Accepted Manuscript
Version of Record at: https://doi.org/10.1002/aic.18392

Ionic liquid binary mixtures: machine learning-assisted modeling, solvent tailoring, process design and optimization

Yuqiu Chen [a,c], Sulei Ma [b], Yang Lei [b,*], Xiaodong Liang [c], Xinyan Liu [b,*], Georgios M. Kontogeorgis [c], Rafiqul Gani [d,e,f]

[a] *Department of Chemical and Biomolecular Engineering, University of Delaware, 150 Academy Street, Newark, Delaware 19716, United States*

[b] *School of Chemistry and Chemical Engineering, Hubei Key Laboratory of Coal Conversion and New Carbon Materials, Wuhan University of Science and Technology, Wuhan 430081, Hubei, China*

[c] *Department of Chemical and Biochemical Engineering, Technical University of Denmark, DK-2800 Lyngby, Denmark*

[d] *PSE for SPEED Company, Charlottenlund, Denmark*

[e] *Sustainable Energy and Environment Thrust, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China*

[f] *Department of Applied Sustainability, Széchenyi István University, Győr, Hungary*

[*]Corresponding author. *E-mail addresses:* leiyang@wust.edu.cn (Yang Lei); liuxinyan@wust.edu.cn (Xinyan Liu)

Abstract

This work conducts a comprehensive modeling study on the viscosity, density, heat capacity, and surface tension of ionic liquid (IL)-IL binary mixtures by combining the group contribution (GC) method with three machine learning algorithms: artificial neural network, XGBoost, and LightGBM. A large number of experimental data from reliable open sources is exhaustively collected to train, validate, and test the proposed

ML-based GC models. Furthermore, the Shapley Additive Explanations technique is employed to quantify the influential factors behind all the studied properties. Finally, these ML-based GC models are sequentially integrated into computer-aided mixed solvent design, process design, and optimization through an industrial case study of recovering hydrogen from raw coke oven gas. Optimization results demonstrate their high computational efficiency and integrability in solvent and process design, while also highlighting the significant potential of IL-IL binary mixtures in practical applications.

Keywords: Ionic liquid mixtures; machine learning; property modeling; solvent tailoring; H_2 recovery.

1. Introduction

Ionic liquids (ILs) are innovative fluids that have garnered significant attention from both academia and industries. Over the past few decades, extensive research has been conducted on the applications of ILs in various fields such as electrochemistry [1], synthetic materials [2], and pharmaceutical manufacturing [3, 4]. In these areas, ILs serve different roles, including as extractants/absorbents in separations, media and/or catalysts in lignocellulosic biomass pretreatment, and functional materials in batteries. With the growing interest in ILs within industrial settings, it is crucial to deeply understand their property behaviors for effective product and process design. Consequently, the characterization of IL properties and the establishment of structure-property relationships for ILs and IL-based mixtures are equally important in investigating their applications.

Viscosity is a crucial transport property that plays a significant role in fluid flow

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/aic.18392

analysis, process optimization, and material characterization [5]. Understanding and controlling viscosity enables engineers, researchers, and industries to design efficient systems, optimize processes, develop high-quality products, and ensure the desired performance and functionality of materials and fluids in a wide range of applications. Density serves as a fundamental property for understanding the behavior of substances. The knowledge of density is crucial for material identification, process design and optimization, fluid mechanics, energy calculations, and material selection [6]. Heat capacity is essential for thermodynamic calculations, energy balance analyses, temperature control systems, material selection, process optimization, and fundamental research. It plays a crucial role in various scientific, engineering, and industrial applications, enabling the efficient and effective design, operation, and optimization of processes involving heat transfer and energy conversion [7,8]. Surface tension is a crucial physicochemical property used in calculations related to fluid flow, gas absorption, distillation, heat transfer, sedimentation, and it plays a vital role in assessing transfer rates for vapor absorption, ensuring stability in pharmaceutical processing, and influencing the properties and performance of refrigerants [9].

Currently, the availability of experimental data on the physical properties of ILs, particularly their mixtures, is primarily limited to well-known IL families. Consequently, the development of empirical or theoretical models that can reliably predict the properties of ILs becomes paramount in the quest to explore high-performance ILs with significant potential in diverse applications. To date, several predictive models have been developed by different researchers to address this need, such as group contribution (GC)-based models [10], quantitative structure-property relationship (QSPR) models [11], Conductor-like Screening Model for Real Solvents (COSMO-RS) models [12], and fragment contribution-corresponding states (FC-CS) models [13]. Among these models, GC-based models are particularly popular due to their ease of use and their role as the foundation for the development of computer-aided molecular design (CAMD) methods [14]. The optimal design of compounds through manipulating properties at the molecular level is often the key to considerable scientific advances and improved process systems performance, as reported [15].

To the best of our knowledge, most of the published works mainly focus on the property prediction of pure ILs, and modeling studies on the physical property prediction of mixtures containing ILs, especially IL-IL mixtures, are still scarce. As reported, IL-IL mixed solvents often exhibit superior performance compared to pure ILs in many cases [16]. Predicting properties of IL-IL mixtures is a fundamental step in the rational design of processes involving these unique solvents. It not only enhances process efficiency but also contributes to the economic and environmental sustainability of industrial applications. Therefore, there is a high demand for predictive models for the properties of IL-IL mixtures. As mentioned above, GC-based models can be easily integrated into computer-aided design methods [17], which have been recognized as time- and cost-efficient techniques for product and process design. Moreover, GC-based models have the advantage of expanding the design space for these compounds and thereby expanding the potential applications of these property models. For these reasons, GC methods are preferred for building models to predict the density, heat capacity, and

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/aic.18392

viscosity of IL-IL mixtures in this study.

When constructing property models based on GC methods, initially, most properties are associated with group numbers using the multiple linear regression (MLR) algorithm [18]. However, some properties, particularly for mixture systems, cannot be adequately described by linear model expressions. In this regard, nonlinear GC models offer viable alternatives. Recently, machine learning (ML) algorithms such as artificial neural networks (ANN) and support vector machine (SVM) have been employed to develop nonlinear GC models for predicting CO_2 solubility in ILs [19], toxicity of ILs [20], and physical properties of IL-H_2O mixtures such as viscosity [21] and surface tension [22]. Additionally, hybrid modeling methods that combine GC methods and ANN algorithms have been introduced to model the phase equilibrium behavior of aqueous two-phase systems (ATPSs) [23-24]. Recent works have also demonstrated the seamless integration of ANN-based GC models with computer-aided design methods for the optimal design of IL-based aqueous biphasic systems [25]. Inspired by these successful applications, this study aims to combine the GC method with three popular ML algorithms, namely ANN, XGBoost, and LightGBM, to build models for predicting the viscosity, density, heat capacity, and surface tension of IL-IL mixtures. Next, the Shapley Additive Explanations (SHAP) technique [26] is employed to investigate the factors influencing these four physical properties of IL-IL mixtures. Furthermore, the integration of these ML-based GC models into computer-aided molecular design and process optimization is demonstrated through a case study of recovering hydrogen from raw coke oven gas. Figure 1 illustrates the framework of a comprehensive study on IL-

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/aic.18392

IL mixtures, encompassing ML-assisted modeling, mixed solvent tailoring, process design and optimization. It is worth mentioning that, while integrating mixed solvent and process design may provide a better solution, these two sub-problems are formulated and solved separately in this work. This is because the integrated design-based MINLP problem would be much more complex than the individual sub-problems. Additionally, demonstrating how the developed ML models can be integrated into solvent and process design, respectively, is another reason for solving the two sub-problems separately.

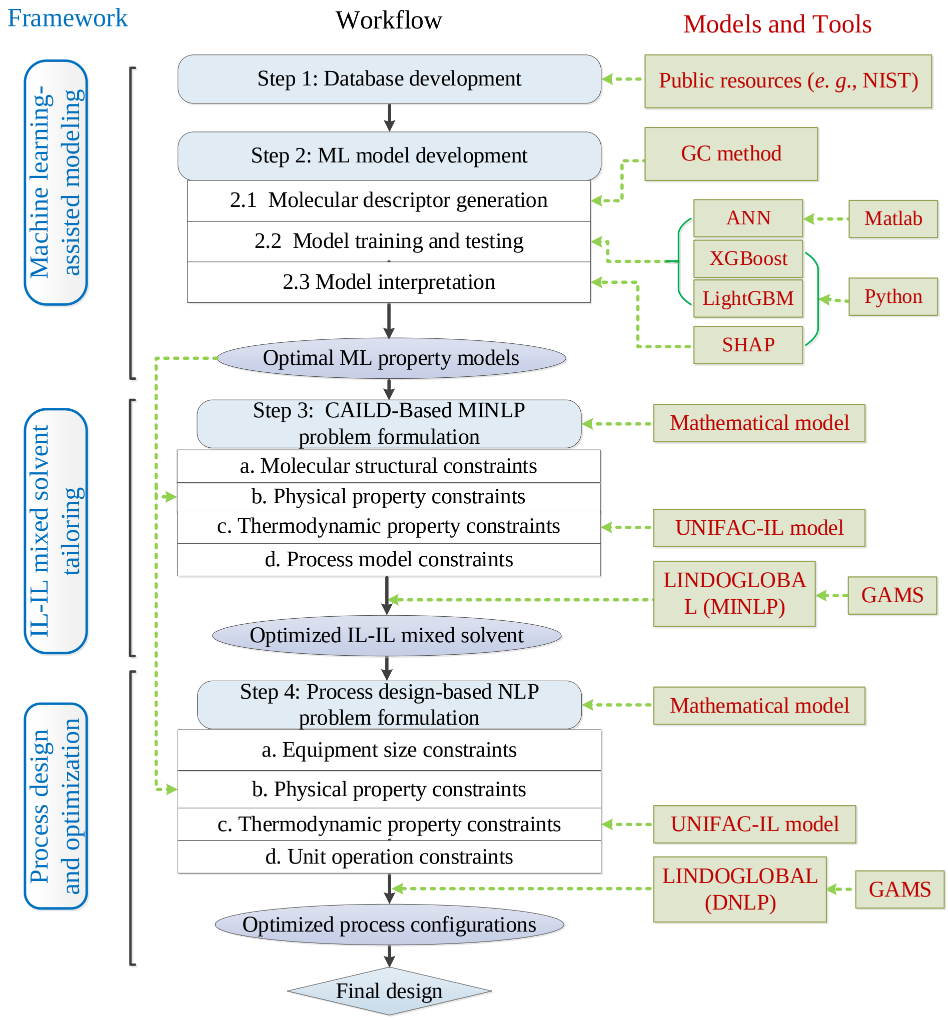


Figure 1. Framework of the machine learning-assisted modeling, solvent tailoring, process design and optimization.

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/aic.18392

2. Machine learning-assisted modeling

2.1 Experimental data

In this work, a total of 6141 experimental data points for viscosity, 3626 experimental data points for density, 730 experimental data points for heat capacity, and 626 experimental data points for surface tension of IL-IL mixtures are collected. These data points encompass 133 different IL-IL binary mixtures at various compositions and temperatures ranging from 263.64K to 533.2K [27-53]. The ILs cover six different family cations, including imidazolium, piperidinium, ammonium, pyridinium, and phosphonium（Im、Pyr、Py、Pip、N、P）. Table S1 provides a summary of all the collected IL-IL mixtures and their corresponding temperature ranges. To ensure the reliability of the model, an unbiased evaluation of model fit on the training data set is conducted. The collected viscosity experimental data points are divided into a training and validation set consisting of 5207 data points for model optimization, and a separate test set of 934 data points is used to evaluate the model's prediction performance, while for density, the training and validation set contains 3080 data points and the test set includes 546 data points. Similarly, for heat capacity, a training and validation set of 619 data points and a test set of 111 data points are used, while for surface tension, they are 549 and 77 data points, respectively. Detailed information about the training, validation, and test datasets for all properties can be found in Tables S2-S4 in the Supporting Information II.

2.2 Methods

GC method

To conduct GC-based modeling studies, the IL molecules are divided into three types of functional groups: cation skeleton groups, anion groups, and substituent groups on the cation skeleton. In this study, 4 types of cation skeletons, 25 anions, and 5 substituents are identified from the decomposition of IL molecules included in the dataset for the viscosity of IL-IL binary mixtures, while for density, 3 cation skeletons, 22 anions, and 5 substituents are determined. Similarly, for heat capacity, 3 cation skeletons, 6 anions, and 4 substituents are obtained, while for surface tension, there are 4 cation skeletons, 25 anions, and 5 substituents. Table S5 lists the input variables used in the ML modeling study for predicting viscosity, density, heat capacity, and surface tension in IL-IL binary mixtures. These variables encompass decomposed functional groups of ILs, mixture composition, the system's temperature and pressure.

In this work, Artificial Neural Network (ANN), XGBoost, and LightGBM models that have been proved to be effective in handling complex relationships within datasets are selected. ANN is known for its ability to capture intricate patterns and nonlinear dependencies. XGBoost and LightGBM, on the other hand, are gradient boosting algorithms that excel in predictive modeling, especially when dealing with large datasets. The choice of these three models was made to ensure a comprehensive evaluation of different machine learning approaches, considering their diverse strengths and capabilities. This allows us to assess and compare their performance in predicting the thermophysical properties of IL-IL binary mixtures accurately.

ANN algorithm

Artificial Neural Network (ANN) is a computational model inspired by the structure and function of the human brain. It consists of interconnected nodes, called artificial

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/aic.18392

neurons or "neurons," organized into layers. Each neuron takes input, performs a computation, and produces an output that is passed on to other neurons. The advantages of ANN in modeling studies can be summarized as: (1) ANNs can learn organically, meaning their outputs are not limited solely by inputs, and they have the ability to generalize from their inputs. (2) Nonlinear systems can find shortcuts to reach computationally expensive solutions. (3) ANNs have high fault tolerance potential. (4) ANNs can perform tasks beyond routing around non-operational parts of the network. The advantages of ANNs are particularly evident when a large experimental dataset with a wide range of variables is available.

In this work, a popular ANN architecture consisting of a three-layer (i.e. input layer, hidden layer, and output layer) feedforward network (see Figure S1.) is employed for building the model. The input layer consists of the fundamental structural fragments of the ILs, along with the temperature and molar fraction of the ILs, resulting in a total of N variables. The hidden layer processes the input information and transmits it to the output layer, which calculates the properties and measures the error between the predicted value and the experimental value. The outputs of the hidden layer (Z_1) and the output layer (Y) are represented by Equations (1) and (2), respectively.

$$Z_1 = f_1(W_1 P + B_1) \tag{1}$$

$$Y = f_2(W_2 Z_1 + B_2) \tag{2}$$

where W_1 and W_2 represent the weights of the neurons in the hidden layer and output layer, respectively. B_1 and B_2 are the bias parameters for the hidden layer and output layer, respectively. The number of neurons in the hidden layer plays a crucial role in achieving accurate predictions. If there are too many neurons, it can result in

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/aic.18392

overfitting, leading to significant biases when extrapolating beyond the training data. On the other hand, too few neurons can result in underfitting and significant deviations from experimental data. To address this issue, a validation set is utilized to optimize the number of neurons in the hidden layer of ANN model, as illustrated in Figures S2-S5 in the Supporting Information file I. These measures contribute to a robust and well-performing ANN model. In this study, the tansig function (Equation (3)) is used as an activation function in the hidden layer and the purelin transfer function (Equation (4)) is used in the output layer.

$$f_1(x) = \frac{2}{1+e^{-2x}} \tag{3}$$

$$f_2(x) = x \tag{4}$$

XGBoost algorithm

XGBoost is a parallel regression tree model based on the boosting technique, where boosting refers to obtaining the final classifier by weighting the sum of existing weak classifiers [54]. The XGBoost model is an improvement over the Gradient Boosted Decision Tree (GBDT) model. Compared to GBDT, XGBoost significantly enhances the speed of model training calculations and improves prediction and classification accuracy, making it an upgraded version of the GBDT algorithm. To prevent model overfitting and improve generalization ability, XGBoost incorporates regularization terms into the loss function of the GBDT model. The traditional GBDT loss function adopts a first-order Taylor expansion and uses the negative gradient value to replace the residual for fitting. In XGBoost, a second-order Taylor expansion is added to the loss function, capturing the second derivative to gather gradient direction information and improve model accuracy. Furthermore, XGBoost employs block-wise and sorted feature processing, enabling parallelization for finding the best-split points and thereby enhancing computation speed. The calculation of the output function is represented by Equation (5).

$$\hat{Y}_i^{T} = \sum_{k=1}^{T} f_k(x_i) = \hat{y}_i^{T-1} + f_T(x_i) \tag{5}$$

where $\hat{y}_i^{T-1}$ represents the generated tree, $f_T(x_i)$ is the newly created tree model, and T is the total number of tree models.

LightGBM algorithm

LightGBM, short for Lightweight Gradient Boosting Machine, is a Gradient Boosting Decision Tree (GBDT) algorithm framework renowned for its rapid training, minimal memory requirements, support for efficient parallel training, enhanced accuracy, and the ability to process extensive datasets swiftly [55]. While XGBoost is a widely recognized GBDT tool, it grapples with significant memory usage. To tackle this challenge, Microsoft developed LightGBM as an optimization of the conventional GBDT algorithm. LightGBM distinguishes itself from XGBoost by implementing the histogram algorithm instead of the pre-sorted algorithm. This choice results in decreased memory consumption and simplifies data separation. However, it's worth noting that the histogram algorithm may not be as time-efficient as the pre-sorted algorithm when processing sparse data. LightGBM incorporates two novel techniques, Exclusive Feature Bundling (EFB) and Gradient-based One-Side Sampling, into the histogram-based GBDT algorithm. EFB enables the fusion and bundling of certain features, reducing the feature dimension without sacrificing accuracy. Additionally, LightGBM employs a Leaf-wise strategy with a capped maximum depth, which serves

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/aic.18392

to prevent overfitting without compromising computational efficiency. You can find the mathematical formulation of LightGBM in Equation (6).

$$F(x) = \sum_{m=1}^{M} f_m(x) \tag{6}$$

In Equation 15, $F(x)$ represents the final output value, while $f_m(x)$ denotes the output value of the m-th weak regression tree.

Model evaluation indicators

Evaluating the accuracy of a model is a crucial step in developing machine learning models to assess their performance in making predictions. In this study, two evaluation metrics, namely the mean absolute error (MAE) and the squared correlation coefficient (R^2), as indicated by Equations (7) and (8), respectively, are employed.

$$R^2 = \frac{\sum_{i=1}^{N}\left(\vartheta_{i=1}^{\exp} - \vartheta_{i=1}^{cal}\right)^2}{\sum_{i=1}^{N}\left(\vartheta_{i=1}^{\exp} - \overline{\vartheta}\right)} \tag{7}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{N}\left|\vartheta_{i=1}^{\exp} - \vartheta_{i=1}^{cal}\right| \tag{8}$$

where N is the total number of the data points. $\vartheta_{i=1}^{\exp}$ and $\vartheta_{i=1}^{cal}$ represent the experimental and model-calculated values, respectively, and $\overline{\vartheta}$ denotes the average value of all included data points.

SHAP Interpretation

SHAP, a Python-based "model interpretation" package, is designed to interpret the outputs of machine learning models [56]. It derives its name from "SHapley Additive exPlanations" and employs a cooperative game theory-inspired approach to construct an additive explanatory model that treats all features as "contributors." In this framework, for every predicted sample, the model generates a predicted value, and the SHAP value quantifies the assigned value to each feature in that specific sample.

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/aic.18392

Notably, SHAP offers an advantage over other interpretation methods by being able to capture the influence of features in each sample and discern their positive and negative effects. The Shapley value is calculated according to Equation (9).

$$\hat{\phi}_j = \frac{1}{K}\sum_{K=1}^{K}\left(\hat{g}\left(x_{+j}^{m}\right)-\hat{g}\left(x_{-j}^{m}\right)\right) \tag{9}$$

where $\hat{g}\left(x_{+j}^{m}\right)$ is the prediction for x.

In this study, the SHAP method is used to interpret the relative importance of features such as temperature, molar fraction of IL, and structural fragments on the prediction results. The features for each property in all developed ML-based models are listed in Table S5.

2.3 Results and discussions

After conducting the initial machine learning calculations, it was observed that certain data points were impacting the overall model performance. To mitigate the adverse effects of these singular samples on the model, data normalization was applied. The normalized data was then split into training sets (70%), validation sets (15%), and test sets (15%) to prevent overfitting. For the ANN model optimization, a range of 1 to 20 neurons was considered, and the number of neurons yielding the best-fit performance was determined (see Figures S2-S5, Table S7). The Levenberg-Marquardt algorithm in MATLAB was employed for the optimization of ANN-based GC models, while the XGBoost- and LightGBM-based modeling studies were performed in Jupyter Notebook using Python programming. The optimized parameters of the ML models for each of the studied properties can be found in Table S6 (Supporting Information II).

Table 1 compares the performance of the GC model using ANN, XGBoost, and

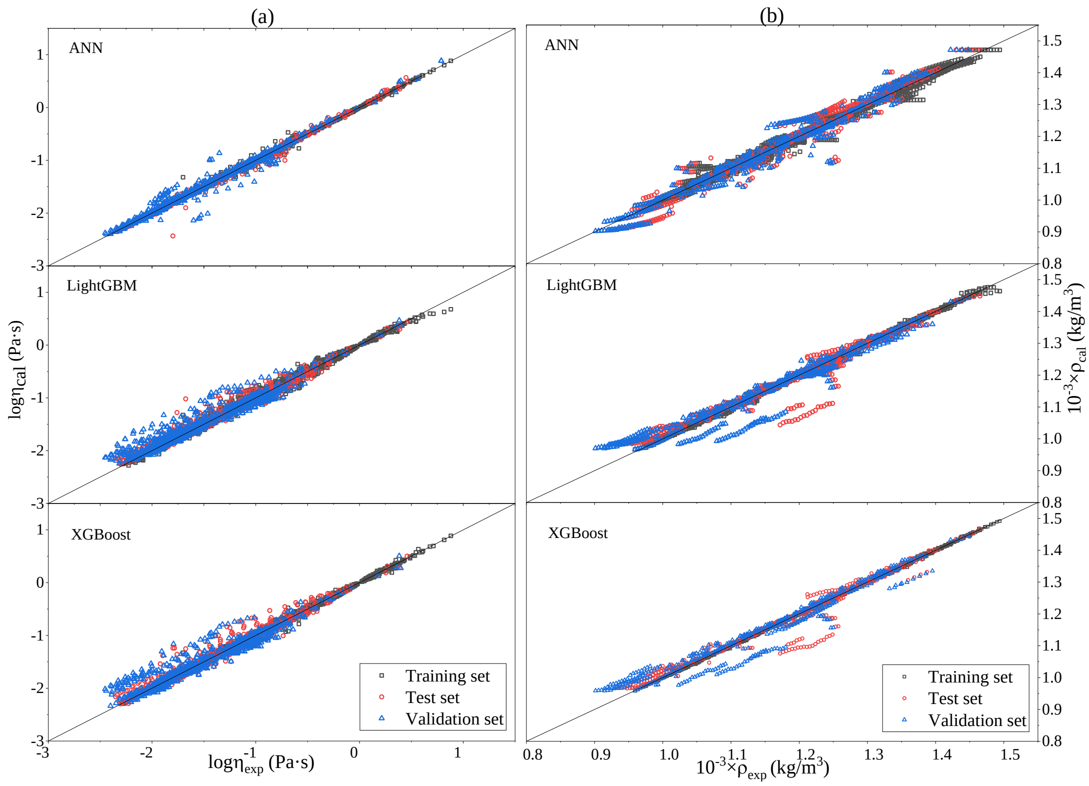
Accepted Manuscript
Version of Record at: https://doi.org/10.1002/aic.18392

LightGBM in the predictions on four studied properties, *i.e.,* viscosity, density, surface tension, and heat capacity. Although all three machine learning algorithms demonstrate strong modeling performance in all four studied properties of IL-IL binary mixtures, it is evident that the ANN-based GC model significantly outperforms its counterparts across the training, validation, and test sets in both viscosity and density predictions. In contrast, the LightGBM-based model provides the best prediction performance for surface tension, while the XGboost-based GC model offers the best prediction performance for heat capacity. To better illustrate the model performance, a comparison based on the estimated values using ANN, XGBoost, and LightGBM is presented in Figure 2. Obviously, the predicted values of viscosity and density from the ANN-based GC model are closest to the experimental values, while surface tension predicted from the LightXBG-based model and heat capacity predicted from the XGBoost-based model have the largest points distributed closer to the diagonal. A direct comparison of the predictive modeling performance using these three algorithms for each property is given in Figure S6. The errors between experimental values and predicted values in all property predictions using GC models with ANN, XGBoost, and LightGBM are demonstrated in Figure S7 (Tables S8 and S9).

Table 1. Performance of the GC model using ANN, XGBoost, and LightGBM in the predictions for viscosity, density, heat capacity, and surface tension

		Training set		Validation set		Test set	
Property	Model	MAE	R^2	MAE	R^2	MAE	R^2
Viscosity	ANN	0.000253	0.9999	0.000538	0.9999	0.000321	0.9999
	XGBoost	0.017319	0.9974	0.082490	0.9487	0.090668	0.8065
	LightGBM	0.038794	0.9872	0.100415	0.9326	0.104012	0.8288

Density	ANN	0.001171	0.9999	0.001246	0.9999	0.001352	0.9998
	XGBoost	0.001991	0.9969	0.011669	0.9545	0.026004	0.8795
	LightGBM	0.005329	0.9932	0.015860	0.9471	0.035892	0.8389
Heat capacity	ANN	0.002561	0.9994	0.002351	0.9990	0.000768	0.9999
	XGBoost	0.000976	0.9999	0.011721	0.9778	0.012095	0.9449
	LightGBM	0.003937	0.9976	0.015597	0.9344	0.025205	0.8816
Surface tension	ANN	0.001490	0.9151	0.001316	0.9473	0.001702	0.8433
	XGBoost	0.000561	0.9846	0.000869	0.9758	0.001170	0.9550
	LightGBM	0.000587	0.9811	0.001305	0.9411	0.002027	0.8751



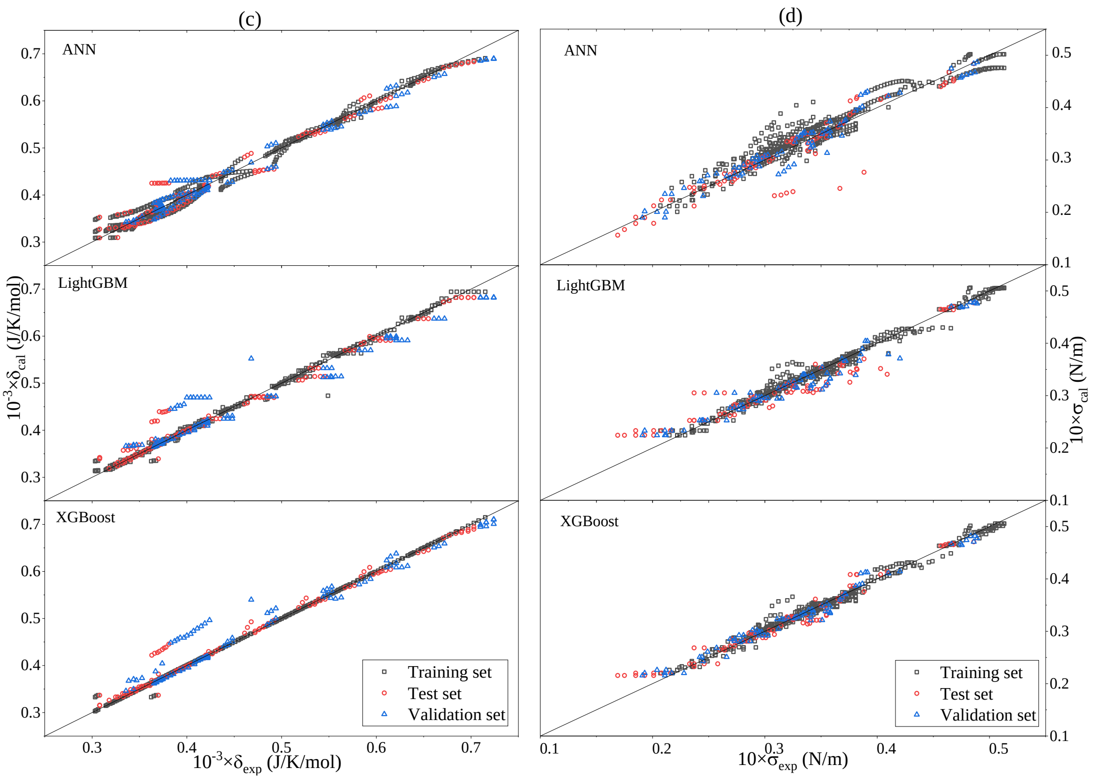


Figure 2. Plots of experimental and predicted values from the GC model using ANN, XGBoost, and LightGBM in the predictions of viscosity, density, heat capacity, and surface tension of IL-IL binary mixtures.

In order to investigate the influence of each feature on the density of IL-IL binary mixtures, SHAP analysis is performed on each influencing factor. The SHAP analysis results for viscosity prediction are presented in Figure S8(a-b), and the degree of interaction of each influencing factor is shown in Figure S8(c). It can be observed that the temperature of ILs has the greatest influence on the viscosity of IL-IL binary systems, followed by the molar fraction and $-CH_2$ group. In addition, the SHAP interaction diagrams indicate that the combination of temperature and molar fraction of the ILs has the greatest influence on the viscosity of IL-IL binary mixtures. For density, the functional group BF_4 of ILs in the binary system has the greatest influence and exhibits a negative correlation trend, followed by the functional group $-CH_2$.

Meanwhile, the combination of the BF_4 and $-CH_2$ groups of ILs has the greatest influence on the density of IL-IL binary mixtures. The temperature of ILs in the binary systems has the greatest influence on the heat capacity, followed by the molar fraction. The combination of temperature and molar fraction of the ILs has the greatest influence on the heat capacity of IL-IL binary mixtures. The surface tension of IL-IL binary mixtures is primarily influenced by $-CH_2$ and Tf_2N groups, with temperature also playing a significant role in determining this property. The SHAP analysis results for density, heat capacity, and surface tension can be found in Figures S9-S11.

3. IL-IL mixed solvent tailoring

In 2021, global coal consumption reached 160.1 exajoules (EJ), showing the highest increase (6.3%) since 2014 and accounting for 26.9% of primary energy consumption [57,58]. The coal coking process, known for its energy-intensive, high energy consumption, and high pollution risk nature, suffers from low energy efficiency and inefficient production practices [59,60].

One valuable byproduct of the coking process in coke production is raw coke oven gas, mainly composed of H_2, CH_4, CO, and minor amounts of CO_2, O_2, and N_2. Raw coke oven gas serves as a valuable source of energy and chemical feedstocks, necessitating its effective utilization within the coal industry [61,62]. Unfortunately, raw coke oven gas is primarily used as a fuel source, with certain chemicals, especially H_2, being underutilized due to the high cost and technical challenges associated with treating and purifying the gas for use in other industrial applications. For instance, H_2 holds potential as an energy carrier in various applications, including electricity generation and as a fuel in fuel cells to power vehicles with minimal emissions. However, the process of recovering H_2 from raw coke oven gas is energy-intensive when using cryogenic techniques, and it becomes highly expensive when employing membrane separation methods. An absorption-based approach offers a cost- and energy-effective separation method, but finding a suitable solvent with high selectivity for CH_4/H_2 and CO/H_2 is a challenging task.

As reported, ILs that comprised ammonium cation group and BF_4 anion group, have shown excellent selectivity for separating CO and H_2, but they tend to have higher viscosities, typically exceeding 0.1 Pa·s [63]. To address this issue, introducing a low-viscosity IL that also has good selectivity for CO/H_2 and CH_4/H_2 as a co-solvent may provide a solution. Currently, two primary approaches are used for solvent screening: experimental data and computational models. The experimental-based method is preferable as it often yields more reliable results. However, it is limited in handling mixed solvent design problems due to the lack of experimental data. On the other hand, model-based design methods, such as the computer-aided molecular design (CAMD) technique, offer alternatives for tailoring mixed solvents. The effectiveness of these methods relies on the availability of physical and thermodynamic property models. In addition to the thermodynamic models developed for IL-gas systems in our previous work [64], the machine learning-based models of physical properties developed for IL-IL mixtures in this study enable the use of the computer-aided design method to tailor high-performance IL-IL mixed solvents for H_2 recovery from raw coke oven gas.

3.1 CAILD-Based MINLP problem formulation

In the design problem of IL-IL mixed solvents for H_2 recovery from raw coke oven gas, a single objective optimization is employed. The objective is to maximize the selectivity of CO over H_2 (as described by Equation (10)), while setting the selectivity of CH_4 over H_2 and the solubility of the key absorbed component CO as design constraints. This approach recognizes that the separation of H_2 and CO is generally more challenging than that of H_2 and CH_4. Meanwhile, employing a single objective optimization with less complexity enhances computational efficiency compared to multi-objective optimization methods.

$$\beta_{CO/H2} = \frac{P_{H_2}^{sat}\gamma_{H_2}^{\infty}}{M_{IL}P_{CO}^{sat}\gamma_{CO}^{\infty}} \quad (10)$$

where $\beta_{CO/H2}$ denote the solvent's absorption selectivity of CO over H_2, and M_{IL} represents the molar mass of tailored IL-IL binary solvent. $P_{H_2}^{sat}$ and P_{CO}^{sat} are the vapor pressure of CO and H_2, while $\gamma_{H_2}{}^{\infty}$ and $\gamma_{CO}{}^{\infty}$ are infinite dilution activity coefficients of H_2 and CO, respectively. The molar weight of different ILs varies a lot but their cost is mostly based on the weight consumption. For this reason, the selectivity is calculated based on the mass-based Henry's law constant, a validated and effective index in solvent screening for gas purification [65]. The infinite dilution activity coefficients of the gas in ILs are estimated using the UNIFAC-IL model developed in the previous work [49]. Along with molecular structural constraints, process model constraints, and physical and thermodynamic property constraints, the computer-aided design of IL-IL mixed solvents is formulated as a mixed-integer nonlinear programming (MINLP) problem, as summarized in Equation (11).

$$\max_{z,y} \beta_{CO/H2} = f\left(z, y, T\right) \quad (11)$$

s.t. Molecular structural constraints

Physical property constraints

Thermodynamic property constraints

Process model constraints

where vector z represents discrete variables representing the IL molecules; vector y denotes continuous variables describing the composition of the IL-IL binary mixed solvent and variable T is the system's temperature. As for the gas separation process, T is specified at 298.15K.

$$\sum_{i\in C} c_i^m = 1 \qquad \forall m \in M \tag{12}$$

$$\sum_{j\in A} a_j^m = 1 \qquad \forall m \in M \tag{13}$$

$$\sum_{l\in N} g_l^m - \sum_{i\in C} c_i^m v_i^m = 0 \qquad \forall m \in M \tag{14}$$

$$\sum_{i\in C}\left(2-v_i^m\right)+\sum_{l\in N}\sum_{k\in S}\left(2-v_{k,l}^m\right)g_l^m n_{k,l}^m = 2 \qquad \forall m \in M \tag{15}$$

$$\sum_{k\in S}\left(2-v_{k,l}^m\right)g_l^m n_{k,l}^m = 1 \qquad \forall m \in M, l \in N \tag{16}$$

$$L_k \le \sum_{l\in N}\sum_{k\in S} g_l^m n_{k,l}^m \le U_k \qquad \forall m \in M \tag{17}$$

$$Ll_{k,l} \le \sum_{k\in S} g_l^m n_{k,l}^m \le Ul_{k,l} \qquad \forall m \in M, l \in N \tag{18}$$

$$T_{mp} = T_0 + \sum_{m\in M} x_m\left(\sum_{i\in C} c_i^m \Delta t_i + \sum_{j\in A} a_j^m \Delta t_j + \sum_{l\in N}\sum_{k\in S} g_l^m n_{k,l}^m \Delta t_k\right) \le 298.15(K) \tag{19}$$

$$\eta = W_2 \cdot \left(\frac{2}{1+e^{-2(W_1\cdot p+b_1)}}-1\right)+b_2 \le 100(cP) \tag{20}$$

$$\frac{P_{H_2}^{sat}\gamma_{H_2}^{\infty}}{P_{CH_4}^{sat}\gamma_{CH_4}^{\infty}} > 50 \tag{21}$$

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/aic.18392

$$\gamma_{CO}^{\infty} < 1 \tag{22}$$

$$\sum_{m \in M} x_m = 1 \tag{23}$$

In this optimization problem, numerous feasible IL candidates are generated from a comprehensive combination of cations, anions, and substitute groups. The feasibility and complexity constraints on the structures of both the first and second ILs are represented by Equations (12)-(18), where the sets of cations, anions, substituents, and IL molecules are described as C, A, S and M, respectively. The binary variables c_i^m, a_j^m, x_l^m represent the cation i, anion j, and side chain l within IL molecule m. The integer variables $n_{k,l}^m$ denote the number of substituent of type k in the side chain l within IL molecule m. The vectors of v_i^m and $v_{k,l}^m$ represent the group valence of cation i and substituent k in the side chain l within IL molecule m, respectively. Equation (12) and Equation (13) indicate that each IL molecule contains only one cation and one anion. The octet rule, expressed in Equation (14), ensures the consistency between the free valence of the cation and the number of side chains. Equations 15 and 16 are imposed to avoid more than one covalent bond between any two adjacent groups. In addition, the minimum and maximum numbers of substituents L_k and U_k for the whole cation (Equation (17)), and $Ll_{k,l}$, $Ul_{k,l}$ (Equation (18)) for each side chain l are specified, respectively. These specifications recognize the size and complexity of the generated IL candidates. The detailed information for the building blocks of the IL molecule in this CAILD-based MINLP problem. The constraints on physical properties including melting point temperature (Equation (19)) and viscosity (Equation (20)) are introduced to ensure the generated IL candidates are feasible in practical operations. The melting point of IL is estimated from a GC-based model proposed by Lazzús, while the viscosity of IL-IL binary mixtures is calculated

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/aic.18392

from the ANN-GC model developed in Section 2. In Equation (19), Δt_i, Δt_j and Δt_k represent the melting temperature contributions of the cation, anion, and substituent groups included in both the first and second IL molecules. T_0 is the temperature adjusting parameter. W_1, W_2, b_1 and b_2 in Equation (20) are the optimized ANN-GC model parameters (see Table S6(a)). The input vector, denoted as p, encompasses the structures of the first and second IL molecules, the system's temperature, and the mixture composition (see Table S5 (a)). In addition, a thermodynamic constraint on the selectivity of CH_4/H_2, as expressed by Equation (21), is applied to ensure the feasibility of separating H_2 and CH_4. Meanwhile, a thermodynamic constraint on the infinite dilute activity coefficient of CO, described by Equation (22), is used to ensure the solubility performance of the tailored mixed solvent. Equation (23) gives a process model constraint describing the composition of a binary mixture in terms of the mole fraction of ILs $m1$ and $m2$.

3.2 Solutions and discussions

The model was implemented in GAMS (24.4.6), and it was solved using a deterministic global optimization solver, LINDOGLOBAL. Model details from GAMS solution of the mixed solvent design-based MINLP optimization problem are given in Table S11. The optimal solution was achieved at 30 seconds on an Intel(R) Xeon(R) E5-1620 3.70 GHz PC, demonstrating the high computational efficiency and integrability of the developed ML-based property models in solvent design. Optimization resutls show that a binary mixed solvent combining 57.2 mol% butylethylammonium tetrafluoroborate ($[N_{4,1,0,0}][BF_4]$) and 42.8 mol% 1-methylimidazolium bis(trifluoromethanesulfonyl)amide ($[C_1Im][Tf_2N]$) meets all the design constraints has

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/aic.18392

the maximum $\beta_{CO/H2}$ (= 0.28 g^{-1}). Figure 3 presents the molecular structure and some important properties of the tailored binary mixed solvent at room temperature. In addition, the impact of the composition on the selectivity performance and viscosity property of the mixed solvent is also investigated and presented in Figure 4 (see Table S10). As expected, both $\beta_{CO/H2}$ and viscosity of the mixed solvent decrease when the mole fraction of [C_1Im][Tf_2N] is increasing. This is due to the fact that IL [$N_{4,1,0,0}$][BF_4] has both higher $\beta_{CO/H2}$ and viscosity compared to [C_1Im][Tf_2N]. This can also be used to explain why the optimal is determined at where the viscosity of the tailored solvent reach its design uplimit.

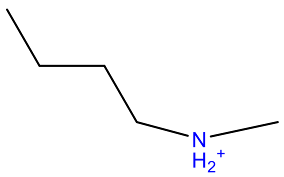
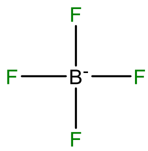
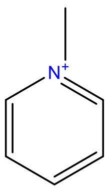
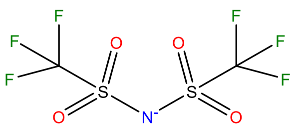
IL (b1)	 [$N_{4,1,0,0}$]	 [BF_4]	[$N_{4,1,0,0}$][BF_4]: 57.2 mol% [C_1Im][Tf_2N]: 42.8 mol%
			$\beta_{CO/H2}$: 0.28 (g^{-1})
			M_{IL} : 254.75 g/mol
IL (b2)	 [C_1Im]	 [Tf_2N]	T_m : 276.7 (K)
			η : 100 (cP)
			$\frac{P_{H_2}^{sat}\gamma_{H_2}^{\infty}}{P_{CH_4}^{sat}\gamma_{CH_4}^{\infty}}$: 57.4

Figure 3. Molecular structure and some important properties of the binary mixed solvent combining 57.2 mol% [$N_{4,1,0,0}$][BF_4] and 42.8 mol% [C_1Im][Tf_2N].

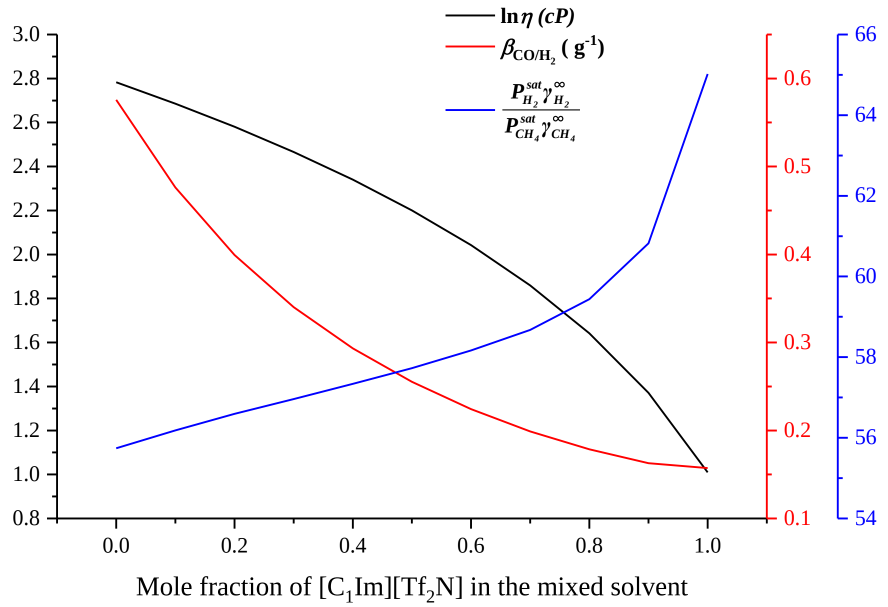


Figure 4. Impact of the composition on the selectivity performance and viscosity property of the binary mixed solvent combining [$N_{4,1,0,0}$][BF_4] and [C_1Im][Tf_2N].

4. Process design and optimization

In this section, a case study on H_2 recovery from raw coke oven gas is presented. The study explores the potential of utilizing the IL-IL mixed solvent discussed in Section 2 and integrating ML-driven models into process design and optimization. We use one of the world's largest coking plants in China as our case study, which has an annual coal coking capacity of 3 million tons with a corresponding production of 12 billion cubic meters of raw coke oven gas. This gas primarily consists of H_2, CO, and CH_4 in a mole ratio of 6:1:3.

The case involves a common gas absorption process, illustrated in Figure 5. The raw coke oven gas is compressed and fed to the absorption column at the bottom, where it contacts the tailored IL-IL binary solvent from the top. The mixed solvent selectively

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/aic.18392

absorbs CH_4 and CO over H_2, resulting in a high-purity H_2 product at the column's top. The CO- and CH_4-enriched solvent collected at the column's bottom is expanded and sent to a flash drum for regeneration. The gas, primarily containing CH_4 and CO, is treated as a byproduct. The regenerated solvent, mixed with a small amount of makeup solvent, is returned to the absorption column via a pump. Expansion valves reduce liquid pressure, and both the pump and compressor operate isentropically. Thermodynamic calculations are performed using the UNIFAC-IL model proposed in our previous work, while physical properties like density, viscosity, surface tension, and heat capacity are calculated using ML-driven models developed in Section 2.

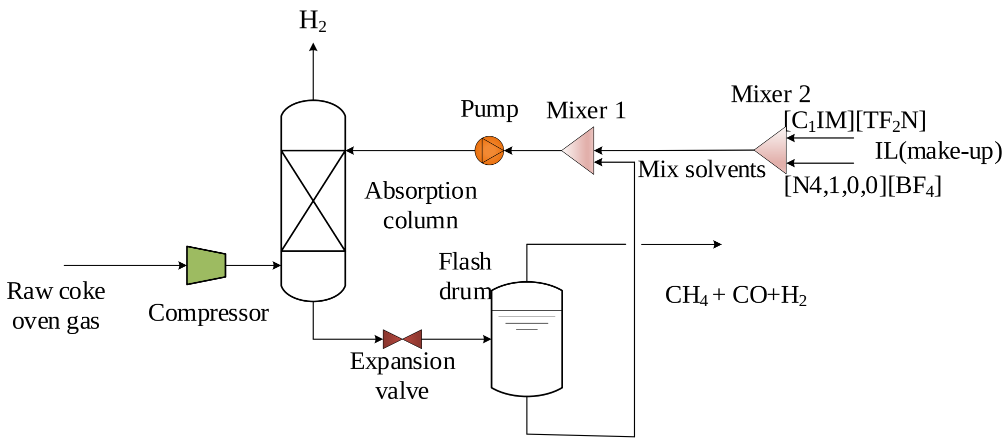


Figure 5. Process flowsheet for recovering H_2 from raw coke oven gas using a binary IL-IL mixed solvent.

As discussed in Section 3, both $\beta_{CO/H2}$ and viscosity of the mixed solvent decrease with the increase of [C_1Im][Tf_2N]. Obviously, a higher $\beta_{CO/H2}$ means a lower solvent input and smaller equipment size demands, and therefore it is able to reduce the cost of the process. In contrast, an increase in the viscosity of the solvent will reduce the process efficiency, especially for the absorption column. For this reason, the composition of the mixed solvent is relaxed and set as a design variable to be optimized, but the viscosity constraint (<100 cP) that is used in the solvent tailoring stage is still activated. In this work, a nonlinear programming (NLP) model coupled with the maximum net present value (NPV), as defined in Equation (23), is employed to optimize the composition, flowrate of the mixed solvent, and equipment and operating variables (*e.g.*, absorption column pressure, flash unit pressure) in the H_2 recovery process from raw coke oven gas. The assumptions and specifications used to derive the detailed formulation of this optimization problem are provided in Table 2.

$$NPV = \sum_{0}^{20} \frac{NCF_n}{(1+r)^n} \qquad (23)$$

where NCF is the non-discounted cash flow for year n, and r is the discount rate.

Table 2: Design specifications and constraints in the process optimization for recovering H_2 from raw coke oven gas

Design specifications and constraints	
Feedstock	The feed flowrate is 16128 Nm^3/h at 5 bar and 298.15 K
Product	The purity of the H_2 product must exceed 95%
Solvent	The sum of the mole fractions of ILs must equal 1
	Solvent viscosity must not exceed 100 cP
	The mole fraction of ILs in the regenerated solvent stream must be higher than 99.8%
Equipment	The height of the absorption column must be less than 50 m
	The diameter of the absorption column must not exceed 5 m

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/aic.18392

Operations	Absorption column pressure should fall within the range of 20 to 80 bar
	Flash unit pressure should fall within the range of 0.01 to 1 bar.

In this work, the performance objective is the Net Present Value (NPV) of the profit over a 20-year lifetime. Calculating NPV involves determining the purchased cost of each unit, which is then used to compute the Annualized Cost of Capital (ACC) for the process. Installation costs are typically higher than the purchase cost and should be factored into ACC. Additionally, costs associated with the operation of the process, including labor, utilities, maintenance, repairs, taxes, and so on, need to be estimated. The dimensions of each unit can be estimated based on flow characteristics and relevant energy duty. The assumptions for the NPV calculation include a discount rate of 10%, a straight-line depreciation method over 7 years, a tax rate of 30%, a financing equity of 100%, and 8000 hours of operation per year. Detailed calculations for ACC, OPEX, and NPV can be found in the Supporting Information I.

The formulated NLP model is implemented in GAMS (version 24.4.6), and LINDOGLOBAL is employed as the DNLP (Discontinuous Nonlinear) solver to solve the model. Model details from GAMS solution of this H_2 recovery process design-based NLP optimization problem is also provided in Table S11. Optimization results, including the composition and flowrate of the mixed solvent, the recovery rate and purity of the H_2 product, the operating pressure, and the size of the main equipment *(i.e.*, adsorption column and flash unit), as well as the Net Present Value (NPV) of profit over a 20-year lifetime, are presented in Table 3. Firstly, the best solution is found within 27 seconds on an Intel(R) Xeon(R) E5-1620 3.70 GHz PC, indicating the high computational efficiency and integrability of our developed ML-based property models in process optimization. Secondly, the column efficiency is only 0.29, which implies that viscosity has a significant impact on equipment operations. Nonetheless, the optimization results related to the composition of the mixed solvent show that the solvent's selectivity performance has a higher impact on economic performance than its viscosity in the process. This indicates that the integrated design for the studied process will have very limited improvement compared to the two-step design strategy used in this work. Furthermore, it is important to point out that the binary mixed solvent provides a promising alternative for hydrogen recovery from raw coke oven gas. Even for this basic flowsheet, the process allows us to obtain an H_2 product with 95% purity.

Table 3. Optimization results for the process design to recover H_2 from raw coke oven gas with [$N_{4,1,0,0}$][BF_4]- [C_1Im][Tf_2N] mixed solvent.

Composition of mixed solvent	[$N_{4,1,0,0}$][BF_4]: 57.2 mol%
	[C_1Im][Tf_2N]: 42.8 mol%
Flowrate of mixed solvent (mol/s)	5394.4
Absorption column pressure (bar)	39.6
Column height (m)	16.3
Column cross section (m^2)	1.8
Colum efficiency	0.29
Flash tank volume (m^3)	600.3
Flash pressure (bar)	0.57
% H_2 in product stream	95.0

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/aic.18392

% H_2 recovery	86.8
% H_2 in outlet raw coke oven gas	16.8
NPV (MM$)	812.7

In terms of the process design itself, this work considers a simple process flowsheet, which is likely to offer less flexibility in process operations but also comes with lower costs. A more complex flowsheet would provide greater flexibility in process configurations and has the potential to enhance process performance. For instance, the addition of cooling utilities to the process could decouple temperature and pressure in the absorber, enabling operation at high pressure-low temperature states. Using multiple columns in series, each at different conditions, would allow each absorption column to operate at its optimal conditions. However, any increase in process complexity must be accompanied by a sufficient improvement in process performance to offset the additional capital and operating costs incurred. On the other hand, many industrial processes require hydrogen with purity levels ranging from 95% to 99.9%, necessitating adjustments in certain cases. It's important to note that increasing the purity level of hydrogen may lead to a higher quantity of H_2 being lost to the waste gas stream, resulting in a reduced NPV. Therefore, a trade-off between the purity level of the H_2 product and the overall economic performance of the process should be carefully considered before implementation.

5. Conclusion

In this work, predictive models that combine the GC method with three well-known machine learning algorithms (*i.e.*, ANN, XGBoost, and LightGBM) are developed for

Accepted Manuscript
Version of Record at: https://doi.org/10.1002/aic.18392

the viscosity, density, heat capacity, and surface tension of IL-IL binary mixtures. A comprehensive database consisting of 13,000 experimental data points from 136 different IL-IL mixture systems, encompassing 9 cation skeleton groups and 29 anion groups, is initially used to train and evaluate the proposed ML-based GC models. The modeling results demonstrate that the GC model with all three machine learning algorithms, can produce reliable predictions. In particular, the ANN-based GC model with optimized neurons in the hidden layer demonstrates the best predictive performance, achieving an R^2 value exceeding 0.999 for all studied properties. These modeling results underscore the potent capabilities of ANN in predicting the properties of IL-IL binary mixtures. Furthermore, SHAP analysis reveals that the viscosity of these mixtures is primarily affected by temperature and molar fraction, while the presence of functional groups BF_4 and $-CH_2$ has the most significant influence on the density of IL-IL binary mixtures. On the other hand, heat capacity is primarily influenced by the functional group BF_4, and surface tension. These insights from the SHAP analysis provide valuable information for understanding the key factors that influence the properties of IL-IL mixtures.

The case study of H_2 recovery from raw coke oven gas demonstrates the high computational efficiency and integrability of our developed ANN-based GC models in computer-aided mixed solvent design, process design, and optimization. Optimization results of the formulated CAILD-based MINLP problem show that a binary mixed solvent, combining 57.2 mol% $[N_{4,2,0,0}][BF_4]$ and 42.8 mol% $[C_1Im][Tf_2N]$, offers the best separation performance for recovering H_2 from raw coke oven gas. Furthermore, optimization results of the process design-based NLP problem indicate that the IL-IL binary solvent provides a promising alternative for hydrogen recovery from raw coke oven gas. The tailored solvent allows us to obtain an H_2 product with 95% purity, even though a simple process. Additionally, solvent viscosity significantly impacts equipment operations, but the solvent's selectivity performance has a higher impact on economic performance than viscosity. In the future, a more complex flowsheet can be considered, as it could provide greater flexibility in process configurations and has the potential to enhance process performance. Overall, this study contributes to advancing the modeling, design, and optimization of IL-IL binary mixtures and their applications in various fields.

Acknowledgements

The authors gratefully acknowledge the financial support from the University of Delaware, the Technical University of Denmark, and the National Natural Science Foundation of China (21706198, 22208253).

Supporting Information

Supporting information associated with this article can be found in the online version.

References

[1] Liu K, Wang Z, Shi L, *et al.* Ionic liquids for high performance lithium metal batteries. Journal of Energy Chemistry. 2021;59:320-333.
[2] Kerche E F, Demétrio da Silva V, da Cruz J A, *et al.* Aramid pulp modified with imidazolium ionic liquids as a potential replacement for carbon black in fluorinated rubber composites. Polymer Composites. 2023;44(5):2887-2897.
[3] Silva A B P, Jesus A R, Agostinho D A S, *et al.* Using dicationic ionic liquids to upgrade the cytotoxicity and solubility of poorly water-soluble drugs. Journal of Ionic Liquids. 2023;3(1): 100052-100060.
[4] Greaves T L, Drummond C J. Protic ionic liquids: properties and applications. Chemical

Reviews. 2008;108(1):206-237.
[5] Rooney D, Jacquemin J, Gardas R. Thermophysical properties of ionic liquids. Ionic liquids. 2010;290:185-212.
[6] Chen J H, Yang X Z, Chen G Z, *et al.* Discussion on modeling method of molecular property prediction model taking ionic liquid density as an example. CIESC Journal. 2023;74(2):630-641. (In Chinese)
[7] Farahani N, Gharagheizi F, Mirkhani S A, *et al.* A simple correlation for prediction of heat capacities of ionic liquids. Fluid Phase Equilibria. 2013;337:73-82.
[8] Dai Z, Chen Y, Liu C, *et al.* Prediction and verification of heat capacities for pure ionic liquids. Chinese Journal of Chemical Engineering. 2021;31:169-176.
[9] Shojaeian A. New experimental and modeling based on the N-Wilson-NRF equation for surface tension of aqueous alkanolamine binary mixtures. Journal of Molecular Liquids. 2018;254: 26-33.
[10] Chen Y, Kontogeorgis G M, Woodley J M. Group contribution based estimation method for properties of ionic liquids. Industrial & Engineering Chemistry Research. 2019;58(10):4277-4292.
[11] Nekoeinia M, Yousefinejad S, Abdollahi-Dezaki A. Prediction of ETN polarity scale of ionic liquids using a QSPR approach. Industrial & Engineering Chemistry Research. 2015;54(50): 12682-12689.
[12] Kahlen J, Masuch K, Leonhard K. Modelling cellulose solubilities in ionic liquids using COSMO-RS. Green Chemistry. 2010;12(12):2172-2181.
[13] Huang Y, Dong H, Zhang X, *et al.* A new fragment contribution-corresponding states method for physicochemical properties prediction of ionic liquids. AIChE Journal. 2013;59(4):1348-1359.
[14] Gani R. Group contribution-based property estimation methods: advances and perspectives. Current Opinion in Chemical Engineering. 2019;23:184-196.
[15] Alshehri A S, Gani R, You F. Deep learning and knowledge-based methods for computer-aided molecular design—toward a unified approach: State-of-the-art and future directions. Computers & Chemical Engineering. 2020;141:107005-107062.
[16] Tariq M, Freire M G, Saramago B, *et al.* Surface tension of ionic liquids and ionic liquid solutions. Chemical Society Reviews. 2012;41(2):829-868.
[17] Chen Y, Koumaditi E, Gani R, *et al*. Computer-aided design of ionic liquids for hybrid process schemes. Computers & Chemical Engineering. 2019;130:106556-106588.
[18] Zhao Y, Zhang X, Deng L, *et al.* Prediction of viscosity of imidazolium-based ionic liquids using MLR and SVM algorithms. Computers & Chemical Engineering. 2016;92:37-42.
[19] Song Z, Shi H, Zhang X, *et al.* Prediction of CO_2 solubility in ionic liquids using machine learning methods. Chemical Engineering Science. 2020;223:115752-115758.
[20] Wang Z, Song Z, Zhou T. Machine learning for ionic liquid toxicity prediction. Processes. 2020;9(1):65-74.
[21] Chen Y, Peng B, Kontogeorgis G M, *et al.* Machine learning for the prediction of viscosity of ionic liquid–water mixtures. Journal of Molecular Liquids. 2022;350:118546-118557.
[22] Fu Y, Chen Y, Zhang C, *et al*. Prediction surface tension of ionic liquid–water mixtures using

a hybrid group contribution and artificial neural network method. Fluid Phase Equilibria. 2022;563: 113571-113581.

[23] Chen Y, Liang X, Woodley J M, *et al*. Modelling study on phase equilibria behavior of ionic liquid-based aqueous biphasic systems. Chemical Engineering Science. 2022;247:116904-116916.

[24] Chen Y, Liang X, Kontogeorgis G M. Artificial neural network modeling on the polymer-electrolyte aqueous two-phase systems involving biomolecules. Separation and Purification Technology. 2023;306:122624-122639.

[25] Chen Y, Meng X, Cai Y, *et al.* Optimal aqueous biphasic systems design for the recovery of ionic liquids. Industrial & Engineering Chemistry Research. 2021;60(43):15730-15740.

[26] Gebreyesus Y, Dalton D, Nixon S, *et al*. Machine learning for data center optimizations: Feature selection using shapley additive explanation (SHAP). Future Internet. 2023;15(3):88-104.

[27] Hiraga Y, Koyama K, Sato Y, *et al.* High pressure densities for mixed ionic liquids having different functionalities: 1-butyl-3-methylimidazolium chloride and 1-butyl-3-methylimidazolium bis (trifluoromethylsulfonyl) imide. The Journal of Chemical Thermodynamics. 2017;108:7-17.

[28] Martins M A R, Sharma G, Pinho S P, *et al.* Selection and characterization of non-ideal ionic liquids mixtures to be used in CO_2 capture. Fluid Phase Equilibria. 2020;518:112621-112633.

[29] Páez M S, Vega Y A, Romero C M. Effect of temperature on the viscosities and the volumetric properties of the binary mixtures of the ionic liquids [bmim][PF_6] and [bmim][CF_3SO_3]. Journal of Molecular Liquids. 2017;243:78-84.

[30] Song D, Chen J. Density and viscosity data for mixtures of ionic liquids with a common anion. Journal of Chemical & Engineering Data. 2014;59(2):257-262.

[31] Song D, Chen J. Densities and viscosities for ionic liquids mixtures containing [eOHmim][BF_4], [bmim][BF_4] and [bpy][BF_4]. The Journal of Chemical Thermodynamics. 2014;77:137-143.

[32] Almeida H F D, Canongia Lopes J N, Rebelo L P N, *et al.* Densities and viscosities of mixtures of two ionic liquids containing a common cation. Journal of Chemical & Engineering Data. 2016;61(8):2828-2843.

[33] Zheng Y, Zheng Y, Wang Q, *et al.* Density and viscosity of binary mixtures of 1-ethyl-3-methylimidazolium heptachlorodialuminate and tetrachloroaluminate ionic liquids. Journal of Chemical & Engineering Data. 2017;62(11):4006-4014.

[34] Vieira N S M, Vázquez-Fernández I, Araújo J M M, *et al.* Physicochemical characterization of ionic liquid binary mixtures containing 1-butyl-3-methylimidazolium as the common cation. Journal of Chemical & Engineering Data. 2019;64(11):4891-4903.

[35] Gouveia A S L, Tomé L C, Marrucho I M. Density, viscosity, and refractive index of ionic liquid mixtures containing cyano and amino acid-based anions. Journal of Chemical & Engineering Data. 2016;61(1):83-93.

[36] Ning H, Hou M Q, Mei Q Q, *et al*. The physicochemical properties of some imidazolium-based ionic liquids and their binary mixtures. Science China-Chemistry. 2012;55(8): 1509-1518.

[37] Kanakubo M, Makino T, Umecky T. CO_2 solubility in and physical properties for ionic liquid mixtures of 1-butyl-3-methylimidazolium acetate and 1-butyl-3-methylimidazolium bis

(trifluoromethanesulfonyl) amide. Journal of Molecular Liquids. 2016;217:112-119.

[38] Andriyko Y O, Reischl W, Nauer G E. Trialkyl-substituted imidazolium-based ionic liquids for electrochemical applications: Basic physicochemical properties. Journal of Chemical & Engineering Data. 2009;54(3):855-860.

[39] Shaahmadi F, Hashemi Shahraki B, Farhadi A. The solubility of carbon dioxide and density for binary mixtures of 1-Butyl-3-methylimidazolium acetate and 1-Butyl-3-methylimidazolium tetrafluoroborate. Journal of Chemical & Engineering Data. 2019;64(2):584-593.

[40] Yang J Z, Jin Y, Xu W G, *et al.* Studies on mixture of ionic liquid $EMIGaCl_4$ and EMIC. Fluid Phase Equilibria. 2005;227(1):41-46.

[41] Navarro P, Larriba M, García J, *et al*. Thermal stability, specific heats, and surface tensions of ([emim][DCA]+[4empy][Tf_2N]) ionic liquid mixtures. The Journal of Chemical Thermodynamics. 2014;76:152-160.

[42] Larriba M, Navarro P, Beigbeder J B, *et al.* Mixing and decomposition behavior of {[4bmpy][Tf2N]+[emim][EtSO4]} and {[4bmpy][Tf2N]+[emim][TFES]} ionic liquid mixtures. The Journal of Chemical Thermodynamics. 2015;82:58-75.

[43] Ge R, Hardacre C, Jacquemin J, *et al.* Heat capacities of ionic liquids as a function of temperature at 0.1 MPa. Measurement and prediction. Journal of Chemical & Engineering Data. 2008;53(9):2148-2153.

[44] Bruce D W, Cabry C P, Canongia Lopes J N, *et al.* Nanosegregation and structuring in the bulk and at the surface of ionic-liquid mixtures. The Journal of Physical Chemistry B. 2017;121(24):6002-6020.

[45] Navia P, Troncoso J, Romaní L. Viscosities for ionic liquid binary mixtures with a common ion. Journal of Solution Chemistry. 2008;37:677-688.

[46] Fillion J J, Brennecke J F. Viscosity of ionic liquid–ionic liquid mixtures. Journal of Chemical & Engineering Data. 2017;62(6):1884-1901.

[47] Larriba M, Garcia S, Navarro P, *et al*. Physical properties of N-butylpyridinium tetrafluoroborate and N-butylpyridinium bis (trifluoromethylsulfonyl) imide binary ionic liquid mixtures. Journal of Chemical & Engineering Data. 2012;57(4):1318-1325.

[48] Lepre L F, Costa Gomes M, Padua A A H, *et al.* On the regular behavior of a binary mixture of ionic liquids. The Journal of Physical Chemistry B. 2019;123(30):6579-6587.

[49] Afzal W, Liu X, Prausnitz J M. High solubilities of carbon dioxide in tetraalkyl phosphonium-based ionic liquids and the effect of diluents on viscosity and solubility. Journal of Chemical & Engineering Data. 2014;59(4):954-960.

[50] Annat G, Forsyth M, MacFarlane D R. Ionic liquid mixtures-variations in physical properties and their origins in molecular structure. The Journal of Physical Chemistry B. 2012;116(28):8251-8258.

[51] Koller T M, Lenahan F D, Schmidt P S, et al. Surface tension and viscosity of binary mixtures of the fluorinated and non-fluorinated ionic liquids [PFBMIm][PF_6] and [C_4C1Im][PF_6] by the pendant drop method and surface light scattering. International Journal of Thermophysics.

2020;41:1-24.

[52] Oliveira M B, Domínguez-Pérez M, Freire M G, *et al.* Surface tension of binary mixtures of 1-alkyl-3-methylimidazolium bis (trifluoromethylsulfonyl) imide ionic liquids: experimental measurements and soft-SAFT modeling. The Journal of Physical Chemistry B. 2012;116(40):12133-12141.

[53] Oliveira M B, Domínguez-Pérez M, Cabeza O, *et al.* Surface tensions of binary mixtures of ionic liquids with bis (trifluoromethylsulfonyl) imide as the common anion. The Journal of Chemical Thermodynamics. 2013;64:22-27.

[54] Ke G, Meng Q, Finley T, *et al*. Lightgbm: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems. 2017;2:30-38.

[55] Lundberg S M, Lee S I. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems. 2017;3:30-39.

[56] Naser J, Mjalli F S, Gano Z S. Molar heat capacity of selected type III deep eutectic solvents. Journal of Chemical & Engineering Data. 2016;61(4):1608-1615.

[57] BP Statistical review of world energy 2022. London: BP; June 2022.

[58] Gonçalves A, Puna JF, Guerra L, *et al.* Towards the development of syngas/ biomethane electrolytic production, using liquefied biomass and heterogeneous catalyst. Energies 2019;12(19):3787-3808.

[59] Zhao Z, Wang R, Ge L, *et al.* Energy utilization of coal-coking wastes via coal slurry preparation: the characteristics of slurrying, combustion, and pollutant emission. Energy. 2019;168:609-618.

[60] Park JK, Young LS, Kim J, *et al*. Efficiency, economic, energy, and safety (3ES) analyses on different configurations of MDEA absorption process for coke oven gas desulfurization. Chem Eng J Adv. 2022;10:100281-100292.

[61] Shen Y, Wang M, Wu Y, *et al*. Role of gas coal in directional regulation of sulfur during coal-blending coking of high organic-sulfur coking coal. Energy Fuels. 2020;34(3):2757-2764

[62] Matyjaszek M, Wodarski K, Krzemien A, *et al*. Coking coal mining investment: boosting European Union's raw materials initiative. Resour Pol. 2018;57:88-97.

[63] Huo M, Peng XW, Zhao J, *et al.* COSMO-RS based solvent screening and H_2/CO separation experiments for CO absorption by ionic liquids. CIESC Journal. 2022;73(12):5305-5313.

[64] Chen Y, Liu X, Woodley J M, *et al.* Gas solubility in ionic liquids: UNIFAC-IL model extension. Industrial & Engineering Chemistry Research. 2020;59(38):16805-16821.

[65] Liu X, Chen Y, Zeng S, *et al.* Structure optimization of tailored ionic liquids and process simulation for shale gas separation. AIChE Journal. 2020;66(2):16794-16833.