ACS Sustainable Chemistry & Engineering

pubs.acs.org/journal/ascecg

Research Article

Artificial Intelligence-Enhanced CO_2 Capture Capacity Predictions of Functional Ionic Liquids Based on Group Contribution Descriptors

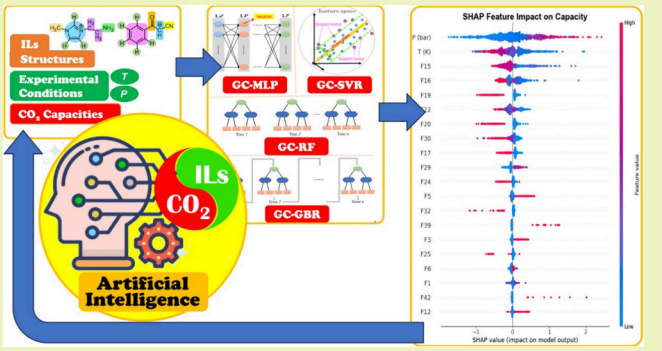
Ruina Zhang, Guokai Cui,* Wei Zhang, Kaikai Li, Lai Li, Xiangyu Shen, Chunliang Ge, Wanxiang Zhang, Wenyang Fan, Ying Zhou, Yuan Tian, Xiaopo Niu, Guoxiong Zhan, Quanli Ke, Shuhang Ren, Xiangping Zhang,* and Hanfeng Lu*

Cite This: *ACS Sustainable Chem. Eng.* 2025, 13, 19977−19994

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Ionic liquids (ILs), especially functional ILs, have attracted much attention as tunable sorbents for CO_2 capture owing to their unique properties. Given the significant role of the structure, an extensive data bank was established containing 2500 experimental capacity data points of CO_2 capture as a function of operational temperature and pressure by 232 kinds of ILs during 2002 to 2025. These total data were randomly divided into an 80% training set and 20% test set. Based on the group contribution (GC), the structures of these ILs are divided into 44 types of fragments, including ionic fragments (IFs). All 44 fragments, *T*, and *P* were used as the input parameters, while the capacity was regarded as the output parameter to establish four kinds of GC-based machine learning (ML) regression models, multilayer perceptron (MLP), support vector regression (SVR), random forest (RF), and gradient boosting regression (GBR), representing a good balance between complexity and performance. Among these models, the GC-GBR model demonstrated the strongest predictive ability. Feature importance and SHAP analysis uncover a quantitative structure−property relationship (QSPR) where abundant aliphatic groups are strong indicators of low performance, while the amine group is identified as a key promoter of high performance. This interpretability transforms the model from a black box into an interpretable tool for molecular design. This is the first time that GBR-SHAP has been used for analysis of CO_2 capture performance by functional ILs, and we hope that these GC-based ML models could be used to develop functional ILs in the future for efficient CO_2 capture.



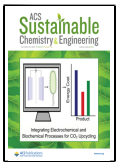
KEYWORDS: *ionic liquids, CO_2 capture, carbon neutrality, quantitative structure-property relationship, machine learning, SHAP*

1. INTRODUCTION

According to the report "State of the Global Climate 2024" recently published by the World Meteorological Organization (WMO), the CO_2 mole fraction reached a new high (417.9 ± 0.2 ppm) in 2022, while the preindustrial mole fraction of CO_2 is only 278.3 ppm.[1] It is clear that a large amount of CO_2 is emitted through the combustion of fossil fuels into the atmosphere every year, resulting in the human-caused global climate changes and leading to widespread adverse impacts on food and water security, human health, and economies and society and related losses and damages.[2] Thus, a series of technologies such as carbon capture, utilization, and storage (CCUS) have been developed as an efficient way of reducing carbon emissions during these decades.[3] Among these technologies, monoethanolamine (30 wt % in water) has been widely used as an efficient CO_2 absorbent in flue gas treatment. Due to their high CO_2 desorption and regeneration temperature (120−140 °C), alkanolamine based sorbents have high energy consumption as well as high solvent loss. Because CO_2 is a cheap and abundant C1 resource, the processes from CO_2 to value-added chemicals have been well developed through conversion and reduction.[4,5] As CO_2 capture is the key process, it is crucial to develop alternative sustainable CCUS technologies for efficient CO_2 capture.[6]

Ionic liquids (ILs) are considered as green and sustainable solvents or advanced materials due to their unique physicochemical properties, including nonflammability, wide electrochemical window, high thermal and chemical stability, and adjustable structures of cations and anions.[7] It is predicted that there are approximately 10^{18} accessible room temperature ILs, including binary and ternary mixtures.[8] Thus, ILs have been developed a lot in the recent decades as solvents,

Received: July 11, 2025
Revised: November 5, 2025
Accepted: November 6, 2025
Published: November 14, 2025



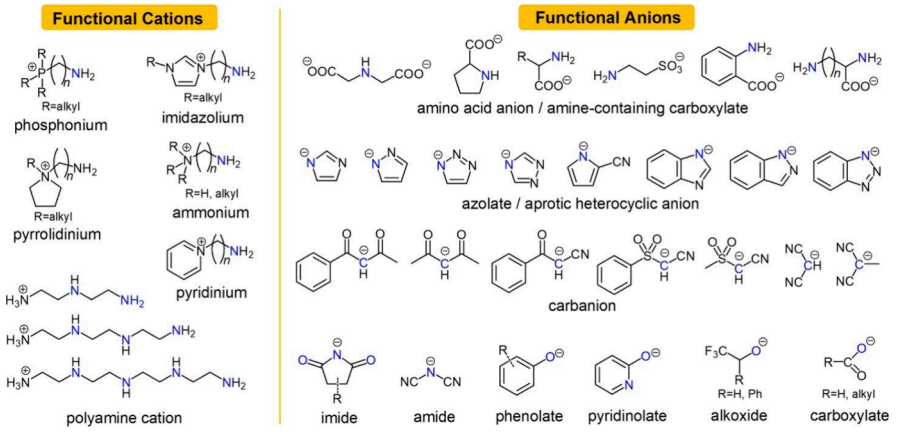


Figure 1. Typical functional cations (a) and functional anions (b).

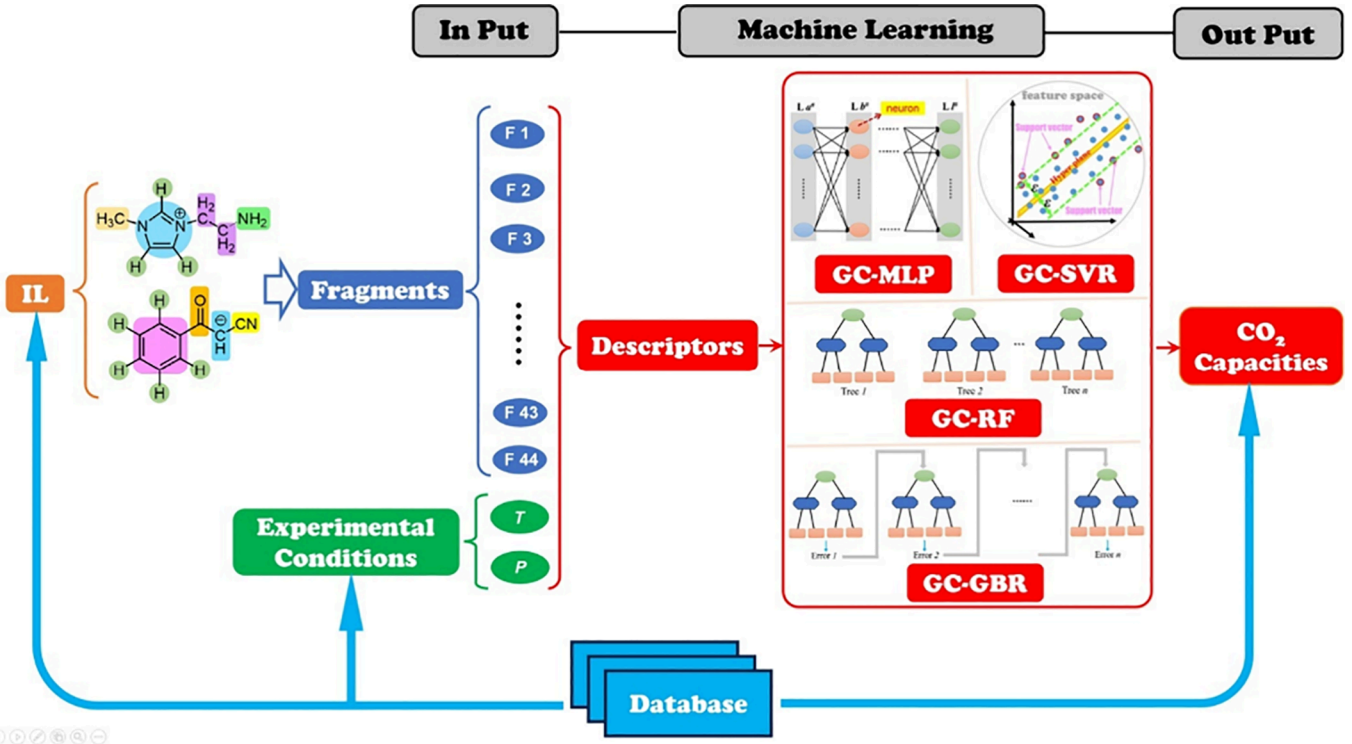


Figure 2. Schematic of generated fragments as descriptors for machine learning based on the database of CO_2 capacities of functional ILs.

sorbents, catalysts, and carriers for diverse applications in many research fields,[9] such as energy storage and conversion,[10] separation/purification and extraction,[11] chemical and materials synthesis,[12] and pharmaceutics and biomedicine.[13] The application of ILs in CO_2 capture is based on a pioneer work reported by Brennecke et al.,[14] exhibiting that CO_2 is highly soluble in ILs (in their work, 1-butyl-3-methylimidazolium hexafluorophosphate [Bmim][PF_6] was used) while the CO_2-rich phase is not significantly contaminated by ILs. Subsequently, more kinds of ILs have been developed for CO_2 or other gas separations.[15] Subsequently, various conventional ILs have been reported, which are composed of cations with different alkyl chain lengths and fluorinated anions such as halogen anions [X] (X = F, Cl, Br, or I), tetrafluoroborate ([BF_4]), hexafluorophosphate ([PF_6]), and bis(trifluoromethanesulfonyl) imide ([Tf_2N]).[16,17] However, conventional ILs improved CO_2 capacity through weak physical interaction, which is not a benefit for postcombustion flue gas CO_2 capture.[18] Compared with conventional ILs, functional ILs with active sites have been developed since 2002 for efficient CO_2 capture at low-concentration conditions or even from ambient air.[19−22] The cations can be functionalized mainly with amino groups,[23,24] while anions have a variety of structures, including amino acid anions,[25−27] alkoxide anions,[28] carboxylate anions,[29,30] azolate anions,[31−33] phenolates,[34−36] pyridinolate,[37−39] imide anions,[40,41] carbanions,[42−44] *etc.* (Figure 1). Although the experimental methods based on experiences and "trial-and-error" process are still mainstream for searching efficient ILs for CO_2 capture, its time consumption is an inevitable drawback.[45]

Along with the development of experiments, several equations and models have been developed to overcome the time-consumption drawback and identify efficient sorbents for CO_2 capture, such as the equation of state (EoS),[46] molecular

https://doi.org/10.1021/acssuschemeng.5c06870
ACS Sustainable Chem. Eng. 2025, 13, 19977−19994

dynamics (MD) simulation,[47] UNIFAC model,[48] Conductor-like Screening Model for Real Solvents (COSMO-RS),[49] machine learning (ML),[50] *etc*. Among these prediction methods, ML, the core of artificial intelligence (AI), performs better in high-dimensional data classification and prediction and has the potential to significantly save costs and improve efficiency. By applying ML to the study of CO_2 capture by ILs, computers can simulate the linear or nonlinear relationship between the structure of ILs and CO_2 capture performance, as well as predict the CO_2 capture performance by ILs with specific structures under specific conditions.[50] Many ML models, including artificial neural network (ANN), support vector machine (SVM), *etc*., have been built for CO_2 capacity prediction based on physical property descriptors of ILs, such as molecular weight, acentric factor, critical temperature, critical pressure, and critical compressibility factor.[51−53] For comparison, models based on molecular structure descriptors only need the structures of ILs and there is no need for the aforementioned physical properties that may be unmeasurable for most ILs.[54] In order to study the quantitative structure−property relationship (QSPR), ionic fragment contribution based on group contribution (GC) was proposed by Huang et al.[55] and has been used for building ML models for gas capture,[56] including CO_2 absorption,[57] by dividing the structures of ILs into fragments, including ionic fragments (IFs) and neutral groups. Besides, the experimental data set affects the accuracy of these data-driven models and the results of judgments. However, the reported data on CO_2 capture capacity by functional ILs have yet to be fully leveraged for building high-accuracy ML models. The development of such models presents a unique opportunity and challenge as they must extract meaningful patterns from a limited number of examples.

In this work, four ML models based on the multilayer perceptron (MLP) neural network, support vector regression machine (SVR), random forest (RF), and gradient boosting regression (GBR) have been constructed using molecular structure descriptors to predict the CO_2 capture by functional ILs (Figure 2). First, we build an experimental data set based on the summarizing of functional ILs for CO_2 capture under different temperatures and partial pressures in 2002 to 2025. This extensive data bank contains 2500 experimental capacity data points of CO_2 capture as a function of operational temperatures and pressures by 232 kinds of ILs including 44 cations and 123 anions. The total experimental data were randomly divided into two data subsets, a training set with 80% of total data and a test set with 20% of total data. Based on the group contribution (GC), the structures of these ILs are divided into 44 types of fragments. All 44 GC-based molecular descriptors, temperature (T), and pressure (P) were used as the input parameters, and the capture capacity was regarded as the desired output parameter to establish the GC-MLP, GC-SVR, GC-RF, and GC-GBR models. With feature importance and SHAP analysis, these ML models provide an opportunity to establish a quantitative structure−property relationship between the capture capacity of CO_2 and the structure of functional ILs.

2. METHODOLOGY

Figure 3 illustrates the overall methodology of building machine learning models in this work. First, data are gathered from the literature, including the structures of ILs, the operation temperature (T) and partial pressure (P), and the corresponding CO_2 capture capacities. Then, group contribution (GC) was selected as the molecular structure descriptors. The input parameters consist of structure descriptors and condition descriptors (T, P). The multilayer perceptron (MLP) neural network, support vector regression (SVR), random forest (RF), and gradient boosting regression (GBR) are selected as the machine learning models. After training the models with a training data set and optimizing models with adjusting parameters, the machine learning models are built for predicting CO_2 capture capacity at given operation conditions for ILs (Figure S1).

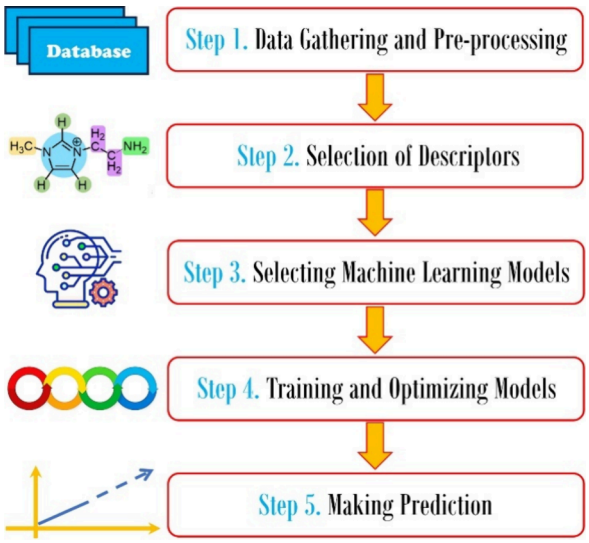


Figure 3. Overall methodology of building machine learning models.

2.1. Data Gathering and Preprocessing. Based on the summarizing of functional ILs for CO_2 capture, the experimental data set with 2500 experimental data points about the capture capacity of CO_2 by 232 kinds of ILs including 44 cations and 123 anions were first collected from 59 literatures reported from 2002 to 2025. The summary of this database is presented in Table 1. The structures and abbreviations of 44 cations and 123 anions are listed in Figure S2, while the full name of these 232 kinds of ILs can be found in Table S1. Detailed information can be found in the Excel file of the Supporting Information. To prevent features with wide ranges from dominating the model and expedite the algorithm's convergence, a data preprocessing strategy was implemented for standardization.[58] The standardized data aligns with the standard normal distribution using the equation provided below:[56]

$$x_{scaler} = \frac{x - \mu}{\sigma}$$

where x represents the original data, x_{scaler} is the preprocessed data, μ is the mean value of samples, and σ is the standard deviation of samples. In application, the StandardScaler (a Python function) was initially fitted to the training set, and two parameters (μ and σ) were readily obtained. Then, these same parameters (μ and σ) were utilized to transform both the training and the test sets. The values of μ and σ of samples are listed in Table S2. This method guaranteed that the models were trained and tested by using consistent standardized criteria, enhancing the reliability and comparability of the results. The training set encompassed 2000 data points (80% of total data points), whereas the test set comprised 500 data

https://doi.org/10.1021/acssuschemeng.5c06870
ACS Sustainable Chem. Eng. 2025, 13, 19977−19994

Table 1. Identification (ID) of IL, the Abbreviations of ILs, and the Number of Data Points (N) for Each CO_2−IL System

ID	IL	N	ID	IL	N	ID	IL	N
1	[Bis(mim)C2][Gly]2	16	79	[N2224][4-Br-Im]	1	157	[P66614][2-Op]	19
2	[Bis(mim)C2][Im]2	14	80	[N2224][Ala]	1	158	[P66614][2-SCH3-BenIm]	19
3	[Bis(mim)C2][Pro]2	14	81	[N2224][BenIm]	4	159	[P66614][3-CF3-Pyrz]	41
4	[Bis(mim)C4][Gly]2	16	82	[N2224][Et2NCOCHCN]	1	160	[P66614][3-CH3−5-CF3-Pyrz]	10
5	[Bis(mim)C4][Im]2	57	83	[N2224][Im]	1	161	[P66614][3-Cl-PhO]	1
6	[Bis(mim)C4][Pro]2	15	84	[N2224][β-Ala]	1	162	[P66614][3-HMPz]	1
7	[Bmim][(CH3)2CHCOO]	9	85	[N4442][2-Op]	1	163	[P66614][3-N(CH3)2-PhO]	1
8	[Bmim][(CH3)3CCOO]	9	86	[N66614][Arg]	1	164	[P66614][3-OCH3−2-Op]	1
9	[Bmim][2-Op]	1	87	[N66614][Asn]	1	165	[P66614][3-Op]	1
10	[Bmim][Ac]	9	88	[N66614][Gln]	1	166	[P66614][4,5−2EF-Im]	1
11	[Bmim][Ala]	61	89	[N66614][His]	1	167	[P66614][4,5-CN-Im]	1
12	[Bmim][Arg]	4	90	[N66614][Lys]	1	168	[P66614][4-Br-Im]	1
13	[Bmim][CH3CH2COO]	9	91	[N66614][Met]	1	169	[P66614][4-CF3-PhO]	1
14	[Bmim][CH3COCH2CH2COO]	19	92	[NH2-emim][BF4]	18	170	[P66614][4-CH3-Im]	1
15	[Bmim][Gly]	56	93	[NH2-emim][CF3SO3]	18	171	[P66614][4-CH3O-PhO]	1
16	[Bmim][His]	4	94	[NH2-emim][N(CN)2]	18	172	[P66614][4-CH3-PhO]	1
17	[Bmim][IAAc]	9	95	[NH2-emim][PF6]	18	173	[P66614][4-CHO-Im]	10
18	[Bmim][Im]	13	96	[NH2-emim][Tf2N]	18	174	[P66614][4-CHO-PhO]	1
19	[Bmim][Leu]	3	97	[NH2-emmim][Tau]	19	175	[P66614][4-Cl-PhO]	3
20	[Bmim][Lys]	3	98	[NH2-pbim][BF4]	1	176	[P66614][4-EF-PhO]	1
21	[Bmim][Met]	4	99	[NH2-pmim][Im]	13	177	[P66614][4-Kt-PhO]	1
22	[Bmim][NH2COCH2CH2COO]	9	100	[OH-emim][Lys]	1	178	[P66614][4-NO2-Im]	1
23	[Bmim][Pro]	5	101	[OH-emmim][PhO]	1	179	[P66614][4-NO2-PhO]	1
24	[Bmim][TFA]	9	102	[P22212][2-CN-Pyrro]	10	180	[P66614][4-Op]	1
25	[Bmim][Val]	52	103	[P2224][1,2,4-Triz]	6	181	[P66614][5-HMPz]	1
26	[Bmim]2[IDA]	10	104	[P2224][2-CN-Pyrro]	7	182	[P66614][6-Br-BenIm]	50
27	[C1C4Pyr][Ac]	2	105	[P2228][2-CN-Pyrro]	15	183	[P66614][Ala]	3
28	[C1C4Pyr][2-Op]	1	106	[P2228][3-CF3-Pyrz]	7	184	[P66614][BenIm]	38
29	[Cho][Arg]	44	107	[P2228][3-CH3−5-CF3-Pyrz]	7	185	[P66614][BenTriz]	2
30	[Cho][Gln]	44	108	[P2228][4-NO2-Pyrz]	5	186	[P66614][CH(CN)2]	1
31	[Cho][Glu]	44	109	[P44412][2-CN-Pyrro]	8	187	[P66614][CH3C(CN)2]	1
32	[Cho][His]	44	110	[P44418][2-CN-Pyrro]	8	188	[P66614][CH3COCHCOCH3]	1
33	[Cho][Pro]	44	111	[P4442][2-Op]	1	189	[P66614][CH3SO2CHCN]	1
34	[Cho][Tyr]	44	112	[P4442][BenIm]	1	190	[P66614][Et2NCOCHCN]	6
35	[DEEDAH][Ac]	1	113	[P4442][CH(CN)2]	1	191	[P66614][Gly]	15
36	[DEEDAH][HCOO]	1	114	[P4442][DAA]	2	192	[P66614][Ile]	9
37	[DMAPAH][1,2,4-Triz]	1	115	[P4442][D-Ser-H]	1	193	[P66614][Im]	2
38	[DMAPAH][Ac]	1	116	[P4442][Gly]	1	194	[P66614][Ind]	40
39	[DMAPAH][HCOO]	1	117	[P4442][H-IDA]	1	195	[P66614][iPrCOCHCN]	1
40	[DMAPAH][Im]	1	118	[P4442][Im]	1	196	[P66614][Lys]	11
41	[DMAPAH][Pyrz]	1	119	[P4442][L-Ser-H]	1	197	[P66614][MA-Tetz]	1
42	[DMEDAH][1,2,4-Triz]	2	120	[P4442][Ph-Suc]	1	198	[P66614][Met]	14
43	[DMEDAH][HCOO]	1	121	[P4442][Suc]	26	199	[P66614][NH2CH2SO3]	1
44	[DMEDAH][Im]	2	122	[P4442]2[D-Ser]	1	200	[P66614][o-AA]	4
45	[DMEDAH][Pyrz]	2	123	[P4442]2[EDDA]	1	201	[P66614][o-ANA]	1
46	[Eeim][Ac]	52	124	[P4442]2[IDA]	31	202	[P66614][p-AA]	5
47	[Emim][1,2,4-Triz]	1	125	[P4442]2[l-HSer]	1	203	[P66614][p-ANA]	1
48	[Emim][Ac]	9	126	[P4442]2[L-Ser]	1	204	[P66614][PCCPhO]	1
49	[Emim][Im]	12	127	[P4442]3[NTA]	1	205	[P66614][PhCOCHCN]	1
50	[Emim][TFA]	9	128	[P4442-OH][2-Op]	1	206	[P66614][PhCOCHCOCH3]	1
51	[EMIM]10[DCP5]	1	129	[P4444][2-F-PhO]	36	207	[P66614][PhO]	4
52	[HO-emim][Im]	13	130	[P4444][3-F-PhO]	49	208	[P66614][PhSO2CHCN]	1
53	[Me2N(CH2CH2OH)2][Tau]	2	131	[P4444][4-F-PhO]	39	209	[P66614][PNNPhO]	13
54	[Me-C4mim][2-Op]	1	132	[P4444][Ac]	1	210	[P66614][PPhO]	1
55	[Me-C6mim][2-Op]	1	133	[P4444][Ala]	1	211	[P66614][Pro]	103
56	[Me-C8mim][2-Op]	1	134	[P4444][Bic]	1	212	[P66614][Pyrz]	1
57	[N1,1,10,2OH][Threo]	3	135	[P4444][CH(CN)2]	1	213	[P66614][Sar]	5
58	[N1,1,2OH,2OH][1,2,4-Triz]	24	136	[P4444][CH3CH2CH2COO]	1	214	[P66614][Tau]	6
59	[N1,1,4,2OH][Threo]	3	137	[P4444][CH3CH2COO]	1	215	[P66614][Tetz]	1
60	[N1,1,6,2O4][Lys]	3	138	[P4444][dmGly]	1	216	[P66614][β-Ala]	1
61	[N1,1,6,2OH][Tau]	3	139	[P4444][Gly]	1	217	[P66614]10[DCP5]	8

https://doi.org/10.1021/acssuschemeng.5c06870
ACS Sustainable Chem. Eng. 2025, 13, 19977−19994

Table 1. continued

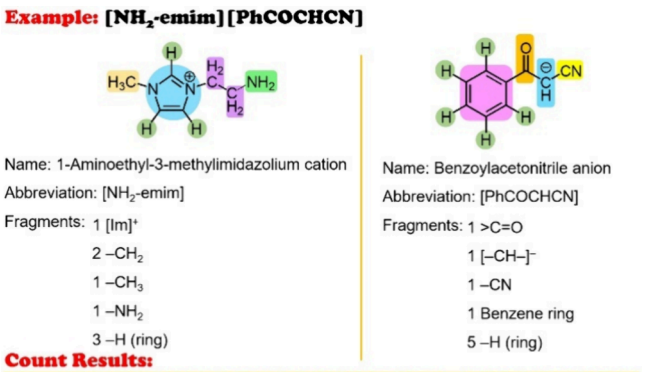
ID	IL	*N*	ID	IL	*N*	ID	IL	*N*
62	[N1,1,7,2OH][Threo]	3	140	[P4444][HCOO]	10	218	[P66614]2[AA-Ac]	1
63	[N1111][1,2,4-Triz]	1	141	[P4444][Ile]	1	219	[P66614]2[AA-Im]	1
64	[N1111][CH(CN)2]	13	142	[P4444][PhO]	43	220	[P66614]2[AA-Su]	1
65	[N1111][CH3NH(CH2)3COO]	14	143	[P4444][Pro]	1	221	[P66614]2[Am-iPA]	1
66	[N1111][Gly]	1	144	[P4444][Val]	1	222	[P66614]2[Asp]	25
67	[N1111][NH2(CH2)3COO]	10	145	[P4446][BenIm]	1	223	[P66614]2[DC]	8
68	[N1111][NH2(CH2)5COO]	39	146	[P4446][Et2NCOCHCN]	1	224	[P8884][2-Op]	1
69	[N111H][CH3CH2CH2COO]	55	147	[P4446][Im]	1	225	[Ph−C8eim][2-Op]	7
70	[N2221][Ala]	1	148	[P66614][1,2,4-Triz]	36	226	[VBIm][Ala]	44
71	[N2222][1,2,4-Triz]	1	149	[P66614][1-HDMPz]	1	227	[VBIm][Arg]	44
72	[N2222][Ac]	1	150	[P66614][1-HMPz]	1	228	[VBIm][Gly]	44
73	[N2222][Ala]	1	151	[P66614][1-Naph]	1	229	[VBIm][His]	44
74	[N2222][CH(CN)2]	53	152	[P66614][2,4,6-Cl-PhO]	1	230	[VBIm][Lys]	44
75	[N2222][PhCOO]	1	153	[P66614][2,4-Cl-PhO]	1	231	[VBIm][Pro]	44
76	[N2222][PhO]	1	154	[P66614][2-Cl-PhO]	1	232	[VBIm][Val]	44
77	[N2222][*β*-Ala]	1	155	[P66614][2-CN-Pyrro]	103			
78	[N2224][1,2,3-Triz]	1	156	[P66614][2-Naph]	1			

Figure 4. Fragment codes, abbreviations, and their relevant structures.

points (20%), with the value of the random state set to be 512. This division ensured no overlap between the training and test data, preventing potential data leakage, and a specific random state seed was assigned to guarantee reproducibility during data splitting.

2.2. Construction of GC-Based Feature Descriptors. Based on the group contribution (GC), the structures of functional ILs applied for CO_2 capture could be divided into 44 fragments, including ionic fragments (IFs), in this work. The fragment codes, abbreviations, and their relevant structures are given in Figure 4. Taking [NH_2-emim]-[PhCOCHCN] as an example, its structure could be divided into 9 fragments: 1 $[Im]^+$, 2 $-CH_2-$, 1 $-CH_3$, 1 $-NH_2$, and 3 −H (ring) for the cation [NH_2-emim], and 1 >C=O, 1 $[-CH-]^-$, 1 −CN, 1 benzene ring, and 5 −H (ring) for the anion [PhCOCHCN], as illustrated in Figure 5. Therefore, the total fragments for $[Im]^+$ (F 1), −CN (F 12), $-CH_3$, (F 15), $-CH_2-$ (F 16), $[-CH-]^-$ (F 17), −H (ring) (F 22), (F 27),

https://doi.org/10.1021/acssuschemeng.5c06870
ACS Sustainable Chem. Eng. 2025, 13, 19977−19994



Count Results:

Fragment	F 1	F 12	F 15	F 16	F 17	F 22	F 27	F 29	F 30
Cation	1	0	1	2	0	3	0	1	0
Anion	0	1	0	0	1	5	1	0	1
Total	1	1	1	2	1	8	1	1	1

Figure 5. Example of a fragment division.

benzene ring (F 29), and >C=O (F 30) of this IL are 1, 1, 1, 2, 1, 8, 1, 1, and 1, respectively.

2.3. Construction of GC-Based ML Models. *2.3.1. GC-RF Model.* The RF model combines multiple independent decision trees and constructs a forest using a random approach, which enhances the model's generalization ability. For regression, the final result is the average of each tree's output (Figure 6a). The RF model includes five key characteristics: first, RF enhances generation capability while reducing load and time; second, it is robust to missing data and maintains accuracy; third, it reduces overfitting risk due to the use of multiple decision trees; fourth, it provides variable importance estimates, including which features are important for prediction. Thus, RF has gained significant traction in the process industry, environment, and structure–property relationship-based property prediction over the past decades. However, increasing the number of trees in the forest resulted in longer prediction times.

2.3.2. GC-GBR Model. GBR is another type of ensemble model, and its fundamental component is also decision trees. Actually, GBR builds trees sequentially, meaning that GBR generates trees one after another, and each tree corrects the errors of its predecessor (Figure 6b). Generally, GBR exhibits four distinct characteristics compared to RF: first, GBR operates sequentially, whereas RF operates in parallel; second, GBR's calculation speed is generally slower than RF's; third, GBR is more susceptible to parameter adjustments, outliers, and noise; fourth, GBR provides important variable estimations due to its high approximation accuracy and execution speed. Thus, GBR has been extensively used for classification and regression tasks and in establishing prediction relationships for structure–property applications.

2.3.3. GC-MLP Model. The MLP model is a feed-forward ANN. It comprises an input layer, one or more hidden layers, and an output layer. Data enters through the input layer, undergoes processing in the hidden layers, and results in predictions at the output layer (Figure 6c). Here, there are 46 neurons in the input layer, including 44 fragments, temperature, and pressure, while one neuron is in the output layer. Because two hidden layers can already satisfy the requirement of using activation functions to represent an arbitrary decision boundary with arbitrary accuracy as well as approximating any

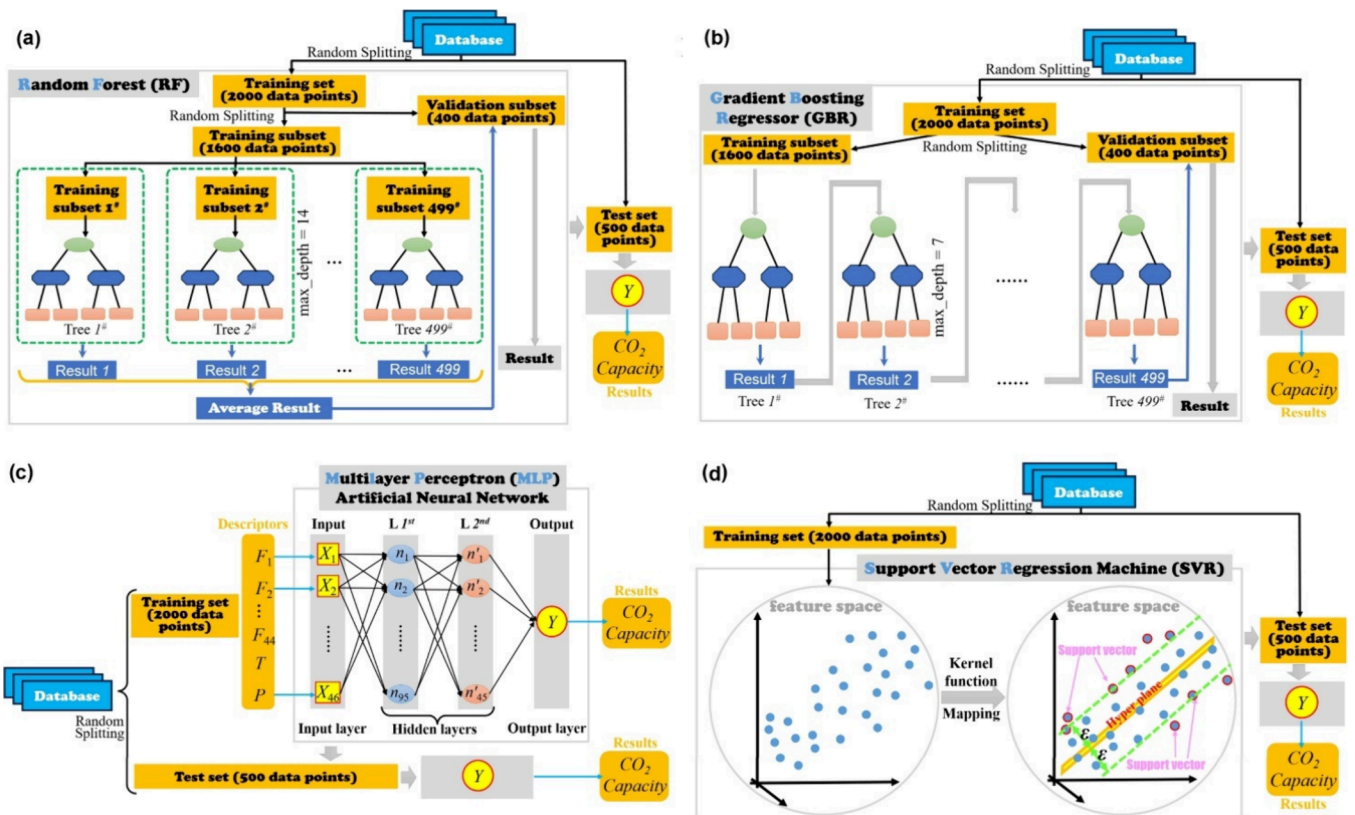


Figure 6. Schematic diagrams of the GC-based optimized ML models. (a) GC-RF, (b) GC-GBR, (c) GC-MLP with two hidden layers and ReLU activation, and (d) GC-SVR.

https://doi.org/10.1021/acssuschemeng.5c06870
ACS Sustainable Chem. Eng. 2025, 13, 19977−19994

smooth mapping to any precision, an MLP model with a (46-N_{L1}-N_{L2}-1) architecture is used to predict the CO_2 capture capacity of ILs, where N_{L1} and N_{L2} are the neuron numbers for the first hidden layer (L_1) and the second hidden layer (L_2), respectively (details see the Supporting Information).

2.3.4. GC-SVR Model. The SVR model is grounded on the principles of SVM, tailored to regression tasks. The fundamental concept of SVR involves using mathematical operations (kernel function) and creating nonlinear decision boundaries, resulting in the disordered nonlinear points that are rearranged into linear and divisible points (Figure 6d). Thus, a hyperplane is obtained, and a majority of data points reside within a specific margin from the plane. The rearrangement process from a lower-dimensional nonlinear space to a higher-dimensional linear space is termed mapping. Subsequently, the regression task is carried out in the reorganized, higher-dimensional linear space. The SVR model, including a mapping and kernel function, is described in detail in the Supporting Information.

2.4. Regression Evaluation Metrics and Parameter Optimization. The coefficient of correlation (R^2), mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and average absolute relative deviation (AARD) are five kinds of statistical parameters that have been adopted as a true test for evaluating these models' performance, and these metrics could be given by the following equations,

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_{pred}(i) - y_{exp}(i))^2}{\sum_{i=1}^{N} (y_{pred}(i) - \overline{y}_{exp})^2}$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_{pred}(i) - y_{exp}(i)|$$

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (y_{exp}(i) - y_{pred}(i))^2$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_{exp}(i) - y_{pred}(i))^2}$$

$$AARD\% = \frac{100\%}{N} \sum_{i=1}^{N} \left| \frac{y_{pred}(i) - y_{exp}(i)}{y_{exp}(i)} \right|$$

where i and N are the i^{th} data point and the total number of data points, respectively, and y_{exp}, y_{pred}, and $-y_{exp}$ are experimental value, predicted value, and average experimental value, respectively. A good machine learning model will give an R^2 that is very close to 1, while its MAE, MSE, and RMSE values are close to 0. It is known that, not affected by extreme errors, MAE is robust and less sensitive to outliers than MSE or RMSE, which give more weight to large errors than small ones. It should be noted that the units of R^2, MAE, MSE, RMSE, and AARD% are dimensionless, mol/kg, $(mol/kg)^2$, mol/kg, and %, respectively.

R^2 Gap (ΔR^2), MSE Ratio, and AARD% Increase are three indicators for evaluating the degree of overfitting. ΔR^2 and AARD% Increase refer to the difference in performance between the training set and test set, while the MSE Ratio refers to the ratio of Training MSE to Test MSE. These metrics can be given by the following equations:

$$R^2\text{Gap}(\Delta R^2) = \text{Training}R^2 - \text{Test}R^2$$

$$\text{MSE Ratio} = \text{Training MSE}/\text{Test MSE}$$

$$\text{AARD\%Increase} = \text{Test AARD\%} - \text{Train AARD\%}$$

A small gap of R^2 or AARD% (close to 0) indicates that the model's performance is consistent between the training and test sets, indicating robust performance. A ratio of MSE (close to 1) suggests that the model's error is similar on both data sets, indicating good generalization.

Parameter tuning is a crucial process in machine learning and model performance optimization. There are several methods for parameter optimization, and grid and random searches are two common approaches used in model evaluation. Figure 7 illustrates the grid search approach (in two-dimensional space), random search approach (in two-dimensional space), and 5-fold cross-validation for each combination. The grid search operates on the principle of exhaustively searching through a predefined set of parameter values in order to determine which combination yields the highest performance metric, while the random search explores the parameter space by sampling a fixed number of parameter combinations from specified distributions, offering a more efficient approach, particularly in high-dimensional spaces where the exhaustive search becomes computationally prohibitive. Thus, the grid search employs a brute-force approach and requires a significant amount of time to execute, especially for a model with too many parameters and a large search space. Besides, it is crucial to define the parameter ranges within which each parameter will vary. These systematic approaches guarantee the best results. Finally, the combination of parameters that yields the best score is selected and can be used for testing.

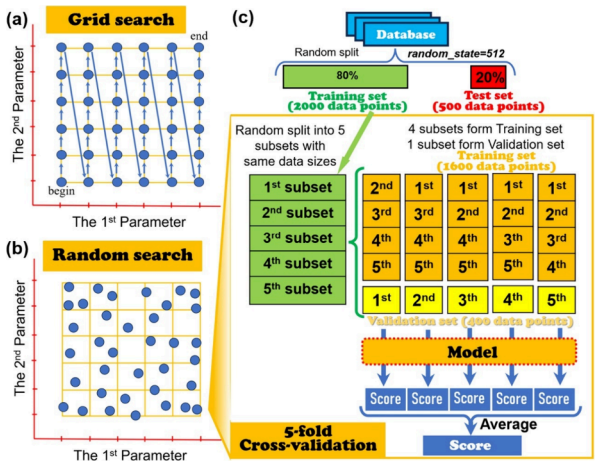


Figure 7. An illustration of the grid search approach (a), random search approach (b), and 5-fold cross-validation (c) for each combination.

During the modeling process, k-fold cross-validation is employed as an additional step to enhance the result reliability (Figure 7b). This method involves initially dividing the data into n equal groups. Subsequently, $n - 1$ groups are used for training, while the remaining group is utilized for testing in the model training process. This process is iterated n times to ensure that all data is trained, ultimately yielding the MSE score for the training set. This entire procedure is then repeated k times with each repetition based on a different

https://doi.org/10.1021/acssuschemeng.5c06870
ACS Sustainable Chem. Eng. 2025, 13, 19977−19994

Table 2. Number of Data Points (N) and Corresponding Percentage in Different Partial Pressure (P, in bar) Ranges, Temperature (T, in K) Ranges, and Absorption Capacity Ranges (Z, in mol/kg) from the Literature

P (bar)	N	percent (%)	T (K)	N	percent (%)	Z (mol/kg)	N	percent (%)
$0 < P \leq 1$	1596	63.84	$280 < T \leq 290$	207	8.28	$0 < Z \leq 1$	870	34.8
$1 < P \leq 2$	315	12.6	$290 < T \leq 300$	733	29.32	$1 < Z \leq 2$	934	37.36
$2 < P \leq 3$	171	6.84	$300 < T \leq 310$	371	14.84	$2 < Z \leq 3$	418	16.72
$3 < P \leq 4$	169	6.76	$310 < T \leq 320$	567	22.68	$3 < Z \leq 4$	177	7.08
$4 < P \leq 5$	41	1.64	$320 < T \leq 330$	229	9.16	$4 < Z \leq 5$	59	2.36
$5 < P \leq 6$	25	1	$330 < T \leq 340$	207	8.28	$5 < Z \leq 6$	20	0.8
$6 < P \leq 7$	31	1.24	$340 < T \leq 350$	79	3.16	$6 < Z \leq 7$	12	0.48
$7 < P \leq 8$	31	1.24	$350 < T \leq 360$	86	3.44	$7 < Z \leq 8$	4	0.16
$8 < P \leq 9$	24	0.96	$360 < T \leq 370$	0	0	$8 < Z \leq 9$	3	0.12
$9 < P \leq 10$	30	1.2	$370 < T \leq 380$	20	0.8	$9 < Z \leq 10$	2	0.08
$10 < P \leq 11$	15	0.6	$380 < T \leq 390$	0	0	$10 < Z$	1	0.04
$11 < P \leq 12$	0	0	$390 < T$	1	0.04			
$12 < P \leq 13$	14	0.56						
$13 < P \leq 14$	2	0.08						
$14 < P \leq 15$	12	0.48						
$15 < P \leq 16$	4	0.16						
$16 < P \leq 17$	0	0						
$17 < P \leq 18$	4	0.16						
$18 < P \leq 19$	0	0						
$19 < P \leq 20$	15	0.6						
$20 < P$	1	0.04						

random division of the data. By utilization of k various partitions, this approach minimizes variance, and the performance estimation becomes less affected by the random division of data. In this study, considering the limited data on the CO_2 capacity in functional ILs (2000 data points for the training set), we chose n to be 5 and k to be 5. Thus, 5-fold cross-validation (cv = 5) is used in the model evaluation process for evaluating the performance of models and optimizing their parameters.[59]

3. RESULTS AND DISCUSSION

3.1. Data Analysis. Before the ML models were constructed, the database was qualitatively analyzed. It is clear that CO_2 absorption capacity of ILs increases as temperature decreases or pressure increases. Thus, high pressure and low temperature favor the absorption of CO_2 in functional ILs. The data encompass a vast array of CO_2 capacity measurements, ranging from 0.0037 to 10.421 mol/kg, and span a wide spectrum of temperatures (283.1−393.15 K) and pressures (0.0004−20.006 bar). Table 2 and Figure 8 provide the distribution of experimental CO_2 data points from the database in the different ranges of CO_2 partial pressures, absorption temperatures, and absorption capacities with distribution counts and percentages. It can be seen that most data points, 63.84% and 66.84%, are in a narrow CO_2 partial pressure (in bar) range ($0 < P \leq 1$) and absorption temperature (in K) range ($290 < T \leq 320$), respectively. This highlights a significant research focus on the capture of CO_2 under mild conditions.

3.2. Model Parameter Optimization. *3.2.1. GC-MLP Model Optimization and Performance.* Four GC-based ML models were developed in this study. For the GC-MLP model, this study employs a four-layer feedforward neural network architecture, comprising an input layer, two hidden layers, and an output layer. The MLP model was implemented by using sklearn in Python. It is known that too few neurons in hidden layers will lead to underfitting, while too many neurons may lead to overfitting. The grid search with the ReLU activation function is used to optimize the number of neurons in each hidden layer, and the neuron count spanning is [10, 110] with an interval of 5. The initial value of the learning rate is set to 0.01 and 0.05. MSE is used to evaluate the performance. Thus, 882 combinations ($21 \times 21 \times 2$) were tested with cv = 5. It can be seen that the optimized first hidden layer contains 95 neurons; the second hidden layer contains 45 neurons, and the learning rate is set to 0.01, resulting in the lowest average MSE (MSE = 0.1956). Thus, the architecture of the MLP model with the ReLU activation function is 46−95−45−1. Besides, the ANN model with one hidden layer was also studied. The number of combinations was 42 (21×2), resulting in the architectures of 46−15−1 (learning rate = 0.05, MSE = 0.862331). It is clear that the model with one hidden layer is not suitable for prediction. Although the ANN model with three hidden layers may obtain a little bit better result than the model with two hidden layers, the search space 18522 ($21 \times 21 \times 21 \times 2$) is very large and requires a long computation time. In most cases, two hidden layers are enough to establish the correlations between model input and output.[60] Therefore, we choose two hidden layers in this work to construct the MLP model. Furthermore, the comparative performance of GC-MLP models with various activation function types, including ReLU, Logistic, TanH, and Identity, has been studied by a random search. The parameters of these models can be found in Table S3, and the performance comparisons were illustrated in Table 3 and Figures S3−S5. As seen for the test set, TanH achieved the best performance on the test set in terms of R^2, MAE, MSE, and RMSE, while ReLU and Logistic demonstrated moderate performance. Moreover, the elevated AARD metric suggests the data noise and outliers. Besides, the Identity performed extremely poorly across all metrics. Therefore, TanH was considered in this work to construct the MLP model.

3.2.2. GC-SVR Model Optimization and Performance. The SVR model was implemented by using scikit-learn in Python.

https://doi.org/10.1021/acssuschemeng.5c06870
ACS Sustainable Chem. Eng. 2025, 13, 19977−19994

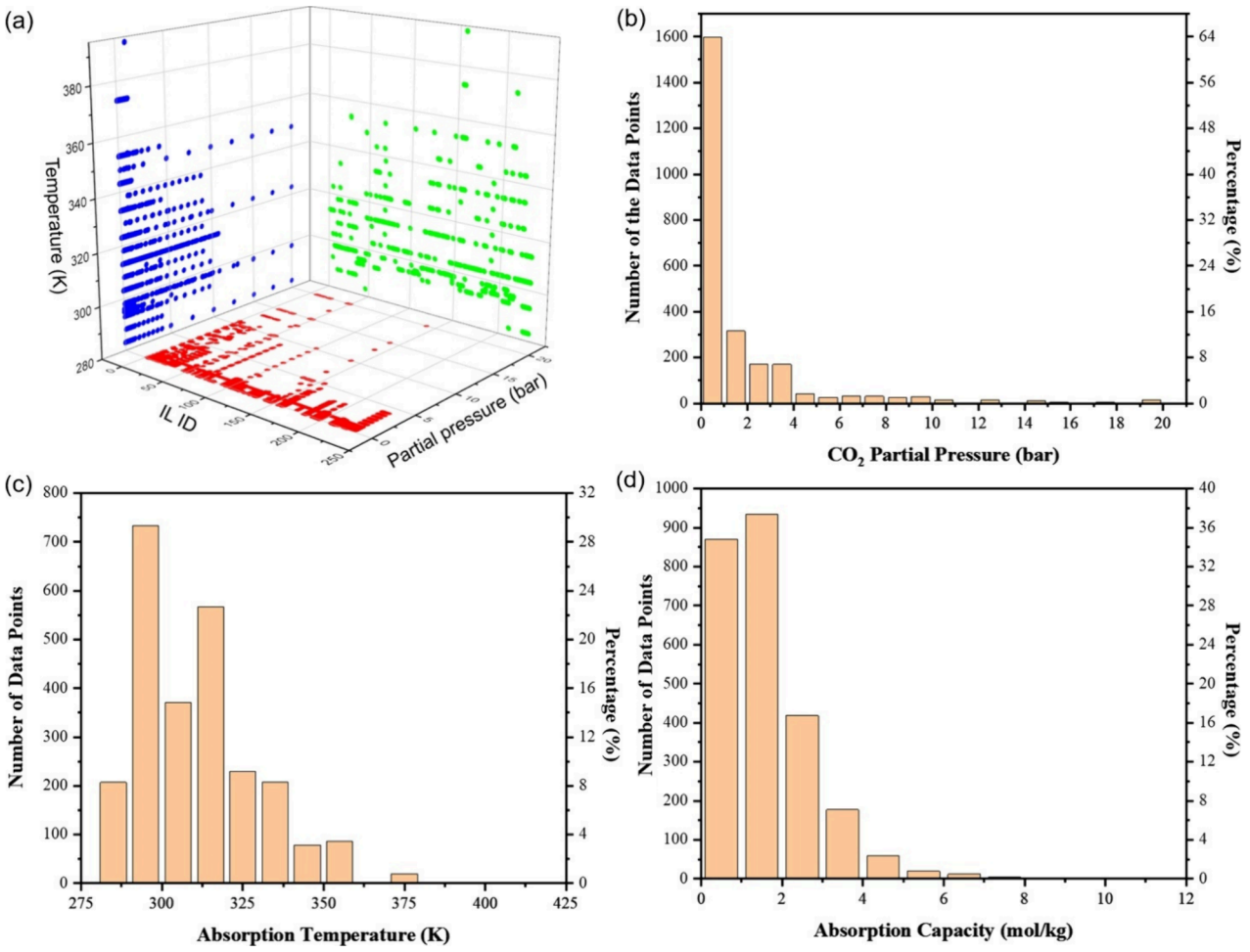


Figure 8. CO_2 capacity data of functional ILs collected from the literature and used in model training. (a) Projection of data points in the IL ID − temperature (blue), IL ID − partial pressure (red), and temperature − partial pressure (green) plots. (b−d) Histogram distributions of experimental CO_2 data points from the database in different ranges of CO_2 partial pressures, absorption temperatures, and absorption capacities with distribution counts (left axis) and percentages (right axis).

Table 3. Statistical Indicator Comparison of GC-MLP Models with Various Activation Function Types

		GC-MLP				
metric		ReLU (grid)	ReLU (random)	Logistic (random)	TanH (random)	Identity (random)
Test set	R^2	0.8439	0.8191	0.8084	0.8495	0.3974
	MAE [mol/kg]	0.2541	0.2621	0.2633	0.2403	0.6393
	MSE [(mol/kg)^2]	0.2153	0.2495	0.2642	0.2075	0.8311
	RMSE [mol/kg]	0.4640	0.4995	0.5140	0.4555	0.9116
	AARD%	36.7336	38.3311	36.8485	43.1605	112.6054
Training set	R^2	0.9388	0.9334	0.9301	0.9321	0.4420
	MAE [mol/kg]	0.1981	0.1967	0.1909	0.1807	0.6054
	MSE [(mol/kg)^2]	0.0895	0.0975	0.1023	0.0994	0.8166
	RMSE [mol/kg]	0.2992	0.3122	0.3189	0.3153	0.9036
	AARD%	31.8461	29.8929	31.7995	39.6505	134.3129
Overfitting	R^2 Gap (ΔR^2)	0.0949	0.1143	0.1217	0.0826	0.0046
	MSE Ratio	0.4157	0.3908	0.3872	0.4790	0.9825
	AARD% Increase	4.8875	8.4382	5.0490	3.5100	−21.7075
	Severity	medium	high	high	low	underfitting

The optimal values for C, γ, d, r (coef0), and ε in SVR models with different kernel functions, including Radial Basis Function (RBF), Linear, Polynomial, and Sigmoid, are typically found through systematic hyperparameter optimization techniques by a random search with n_iter = 100 and cv = 5. The parameter search ranges are as follows: [−3, 3] for $\log_{10}C$, [−5, 2] for $\log_{10}\gamma$, (2, 3, 4) for d, (0.0, 0.1, 1.0) for r (coef0), and [0.001, 0.1] for ε. The parameters of these models with different kernel types can be found in Table S4, and the performance comparisons are illustrated in Table 4 and Figures S6−S8. For the optimized GC-SVR model with the Linear kernel function, the value of C was 0.1026 and the lowest average MSE of

https://doi.org/10.1021/acssuschemeng.5c06870
ACS Sustainable Chem. Eng. 2025, 13, 19977−19994

Table 4. Statistical Indicator Comparison of GC-SVR Models with Various Kernel Types

metric		GC-SVR				
		RBF (grid)	RBF (random)	Linear (random)	Polynomial (random)	Sigmoid (random)
Test set	R^2	0.8008	0.8440	0.2601	0.7761	0.2766
	MAE [mol/kg]	0.2312	0.2138	0.6169	0.2861	0.6190
	MSE [(mol/kg)^2]	0.2747	0.2152	1.0204	0.3088	0.9977
	RMSE [mol/kg]	0.5241	0.4639	1.0101	0.5557	0.9989
	AARD%	27.9603	32.3137	100.4916	40.4569	107.8958
Training set	R^2	0.9702	0.9383	0.3888	0.8726	0.3661
	MAE [mol/kg]	0.0678	0.1268	0.5667	0.1956	0.5818
	MSE [(mol/kg)^2]	0.0436	0.0904	0.8945	0.1865	0.9277
	RMSE [mol/kg]	0.2089	0.3006	0.9458	0.4318	0.9632
	AARD%	12.3307	27.7841	104.7328	29.7616	126.8622
Overfitting	R^2 Gap (ΔR^2)	0.1694	0.0943	0.1287	0.0965	0.0895
	MSE Ratio	0.1587	0.4201	0.8766	0.6039	0.9298
	AARD% Increase	15.6296	4.5296	−4.2412	10.6953	−18.9664
	Severity	severe	moderate	moderate	moderate	mild

Table 5. Statistical Indicator Comparison of the Original GC-RF Model (Orig.) with Different Further Optimized GC-RF Models (Opt. 1−5)

metric		GC-RF					
		Orig.	Opt. 1	Opt. 2	Opt. 3	Opt. 4	Opt. 5
Test set	R^2	0.8629	0.8627	0.8865	0.7668	0.8823	0.8430
	MAE [mol/kg]	0.2362	0.2362	0.2223	0.3471	0.2168	0.2706
	MSE [(mol/kg)^2]	0.1891	0.1893	0.1565	0.3216	0.1622	0.2165
	RMSE [mol/kg]	0.4348	0.4351	0.3956	0.5671	0.4028	0.4653
	AARD%	28.3162	38.4276	39.9657	73.3636	37.5583	52.6181
Training set	R^2	0.9775	0.9776	0.9830	0.8565	0.9872	0.9431
	MAE [mol/kg]	0.1154	0.1155	0.0936	0.2834	0.0790	0.1744
	MSE [(mol/kg)^2]	0.0329	0.0328	0.0249	0.2100	0.0187	0.0833
	RMSE [mol/kg]	0.1815	0.1812	0.1579	0.4582	0.1368	0.2886
	AARD%	27.5241	32.4874	26.2136	88.5886	23.8670	51.2349
Overfitting	R^2 Gap (ΔR^2)	0.1146	0.1149	0.0965	0.0897	0.1049	0.1001
	MSE Ratio	0.1740	0.1733	0.1591	0.6529	0.1153	0.3847
	AARD% Increase	0.7291	0.5492	13.7521	−15.2250	13.6913	1.3833
	Severity	severe	severe	moderate to severe	moderate	severe	moderate

1.0204 was obtained for the test set. Compared with GC-SVR models with Linear and Sigmoid kernels, the performances of models with RBF and Polynomial are better, and RBF (C = 34.8793, γ = 0.2120, ε = 0.0069) exhibited the best performance of all (MSE = 0.2152). The $\log_{10}C$ and $\log_{10}\gamma$ have been further optimized by a random search with n_iter = 10000 and cv = 5 based on ε = 0.0069, resulting in C = 41.2153, γ = 0.3064, and MSE = 0.2167. For comparison, they also have been optimized by a grid search with 10000 candidates (100 × 100) and cv = 5, resulting in C = 46.415888, γ = 0.284804, and MSE = 0.216972. These further optimizations exhibited no significant improvement. Thus, the former results were used for further comparison.

3.2.3. GC-RF Model Optimization and Performance. The RF model is an ensemble ML algorithm comprising decision trees with quantitative analysis capabilities. The optimal RF model could be obtained through adjusting the number of decision trees (n_estimators) and the maximum depth of the tree (max_depth). In this study, the values of n_estimators and max_depth are in the ranges of [50, 500) and [5, 15), respectively, both with interval 1. After candidates (450 × 10) were tested by grid search with cv = 5, the optimal model was obtained with a max_depth of 14, an n_estimators of 499, and a test set MSE of 0.1891. The results demonstrate a near-perfect performance on the training data (R^2 = 0.9775, MAE = 0.1154) but a significant drop in performance on the test set (R^2 = 0.8629, MAE = 0.2362), indicating the severe overfitting of the model (MSE Ratio = 0.1740). Thus, further optimizations of the parameters should be addressed to overcome the issue. We named the original RF model as RF Orig. and the optimized models as RF Opt. 1 ∼ 5.

The parameters of RF Orig. and RF Opt. 1 ∼ 5 models can be found in Table S5, and the performance comparisons are illustrated in Table 5 and Figures S9−S11. For RF Opt. 1, the only significant change is the search strategy, shifting from a computationally expensive grid search (450 × 10) combinations to a more efficient random search with 100 iterations. Although the number of estimators (n_estimators) sees a trivial decrease from 499 to 449, the model's predictive capability and the overfitting problem are unchanged (test set R^2 = 0.8627, MSE Ratio = 0.1733) because the other parameters (max_depth, max_features, min_samples_leaf, min_samples_split) remain at their default. Therefore, the GC-RF model was further optimized by introducing the aforementioned regularization parameters, resulting in the RF Opt. 2 model. The results showed that the test set R^2 increased to 0.8865 while the MSE Ratio decreased to 0.1591, indicating that the test performance improved significantly and the

https://doi.org/10.1021/acssuschemeng.5c06870
ACS Sustainable Chem. Eng. 2025, 13, 19977−19994

Table 6. Statistical Indicator Comparison of the Original GC-GBR Model with Five Optimized Models

		GC-GBR					
metric		Orig.	Opt. 1	Opt. 2	Opt. 3	Opt. 4	Opt. 5
Test set	R^2	0.9186	0.9234	0.9158	0.8618	0.8836	0.8916
	MAE [mol/kg]	0.1763	0.1766	0.1881	0.2551	0.2289	0.2261
	MSE [(mol/kg)^2]	0.1122	0.1056	0.1162	0.1906	0.1605	0.1494
	RMSE [mol/kg]	0.3350	0.3250	0.3409	0.4366	0.4007	0.3866
	AARD%	10.1582	26.9097	28.7527	45.1301	33.7770	39.1776
Training set	R^2	0.9989	0.9971	0.9915	0.9401	0.9637	0.9756
	MAE [mol/kg]	0.0246	0.0431	0.0772	0.1790	0.1456	0.1217
	MSE [(mol/kg)^2]	0.0017	0.0042	0.0124	0.0876	0.0532	0.0357
	RMSE [mol/kg]	0.0409	0.0651	0.1112	0.2960	0.2306	0.1889
	AARD%	0.5321	8.8225	18.1337	46.3691	37.3513	32.6223
Overfitting	R^2 Gap (ΔR^2)	0.0802	0.0737	0.0758	0.0783	0.0801	0.0840
	MSE Ratio	0.0152	0.0398	0.1067	0.4596	0.3315	0.2390
	AARD% Increase	9.6261	18.0872	10.6190	−1.2390	−3.5743	6.5553
	Severity	severe	severe	moderate to severe	moderate	moderate	moderate to severe

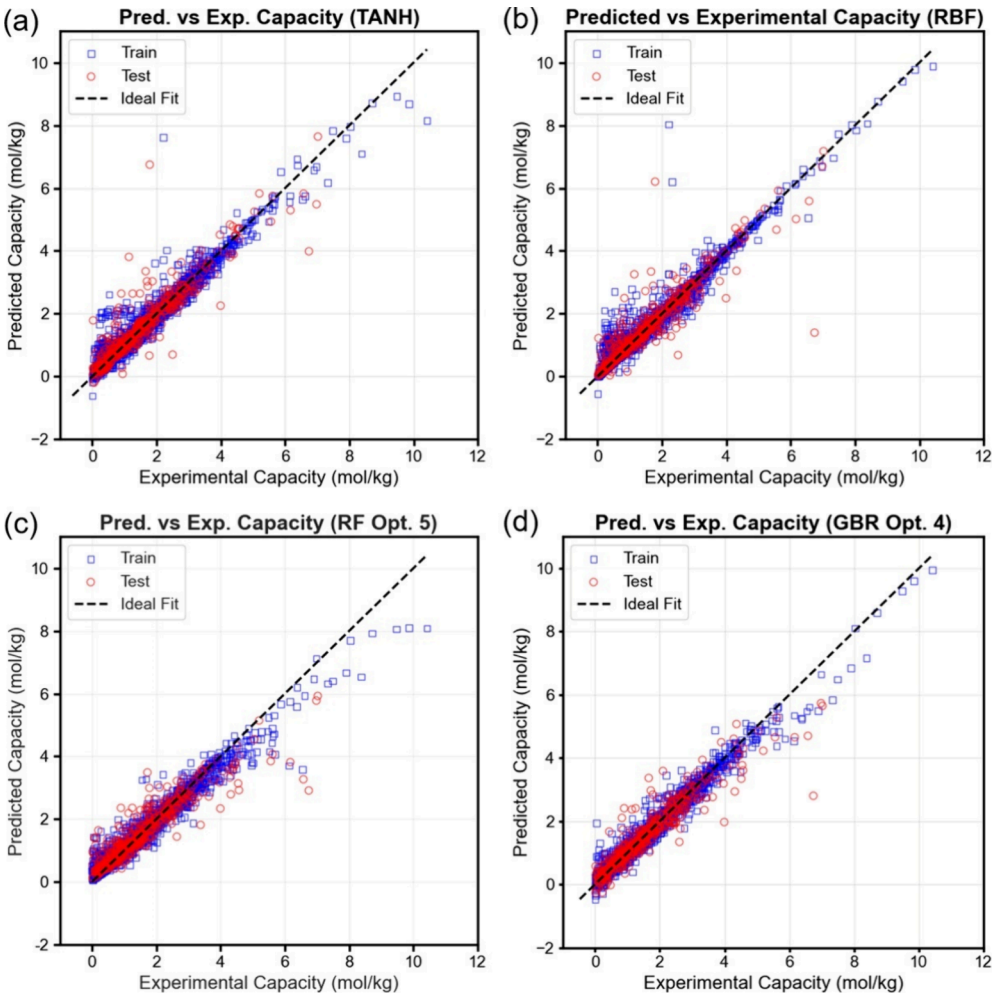


Figure 9. Comparison of the selected GC-based ML models for predicting CO_2 capacities. (a) GC-MLP with TanH activation function, (b) GC-SVR with RBF kernel, (c) GC-RF, and (d) GC-GBR.

overfitting gap decreased slightly. This is the result of three parameters mutually constraining each other. The max_depth and min_samples_split, higher than those of RF Orig. and RF Opt. 1, refer to the more complex model, which will result in overfitting. The max_features drastically reduced to 0.3181, forcing each tree to consider only a random subset of features (≈32%) for each split and making the model less likely to overfit. Thus, RF Opt. 2 improved the performance but did not fully solve the overfitting issue. Further optimized model RF Opt. 3 with max_depth = 17 and max_features = 0.5662 results in the highest MSE Ratio of 0.6529 but lowest test set R^2 of 0.7668. This is not a desirable outcome. RF Opt. 4 synchronously increases both max_depth and max_features to 30 and 0.6084, respectively, resulting in a drastic increase in

https://doi.org/10.1021/acssuschemeng.5c06870
ACS Sustainable Chem. Eng. 2025, 13, 19977−19994

Table 7. Statistical Indicator Comparison of Selected GC-Based ML Models

ML model	data set	R^2	MAE[mol/kg]	MSE[(mol/kg)^2]	RMSE[mol/kg]	AARD%
GC-MLP (TanH)	Training set	0.9321	0.1807	0.0994	0.3153	39.6505
	Test set	0.8495	0.2403	0.2075	0.4555	43.1605
GC-SVR (RBF)	Training set	0.9383	0.1268	0.0904	0.3006	27.7841
	Test set	0.8440	0.2138	0.2152	0.4639	32.3137
GC-RF (Opt. 5)	Training set	0.9431	0.1744	0.0833	0.2886	51.2349
	Test set	0.8430	0.2706	0.2165	0.4653	52.6181
GC-GBR (Opt. 4)	Training set	0.9637	0.1456	0.0532	0.2306	37.3513
	Test set	0.8836	0.2289	0.1605	0.4007	33.7770

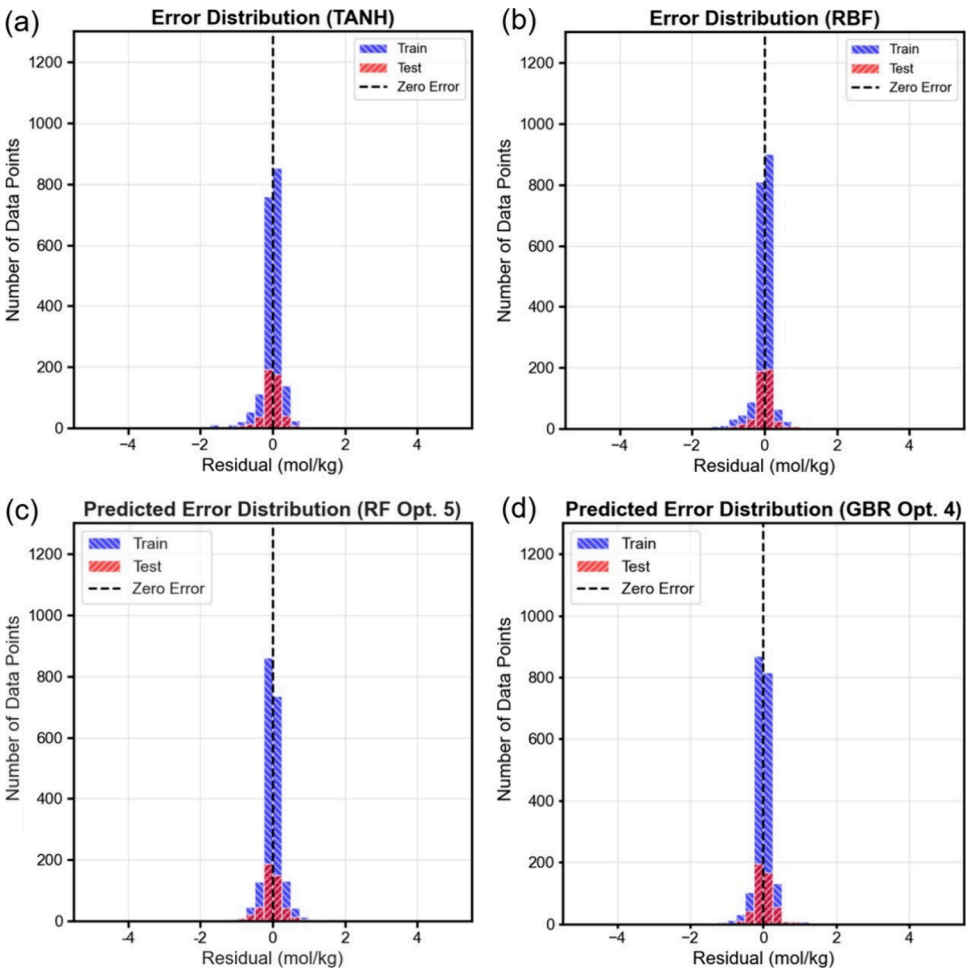


Figure 10. Distribution of the predicted errors in the training and test sets using selected GC-ML models: (a) GC-MLP with the TanH activation function, (b) GC-SVR with RBF kernel, (c) GC-RF, and (d) GC-GBR.

performance (test set R^2 = 0.8823) and a consequent return of overfitting (MSE Ratio = 0.1153). In order to seek a middle ground between the high-performance but overfit model (RF Opt. 4) and the well-generalized but underfit model (RF Opt. 3), mild constraints were introduced, and RF Opt. 5 was obtained with test set R^2 = 0.8430 and MSE Ratio = 0.3847. Therefore, RF Opt. 5 was selected for further study.

3.2.4. GC-GBR Model Optimization and Performance. Different from the RF model, the GBR model constructs additive regression models by sequentially fitting simple decision trees based on the residuals at each iteration. In the beginning, the values of n_estimators and max_depth are selected for optimization in the same range [50, 500) and [5, 15), respectively, with interval 1 and cv = 5. After candidates (450 × 10) were tested by a grid search, the optimal model was obtained with a max_depth of 7, an n_estimators of 499, and a test set MSE of 0.1122. A reasonable speculation is that increasing the number of trees can reduce max_depth and achieve good accuracy. However, the R^2 Gap ($\Delta\overline{R^2}$) of ~0.08 and an extremely low MSE Ratio of 0.0152 confirm a severe overfitting issue. The model had memorized the training data almost perfectly (training set R^2 = 0.9989) but failed to generalize equally well to the unseen test data (test set R^2 = 0.9186). Therefore, a re-evaluation is necessary and has been performed. It should be noted that the purpose of adjusting parameters (e.g., learning_rate, max_depth, min_samples_leaf, min_samples_split, n_iter_no_change, validation_fraction, and subsample) is the relentless pursuit of an optimal trade-off: mitigating the model's pronounced overfitting while preserving, or even enhancing, its predictive performance on

https://doi.org/10.1021/acssuschemeng.5c06870
ACS Sustainable Chem. Eng. 2025, 13, 19977−19994

unseen data. We named the original GBR model GBR Orig. and the optimized models GBR Opt. 1 ∼ 5.

The parameters of GBR Orig. and GBR Opt. 1 ∼ 5 models can be found in Table S6, and the performance comparisons are illustrated in Table 6 and Figures S12−S14. Based on the similar metrics, it can be seen from the results of GBR Orig. and GBR Opt. 1 that shifting from an exhaustive grid search over a massive 450 × 10 parameter space to a more efficient random search with 100 iterations allows for a broader exploration of the hyperparameter space with less computational cost. Therefore, random research was used for further optimization. Besides, a higher learning_rate allows each tree to contribute more, often leading to models that require fewer trees, while an increased min_samples_split makes it harder for the model to split nodes, thus creating simpler trees. Thus, the MSE Ratio of 0.1067 was obtained by GBR Opt. 2, which was 7 times that of GBR Orig. In the model GBR Opt. 3 ∼ 5, early stopping parameters, including n_iter_no_change = 15, tol = 0.00001, and validation_fraction = 0.2, were introduced, which is the most important change in the optimization. Instead of predefining the number of trees (n_estimators), the algorithm now uses a held-out validation set (20% of the training data) to stop adding trees once performance on this validation set fails to improve by at least tol for 15 consecutive rounds. This directly combats overfitting by preventing the model from learning noise after a certain point. As expected with such a strong regularization, the effect is immediate and noticeable. For GBR Opt. 3, the MSE Ratio increased to 0.4596, but the test performance R^2 dropped significantly from 0.9158 to 0.8618. After further reducing validation_fraction from 0.2 to 0.1 and increasing learning_rate from 0.0708 to 0.1290, the models GBR Opt. 4 and GBR Opt. 5 changed slightly more powerful and less constrained. Based on the metrics of these two models, GBR Opt. 4 with a test set R^2 of 0.8863 and an MSE Ratio of 0.3315 is better than GBR Opt. 5 (test set R^2 = 0.8916, MSE Ratio = 0.2390). Therefore, GBR Opt. 4 was chosen for further study.

3.3. Model Performance Evaluation. *3.3.1. Predicted CO_2 Capacities.* Following the development of four GC-based ML models, a systematic performance evaluation was conducted based on a comparative analysis. Figure 9 illustrates the comparison of the different GC-based ML models for the training set and test set splinted from the whole database randomly in different models. It is known that $y = x$ is the ideal regression line, where x is the experimental CO_2 capacities and y is the predicted CO_2 capacities by different models. Thus, enhanced model accuracy according to the order GC-MLP (TanH) < GC-RF (Opt. 5) < GC-SVR (RBF) < GC-GBR (Opt. 4) was evidenced by data points converging toward the $y = x$ ideal regression line, where improved clustering along this diagonal corresponds to higher predictive fidelity. Such behavior directly correlates with the data set's correlation coefficient (R^2), serving as a quantitative indicator of model efficacy. In Table 7, where the comprehensive performance metrics for GC-based ML models are tabulated, the comparative analysis revealed the GC-GBR model's superior predictive capability, achieving the highest R^2 values, 0.9637 for the training set and 0.8836 for the test set. The results suggested that the GC-GBR model confirms exceptional concordance between computational projections and experimental observations. These results are confirmed by the residual analysis (see the Supporting Information for details).

Additionally, a comparison of the frequency distributions of errors between the predicted and experimental CO_2 capacity values in the training set and test set by the developed different GC-based ML models is presented in Figure 10, where the vertical axis shows the number of data points for each variable's specific range (horizontal axis). It can be seen that 80.5, 85.5, 79.8, and 84.2% training errors and 73.8, 77.0, 67.6, and 72% test errors are inside the range of [−0.25, 0.25] by GC-MLP with TanH activation, GC-SVR with RBF kernel, GC-RF, and GC-GBR models, respectively. These results demonstrate that the GC-GBR model has high accuracy and reliability to predict the CO_2 capacities of functional ILs.

3.3.2. Comparison of Predicted Capacity with Thermodynamic Equations. To obtain a more thorough grasp of these models, it is important to present a comparison of experimental and predicted CO_2 capacity data by different GC-based ML models and thermodynamic methods. CO_2 capture by typical ILs, such as amine-based IL [Bmim][Ala] and amine-free IL [P66614][2-CN-Pyrro], under different partial pressures is applied to evaluate these ML models. The results are illustrated in Figure 11, and the corresponding thermodynamic equation ("deactivated model") for the absorptions of these ILs in detail can be found in the Supporting Information and references.[61,62] It is evident that

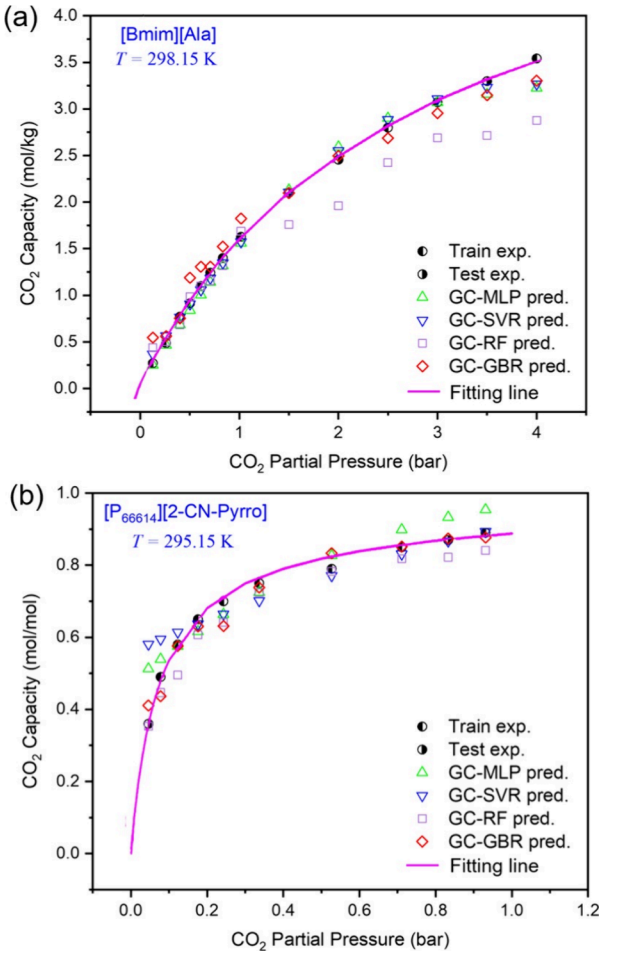


Figure 11. Comparison of experimental data (exp.) and predicted data (pred.) by selected GC-based ML models with thermodynamic methods (fitting lines) for CO_2 capture by (a) [Bmim][Ala] and (b) [P66614][2-CN-Pyrro].

https://doi.org/10.1021/acssuschemeng.5c06870
ACS Sustainable Chem. Eng. 2025, 13, 19977−19994

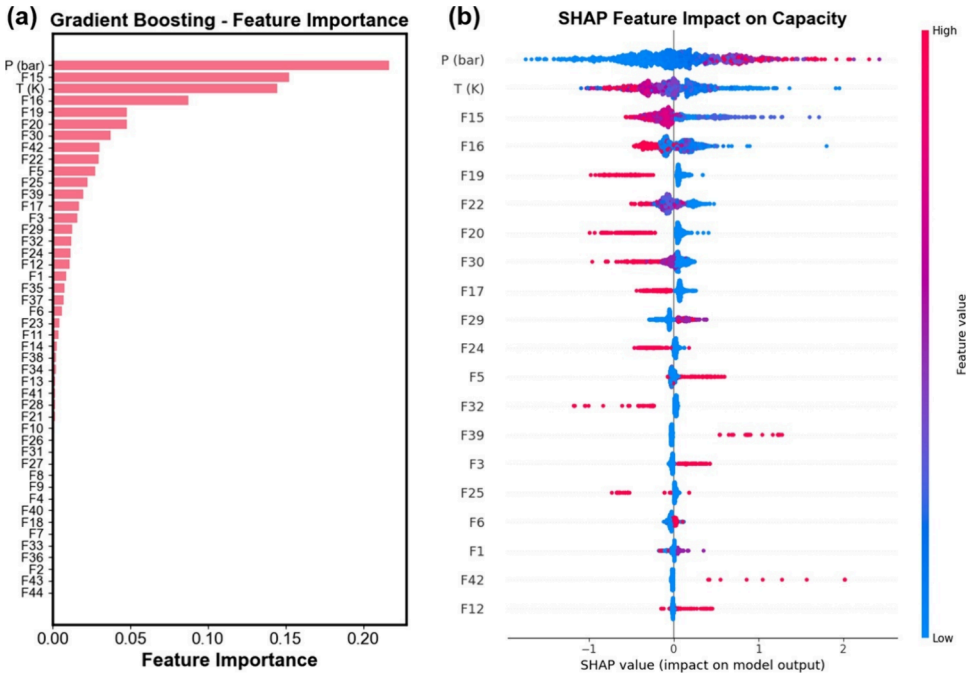


Figure 12. Feature importance data (a) and SHAP plots (b) of the GC-GBR model (GBR Opt. 4).

the predicted data from all these ML models demonstrate consistent trends of change in these ILs, aligning well with the experimental data and conforming to the general strong interaction between CO_2 and active sites reported in previous works, where the CO_2 capacity increases with pressure at a constant temperature. Among these GC-based ML models, it is clear to see that the GC-GBR model can simulate the experimental data precisely.

3.3.3. Feature Importance and SHAP Analysis. The feature importance analysis reveals the key descriptors that the model identified as most predictive for the CO_2 absorption capacity. Feature importance is calculated as the mean reduction in impurity (e.g., MSE) achieved by splits using each feature across all trees in the ensemble. In contrast, the SHapley Additive exPlanations (SHAP) plot is obtained by calculating the marginal contribution of each feature to every individual prediction relative to a baseline average, providing a deep insight into how each feature influences the model's predictions. The results from the GC-GBR model (GBR Opt. 4) are illustrated in Figure 12. The importance distribution is highly nonuniform, indicating that a small subset of features dominates the predictive power of the model. The top five features (P, T, F 15, F 16, and F 19) account for a significantly larger share of the total importance, suggesting that the capacity is primarily driven by a few critical properties or conditions. The operating conditions P (Importance = 0.2169) and T (Importance = 0.1449) are among the most influential features. The SHAP plot shows a strong positive correlation for P and a strong negative correlation for T, which means that high pressure values (red dots) are associated with positive SHAP values, while high temperature values (red dots) are associated with negative SHAP values. The results indicated that increased pressure causes CO_2 capture, while increased thermal energy promotes desorption. Interestingly, saturated aliphatic groups, such as F 15 ($-CH_3$) and F 16 ($-CH_2-$), are assigned high importance. However, the SHAP plot reveals that this is a spurious correlation. The SHAP plot for F15 ($-CH_3$), F16 ($-CH_2-$), F19 ($=CH_2$), F17 ($>CH-$), F20 ($=CH-$), and F22 (−H, ring) shows clearly that high counts (red dots) are clustered on the left, in the negative SHAP value region. This indicates that the model has learned the association between abundant aliphatic chains and lower CO_2 absorption capacity by using them as proxies for poor performance. In contrast, F29 ($-NH_2$) with Importance = 0.0129 exhibits the opposite behavior: high counts (red dots) are associated with positive SHAP values. Thus, the model correctly identifies that an increasing number of amine groups increases the predicted CO_2 capacity. This is a pivotal finding. The model has learned a logically consistent hierarchy that abundant inert, saturated carbon−hydrogen groups are the strongest indicators of low performance, while the amine group is the key indicator of high performance. The results move the model beyond a mere "statistical black-box" to an interpretable tool that aligns with chemical intuition.

4. CONCLUSIONS

This study established an efficient and rapid approach to predict the CO_2 capture capacity by functional ionic liquid (IL). Initially, we summarized the structures of functional ILs and established an extensive data bank containing 2500 experimental capacity data points of CO_2 capture as a function of operational pressure (P) and temperature (T) by 232 ILs including 44 cations and 123 anions. Based on 44 types of fragments, including ionic fragments (IFs), four group contribution (GC) based ML regression models, GC-MLP, GC-SVR, GC-RF, and GC-GBR, were developed. A combination of 5-fold cross-validation and grid search/random search was employed as an effective way to optimize model parameters and enhance their performance. Besides, managing the trade-off between model complexity and generalization performance is a fundamental challenge. GC-based MLP with TanH, SVR with RBF, RF Opt. 5, and GBR Opt. 4 are four ML models, representing a good balance between complexity

https://doi.org/10.1021/acssuschemeng.5c06870
ACS Sustainable Chem. Eng. 2025, 13, 19977−19994

and performance. Among these models, the GC-GBR model demonstrated the strongest predictive ability. Using the optimized GC-GBR model (GBR Opt. 4) with lower overfitting, superior accuracy was achieved on the training set (R^2 = 0.9637, MAE = 0.1456, MSE = 0.0532, RMSE = 0.2306, AARD% = 37.35%) and the test set (R^2 = 0.8836, MAE = 0.2289, MSE = 0.1605, RMSE = 0.4007, AARD% = 33.78%). Beyond its predictive performance, the GC-GBR model offered valuable interpretability through the SHAP analysis. This analysis revealed that higher pressure and lower temperature increased CO_2 capacity. It also identified that abundant aliphatic groups served as indicators of lower absorption capacity, whereas the amine group was correctly recognized as a key promoter of higher CO_2 uptake. The GBR-SHAP analysis transitions the model from a black-box predictor to an interpretable tool that can guide molecular design. Furthermore, typical ILs [Bmim][Ala] and [P66614][2-CN-Pyrro] are selected to compare the experimental and predicted CO_2 capacity data by different IF-based ML models and thermodynamic methods, and the predictions of the GC-GBR model are in excellent agreement with both the experimental data and the thermodynamic fitting curve. This is the first time that GBR-SHAP has been used for analysis of CO_2 capture performance by functional ILs, and we hope that these GC-based ML models could be used to develop functional ILs in the future for efficient CO_2 capture.

■ ASSOCIATED CONTENT

Data Availability Statement

The data that support the findings of this study are available as xlsx and pdf files in the Supporting Information. The code for model training and evaluation is available at https://github.com/chemcgk/zhang/tree/main/IL-CO2-model

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssuschemeng.5c06870.

Supporting Information-01: the method of construction of the GC-MLP model and GC-SVR model, the abbreviations and full-names of 232 kinds of ILs (Table S1), the mean value and the standard deviation of samples for transforming both the training and test sets (Table S2), the results of the optimized GC-MLP models (Table S3), GC-SVR models (Table S4), GC-RF models (Table S5), and GC-GBR models (Table S6), the detailed methodology of building ML models (Figure S1), the structures and abbreviations of 47 cations and 123 anions (Figure S2), the model performance, residual, and errors of the GC-MLP models (ReLU, Logistic, TanH), GC-SVR models (RBF, Linear, Polynomial, Sigmoid), GC-RF models (Orig., Opt. 1 ~ 5), and GC-GBR models (Orig., Opt. 1 ~ 5) on the training set and test set (Figures S3−S14), and the residual analysis of the selected four models (Figures S15−S16) (PDF)

Supporting Information-02: CO_2 capacity data and reference, predicted results, and predicted data comparison of [Bmim][Ala] and [P66614][2-CN-Pyrro] (XLSX)

■ AUTHOR INFORMATION

Corresponding Authors

Guokai Cui − *Innovation Team of Air Pollution Control, Institute of Catalytic Reaction Engineering, Zhejiang Key Laboratory of Surface and Interface Science and Engineering for Catalysts, State Key Laboratory of Green Chemical Synthesis and Conversion, College of Chemical Engineering, Zhejiang University of Technology, Hangzhou 310014, China*; orcid.org/0000-0002-7223-2869; Email: chemcgk@163.com

Xiangping Zhang − *Center of Ionic Liquids and Green Energy, Beijing Key Laboratory of Solid State Battery and Energy Storage Process, State Key Laboratory of Mesoscience and Engineering, Institute of Process Engineering, Chinese Academy of Sciences, Beijing 100190, China*; orcid.org/0000-0002-1431-0873; Email: xpzhang@ipe.ac.cn

Hanfeng Lu − *Innovation Team of Air Pollution Control, Institute of Catalytic Reaction Engineering, Zhejiang Key Laboratory of Surface and Interface Science and Engineering for Catalysts, State Key Laboratory of Green Chemical Synthesis and Conversion, College of Chemical Engineering, Zhejiang University of Technology, Hangzhou 310014, China*; orcid.org/0000-0001-8934-8270; Email: luhf@zjut.edu.cn

Authors

Ruina Zhang − *Innovation Team of Air Pollution Control, Institute of Catalytic Reaction Engineering, Zhejiang Key Laboratory of Surface and Interface Science and Engineering for Catalysts, State Key Laboratory of Green Chemical Synthesis and Conversion, College of Chemical Engineering, Zhejiang University of Technology, Hangzhou 310014, China*

Wei Zhang − *Zhejiang Zheneng Technology & Environment Group Co., Ltd., Hangzhou 310012, China*

Kaikai Li − *Center of Ionic Liquids and Green Energy, Beijing Key Laboratory of Solid State Battery and Energy Storage Process, State Key Laboratory of Mesoscience and Engineering, Institute of Process Engineering, Chinese Academy of Sciences, Beijing 100190, China*

Lai Li − *Zhejiang Zheneng Technology & Environment Group Co., Ltd., Hangzhou 310012, China*

Xiangyu Shen − *Innovation Team of Air Pollution Control, Institute of Catalytic Reaction Engineering, Zhejiang Key Laboratory of Surface and Interface Science and Engineering for Catalysts, State Key Laboratory of Green Chemical Synthesis and Conversion, College of Chemical Engineering, Zhejiang University of Technology, Hangzhou 310014, China*

Chunliang Ge − *Zhejiang Zheneng Technology & Environment Group Co., Ltd., Hangzhou 310012, China*

Wanxiang Zhang − *State Key Laboratory of Chemical Resource Engineering, Beijing University of Chemical Technology, Beijing 100029, China*; orcid.org/0000-0001-5914-2001

Wenyang Fan − *Zhejiang Zheneng Technology & Environment Group Co., Ltd., Hangzhou 310012, China*

Ying Zhou − *Innovation Team of Air Pollution Control, Institute of Catalytic Reaction Engineering, Zhejiang Key Laboratory of Surface and Interface Science and Engineering for Catalysts, State Key Laboratory of Green Chemical Synthesis and Conversion, College of Chemical Engineering,*

https://doi.org/10.1021/acssuschemeng.5c06870
ACS Sustainable Chem. Eng. 2025, 13, 19977−19994

Zhejiang University of Technology, Hangzhou 310014, China

Yuan Tian − *Center of Ionic Liquids and Green Energy, Beijing Key Laboratory of Solid State Battery and Energy Storage Process, State Key Laboratory of Mesoscience and Engineering, Institute of Process Engineering, Chinese Academy of Sciences, Beijing 100190, China*

Xiaopo Niu − *Innovation Team of Air Pollution Control, Institute of Catalytic Reaction Engineering, Zhejiang Key Laboratory of Surface and Interface Science and Engineering for Catalysts, State Key Laboratory of Green Chemical Synthesis and Conversion, College of Chemical Engineering, Zhejiang University of Technology, Hangzhou 310014, China*

Guoxiong Zhan − *Center of Ionic Liquids and Green Energy, Beijing Key Laboratory of Solid State Battery and Energy Storage Process, State Key Laboratory of Mesoscience and Engineering, Institute of Process Engineering, Chinese Academy of Sciences, Beijing 100190, China*

Quanli Ke − *Innovation Team of Air Pollution Control, Institute of Catalytic Reaction Engineering, Zhejiang Key Laboratory of Surface and Interface Science and Engineering for Catalysts, State Key Laboratory of Green Chemical Synthesis and Conversion, College of Chemical Engineering, Zhejiang University of Technology, Hangzhou 310014, China;* orcid.org/0000-0002-9123-5803

Shuhang Ren − *State Key Laboratory of Chemical Resource Engineering, Beijing University of Chemical Technology, Beijing 100029, China;* orcid.org/0000-0003-3253-8852

Complete contact information is available at: https://pubs.acs.org/10.1021/acssuschemeng.5c06870

Author Contributions

Ruina Zhang: Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing − original draft. Guokai Cui: Conceptualization, Project administration, Funding acquisition, Resources, Writing − review and editing. Wei Zhang: Formal analysis. Kaikai Li: Formal analysis, Validation, Methodology. Lai Li: Formal analysis. Xiangyu Shen: Formal analysis, Data curation. Chunliang Ge: Formal analysis. Wanxiang Zhang: Formal analysis, Methodology. Wenyang Fan: Formal analysis. Ying Zhou: Resources, Formal analysis, Funding acquisition. Yuan Tian: Formal analysis, Methodology. Xiaopo Niu: Formal analysis. Guoxiong Zhan: Formal analysis. Quanli Ke: Formal analysis, Funding acquisition. Shuhang Ren: Formal analysis, Methodology. Xiangping Zhang: Formal analysis, Conceptualization. Hanfeng Lu: Resources, Project administration, Funding acquisition, Conceptualization, Supervision, Writing − review and editing.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Key Research and Development Project in Zhejiang Province (No. 2024C03108, No. 2024C03114, No. 2023C03127), the National Natural Science Foundation of China (No. 22378353), the Zhejiang Provincial Natural Science Foundation of China (No. LTGS24E080008), the Zhejiang Planned Projects for Postdoctoral Research Funds (No. ZJ2023145), and the Zhejiang Zheneng Technology & Environment Group Co., Ltd. "Research and Development of Technology and Equipment based on Ionic Liquids for Carbon Capture from Flue Gas with Low Energy Consumption" Technology Project (No. ZNKJ-2024-077).

■ REFERENCES

(1) W.M.O. *State of the global climate 2023 (WMO-No. 1347)*; 978−92−63−11347−4; World Meteorological Organization: Geneva 2, Switzerland, **2023**, 2024; pp 2. https://library.wmo.int/zh/records/item/68835-state-of-the-global-climate-2023.

(2) I.P.C.C, *Climate Change 2023: Synthesis Report*. World Meteorological Organization: Geneva, Switzerland, 2024; p 42. doi: .

(3) Wu, C.; Huang, Q.; Xu, Z.; Sipra, A. T.; Gao, N.; Vandenberghe, L. P. d. S.; Vieira, S.; Soccol, C. R.; Zhao, R.; Deng, S.; Boetcher, S. K. S.; Lu, S.; Shi, H.; Zhao, D.; Xing, Y.; Chen, Y.; Zhu, J.; Feng, D.; Zhang, Y.; Deng, L.; Hu, G.; Webley, P. A.; Liang, D.; Ba, Z.; Mlonka-Mędrala, A.; Magdziarz, A.; Miskolczi, N.; Tomasek, S.; Lam, S. S.; Foong, S. Y.; Ng, H. S.; Jiang, L.; Yan, X.; Liu, Y.; Ji, Y.; Sun, H.; Zhang, Y.; Yang, H.; Zhang, X.; Sun, M.; Tsang, D. C. W.; Shang, J.; Muller, C.; Rekhtina, M.; Krödel, M.; Bork, A. H.; Donat, F.; Liu, L.; Jin, X.; Liu, W.; Saqline, S.; Wu, X.; Xu, Y.; Khan, A. L.; Ali, Z.; Lin, H.; Hu, L.; Huang, J.; Singh, R.; Wang, K.; He, X.; Dai, Z.; Yi, S.; Konist, A.; Baqain, M. H. S.; Zhao, Y.; Sun, S.; Chen, G.; Tu, X.; Weidenkaff, A.; Kawi, S.; Lim, K. H.; Song, C.; Yang, Q.; Zhao, Z.; Gao, X.; Jiang, X.; Ji, H.; Akinola, T. E.; Lawal, A.; Otitoju, O. S.; Wang, M.; Zhang, G.; Ma, L.; Sempuga, B. C.; Liu, X.; Oko, E.; Daramola, M.; Yu, Z.; Chen, S.; Kang, G.; Li, Q.; Gao, L.; Liu, L.; Zhou, H. A comprehensive review of carbon capture science and technologies. *Carbon Capture Sci. Technol.* **2024**, *11*, No. 100178.

(4) Khan, B.; Faheem, M. B.; Peramaiah, K.; Nie, J.; Huang, H.; Li, Z.; Liu, C.; Huang, K. W.; He, J. H. Unassisted photoelectrochemical CO2-to-liquid fuel splitting over 12% solar conversion efficiency. *Nat. Commun.* **2024**, *15* (1), 6990.

(5) Zhang, R.; Cui, G.; Wang, X.; Chen, Y.; Qiu, X.; Ke, Q.; Deng, D.; Ge, C.; Lu, H.; Dai, S. Ionic Liquid-based Advanced Porous Organic Hyper-Crosslinked Polymers (ILHCPs) for CO_2 Capture and Conversion. *Chem. Eng. J.* **2024**, *489*, No. 151102.

(6) Gao, W.; Liang, S.; Wang, R.; Jiang, Q.; Zhang, Y.; Zheng, Q.; Xie, B.; Toe, C. Y.; Zhu, X.; Wang, J.; Huang, L.; Gao, Y.; Wang, Z.; Jo, C.; Wang, Q.; Wang, L.; Liu, Y.; Louis, B.; Scott, J.; Roger, A.-C.; Amal, R.; He, H.; Park, S.-E. Industrial carbon dioxide capture and utilization: state of the art and future challenges. *Chem. Soc. Rev.* **2020**, *49* (23), 8584−8686.

(7) Pei, Y.; Zhang, Y.; Ma, J.; Fan, M.; Zhang, S.; Wang, J. Ionic liquids for advanced materials. *Mater. Today Nano* **2022**, *17*, No. 100159.

(8) Holbrey, J. D.; Seddon, K. R. Ionic Liquids. *Clean Prod. Process.* **1999**, *1* (4), 223−236.

(9) Lei, Z.; Dai, C.; Hallett, J.; Shiflett, M. Introduction: Ionic Liquids for Diverse Applications. *Chem. Rev.* **2024**, *124* (12), 7533−7535.

(10) Zhou, T.; Gui, C.; Sun, L.; Hu, Y.; Lyu, H.; Wang, Z.; Song, Z.; Yu, G. Energy Applications of Ionic Liquids: Recent Developments and Future Prospects. *Chem. Rev.* **2023**, *123* (21), 12170−12253.

(11) Yu, G.; Dai, C.; Liu, N.; Xu, R.; Wang, N.; Chen, B. Hydrocarbon Extraction with Ionic Liquids. *Chem. Rev.* **2024**, *124* (6), 3331−3391.

(12) Dupont, J.; Leal, B. C.; Lozano, P.; Monteiro, A. L.; Migowski, P.; Scholten, J. D. Ionic Liquids in Metal, Photo-, Electro-, and (Bio) Catalysis. *Chem. Rev.* **2024**, *124* (9), 5227−5420.

(13) Shamshina, J. L.; Rogers, R. D. Ionic Liquids: New Forms of Active Pharmaceutical Ingredients with Unique. *Tunable Properties. Chem. Rev.* **2023**, *123* (20), 11894−11953.

(14) Blanchard, L. A.; Hancu, D.; Beckman, E. J.; Brennecke, J. F. Green processing using ionic liquids and CO_2. *Nature* **1999**, *399* (6731), 28−29.

(15) Yu, G.; Dai, C.; Gao, H.; Zhu, R.; Du, X.; Lei, Z. Capturing Condensable Gases with Ionic Liquids. *Ind. Eng. Chem. Res.* **2018**, *57* (36), 12202−12214.
(16) Bara, J. E.; Carlisle, T. K.; Gabriel, C. J.; Camper, D.; Finotello, A.; Gin, D. L.; Noble, R. D. Guide to CO2 Separations in Imidazolium-Based Room-Temperature Ionic Liquids. *Ind. Eng. Chem. Res.* **2009**, *48* (6), 2739−2751.
(17) Hillesheim, P. C.; Singh, J. A.; Mahurin, S. M.; Fulvio, P. F.; Oyola, Y.; Zhu, X.; Jiang, D.-e.; Dai, S. Effect of alkyl and aryl substitutions on 1,2,4-triazolium-based ionic liquids for carbon dioxide separation and capture. *RSC Adv.* **2013**, *3* (12), 3981−3989.
(18) Blanchard, L. A.; Gu, Z.; Brennecke, J. F. High-Pressure Phase Behavior of Ionic Liquid/CO2 Systems. *J. Phys. Chem. B* **2001**, *105* (12), 2437−2444.
(19) Bates, E. D.; Mayton, R. D.; Ntai, I.; Davis, J. H. CO_2 Capture by a Task-Specific Ionic Liquid. *J. Am. Chem. Soc.* **2002**, *124* (6), 926−927.
(20) Xu, Y.; Zhang, R.; Zhou, Y.; Hu, D.; Ge, C.; Fan, W.; Chen, B.; Chen, Y.; Zhang, W.; Liu, H.; Cui, G.; Lu, H. Tuning ionic liquid-based functional deep eutectic solvents and other functional mixtures for CO_2 capture. *Chem. Eng. J.* **2023**, *463*, No. 142298.
(21) Zhang, R.; Ke, Q.; Zhang, Z.; Zhou, B.; Cui, G.; Lu, H. Tuning Functionalized Ionic Liquids for CO_2 Capture. *Int. J. Mol. Sci.* **2022**, *23* (19), 11401.
(22) Cui, G.; Wang, J.; Zhang, S. Active chemisorption sites in functionalized ionic liquids for carbon capture. *Chem. Soc. Rev.* **2016**, *45* (15), 4307−4339.
(23) Wen, S.; Wang, T.; Zhang, X.; Xu, W.; Hu, X.; Wu, Y. Novel amino acid ionic liquids prepared via one-step lactam hydrolysis for the highly efficient capture of CO_2. *AIChE J.* **2023**, *69* (11), No. e18206.
(24) Zhang, J.; Jia, C.; Dong, H.; Wang, J.; Zhang, X.; Zhang, S. A Novel Dual Amino-Functionalized Cation-Tethered Ionic Liquid for CO2 Capture. *Ind. Eng. Chem. Res.* **2013**, *52* (17), 5835−5841.
(25) Chen, F.-F.; Huang, K.; Zhou, Y.; Tian, Z.-Q.; Zhu, X.; Tao, D.-J.; Jiang, D.-e.; Dai, S. Multi-Molar Absorption of CO_2 by the Activation of Carboxylate Groups in Amino Acid Ionic Liquids. *Angew. Chem., Int. Ed.* **2016**, *55* (25), 7166−7170.
(26) Gurkan, B. E.; de la Fuente, J. C.; Mindrup, E. M.; Ficke, L. E.; Goodrich, B. F.; Price, E. A.; Schneider, W. F.; Brennecke, J. F. Equimolar CO_2 Absorption by Anion-Functionalized Ionic Liquids. *J. Am. Chem. Soc.* **2010**, *132* (7), 2116−2117.
(27) Yoon, B.; Chen, S.; Voth, G. A. On the Key Influence of Amino Acid Ionic Liquid Anions on CO_2 Capture. *J. Am. Chem. Soc.* **2024**, *146* (2), 1612−1618.
(28) Wang, C.; Luo, H.; Jiang, D.-e.; Li, H.; Dai, S. Carbon Dioxide Capture by Superbase-Derived Protic Ionic Liquids. *Angew. Chem., Int. Ed.* **2010**, *49* (34), 5978−5981.
(29) Gurau, G.; Rodríguez, H.; Kelley, S. P.; Janiczek, P.; Kalb, R. S.; Rogers, R. D. Demonstration of Chemisorption of Carbon Dioxide in 1,3-Dialkylimidazolium Acetate Ionic Liquids. *Angew. Chem., Int. Ed.* **2011**, *50* (50), 12024−12026.
(30) Anderson, K.; Atkins, M. P.; Estager, J.; Kuah, Y.; Ng, S.; Oliferenko, A. A.; Plechkova, N. V.; Puga, A. V.; Seddon, K. R.; Wassell, D. F. Carbon dioxide uptake from natural gas by binary ionic liquid-water mixtures. *Green Chem.* **2015**, *17* (8), 4340−4354.
(31) Wang, C.; Luo, X.; Luo, H.; Jiang, D.-e.; Li, H.; Dai, S. Tuning the Basicity of Ionic Liquids for Equimolar CO_2 Capture. *Angew. Chem., Int. Ed.* **2011**, *50* (21), 4918−4922.
(32) Wang, K.; Zhang, Z.; Wang, S.; Jiang, L.; Li, H.; Wang, C. Dual-Tuning Azole-Based Ionic Liquids for Reversible CO2 Capture from Ambient Air. *ChemSusChem* **2024**, *17* (16), No. e202301951.
(33) Seo, S.; Quiroz-Guzman, M.; DeSilva, M. A.; Lee, T. B.; Huang, Y.; Goodrich, B. F.; Schneider, W. F.; Brennecke, J. F. Chemically Tunable Ionic Liquids with Aprotic Heterocyclic Anion (AHA) for CO2 Capture. *J. Phys. Chem. B* **2014**, *118* (21), 5740−5751.
(34) Wang, C.; Guo, Y.; Zhu, X.; Cui, G.; Li, H.; Dai, S. Highly efficient CO_2 capture by tunable alkanolamine-based ionic liquids with multidentate cation coordination. *Chem. Commun.* **2012**, *48* (52), 6526−6528.
(35) Lee, T. B.; Oh, S.; Gohndrone, T. R.; Morales-Collazo, O.; Seo, S.; Brennecke, J. F.; Schneider, W. F. CO_2 Chemistry of Phenolate-Based Ionic Liquids. *J. Phys. Chem. B* **2016**, *120* (8), 1509−1517.
(36) Zhao, T.; Zhang, X.; Tu, Z.; Wu, Y.; Hu, X. Low-viscous diamino protic ionic liquids with fluorine-substituted phenolic anions for improving CO2 reversible capture. *J. Mol. Liq.* **2018**, *268*, 617−624.
(37) Luo, X.; Guo, Y.; Ding, F.; Zhao, H.; Cui, G.; Li, H.; Wang, C. Significant Improvements in CO_2 Capture by Pyridine-Containing Anion-Functionalized Ionic Liquids through Multiple-Site Cooperative Interactions. *Angew. Chem., Int. Ed.* **2014**, *53* (27), 7053−7057.
(38) Zhu, X.; Liu, S.; Shang, R.; Chen, X.; Xu, Y.; Guo, Y.; Ling, B. Understanding on the structure of novel hydroxypyridine anion-based ionic liquids and their effect on CO2 absorption behavior. *J. Environ. Chem. Eng.* **2024**, *12* (5), No. 113552.
(39) Luo, X.-Y.; Chen, X.-Y.; Qiu, R.-X.; Pei, B.-Y.; Wei, Y.; Hu, M.; Lin, J.-Q.; Zhang, J.-Y.; Luo, G.-G. Enhanced CO_2 capture by reducing cation−anion interactions in hydroxyl-pyridine anion-based ionic liquids. *Dalton Trans.* **2019**, *48* (7), 2300−2307.
(40) Huang, Y.; Cui, G.; Zhao, Y.; Wang, H.; Li, Z.; Dai, S.; Wang, J. Preorganization and Cooperation for Highly Efficient and Reversible Capture of Low-Concentration CO_2 by Ionic Liquids. *Angew. Chem., Int. Ed.* **2017**, *56* (43), 13293−13297.
(41) Chen, T.; Zhang, Y.; Xu, Y. Efficient Synthesis of Quinazoline-2,4(1*H*,3*H*)-dione via Simultaneous Activated CO_2 and 2-Aminobenzonitrile by 1-Methylhydantoin Anion-Functionalized Ionic Liquid through the Multiple-Site Cooperative Interactions. *ACS Sustainable Chem. Eng.* **2022**, *10* (32), 10699−10711.
(42) Suo, X.; Fu, Y.; Do-Thanh, C.-L.; Qiu, L.-Q.; Jiang, D.-e.; Mahurin, S. M.; Yang, Z.; Dai, S. CO_2 Chemisorption Behavior in Conjugated Carbanion-Derived Ionic Liquids via Carboxylic Acid Formation. *J. Am. Chem. Soc.* **2022**, *144* (47), 21658−21663.
(43) Chen, M.; Xiong, W.; Chen, W.; Li, S.; Zhang, F.; Wu, Y. Synergy of carbanion siting and hydrogen bonding in super-nucleophilic deep eutectic solvents for efficient CO_2 capture. *AIChE J.* **2024**, *70* (4), No. e18319.
(44) Zhang, Z.; Mao, W.; Wang, K.; Jiang, L.; Wang, S.; Li, H.; Wang, C. Highly Efficient and Reversible Carbon Dioxide Capture by Carbanion-Functionalized Ionic Liquids. *ChemSusChem* **2024**, *17* (24), No. e202401111.
(45) Yang, Z.; Chen, B.; Chen, H.; Li, H. A critical review on machine-learning-assisted screening and design of effective sorbents for carbon dioxide (CO2) capture. *Front. Energy Res.* **2023**, *10*, 1043064.
(46) Bazargani, Z.; Sabzi, F. Prediction of CO2 solubility in ionic liquids with [HMIM] and [OMIM] cations by equation of state. *J. Mol. Liq.* **2016**, *216*, 87−93.
(47) Feider, N. O.; Mahurin, S. M.; Do-Thanh, C.-L.; Dai, S.; Jiang, D.-e. Molecular dynamics simulations of a dicationic ionic liquid for CO2 capture. *J. Mol. Liq.* **2021**, *335*, No. 116163.
(48) Lei, Z.; Dai, C.; Wang, W.; Chen, B. UNIFAC model for ionic liquid-CO2 systems. *AIChE J.* **2014**, *60* (2), 716−729.
(49) Palomar, J.; Gonzalez-Miquel, M.; Polo, A.; Rodriguez, F. Understanding the Physical Absorption of CO2 in Ionic Liquids Using the COSMO-RS Method. *Ind. Eng. Chem. Res.* **2011**, *50* (6), 3452−3463.
(50) Song, Z.; Shi, H.; Zhang, X.; Zhou, T. Prediction of CO2 solubility in ionic liquids using machine learning methods. *Chem. Eng. Sci.* **2020**, *223*, No. 115752.
(51) Eslamimanesh, A.; Gharagheizi, F.; Mohammadi, A. H.; Richon, D. Artificial Neural Network modeling of solubility of supercritical carbon dioxide in 24 commonly used ionic liquids. *Chem. Eng. Sci.* **2011**, *66* (13), 3039−3044.
(52) Baghban, A.; Ahmadi, M. A.; Hashemi Shahraki, B. Prediction carbon dioxide solubility in presence of various ionic liquids using computational intelligence approaches. *Journal of Supercritical Fluids* **2015**, *98*, 50−64.

https://doi.org/10.1021/acssuschemeng.5c06870
ACS Sustainable Chem. Eng. 2025, 13, 19977−19994

(53) Ouaer, H.; Hosseini, A. H.; Nait Amar, M.; El Amine Ben Seghier, M.; Ghriga, M. A.; Nabipour, N.; Andersen, P. Ø.; Mosavi, A.; Shamshirband, S. Rigorous Connectionist Models to Predict Carbon Dioxide Solubility in Various Ionic Liquids. *Applied Sciences* **2020**, *10* (1), 304.
(54) Yang, A.; Sun, S.; Su, Y.; Kong, Z. Y.; Ren, J.; Shen, W. Insight to the prediction of CO2 solubility in ionic liquids based on the interpretable machine learning model. *Chem. Eng. Sci.* **2024**, *297*, No. 120266.
(55) Huang, Y.; Dong, H.; Zhang, X.; Li, C.; Zhang, S. A new fragment contribution-corresponding states method for physicochemical properties prediction of ionic liquids. *AIChE J.* **2013**, *59* (4), 1348−1359.
(56) Li, K.; Zhu, Y.; Shi, S.; Song, Y.; Jiang, H.; Zhang, X.; Zeng, S.; Zhang, X. Machine learning models coupled with Ionic Fragment σ-profiles to predict ammonia solubility in ionic liquids. *Green Chem. Eng.* **2025**, 223.
(57) Tian, Y.; Wang, X.; Liu, Y.; Hu, W. Prediction of CO2 and N2 solubility in ionic liquids using a combination of ionic fragments contribution and machine learning methods. *J. Mol. Liq.* **2023**, *383*, No. 122066.
(58) Maharana, K.; Mondal, S.; Nemade, B. A review: Data preprocessing and data augmentation techniques. *Global Transitions Proceedings* **2022**, *3* (1), 91−99.
(59) Lemaoui, T.; Boublia, A.; Lemaoui, S.; Darwish, A. S.; Ernst, B.; Alam, M.; Benguerba, Y.; Banat, F.; AlNashef, I. M. Predicting the CO2 Capture Capability of Deep Eutectic Solvents and Screening over 1000 of their Combinations Using Machine Learning. *ACS Sustainable Chem. Eng.* **2023**, *11* (26), 9564−9580.
(60) Yusuf, F.; Olayiwola, T.; Afagwu, C. Application of Artificial Intelligence-based predictive methods in Ionic liquid studies: A review. *Fluid Phase Equilib.* **2021**, *531*, No. 112898.
(61) Noorani, N.; Mehrdad, A. CO_2 solubility in some amino acid-based ionic liquids: Measurement, correlation and DFT studies. *Fluid Phase Equilib.* **2020**, *517*, No. 112591.
(62) Gurkan, B.; Goodrich, B. F.; Mindrup, E. M.; Ficke, L. E.; Massel, M.; Seo, S.; Senftle, T. P.; Wu, H.; Glaser, M. F.; Shah, J. K.; Maginn, E. J.; Brennecke, J. F.; Schneider, W. F. Molecular Design of High Capacity, Low Viscosity, Chemically Tunable Ionic Liquids for CO_2 Capture. *J. Phys. Chem. Lett.* **2010**, *1* (24), 3494−3499.

https://doi.org/10.1021/acssuschemeng.5c06870
ACS Sustainable Chem. Eng. 2025, 13, 19977−19994