

Article

XGBoost and Artificial Neural Networks as Surrogate Models for Vapor–Liquid Equilibrium in PC-SAFT

Yiwen Pang *, Zhongwei Ding * and Qunsheng Li

Department of Chemical Engineering, Beijing University of Chemical Technology, Beijing 100029, China

* Correspondence: 2021410008@buct.edu.cn (Y.P.); dingzw@mail.buct.edu.cn (Z.D.)

Abstract

Phase equilibrium calculations are crucial in chemical engineering design and optimization processes. The PC-SAFT equation of state (EoS) can precisely calculate phase equilibrium, but is relatively complex and computationally intensive. Surrogate models are mathematically simple models that map or regress the input–output relationships of more complex, computationally demanding models. This work employs XGBoost and a hybrid XGBoost-artificial neural networks (XGBoost-ANN) model as surrogate models to replace PC-SAFT EoS calculations for the vapor–liquid equilibrium (VLE) of binary associating systems. This work investigates the VLE of five binary associating systems using data generated by the PC-SAFT EoS. The surrogate models take temperature, pressure, liquid phase mole fractions, and the PC-SAFT parameters for binary associating systems as inputs, and predict the vapor phase mole fractions. Both surrogate models significantly reduce the computational time for calculating VLE data compared to the PC-SAFT EoS, while achieving good prediction results.

Keywords: surrogate models; XGBoost; XGBoost-ANN; PC-SAFT

Academic Editor: Chi-Min Shu

Received: 15 October 2025

Revised: 22 November 2025

Accepted: 27 November 2025

Published: 4 December 2025

Citation: Pang, Y.; Ding, Z.; Li, Q. XGBoost and Artificial Neural Networks as Surrogate Models for Vapor–Liquid Equilibrium in PC-SAFT. *Processes* **2025**, *13*, 3918. <https://doi.org/10.3390/pr13123918>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The accuracy of phase equilibrium calculations using equations of state (EoS) is crucial for designing and optimizing processes in the process industries. Gross and Sadowski [1] proposed the perturbed-chain statistical associating fluid theory (PC-SAFT), which accurately predicts the phase equilibrium distribution of substances. However, the mathematical formulation of PC-SAFT is significantly more complex than that of the cubic EoSs commonly used in petrochemical processes, such as Soave–Redlich–Kwong (SRK) [2] and Peng–Robinson (PR) [3].

Surrogate models are mathematically simplified representations that map or regress the input–output relationships of more complex, computationally intensive models [4]. Surrogate models can be evaluated more rapidly than rigorous thermodynamic models, yielding a simplified thermodynamic model; however, this approach is applicable only within a limited physical range [5]. In process engineering, surrogate models are widely employed to simplify EoSs and predict thermodynamic parameters [6–9]. Sun et al. [10] employed graph neural networks (GNNs) to predict vapor–liquid phase equilibrium in binary mixtures (e.g., acid + water, water + ethylene glycol), achieving predictive performance comparable to that of NRTL and Wilson with their developed surrogate model. Ihunde and Olorode [11] employed Physics-Informed Neural Networks (PINN), which are neural networks trained to solve supervised learning tasks while adhering to physical constraints described by general nonlinear partial differential equations [12],

and standard deep learning models as surrogate models for the PR EoS, to predict the flash of the methane–propane–tetradecane system. They found that the PINN model achieved a root mean square error (RMSE) approximately two times lower than the standard deep learning model. Recent advances in PINN have largely focused on overcoming the loss-imbalance problem through sophisticated self-adaptive weighting strategies, such as maximum-likelihood-based Gaussian modelling of individual loss terms [13], dynamic minimax-trained weights that act as learnable attention-like masks [14], and the deliberate use of radial basis function activations to faithfully represent nonlocal integral operators in inverse problems [15]. These techniques have dramatically improved convergence, accuracy, and physical consistency for challenging equations, including Navier–Stokes, Allen–Cahn, and peridynamic models. Hakim et al. [16] investigated the liquid–liquid equilibrium of ternary systems composed of aliphatic compounds + aromatic compounds + ionic liquids using a hybrid artificial neural network (ANN) approach that combined backpropagation (BP) and the Group Method of Data Handling (GMDH). The data obtained from the trained neural network were compared with the calculated values from the NRTL and UNIQUAC. The results indicated that the predictions from the BP neural network were more accurate. Despite the promising results achieved by neural networks or machine learning methods in phase equilibrium calculations, most existing studies on surrogate models for phase equilibrium have primarily employed cubic EoSs and activity coefficient models [8,17–22], with relatively fewer investigations into SAFT-type EoSs.

Tree boosting is a widely applied and highly effective machine learning method. XGBoost is a scalable machine learning system for tree boosting, frequently integrated with neural networks [23]. In XGBoost, a novel sparsity-aware algorithm is employed to handle sparse data, and a built-in weighted quantile sketch is utilized for approximate learning. These algorithmic improvements have enabled its widespread adoption by data scientists. Despite the growing number of machine learning-based surrogate models for vapour–liquid equilibrium calculations, the vast majority of reported studies have focused on activity coefficient models (NRTL, UNIQUAC, etc.) and cubic EoS. Applications to advanced molecular-based equations such as SAFT-family models remain extremely scarce, primarily because of their considerably higher computational cost and complex residual thermodynamics. The present work fills this gap by developing highly efficient XGBoost- and XGBoost-ANN-based surrogate models trained exclusively on PC-SAFT calculated data. This work investigated the isotherm VLE of five binary associating systems (water + methanol, water + ethanol, water + 1-propanol, water + 2-propanol, and water + 1-butanol) using XGBoost and a hybrid XGBoost-ANN approach (hereafter referred to as XGBoost-ANN). Surrogate models were developed with temperature (T), pressure (P), the mole fraction of the liquid phase (x_1), and the PC-SAFT parameters for binary associating systems (m_i , σ_i , ε_i/k_B , $\kappa^{A_iB_i}$, $\varepsilon^{A_iB_i}/k_B$, N_{A_i} , N_{B_i}) as inputs, and the mole fraction of the vapor phase as the output. The output is represented by y_1 . The data used for training and testing the PC-SAFT surrogate models were calculated using PC-SAFT and verified for thermodynamic consistency (comprising a total of 265 data points). Additional details regarding the data are provided in the Supplementary Materials. Furthermore, this work imposed the physical constraint that the output values should lie within the range $[0, 1]$ and customized a loss function for the neural network training in XGBoost-ANN to incorporate this constraint.

2. Method

2.1. PC-SAFT

For the thermodynamic phase equilibrium of two phases (I and II),

$$\mu_i^I = \mu_i^{II} (\forall i = 1 \dots n_c) \quad (1)$$

$$\hat{f}_i^I = \hat{f}_i^{II} (\forall i = 1 \dots n_c) \quad (2)$$

In Equations (1) and (2), n_c is the number of components, μ_i represents the chemical potential of component i , and $\hat{f}_i$ denotes the fugacity of the i component. The phase equilibrium relationship expressed in terms of fugacity coefficients is given by

$$y_i \varphi_i^I = x_i \varphi_i^{II} \quad (3)$$

In Equation (3), φ_i is the fugacity coefficient of component i . For the binary associating systems investigated in this work, the residual Helmholtz energy ($\tilde{a}^{res}$) of the PC-SAFT EoS [1] can be expressed as

$$\tilde{a}^{res} = \tilde{a}^{hc} + \tilde{a}^{disp} + \tilde{a}^{assoc} \quad (4)$$

In Equation (4), $\tilde{a}^{hc}$ represents the hard-chain reference contribution, $\tilde{a}^{disp}$ is the dispersion contribution. $\tilde{a}^{assoc}$ denotes the association contribution.

In the PC-SAFT EoS, the fugacity coefficient is expressed as

$$\ln \varphi_i = \frac{\mu_i^{res}}{k_B T} - \ln Z \quad (5)$$

$$\frac{\mu_i^{res}}{k_B T} = \tilde{a}^{res} + (Z - 1) + \left(\frac{\partial \tilde{a}^{res}}{\partial x_i} \right)_{T,V} - \sum_j^{n_c} \left[x_j \left(\frac{\partial \tilde{a}^{res}}{\partial x_j} \right)_{T,V} \right] \quad (6)$$

In Equations (5) and (6), Z is the compression factor,

$$Z = 1 + \eta \left(\frac{\partial \tilde{a}^{res}}{\partial \eta} \right)_{T,x_i} \quad (7)$$

In Equation (7), η represents the packing fraction. This work utilized the FeOs library [24] in Python 3.10.11 to compute the VLE of binary associating systems using the PC-SAFT EoS. The specific equations of the PC-SAFT EoS and the binary interaction parameters (k_{ij}) regressed for each dataset in this work are provided in the Supplementary Materials. In this work, water and alcohols in the considered binary associating systems are modeled using the 2B association scheme, where all site types (A and B) are equally likely to participate in hydrogen bonding [25]. The number of A sites is set to 1 (i.e., $N_A = 1$) and the number of B sites to 1 (i.e., $N_B = 1$), with the A and B sites illustrated in Figure 1.

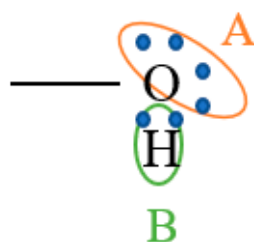


Figure 1. Schematic diagram of the 2B association scheme for water and alcohols.

2.2. XGBoost

2.2.1. Tree Ensemble Model

For regression problems, each regression tree assigns a continuous score to every leaf, with w_i representing by the score associated with the i th leaf. The tree structure is illustrated in Figure 2.

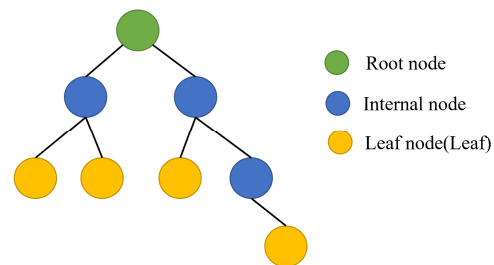


Figure 2. Tree structure.

Figure 3 illustrates the flowchart of the tree ensemble model for regression problems. Initially, training data are input into each tree. The regression trees then assign the corresponding w_i values to the combiner. The combiner performs a weighted sum or average of the w_i values to produce the final output. The regularized objective function of the tree ensemble model is defined as

$$L = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (8)$$

where

$$\Omega(f) = \gamma \text{Leaves}N + \frac{1}{2} \lambda \|w\|^2 \quad (9)$$

In Equation (8), y_i and $\hat{y}_i$ represent the true and predicted values, respectively. l is the differentiable convex loss function quantifies the discrepancy between the $\hat{y}_i$ and y_i . Ω is used to penalize the complexity of the tree ensemble model. In Equation (9), $\text{Leaves}N$ is the number of leaves in the tree. γ and λ are penalization parameters.

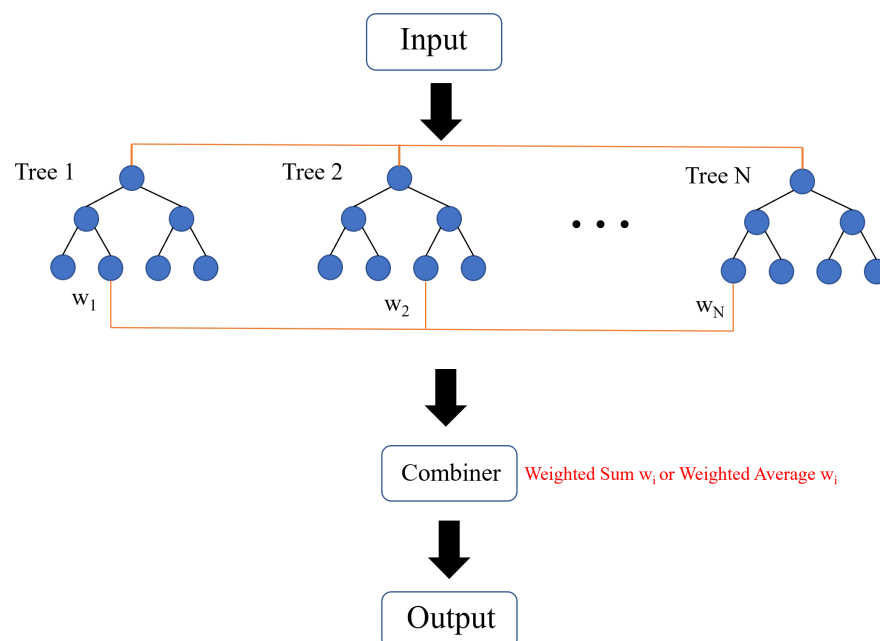


Figure 3. Schematic diagram of the tree ensemble model.

2.2.2. Gradient Tree Boosting

The objective function in Equation (8) cannot be optimized using traditional optimization approaches in Euclidean space. In XGBoost, an iterative objective function based on Equation (8) is employed.

$$L^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \quad (10)$$

In Equation (10), $\hat{y}_i^{(t-1)}$ is the predicted value of the i th instance at the t th iteration. f_t represents an additional perturbation function introduced. To facilitate rapid optimization of the objective function, a second-order approximation is applied to the first term in Equation (10).

$$L^{(t)} = \sum_{i=1}^n \left[l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \quad (11)$$

where

$$g_i = \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)}) \quad (12)$$

$$h_i = \partial_{\hat{y}_i^{(t-1)}}^2 l(y_i, \hat{y}_i^{(t-1)}) \quad (13)$$

By further simplifying the constant term $l(y_i, \hat{y}_i^{(t-1)})$, the simplified objective function in XGBoost can be obtained.

$$L^{(t)} = \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \quad (14)$$

If f_t is set to w_j , $I_j = \{i | q(x_i) = j\}$ is the instance set of leaf j , Equation (14) can be rewritten as

$$L^{(t)} = \sum_{j=1}^{LeavesN} \left[\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma LeavesN \quad (15)$$

replace w_j in Equation (15) with

$$w_j^* = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad (16)$$

Hence, Equation (15) can be rewritten as

$$L^{(t)} = - \frac{1}{2} \sum_{j=1}^{LeavesN} \frac{\left(\sum_{i \in I_j} g_i \right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma LeavesN \quad (17)$$

The schematic diagram of the XGBoost algorithm is presented in Figure 4.

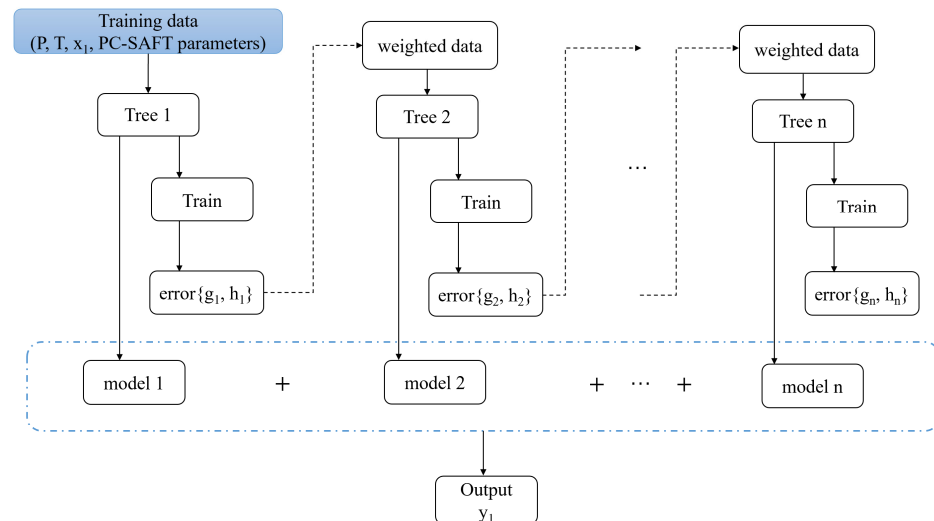


Figure 4. Schematic diagram of XGBoost in this work. The PC-SAFT parameters include m_i , σ_i , ε_i/k_B , $\kappa^{A_i B_i}$, $e^{A_i B_i}/k_B$, N_{A_i} , N_{B_i} .

2.3. Artificial Neural Networks

In this work, we constructed a custom loss function utilizing the ReLU function to address the scenario where the training data exceeds the physical constraint range of $[0, 1]$. The mathematical expression of the ReLU function is presented in Equation (18), while Figure 5 illustrates the graphical representation of the ReLU function.

$$\text{ReLU}(x) = \max(0, x) \quad (18)$$

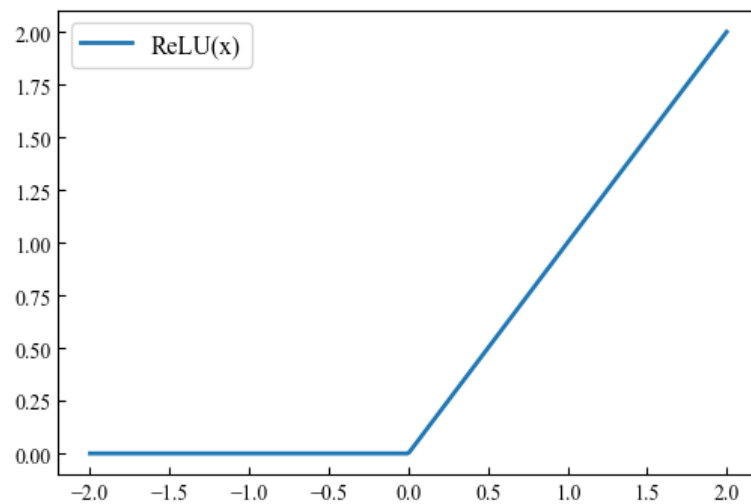


Figure 5. ReLU function.

In this work, the custom loss function is called ConstrainedMSE,

$$\text{ConstrainedMSE} = \text{MSE} + \mu \sum_{i=1}^N [\text{ReLU}(-\hat{y}_i) + \text{ReLU}(\hat{y}_i - 1)] \quad (19)$$

In Equation (19), N is the number of experimental data. μ represents the penalty weight. MSE is the mean squared error,

$$\text{MSE} = \frac{1}{N} \sum_i^N (y_i - \hat{y}_i)^2 \quad (20)$$

In this work, the ANN was constructed using PyTorch [26], with the Adam optimizer employed and a learning rate set at 1×10^{-4} . L2 regularization was implemented with a weight decay coefficient of 1×10^{-5} .

2.4. XGBoost-ANN

The other surrogate model in this work is XGBoost-ANN. The XGBoost-ANN first trains the input data (T, P, x_1 , PC-SAFT parameters) using XGBoost, and then employs the output from XGBoost as the input for ANN training to obtain the predicted values (y_1). The algorithmic schematic of XGBoost-ANN is shown in Figure 6.

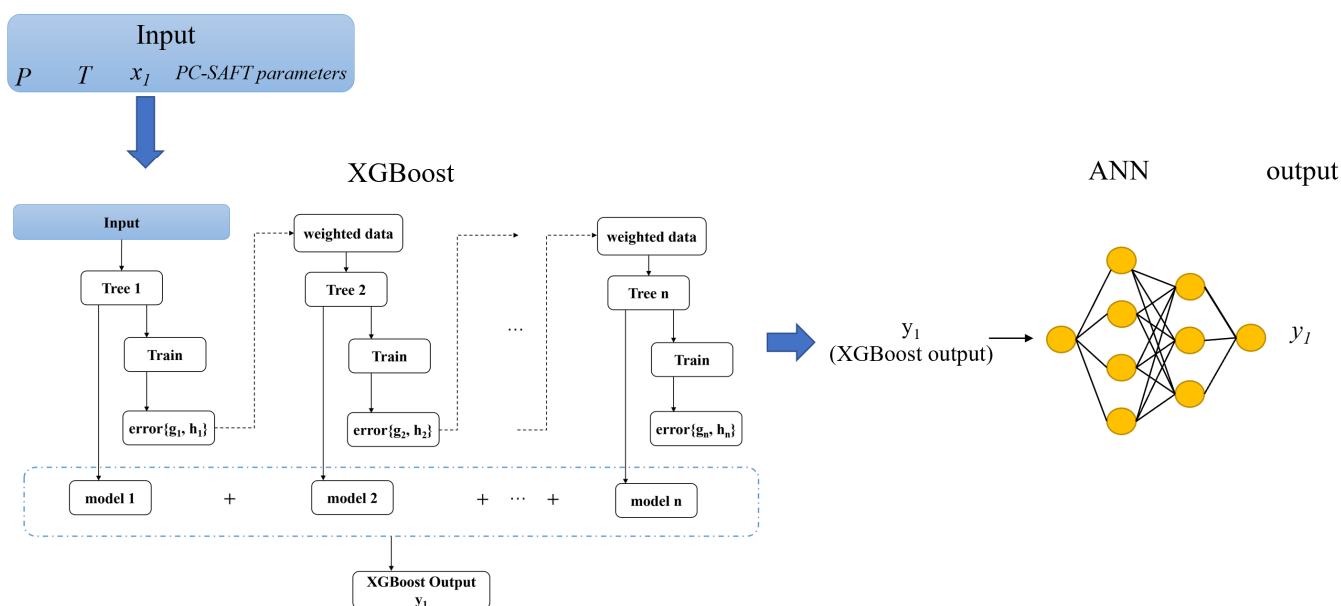


Figure 6. Schematic diagram of XGBoost-ANN. The PC-SAFT parameters include $m_i, \sigma_i, \varepsilon_i/k_B, \kappa^{A_i B_i}, \varepsilon^{A_i B_i}/k_B, N_{A_i}, N_{B_i}$.

Integrating ANN with ConstrainedMSE into XGBoost in the XGBoost-ANN ensures that the final trained results adhere to the physically constrained range of $[0, 1]$. Alternative and potentially more principled approaches to enforce the $[0, 1]$ bounds include employing a sigmoid activation at the output layer or predicting logit-transformed targets. These strategies, which eliminate the need for an explicit penalty hyperparameter μ , will be systematically investigated in our future work.

2.5. Thermodynamic Consistency

Thermodynamic data measured under identical conditions may vary among different researchers. Conducting thermodynamic consistency tests on experimental values is essential [27]. At constant temperature, the Gibbs–Duhem equation for a binary system can be expressed as [28]

$$\left(\frac{V^E}{RT}\right)dP = x_1 d(\ln \gamma_1) + x_2 d(\ln \gamma_2) \quad (21)$$

In Equation (21), V^E represents the reduced molar volume, γ_i denotes the activity coefficient of component i , R is the ideal gas constant, and x_i is the mole fraction of component i . This work employed the PC-SAFT EoS to calculate VLE using the $\varphi - \varphi$ approach, and Equation (21) can be rewritten as [29,30]

$$\left(\frac{Z-1}{P}\right)dP = x_1 d(\ln \varphi_1) + x_2 d(\ln \varphi_2) \quad (22)$$

In this work, we employed the integral form of Equation (22) to conduct thermodynamic consistency tests on isothermal vapor–liquid equilibrium data [27,31].

$$\int \frac{1}{Py_2} dP = \int \frac{1-y_2}{y_2(Z-1)} \frac{d\varphi_1}{\varphi_1} + \int \frac{1}{Z-1} \frac{d\varphi_2}{\varphi_2} \quad (23)$$

In Equation (23), the left-hand side term P represents the experimentally measured system pressure, and y_2 denotes the experimentally determined vapor-phase mole fraction. On the right-hand side, y_2 , Z and φ_i are calculated using the PC-SAFT EoS. The basis for deciding thermodynamic consistency is [27,31]

$$\% \Delta A = 100 \left| \frac{A_\varphi - A_P}{A_P} \right| < 20 \quad (24)$$

where

$$A_P = \int \frac{1}{Py_2} dP \quad (25)$$

$$A_\varphi = \int \frac{1-y_2}{y_2(Z-1)} \frac{d\varphi_1}{\varphi_1} + \int \frac{1}{Z-1} \frac{d\varphi_2}{\varphi_2} \quad (26)$$

The Supplementary Materials provides additional details on the thermodynamic consistency tests.

3. Results and Discussion

This work optimized the hyperparameters `n_estimators`, `max_depth`, `learning_rate`, `subsample`, `colsample_bytree`, `colsample_bylevel`, `colsample_bynode`, `reg_alpha`, `reg_lambda`, `min_child_weight`, `gamma`, and `max_delta_step` of XGBoost using the randomized grid search method from the scikit-learn library [32]. The hyperparameter optimization range for XGBoost is presented in Table 1. Table 2 presents the optimized parameter values obtained through random grid search in this work. Moreover, the neural network architecture of the ANN was optimized using Optuna [33]. A detailed introduction to the Optuna library is provided in the Supplementary Materials. The optimization ranges were set as the number of neural network layers within [1, 5], the number of neurons per layer within [4, 64], with an optimization step size of 4, and the penalty weight μ within [1, 10,000]. The batch size was selected from the set {16, 32, 64, 128}. The hyperparameters of XGBoost in the XGBoost-ANN framework were adopted from the previously optimized XGBoost hyperparameters, while the optimized parameters for the ANN are presented in Table 3. We utilized thermodynamic consistency data from five binary mixture systems (water + methanol, water + ethanol, water + 1-propanol, water + 2-propanol, and water + 1-butanol) for training the XGBoost and XGBoost-ANN surrogate models. The complete dataset consists of 265 data points. To prevent data leakage and ensure fair evaluation across different binary systems, the dataset was split in a system-stratified manner. For each of the five binary associating systems, 80% of the data points were randomly selected for training (total: 212 points), and the remaining 20% were reserved as the test set (total: 53 points). This stratified splitting guarantees that both training and test sets contain data from all five binary systems in equal proportion.

Table 1. Hyperparameter range for XGBoost optimization in this work.

Hyperparameter	The Range of Hyperparameters	Parameter Usage
n_estimators	[60, 400]	The number of decision trees
max_depth	[3, 10]	The maximum depth of each decision tree
learning_rate	[0.01, 0.06]	Learning rate
subsample	[0.5, 0.9]	The fraction of training data randomly sampled for training each tree
colsample_bytree	[0.6, 0.9]	The fraction of features randomly sampled for constructing each tree
colsample_bylevel	[0.7, 0.9]	The fraction of features sampled at each split within a tree
colsample_bynode	[0.7, 0.9]	The fraction of features sampled for each node split in a tree
reg_alpha	[0.1, 20]	The L1 regularization term on weights
reg_lambda	[0.1, 30]	The L2 regularization term on weights
min_child_weight	[2, 10]	The minimum sum of instance weights needed in a child node
gamma	[0.01, 1.0]	The minimum loss reduction required to make a split on a leaf node of the tree
max_delta_step	[1, 5]	The maximum weight change for each tree's output

Table 2. The optimized parameters for XGBoost in this work.

Hyperparameter	The Optimized Value
n_estimators	81
max_depth	6
learning_rate	0.0592
subsample	0.7099
colsample_bytree	0.8816
colsample_bylevel	0.7113
colsample_bynode	0.8444
reg_alpha	0.1038
reg_lambda	0.1141
min_child_weight	3
gamma	0.0100
max_delta_step	1

Table 3. The optimized parameters for ANN in this work.

Hyperparameter	The Optimized Value
The number of layers	4
The number of neurons per layer	{64,52,60,8}
Batch size	16
μ	1023.1298

Details of the data used for training and validating the surrogate model are provided in Table S3 of the Supplementary Materials. The binary interaction parameters k_{ij} were regressed based on the PC-SAFT EoS, and the average absolute percentage deviation (%AAD) is presented in Tables S4–S8 of the Supplementary Materials. The loss function for the ANN training set is presented in Figure S1 of the Supplementary Materials.

A comparison of Figures 7 and 8 reveals that the surrogate model based on XGBoost-ANN outperforms the XGBoost surrogate model in terms of R^2 , RMSE, and MAE (Mean Absolute Error) for both the training and test sets. Additionally, we compared the computational time required to process all data points (265 points) using PC-SAFT, the surrogate model based on XGBoost, and the surrogate model based on XGBoost-ANN, with specific timings presented in Table 4. The computer configuration used for testing computational

speed is detailed in Table S9 of the Supplementary Materials. As shown in Table 4, the computational time of the surrogate models is significantly reduced compared to that of the PC-SAFT EoS. The final computational time was determined by averaging the results of three independent measurements. The surrogate model based on XGBoost achieves a computational time reduction of 2918.5 times, while the surrogate model based on XGBoost-ANN achieves a computational time reduction of 973 times.

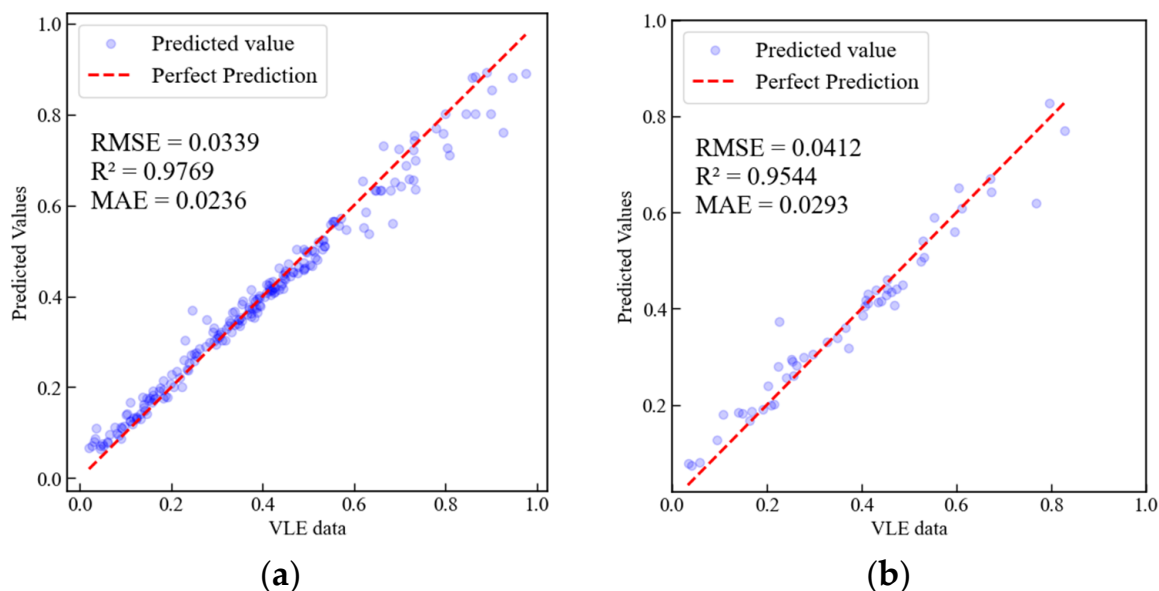


Figure 7. Comparison of predicted values from the surrogate model based on XGBoost with VLE data calculated using PC-SAFT. (a) Training set, (b) Test set.

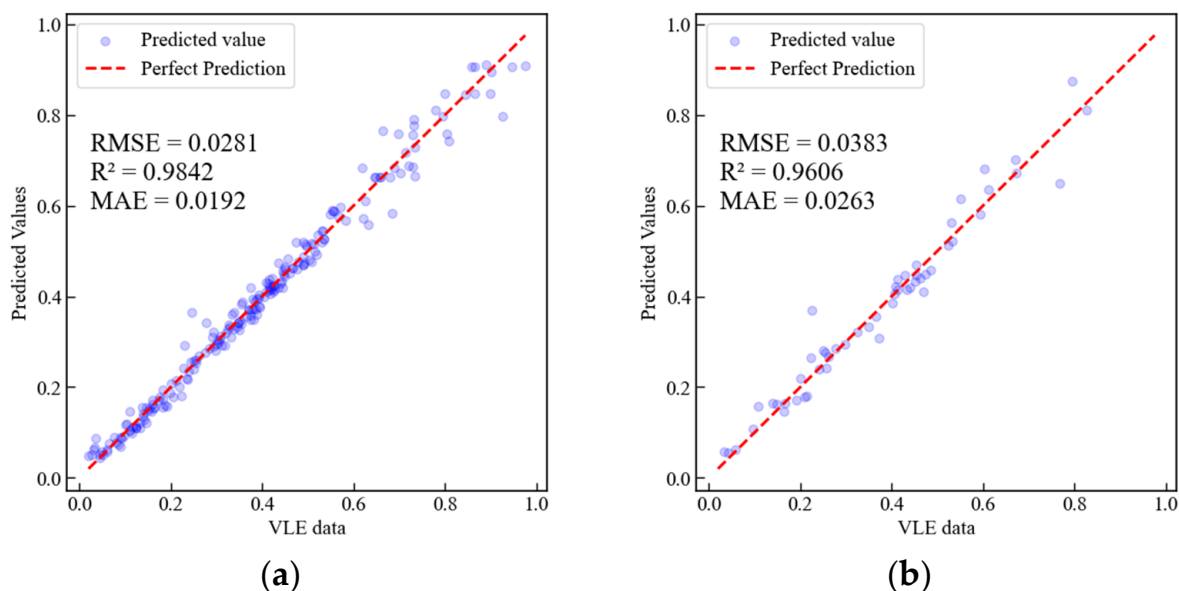


Figure 8. Comparison of predicted values from the surrogate model based on XGBoost-ANN with VLE data calculated using PC-SAFT. (a) Training set, (b) Test set.

Table 4. Computational time required for processing all data points using PC-SAFT, the surrogate model based on XGBoost, and the surrogate model based on XGBoost-ANN.

Model	Computation Time/s
PC-SAFT	2.9185
XGBoost	0.0010
XGBoost-ANN	0.0030

3.1. Water + Methanol

Figure 9 presents the experimental values (y_1), PC-SAFT calculated values, and surrogate model predictions at temperatures of 298.144 K, 318 K, 328.126 K, 373.124 K, 388.150 K, and 403.15 K for the water + methanol system. As observed in Figure 9d,f, when the liquid-phase mole fraction (x_1) exceeds 0.7, the predictions of the surrogate model based on XGBoost and the surrogate model based on XGBoost-ANN are lower than the PC-SAFT calculated values, with the surrogate model based on XGBoost-ANN predictions being closer to the PC-SAFT values. Table 5 presents the R^2 and MSE values for the surrogate model and PC-SAFT calculated values at the temperatures shown in Figure 9. From Table 5, it can be observed that only at temperatures of 298.144 K, 328.136 K, and 388.150 K does the surrogate model based on XGBoost slightly outperform the surrogate model based on XGBoost-ANN, while at all other temperatures, the surrogate model based on XGBoost-ANN yields superior prediction results compared to XGBoost. Table 6 reports the %AAD achieved with PC-SAFT and with the surrogate model, respectively. The $\Delta\%$ AAD is the absolute deviation between the %AAD of the PC-SAFT calculations and the %AAD of the surrogate model predictions. It is evident from Table 6 that the $\Delta\%$ AAD for all systems displayed in Figure 9 remains below 5%.

Table 5. Comparison of surrogate models and PC-SAFT calculated values at different temperatures for the water + methanol system.

Temperature/K	XGBoost		XGBoost-ANN	
	MSE	R^2	MSE	R^2
298.144	0.0001	0.9929	0.0002	0.9884
318.000	0.0002	0.9832	2.3384×10^{-5}	0.9979
328.136	0.0002	0.9918	0.0003	0.9867
373.124	0.0044	0.9328	0.0025	0.9617
388.150	0.0001	0.9979	0.0003	0.9885
403.150	0.0028	0.9563	0.0017	0.9727

Table 6. The %AAD values of the PC-SAFT calculations and those of the surrogate models in Figure 9.

Temperature/K	PC-SAFT	XGBoost		XGBoost-ANN	
	%AAD	%AAD	$\Delta\%$ AAD	%AAD	$\Delta\%$ AAD
298.144	4.9033	8.3390	3.4357	6.184	1.4151
318.000	6.6105	11.2727	4.6622	6.4610	0.1495
328.136	5.3121	8.4368	3.1247	3.7656	1.5464
373.124	2.1875	6.4645	4.2773	5.9824	3.7952
388.150	1.6987	2.6912	0.9925	2.7253	1.0266
403.150	5.1908	7.2112	2.2004	9.6721	4.4814

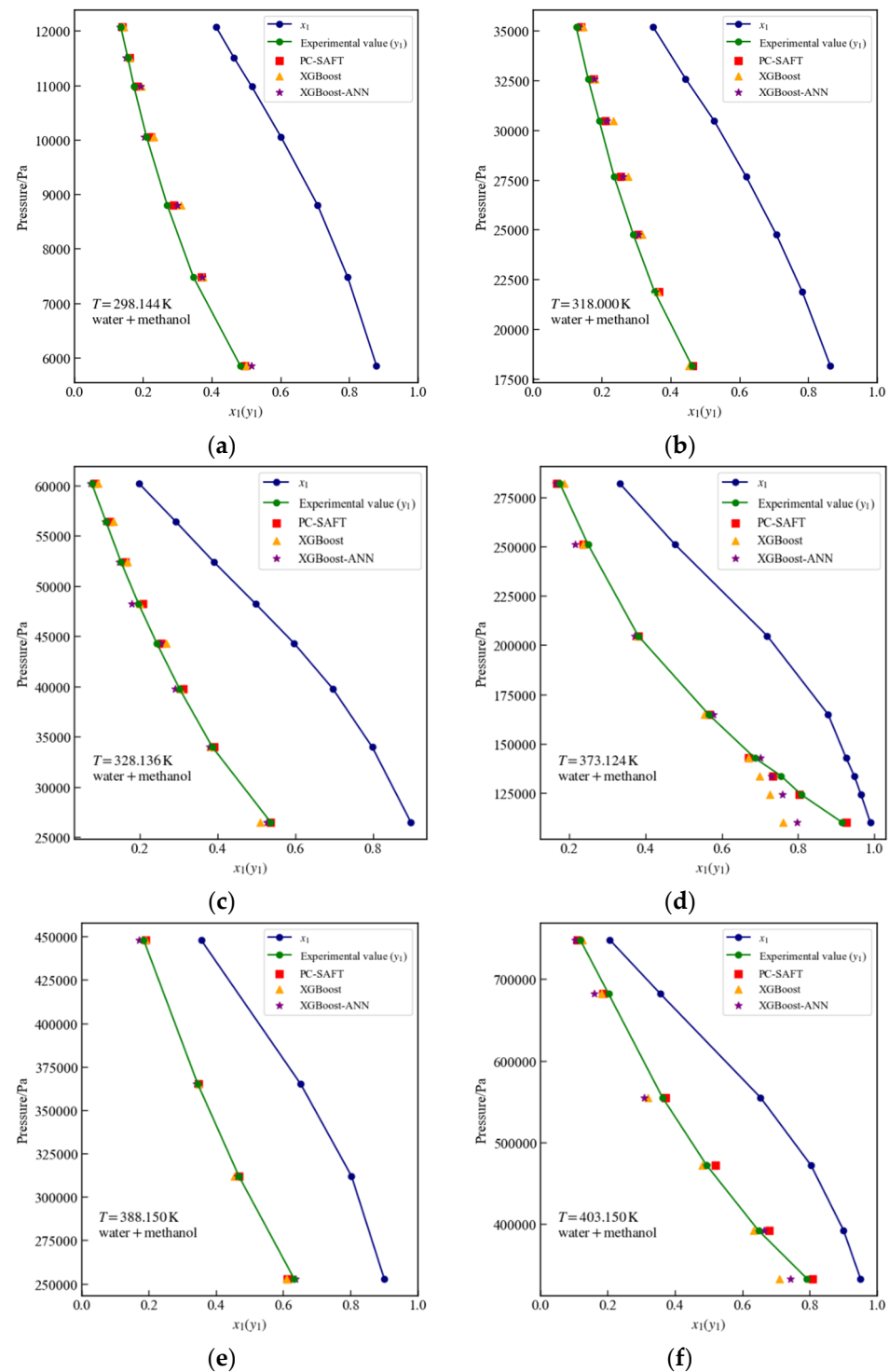


Figure 9. Vapor–liquid equilibrium data for the water + methanol system at different temperatures: comparison of PC-SAFT calculated values with predictions from XGBoost and XGBoost-ANN surrogate models. (a) 298.144 K, (b) 318 K, (c) 328.126 K, (d) 373.124 K, (e) 388.15 K, (f) 403.15 K.

3.2. Water + Ethanol

Figure 10 illustrates the liquid-phase mole fraction (x_1), experimental values (y_1), PC-SAFT calculated values, and surrogate models' predictions at temperatures of 323.15 K, 328.15 K, 333.15 K, and 423.7 K. As observed in Figure 10d, the surrogate model based on XGBoost-ANN exhibits larger prediction deviations for three points where y_1 lies within the range [0.6, 0.8]. Table 7 indicates that at 423.7 K, the surrogate model based on XGBoost provides more accurate predictions, whereas at the other temperatures, the surrogate model

based on XGBoost-ANN predictions is closer to the PC-SAFT calculated values. Table 8 indicates that all $\Delta\%$ AAD values associated with Figure 10 remain below 4%, with the only exception being the surrogate model based on XGBoost-ANN at 423.7 K, which shows substantially greater discrepancies relative to PC-SAFT and experimental data.

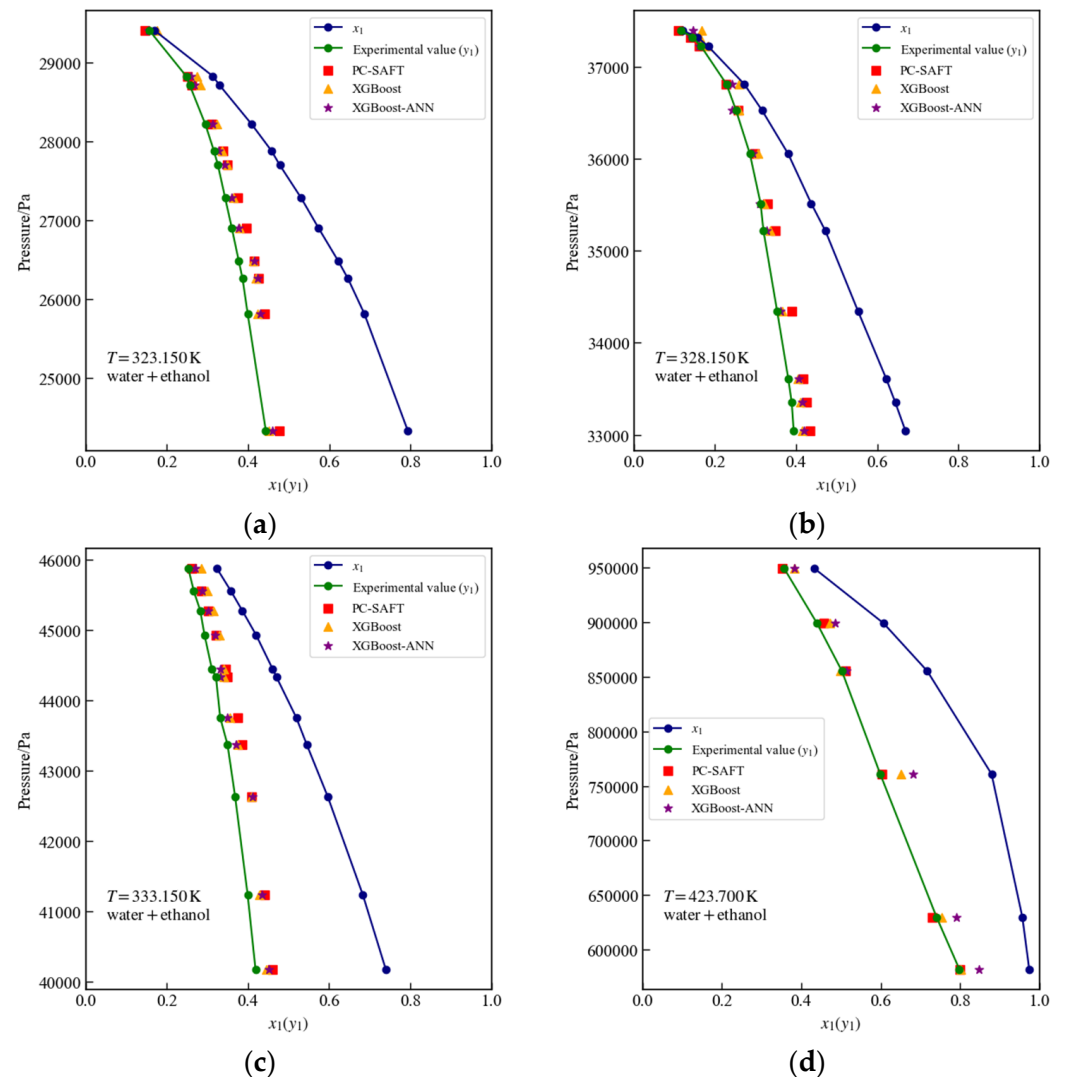


Figure 10. Vapor-Liquid equilibrium data for the water + ethanol system at different temperatures: comparison of PC-SAFT calculated values with predictions from XGBoost and XGBoost-ANN surrogate models. (a) 323.15 K, (b) 328.15 K, (c) 333.15 K, (d) 423.7 K.

Table 7. Comparison of surrogate models and PC-SAFT calculated values at different temperatures for the water + ethanol system.

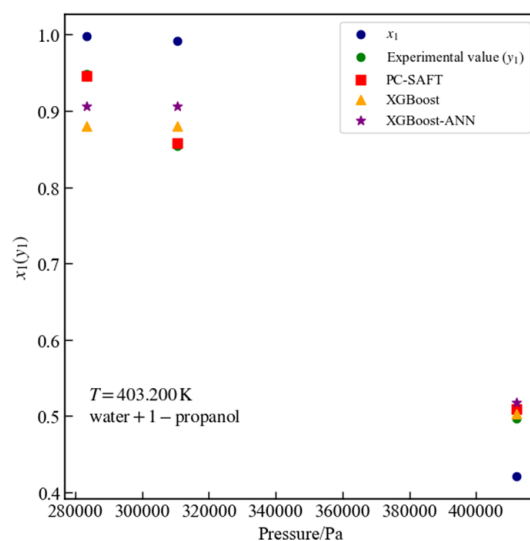
Temperature/K	XGBoost		XGBoost-ANN	
	MSE	R^2	MSE	R^2
323.150	0.0003	0.9658	0.0001	0.9873
328.150	0.0006	0.9519	0.0003	0.9732
333.150	0.0002	0.9510	0.0001	0.9640
423.700	0.0007	0.9703	0.0023	0.9045

Table 8. The %AAD values of the PC-SAFT calculations and those of the surrogate models in Figure 10.

Temperature/K	PC-SAFT	XGBoost		XGBoost-ANN	
	%AAD	%AAD	$\Delta\%$ AAD	%AAD	$\Delta\%$ AAD
323.150	7.0501	7.7871	0.7380	5.4539	1.5962
328.150	6.0784	9.9871	3.9087	5.6462	0.4322
333.150	9.0990	9.2426	0.1437	7.2375	1.8614
423.700	1.4961	4.4587	2.9626	7.6918	6.1957

3.3. Water + 1-Propanol

Figure 11 presents three data points for the water + 1-propanol system at a temperature of 403.2 K that passed thermodynamic consistency tests. For these points, the surrogate model based on XGBoost yields an MSE of 0.0016 and an R^2 of 0.9547, while the surrogate model based on XGBoost-ANN yields an MSE of 0.0013 and an R^2 of 0.9627. For all experimental data points of the water + 1-propanol system, the surrogate model based on XGBoost achieves an MSE of 0.0016 and an R^2 of 0.9482, whereas the surrogate model based on XGBoost-ANN achieves an MSE of 0.0012 and an R^2 of 0.9586. For the water + 1-propanol binary system across all data points, the %AAD values are 4.3023% for PC-SAFT, 4.9932% for the surrogate model based on XGBoost, and 5.1969% for the surrogate model based on XGBoost-ANN, respectively.

**Figure 11.** Vapor–liquid equilibrium data for the water + 1-propanol system at 403.2 K: comparison of PC-SAFT calculated values with predictions from XGBoost and XGBoost-ANN surrogate models.

3.4. Water + 2-Propanol

Figure 12 presents three data points for the water + 2-propanol system at a temperature of 353.129 K that passed thermodynamic consistency tests. For these points, the surrogate model based on XGBoost yields an MSE of 0.0007 and an R^2 of 0.9669, while the surrogate model based on XGBoost-ANN yields an MSE of 0.0001 and an R^2 of 0.9943. For all experimental data points of the water + 2-propanol system, the surrogate model based on XGBoost achieves an MSE of 0.0027 and an R^2 of 0.8778, whereas the surrogate model based on XGBoost-ANN achieves an MSE of 0.0025 and an R^2 of 0.8865. For the water + 2-propanol binary system across all data points, the %AAD values are 11.7714% for PC-SAFT, 13.2330% for the surrogate model based on XGBoost, and 14.8303% for the surrogate model based on XGBoost-ANN, respectively. Both PC-SAFT and the surrogate models show considerably larger deviations for the VLE experimental data of the water + 2-propanol system.

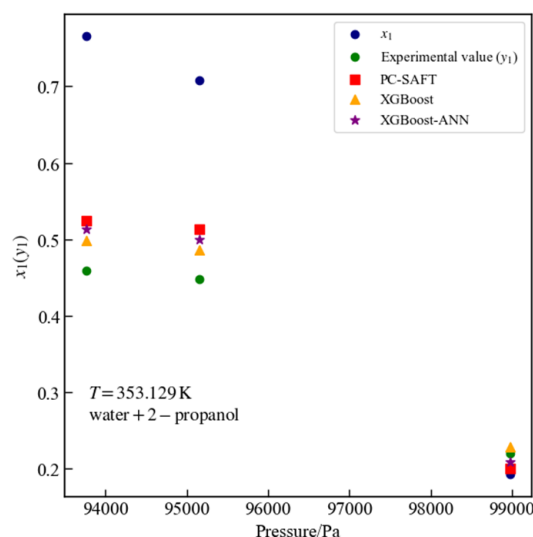


Figure 12. Vapor–Liquid equilibrium data for the water + 2-propanol system at 353.129 K: comparison of PC-SAFT calculated values with predictions from XGBoost and XGBoost-ANN surrogate models.

3.5. Water + 1-Butanol

Figure 13 presents three data points for the water + 1-butanol system at a temperature of 403.118 K that passed thermodynamic consistency tests. For these points, the surrogate model based on XGBoost yields an MSE of 0.0055 and an R^2 of 0.7970, while the surrogate model based on XGBoost-ANN yields an MSE of 0.0051 and an R^2 of 0.8098. For all experimental data points of the water + 1-butanol system, the surrogate model based on XGBoost achieves an MSE of 0.0057 and an R^2 of 0.8670, whereas the surrogate model based on XGBoost-ANN achieves an MSE of 0.0040 and an R^2 of 0.9071. For all data of the water + 1-butanol system, the predictions of the surrogate model based on XGBoost-ANN are significantly closer to the PC-SAFT calculated values. For the water + 1-butanol binary system across all data points, the %AAD values are 7.1175% for PC-SAFT, 11.6284% for the surrogate model based on XGBoost, and 9.7374% for the surrogate model based on XGBoost-ANN, respectively.

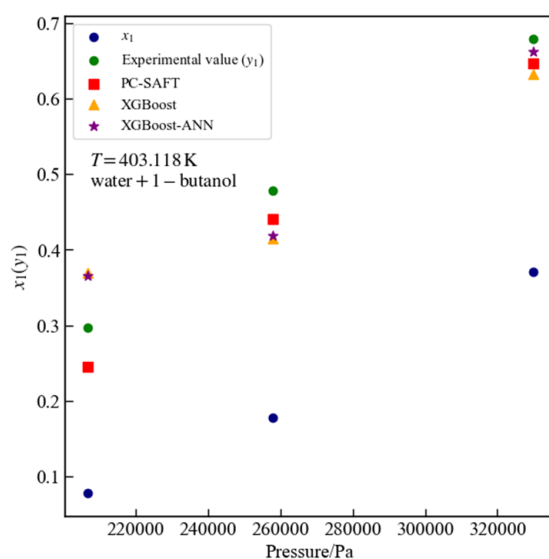


Figure 13. Vapor–liquid equilibrium data for the water + 1-butanol system at 403.118 K: comparison of PC-SAFT calculated values with predictions from XGBoost and XGBoost-ANN surrogate models.

The present study is restricted to binary mixtures. Validation on multicomponent and reactive systems, which is essential for real industrial applications, remains to be conducted and is highlighted as the most important direction for future work. Another important limitation of the current work is the absence of uncertainty quantification. In future work, we will conduct a more comprehensive investigation of PC-SAFT surrogate models by incorporating a broader range of machine learning algorithms as well as uncertainty quantification and analysis. The surrogate models presented in this paper are trained and validated on only five binary associating systems. Defining the precise domain of applicability and safe extrapolation boundaries of the proposed framework will therefore constitute a major focus of our future work, in which a significantly larger and more diverse set of binary and multicomponent systems will be incorporated.

4. Conclusions

This work constructed surrogate models for the VLE of the water + methanol, water + ethanol, water + 1-propanol, water + 2-propanol, and water + 1-butanol binary systems using XGBoost and XGBoost-ANN methods. The models were optimized using randomized grid search and the Optuna library. Both surrogate models significantly reduce the computational time for calculating VLE data compared to the PC-SAFT EoS, while achieving good prediction results. Among the systems investigated, the water + 2-propanol and water + 1-butanol binary mixtures exhibit the largest deviations for both PC-SAFT and all surrogate models. These two systems possess only limited thermodynamically consistent experimental VLE data points 13 for water + 2-propanol and 14 for water + 1-butanol after application of rigorous thermodynamic consistency tests. Even when the binary interaction parameters k_{ij} of PC-SAFT were regressed against these experimental data, substantial deviations from the measurements persist. Consequently, the relatively poor performance of the surrogate models for these two systems is largely attributable to the scarcity or questionable quality of the underlying experimental database. Improved predictive accuracy of the surrogate models is expected in the future once more reliable and extensive VLE measurements become available for these two challenging binary systems.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/pr13123918/s1>. The supplementary materials are in the supplementary_material.docx [34–65]. The raw data is available at: https://github.com/ywpang-ren/XGb_pang_paper/tree/main/paper_data (accessed on 26 November 2025).

Author Contributions: Y.P.: Conceptualization, Data Curation, Investigation, Methodology, Writing—original draft. Z.D.: Data Curation, Writing—review and editing, Resource. Q.L.: Writing—review and editing, Resource. All authors have read and agreed to the published version of the manuscript.

Funding: This work is in part supported by the Research Fund from the National Key Research and Development Program of China (No. 2022YFC2106300). The APC was funded by the National Key Research and Development Program of China.

Data Availability Statement: Data will be made available on request.

Conflicts of Interest: The author declares no competing financial interests.

Nomenclature

%AAD	average absolute percentage deviation
$\hat{a}^{assoc}$	the association contribution term
$\hat{a}^{disp}$	the dispersion contribution term
$\hat{a}^{hc}$	the hard-chain reference contribution term
ANN	Artificial Neural Network

Ap	pressure integral term
A_φ	the integral of the fugacity coefficient
$\tilde{a}^{res}$	the residual Helmholtz energy
EoS	Equation of State
f_t	additional perturbation function
$\hat{f}_i$	the fugacity of the component i
k_B	Boltzmann constant
k_{ij}	binary interaction parameters
L	the objective function of XGBoost
$LeavesN$	the number of leaves in the tree
MAE	mean absolute error
m_i	the number of segments
MSE	mean square error
N_{Ai}, N_{Bi}	the number of association sites
n_c	the number of components
P	pressure
PINN	Physics-Informed Neural Network
RMSE	root mean square error
T	temperature
V^E	reduced molar volume
VLE	vapor–liquid equilibrium
w_i	the score associated with the ith leaf
y_i	true value
$\hat{y}_i$	predict value
Z	compression factor
σ_i	segment diameter
ε_i/k_B	depth of pair potential, K
$\kappa^{A_iB_j}$	associative volume
$\varepsilon^{A_iB_j}/k_B$	associative energy
μ	penalty weight
μ^{res}	the residual chemical potential
μ_i	the chemical potential of component i
φ_i	the fugacity coefficient of the component i
η	packing fraction
Ω	penalize function
γ, λ	penalization parameters
γ_i	activity coefficient of component i

References

1. Gross, J.; Sadowski, G. Perturbed-Chain SAFT: An Equation of State Based on a Perturbation Theory for Chain Molecules. *Ind. Eng. Chem. Res.* **2001**, *40*, 1244–1260. [\[CrossRef\]](#)
2. Soave, G. Equilibrium constants from a modified Redlich-Kwong equation of state. *Chem. Eng. Sci.* **1972**, *27*, 1197–1203. [\[CrossRef\]](#)
3. Peng, D.; Robinson, D.B. A rigorous method for predicting the critical properties of multicomponent systems from an equation of state. *AIChE J.* **1977**, *23*, 137–144. [\[CrossRef\]](#)
4. McBride, K.; Sundmacher, K. Overview of Surrogate Modeling in Chemical Process Engineering. *Chem. Ing. Tech.* **2019**, *91*, 228–239. [\[CrossRef\]](#)
5. Winz, J.; Engell, S. Data-efficient surrogate modeling of thermodynamic equilibria using Sobolev training, data augmentation and adaptive sampling. *Chem. Eng. Sci.* **2024**, *299*, 120461. [\[CrossRef\]](#)
6. Rittig, J.G. Gibbs–Duhem-informed neural networks for binary activity coefficient prediction. *Digit. Discov.* **2023**, *2*, 1752–1767. [\[CrossRef\]](#)
7. Li, Y.; Zhang, T.; Sun, S.; Gao, X. Accelerating flash calculation through deep learning methods. *J. Comput. Phys.* **2019**, *394*, 153–165. [\[CrossRef\]](#)
8. Ottaiano, G.Y.; Martins, T.D. Machine learning models for vapor-liquid equilibrium of binary mixtures: State of the art and future opportunities. *Chem. Eng. Res. Des.* **2024**, *211*, 66–77. [\[CrossRef\]](#)

9. Wang, S.; Sobiecki, N.; Ding, D.; Zhu, L.; Wu, Y.-S. Accelerating and stabilizing the vapor-liquid equilibrium (VLE) calculation in compositional simulation of unconventional reservoirs using deep learning based flash calculation. *Fuel* **2019**, *253*, 209–219. [\[CrossRef\]](#)
10. Sun, J.; Xue, J.; Yang, G.; Li, J.; Zhang, W. Vapor-liquid phase equilibrium prediction for mixtures of binary systems using graph neural networks. *AIChE J.* **2025**, *71*, e18637. [\[CrossRef\]](#)
11. Ihunde, T.A.; Olorode, O. Application of physics-informed neural networks to compositional modeling. *J. Pet. Sci. Eng.* **2022**, *211*, 110175. [\[CrossRef\]](#)
12. Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **2019**, *378*, 686–707. [\[CrossRef\]](#)
13. Xiang, Z.; Peng, W.; Liu, X.; Yao, W. Self-adaptive loss balanced Physics-informed neural networks. *Neurocomputing* **2022**, *496*, 11–34. [\[CrossRef\]](#)
14. Li, S.; Feng, X. Dynamic Weight Strategy of Physics-Informed Neural Networks for the 2D Navier–Stokes Equations. *Entropy* **2022**, *24*, 1254. [\[CrossRef\]](#)
15. Difonzo, F.V.; Lopez, L.; Pellegrino, S.F. Physics informed neural networks for an inverse problem in peridynamic models. *Eng. Comput.* **2024**, *5*, 1–10. [\[CrossRef\]](#)
16. Hakim, M.; Behmardikalantari, G.; Abedini Najafabadi, H.; Pazuki, G.; Vosoughi, A.; Vossoughi, M. Prediction of liquid-liquid equilibrium behavior for aliphatic+aromatic+ionic liquid using two different neural network-based models. *Fluid Phase Equilibria* **2015**, *394*, 140–147. [\[CrossRef\]](#)
17. Petersen, R.; Fredenslund, A.; Rasmussen, P. Artificial neural networks as a predictive tool for vapor-liquid equilibrium. *Comput. Chem. Eng.* **1994**, *18*, S63–S67. [\[CrossRef\]](#)
18. Ghanadzadeh, H.; Ahmadifar, H. Estimation of (vapour+liquid) equilibrium of binary systems (tert-butanol+2-ethyl-1-hexanol) and (n-butanol+2-ethyl-1-hexanol) using an artificial neural network. *J. Chem. Thermodyn.* **2008**, *40*, 1152–1156. [\[CrossRef\]](#)
19. Urata, S.; Takada, A.; Murata, J.; Hiaki, T.; Sekiya, A. Prediction of vapor-liquid equilibrium for binary systems containing HFes by using artificial neural network. *Fluid Phase Equilibria* **2002**, *199*, 63–78. [\[CrossRef\]](#)
20. Yamamoto, H.; Tochigi, K. Prediction of vapor-liquid equilibria using reconstruction—Learning neural network method. *Fluid Phase Equilibria* **2007**, *257*, 169–172. [\[CrossRef\]](#)
21. Reyhani, S.Z.; Ghanadzadeh, H.; Puigjaner, L.; Recances, F. Estimation of Liquid–Liquid Equilibrium for a Quaternary System Using the GMDH Algorithm. *Ind. Eng. Chem. Res.* **2009**, *48*, 2129–2134. [\[CrossRef\]](#)
22. Moghadam, M.; Asgharzadeh, S. On the application of artificial neural network for modeling liquid-liquid equilibrium. *J. Mol. Liq.* **2016**, *220*, 339–345. [\[CrossRef\]](#)
23. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the KDD'16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [\[CrossRef\]](#)
24. Rehner, P.; Bauer, G.; Gross, J. FeO₃: An Open-Source Framework for Equations of State and Classical Density Functional Theory. *Ind. Eng. Chem. Res.* **2023**, *62*, 5347–5357. [\[CrossRef\]](#)
25. Kontogeorgis, G.M. *Thermodynamic Models for Industrial Applications: From Classical and Advanced Mixing Rules to Association theories*; John Wiley & Sons: Hoboken, NJ, USA, 2009.
26. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *arXiv* **2019**, arXiv:1912.01703. [\[CrossRef\]](#)
27. Arce, P.F.; Vieira, N.F.; Igarashi, E.M.S. Thermodynamic Modeling and Simulation of Biodiesel Systems at Supercritical Conditions. *Ind. Eng. Chem. Res.* **2018**, *57*, 751–767. [\[CrossRef\]](#)
28. Smith, J.M.; Van Ness, H.C.; Abbott, M.M. *Introduction to Chemical Engineering Thermodynamics*, 6th ed.; McGraw-Hill: New York, NY, USA, 2014.
29. Valderrama, J.O.; Zavaleta, J. Thermodynamic consistency test for high pressure gas–solid solubility data of binary mixtures using genetic algorithms. *J. Supercrit. Fluids* **2006**, *39*, 20–29. [\[CrossRef\]](#)
30. Saali, A.; Shokouhi, M.; Sakhaeinia, H.; Kazemi, N. Thermodynamic Consistency Test of Vapor-liquid Equilibrium Data of Binary Systems Including Carbon Dioxide (CO₂) and Ionic Liquids Using the Generic Redlich–Kwong Equation of State. *J. Solut. Chem.* **2020**, *49*, 383–404. [\[CrossRef\]](#)
31. Alvarez, V.H.; Aznar, M. Application of a Thermodynamic Consistency Test to Binary Mixtures Containing an Ionic Liquid. *Open Thermodyn. J.* **2008**, *2*, 25–38. [\[CrossRef\]](#)
32. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
33. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-generation Hyperparameter Optimization Framework. *arXiv* **2019**, arXiv:1907.10902. [\[CrossRef\]](#)

34. Gross, J.; Sadowski, G. Application of the perturbed-chain SAFT equation of state to associating systems. *Ind. Eng. Chem. Res.* **2002**, *41*, 5510–5515. [[CrossRef](#)]
35. Huang, S.H.; Radosz, M. Equation of state for small, large, polydisperse, and associating molecules. *Ind. Eng. Chem. Res.* **1990**, *29*, 2284–2294. [[CrossRef](#)]
36. Hall, D.J.; Mash, C.J.; Pemberton, R.C. *NPL Report Chemistry*; National Physical Laboratory: Middlesex, UK, 1979.
37. McGlashan, M.L.; Williamson, A.G. Isothermal Liquid-Vapor Equilibria for System Methanol-Water. *J. Chem. Eng. Data* **1976**, *21*, 196–199. [[CrossRef](#)]
38. Ferguson, J.B.; Funnel, W.S. The Determination of Vapor and Liquid Compositions in Binary Systems. I. Methyl Alcohol-Water. *J. Phys. Chem.* **1929**, *33*, 1–8. [[CrossRef](#)]
39. Yao, J.; Li, H.; Han, S. Vapor–Liquid Equilibrium Data for Methanol–Water–NaCl at 45 °C. *Fluid Phase Equilibria* **1999**, *4*, 15. [[CrossRef](#)]
40. Fu, J.; Wang, K.; Hu, Y. Studies on vapor-liquid and liquid-liquid-vapor equilibria for the ternary system methanol-methyl methacrylate-water (II) ternary system. *Chin. J. Chem. Eng.* **1989**, *4*, 14.
41. Broul, M.; Hlavaty, K.; Linek, J. Liquid-Vapour Equilibrium in Systems of Electrolytic Components: V the System Methanol + Water + Lithium Chloride at 60 °C. *Collect. Czech. Chem. Commun.* **1969**, *34*, 28–34. [[CrossRef](#)]
42. Schubert, H. Influence of Simple Salts on the Isothermal Vapor-Liquid Equilibrium Behavior of the Methanol-Water System. *Z. Für Phys. Chem.* **1974**, *255*, 165–179. [[CrossRef](#)]
43. Bao, Z.; Liu, M.; Yang, J. Measurement and Correlation of Moderate Pressure Vapor-Liquid Equilibrium Data for Methanol-Water Binary System. *J. Chem. Ind. Eng. China* **1995**, *46*, 230.
44. Griswold, J. Phase-Equilibria of the Acetone-Methanol-Water System from 100°C into the Critical Region. *Chem. Eng. Progr., Symp. Ser.* **1952**, *48*, 18–34.
45. Schröder, W. Messungen von Siedegleichgewichten Bei Überdruck. *Chem. Ing. Tech.* **1958**, *30*, 523–525. [[CrossRef](#)]
46. Phutela, R.; Kooner, Z.; Fenby, D. Vapour Pressure Study of Deuterium Exchange Reactions in Water-Ethanol Systems: Equilibrium Constant Determination. *Aust. J. Chem.* **1979**, *32*, 2353. [[CrossRef](#)]
47. Yamamoto, H.; Terano, T.; Nishi, Y.; Tokunaga, J. Vapor-Liquid Equilibria for Methanol + Ethanol + Calcium Chloride, + Ammonium Iodide, and + Sodium Iodide at 298.15 K. *J. Chem. Eng. Data* **1995**, *40*, 472–477. [[CrossRef](#)]
48. Herraiz, J.; Shen, S.; Coronas, A. Vapor–Liquid Equilibria for Methanol + Poly(Ethylene Glycol) 250 Dimethyl Ether. *J. Chem. Eng. Data* **1998**, *43*, 191–195. [[CrossRef](#)]
49. Vu, D.T.; Lira, C.T.; Asthana, N.S.; Kolah, A.K.; Miller, D.J. Vapor–Liquid Equilibria in the Systems Ethyl Lactate + Ethanol and Ethyl Lactate + Water. *J. Chem. Eng. Data* **2006**, *51*, 1220–1225. [[CrossRef](#)]
50. Kurihara, K.; Minoura, T.; Takeda, K.; Kojima, K. Isothermal Vapor-Liquid Equilibria for Methanol + Ethanol + Water, Methanol + Water, and Ethanol + Water. *J. Chem. Eng. Data* **1995**, *40*, 679–684. [[CrossRef](#)]
51. Kang, J.W.; Diky, V.; Chirico, R.D.; Magee, J.W.; Muzny, C.D.; Abdulagatov, I.; Kazakov, A.F.; Frenkel, M. Quality Assessment Algorithm for Vapor–Liquid Equilibrium Data. *J. Chem. Eng. Data* **2010**, *55*, 3631–3640. [[CrossRef](#)]
52. Schubert, H. Isothermic behavior of phase-equilibrium in the system N-alcohol-water-urea at 60-degrees Celsius. *Z. Phys. Chem.-Leipz.* **1980**, *261*, 777–790. [[CrossRef](#)]
53. Arango, I.C.; Villa, A.L. Isothermal Vapor–Liquid and Vapor–Liquid–Liquid Equilibrium for the Ternary System Ethanol+water+diethyl Carbonate and Constituent Binary Systems at Different Temperatures. *Fluid Phase Equilibria* **2013**, *339*, 31–39. [[CrossRef](#)]
54. Cristino, A.F.; Rosa, S.; Morgado, P.; Galindo, A.; Filipe, E.J.M.; Palavra, A.M.F.; Nieto De Castro, C.A. High-Temperature Vapour–Liquid Equilibrium for the Water–Alcohol Systems and Modeling with SAFT-VR: 1. Water–Ethanol. *Fluid Phase Equilibria* **2013**, *341*, 48–53. [[CrossRef](#)]
55. Niesen, V.; Palavra, A.; Kidnay, A.J.; Yesavage, V.F. An Apparatus for Vapor–Liquid Equilibrium at Elevated Temperatures and Pressures and Selected Results for the Water–Ethanol and Methanol–Ethanol Systems. *Fluid Phase Equilibria* **1986**, *31*, 283–298. [[CrossRef](#)]
56. Barr-David, F.; Dodge, B.F. Vapor-Liquid Equilibrium at High Pressures. The Systems Ethanol-Water and 2-Propanol-Water. *J. Chem. Eng. Data* **1959**, *4*, 107–121. [[CrossRef](#)]
57. Murti, P.; Van Winkle, M. Vapor-Liquid Equilibria for Binary Systems of Methanol, Ethyl Alcohol, 1-Propanol, and 2-Propanol with Ethyl Acetate and 1-Propanol-Water. *Ind. Eng. Chem. Chem. Eng. Data Ser.* **1958**, *3*, 72–81. [[CrossRef](#)]
58. Schreiber, E.; Schüttau, E.; Rant, D.; Schubert, H. Die Beeinflussbarkeit des isothermen Phasengleichgewichtsverhaltens der Systeme 1-Propanol/Wasser und n-Butanol/Wasser durch einige Metallchloride. *Z. Für Phys. Chem.* **1971**, *247*, 23–40. [[CrossRef](#)]
59. Woerpel, U.; Vohland, P.; Schubert, H. The Effect of Urea on the Vapor-Liquid Equilibrium Behavior of n-Propanol/Water at 60 °C. *Z. Phys. Chem.* **1977**, *258*, 905–912.
60. Wrewsky, M.Z. Composition and Vapor Pressure of Binary Mixtures. *Phys. Chem. Stoichiomet. Verwandtschaftsl.* **1913**, *81*, 1–29.

61. Ratcliff, G.A.; Chao, K.C. Prediction of Thermodynamic Properties of Polar Mixtures by a Group Solution Model. *Can. J. Chem. Eng.* **1969**, *47*, 148–153. [[CrossRef](#)]
62. Udovenko, V.V.; Mazanko, T.F. Liquid-Vapour Equilibrium in Propan-2-Ol-Water and Propan-2-Ol-Benzene Systems. *Russ. J. Phys. Chem. USSR.* **1967**, *41*, 863.
63. Wu, H.S.; Hagewiesche, D.; Sandler, S.I. Vapor—Liquid Equilibria of 2-Propanol + Water + N,N-Dimethyl Formamide. *Fluid Phase Equilibria* **1988**, *43*, 77–89. [[CrossRef](#)]
64. Lyzlova, R.V.; Zaiko, L.N.; Susarev, M.P. Experimental Investigation and Calculation of Liquid-Vapor Equilibrium in the Ternary System n-Butyl Alcohol + Isobutyl Alcohol + Water at 35 c. *Zhurnal Prikl. Khimii* **1979**, *52*, 551–555.
65. Kharin, S.E.; Pereygin, V.M. Liquid-Vapor Phase Equilibrium in Ethanol-n-Butanol and Water-n-Butanol Systems. *Izv. Vyssh. Ucheb. Zaved. Khim. Khim. Tekhnol.* **1969**, *12*, 424–428.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.