Title

Physics-Based Machine Learning Models Predict Carbon Dioxide Solubility in Chemically Reactive Deep Eutectic Solvents.

Permalink

https://escholarship.org/uc/item/6s99q12f

Journal

ACS Omega, 9(17)

Authors

Mohan, Mood
Demerdash, Omar
Simmons, Blake
et al.

Publication Date

2024-04-30

DOI

10.1021/acsomega.4c01175

Peer reviewed

Physics-Based Machine Learning Models Predict Carbon Dioxide Solubility in Chemically Reactive Deep Eutectic Solvents

Mood Mohan,* Omar N. Demerdash, Blake A. Simmons, Seema Singh, Michelle K. Kidder,* and Jeremy C. Smith*

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Carbon dioxide (CO_2) is a detrimental greenhouse gas and is the main contributor to global warming. In addressing this environmental challenge, a promising approach emerges through the utilization of deep eutectic solvents (DESs) as an ecofriendly and sustainable medium for effective CO_2 capture. Chemically reactive DESs, which form chemical bonds with the CO_2, are superior to nonreactive, physically based DESs for CO_2 absorption. However, there are no accurate computational models that provide accurate predictions of the CO_2 solubility in chemically reactive DESs. Here, we develop machine learning (ML) models to predict the solubility of CO_2 in chemically reactive DESs. As training data, we collected 214 data points for the CO_2 solubility in 149 different chemically reactive DESs at different temperatures, pressures, and DES molar ratios from published work. The physics-driven input features for the ML models include σ-profile descriptors that quantify the relative probability of a molecular surface segment having a certain screening charge density and were calculated with the first-principle quantum chemical method COSMO-RS. We show here that, although COSMO-RS does not explicitly calculate chemical reaction profiles, the COSMO-RS-derived σ-profile features can be used to predict bond formation. Of the models trained, an artificial neural network (ANN) provides the most accurate CO_2 solubility prediction with an average absolute relative deviation of 2.94% on the testing sets. Overall, this work provides ML models that can predict CO_2 solubility precisely and thus accelerate the design and application of chemically reactive DESs.
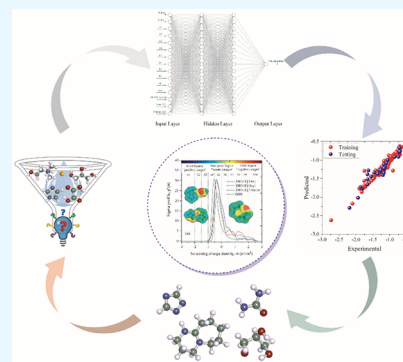
INTRODUCTION

Carbon dioxide (CO_2) is the primary contributor to greenhouse gas emissions, with ca. 80% resulting from the combustion of fossil fuels in transportation, electric power, and industrial sectors.[1] With the increasing utilization of fossil fuels in recent decades, the amount of CO_2 emitted into the atmosphere has been rising, resulting in climate change and severe impacts, including extreme weather events (floods, blizzards, and storms), drought, sea-level rise, and disturbed water systems.[2] With the aim of reducing carbon emissions, many technologies for carbon capture have been examined. However, the high cost of conventional technologies, including absorption, adsorption, membrane, and cryogenics, together with other challenges (e.g., material corrosion and secondary-pollution for solvent scrubbing) imposes great challenges for their practical implementation.[3,4] Therefore, there is an urgency to develop high-efficiency and novel processes for CO_2 capture. Recently, ionic liquids (ILs) have been demonstrated as potential solvents for CO_2 capture[5,6] and have been extensively studied due to their attractive properties.[6−8] However, due to multiple steps involved in the synthesis and purification processes, ILs are expensive solvents. For this reason, deep eutectic solvents (DESs) have emerged as a promising alternative in a wide variety of research areas

and industries, including CO_2 capture, biomass processing, nanotechnology, extraction processes, electrochemistry, and catalysis.[9,10] Using DES for carbon dioxide capture holds great promise for a sustainable and environmentally responsible approach to reducing greenhouse gas emissions from various industrial processes, as they can be made from renewable resources, where their synthesis and use have low energy requirements, do not require oxygen-free environments, have little environmental impact for waste disposal, and most are recyclable without the need for toxic solvents or energy requirements.

DESs first appeared in the literature in 2003, thanks to pioneering work by Abbott et al.[11] When compared to ILs, DESs offer a few primary advantages, the most notable of which is perhaps that the preparation of DESs is simple and economical, requiring no additional purification steps, little to no heating, and no additional costly or toxic solvents are

19548

https://doi.org/10.1021/acsomega.4c01175
ACS Omega 2024, 9, 19548−19559

required.[12,13] A fascinating property of DESs is their structural and chemical diversity, which allows for tuning of desirable properties needed for sustainable processes under varied environmental conditions. DESs are prepared by mixing a hydrogen bond acceptor (HBA) and hydrogen bond donor (HBD) at a specific molar ratio, and the resulting mixture turns into a liquid that is driven by strong interactions between the HBA and the HBD.[13,14]

In recent years, DESs have been demonstrated as potential solvents for CO_2 absorption.[15−17] However, depending on whether DESs are physically based or chemically reactive (i.e., involving chemical bond formation), the CO_2 gas absorption capacity, selectivity, and absorption enthalpy behaviors are different. For the physically based DES, the gas absorption capacity is in accordance with Henry's law constant and the structure of HBA and HBD. The CO_2 absorption enthalpy is low for physically based DES. In one example, Li et al.[18] studied a series of [Ch]Cl-based DESs and found relatively low CO_2 solubility. Physically based DES was suggested to limit the absorption of CO_2 in these [Ch]Cl-based DESs. Furthermore, Wang et al.[19] investigated the influence of temperature and pressure on the CO_2 absorption capacity of physically based DES, i.e., [ATPP]Br-phenol, [TBP]Br-phenol, and [TBP]Br-diethylene glycol. The highest absorption amount was obtained by [ATPP]Br-phenol (1:4) at 313.15 K and high pressure (13.3 bar) with a mole fraction of 0.1974 (1.62 mol of CO_2 per kg of DES).

Considering the needs of practical application, the absorbent needs to be further improved from the physically based variety to increase the absorption performance. Therefore, chemically reactive DESs (i.e., DBN-EU, [HDBU][Triz]-EG, [DBNH][2-MeIm]-EG, L-arginine-triethanolamine, etc.) were introduced and prepared to enhance the CO_2 solubility. With the introduction of amine-functionalized groups in the DESs (examples: superbases and amines) that react with CO_2, enhanced CO_2 absorption capacity was realized. In one example, Garcia-Argüelles et al.[20] prepared a novel super-base−based DES by mixing superbase 1,8-diazabicyclo[5.4.0]-undec-7-ene (DBU) and 1,5,7-triazabicyclo[4.4.0]dec-5-ene (TBD) with benzyl alcohol and ethylene glycol at different molar ratios and showed this is a potential solvent for CO_2 capture. Later, Jiang et al.[21] and Yan et al.[22] obtained high CO_2 absorption uptake by preparing a DESs using superbase solvents to increase the number of active sites in the absorption system. Recently, Wang et al.[23] reported that the DES composed of biophenol-derived superbase ionic liquids and ethylene glycol (EG) exhibits a high CO_2 capacity, up to 1.0 mol CO_2/mol DESs, which is much better than both those of the parent ILs and physically based DES.

Recently, Lemaoui et al.,[24] Wang et al.,[25] and us[26] developed ML and QSPR models to predict the CO_2 solubility in physically based DESs. However, unlike physically based DESs, the chemically reactive DESs have different mechanism for CO_2 absorption, and the absorption of CO_2 is much higher in chemically reactive DESs than physically based DES. Unfortunately, to date, the majority of the research into CO_2 absorption using chemically reactive DESs has relied on experimental methods, which have only been able to address a small fraction of potential DES candidates.[20,22,23,27,28] Because of high structural diversity, the experimental screening of a very large number of DES combinations for their capacity to solubilize CO_2 is impractical. Therefore, in this context, it is highly desirable to have a reliable computational model for

predicting CO_2 solubilities in chemically based DESs. This would reduce both the cost and the time required to develop effective solvent systems for carbon capture and utilization. In a first attempt, recently, Liu et al.[29] used the COSMO-RS (conductor-like screening model for real solvents), an effective quantum chemical computational method for calculating thermodynamic properties and for screening solvents for gas solubilities, to calculate the solubility of CO_2 in 39 different chemically reactive DESs. However, high deviation (195%) was found between experimental and COSMO-RS-calculated CO_2 solubilities in chemically reactive DESs, and none of the other previously reported theoretical models investigated chemically reactive DESs.[29]

Given the limitations of the COSMO-RS model in predicting CO_2 solubilities in chemically reactive DESs, a potentially useful approach to obtaining an accurate and cost-effective tool is to integrate COSMO-RS with machine-learning (ML) models based on quantitative structure−property relationships (QSPR). Further, there is no computational tool to accurately calculate the solubility of CO_2 in chemically reactive DESs. The present study explores different ML models in this regard.

To begin with, a comprehensive survey of the published experimental results of the CO_2 solubility was carried out for different chemically reactive DESs under different experimental conditions (temperature, pressure, and molar ratios). A range of ML algorithms, namely, artificial neural networks (ANN), support-vector regression (SVR), random forest (RF), and gradient-boosted trees (GBT), were investigated in developing ML models for the prediction of CO_2 solubilities. The ML models were ranked based on metrics of model accuracy and precision. Moreover, a *post hoc* rationalization of the ML models' performance is also performed. We also performed a Shapley additive explanation (SHAP) analysis to interpret the ML results and characterize the feature importance.

■ COMPUTATIONAL DETAILS

COSMO-RS Model. COSMO-RS calculations were performed to calculate the carbon dioxide (CO_2) solubility in chemically reactive DESs. The COSMO-RS model predicts thermodynamic properties of any chemical compound by creating a virtual conductor around each molecule, upon which the surface area and screening charge density of each molecular surface segment are calculated, and based on this, the σ-profiles are measured.[30] The σ-profile of a molecule is a probability distribution that quantifies the relative probability of a molecular surface segment having a certain screening charge density.[31] Recent studies demonstrated that the COSMO-RS-derived σ-profile features can be used to build ML models that predict thermodynamic properties of solvents as well as predict the CO_2 solubilities in physically based DESs.[26,32,33] The detailed discussion on the calculation of COSMO and CO_2 solubility predictions in chemically reactive DESs are provided in Section S1.

Database of CO_2 Solubility in DESs. We carried out a comprehensive survey of the published experimental results for CO_2 solubility. These studies were carried out for a variety of chemically reactive DESs at different temperatures (293.15 to 353.15 K), pressures (10.1 to 160.3 kPa), and molar ratios (3:1 to 1:8); altogether, we found 214 data points for 149 DESs. As per the literature,[20,29,34−36] DES with amine groups and amine-functionalized groups (examples: alkanolamines, superbases, imidazole, and amines) that are highly reactive with CO_2 are
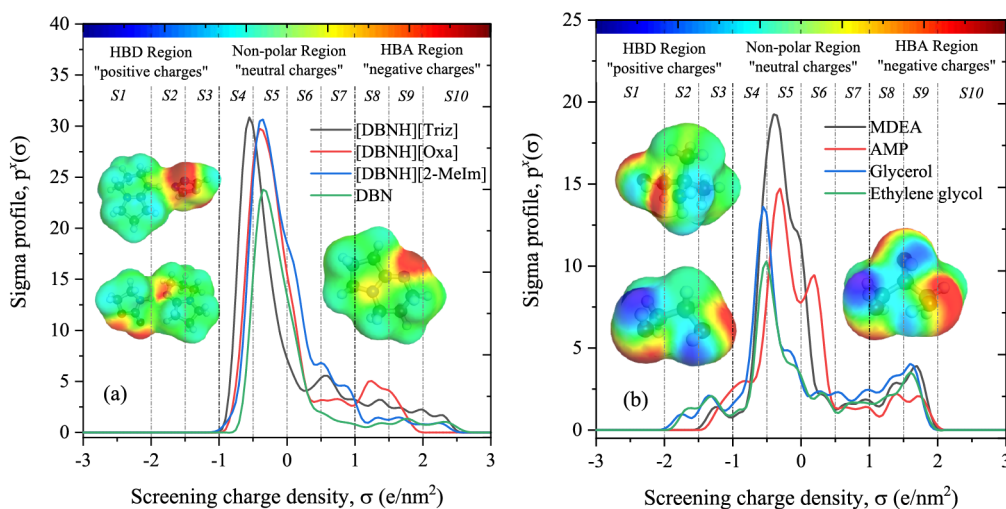
Figure 1. Representation of the ten $S_{\sigma\text{-profile}}$ descriptors in the σ-range for the (a) HBA and (b) HBD of DESs along with their COSMO cavities. The molecular polarity is graphically represented by the colors blue and red, where blue is the negative screening charge density ("*i.e.*, hydrogen bond donating capability"), while red represents the positive screening charge density ("*i.e.*, hydrogen bond accepting capability"). The green and yellow color regions characterize "neutral or nonpolar" molecular surfaces.

considered in this study. It is important to mention that as compared to previous ML data sets reported in the literature, the number of data points in this study is lower; however, the number of individual DESs in the present study is higher (149 DESs). Recently, Lemaoui et al.[24] developed a QSPR model to predict the solubility of CO_2 in physically based DESs. They used 2327 data points with 94 individual DESs. Wang et al.[25] developed a random forest ML model to predict the CO_2 solubility in 59 physically based DESs with 1011 data points. In contrast, our data set contains a larger number of DESs (149), with high structural diversity of the HBA and HBD and DESs molar ratios. All the DES constituents, involving 43 HBAs and 18 HBDs, are given in Figures S1 and S2. The COSMO files for all the molecules were generated based on the procedure outlined in the first paragraph of Section S1.

Calculation of COSMO-RS Input Features for Machine Learning Model. As outlined in the section "COSMO-RS Model", the COSMO files of the investigated molecules were generated and used for calculation of the σ-profiles. Figure 1 shows the σ-profiles of several examples of the HBA and HBD along with their COSMO cavities. As a result, the integrated area under the σ-profile curve over discrete segments may be used to obtain a physiochemically informative description of the surface of a molecule, which is designated as the $S_{\sigma\text{-profiles}}$. The $S_{\sigma\text{-profiles}}$ molecular parameter is an *a priori* quantum chemistry parameter that characterizes the probability of a molecule having a surface polarization charge within a discrete bin, the σ-range. More information on the $S_{\sigma\text{-profiles}}$ molecular descriptor can be found in the work of Torrecilla et al.[37]

Figure 1 displays the σ-profiles of the HBAs and HBDs of several DESs. It has been seen previously that the σ-profile distributions in hydrogen bond donor and acceptor regions as well as the σ-profile of the molecules vary widely, revealing a unique σ-profile property for each molecule.[38] The σ-profiles are categorized into three regions: H-bond acceptor ($\sigma > 1$ e/ nm^2), H-bond donor ($\sigma < -1$ e/nm^2), and nonpolar, i.e., hydrophobic (-1 e/nm$^2 < \sigma > +1$ e/nm^2) regions. To determine the input features based on the σ-profiles for the machine learning model, the σ-profiles of DES constituents were divided into 10 fractions (i.e., S1–S10) by integrating σ-

profile $p^x(\sigma)$ curves over the discrete bins of the screening charge density, σ. As exemplified by HBA and HBD in Figure 1a, b, the fractions of the $S_{\sigma\text{-profiles}}$ are classified into five classes depending on the screening charge densities: (1) The strong donor region [S1 and S2], (2) the weak donor region [S3], (3) the nonpolar region [S4, S5, S6, and S7], (4) the weak acceptor region [S8], and (5) the strong acceptor region [S9 and S10]. The $S_{\sigma\text{-profiles}}$ of modeled DESs are discussed Section S2.

■ MACHINE LEARNING MODELS

Artificial Neural Network. In the present study, an artificial neural network (ANN)-based ML model was developed using the JMP Pro statistical software (JMP Pro SAS 17.2.0)[39] by utilizing the temperature, pressure, dissociation constant (pK_a) of acidic and basic, viscosity of DESs, molecular surface area, and the 10 $S_{\sigma\text{-profiles}}$ molecular descriptors as input features to predict the solubility of CO_2 in DESs as an output variable. In training the ANN model, a nonparametric function f is sought that maps the input features to the predicted solubility of CO_2 as shown below:

$$x_{CO_2}^{DES} = f\left(T, P, pK_a, \eta, A, S_{1,\sigma\text{-profile}}^{DES}, S_{2,\sigma\text{-profile}}^{DES}, \dots \right.$$
$$\left. , S_{10,\sigma\text{-profile}}^{DES}\right) \tag{1}$$

where $x_{CO_2}^{DES}$ is the solubility of CO_2 in DES, T and P are the temperature (K) and pressure (kPa). pK_a is the dissociation constant for HBA and HBD under basic and acidic conditions. η is the viscosity (in mPa·s) of DES. A is the molecular surface area of the DES. $S_{i,\sigma\text{-profile}}$ is the descriptor in the σ-profile region "i" i.e., from S1 to S10. For the ANN, 80% of the data were used for training, and 20% of the data were used for testing. The ANN model was developed with 2 hidden layers and 20 neurons in each hidden layer. The network's learning rate was fixed to 0.1, the number of tours was set to 10,000, and a squared penalty method was used for optimization. All other options in the JMP Pro SAS 17.2.0 software was kept as default.[39] The principle of the ANN algorithm and more details of the ANN model are given in Section S3.1.

Support Vector Machine. Support vector machine (SVM) is based on a statistical learning theory developed by Vapnik (Vapnik, 1995).[40] SVM is a popular supervised ML method that can be applied to classification and regression problems. Its popularity is largely based on its property of establishing nonlinear relationships between the feature set and the prediction target through the use of kernel functions.[41] When the prediction target is a continuous real value, the models that are fitted are known as support vector regression (SVR) models. In a regression problem, the goal is to fit a model that minimizes the error between a prediction and the target, and the response falls within a range of $-\varepsilon$ to ε.

In this work, statistical software JMP Pro SAS 17.2.0 was also used to develop the SVR model to predict the CO_2 solubilities. The hyperparameters such as kernel (linear and radial basis function [RBF]), width of the RBF kernel, gamma (0 to 0.5), and cost (0 to 100) were tuned, and optimal values were used for the development of SVR model. The RBF kernel was used for the SVR model with gamma and cost values of 0.05 and 43.2, respectively.

Random Forest and Gradient-Boosted Tree. In addition to ANN and SVR, JMP Pro 17.2.0 was used to develop random forest (RF) and Gradient Booster Tree (GBT) models. RF is a set of classification or regression trees, first proposed by Breiman[42] in 2001, that comprise an ensemble of multiple regressors or classifiers. Each decision tree in an RF is independent and can be developed in parallel during the data regression, thus reducing the computational cost of model development. The random selection of features to be used at splitting nodes enables fast training of this algorithm, even in the case of the large dimensionality of the feature vector. Each split in a tree considers a random subset of the predictors. In such a way, many weak tree models are combined to produce a powerful RF model. The final prediction for an observation is the average of the predicted values for that observation over all of the decision trees. All the decision tree's outputs are averaged (b is the number of trees), providing an even more accurate result than the single-tree structure.

$$y = \frac{1}{b} \sum_{j=1}^{b} y_b \qquad (2)$$

where y_b is the prediction of the bth tree and y is the average over all b trees.

Different algorithms based on ensembles of decision trees have been proposed, and among those gradient-boosted trees (GBT) have been considered as one of the most important advances in ML over the last 20 years.[43,44] Boosting a tree is the process of building a large, additive decision tree by fitting a sequence of smaller decision trees, called *layers*.[45] The tree at each layer consists of a small number of splits. The tree is fitted based on the residuals of the previous layers, which allows each layer to correct the fit for bad fitting data from the previous layers. GBT follows three main steps sequentially: It optimizes the loss function, spots the weaker learner, and improves it by adding more trees to increase the accuracy. The final prediction for an observation is the sum of the predictions for that observation over all of the layers. More details of hyperparameter optimization for RF and GBT are provided in Section S3.2. The SHAP analysis for ML model interpretation is also discussed in Section S3.3.

To assess the predictive capability of the developed ML models, performance metrics including the determination coefficient (R^2), average absolute relative deviation (AARD), mean absolute error (MAE), root-mean-square error (RMSE), and Pearson's correlation coefficient (r) were calculated. The best ML model was selected based on the lowest AARD, MAE, and RMSE and highest R^2 values.

RESULTS AND DISCUSSION

Development of ML Models for CO_2 Solubility Predictions. We performed an exhaustive literature survey and found that the available data consists of 214 data points for 149 DESs. This was sufficient to develop an accurate model. In the literature, ML and deep learning (DL) models have been often developed in related fields with number of data points similar to ours and achieved good prediction capability. For example, Kartal and Özveren[46] (178 data points), Kardani et al.[47] (81 data points), and Gao et al.[48,49] (125−150 data points) developed ML/DL models with 81−178 data points for the prediction of lignocellulosic biomass composition and conversion of biomass during the hydrothermal carbonization. Recently, Zhang et al.[50] developed ML models with 132 experimental data points for the prediction of the Henry's law constant for CO_2 solubility in ionic liquids (ILs). However, recognizing the fact that our data set was small, we took steps to avoid overfitting in training the ML model. First, the training and testing data were randomly split by using an 80:20 ratio of training to testing data. Second, the data splitting using this same ratio of training/testing was repeated for 15−20 times to generate different compositions of the training and testing set and yielded similar performance of the ML models. The input features for the ML model are COSMO-RS-calculated σ-profile descriptors ($S_{\sigma\text{-profiles}}$-1 to $S_{\sigma\text{-profiles}}$-10), pK_a, ML-predicted viscosity, and the experimental temperatures and pressures. As outlined in the section "Calculation of COSMO-RS Input Features for Machine Learning Model", the COSMO files of investigated molecules were generated and used for the calculation of σ-profiles. The COSMO-RS-derived σ-profile captures the molecular polarity through screening of charge densities on the molecule. Further, to incorporate the chemical reactivity, we calculated the dissociation constants (pK_a) of HBAs and HBDs. The commercial package ChemAxon was utilized for the calculation of pK_a values of HBA and HBD.[51,52] We have also calculated the viscosity of DESs using our in-house ML models[53] and used this as an input feature to study the effect of viscosity in predicting solubility. First, we calculated the solubility of CO_2 in chemically reactive (DESs) using the COSMO-RS and multilinear regression (MLR) models. However, these models demonstrated weaker predictive capabilities and significant deviations. For a detailed discussion of these models, please refer to Section S4.1. in the Supporting Information.

Figure 2 illustrates the correlation of experimental and ML-predicted CO_2 solubilities in the training and testing sets for different ML models. As depicted in the parity plot in Figure 2, the predictions for the training sets are in excellent agreement with the experimental data with high accuracy. However, the RF, SVR, and GBT models show weaker predictions on the test sets, with lower accuracy, i.e., $R^2 = 0.790-0.837$, while the ANN model predictions show an excellent agreement with experiment ($R^2 = 0.989$), indicating that ANN model predictions are more accurate than the other models. Table S1 lists the statistical parameters (R^2, AARD, MAE, and RMSE)
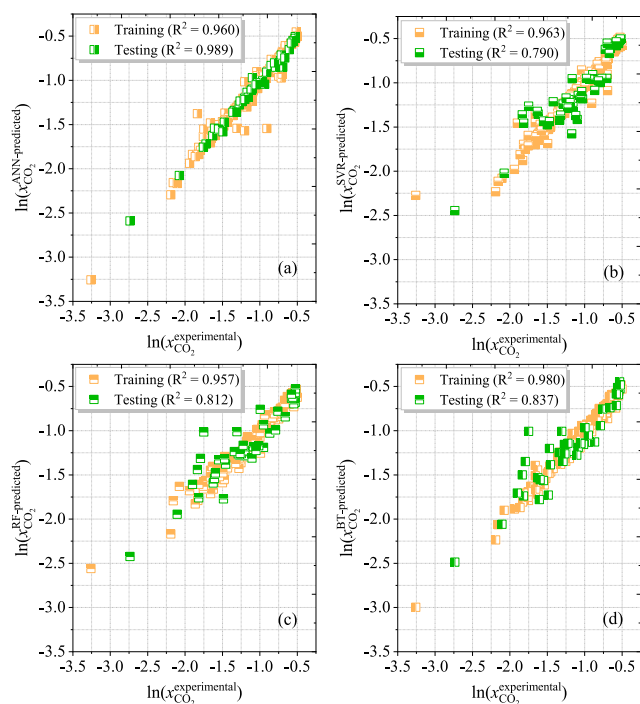
Figure 2. Experimental and predicted CO_2 solubility in chemically reactive DESs using (a) an ANN, (b) SVR, (c) RF, and (d) GBT ML models on training set and testing sets.

for the ML models. R^2, AARD, MAE, and RMSE values for the ANN model are 0.989, 2.94%, 0.029, and 0.051, respectively, for the test data. Again, RF, SVR, and GBT models show desirable levels of accuracy for the training data sets but lower accuracy on the test set, with higher RMSE and AARD values. In the statistical parameter estimations, the ANN model shows

low bias and low variance, while the RF, SVR, and GBT models have low bias but high variance. A model with minimal bias and variance is regarded as an optimal ML model.

Furthermore, statistical residual analysis was also performed for all the ML models and confirmed the goodness-of-fit through a probability plot of the relative deviations, relative deviations vs experimental CO_2 solubility, and standard residuals vs predicted CO_2 solubility values. Figures 3 and S6 depict the statistical analysis plots for all of the ML models. The ANN model shows that the majority of the CO_2 solubility relative deviations are within 15% with an AARD of 3% and RMSE of 0.07. The distribution of the relative deviations in different ARD ranges is also shown in Figure 4; the majority of
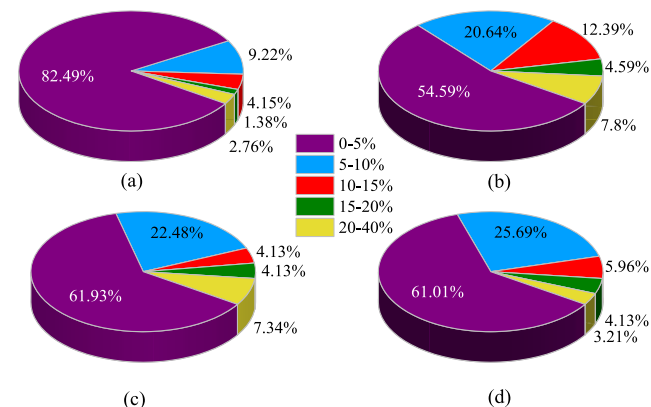

Figure 4. Distribution of the absolute relative deviation in different deviation ranges: (a) ANN, (b) RF, (c) SVR, and (d) GBT.

the CO_2 solubility prediction data (~92%) lies within 10% of the AARD and 96% of the data within 15% of the AARD. Only 2.76% of the data lie beyond 20% of the AARD. These results
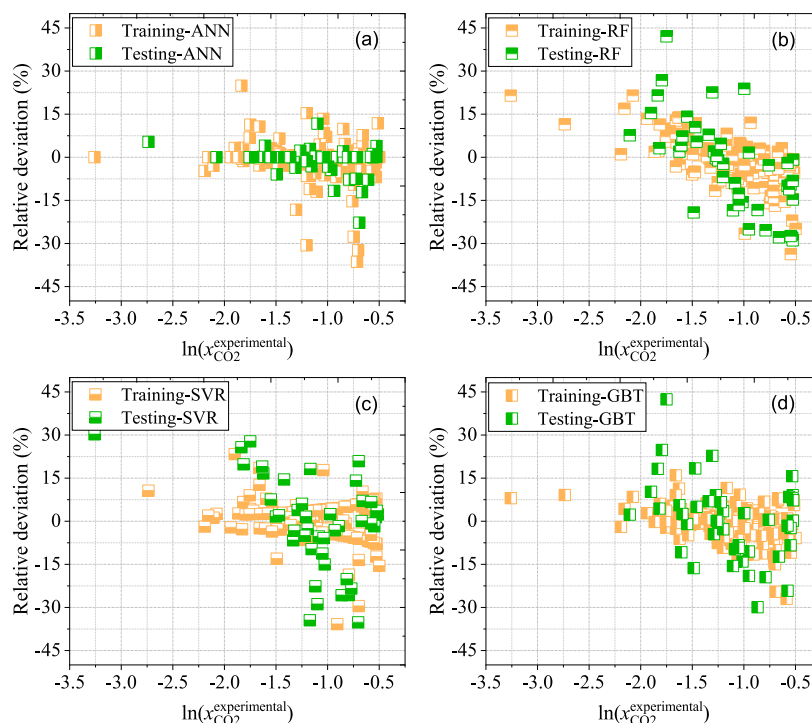

Figure 3. Relative deviation between the experimental and predicted CO_2 solubilities in chemical-based DESs: (a) ANN, (b) RF, (c) SVR, and (d) GBT.

clearly demonstrate the accuracy of the developed ANN model. From Figure 4, it is also seen that the GBT model shows ~93% of CO_2 solubility data within 15% of AARD. However, the GBT model shows higher deviations on the test data (Figure S6). For RF and SVR models, only 88% of the data lies within 15% of AARD.

A Taylor diagram was generated to further analyze the statistical metrics for a more understandable representation of ML model performances (Figure 5a). Three performance metrics (the standard deviation (SD), the RMSE, and the Pearson correlation coefficient (r)) of each ML model, including ANN, SVR, RF, and GBT, are utilized to quantify the degree of discrepancy between the model predictions and the related experimental values. Figure 5a shows that the ANN
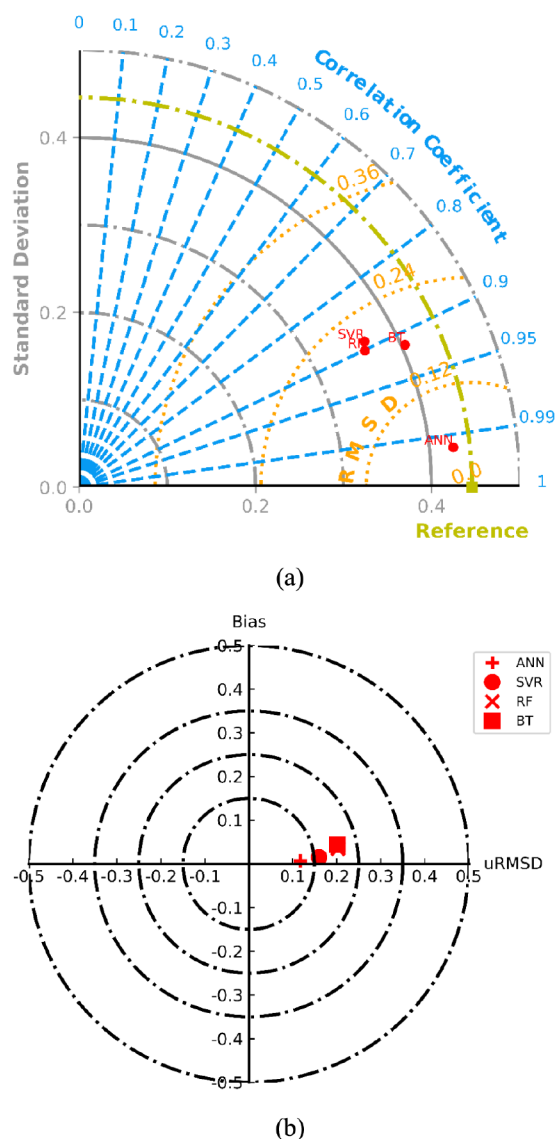

(a)

(b)

Figure 5. (a) Taylor diagram showing the performance of the different ML models on testing data sets. The sky blue dashed line indicates pearson's correlation coefficient (r), the gray dashed line indicates the standard deviation (SD), and the orange dots represent the RMSD values. The red circle represents four different ML models. The green line represents the experimental measured reference SD. (b) Bull's eye plot showing the correlation between bias and uRMSD for evaluating the performance of ML models on testing data sets.
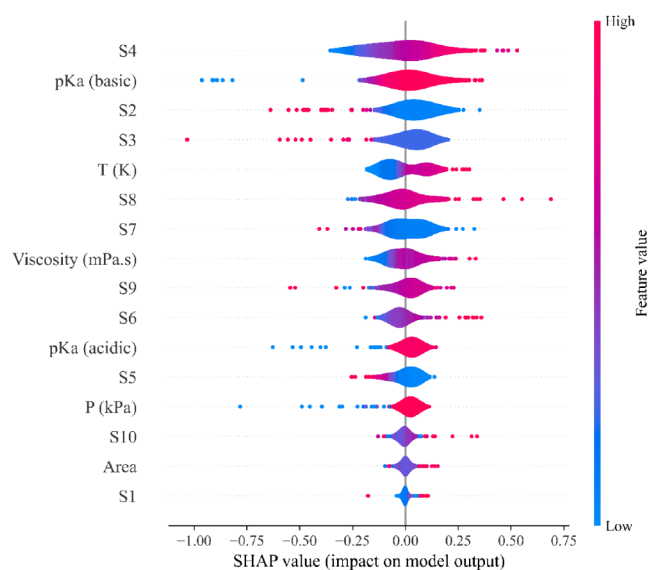
model has a lower RMSD and higher r values than the RF, SVR, and GBT models. Moreover, the SD of the ANN is closer to the experimental SD. The RF and SVR models showed larger deviations with respect to experiment as compared to ANN, which is in accord with the ML predictions. The ANN model showed a lower RMSD, higher correlation coefficient (r), and closer SD values, which confirm that the ANN has superior predictive power compared with models trained with the other algorithms.

The bull's eye plot in Figure 5b depicts the correlation between bias and unbiased root-mean-square deviation (uRMSD) for evaluating prediction errors of ML models on testing data sets. Figure 5b provides three critical insights: (1) whether the model overestimates or underestimates (positive or negative values of the bias on the y-axis, respectively), (2) whether the model standard deviation is larger or smaller than the standard deviation of the experimental measurements (positive or negative values on the x-axis, respectively), and (3) the error performance as quantified by the uRMSD represented as the distance to the coordinates' origin. Models with high uRMSD are overtrained and do not generalize to the test data. The first dashed circle line near to the origin that forms a "bull's eye" indicates the observational uncertainty and communicates the estimated limits of model performance. Figure 5b reveals that model predictions from the ANN have a smaller uRMSD and a lower bias than SVR, RF, and GBT. Therefore, it is evident that the ANN provides the best ML model prediction because it is closest to the origin and has the smallest bias.

Importance of Input Features and Rationality of Developed ML Models. A covariance matrix analysis was performed between ML input features to investigate the correlation between pairs of ML features, as well as between individual features and the experimental CO_2 solubility (Figure S7). In the context of a covariance matrix, each element of the matrix represents the covariance between two features. Covariance is a measure of how much two variables vary together; a positive covariance indicates that as one variable increases, the other tends to increase, while a negative covariance indicates that as one variable increases, the other tends to decrease. A linear correlation between two features is high when the corresponding covariance is close to 1 (positively correlated) or −1 (negatively correlated). From Figure S7, there is no significant linear correlation between input features of ML except for $S_{\sigma\text{-profiles}}$−5 (S5), $S_{\sigma\text{-profiles}}$−6 (S6) of the sigma profile descriptor, and surface area. The lack of linear correlation indicates that the features are non-redundant, indicating that they may each make a unique contribution to a ML model. A positive influence of the input features on the CO_2 solubility is indicated by a positive covariance matrix value, while a negative covariance matrix value indicates a negative influence. Pressure (P), S2, S4, S5, S6, S7, S8, area, and pK_a show a modest positive correlation with the CO_2 solubility, indicating that as the value of these parameters increases, the solubility of CO_2 tends to increase as well.

Furthermore, the importance of input features on the ANN model is interpreted using the SHAP analysis, and the results are depicted in Figure 6. An input feature with SHAP value >0 implies the model uses that feature to increase the magnitude of CO_2 solubility, while SHAP value <0 decreases the CO_2 solubility. The polar and nonpolar σ-profile regions of DES (features S2, S3, S4, and S8), pK_a at basic conditions,

Figure 6. Importance of input features on the solubility of CO_2 predictions using the ANN model.

der Waals (vdW) interactions would also correlate with CO_2 solubility, given that nonpolar features are known to correlate with these two properties. pK_a in basic conditions also plays an important role in the CO_2 solubility, and higher pK_a (basic) value has an influence on the CO_2 solubility. Sang et al.[35] reported the solubility of CO_2 in 1,5-diazabicyclo [4.3.0]-non-5-ene (DBN)-based DESs ([DBNH][Triz]-EG, [DBNH]-[Oxa]-EG, and [DBNH][2-MeIm]-EG). The DES [DBNH]-[Oxa]-EG at a 2:1 molar ratio shows a higher solubility of CO_2 than [DBNH][2-MeIm]-EG and [DBNH][Triz]-EG at a 2:1 molar ratio. This is mainly due to the higher pK_a (basic) value of [DBNH][Oxa]-EG (pK_a = 20.8 in DMSO) as compared with those of [DBNH][2-MeIm] (pK_a = 19.3 in DMSO) and [DBNH][Triz] (pK_a = 14.75 in DMSO).

It is interesting to mention that generally, the CO_2 solubility decreases with increasing temperature. In contrast, in chemically reactive DESs, the solubility of CO_2 increases with the temperature. Thus, from the SHAP analysis (Figure 6), a higher temperature leads to a higher CO_2 solubility. Ren et al.[54] reported the solubility of CO_2 in amino acid-based DESs. Indeed, as the temperature increases from 313.15 to 333.15 K, the solubility of CO_2 increases. Also, Sang et al.[35] measured the solubility of CO_2 in [DBNH][Oxa]-EG (1:0.5) DESs and reported that the solubility of CO_2 increased from 293.15 to 313.15 K. The results indicate that too high or too low temperatures are not beneficial for high CO_2 solubility in

temperature, and viscosity are found to be the most important features. These observations are consistent with experimental trends, as the DES with higher nonpolar regions (S4) tend to increase CO_2 solubility. One can infer that free volume and van
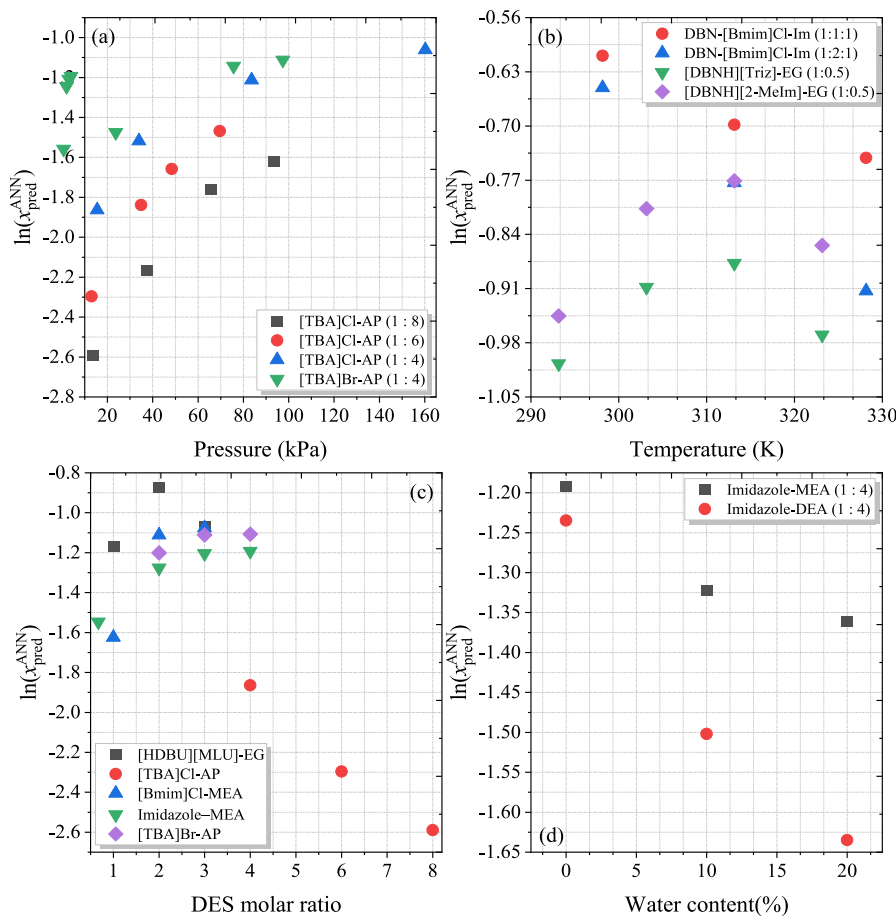

Figure 7. ANN-based ML predicted CO_2 solubilities in (a) [TBA]-based DESs at different pressures at 298.15 K, (b) DBN-based DES at different temperatures at 100 kPa, (c) effect of DES molar ratio on CO_2 solubility, [HDBU][MLU]-EG at 313.15 K, Imidazole-MEA at 303.15 K, and [TBA]Cl-AP, [Bmim]Cl-MEA, and [TBA]Br-AP data at 298.15 K, and (d) effect of water content on the CO_2 solubility.

Table 1. Comparison of Experimental and ML-Predicted CO_2 Solubilities ($\ln(x)$) in External Chemical-Based DESs[a][b]

DES	experimental, $\ln(x)$	ANN	RF	SVR	GBT	MLR
TMG-glycerol (1:1)	−1.633	−1.632	−1.408	−1.323	−1.583	−0.982
TMG-glycerol (1:3)	−1.660	−1.707	−1.524	−1.450	−1.511	−0.952
DBU-glycerol (1:3)	−1.537	−1.526	−1.386	−1.453	−1.540	−0.988
DBN-glycerol (1:3)	−1.551	−1.551	−1.441	−1.509	−1.547	−1.101

[a]These DESs were not considered in the training and testing data sets. [b]The experimental data is taken from Huang et al.[64]
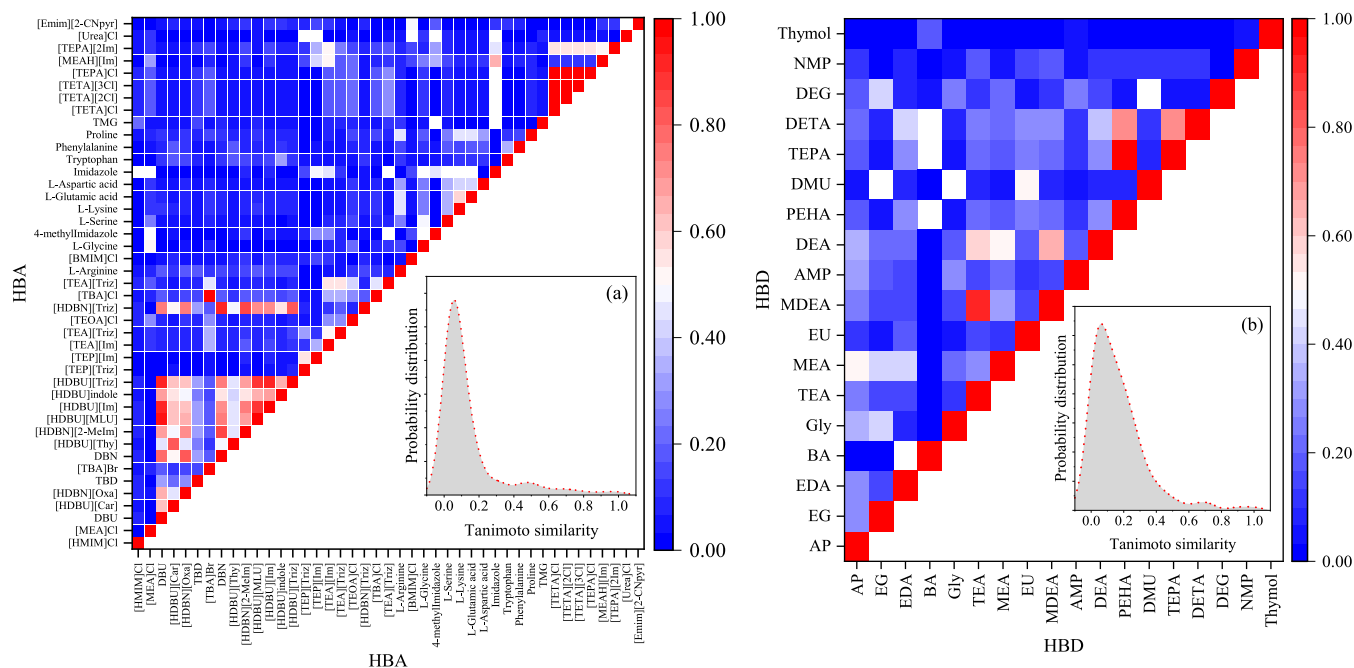

Figure 8. Covariance matrix of Tanimoto similarities for (a) hydrogen bond acceptors (HBA) and (b) hydrogen bond donors (HBD).

chemically reactive DESs. Features S2 and S3 are also shown to influence the CO_2 solubility. In contrast, DESs with high values of σ-profile polar features S8, S9, and S10 tend to exhibit increased CO_2 solubility. Also, DESs with lower electron donor regions (S2 and S3) and higher electron accepting regions (S8, S9, and S10) are seen to be better solvents for CO_2 because the intramolecular interaction within these DESs will be weaker and leave greater capability for nucleophilic attack.[35,36] A larger value of the nonpolar feature S4 is advantageous for physical absorption, while higher values of the polar features (S8, S9, and S10) and pK_a are favorable for chemical absorption.

Rainbolt et al.[55] reported the solubility of CO_2 in tertiary alkanolamines. The reported alknolamines were found to absorb CO_2 via both chemical binding and physical absorption. For example, N,N-dimethylethanolamine (DMEA) captures 8.6% of CO_2 through physical absorption and 12% through chemical absorption at 3447.37 kPa and 298.15 K. Further, the viscosity of DESs is also a critical feature for CO_2 solubility predictions. In general, DESs with a low viscosity have high CO_2 solubility. However, according to the SHAP analysis, for chemical DESs, a higher value of viscosity benefits the CO_2 solubility. The DESs with larger free volume and stronger vdW interactions result in increased CO_2 solubility; on the other hand, the solvent with higher free volume and stronger interactions leads to higher viscosity of DESs.[56,57] It is well-known that pressure is directly proportional to gas solubility.[26,58] From our predictions, pressure shows a positive

effect on the CO_2 solubility predictions, which is consistent with experimental measurements and theory.

To further evaluate the reliability of the ML models developed in this work, the effects of experimental parameters such as temperature, pressure, DES molar ratio, HBA, HBD, and water content on the CO_2 solubility predictions were investigated and compared to those of experimental measurements. Figure 7 shows the rationality of ANN-based model on the solubility of CO_2 in chemically reactive DESs, and it is clear that the developed ANN model shows excellent performance and explain the ability to predict CO_2 solubility and reproduce experimentally observed trends. Further the rationality of ML models is discussed in Section S4.2.

Apart from the rationality of ML models, we have validated our ML models on an external data set to ensure robustness. Again, the ANN model predictions are more accurate than the other models and closer to the experimental CO_2 solubilities with an AARD of 1% (see Table 1). As we discussed earlier, the RF, SVR, and GBT models are apparently more biased toward the observed higher deviations of their corresponding predictions from experimental data; thus, the predictions of these models are not as reliable as ANN.

Tanimoto Similarities. An important question concerns how different from each other the DESs were used as input. To search the chemical structural similarities, we performed Tanimoto similarity[59,60] analysis for HBAs and HBDs based on the Daylight fingerprints[61,62] (2048 bits) using an RDKit Python tool.[63] For each pair of molecular fingerprints, corresponding to two molecules, Tanimoto scores were
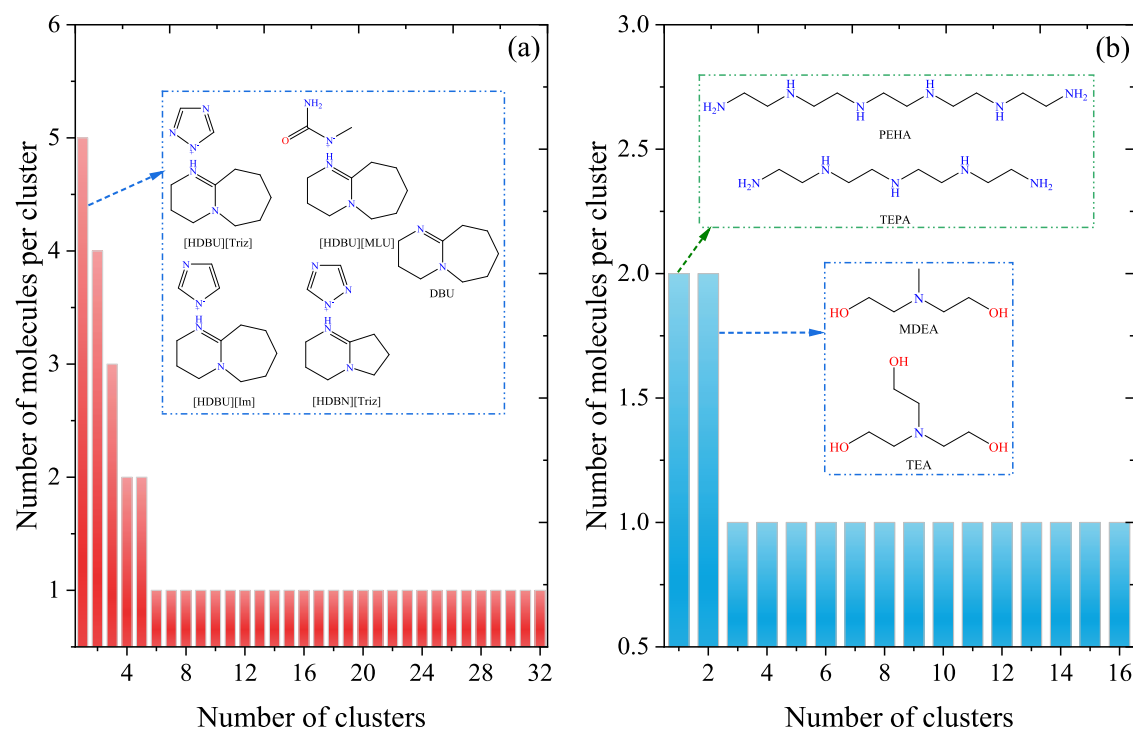
Figure 9. Clustering analysis of DES components (a) HBA and (b) HBDs along with their largest cluster molecular structure.

calculated, providing the structural similarity of that pair. If the Tanimoto score is above 0.85 or closer to 1, the two molecular structures are deemed highly similar, and if lower than 0.5, they are highly dissimilar. Figure 8 shows the Tanimoto similarity scores for HBAs and HBDs. In Figure 8a, only superbase—based HBAs are similar to scores greater than 0.85. A large peak is seen at 0.1, which indicates that many HBAs are highly dissimilar. Smaller propensities are seen at higher similarity scores, implying that identical chemical structural space of HBA is very low.

Further, we also calculated the number of clusters based on the chemical structures and Tanimoto similarity scores using Butina algorithm[61] and found that 32 clusters were formed from 43 molecules at a cutoff distance of 0.2. Twenty-seven clusters had only one compound, two clusters contained two compounds, and one cluster each contained three, four, and five HBAs, respectively (Figure 9a). Figure 8b shows the Tanimoto similarity scores of the HBDs. Only TEA and MDEA, and TEPA and PEHA are highly similar to each other, and the remaining HBDs are dissimilar (Figure 9b). Based on the Tanimoto similarity scores, HBD forms 16 clusters where 14 clusters contain only one molecule and two clusters contain two molecules (i.e., TEA—MDEA and TEPA—PEHA). From the Tanimoto similarity and associated clustering analysis, it is clear that the HBAs and HBDs comprise a diverse set covering a large chemical structural space of DESs.

Development of Novel DESs for Improving CO_2 Solubility. After the successful development of ML models and the careful evaluation of CO_2 solubility predictions in 149 chemically reactive DESs, the ANN model is used to predict the solubility of CO_2 in new HBA and HBD combinations whose CO_2 solubilities have not been tested and reported in the literature. We utilized SHAP analysis predictions (Figure 6) to develop the novel combination of DESs. Figure 10 shows the solubilities of CO_2 in 11 different DESs at 298.15 K and 100 kPa. Among 11 DESs, the HBA ethanolamine (MEA) with
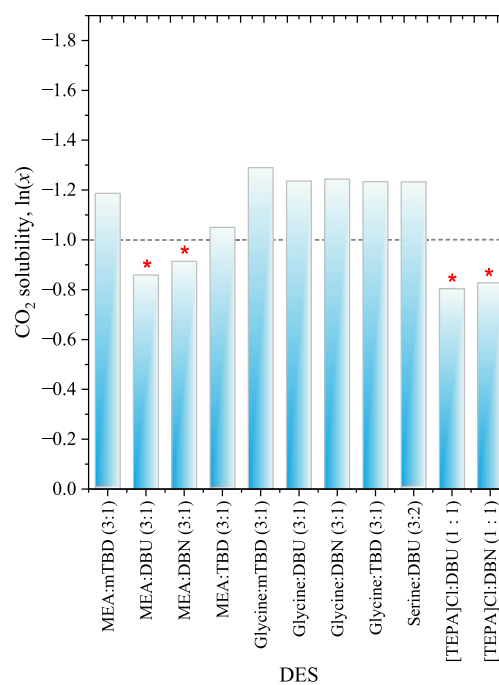

Figure 10. Development of new DESs combination for improving CO_2 solubilities in chemically reactive DESs using the ANN model. *These DESs have highest solubility of CO_2 at 298.15 K and 100 kPa.

superbase DBN and DBU (3:1 molar ratio) and [TEPA]Cl with DBN and DBU (1:1 molar ratio) have shown to be better solvents for improved CO_2 solubility. The higher solubilities in MEA:DBU, MEA:DBN, [TEPA]Cl:DBU, and [TEPA]-Cl:DBN are due to the higher pK_a and stronger chemical reactivity of these solvents. Furthermore, the COSMO-RS calculations were performed to confirm the molar ratio and eutectic point of newly developed DES combinations. Since

the phase transition properties (i.e., melting point and heat fusion values) of all the HBAs and HBDs were not available in the literature, we performed COSMO-RS calculations for ethanolamine and mTBD as an example of predicting the eutectic point composition. Figure S11 shows the COSMO-RS-calculated eutectic point composition of ethanolamine and mTBD and forms a eutectic point at 263.5 K with an mTBD composition of 0.59. The melting temperature of 3:1 molar ratio of ethanolamine ($x = 0.75$) and mTBD ($x = 0.25$) was 275.15 K.

CONCLUSIONS

This work shows that the ANN model accurately predicts the solubility of CO_2 in chemically reactive DESs. To the best of our knowledge, this is the first attempt to apply ML models to predict the solubility of CO_2 in chemically reactive DESs. We established a database containing 214 experimental data points of CO_2 solubility in 149 chemically reactive DESs at different temperatures, pressures, and molar ratios. First, the COSMO-RS model was employed to calculate the solubility of CO_2 in chemical-based DESs, and a large deviation from experiment (AARD of 329%) was obtained. To address this, advanced ML algorithms were then investigated to determine whether improved performance could be obtained based on the COSMO-RS-derived σ-profile features, pK_a, viscosity, T, and P. Among the four ML methods investigated, ANN shows the optimal performance on the CO_2 solubility predictions, and the overall R^2, AARD, MAE, and RMSE values of this model are 0.966, 3.23%, 0.034, and 0.08, respectively.

From the present study, the following observations were critical and significant for CO_2 capture research:

1. Generally, CO_2 solubility decreases with increasing temperature. However, chemically reactive DESs show an intriguing increase in CO_2 solubility with temperature. Too high or too low temperatures compromise CO_2 solubility in chemically reactive DESs.

2. pK_a of DESs also plays an important role in the CO_2 solubility, as higher pK_a (basic) has greater influence and tends to increase CO_2 solubility.

3. DESs with lower electron donor regions (S2 and S3) and higher electron accepting regions (S8, S9, and S10) are better solvents for high CO_2 solubility because the intermolecular interactions between HBA and HBD are weaker, leading to greater capability for nucleophilic attack between CO_2 and chemical DESs. A larger value of the nonpolar feature S4 is advantageous for physical absorption, while higher values of the polar features (S8, S9, and S10) and pK_a values are favorable for chemical absorption.

4. Viscosity emerges as a critical factor, with lower viscosity generally favored for higher CO_2 solubility, but our SHAP analysis highlights exceptions where higher viscosity positively correlates with increased CO_2 solubility.

In summary, our comprehensive ML models, incorporating diverse features, not only enhance the accuracy of CO_2 solubility predictions in chemically reactive DESs but also unveil unprecedented insights into the intricate interplay of temperature, molecular structure, and viscosity in the CO_2 capture process. We show here that a quantum chemical method that does not explicitly calculate chemical reaction profiles (COSMO-RS) can nevertheless be used to predict

bond formation. Overall, the developed ANN-based ML model accurately predicts the solubility of CO_2 in chemically reactive DESs and thus will permit accurate accelerated screening for CO_2 solubility while optimizing conditions, thus liberating researchers from costly and time-consuming experimental trials.

ASSOCIATED CONTENT

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.4c01175.

> The generation of COSMO files for the COSMO-RS calculations and he hyperparameter optimization for different ML models. The details of COSMO-RS model calculations, MLR predicted CO_2 solubility, and rationality of ML models are also provided along with ML model-predicted CO_2 solubilities. In addition, residual analysis and the rationality of the ML models are provided (PDF)

AUTHOR INFORMATION

Corresponding Authors

Mood Mohan — *Biosciences Division and Center for Molecular Biophysics, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, United States;* ⓞ orcid.org/0000-0001-5937-9746; Email: moodm@ornl.gov

Michelle K. Kidder — *Manufacturing Science Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831-6201, United States;* ⓞ orcid.org/0000-0003-0851-835X; Email: kidderm@ornl.gov

Jeremy C. Smith — *Biosciences Division and Center for Molecular Biophysics, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, United States; Department of Biochemistry and Cellular and Molecular Biology, University of Tennessee, Knoxville, Tennessee 37996, United States;* ⓞ orcid.org/0000-0002-2978-3227; Email: smithjc@ornl.gov

Authors

Omar N. Demerdash — *Biosciences Division and Center for Molecular Biophysics, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, United States*

Blake A. Simmons — *Deconstruction Division, Joint BioEnergy Institute, Emeryville, California 94608, United States; Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States;* ⓞ orcid.org/0000-0002-1332-1810

Seema Singh — *Deconstruction Division, Joint BioEnergy Institute, Emeryville, California 94608, United States*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsomega.4c01175

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

■ REFERENCES

(1) Pishro, K. A.; Murshid, G.; Mjalli, F. S.; Naser, J. Investigation of CO_2 solubility in monoethanolamine hydrochloride based deep eutectic solvents and physical properties measurements. *Chin. J. Chem. Eng.* **2020**, *28* (11), 2848−2856.

(2) Cianconi, P.; Betrò, S.; Janiri, L. The impact of climate change on mental health: a systematic descriptive review. *Front. Psychiatry* **2020**, *11*, 490206.

(3) Gao, W.; Liang, S.; Wang, R.; Jiang, Q.; Zhang, Y.; Zheng, Q.; Xie, B.; Toe, C. Y.; Zhu, X.; Wang, J.; Huang, L. Industrial carbon dioxide capture and utilization: state of the art and future challenges. *Chem. Soc. Rev.* **2020**, *49* (23), 8584−8686.

(4) Dubey, A.; Arora, A. Advancements in carbon capture technologies: A review. *J. Cleaner Prod.* **2022**, 133932. .

(5) Zhang, X.; Zhang, X.; Dong, H.; Zhao, Z.; Zhang, S.; Huang, Y. Carbon capture with ionic liquids: overview and progress. *Energy Environ. Sci.* **2012**, *5* (5), 6668−6681.

(6) Yan, F.; Dhumal, N. R.; Kim, H. J. CO_2 capture in ionic liquid 1-alkyl-3-methylimidazolium acetate: a concerted mechanism without carbene. *Phys. Chem. Chem. Phys.* **2017**, *19* (2), 1361−1368.

(7) Tamilarasan, P.; Ramaprabhu, S. Integration of polymerized ionic liquid with graphene for enhanced CO_2 adsorption. *J. Mater. Chem. A* **2015**, *3* (1), 101−108.

(8) Prakash, P.; Venkatnathan, A. Molecular mechanism of CO_2 absorption in phosphonium amino acid ionic liquid. *RSC Adv.* **2016**, *6* (60), 55438−55443.

(9) Smith, E. L.; Abbott, A. P.; Ryder, K. S. Deep eutectic solvents (DESs) and their applications. *Chem. Rev.* **2014**, *114* (21), 11060−11082.

(10) Hansen, B. B.; Spittle, S.; Chen, B.; Poe, D.; Zhang, Y.; Klein, J. M.; Horton, A.; Adhikari, L.; Zelovich, T.; Doherty, B. W.; et al.et al. Deep eutectic solvents: A review of fundamentals and applications. *Chem. Rev.* **2021**, *121*, 1232−1285.

(11) Abbott, A. P.; Capper, G.; Davies, D. L.; Rasheed, R. K.; Tambyrajah, V. Novel solvent properties of choline chloride/urea mixtures. *Chem. Commun.* **2003**, *1*, 70−71.

(12) Verma, R.; Mohan, M.; Goud, V. V.; Banerjee, T. Operational strategies and comprehensive evaluation of menthol based deep eutectic solvent for the extraction of lower alcohols from aqueous media. *ACS Sustainable Chem. Eng.* **2018**, *6* (12), 16920−16932.

(13) Naik, P. K.; Mohan, M.; Banerjee, T.; Paul, S.; Goud, V. V. Molecular dynamic simulations for the extraction of quinoline from heptane in the presence of a low-cost phosphonium-based deep eutectic solvent. *J. Phys. Chem. B* **2018**, *122* (14), 4006−4015.

(14) Mohan, M.; Naik, P. K.; Banerjee, T.; Goud, V. V.; Paul, S. Solubility of glucose in tetrabutylammonium bromide based deep eutectic solvents: Experimental and molecular dynamic simulations. *Fluid Phase Equilib.* **2017**, *448*, 168−177.

(15) Chen, Y.; Ai, N.; Li, G.; Shan, H.; Cui, Y.; Deng, D. Solubilities of carbon dioxide in eutectic mixtures of choline chloride and dihydric alcohols. *J. Chem. Eng. Data* **2014**, *59* (4), 1247−1253.

(16) Alhadid, A.; Safarov, J.; Mokrushina, L.; Müller, K.; Minceva, M. Carbon Dioxide Solubility in Nonionic Deep Eutectic Solvents Containing Phenolic Alcohols. *Front. Chem.* **2022**, 300. .

(17) Pelaquim, F. P.; Barbosa Neto, A. M.; Dalmolin, I. A. L.; Costa, M. C. A. O. D. Gas solubility using deep eutectic solvents: review and analysis. *Ind. Eng. Chem. Res.* **2021**, *60* (24), 8607−8620.

(18) Li, G.; Deng, D.; Chen, Y.; Shan, H.; Ai, N. Solubilities and thermodynamic properties of CO_2 in choline-chloride based deep eutectic solvents. *J. Chem. Thermodyn.* **2014**, *75*, 58−62.

(19) Wang, J.; Cheng, H.; Song, Z.; Chen, L.; Deng, L.; Qi, Z. Carbon dioxide solubility in phosphonium-based deep eutectic solvents: an experimental and molecular dynamics study. *Ind. Eng. Chem. Res.* **2019**, *58* (37), 17514−17523.

(20) García-Argüelles, S.; Ferrer, M. L.; Iglesias, M.; Del Monte, F.; Gutiérrez, M. C. Study of superbase-based deep eutectic solvents as the catalyst in the chemical fixation of CO_2 into cyclic carbonates under mild conditions. *Materials* **2017**, *10* (7), 759.

(21) Jiang, B.; Ma, J.; Yang, N.; Huang, Z.; Zhang, N.; Tantai, X.; Sun, Y.; Zhang, L. Superbase/acylamido-based deep eutectic solvents for multiple-site efficient CO_2 absorption. *Energy Fuel* **2019**, *33* (8), 7569−7577.

(22) Yan, H.; Zhao, L.; Bai, Y.; Li, F.; Dong, H.; Wang, H.; Zhang, X.; Zeng, S. Superbase ionic liquid-based deep eutectic solvents for improving CO_2 absorption. *ACS Sustainable Chem. Eng.* **2020**, *8* (6), 2523−2530.

(23) Wang, Z.; Wang, Z.; Huang, X.; Yang, D.; Wu, C.; Chen, J. Deep eutectic solvents composed of bio-phenol-derived superbase ionic liquids and ethylene glycol for CO_2 capture. *Chem. Commun.* **2022**, *58* (13), 2160−2163.

(24) Lemaoui, T.; Boublia, A.; Lemaoui, S.; Darwish, A. S.; Ernst, B.; Alam, M.; Benguerba, Y.; Banat, F.; AlNashef, I. M. Predicting the CO_2 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.

(25) Wang, J.; Song, Z.; Chen, L.; Xu, T.; Deng, L.; Qi, Z. Prediction of CO_2 solubility in deep eutectic solvents using random forest model based on COSMO-RS-derived descriptors. *Green Chem. Eng.* **2021**, *2* (4), 431−440.

(26) Mohan, M.; Demerdash, O.; Simmons, B. A.; Smith, J. C.; Kidder, M. K. K.; Singh, S. Accurate prediction of carbon dioxide capture by deep eutectic solvents using quantum chemistry and a neural network. *Green Chem.* **2023**, *25*, 3475−3492.

(27) Liu, X.; Gao, B.; Jiang, Y.; Ai, N.; Deng, D. Solubilities and thermodynamic properties of carbon dioxide in guaiacol-based deep eutectic solvents. *J. Chem. Eng. Data* **2017**, *62* (4), 1448−1455.

(28) Liu, Y.; Yu, H.; Sun, Y.; Zeng, S.; Zhang, X.; Nie, Y.; Zhang, S.; Ji, X. Screening deep eutectic solvents for CO_2 capture with COSMO-RS. *Front. Chem.* **2020**, *8*, 82.

(29) Liu, Y.; Dai, Z.; Zhang, Z.; Zeng, S.; Li, F.; Zhang, X.; Nie, Y.; Zhang, L.; Zhang, S.; Ji, X. Ionic liquids/deep eutectic solvents for CO_2 capture: Reviewing and evaluating. *Green Energy Environ.* **2021**, *6* (3), 314−328.

(30) Klamt, A. The COSMO and COSMO-RS solvation models. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1* (5), 699−709.

(31) Mohan, M.; Keasling, J. D.; Simmons, B. A.; Singh, S. In silico COSMO-RS predictive screening of ionic liquids for the dissolution of plastic. *Green Chem.* **2022**, *24* (10), 4140−4152.

(32) Mohan, M.; Smith, M. D.; Demerdash, O. N.; Kidder, M. K.; Smith, J. C. Predictive Understanding of the Surface Tension and Velocity of Sound in Ionic Liquids using Machine Learning. *J. Chem. Phys.* **2023**, *158*, 21.

(33) Mohan, M.; Smith, M. D.; Demerdash, O. N.; Simmons, B. A.; Singh, S.; Kidder, M. K.; Smith, J. C. Quantum Chemistry-Driven Machine Learning Approach for the Prediction of the Surface Tension

and Speed of Sound in Ionic Liquids. *ACS Sustainable Chem. Eng.* **2023**, *11* (20), 7809−7821.

(34) Cheng, J.; Wu, C.; Gao, W.; Li, H.; Ma, Y.; Liu, S.; Yang, D. CO2 Absorption Mechanism by the Deep Eutectic Solvents Formed by Monoethanolamine-Based Protic Ionic Liquid and Ethylene Glycol. *Int. J. Mol. Sci.* **2022**, *23* (3), 1893.

(35) Sang, H.; Su, L.; Han, W.; Si, F.; Yue, W.; Zhou, X.; Peng, Z.; Fu, H. Basicity-controlled DBN-based deep eutectic solvents for efficient carbon dioxide capture. *J. CO2 Util.* **2022**, *65*, 102201.

(36) Wang, Z.; Wu, C.; Wang, Z.; Zhang, S.; Yang, D. CO2 capture by 1, 2, 3-triazole-based deep eutectic solvents: the unexpected role of hydrogen bonds. *Chem. Commun.* **2022**, *58*, 7376−7379.

(37) Torrecilla, J. S.; Palomar, J.; Lemus, J.; Rodríguez, F. A quantum-chemical-based guide to analyze/quantify the cytotoxicity of ionic liquids. *Green Chem.* **2010**, *12* (1), 123−134.

(38) Abranches, D. O.; Zhang, Y.; Maginn, E. J.; Colón, Y. J. Sigma profiles in deep learning: towards a universal molecular descriptor. *Chem. Commun.* **2022**, *58* (37), 5630−5633.

(39) *JMP® Pro*, 17.2.0.; SAS Institute Inc: Cary, NC, 1989−2023. https://www.jmp.com/en_us/home.html (accessed 12/March/2023).

(40) Vapnik, V. N. *The nature of statistical learning theory*; Springer science & business media, 1999.

(41) Suykens, J. A.; Vandewalle, J. Least squares support vector machine classifiers. *Neural Process. Lett.* **1999**, *9* (3), 293−300.

(42) Breiman, L. Random forests. *Mach. Learn.* **2001**, *45* (1), 5−32.

(43) Hastie, T.; Tibshirani, R.; Friedman, J.. The elements of statistical learning: Data mining, inference, and prediction. In *Springer series in statistics*; Springer: New York, NY, 2009.

(44) Freund, Y.; Schapire, R.; Abe, N. A short introduction to boosting. *J. Jpn. Soc. Artif. Intell.* **1999**, *14* (771−780), 1612.

(45) Yu, L.-Y.; Ren, G.-P.; Hou, X.-J.; Wu, K.-J.; He, Y. Transition State Theory-Inspired Neural Network for Estimating the Viscosity of Deep Eutectic Solvents. *ACS Cent. Sci.* **2022**, *8* (7), 983−995.

(46) Kartal, F.; Özveren, U. An improved machine learning approach to estimate hemicellulose, cellulose, and lignin in biomass. *Carbohydr. Polym. Technol. Appl.* **2021**, *2*, 100148.

(47) Kardani, N.; Hedayati Marzbali, M.; Shah, K.; Zhou, A. Machine learning prediction of the conversion of lignocellulosic biomass during hydrothermal carbonization. *Biofuels* **2022**, *13* (6), 703−715.

(48) Gao, W.; Zhou, L.; Liu, S.; Guan, Y.; Gao, H.; Hui, B. Machine learning prediction of lignin content in poplar with Raman spectroscopy. *Bioresour. Technol.* **2022**, *348*, 126812.

(49) Gao, W.; Zhou, L.; Liu, S.; Guan, Y.; Gao, H.; Hu, J. Machine learning algorithms for rapid estimation of holocellulose content of poplar clones based on Raman spectroscopy. *Carbohydr. Polym.* **2022**, *292*, 119635.

(50) Zhang, W.; Wang, Y.; Ren, S.; Hou, Y.; Wu, W. Novel Strategy of Machine Learning for Predicting Henry's Law Constants of CO2 in Ionic Liquids. *ACS Sustainable Chem. Eng.* **2023**, *11* (15), 6090−6099.

(51) ChemAxon. Chemicalize. https://chemicalize.com/app/calculation. 2020. (accessed 26 November 2023).

(52) Swain, M. Chemicalize. org. *J. Chem. Inf. Model.* **2012**, *52* (2), 613−615.

(53) Mohan, M.; Demerdash, K. D.; Smith, M. D.; Demerdash, O. N.; Kidder, M. K.; Smith, J. C. Accurate Machine Learning for Predicting the Viscosities of Deep Eutectic Solvents. *J. Chem. Theory Comput.* **2024**.

(54) Ren, H.; Lian, S.; Wang, X.; Zhang, Y.; Duan, E. Exploiting the hydrophilic role of natural deep eutectic solvents for greening CO2 capture. *J. Cleaner Prod.* **2018**, *193*, 802−810.

(55) Rainbolt, J. E.; Koech, P. K.; Yonker, C. R.; Zheng, F.; Main, D.; Weaver, M. L.; Linehan, J. C.; Heldebrant, D. J. Anhydrous tertiary alkanolamines as hybrid chemical and physical CO 2 capture reagents with pressure-swing regeneration. *Energy Environ. Sci.* **2011**, *4* (2), 480−484.

(56) Zuo, Y.; Chen, X.; Wei, N.; Tong, J. Effect of water or ethanol on the excess properties of deep eutectic solvents (Tetrabutylammo-nium bromide+ Formic acid/Propionic acid). *J. Mol. Liq.* **2023**, *122034. .

(57) Singh, K.; Shibu, R. P.; Mehra, S.; Kumar, A. Insights into the physicochemical properties of newly synthesized benzyl triethylam-monium chloride-based deep eutectic solvents. *J. Mol. Liq.* **2023**, *386*, 122589.

(58) Pelaquim, F. P.; Neto, A. M. B.; Dalmolin, I. A. L.; Costa, M. C. D. Gas solubility using deep eutectic solvents: review and analysis. *Ind. Eng. Chem. Res.* **2021**, *60* (24), 8607−8620.

(59) Salim, N.; Holliday, J.; Willett, P. Combination of fingerprint-based similarity coefficients using data fusion. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (2), 435−442.

(60) Willett, P. Similarity-based virtual screening using 2D fingerprints. *Drug Discovery Today* **2006**, *11* (23−24), 1046−1053.

(61) Butina, D. Unsupervised data base clustering based on Daylight's fingerprint and Tanimoto similarity: A fast and automated way to cluster small and large data sets. *J. Chem. Inf. Comput. Sci.* **1999**, *39* (4), 747−750.

(62) *Daylight Chemical Information Systems, Daylight*. https://www.daylight.com/about/index.html

(63) Landrum, G. RDKit: A software suite for cheminformatics, computational chemistry, and predictive modeling. *Greg Landrum* **2013**, *8*, 31.

(64) Huang, Z.; Jiang, B.; Yang, H.; Wang, B.; Zhang, N.; Dou, H.; Wei, G.; Sun, Y.; Zhang, L. Investigation of glycerol-derived binary and ternary systems in CO2 capture process. *Fuel* **2017**, *210*, 836−843.