and the calculation of attention are given in eqs 4 and 5

$$h_v^{(k)} = \alpha_{v,v}\mathbf{W}_{h_v} + \sum_{j\in\mathcal{N}_{(v)}} \alpha_{v,j}\mathbf{W}_{h,j} \tag{4}$$

$$\alpha_{v,j} = \frac{\exp(\mathbf{a}^{\mathrm{T}}\ \text{leaky}Re\ LU(\mathbf{W}[h_v \| h_j]))}{\sum_{k\in\mathcal{N}_{(v)}\cup\{v\}} \exp(\mathbf{a}^{\mathrm{T}}\ \text{leaky}Re\ LU(\mathbf{W}[h_v \| h_k]))} \tag{5}$$

In this work, we include GCN, GIN, and GAT for CO_2 solubility prediction. All of the GNN models are developed with PyTorch and PyTorch Geometric.[65] Details can be found in the GNN Models section in the Supporting Information.

IL Explainer. Subgraph-based GNN explaining method provides an explanation for the importance of edges and nodes in a graph by optimizing the mask value that is added to the adjacent matrix.[56] We want to develop an explainer for the IL molecule graph based on this method so that the explainer can provide an importance score that reaches the atom and fragment levels. Let G denote a computational graph, and X denotes the node feature information. Consider the CO_2 solubility prediction as a graph classification problem, then a prediction process can be seen as $\hat{y} = \Phi(G, X)$, where Φ is the GNN model and $\hat{y}$ is the predicted result of CO_2 solubility in a specific ionic molecule pair. An important question is which fragments in the IL molecule contribute most to the CO_2 absorption in ILs. To get this insight from trained GNNs in a data-driven manner, we find a subgraph G_s, where $G_s \subseteq G$ such that G_s is important to the prediction for $\hat{y}$. Mathematically, we can formalize the above process as an optimization framework below

$$\max_{G_s} \text{MI}(Y, G_s) = H(Y) - H(Y|G = G_s) \tag{6}$$

Here, MI denotes mutual information (MI) that can quantify the change in probability for the prediction $\hat{y} = \Phi(G, X)$ between the unconditional prediction and the prediction that condition on subgraph G_s. If we find a subset G_s that maximizes the mutual information, we can say the subset has importance for the GNN prediction process. Since the entropy term $H(Y)$ is fixed as Φ is fixed for a trained GNN, our optimization process is equivalent to

$$\min_{G_s} H(Y|G = G_s) \tag{7}$$

However, it can be hard to traverse all of the subgraphs for a whole molecular graph and compute the MI because the number of the subgraph grows exponentially with the numbers of the nodes and edges. To solve the problem, instead of straightly masking off the part that does not belong to the subgraph, a learnable mask matrix with all of its element values between 0 and 1 is assigned to each edge in the graph. Based on that, we further optimize the mask value on each edge with gradient descent to maximize the mutual information. Finally, a traversal problem is reduced to an optimization problem that can be solved with gradient descent. Let $\sigma(M)$ denote the mask for the graph adjacent matrix, where M is the original mask $M \in \mathbb{R}^{n\times n}$, and sigmoid function σ is to map the element in M to (0,1). Finally, the optimization framework can be written as

$$\min_{G_s} H(Y|G_s = A \odot \sigma(M)) \tag{8}$$

https://doi.org/10.1021/acssuschemeng.2c05985
ACS Sustainable Chem. Eng. 2022, 10, 16681−16691

Here, A denotes the adjacent matrix for graph G, and the objective is to learn a mask $\sigma(M)$ such that it can minimize the target function above. In our experiment, to optimize a mask for a single molecule pair, we run 100 epochs for each pair to let the target function eq 8 converge and use the optimized mask for importance score computing.

This method provides an explanation of the importance of each edge (bond) for the molecule graph but could not provide that for node (atom). Thus, it is still challenging to derive the importance of fragments in the CO_2 solubility prediction of ILs since some of the fragments may only contain one atom and do not have edges in their molecule graph. To gain fragment importance from a node perspective, we leverage the global node added to the cation and anion pair. Since the global node feature is the representation of the IL and is connected to all atoms in the molecular graph, we can consider the importance of edges that are connected between the global node and atom as the importance of each atom to the prediction result. We can then use this atom-level importance to formalize importance analysis for a specific functional group for the whole data set further. Figure 3 illustrates the explanation process and how to extract fragment importance score from a single ionic molecule pair using 1,3-dimethyl-imidazolium as an example.

Following the idea, we developed a score function for quantifying the importance of a specific functional group with respect to the whole data set. Let M_{mol} denote a set of edge mask values (importance value) that includes all edges between the global node and each atom node within the same molecule and $\overline{M_{mol}}$ denote the mean value of the set M_{mol}. Besides, let M_{frag} denote a set of edge mask values that only contain edges between the global node and the atom for a specific functional group fragment that we are interested in and $\overline{M_{frag}}$ denotes the mean value of the set M_{frag}. We compute the importance score of a single type of fragment i within a single ionic molecule pair j as follows

$$\text{score}^{i}_{\text{mol}(j)} = \overline{M^{i}_{\text{frag}(j)}} - \overline{M^{i}_{\text{mol}(j)}} \tag{9}$$

Further, the importance score of fragment i within the whole data set score^{i} is computed as

$$\text{score}^{i} = \frac{\sum_{j=1}^{N_{\text{frag}}} \text{score}^{i}_{\text{mol}(j)}}{N_{\text{frag}}} = \frac{\sum_{j=1}^{N_{\text{frag}}} (\overline{M^{i}_{\text{frag}(j)}} - \overline{M^{i}_{\text{mol}(j)}})}{N_{\text{frag}}} \tag{10}$$

where N_{frag} denotes the number of ionic molecule pairs within the whole data set that contain the certain fragment i. By this means, the score variance is normalized between different IL pairs to an averaged importance score for each functional group. Our goal is to utilize the score function above to analyze the ranking of importance of different functional group fragments in the data set's IL molecule space. We collected a set of functional groups that exists in the cations of the data set and compute the importance ranking of those functional group. Then, we compare the ranking with the theoretical evidence for CO_2 solubility in IL to test if the IL explainer can provide a reasonable explanation for the IL molecule.

RESULTS AND DISCUSSION

CO_2 Solubility Prediction. The result of the CO_2 solubility prediction tasks is shown in Table 1. Model performance is measured by MAE and R^2 on the test set. MAE can quantify the accuracy of the predicted result; the lower MAE is the more accurate the predicted results are. R^2 quantifies the proportion of the variation in the predicted solubility from the input data, and higher R^2 demonstrates better model performance. The result shows that FP outperforms GC on SVM, RF, as well as MLP with the evaluation through MAE and R^2. Figure 4 shows that the model with FP representation also converges much faster than that with GC representation. After 300 epochs of training, FP still reaches a higher accuracy than the GC method. Though

Table 1. Performance of Different Models Combined with Different Feature Representations on CO_2 Solubility Prediction Task, MAE, and R^2 for Different Combinations Are Reported Here[a]

model	MAE ↓	R^2 ↑
SVM + GC	0.0753	0.8240
SVM + FP	0.0655	0.8633
RF + GC	0.0223	0.9774
RF + FP	0.0209	0.9802
XGBoost + GC	0.0182	0.9865
XGBoost + FP	0.0189	0.9847
MLP + GC	0.0170	0.9873
MLP + FP	0.0151	0.9883
GCN	0.0723	0.8197
GAT	0.0253	0.9767
GIN	**0.0137**	**0.9884**

[a]Noted that: to make sure the model converges, we run 300 epochs to train each model below and finally take the one with the lowest validation loss as the result (↑: the higher the better,↓: the lower the better).

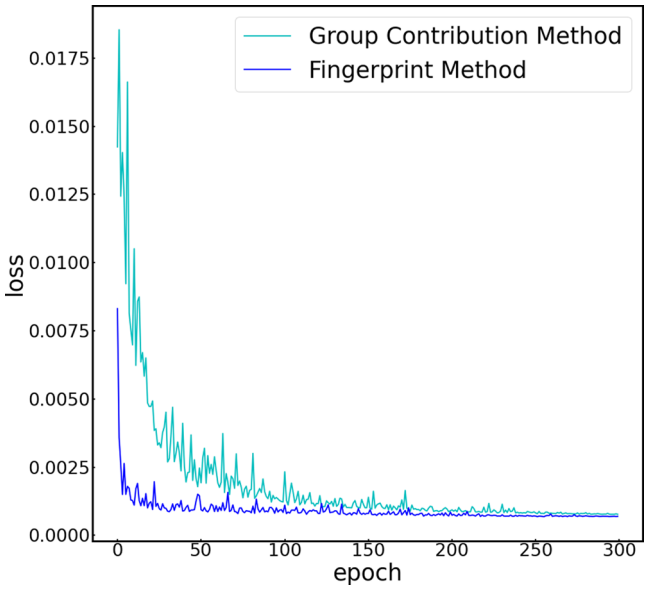


Figure 4. Validation loss-epoch curve for both GC and FP.

Morgan FP does not surpass GC using XGBoost, the differences between them on MAE and R^2 are trivial. The reason that FP generally outperforms the GC method is that the GC method can lead to information loss during the representation process since it manually sliced molecules into a fragment and ignore the structure details within a fragment. As for GNN models, GIN is the best-performing model compared with GCN and GAT. GIN also achieves the best performance among all of the ML models. First, except for GCN, GAT and GIN outperform most of the other descriptor-based machine learning groups. This proves the power of the GNN model. One possible explanation for the uncompetitive performance of GCN is that although GCN uses graph representation like the other two GNN models, the feature aggregation and update process are much simpler than the other two models; also, the update function of GCN is a surjective function. These will lead to a relative loss of information during the graph convolution process and finally result in a worse performance. Second, though both FP representation and the

https://doi.org/10.1021/acssuschemeng.2c05985
ACS Sustainable Chem. Eng. 2022, 10, 16681−16691

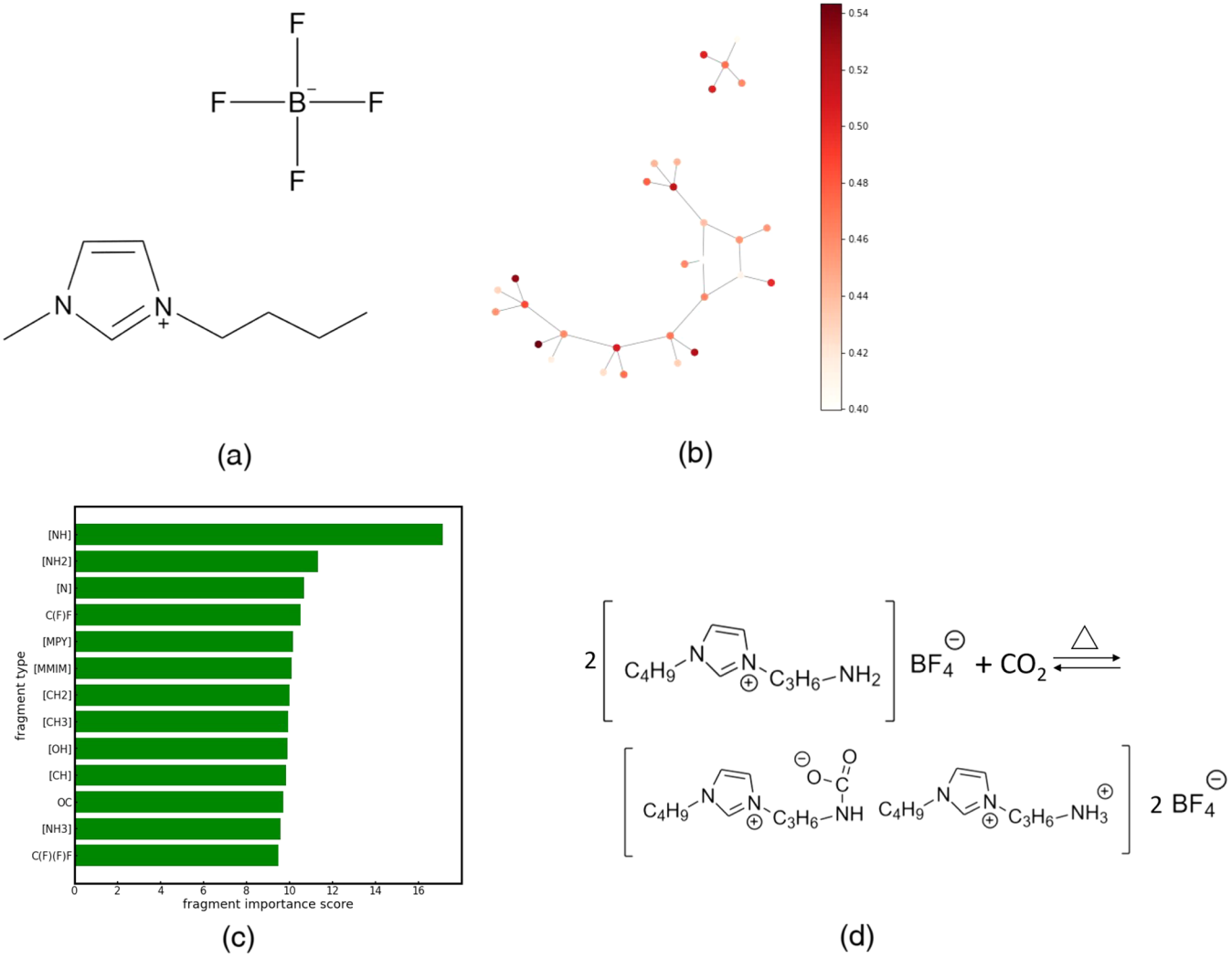


Figure 5. (a) Structure for the example molecule; here, we use [PMIM] and [BF_4]; (b) the importance hot map for [BMIM] and [BF_4]; (c) fragment importance of cations with respect to the whole data set; and (d) proposed reaction of CO_2 with [$NH_{2p\text{-}bim}$][BF_4].

GNN model work on the graph structure of the IL molecule, FP only handles graph structure at the feature representation stage, while both feature representation and the post-feature-extraction stage for GNN are all designed to handle graph-structured data. For these reasons, a GNN with good expressiveness (here like GIN) is likely to beat the descriptor-based method on the task related to molecule and this is also proved by the experimental result in Table 1.

Model Explanation Result Analysis. *Theoretical Understanding of CO_2 Solubility in ILs.* Before diving into the explanation of the machine learning model in a data-driven manner, it is necessary to check the theories about the CO_2 absorption process in ILs. Such theories validate the explanation results from our IL explainer. However, there are two major challenges in validating the data-driven explanation results. First, the existing ground truth with respect to some of the fragments distributed sparsely in various studies. Second, the factor that influences the CO_2 in solubility in a certain IL system is complex. One fragment can influence the solubility in a physical way or chemical way to different extents, which means the importance of one fragment can be composed of various mechanisms. In the following section, we summarize and categorize the theoretical explanation in the literature.

The mechanism of CO_2 solvation in IL can be divided into physical and chemical absorptions depending on the interaction between CO_2 and the ILs. Physical absorption considers anions play the primary role in CO_2 storage.[66] However, according to the experimental record, when the anion is unchanged, the fluorination of the alkyl chain and ester groups on cations is favorable for improving CO_2 solubility.[67−69] Another model based on physical absorption is called the free volume model. This model suggests that ionic molecules, especially cations with longer alkyl chains, tend to create more free volume space for trapping CO_2 inside the liquid. Thus, the carbon alkyl chain is favorable for the CO_2 storage process as well. Chemical absorption, which depends on the chemical interaction between CO_2 and IL molecules, on the other hand, usually results in a stronger combination between CO_2 and IL than physical absorption. This is because the chemical bond is usually stronger than the physical bond. Generally, the chemical bond is mainly composed of Coulomb interaction, while the major contribution of physical interaction is the van der Waals force.[69] One of the major chemical interactions for CO_2 and IL is between amine and CO_2. CO_2 tends to react with the amine in IL in a similar manner as aqueous amine by forming a carbamate salt. Amine−CO_2 interaction is one of the most strengthening

https://doi.org/10.1021/acssuschemeng.2c05985
ACS Sustainable Chem. Eng. 2022, 10, 16681−16691

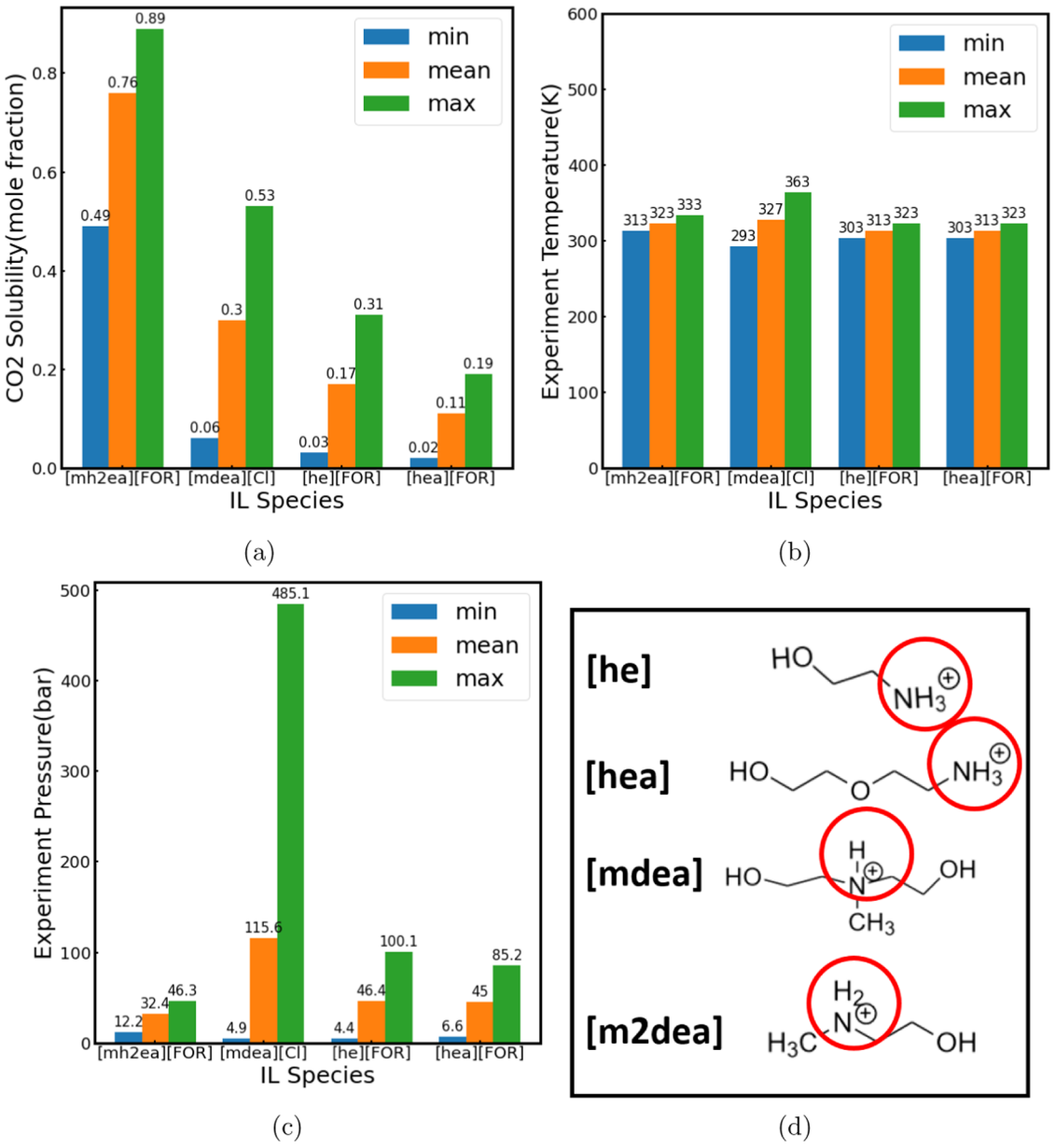


Figure 6. (a) Summary of the solubility of amine ILs, (b) summary of the temperature of amine ILs, (c) summary of the pressure of amine ILs, and (d) molecular structures of amine ILs mentioned in the figure.

interactions between IL and CO_2. There are also non-amine–CO_2 chemical interactions such as CO_2 with superbased-derived protic ILs,[70] phenolic ILs, and carboxylation of imidazolium acetate by CO_2.[71] A more specific version of reactions for IL and CO_2 can be found in the Reactions for IL and CO_2 section in the Supporting Information.

Analyzing the Explanation Result from IL Explainer. Figure 5 shows the importance analysis result from IL Explainer. Figure 5a shows an example of the molecular structure of an IL molecule pair [PMIM] (cation) and [BF_4] (anion). For this IL molecule pair, we use IL Explainer to get a molecule graph with node importance, as shown in Figure 5b. Each node (atom) importance is shown through a hot map; the importance value of each node is mapped to the color deepness. As for the data set level explanation, with the importance score function discussed in the IL Explainer section, we compute the importance score of different fragments with respect to the data set and rank them according to the score. Due to the reason that the importance of anion and cation are usually different according to the experimental record,[66] we only summarize fragments that appear in cations to make the result more comparable. Finally, we have an importance ranking of fragments existing in cations with respect to the whole data set in Figure 5c.

In the following part, the result will be discussed case by case. First, let us take a look at fragments with medium and low ranks in the result (Figure 5c). Based on the physical absorption theories, the type of interaction between most of the fragments in this region and CO_2 belongs to physical interaction. Specifically, [CH_2], [CH_3], and [CH] are the components of the alkyl chain; their contribution is creating more free volume for CO_2 to solve based on the free volume theory.[14,72] Fragments like C(F)F, C(F)(F)F, and OC are also favorable for improving the CO_2 solubility with physical interaction.[69] [OH] can form a hydrogen bond with CO_2 to strengthen their interaction. However, physical interactions are

https://doi.org/10.1021/acssuschemeng.2c05985
ACS Sustainable Chem. Eng. 2022, 10, 16681−16691

weaker than chemical interactions between ILs and CO_2. For this reason, it is hard for the fragments that forms physical interaction with ILs to reach a high rank among all fragments. To take a step further, we can see that most of the fragments with a high ranking tend to form a chemical interaction with CO_2. One of the most interesting results is the amine group; the amine group is said to have a strong chemical interaction with CO_2. Usually, when amine reacts with CO_2, one of the N−H bonds of nitrogen will be broken and the hydrogen can be substituted with O−C═O (refer to the mechanism in Figure 5d). However, in our result, [NH] and [NH_2] take up the top 2 rankings, but [NH_3] receives a very low ranking. To further explore the reason behind it, we combine the data points related to amine groups and the reaction condition to find the reason.

Figure 6 contains details of the data points related to the amine group; there are four kinds of IL pairs that contain the amine group [m2hea][FOR] (78 data points), [mdea][CL] (35 data points), [he][FOR] (27 data points), and [hea][FOR] (21 data points). Figure 6a−c includes the max, mean, min of CO_2 solubility, temperature, and pressure of data points of four kinds of IL molecule pair. Figure 6d shows the type of amine group that each cation contains, including [NH_3] in [hea] and [he], [NH] in [mdea], and [NH_2] in [mh2ea]. [hea], [he], and [mh2ea] share the same anion, and the mean temperature and pressure are close. Under those conditions, [mh2ea] shows a higher CO_2 solubility than the other two in a large amount. Considering the major structural difference of these three cations is the amine group; IL Explainer successfully captures the information where [NH_2 importance is much higher than NH_3]. Besides, although [mdea] is paired with [Cl] that is different from [FOR] in the other three IL pairs, it still shows a higher CO_2 solution than other NH_3-based IL pairs under similar average pressure and temperature conditions. Since [mdea] contains [NH] instead of [NH_3], we can see that IL Explainer also detects that [NH] bears higher importance than [NH_3]. Furthermore, if we think about this problem from an energy perspective, there is still a more fundamental reason behind it. Figure 5d shows a possible reaction for [NH_2] and CO_2. In this reaction, [NH_2] plays as a reactant, while its product contains [NH_3]. This reaction takes place with a heating condition, which means the product is more stable than the reactant in energy. In this way, we can understand [NH_3] as a more stable structure than [NH_2], and thus, it is harder for [NH_3] to react with CO_2 when compared with other amine groups with less hydrogen. The most exciting finding here is that IL Explainer captures the fact that [NH] and [NH_2] are more important than [NH_3] in CO_2 absorption for structure stability reasons from a purely data-driven manner. This demonstrates the power of IL Explainer in providing insight into a chemical problem from a data perspective and the potential on assisting the understanding and designing of new functional IL molecules in the future. Besides, IL Explainer is highly scalable. Since the importance score for each atom within the molecule is accessible, we can gain the importance of every substructure in the molecule. For example, by collecting the importance score of the alkyl chain of different lengths, we can study how the alkyl chain length affects CO_2 solubility in IL. Moreover, IL Explainer can be extended to periodic systems by considering the periodic interaction of nodes in the adjacency matrix as well. By this means, nodes that are neighboring under the periodic boundary condition are modeled in the interpretable GNN.

CONCLUSIONS

To summarize, we develop two categories of ML methods (descriptor-based ML models and GNNs) for predicting CO_2 solubility in ILs and an explanation method to detect the importance of the functional groups in a data-driven manner. For descriptor-based ML models, the result shows that FP outperforms GC in most model-descriptor combinations. Among all of the experiments in this part, MLP with FP reaches the best performance of MAE (0.0151) and R^2 (0.9883). For the GNN model, we develop GCN, GAT, and GIN models with a virtual global node. GIN reaches the best performance with MAE (0.0137) and R^2 (0.9884), which demonstrates the effectiveness of GNNs in learning from graphical data. Furthermore, for the explanation method, the IL explainer takes a trained GNN and an IL molecule data point as the input. By learning a mask to maximize the mutual information change, we can gain a node-level importance explanation. Through statistical counting and normalization, we make a fragment importance rank for cations across the whole data set. The ranking result shows that fragments that have physical interaction with CO_2 tend to have less importance than those that have chemical interaction. This can be explained as the chemical interactions are usually more powerful compared to the physical interactions. Besides that, for chemical interaction fragments, we find that amine groups with different numbers of hydrogen can be differently favorable for the absorption process. Results have shown that the amine group with less hydrogen connected to nitrogen could be more favorable in formalizing a stable chemical interaction with CO_2. The accurate ML models and importance ranking obtained from explainable GNN can provide insights into designing new functional ILs in the future.

ASSOCIATED CONTENT

Data Availability Statement

Code and data can be found through the following GitHub link: https://github.com/ftyuejian/Predicting-CO2-Absorption-in-Ionic-Liquid-with-Molecular-Descriptors-and-Explainable-GNN.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssuschemeng.2c05985.

Experiment details of descriptor-based machine learning, GNN models, prediction-label plots, and the reaction condition and mechanism mentioned in the paper (PDF)

AUTHOR INFORMATION

Corresponding Author

Amir Barati Farimani − *Department of Materials Science and Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States; Department of Mechanical Engineering and Machine Learning Department, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States;* orcid.org/0000-0002-2952-8576; Email: barati@cmu.edu

Authors

Yue Jian − *Department of Materials Science and Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States;* orcid.org/0000-0003-2464-3337

https://doi.org/10.1021/acssuschemeng.2c05985
ACS Sustainable Chem. Eng. 2022, 10, 16681−16691

Yuyang Wang − *Department of Mechanical Engineering and Machine Learning Department, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States;* orcid.org/0000-0003-0723-6246

Complete contact information is available at: https://pubs.acs.org/10.1021/acssuschemeng.2c05985

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank the start-up fund provided by the Department of Mechanical Engineering at Carnegie Mellon University. This work was also funded, in part, by the Advanced Research Projects Agency-Energy (ARPA-E), U.S. Department of Energy, under Award No. DE-AR0001221.

■ REFERENCES

(1) Stewart, C.; Hessami, M. A. A study of methods of carbon dioxide capture and sequestration - The sustainability of a photosynthetic bioreactor approach. *Energy Convers. Manage.* **2005**, *46*, 403−420.
(2) IPCC. Carbon Dioxide Capture and Storage. https://www.ipcc.ch/report/carbon-dioxide-capture-and-storage/ (accessed November 11, 2022).
(3) Mandal, B. P.; Kundu, M.; Bandyopadhyay, S. S. Physical solubility and diffusivity of N2O and CO2 into aqueous solutions of (2-Amino-2-methyl-1-propanol + monoethanolamine) and (N-methyldiethanolamine + Monoethanolamine). *J. Chem. Eng. Data* **2005**, *50*, 352−358.
(4) Barzagli, F.; Mani, F.; Peruzzini, M. A 13C NMR study of the carbon dioxide absorption and desorption equilibria by aqueous 2-aminoethanol and N-methyl-substituted 2-aminoethanol. *Energy Environ. Sci.* **2009**, *2*, 322−330.
(5) Ebner, A. D.; Ritter, J. A. State-of-the-art Adsorption and Membrane Separation Processes for Carbon Dioxide Production from Carbon Dioxide Emitting Industries. *Sep. Sci. Technol.* **2009**, *44*, 1273−1421.
(6) Zeleňák, V.; Badaničová, M.; Halamová, D.; Čejka, J.; Zukal, A.; Murafa, N.; Goerigk, G. Amine-modified ordered mesoporous silica: Effect of pore size on carbon dioxide capture. *Chem. Eng. J.* **2008**, *144*, 336−342.
(7) Chaffee, A. L.; Knowles, G. P.; Liang, Z.; Zhang, J.; Xiao, P.; Webley, P. A. CO2 capture by adsorption: Materials and process development. *Int. J. Greenhouse Gas Control* **2007**, *1*, 11−18.
(8) Serna-Guerrero, R.; Da'na, E.; Sayari, A. New insights into the interactions of CO2 with amine-functionalized silica. *Ind. Eng. Chem. Res.* **2008**, *47*, 9406−9412.
(9) Liu, N.; Bond, G. M.; Abel, A.; McPherson, B. J.; Stringer, J. Biomimetic sequestration of CO2 in carbonate form: Role of produced waters and other brines. *Fuel Process. Technol.* **2005**, *86*, 1615−1625.
(10) Druckenmiller, M. L.; Maroto-Valer, M. M. Carbon sequestration using brine of adjusted pH to form mineral carbonates. *Fuel Process. Technol.* **2005**, *86*, 1599−1614.
(11) Hu, G.; Smith, K. H.; Wu, Y.; Mumford, K. A.; Kentish, S. E.; Stevens, G. W. Carbon dioxide capture by solvent absorption using amino acids: A review. *Chin. J. Chem. Eng.* **2018**, *26*, 2229−2237.
(12) Zhang, Z.; Borhani, T. N.; Olabi, A. G. Status and perspective of CO2 absorption process. *Energy* **2020**, *205*, 118057.
(13) Dawodu, O. F.; Meisen, A. Degradation of Alkanolamine Blends by Carbon Dioxide. *Can. J. Chem. Eng.* **1996**, *74*, 960−966.
(14) Lei, Z.; Dai, C.; Chen, B. Gas solubility in ionic liquids. *Chem. Rev.* **2014**, *114*, 1289−1326.
(15) Domańska, U.; Króikowski, M.; Acree, W. E. Thermodynamics and activity coefficients at infinite dilution measurements for organic solutes and water in the ionic liquid 1-butyl-1-methylpyrrolidinium tetracyanoborate. *J. Chem. Thermodyn.* **2011**, *43*, 1810−1817.
(16) Seiler, M.; Jork, C.; Kavarnou, A.; Arlt, W.; Hirsch, R. Separation of azeotropic mixtures using hyperbranched polymers or ionic liquids. *AIChE J.* **2004**, *50*, 2439−2454.
(17) Lei, Z.; Chen, B.; Li, C.; Liu, H. Predictive molecular thermodynamic models for liquid solvents, solid salts, polymers, and ionic liquids. *Chem. Rev.* **2008**, *108*, 1419−1455.
(18) Diedenhofen, M.; Klamt, A. COSMO-RS as a tool for property prediction of IL mixtures-A review. *Fluid Phase Equilib.* **2010**, *294*, 31−38.
(19) Dong, Q.; Muzny, C. D.; Kazakov, A.; Diky, V.; Magee, J. W.; Widegren, J. A.; Chirico, R. D.; Marsh, K. N.; Frenkel, M. ILThermo: A free-access web database for thermodynamic properties of ionic liquids. *J. Chem. Eng. Data* **2007**, *52*, 1151−1159.
(20) Marsh, K. N.; Boxall, J. A.; Lichtenthaler, R. Room temperature ionic liquids and their mixtures - A review. *Fluid Phase Equilib.* **2004**, *219*, 93−98.
(21) Meyer, R.; Kvhler, J.; Wiley, A. H. Ionic Liquids in Synthesis Edited by P. Wasserscheid and T. Welton. Wiley-VCH: Weinheim. 2003. xvi + 364 pp. £80. ISBN 3-527-30515-7. *Org. Process Res. Dev.* **2003**, *7*, 223−224.
(22) Welton, T. Room-Temperature Ionic Liquids. Solvents for Synthesis and Catalysis. *Chem. Rev.* **1999**, *99*, 2071−2083.
(23) Das, R. N.; Roy, K.; Das, R. N.; Roy, K. Advances in QSPR/QSTR models of ionic liquids for the design of greener solvents of the future. *Mol. Diversity* **2013**, *17*, 151−196.
(24) Cherkasov, A.; Muratov, E. N.; Fourches, D.; et al. QSAR modeling: Where have you been? Where are you going to? *J. Med. Chem.* **2014**, *57*, 4977−5010.
(25) Koutsoukos, S.; Philippi, F.; Malaret, F.; Welton, T. A review on machine learning algorithms for the ionic liquid chemical space. *Chem. Sci.* **2021**, *12*, 6820−6843.
(26) Abdulfatai, U.; Uzairu, A.; Uba, S.; Shallangwa, G. A. Molecular design of antioxidant lubricating oil additives via QSPR and analysis dynamic simulation method. *Heliyon* **2019**, *5*, e02880.
(27) Puzyn, T.; Suzuki, N.; Haranczyk, M.; Rak, J. Calculation of quantum-mechanical descriptors for QSPR at the DFT level: Is it necessary? *J. Chem. Inf. Model.* **2008**, *48*, 1174−1180.
(28) Li, Z.; Meidani, K.; Yadav, P.; Barati Farimani, A. Graph neural networks accelerated molecular dynamics. *J. Chem. Phys.* **2022**, *156*, 144103.
(29) Faber, F. A.; Hutchison, L.; Huang, B.; Gilmer, J.; Schoenholz, S. S.; Dahl, G. E.; Vinyals, O.; Kearnes, S.; Riley, P. F.; von Lilienfeld, O. A. Machine learning prediction errors better than DFT accuracy. *J. Chem. Theory Comput.* **2017**, *13*, 5255−5264.
(30) Lawler, R.; Liu, Y. H.; Majaya, N.; Allam, O.; Ju, H.; Kim, J. Y.; Jang, S. S. DFT-Machine Learning Approach for Accurate Prediction of pKa. *J. Phys. Chem. A* **2021**, *125*, 8712−8722.
(31) Xu, C.; Wang, Y.; Farimani, A. B.TransPolymer: a Transformer-based Language Model for Polymer Property Predictions, 2022. arXiv:2209.01307. https://arxiv.org/abs/2209.01307.
(32) Valderrama, J. O.; Reátegui, A.; Rojas, R. E. Density of ionic liquids using group contribution and artificial neural networks. *Ind. Eng. Chem. Res.* **2009**, *48*, 3254−3259.
(33) Paduszyński, K.; Domańska, U. Viscosity of ionic liquids: An extensive database and a new group contribution model based on a feed-forward artificial neural network. *J. Chem. Inf. Model.* **2014**, *54*, 1311−1324.
(34) Gardas, R. L.; Coutinho, J. A. A group contribution method for viscosity estimation of ionic liquids. *Fluid Phase Equilib.* **2008**, *266*, 195−201.
(35) Matsuda, H.; Yamamoto, H.; Kurihara, K.; Tochigi, K. Computer-aided reverse design for ionic liquids by QSPR using descriptors of group contribution type for ionic conductivities and viscosities. *Fluid Phase Equilib.* **2007**, *261*, 434−443.
(36) Venkatraman, V.; Alsberg, B. K. Predicting CO2 capture of ionic liquids using machine learning. *J. CO_2 Util.* **2017**, *21*, 162−168.

(37) Eike, D. M.; Brennecke, J. F.; Maginn, E. J. Predicting melting points of quaternary ammonium ionic liquids. *Green Chem.* **2003**, *5*, 323−328.
(38) Mehrkesh, A.; Karunanithi, A. T. New quantum chemistry-based descriptors for better prediction of melting point and viscosity of ionic liquids. *Fluid Phase Equilib.* **2016**, *427*, 498−503.
(39) Venkatraman, V.; Evjen, S.; Knuutila, H. K.; Fiksdahl, A.; Alsberg, B. K. Predicting ionic liquid melting points using machine learning. *J. Mol. Liq.* **2018**, *264*, 318−326.
(40) Deng, T.; hai Liu, F.; zhu Jia, G. Prediction carbon dioxide solubility in ionic liquids based on deep learning. *Mol. Phys.* **2020**, *118*, 1652367.
(41) Morgan, H. L. The Generation of a Unique Machine Description for Chemical Structures−A Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* **1965**, *5*, 107−113.
(42) Rogers, D.; Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742−754.
(43) Atz, K.; Grisoni, F.; Schneider, G. Geometric deep learning on molecular representations. *Nat. Mach. Intell.* **2021**, *3*, 1023−1032.
(44) Wang, Y.; Magar, R.; Liang, C.; Farimani, A. B. Improving Molecular Contrastive Learning via Faulty Negative Mitigation and Decomposed Fragment Contrast. *J. Chem. Inf. Model.* **2022**, *62*, 2713−2725.
(45) Wang, Y.; Wang, J.; Cao, Z.; Barati Farimani, A. Molecular contrastive learning of representations via graph neural networks. *Nat. Mach. Intell.* **2022**, *4*, 279−287.
(46) Kipf, T. N.; Welling, M. In *Semi-Supervised Classification with Graph Convolutional Networks*, International Conference on Learning Representations, 2017.
(47) Magar, R.; Wang, Y.; Farimani, A. B.Crystal Twins: Self-supervised Learning for Crystalline Material Property Prediction, 2022. https://arxiv.org/abs/2205.01893.
(48) Yadav, P.; Mollaei, P.; Cao, Z.; Wang, Y.; Barati Farimani, A. Prediction of GPCR activity using machine learning. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 2564−2573.
(49) Karamad, M.; Magar, R.; Shi, Y.; Siahrostami, S.; Gates, I. D.; Farimani, A. B. Orbital graph convolutional neural network for material property prediction. *Phys. Rev. Mater.* **2020**, *4*, 093801.
(50) Brody, S.; Alon, U.; Yahav, E. In *How Attentive are Graph Attention Networks?*, International Conference on Learning Representations, 2022.
(51) Xu, K.; Hu, W.; Leskovec, J.; Jegelka, S. In *How Powerful are Graph Neural Networks?*, International Conference on Learning Representations, 2019.
(52) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E.In *Neural Message Passing for Quantum Chemistry*, International Conference on Machine Learning, 2017; pp 1263−1272.
(53) Ruza, J.; Wang, W.; Schwalbe-Koda, D.; Axelrod, S.; Harris, W. H.; Gómez-Bombarelli, R. Temperature-transferable coarse-graining of ionic liquids with dual graph convolutional neural networks. *J. Chem. Phys.* **2020**, *153*, 164501.
(54) Wang, W.; Yang, T.; Harris, W. H.; Gómez-Bombarelli, R. Active learning and neural network potentials accelerate molecular screening of ether-based solvate ionic liquids. *Chem. Commun.* **2020**, *56*, 8920−8923.
(55) Jiménez-Luna, J.; Grisoni, F.; Schneider, G. Drug discovery with explainable artificial intelligence. Nature Machine Intelligence. *Nat. Mach. Intell.* **2020**, *2*, 573−584.
(56) Ying, Z.; Bourgeois, D.; You, J.; Zitnik, M.; Leskovec, J. In *Gnnexplainer: Generating Explanations for Graph Neural Networks*, Advances in Neural Information Processing Systems, 2019.
(57) Song, Z.; Shi, H.; Zhang, X.; Zhou, T. Prediction of CO2 solubility in ionic liquids using machine learning methods. *Chem. Eng. Sci.* **2020**, *223*, 115752.
(58) Landrum, G.RDKit: Open-Source Cheminformatics. http://www.rdkit.org (accessed November 11, 2022).
(59) Boser, B. E.; Guyon, I. M.; Vapnik, V. N.In *A Training Algorithm for Optimal Margin Classifiers*, Proceedings of the Fifth Annual Workshop on Computational Learning Theory, New York, NY, USA, 1992; pp 144−152.
(60) Ho, T. K.In *Random Decision Forests*, Proceedings of 3rd International Conference on Document Analysis and Recognition, 1995; pp 278−282.
(61) Chen, T.; Guestrin, C.et al. In *XGBoost: A Scalable Tree Boosting System*, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016; pp 785−794.
(62) Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533−536.
(63) Hu, W.; Liu, B.; Gomes, J.; Zitnik, M.; Liang, P.; Pande, V.; Leskovec, J. In *Strategies for Pre-training Graph Neural Networks*, International Conference on Learning Representations, 2020.
(64) Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. In *Graph Attention Networks*, International Conference on Learning Representations, 2018.
(65) Fey, M.; Lenssen, J. E. In *Fast Graph Representation Learning with PyTorch Geometric*, ICLR Workshop on Representation Learning on Graphs and Manifolds, 2019.
(66) Anthony, J. L.; Anderson, J. L.; Maginn, E. J.; Brennecke, J. F. Anion effects on gas solubility in ionic liquids. *J. Phys. Chem. B* **2005**, *109*, 6366−6374.
(67) Muldoon, M. J.; Aki, S. N.; Anderson, J. L.; Dixon, J. K.; Brennecke, J. F. Improving carbon dioxide solubility in ionic liquids. *J. Phys. Chem. B* **2007**, *111*, 9001−9009.
(68) Switzer, J. R.; Ethier, A. L.; Hart, E. C.; Flack, K. M.; Rumple, A. C.; Donaldson, J. C.; Bembry, A. T.; Scott, O. M.; Biddinger, E. J.; Talreja, M.; Song, M. G.; Pollet, P.; Eckert, C. A.; Liotta, C. L. Design, Synthesis, and Evaluation of Nonaqueous Silylamines for Efficient CO2 Capture. *ChemSusChem* **2014**, *7*, 299−307.
(69) Zeng, S.; Zhang, X.; Bai, L.; Zhang, X.; Wang, H.; Wang, J.; Bao, D.; Li, M.; Liu, X.; Zhang, S. Ionic-Liquid-Based CO2 Capture Systems: Structure, Interaction and Process. *Chem. Rev.* **2017**, *117*, 9625−9673.
(70) Wang, C.; Luo, H.; Jiang, D.-E.; Li, H.; Dai, S. Carbon Dioxide Capture by Superbase-Derived Protic Ionic Liquids. *Angew. Chem.* **2010**, *122*, 6114−6117.
(71) Wang, C.; Luo, H.; Li, H.; Zhu, X.; Yu, B.; Dai, S. Tuning the Physicochemical Properties of Diverse Phenolic Ionic Liquids for Equimolar CO2 Capture by the Substituent on the Anion. *Chem. − Eur. J.* **2012**, *18*, 2153−2160.
(72) Huang, X.; Margulis, C. J.; Li, Y.; Berne, B. J. Why is the partial molar volume of CO2 so small when dissolved in a room temperature ionic liquid? Structure and dynamics of CO2 dissolved in [Bmim+][PF6−]. *J. Am. Chem. Soc.* **2005**, *127*, 17842−17851.

https://doi.org/10.1021/acssuschemeng.2c05985
ACS Sustainable Chem. Eng. 2022, 10, 16681−16691