

Predicting Fluorinated Gas Solubility in Deep Eutectic Solvents Using Machine Learning and Deep Learning Approaches

Amir Hossein Sheikhshoei

Hakim Sabzevari University

Jan J. Weigand

TU Dresden

Ali Sanati

a.sanati@hsu.ac.ir

Hakim Sabzevari University

Article

Keywords: Machine learning, solubility prediction, deep eutectic solvents, fluorinated gases, SHAP analysis

Posted Date: June 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9985938/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Predicting Fluorinated Gas Solubility in Deep Eutectic Solvents Using Machine Learning and Deep Learning Approaches

Amir Hossein Sheikhshoei¹, Jan J. Weigand^{2,3*}, Ali Sanati^{1*}

¹ Petroleum and Petrochemical Engineering School, Hakim Sabzevari University, Sabzevar, Iran.

² Faculty of Chemistry and Food Chemistry, Chair of Inorganic Molecular Chemistry, Technische Universität Dresden, Dresden, Germany.

³ Department of Chemistry and Polymer Science, Stellenbosch University, Stellenbosch, South Africa.

*Corresponding authors: jan.weigand@tu-dresden.de, a.sanati@hsu.ac.ir

Abstract

This paper introduces a data-driven framework by deploying machine learning (ML) and deep learning (DL) architectures to estimate fluorinated gas (F-gas) absorption capacities in deep eutectic solvents (DESs). The proposed approach aims to support the screening and design of DESs for gas capture applications by correlating thermodynamic and molecular descriptors with experimentally measured solubility data. Four predictive models, namely Gradient Boosting (GBoost) and Random Forest (RF) as ML techniques, and TabNet and GrowNet as DL architectures, were developed and compared using a dataset of 440 experimental data points, incorporating temperature, pressure, and critical molecular descriptors of both F-gases and DESs as input parameters. Further evaluation was conducted using SHapley Additive exPlanations

(SHAP) analysis and the leverage approach. DL models, especially GrowNet, demonstrated high accuracy and strong generalization on test data. GrowNet outperformed all others, achieving the lowest test errors ($MSE = 0.0046\%$, $R^2 = 0.9985$). SHAP analysis revealed pressure and temperature as the most important variables affecting the model predictions, with critical temperature and acentric factor of F-gases also playing significant secondary roles. The leverage method confirmed model reliability, with over 98% of data points lying within its applicability domain. The results demonstrate that ML- and DL-based methods can provide effective predictive tools for estimating F-gas solubility in DES systems.

Keywords: Machine learning, solubility prediction, deep eutectic solvents, fluorinated gases, SHAP analysis.

1. Introduction

The escalating atmospheric levels of F-gases, extensively utilized in sectors such as electronics and cooling systems, pose a critical threat to global environmental stability. Possessing a global warming potential (GWP) that eclipses that of carbon dioxide (CO_2) by several orders of magnitude, these synthetic greenhouse gases are highly detrimental [1]. Driven by stringent global treaties, notably the Kigali Amendment to the Montreal Protocol [2], discovering effective separation and capture mechanisms has become a pressing priority to facilitate their systematic phase-down [3].

Absorption-based separation using advanced solvents presents a promising pathway. In this context, DESs have emerged as promising, sustainable absorbents due to their low volatility, tunable properties, and potential for high affinity towards F-gases. However, the experimental screening of countless possible DES formulations is prohibitively time-consuming and costly [4].

ML/DL techniques can facilitate the prediction of DES performance and enable the rapid screening of a vast number of candidate formulations, as demonstrated in previous studies [5, 6].

Castro et al. (2020) [4] introduced an empirical approach to F-gas absorption utilizing DESs. Their methodology involved formulating DESs from perfluorinated acids and fluorinated ionic liquids (FILs) to assess the solubility of three specific refrigerants (R-32, R-125, and R-134a) at 303.15, 313.15, and 323.15 K. The findings revealed that the absorption efficiency is heavily dependent on the FIL's cation, with imidazolium- and tetrabutylammonium-based formulations outperforming others. Additionally, the inherent preference of these DESs for R-134a over its counterparts illustrates their promising capability in separating complex F-gas systems, such as the R-32/R-125 azeotrope.

Demirbek et al. (2024) [7] created a computational model using the soft-SAFT (Statistical Associating Fluid Theory) equation of state for five different DESs to evaluate their F-gases solubility potential. Their study revealed that gas solubility is a strong function of hydrogen bond acceptor (HBA) to hydrogen bond donor (HBD) molar ratio, temperature, the molecular structure of the refrigerant, and the DES ionic composition. Furthermore, they calculated absorption enthalpies and ideal selectivity, identifying 1:2 as the optimal DESs' HBA to HBD ratio for separating R-134a from R-125 or R-32, though the separation of R-32 from R-125 was difficult. This work demonstrates soft-SAFT as a powerful screening tool for rapidly and efficiently screening novel solvents, guiding experimental efforts, and advancing the development of sustainable technologies.

Complementing the modeling approach of Demirbek et al., Zuo et al. (2025) [8] used a combined computational and experimental method to screen DESs for the selective separation of the F-gas blend R-410A, comprising R32 and R125 gases. They first utilized the COSMO-RS model to

screen 132 DESs, where three top-performing DESs, [Choline Chloride][Malonic Acid], [Tetramethylammonium Chloride][Malonic Acid], and [Tri-methylbenzylammonium Chloride][Malonic Acid] were selected for further experimental validation. The results of this study confirmed that all three DESs exhibited significantly higher solubility for R32 over R125, with [Tri-methylbenzylammonium Chloride][Malonic Acid] identified as the optimal absorbent. Crucially, their study provided profound molecular-level insights through quantum chemical calculations. They identified that multiple non-covalent interactions govern the absorption, with the C–H $\cdots$ Cl interaction between the HBA and the F-gas as the dominant mechanism. This work demonstrates the power of combining computational screening (COSMO-RS) with molecular-level analysis to rationally design and select DESs for targeted gas separations.

In a comprehensive computational approach, Wang et al. (2026) [9] advanced the UNIFAC framework to predict the absorption of F-gases in DESs. Their methodology screened 2160 mixtures derived from 15 HBAs and 48 HBDs at 1:1, 1:2, and 1:3 molar proportions. By fitting 64 new group interaction parameters to 525 empirical data points for R-32, R-125, and R-134a, the model achieved a high precision level ($ARD = 0.1823$, $R^2 = 0.99$), proving superior to COSMO-RS. This robust model facilitated the identification of three prime DESs [Choline Chloride][1,3 propanediol], [Choline Chloride][1,4 butanediol], and [Choline Chloride][glycol] for fractionating the R-32/R-125 azeotropic blend, balancing thermodynamic capabilities with viable melting points and viscosities. Subsequent mechanistic investigations using sigma-profiles and molecular dynamics revealed that the selective affinity for R-32 is predominantly governed by the HBD component.

This study, as far as the authors are aware, is the first to use interpretable ML and DL models to predict F-gas solubility in DESs. Models include random forest (RF), gradient boosting

(GBoost), GrowNet, and TabNet to predict F-gas solubility in DESs, with the aim of supporting the rational design of separation systems for fluorinated gases. To integrate critical properties with ML/DL techniques, a new framework was provided. A comparison of model performance revealed that the DL models, unlike traditional ML models, demonstrated superior predictive capabilities on the test data, enabling accurate predictions for DES systems not included in the training set. However, it should be noted that the investigated dataset remains limited in size and chemical diversity, and therefore the predictive capability demonstrated here reflects the investigated chemical space rather than universal transferability to unrelated systems.

Assessing the models across diverse operational scenarios confirmed that the GrowNet algorithm provides high predictive accuracy and interpretability for determining the solubility of F-gases in DESs. Furthermore, the application of SHapley Additive exPlanations (SHAP) to the GrowNet framework identified operating pressure and temperature as the most influential driving forces governing solubility. Meanwhile, the structural and molecular descriptors associated with both the DESs and the F-gases exerted a secondary, albeit crucial, impact on the final prediction outcomes. The key molecular descriptors include critical temperature of DES (T_c DES), critical pressure of DES (P_c DES), critical volume of DES (V_c DES), acentric factor of DES (ω DES), density of DES (ρ DES), critical temperature of F-gas (T_c F-gas), critical pressure of F-gas (P_c F-gas), and acentric factor of F-gas (ω F-gas).

It is important to emphasize that SHAP analysis explains the behavior of the trained ML/DL model and should not be interpreted as direct thermodynamic proof of causality. Rather, the SHAP values indicate how strongly the trained model associates individual descriptors with the predicted output within the investigated dataset.

Beyond merely identifying the critical descriptors that govern solubility, this strategy provides additional insight into the relative importance of macroscopic operating conditions and molecular descriptors within the trained models.. To conclude, the models' dependability and broad applicability were validated via outlier detection, supporting their potential utility for future screening applications.

2. Dataset Compilation and Feature Selection

Utilizing a repository of 440 experimental measurements, the solubility of F-gases in DESs was estimated through a specific set of input features: T , P , $T_{c,DES}$, $P_{c,DES}$, $V_{c,DES}$, ω_{DES} , ρ_{DES} , $T_{c,F-gas}$, $P_{c,F-gas}$, and ω_{F-gas} . Determinations of DESs' critical parameters were based on previously published values [10-12]. Table 1 provides the core statistical characteristics of the employed dataset.

Table 1. Descriptive statistics of the utilized dataset.

Statistical parameters	Input										Output
	T (K)	P (MPa)	Tc DES (K)	Pc DES (MPa)	Vc DES (mL/mol)	ω DES	ρ DES (g/cm ³)	Tc F-gas (K)	Pc F-gas (MPa)	ω F-gas	X(g/kg)
Minimum	303.15	0.03	696.77	1.46	700.72	0.63	1.27	339.16	3.61	0.27	0.0053
Maximum	323.15	1.13	866.92	1.95	1102.04	0.78	1.67	374.18	5.78	0.32	0.6223
Mean	313.30	0.41	773.93	1.71	882.14	0.69	1.48	353.61	4.52	0.30	0.26
Median	313.15	0.39	749.93	1.77	828.98	0.69	1.40	351.25	4.05	0.30	0.26
skewness	-0.02	0.44	0.30	-0.02	0.29	0.46	0.01	0.51	0.48	-0.04	0.12
Kurtosis	-1.49	-0.86	-1.09	-1.13	-1.12	-1.19	-1.85	-1.28	-1.63	-1.49	-1.15

The selected descriptors were chosen to represent both macroscopic operating conditions and molecular characteristics associated with the DESs and F-gases. Pressure and temperature

describe the thermodynamic conditions governing gas absorption, whereas the critical properties and acentric factors provide simplified descriptors related to molecular interactions, phase behavior, and non-ideality.

Although the employed descriptor set captures important thermodynamic trends, it should be noted that these parameters provide only an indirect representation of the underlying molecular interactions governing gas solubility. Consequently, incorporation of additional molecular descriptors, such as sigma-profiles, electrostatic descriptors, or quantum chemical parameters, may further improve predictive capability and physicochemical interpretability in future studies.

3. Models Development

3.1. Random Forest (RF)

Operating as a widely used ensemble-based supervised learning technique, the Random Forest (RF) algorithm aggregates the predictive power of numerous decision trees [13, 14]. To counteract the overfitting vulnerability inherent in standalone trees, RF employs a stochastic approach, feeding random subsets of samples and variables to each tree during the training phase. The ultimate model output is derived by consolidating the predictions of all trees, typically via mean aggregation for continuous targets or majority consensus for discrete classes, thereby improving generalization performance. This structural framework makes RF well suited for modeling intricate, non-linear relationships within datasets containing high dimensionality and heterogeneous feature types.

By fine-tuning critical hyperparameters, notably the maximum tree depth and the total ensemble size, practitioners can balance predictive accuracy against computational overhead. Furthermore, because the algorithm grows each tree in isolation, it inherently supports

parallelized computation, significantly accelerating the training workflow while preserving resistance against dataset noise and anomalous points. Consequently, RF has been widely applied in regression and classification tasks involving complex datasets [15-18].

3.2. Gradient Boosting (GBoost)

Operating as a sequential ensemble framework, the Gradient Boosting (GBoost) regressor iteratively builds decision trees specifically aimed at correcting the residuals of previous estimators. It relies on an additive forward-stepping procedure that gradually introduces weak learners to optimize the overall predictive capacity [19]. The model is inherently self-correcting, achieving incremental performance gains with the addition of each new base learner [20]. This stagewise evolution provides an effective strategy for navigating the bias-variance dilemma. GBoost commences with a basic initial prediction and systematically refines its output by appending new trees that directly address the predictive flaws of their predecessors [21]. Furthermore, it incorporates an adaptive weighting protocol, ensuring that instances miscalculated in earlier iterations are assigned higher significance during subsequent training phases [22].

Due to its ability to model nonlinear relationships, GBoost has been extensively employed in regression problems involving thermophysical and engineering datasets.

3.3. Tabular Neural Network (TabNet)

Developed specifically for structured tabular information, TabNet represents a deep learning architecture built upon an attention-driven paradigm [23]. A defining characteristic of this algorithm is its implementation of “masked attention,” a mechanism that dynamically prioritizes

the most critical input parameters throughout successive learning phases. By systematically suppressing irrelevant variables, this approach simultaneously improves predictive performance and supports algorithmic interpretability.

Structurally, TabNet operates through a series of sequential decision blocks, each coupling an attention module with a feature transformer to iteratively refine data representations. This continuous filtration of extraneous data contributes to efficient learning behavior and improved generalization performance. Consequently, the model’s adaptive feature routing and robustness against high-dimensional datasets make it a promising approach for intricate regression and classification challenges [22, 24-26].

3.4. Gradient Boosting Neural Networks (GrowNet)

GrowNet, developed by Sorkhan et al. (2020), represents a boosting-based neural network framework that deploys shallow neural networks in the role of base learners [27]. The model operates sequentially by adding neural network learners that iteratively reduce the residual errors generated by preceding stages. Compared to standard boosting algorithms, GrowNet has been reported to improve regression accuracy while reducing sensitivity toward hyperparameter selection [28]. The model is conceptually influenced by Gradient Boosting Decision Trees (GBDT), iteratively minimizing the loss function to improve predictive performance. A key feature is its use of Newton-Raphson-inspired optimization strategies, which leverages second-order gradient information to accelerate the convergence rate while keeping computational expenses manageable. Ultimately, GrowNet provides an efficient framework for nonlinear regression problems and has shown promising performance in recent ML applications [29, 30]. However, it should be noted that the predictive advantages of DL-based methods strongly

depend on dataset size, diversity, and descriptor quality. Since the present study is based on a relatively moderate-sized dataset, the observed performance improvements should primarily be interpreted within the investigated chemical space rather than as universally transferable behavior for unrelated systems.

3.5. Data Partitioning and Hyperparameter Optimization

Data allocation in this research involved a randomized 80:20 split, dedicating 80% of the samples to model training and the remaining 20% to performance testing. To ensure comparative validity, this specific data partition was uniformly applied to all evaluated algorithms.

The optimization of model hyperparameters was conducted using the Grid Search Cross-Validation (GridSearchCV) technique, structured with a 5-fold cross-validation setup ($cv = 5$). GridSearchCV is widely recognized as a systematic hyperparameter optimization strategy that evaluates combinations of predefined hyperparameters to identify the configuration associated with the best predictive performance [31]. Integration with a k-fold strategy, where data is segmented into k folds and validated iteratively k times, further improves the robustness of the evaluation by reducing dependence on a single train-test split.

Utilizing the scikit-learn library [32], this automated procedure scans a specified parameter grid, evaluates each combination through the estimator, and isolates the configuration that delivers the best predictive accuracy for the investigated dataset. A comprehensive breakdown of the optimal hyperparameters identified through this process is presented in Table 2.

Table 2. Optimized hyperparameters for the GBoost, RF, GrowNet, and TabNet models.

Model	Hyperparameter	Best Value
TabNet	n_d	16
	n_a	8

	n_steps	5
	gamma	1.0
	lambda_sparse	0.001
	optimizer_params (learning rate)	0.1
GrowNet	hidden_dim	64
	n_blocks	3
	lr	0.001
	batch_size	32
GBoost	n_estimators	300
	max_depth	3
	subsample	0.5
	learning_rate	0.05
RF	max_features	8
	max_depth	10
	min_samples_split	2
	min_samples_leaf	1
	n_estimators	300

218

219 4. Results and Discussion

220 We utilized graphical representations alongside statistical error metrics to gauge the models’
 221 proficiency in estimating the solubility of F-gases in DESs. This dual approach provides an in-
 222 depth performance review, allowing us to identify the model exhibiting the highest precision and
 223 the lowest error margin. The outcomes of these assessments are elaborated below.

224 4.1. Statistical Metrics for Accuracy Validation

225 The precision of the developed models was assessed by contrasting the estimated solubility of F-
 226 gases in DESs (y_{pred}) against the actual experimental data (y_{exp}). To ensure a rigorous
 227 assessment of model efficacy, a variety of statistical error metrics were calculated, which are
 228 outlined in the subsequent sections:

229 Mean Squared Error (MSE)

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

Mean Absolute Error (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_{i,pred} - y_{i,exp}| \quad (2)$$

Standard Deviation (SD)

$$SD = \sqrt{\frac{\sum_{i=1}^n (\frac{y_{i,exp} - y_{i,pred}}{y_{i,exp}})^2}{n - 1}} \quad (3)$$

Coefficient of Determination (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{i,exp} - y_{i,pred})^2}{\sum_{i=1}^n (y_{i,exp} - \bar{y}_{i,exp})^2} \quad (4)$$

230 Table 3 presents the statistical evaluation metrics for the TabNet, GrowNet, RF, and GBoost
231 models. The findings reveal that GrowNet provides the highest predictive accuracy among the
232 evaluated models for the present dataset. Exhibiting the lowest $MSE\%$ (0.0046), $MAE\%$ (0.4900),
233 and SD (0.1343) alongside an outstanding R^2 of 0.9985, GrowNet displays strong generalization
234 capability coupled with robust training efficacy. This superiority may be associated with
235 GrowNet’s architecture, which combines deep learning components with gradient boosting and
236 can capture intricate nonlinear correlations compared to conventional ML approaches.

237 Furthermore, DL algorithms like GrowNet and TabNet exhibit high predictive performance on
238 both the training and test datasets. The minimal discrepancy between training and testing
239 deviations suggest limited overfitting within the investigated dataset, supporting reliable
240 forecasts for unseen data from the same chemical and descriptor space. Specifically, TabNet
241 records a test $MSE\%$ of 0.0285 and an R^2 of 0.9909, showing only a marginal deviation from its

training metrics and confirming its high precision. This underscores a primary benefit of DL frameworks: their capacity for robust generalization while mapping complex feature dynamics.

Conversely, the RF model yields a moderate outcome, registering a test $MSE\%$ of 0.0807, $MAE\%$ of 2.0853, SD of 0.2553, and an R^2 of 0.9742. Despite RF's solid learning on the training subset, its elevated testing errors relative to GrowNet and TabNet suggest a diminished ability to model complicated nonlinear trends, leading to slightly weaker generalization.

Lastly, GBoost delivers a competitive output with a test $MSE\%$ of 0.0379, $MAE\%$ of 1.3781, SD of 0.1918, and an R^2 of 0.9878. Although sequential boosting effectively minimizes bias, GBoost's accuracy falls marginally short of GrowNet and parallels TabNet, reinforcing the advantages of the DL-based architecture for the investigated datasets.

However, these results should not be overinterpreted as universal evidence for the general superiority of DL models. The dataset contains 440 data points and covers a defined range of F-gases, DESs, temperatures, and pressures. Therefore, the observed model ranking primarily reflects performance within the investigated data domain. Independent external validation with additional DES/F-gas systems would further strengthen the generality of the conclusions.

Table 3. Evaluation of model accuracy via statistical criteria.

Models		Statistical parameters			
		MSE (%)	MAE (%)	SD	R^2
TabNet	Test	0.0285	1.2680	0.3524	0.9909
	Train	0.0141	0.9055	0.2552	0.9948
	Total	0.0169	0.9780	0.2769	0.9940
GrowNet	Test	0.0046	0.4900	0.1343	0.9985
	Train	0.0017	0.2434	0.0854	0.9993
	Total	0.0022	0.2927	0.0970	0.9991

RF	Test	0.0867	2.1890	0.2553	0.9723 258
	Train	0.0107	0.7797	0.1090	0.9961
	Total	0.0259	1.0615	0.1497	0.9908
GBoost	Test	0.0379	1.3781	0.1918	0.9878
	Train	0.0053	0.5847	0.1068	0.9980
	Total	0.0118	0.7434	0.1281	0.9958

259

260 As illustrated in Figure 1, the average absolute relative deviation (AARD) values for the ML and

261 DL models are evaluated against those of the UNIFAC-DES and COSMO-RS frameworks. The

262 findings clearly indicate that the data-driven approaches deliver notably superior precision for

263 the evaluated dataset. Furthermore, when comparing the performance of TabNet, GrowNet, RF,

264 and GBoost, the GrowNet architecture exhibits the minimum predictive error, establishing itself

265 as the most accurate model overall within this comparison.

266 Direct comparison with thermodynamic models should be interpreted carefully, since UNIFAC-

267 DES, COSMO-RS, and ML/DL approaches differ in their underlying assumptions, input

268 information, and validation strategies. While the present data-driven models provide lower

269 deviations for the available dataset, thermodynamic models may offer broader extrapolation

270 capability when applied outside the trained chemical space.

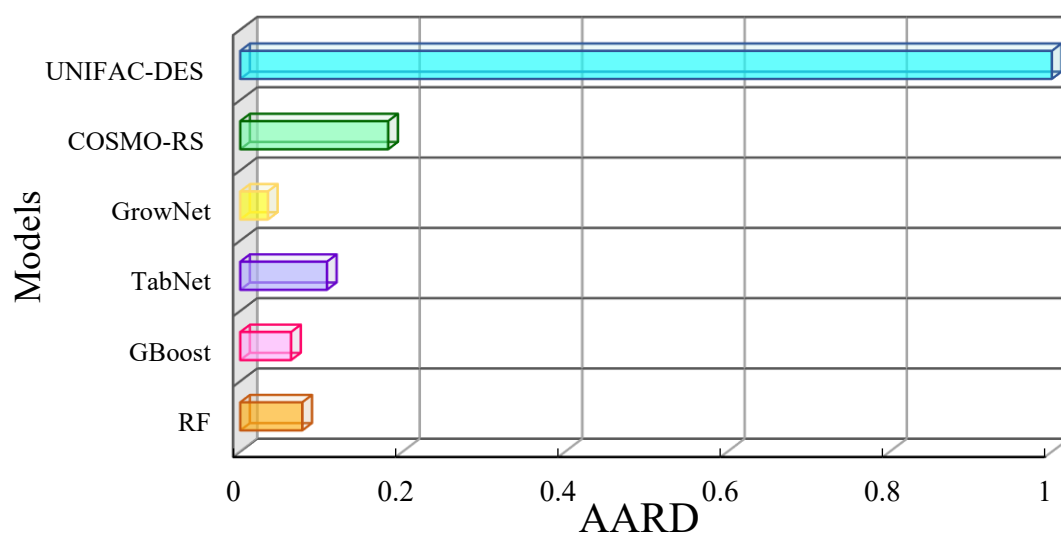


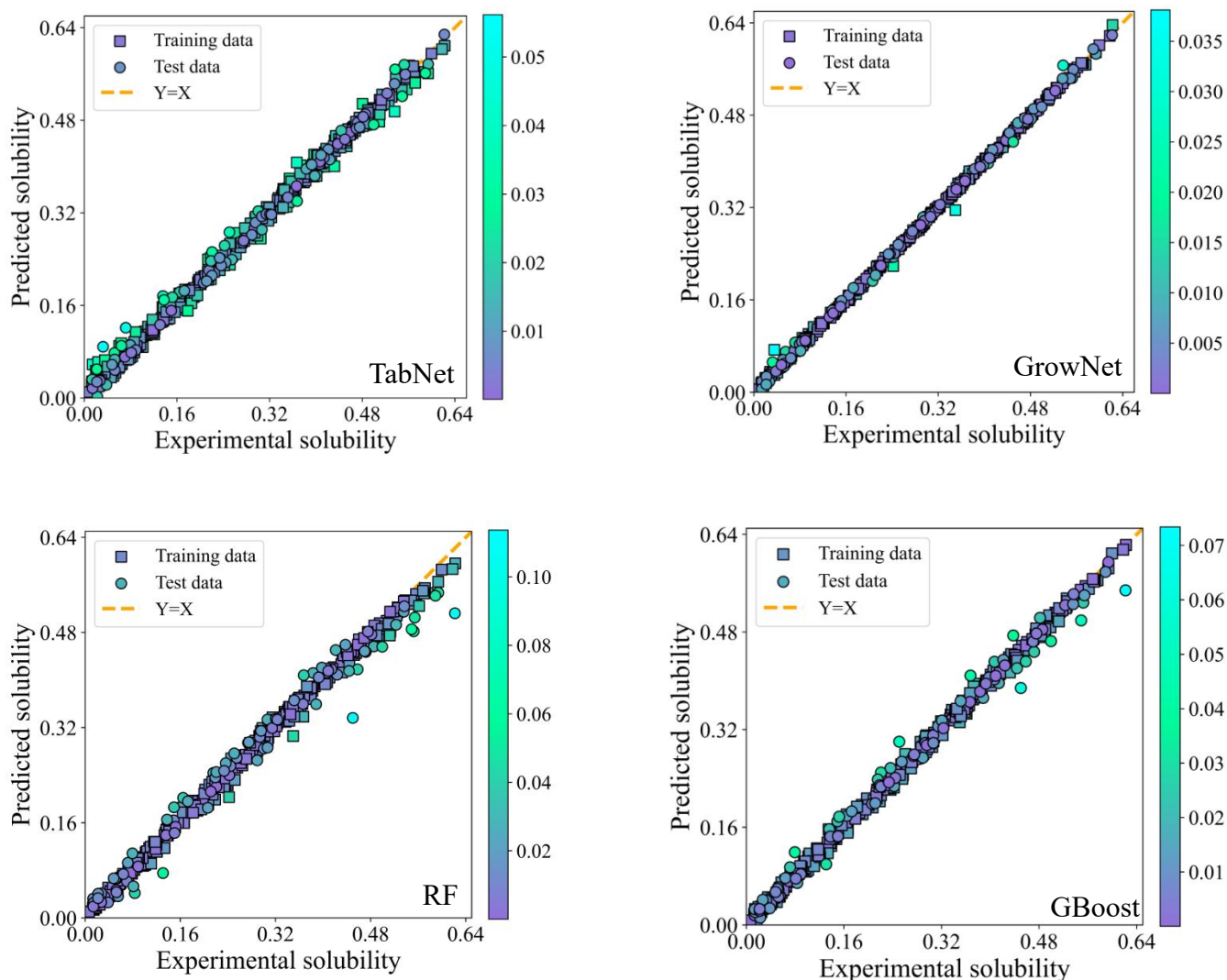
Figure 1. Comparison of the models studied in this paper with UNIFAC-DES and COSMO-RS models.

4.2. Visual Representation of Model Discrepancies

In this study, graphical error analysis was utilized to comprehensively assess and compare the reliability and precision of the predictive models. By illustrating the distribution, trends, and variations of predictive discrepancies, these visual instruments offer additional insight into model performance. Consequently, they serve as a useful complement to the quantitative evaluation metrics, further validating the superiority of the developed frameworks. A cross-plot serves as an effective graphical tool in this context, plotting the predicted data (Pred) versus the experimental measurements (Exp). Within this diagram, the $Y = X$ line denotes the theoretical ideal, where predictions are identical to the actual observations. The proximity and concentration of data points relative to this bisector line reflect the accuracy of the models.

As depicted in Figure 2, the GrowNet algorithm exhibits the tightest clustering of points along the $Y = X$ reference line, with minimal deviations observed in the test cases. This tight grouping

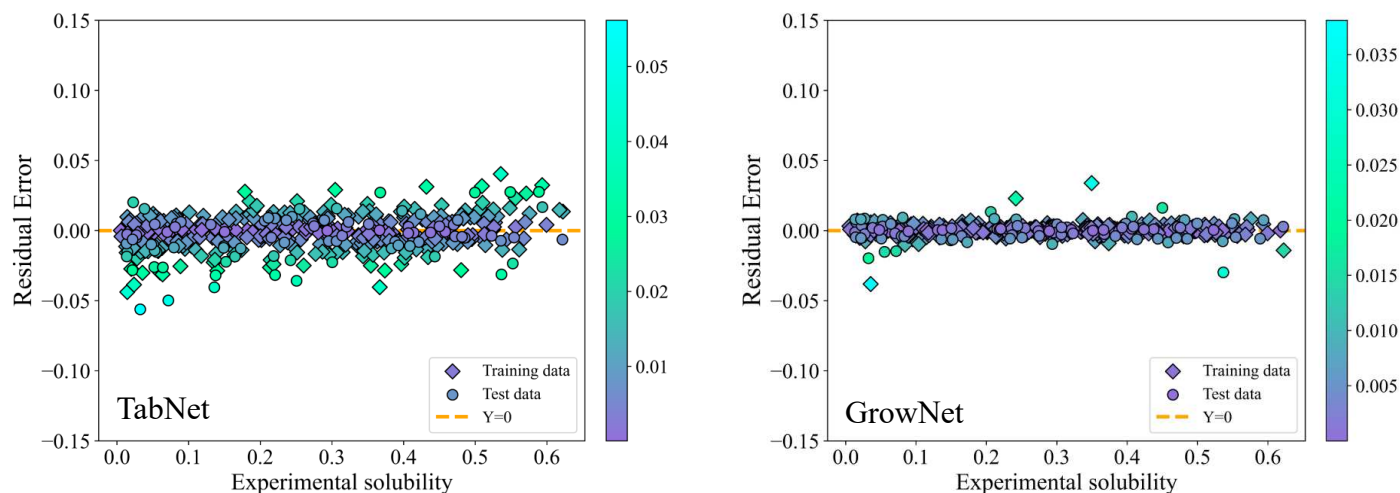
286 demonstrates its high predictive accuracy and good generalization potential for unseen data
 287 within the investigated dataset. Subsequently, the TabNet model ranks second, displaying good
 288 predictive alignment and moderate point density, albeit with a marginally wider scatter compared
 289 to GrowNet.



290 **Figure 2.** Cross-plots comparing experimental solubility values with model predictions.
 291 Residual error plots serve as critical diagnostic tools for assessing the variance between
 292 empirical measurements and model predictions. Instances where data points tightly converge
 293 along the $Y = 0$ axis signify negligible deviations, thereby validating the framework's predictive

consistency. In these visual representations, the ordinate (Y-axis) denotes the relative residual error percentage, whereas the abscissa (X-axis) plots the observed experimental values. Such graphical evaluations are instrumental in mapping error distributions across the entire spectrum of outputs and pinpointing any inherent predictive biases or systematic anomalies.

Figure 3 delineates the residual distributions for the various algorithms forecasting F-gas solubility within DESs. The graphical evidence clearly indicates that the GrowNet architecture achieves the densest clustering adjacent to the $Y = 0$ reference, supporting its high predictive precision. The relatively small number of outliers in the test subset suggest reasonable generalization capabilities for unseen DES configurations. Conversely, alternative algorithms exhibit broader error margins; TabNet and GBoost present moderate variances, while the RF model demonstrates the maximum scatter, aligning with the established metices summarized in Table 3.



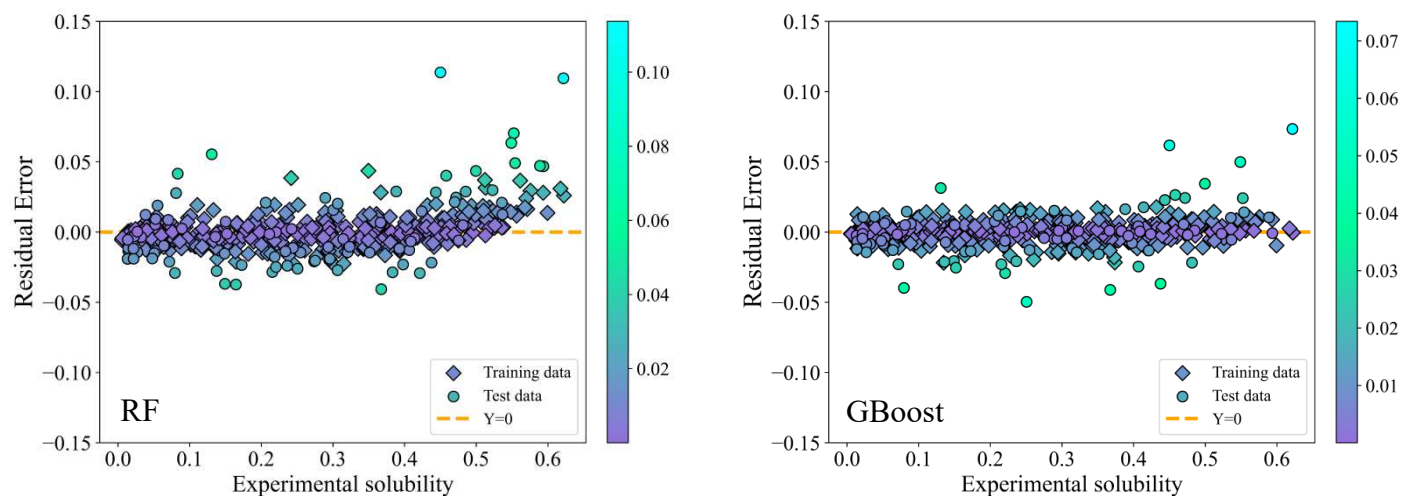


Figure 3. Scatter plots of residuals to evaluate the prediction error of each model.

306

307 Figure 4 illustrates the residual error distributions, offering further insights into the predictive

308 efficacy of the four evaluated frameworks. Among the tested algorithms, GrowNet demonstrates

309 the most concentrated residual profile, signifying the lowest overall prediction error among the

310 investigated models. The majority of its residual metrics are confined within a tight interval of

311 approximately -0.2 to 0.2 , with a pronounced density around the zero mark. This clustering

312 denotes minimal variance from empirical measurements and supports the stability of GrowNet

313 across both the training and testing subsets.

314 The TabNet framework ranks second in accuracy, with its residuals predominantly localized

315 between -0.05 and 0.05 . Conversely, the RF model displays the most extensive error scatter.

316 This expanded dispersion underscores its comparatively weaker predictive capability and

317 reduced stability in modeling the underlying nonlinear trends of the investigated dataset.

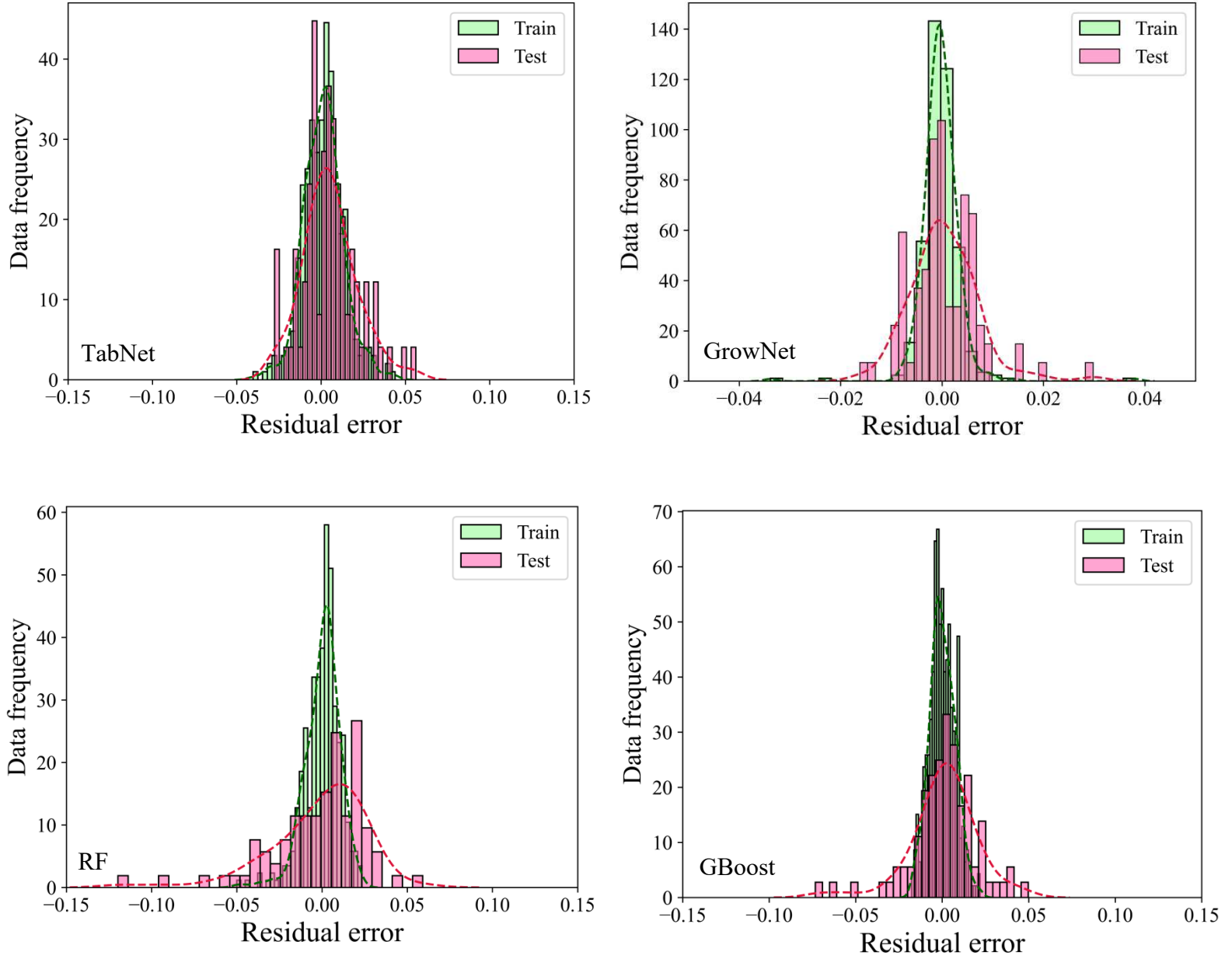


Figure 4. Histograms displaying the frequency distribution of prediction errors.

318 Figure 5 illustrates the cumulative frequency distribution as a function of the percentage residual
319 error for all evaluated algorithms. This graphical representation maps the proportion of data
320 points falling within specific error margins, providing an additional measure for assessing
321 algorithmic precision and consistency. A steep ascent in cumulative frequency at minimal error

thresholds indicates that a substantial fraction of predictions closely aligns with the actual experimental values, underscoring the model’s reliability.

As depicted in Figure 5, GrowNet achieves the maximum cumulative frequency across all error boundaries, demonstrating that the overwhelming majority of its outputs are characterized by negligible deviations. This observation further supports GrowNet’s predictive capability within the investigated dataset. Conversely, TabNet and GBoost exhibit marginally lower cumulative frequencies, indicative of intermediate predictive performance. The RF model records the broadest error dispersion and the lowest overall predictive consistency among the tested frameworks.

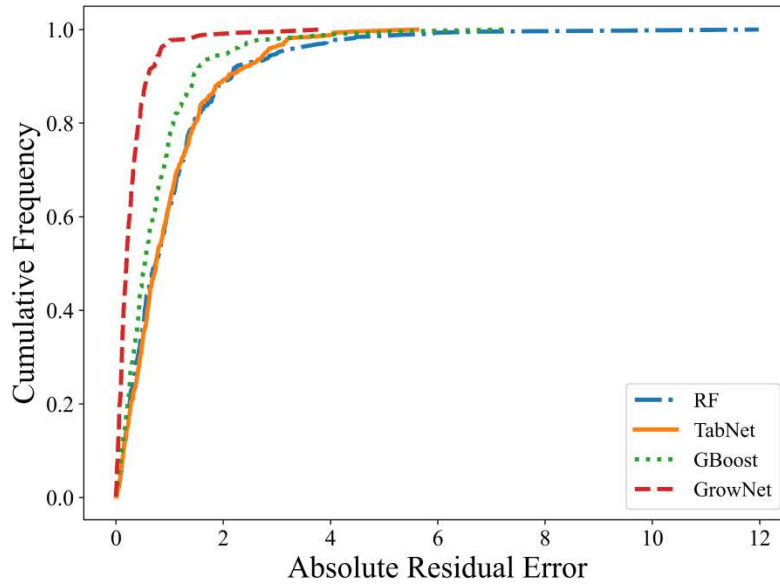


Figure 5. Cumulative distribution analysis of model residuals for performance comparison.

Figure 6 illustrates a visual comparison of the residual errors for all tested models, highlighting the differences between the models’ predictions and the actual experimental data. Among the evaluated algorithms, GrowNet features the tightest error distribution. This indicates that most of

its calculated values are in close agreement with the experimental results, supporting its strong predictive accuracy. On the other hand, the other models show wider error margins, corresponding to larger discrepancies from the measured data and lower overall predictive precision.

The graphical analyses presented in Figures 2–6 consistently support the statistical results summarized in Table 3. However, it should be noted that several of these visualizations provide partially overlapping information regarding model performance. Collectively, the results indicate that GrowNet achieved the lowest overall prediction error for the investigated dataset, while TabNet also demonstrated strong predictive behavior.

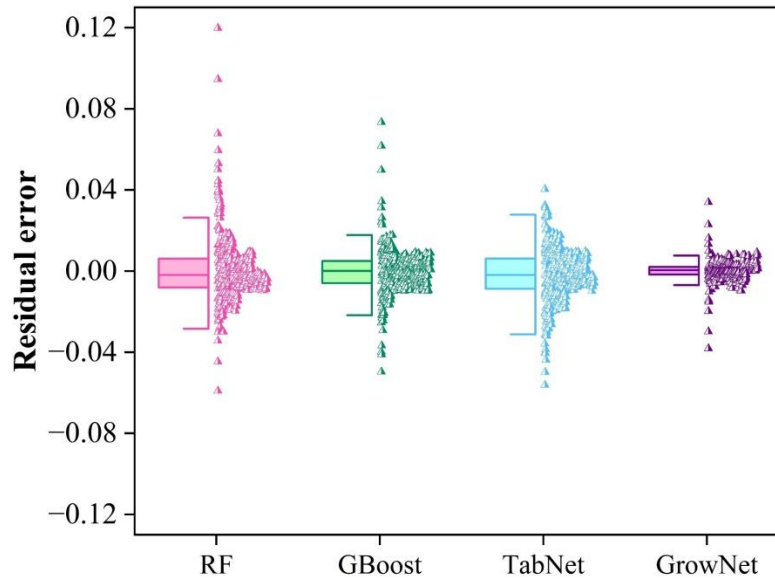


Figure 6. Evaluation of error trends and overall predictive behavior of the models.

Group error plots serve as an effective mechanism for assessing model efficacy across varying experimental states. Figure 7 delivers a comparative evaluation of prediction deviations for all implemented models, enabling a comprehensive analysis of their behavior under different P and T conditions. The findings reveal that the GrowNet architecture consistently yields minimized

errors and superior accuracy across the entire parameter spectrum. Moreover, its predictive stability persists despite fluctuations in input variables, suggesting that its forecasting competence is not confined to restricted intervals. Ultimately, these findings highlight the predictive consistency and robustness of the GrowNet framework under a wide array of thermodynamic conditions.

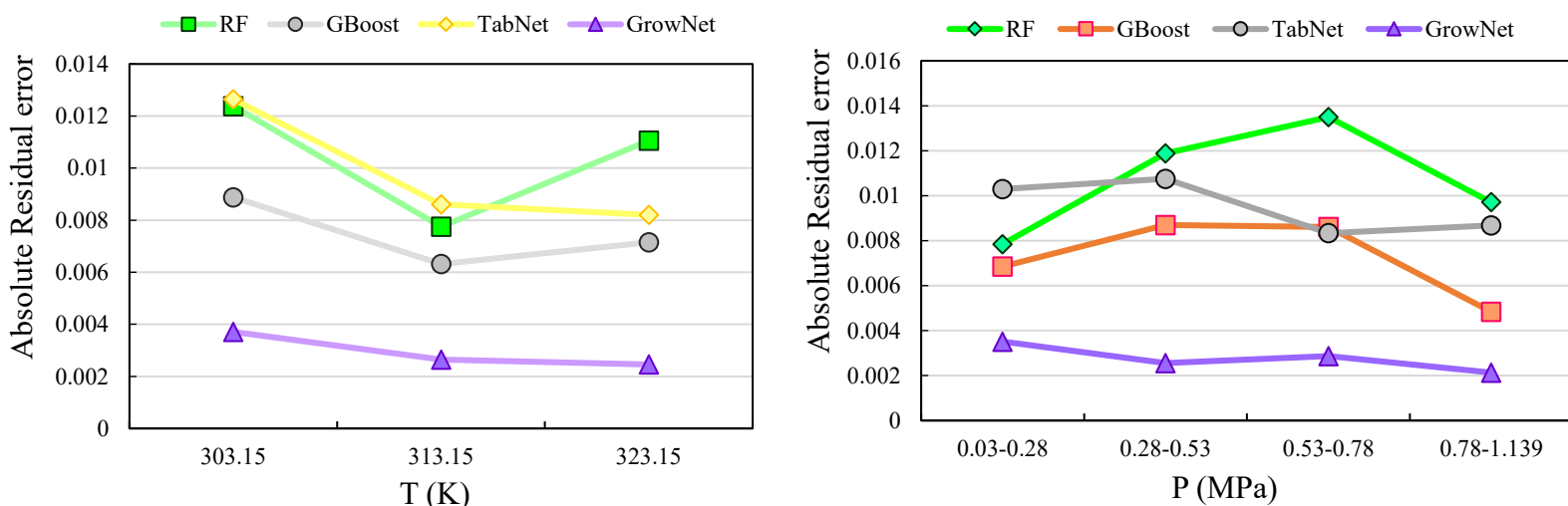


Figure 7. Distribution of model errors categorized by individual input parameters.

Figure 8 compares the absolute residual errors of various models for the three F-gases: R32, R134a, and R125. The plot clearly shows that the GrowNet model exhibits lower error values than the other models, confirming its superior accuracy in predicting solubility for all three gases. Notably, the GrowNet model records a minimum error value of 0.0035 across all gases, underscoring its reliability and robustness in predicting the solubility of F-gases in DESs.

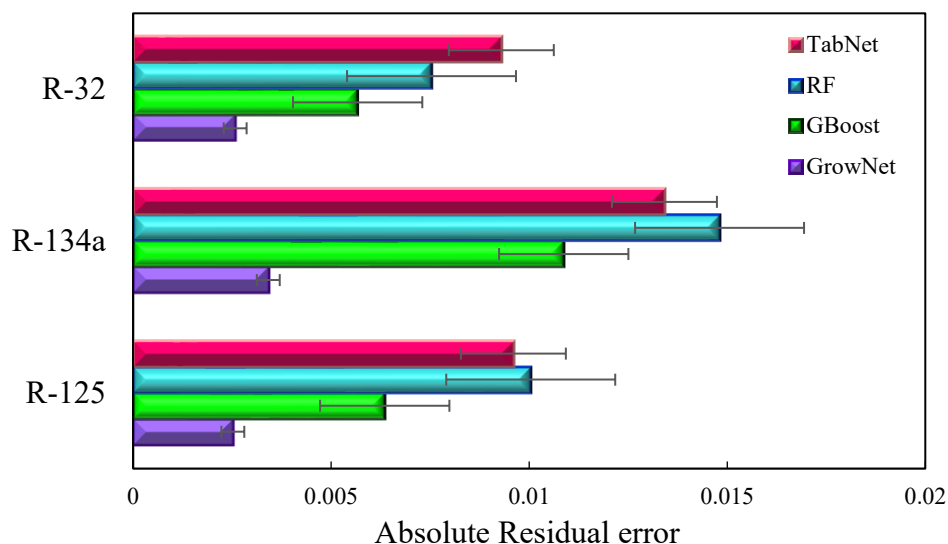


Figure 8. Absolute residual error comparison of different models for predicting different F-gases solubility (R32, R134a, and R125) in DESs.

4.3. Model Trend Analysis

Figure 9 displays the trend analysis comparing the GrowNet predictions against experimental data for different variables, illustrating its proficiency in capturing the essential thermodynamics of F-gas solubility in DESs. Specifically, Figure 9a illustrates the variation of F-gas solubility with pressure at different temperatures for the $[N_{4444}][C_4F_9SO_3]/C_4F_9CO_2H$ (1:1) DES. Along with established thermodynamic theories, the solubility of F-gases increases with pressure due to enhanced gas-liquid interactions and a reduction in system free volume, which promotes gas dissolution. Conversely, solubility declines at higher temperatures, where elevated thermal energy weakens intermolecular forces and favors the release of gas molecules from the solvent phase. The GrowNet model accurately reproduces these experimental trends, confirming its ability to capture pressure- and temperature-dependent solubility variations.

378 Figure 9b presents the solubility behavior of F-gases against pressure at a fixed temperature of
379 323.15 K, examining $[\text{C}_2\text{C}_1\text{Im}][\text{C}_8\text{F}_{17}\text{SO}_3]/\text{C}_4\text{F}_9\text{CO}_2\text{H}$ DES at three different molar ratios. The
380 results reveal that solubility increases with increasing proportions of $[\text{C}_2\text{C}_1\text{Im}][\text{C}_8\text{F}_{17}\text{SO}_3]$,
381 indicating that the fluorinated ionic component enhances the affinity of the DES for F-gases.
382 This effect can be attributed to stronger fluorophilic and van der Waals interactions between the
383 gas molecules and the fluorinated moieties of the DES. The GrowNet model effectively captures
384 this compositional dependency, confirming its predictive consistency across different mixture
385 ratios.

386 Figure 9c compares the solubility of three F-gases, R-32, R-125, and R-134a in
387 $[\text{C}_2\text{C}_1\text{Im}][\text{C}_8\text{F}_{17}\text{SO}_3]/\text{C}_4\text{F}_9\text{CO}_2\text{H}$ (1:1) DES at 313.15 K. Experimental measurements indicate
388 that R-125 and R-32 exhibit similar solubility values, whereas R-134a shows higher solubility,
389 likely due to its larger molecular size and stronger polarizability, which enhance dispersion and
390 dipole-induced interactions with the DES. The GrowNet model accurately predicts these relative
391 trends, confirming its capability to differentiate between molecular characteristics and their
392 corresponding solubility behaviors.

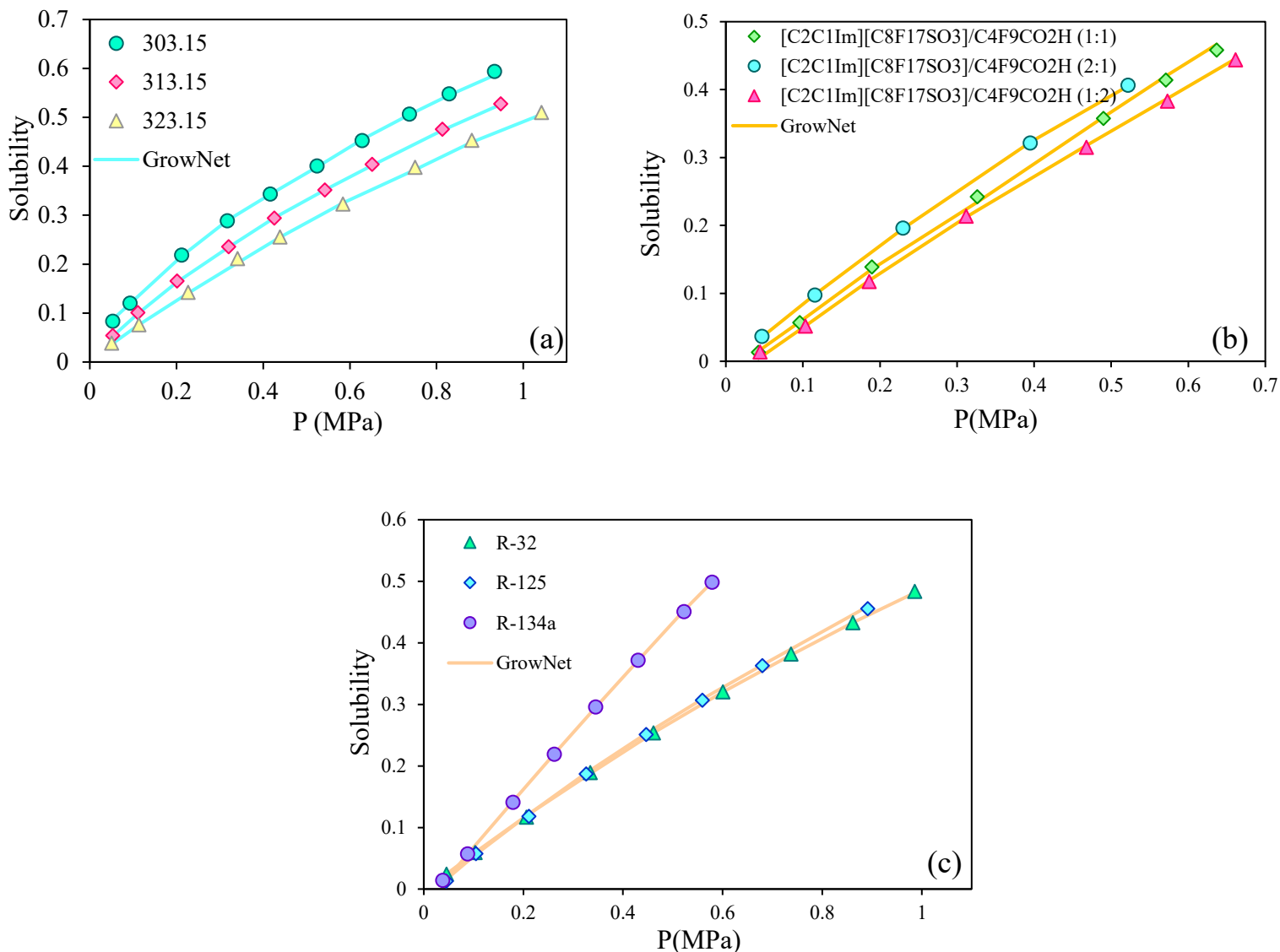


Figure 9. Validation of GrowNet predictions against experimental data (a) temperature and pressure effects, (b) influence of DES molar ratios, and (c) specific behavior of R-32, R-125, and R-134a.

4.4. Model Interpretability via SHAP Values

Figure 10 presents the SHAP summary analysis, detailing the isolated effects of each input variable on the GrowNet algorithm's prediction of F-gas solubility within DESs. Specifically, Figure 10a exhibits a detailed scatter plot of SHAP values, illustrating how fluctuations in feature magnitudes directionally dictate the solubility outcomes. The applied color spectrum,

transitioning from blue, representing low feature values, to red, indicating high ones, enables an intuitive visual interpretation of parameter dynamics. Furthermore, Figure 10b establishes a hierarchical ranking of these variables utilizing their mean absolute SHAP scores. This particular plot isolates the magnitude of global feature importance, independent of the directional impact, affirming that parameters with elevated mean values exert a dominant influence on the model's predictive behavior.

Among the evaluated parameters, pressure (P) stands out as the most influential variable, evidenced by its broad distribution and maximum mean SHAP value. Elevated pressures (as represented by the SHAP distribution in Figure 10a) strongly influence the model predictions. However, SHAP values describe the behavior of the trained ML/DL model and should not be interpreted directly as universal thermodynamic causality. Therefore, the SHAP trends should be interpreted cautiously and in conjunction with the experimentally observed solubility behavior discussed in Figure 9. Conversely, temperature (T) acts as the second most critical determinant within the trained model and exerts an opposing effect; higher temperatures (red markers on the right) increase the predicted solubility.

Furthermore, the acentric factor ($\omega_{\text{F-gas}}$) and critical temperature ($T_{c,\text{F-gas}}$) of the F-gases are pivotal in governing solubility dynamics. Elevated $\omega_{\text{F-gas}}$ values are associated with increased predicted solubility within the trained model, suggesting that gases exhibiting greater molecular asymmetry and stronger non-ideal interactions may interact more favorably with the investigated DES systems. Similarly, an increase in $T_{c,\text{F-gas}}$ correlates with increased predicted solubility, indicating that gases characterized by robust intermolecular interactions or elevated cohesive energy densities may exhibit enhanced affinity toward DES environments. The critical volume of

423 the DES ($V_{c,DES}$) also demonstrates a moderate positive correlation, implying that solvents with
424 expansive molar volumes may provide greater structural flexibility and free volume for gas
425 incorporation.

426 On the other hand, the density of the DES (ρ_{DES}) exerts a marginal positive influence, suggesting
427 that slight increases in solvent density may optimize gas–solvent interactions up to a certain
428 threshold before steric hindrance becomes restrictive. The critical properties of the DES ($P_{c,DES}$
429 and $T_{c,DES}$), alongside the critical pressure of the F-gas ($P_{c,F-gas}$), offer minor yet constructive
430 contributions. This indicates that solvents possessing higher critical parameters generally foster
431 more stable matrices for gas integration across varying thermodynamic states. Lastly, the DES
432 acentric factor (ω_{DES}) exhibits a negligible effect, revealing that the solvent’s molecular non-
433 ideality is of secondary importance compared to dominant external thermodynamic variables.

434 In conclusion, the SHAP analysis underscores that pressure and temperature represent the
435 dominant variables influencing the model predictions for the investigated dataset. Meanwhile,
436 molecular descriptors specifically ω_{F-gas} , $T_{c,F-gas}$, and $V_{c,DES}$ provide essential secondary
437 adjustments that contribute to the nonlinear relationships learned by the model. Ultimately, these
438 insights confirm that the synergistic interplay between macroscopic states and molecular-level
439 characteristics comprehensively dictates F-gas solubility in DES frameworks.

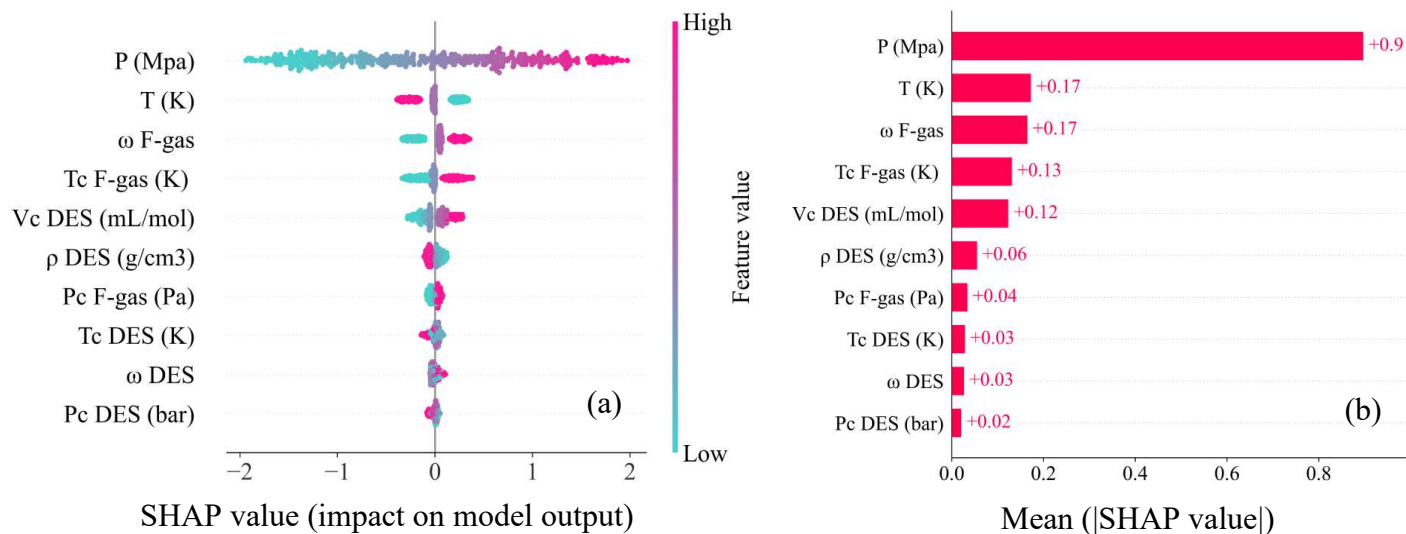


Figure 10. Feature importance analysis evaluating the impact of each input variable on the solubility results.

Figure 11 demonstrates the SHAP-based interpretability analysis of the proposed GrowNet framework for estimating F-gas solubility in DESs. It is important to note that the input features and the corresponding predicted values discussed in this section represent scaled (normalized) data. Figure 11a displays the SHAP summary plot, highlighting the overall impact of individual input variables across ten selected samples, whereas Figures 11b and 11c offer an in-depth local analysis specifically for the seventh data point. The investigated features comprise P , T , ω_{F-gas} , $V_{c,DES}$, $T_{c,F-gas}$, $P_{c,F-gas}$, ρ_{DES} , $T_{c,DES}$, $P_{c,DES}$, and ω_{DES} . In Figure 11a, parameters that positively drive the predicted solubility are colored pink, while negative contributors appear in blue. The visual evidence confirms that pressure (P) and temperature (T) exhibit the most substantial variability and dominant impact, verifying that macroscopic thermodynamic conditions primarily dictate the model prediction of the investigated dataset. Conversely, features like the acentric factor and DES density exert minor influences, implying that structural and molecular-shape attributes are secondary to bulk thermodynamic drivers.

Figure 11b breaks down the local SHAP explanations for the seventh sample, quantifying the exact scaled contribution of each feature to the target prediction. The bar chart demonstrates that pressure delivers the strongest negative influence (-1.38), drastically reducing the predicted model output. On the other hand, temperature and the F-gas critical temperature ($T_{c,F-gas}$) add positive increments ($+0.17$ and $+0.10$, respectively), thereby promoting solubility. These SHAP contributions describe the local behavior of the trained model for this specific observation and should not be interpreted directly as universal thermodynamic causality. Furthermore, the positive impact of $T_{c,F-gas}$ suggests stronger intermolecular interactions between the gas and DES molecules. Minor negative shifts are caused by $V_{c,DES}$ (-0.08) and ω_{F-gas} (-0.05), indicating that these descriptors contribute negatively to the predicted output for this specific data point within the trained model. Slight positive additions from $T_{c,DES}$ and $P_{c,DES}$ ($+0.04$ each) indicate that solvents with higher critical coordinates are associated with slightly higher predicted solubility values in this local SHAP analysis.

The SHAP force plot for the seventh observation is visualized in Figure 11c, mapping how the competing input features push the GrowNet model from its expected base value to the final scaled prediction of $f(x) = -1.48$. Pressure acts as the dominant suppressing force, within the local prediction generated by the trained model. In opposition, T and $T_{c,DES}$ push the prediction higher for the specific observation. Additional parameters, including $T_{c,F-gas}$, $V_{c,DES}$, ω_{F-gas} , and $P_{c,F-gas}$, introduce smaller opposing adjustments. Ultimately, the final output results from the interplay of these contrasting effects, where the strong suppressive role of pressure is only partially mitigated by temperature-driven properties, illustrating how the trained model combines the different descriptors to generate the final prediction.

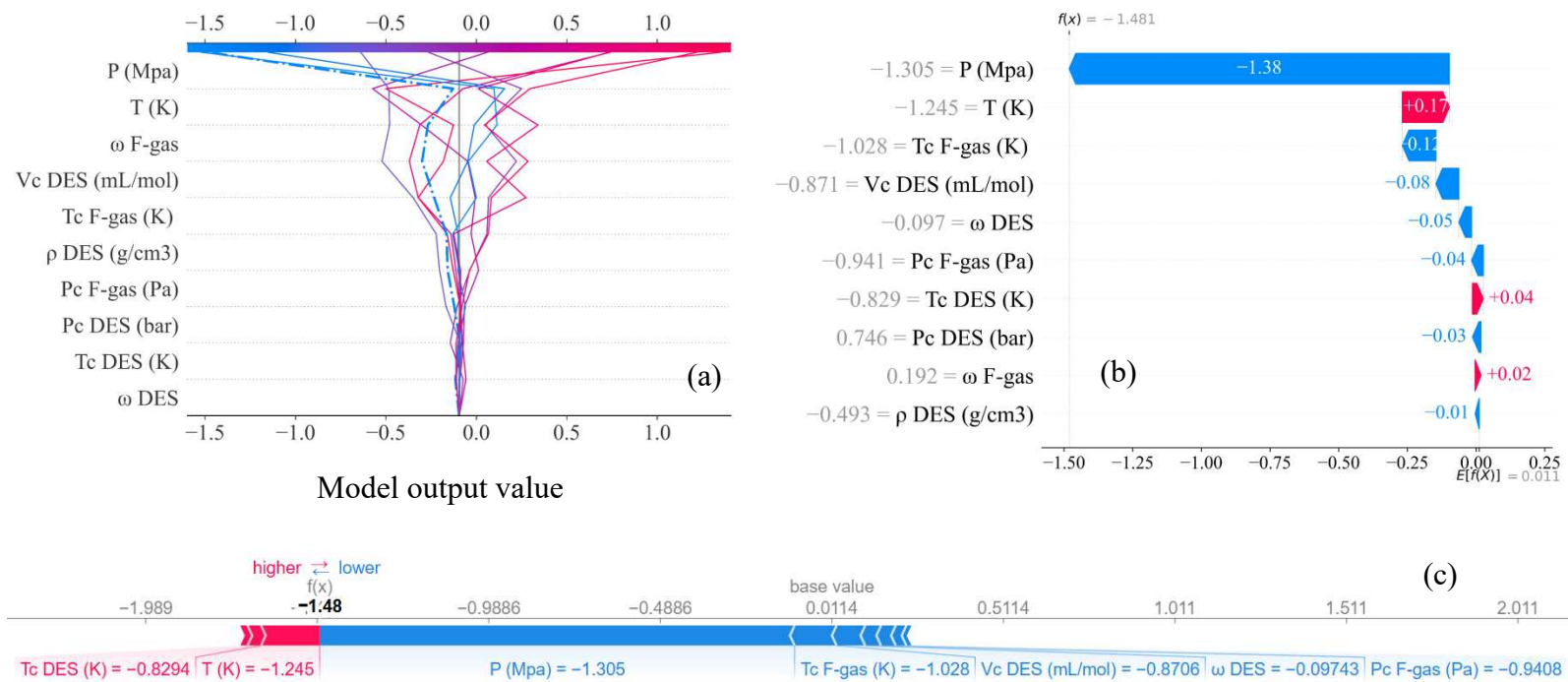


Figure 11. SHAP-based explainability of the GrowNet model: (a) Global importance of features derived from ten typical samples; (b) Local feature effects for observation 7; and (c) Force plot depicting the additive influence of all variables on predicted solubility.

4.5. Identification of Anomalies and Model Application Bounds

The leverage statistical method serves as a widely used diagnostic tool for establishing the applicability domain of predictive models and detecting abnormal or outlier data points. This assessment relies on the generation of a Williams plot, as illustrated in Figure 12. The plot is created by mapping standardized residuals (R) against the leverage values obtained from the Hat matrix (H), which allows for the clear identification of predictions falling within a secure and reliable zone. The foundational equations used to derive these metrics are detailed below [33, 34]:

- Hat matrix (H):

$$H = X(X^T X)^{-1}X^T \quad (5)$$

489 Within these mathematical expressions, H stands for the Hat matrix, while X designates the input
 490 (or design) matrix. The superscript T is used to signify the matrix transpose operation.
 491 Furthermore, the specific entries located along the main diagonal of H are known as the Hat
 492 values.

493 • Leverage limit (H^*):

494

$$H^* = \frac{3 \times (p+1)}{n} \quad (6)$$

495 To assess the reliability of the employed databank and define the usability domain of the
 496 GrowNet model, the Leverage approach and the corresponding Williams plot (Figure 12) were
 497 utilized. The critical hat value (H^*) is calculated based on the dataset size (n) and the number of
 498 input features (p). In this plot, data points located in the range of $0 \leq H \leq H^*$ and $-3 \leq R \leq$
 499 3 are considered valid data, confirming the statistical dependability of these measurements.
 500 Points that maintain valid leverage ($H \leq H^*$) but exceed the residual boundaries of ± 3 are
 501 flagged as suspected data (outliers). Furthermore, data showing $H > H^*$ are considered out of
 502 leverage limits; if their residuals remain within $-3 \leq R \leq 3$, they are well-predicted but
 503 structurally outside the model's domain (good high leverage), whereas exceeding the residual
 504 limits classifies them as bad high leverage.

505 As depicted in Figure 12, an overwhelming majority of the dataset (432 points, or 98.19%) falls
 506 within the valid region. Consequently, with only 8 points (1.81%) identified as suspected data,

the results demonstrate the good predictive reliability of the GrowNet system and confirm that the utilized dataset is of acceptable statistic quality [35, 36].

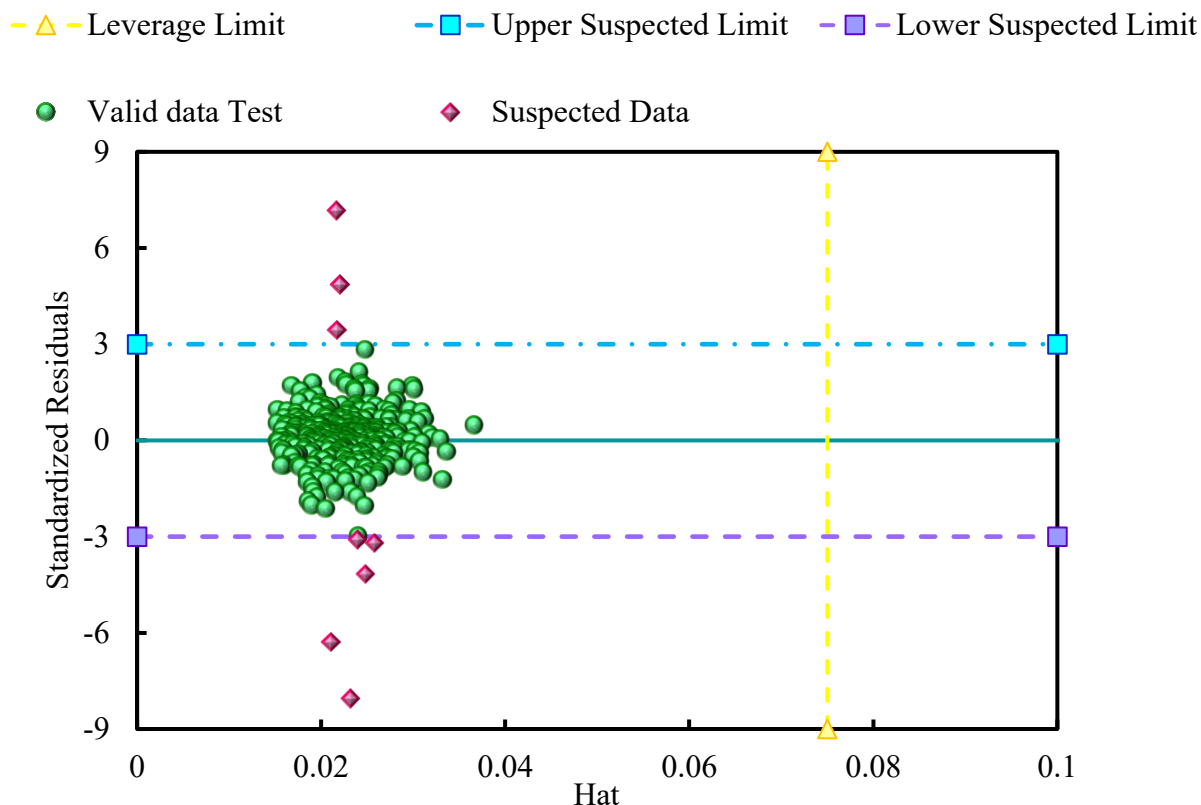


Figure 12. William's plot illustrating the leverage analysis for evaluating the reliability of the GrowNet model predictions.

5. Conclusions

This study demonstrates, for the first time, the application of ML and DL models for predicting the solubility of fluorinated gases in DESs. The principal novelty of this work lies in the development of a predictive framework that integrates fundamental critical properties with advanced ML/DL algorithms, specifically the GrowNet model, to capture solubility behavior in DES systems included within the investigated descriptor and chemical space.

Analysis of four models, including TabNet, GBoost, GrowNet, and Random Forest, demonstrated that the DL models, particularly GrowNet, outperformed the conventional ML methods across the investigated range of temperatures, pressures, and DES compositions. The interpretability of the GrowNet model, evaluated through SHAP analysis, provided additional insight into the variables influencing the model predictions. It confirmed that macroscopic thermodynamic conditions, specifically pressure (negative effect) and temperature (positive effect), are the primary drivers of F-gas solubility within the trained model and investigated dataset. Furthermore, it quantified the significant secondary influence of molecular descriptors, particularly the acentric factor and critical temperature of the F-gas, highlighting the importance of gas-solvent interactions at the molecular level.

Finally, the leverage approach confirmed the model's broad applicability, with over 98% of the data identified as reliable, underscoring its robustness for practical implementation within the investigated descriptor space. This work establishes a data-driven framework for screening and designing task-specific DESs by connecting data-driven prediction and physicochemical understanding, enabling future development of efficient and sustainable separation processes in the chemical industry.

Future research should focus on expanding the database to encompass a broader spectrum of F-gases and DESs. Furthermore, incorporation of advanced molecular descriptors, such as sigma-profiles from COSMO-RS or quantum chemical parameters, could provide improved molecular-level understanding of the specific intermolecular interactions governing solubility, thereby enhancing predictive accuracy and physicochemical interpretability.

Declarations

Funding Statement

The author received no financial support for the research, authorship, and publication of this article.

Competing interests

The authors declare no competing interests.

Data Availability Statement

Data is provided within the manuscript. Further analyses or data will be available upon request via corresponding author's contact info, a.sanati@hsu.ac.ir.

References

1. Parmesan, C., M.D. Morecroft, and Y. Trisurat, Climate change 2022: Impacts, adaptation and vulnerability. 2022, GIEC.
2. Heath, E.A., Amendment to the Montreal protocol on substances that deplete the ozone layer (Kigali amendment). International Legal Materials, 2017. **56**(1): p. 193-205.
3. Khalili, Y., et al., Advancements in hydrogen storage technologies: Integrating with renewable energy and innovative solutions for a sustainable future. Energy Geoscience, 2025: p. 100408.
4. Castro, P.J., et al., Absorption of fluorinated greenhouse gases in deep eutectic solvents. Industrial & Engineering Chemistry Research, 2020. **59**(29): p. 13246-13259.
5. Sheikshoei, A.H., et al., Predicting ammonia solubility in ionic liquids using machine learning models based on critical properties. Results in Engineering, 2025: p. 106951.
6. Sheikshoei, A.H. and A. Sanati, Data-driven insights into CO₂ solubility in deep eutectic solvents. Journal of CO₂ Utilization, 2025. **102**: p. 103285.
7. Demireb, M.G., S.B.R. Reartes, and F. Llorell, Thermodynamic analysis of the absorption of common refrigerants in fluorinated deep eutectic solvents. Fluid Phase Equilibria, 2024. **581**: p. 114077.
8. Zuo, Z., et al., Selective absorption of fluorinated gases by deep eutectic solvents: Thermodynamics and molecular insights. Separation and Purification Technology, 2025. **364**: p. 132342.
9. Wang, J., F. Wei, and H. Qin, Extending the UNIFAC model to facilitate deep eutectic solvent screening for fluorinated gas absorption with molecular-level insights. Chemical Engineering Science, 2025: p. 122493.
10. Boubia, A., et al., Critical Properties of Ternary Deep Eutectic Solvents Using Group Contribution with Extended Lee–Kesler Mixing Rules. ACS Omega, 2023. **8**(14): p. 13177-13191.
11. Darwish, A.S., et al., Machine learning approach for mapping the heat capacity of deep eutectic solvents for sustainable energy applications. Fuel, 2025. **381**: p. 133278.
12. Lemaoui, T., et al., Machine learning approach to map the thermal conductivity of over 2,000 neoteric solvents for green energy storage applications. Energy Storage Materials, 2023. **59**: p. 102795.
13. Pourghasemi, H.R. and N. Kerle, Random forests and evidential belief function-based landslide susceptibility assessment in Western Mazandaran Province, Iran. Environmental Earth Sciences, 2016. **75**(3): p. 185.
14. Breiman, L., Classification and regression trees. 2017: Routledge.
15. Breiman, L., Random Forests. Machine Learning, 2001. **45**(1): p. 5-32.

16. Erwan, S., B. Gérard, and V. Jean-Philippe, Consistency of random forests. *The Annals of Statistics*, 2015. **43**(4): p. 1716-1741.
17. Wang, S. and S. Chen, Insights to fracture stimulation design in unconventional reservoirs based on machine learning modeling. *Journal of Petroleum Science and Engineering*, 2019. **174**: p. 682-695.
18. Sheikhshoeai, A.H. and A. Sanati, New insight into viscosity prediction of imidazolium-based ionic liquids and their mixtures with machine learning models. *Scientific Reports*, 2025. **15**(1): p. 22672.
19. Kumar, R., B. Rai, and P. Samui, Machine learning techniques for prediction of failure loads and fracture characteristics of high and ultra-high strength concrete beams. *Innovative Infrastructure Solutions*, 2023. **8**(8): p. 219.
20. Elith, J., J.R. Leathwick, and T. Hastie, A working guide to boosted regression trees. *Journal of animal ecology*, 2008. **77**(4): p. 802-813.
21. James, G., et al., *An introduction to statistical learning*. Vol. 112. 2013: Springer.
22. Sheikhshoeai, A.H., A. Khoshshima, and D. Zabihzadeh, Predicting the heat capacity of strontium-praseodymium oxysilicate $\text{SrPr}_4(\text{SiO}_4)_3\text{O}$ using machine learning, deep learning, and hybrid models. *Chemical Thermodynamics and Thermal Analysis*, 2025. **17**: p. 100154.
23. Arik, S.Ö. and T. Pfister, Tabnet: Attentive interpretable tabular learning. in *Proceedings of the AAAI conference on artificial intelligence*. 2021.
24. Sheikhshoeai, A.H. and A. Khoshshima, Machine and deep learning models for predicting high pressure density of heterocyclic thiophenic compounds based on critical properties. *Scientific Reports*, 2025. **15**(1): p. 25465.
25. McDonnell, K., et al., Deep learning in insurance: Accuracy and model interpretability using TabNet. *Expert Systems with Applications*, 2023. **217**: p. 119543.
26. Sheikhshoeai, A.H., A. Sanati, and A. Khoshshima, Deep learning models to predict CO_2 solubility in imidazolium-based ionic liquids. *Scientific Reports*, 2025. **15**(1): p. 26445.
27. Badirli, S., et al., Gradient boosting neural networks: Grownnet. *arXiv preprint arXiv:2002.07971*, 2020.
28. Shen, Y. and S. Zhang, A Cox-optimized survival model based on GrowNet. *Journal of Statistical Computation and Simulation*, 2024. **94**(3): p. 624-641.
29. Li, F., et al., Boostxml: gradient boosting for extreme multilabel text classification with tail labels. *IEEE Transactions on Neural Networks and Learning Systems*, 2023. **35**(11): p. 15292-15305.
30. Sheikhshoeai, A.H. and A. Sanati, Interfacial Tension Modeling of Aqueous Ionic Liquids via Machine and Deep Learning. *Energy & Fuels*, 2025. **39**(36): p. 17506-17521.
31. Yan, T., et al., Prediction of geological characteristics from shield operational parameters by integrating grid search and K-fold cross validation into stacking classification algorithm. *Journal of Rock Mechanics and Geotechnical Engineering*, 2022. **14**(4): p. 1292-1303.
32. Pedregosa, F., et al., Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 2011. **12**: p. 2825-2830.
33. Gramatica, P., *Principles of QSAR models validation: internal and external*. *QSAR & combinatorial science*, 2007. **26**(5): p. 694-701.
34. Rousseeuw, P.J. and A.M. Leroy, *Robust regression and outlier detection*. 2003: John wiley & sons.

- 632 35. Sheikhshoei, A.H., et al., Predicting ammonia solubility in ionic liquids using machine
633 learning models based on critical properties. Results in Engineering, 2025. **27**: p. 106951.
634 36. Sheikhshoei, A.H. and R. Zabihi, Machine learning approach to predict the viscosity of
635 perfluoropolyether oils. Scientific Reports, 2025. **15**(1): p. 35273.

636