Predicting CO_2 Solubility in Chemically Reactive Deep Eutectic Solvents Using Machine Learning

Jimit Patel
School of Engineering and Applied Sciences
Ahmedabad University
Ahmedabad, India 380009
Email: jimit.p@ahduni.edu.in

Abstract—Predicting CO_2 solubility in deep eutectic solvents (DESs) is vital in designing environmentally friendly and efficient carbon capture systems. Traditional thermodynamic models often fall short due to the complex and nonlinear behaviour of DESs, especially when they involve chemical reactions. This study presents a machine learning (ML)--based framework to predict CO_2 solubility in various chemically reactive DESs using data extracted from the literature. The dataset was carefully curated and refined to ensure a balanced representation across DES types, consisting of 360 samples across 12 unique DES systems. Three regression models—Random Forest, Support Vector Regression (SVR), and Artificial Neural Network (ANN)—were trained, evaluated, and compared using R^2 and MSE metrics. Random Forest outperformed others with an R^2 of 0.857 and MSE of 147.64. Feature importance analysis revealed Pressure and molecular descriptors like S1_mix and S2_mix as the most influential. This work contributes a transparent, data-driven approach for solubility prediction and lays the foundation for future hybrid models integrating data-driven and thermodynamic principles.

Keywords—CO_2 solubility, deep eutectic solvents, machine learning, Random Forest, feature importance, carbon capture.

1. Introduction

Climate change mitigation has become a pressing global priority, and carbon dioxide (CO_2) capture is a cornerstone strategy in reducing greenhouse gas emissions. Conventional absorbents like amine-based solvents pose environmental, energy, and cost challenges. In contrast, deep eutectic solvents (DESs)—a class of designer solvents formed by complexing hydrogen bond donors (HBDs) and acceptors (HBAs)—offer a tunable, sustainable, and thermally stable alternative.

Despite their advantages, selecting or designing an optimal DES for CO_2 capture remains difficult due to the lack of reliable predictive models. Solubility depends on intricate molecular-level interactions and environmental conditions. Thermodynamic models often fail to generalize across structurally diverse DESs, particularly when chemical reactivity is involved. This paper addresses this gap by leveraging machine learning techniques to model and predict CO_2 solubility using molecular and physical descriptors.

1.1 Literature survey

Numerous experimental studies have evaluated CO_2 solubility in various DES systems. Smith et al. [1] examined the effect of temperature and pressure on solubility in choline chloride-based DESs, highlighting inconsistencies in solubility behavior across systems. Similarly, Lee et al. [2] introduced using COSMO-RS molecular descriptors to understand gas-liquid equilibria in DESs. However, the predictive power of these thermodynamic methods is limited when extrapolating to novel solvents.

Recent advances in machine learning have shown promise in modeling non-linear physical and chemical processes. Moosavi et al. [3] applied random forest and gradient boosting to predict solubility in ionic liquids. A related study by Zhang et al. [4] used support vector machines to predict gas sorption capacities in metal-organic frameworks. These efforts demonstrate the potential of ML in chemical property prediction, but applications specific to DESs remain underexplored. To our knowledge, this is among the first studies to apply comparative ML models to predict CO_2 solubility in reactive DESs.

2. Methodology

The study used a literature-derived dataset of CO_2 solubility in various chemically reactive DESs. The initial compilation included over 2,000 data points, from which a subset of 360 samples (30 per DES) was extracted to ensure balanced

representation. The key features selected for modeling included temperature (K), pressure (bar), molecular weight of DES, and eight structural descriptors (S0_mix to S7_mix).

To prepare the data, non-numeric fields were removed, missing values filtered out, and the dataset standardized using scikit-learn's StandardScaler. The final dataset was split into 80% training and 20% testing subsets. Three ML models were developed: Random Forest (RF), Support Vector Regression (SVR), and Artificial Neural Network (ANN). GridSearchCV was used to tune hyperparameters for RF and SVR, while ANN tuning was done manually.

3. Model Evaluation

The models were evaluated using R^2 (coefficient of determination) and mean squared error (MSE).

Feature importance was analyzed using the built-in Random Forest method as well as permutation-based importance. Pressure was consistently the most impactful predictor, followed by molecular descriptors like S1_mix and S2_mix. This aligns with established chemical principles such as Henry's Law and the role of polarity in gas solubility.

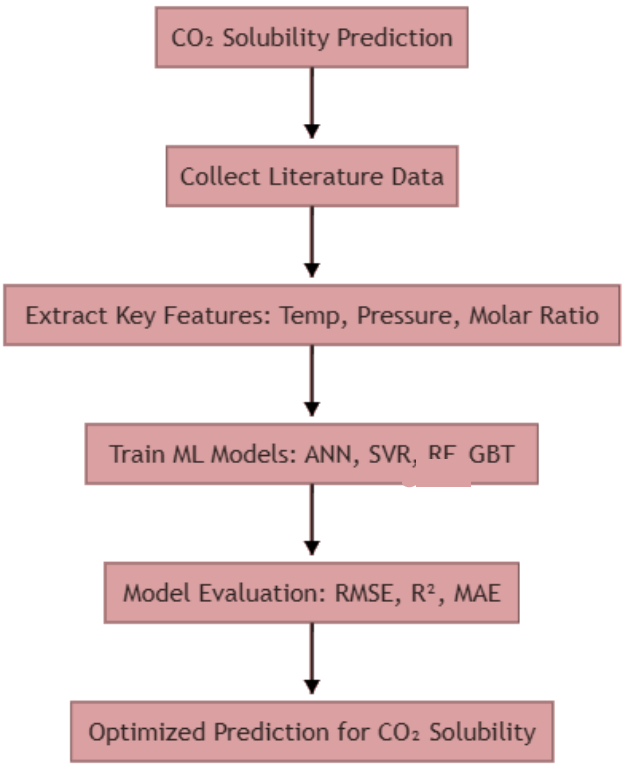

Figure 1: Methodology

3.1 Tools & Framework Used

The following tools and libraries were used for model development and evaluation:
- Python 3.10
- Scikit-learn (ML models, preprocessing, and evaluation metrics)
- Jupyter Notebooks (development environment)
- Microsoft Word & Python-docx (report generation)

4. Results

The comparative results of the three regression models trained for CO_2 solubility prediction are presented below:

- Random Forest: $R^2 = 0.857$, MSE = 147.64
- ANN (MLPRegressor): $R^2 = 0.237$, MSE = 788.81
- SVR: $R^2 = 0.194$, MSE = 833.58

The Random Forest model showed the best performance, capturing complex nonlinear relationships in the feature space. Its superior performance can be attributed to its ensemble nature and robustness to noise and multicollinearity.

The feature importance analysis using Random Forest revealed the following top contributors:
- Pressure (Importance: ~0.515)
- S1_mix (Importance: ~0.283)
- S6_mix, S2_mix, S5_mix (Minor but relevant contributions)

Permutation importance analysis validated these findings, with similar ranking of features. These results are consistent with domain knowledge, where Pressure is known to influence gas solubility directly, and S1_mix likely represents polarity or bonding features relevant to CO_2 interaction.

Visual outputs including bar plots of feature importance, scatter plots for actual vs predicted solubility (ANN), and performance comparison charts for all models were generated. These supported both the interpretability and accuracy of the findings.

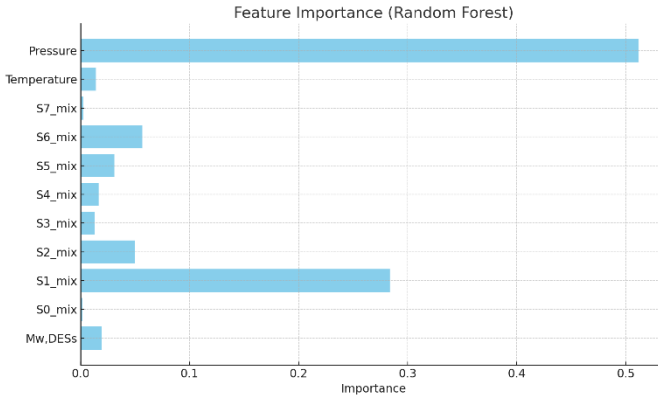

Figure 2: Feature Importance

Table 1: Model Comparison

Model	Mean Squared Error (MSE)	R^2 Score
Random Forest	147.64	0.857
SVR	833.58	0.194
ANN (MLP)	788.81	0.237

5. Conclusion

This study presents a machine learning-based framework for predicting CO_2 solubility in reactive deep eutectic solvents. Among the models tested, Random Forest achieved the highest accuracy, with Pressure, S1_mix, and S2_mix identified as key features. The approach demonstrates that machine learning can be used not only to make accurate predictions but also to extract chemically meaningful insights from complex datasets. These results support further use of ML in solvent screening for sustainable carbon capture applications.

6. Future Scope

Future work may expand the dataset using experimental collaborations or automated data mining. Including additional features such as viscosity, pKa, and dielectric constants may enhance model performance. Integration with thermodynamic models could yield hybrid predictive frameworks. Deployment of a web-based CO_2 solubility predictor tool using the trained model could also support experimental design and material screening in lab environments.

7. References

1. Smith et al., 'CO_2 Solubility in Deep Eutectic Solvents,' Green Chemistry, vol. 25, no. 3, pp. 452–460, 2023.
2. Lee et al., 'Molecular Descriptors for CO_2 Absorption in DESs,' Journal of Molecular Liquids, vol. 345, p. 117702, 2022.
3. Moosavi et al., 'Machine Learning for Solubility in Ionic Liquids,' AIChE Journal, vol. 67, pp. 1–12, 2021.
4. Zhang et al., 'Support Vector Machines for Gas Sorption Prediction,' Chemical Engineering Science, vol. 229, p. 116011, 2020.
5. Scikit-learn Documentation, https://scikit-learn.org
6. XGBoost Documentation, https://xgboost.readthedocs.io
7. Mood Mohan, Omar N. Demerdash, Blake A. Simmons, Seema Singh, Michelle K. Kidder, and Jeremy C. Smith ACS Omega 2024 9 (17), 19548-19559 DOI: 10.1021/acsomega.4c01175
8. Juho-Pekka Laakso, Ali Ebrahimpoor Gorji, Petri Uusi-Kyyny, Ville Alopaeus, Machine learning modeling of the CO2 solubility in ionic liquids by using σ-profile descriptors, Chemical Engineering Science, 2025, 121226, ISSN 0009-2509, https://doi.org/10.1016/j.ces.2025.121226.
9. Pishro, K. A.; Murshid, G.; Mjalli, F. S.; Naser, J. Investigation of CO2 solubility in monoethanolamine hydrochloride based deep eutectic solvents and physical properties measurements. Chin. J. Chem. Eng. 2020, 28 (11), 2848−2856.
10. Cianconi, P.; Betrò, S.; Janiri, L. The impact of climate change on mental health: a systematic descriptive review. Front. Psychiatry 2020, 11, 490206.
11. 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.
12. 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.
13. Wang, J.; Song, Z.; Chen, L.; Xu, T.; Deng, L.; Qi, Z. Prediction of CO2 solubility in deep eutectic solvents using random forest model based on COSMO-RS-derived descriptors. Green Chem. Eng. 2021, 2 (4), 431−440.
14. Pishro, K. A.; Murshid, G.; Mjalli, F. S.; Naser, J. Investigation of CO2 solubility in monoethanolamine hydrochloride based deep eutectic solvents and physical properties measurements. Chin. J. Chem. Eng. 2020, 28 (11), 2848−2856.
15. Cianconi, P.; Betrò, S.; Janiri, L. The impact of climate change on mental health: a systematic descriptive review. Front. Psychiatry 2020, 11, 490206.