

# UltraModel: A Modeling Paradigm for Industrial Objects

## Appendix

**Haoran Yang, Yinan Zhang, Qunshan He, Yuqi Ye, Jing Zhao and Wenhai Wang\***

College of Control Science and Engineering, Zhejiang University, Hangzhou, China

{22332111, zhangyinan, heqs, 0923144, zdzzlab}@zju.edu.cn, zhaojing9532@163.com,

## 1 A Methodology

### 2 A.1 Graph construction for the twin model

In our proposed UltraModel, we leverage graph neural networks (GNNs) methods to enhance the dimensionality and aggregate information from industrial data for MIO tasks. The construction of graphs is crucial in GNNs. Since our UltraModel aims to serve as a paradigm for industrial modeling, it must be adaptable to handle a wide variety of industrial objects. To achieve this, we classify the graph construction methods into the following three cases based on the distinct mechanisms inherent in industrial objects:

- (i) **Fully clear mechanisms of industrial object ( $\mathcal{A}_1$ )**. When the mechanisms of industrial object are fully clear, the functional mapping relationships between variables are well-defined. In this case, for a given node  $v_i$ , we identify all nodes  $v_j$  that satisfy the functional mapping  $v_i = f(v_j)$ , forming the set  $\mathcal{N}(v_i)$ . For each  $v_j \in \mathcal{N}(v_i)$ , we add a directed edge  $e_{ji}$  from  $v_j$  to  $v_i$ . This process will result in the twin model graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , where  $\mathcal{E}$  denotes the set of all edges, which also corresponds to the adjacency matrix.
- (ii) **Partially clear mechanisms of industrial object ( $\mathcal{A}_2$ )**. When the mechanisms of the industrial object are partially clear, certain variables exhibit well-defined functional mapping relationships, while others do not. We process the well-defined part using the method described above. For the other part, considering the highly coupled nature of the variables in industrial data, we adopt a fully connected strategy. Specifically, for such nodes  $v_i$ , its  $\mathcal{N}(v_i)$  will be the entire nodes set  $\mathcal{V}$ , i.e.  $\mathcal{N}(v_i) = \mathcal{V}$ , meaning that the edges  $e_{ji}$  connected to  $v_i$  originate from all nodes  $j = 1, 2, \dots, N$ . By combining all nodes, the graph is constructed.
- (iii) **Unclear mechanisms of industrial object ( $\mathcal{A}_3$ )**. When the mechanisms of industrial object are unclear, we also use a fully connected strategy to construct the graph.

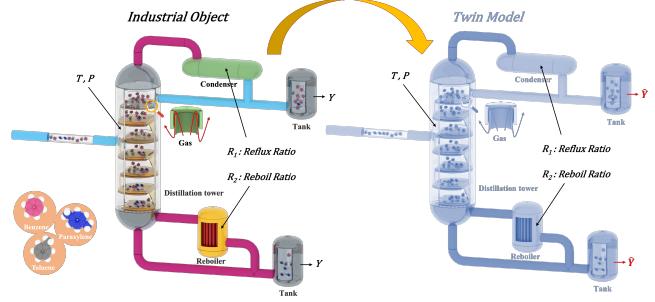


Figure 1: Schematic of modeling for distillation column. It also serves as the industrial background for the first dataset, DIS-COL, in the Experiment section of this paper.

## B Experimental Setup

### B.1 Datasets Details

#### DIS-COL

Fig. 1 illustrates the task of constructing a twin model for a distillation column designed to separate a mixture of benzene, toluene, and xylene. In fact, the model developed in this work is not limited to the distillation column itself but represents a complete distillation system centered on the column, which adds practical significance to our research. The distillation system involves various types of variables, as shown in Table 1. Among them, operational variables (OV) are those that can be directly controlled by operators; state variables (SV) describe the internal operating conditions of the system; and final output variables (FV) represent the target outputs of the distillation system. Once the operational variables are determined, the entire distillation system is also defined, and thus we treat these operational variables as input variables for the system. State variables, which are not the focus of this study, are not discussed in detail. We select four critical operational variables—feed temperature, feed pressure, reflux ratio, and reboil ratio—as the input variables and construct a twin model of the corresponding distillation system based on these inputs. Other operational variables are set to fixed values, as shown in Table 1.

In the DIS-COL dataset, the feed temperature range is set to 350K-410K, with a step size of 5K; the feed pressure range is 115,000Pa-150,000Pa, with a step size of 2,500Pa; and the reflux ratio and reboil ratio both range from 1 to 20, with a

\*Corresponding author

Variable name (Unit)	Value	Type	Variable name (Unit)	Value	Type
Feed composition	{0.3,0.3,0.4}	OV	Total number of plates	30	OV
Feed flow rate	100	OV	Feed plate location	15	OV
<b>Feed temperature (K)</b>	350 - 510	OV	Plate pressure distribution	—	SV
<b>Feed pressure (Pa)</b>	115000 - 150000	OV	Plate temperature distribution	—	SV
<b>Reflux ratio</b>	1 - 20	OV	.....	—	SV
<b>Reboil ratio</b>	1 - 20	OV	<i>Percentage Composition of Substances</i>	{y <sub>1</sub> , y <sub>2</sub> , y <sub>3</sub> }	FV

Table 1: Variables in distillation system. The **variables** in bold represent the input variables for DIS-COL, while the italicized *variables* denote the output variables.

65 step size of 1. Each combination of the values from these four  
 66 variables constitutes a sample, resulting in a total of 78,000  
 67 samples. However, collecting the system output for each of  
 68 these operational variable combinations in a real-world distil-  
 69 lation system would incur an unaffordable cost. Therefore, in  
 70 addition to the work presented in this paper, we undertake an-  
 71 other highly challenging task: building a simulation system  
 72 based on Modelica, an open-source, object-oriented, multi-  
 73 domain modeling language [Zhang *et al.*, 2024], following  
 74 the distillation system shown in Fig. 1. This simulation sys-  
 75 tem can operate under any combination of operational vari-  
 76 ables to generate the system’s output, which includes the per-  
 77 centage composition of the three substances: benzene (y<sub>1</sub>),  
 78 toluene (y<sub>2</sub>), and xylene (y<sub>3</sub>). By simulating all possible  
 79 samples, we obtained the corresponding simulation data, thus  
 80 forming the DIS-COL dataset. It is worth noting that this part  
 81 of the experiment aims to validate the effectiveness of our  
 82 model on the first type of industrial object (tangible industrial  
 83 equipment) described in this paper. Therefore, using simula-  
 84 tion data as a substitute for the costly real-world data does not  
 85 compromise the motivation of our experiments.

## ACE-FAC

87 The acentric factor ( $\omega$ ) is an unmeasurable dimensionless pa-  
 88 rameter used to describe the extent to which real gas deviate  
 89 from ideal gas behavior. It plays a significant role in ther-  
 90 modynamics and fluid mechanics [Biswas *et al.*, 2023]. Al-  
 91 though there are numerous methods for calculating the acen-  
 92 tric factor, such as the Edmister method [Edmister and Lee,  
 93 1984], Nath method [Nath *et al.*, 1976], and Lee-Kesler  
 94 method [Lee and Kesler, 1975], these mechanistic formulas  
 95 are derived from empirical data and thus have limitations in  
 96 their applicability. When the values fall within the range for  
 97 which these formulas are valid, the deviations are minimal;  
 98 however, outside this range, the deviations can be significant  
 99 [Abakporo, 2021]. Therefore, we select the acentric factor as  
 100 a representative of the second type of industrial object (ab-  
 101 stract variable) described in this paper and construct its twin  
 102 model to generate accurate acentric factor values. Among all  
 103 these methods, the Lee-Kesler method is the most commonly  
 104 used due to its simple structure and relatively accurate results.  
 105 Therefore, we adopt the Lee-Kesler method as the mechanis-  
 106 tic model for the acentric factor and use it as the benchmark  
 107 model in this study. Meanwhile, the functional mapping re-  
 108 lationships between variables in the Lee-Kesler method are  
 109 utilized to assist our UltraModel in predicting the real values  
 110 of the acentric factor. The Lee-Kesler method is presented as

follows:

111

$$\begin{aligned} \psi = & -\ln(P_c/1.01325) - 5.92714 + 6.09648/T_{br} + \\ & + 1.28862 \ln T_{br} - 0.169347T_{br}^6, \end{aligned} \quad (1)$$

$$\begin{aligned} \phi = & 15.2518 - 15.6875/T_{br} - 13.4721 \ln T_{br} + \\ & + 0.43577T_{br}^6, \end{aligned} \quad (2)$$

$$\begin{aligned} \rho = & -7.904 + 0.1352K_w - 0.007465K_w^2 + 8.359T_{br} + \\ & +(1.408 - 0.01063K_w)/T_{br}, \end{aligned} \quad (3)$$

$$\omega = \begin{cases} \frac{\psi}{\phi}, & T_{br} \leq 0.8, \\ \rho, & T_{br} > 0.8, \end{cases} \quad (4)$$

where  $P_c$ ,  $T_{br}$ , and  $K_w$  represent the critical pressure, re-  
 112duced boiling temperature, and characteristic factor, respec-  
 113tively. All of them can be calculated from the boiling point  
 114( $T_b$ ) and specific gravity ( $SG$ ), with the specific calculation  
 115formulas to be introduced later.

116

In this part of the experiment, a pseudo-component ap-  
 117proach [Riazi, 2005] is applied to petroleum with boiling  
 118points ranging from 300K to 1100K, resulting in pseudo-  
 119components characterized by their boiling point ( $T_b$ ) and spe-  
 120cific gravity ( $SG$ ). Each pseudo-component represents a sam-  
 121ple in this dataset, with the ground truth  $\omega_{gt}$  obtained using  
 122Aspen Plus, a famous and efficient simulation software [Liu  
 123*et al.*, 2022; Al-Malah, 2022]. The input for each sample  
 124consists of six variables  $\{T_b, SG, K_w, T_{br}, T_c, P_c\}$ , where  
 125the characteristic factors ( $K_w$ ), reduced boiling point ( $T_{br}$ ),  
 126critical temperature ( $T_c$ ), and critical pressure ( $P_c$ ) for each  
 127pseudo-component can be calculated from  $T_b$  and  $SG$  using  
 128the equations provided below [Riazi, 2005],

$$T_{br} = \frac{T_b}{T_c}, \quad (5)$$

$$K_w = \frac{(1.8T_b)^{1/3}}{SG}, \quad (6)$$

$$T_c = 19.06232 \times T_b^{0.58848} \times SG^{0.3596}, \quad (7)$$

$$P_c = 5.53027 \times 10^7 \times T_b^{-2.3125} \times SG^{2.3201}. \quad (8)$$

Model	$y_1$ (Benzene)			$y_2$ (Toluene)			$y_3$ (Xylene)		
	MAE	RMSE	R <sup>2</sup>	MAE	RMSE	R <sup>2</sup>	MAE	RMSE	R <sup>2</sup>
XGBoost	0.0118±0.0001	0.0302±0.0006	0.9800±0.0007	0.0128±0.0001	0.0290±0.0003	0.9748±0.0005	0.0044±0.0001	0.0142±0.0002	0.9584±0.0012
LightGBM	0.0130±0.0002	0.0329±0.0002	0.9761±0.0005	0.0145±0.0001	0.0317±0.0001	0.9698±0.0004	0.0049±0.0001	0.0157±0.0002	0.9492±0.0014
Transformer	<u>0.0074±0.0024</u>	<u>0.0131±0.0046</u>	<u>0.9957±0.0023</u>	<u>0.0072±0.0020</u>	<u>0.0128±0.0040</u>	<u>0.9946±0.0027</u>	0.0047±0.0006	0.0078±0.0009	0.9871±0.0027
GCN	0.0204±0.0009	0.0577±0.0007	0.9192±0.0045	0.0180±0.0007	0.0500±0.0012	0.9167±0.0058	0.0073±0.0006	0.0250±0.0004	0.8513±0.0128
GAT	0.0200±0.0008	0.0567±0.0005	0.9209±0.0005	0.0181±0.0006	0.0492±0.0003	0.9178±0.0011	0.0075±0.0009	0.0247±0.0008	0.8567±0.0075
RADA	0.0202±0.0009	0.0547±0.0038	0.9278±0.0123	0.0180±0.0005	0.0480±0.0028	0.9221±0.0126	0.0084±0.0009	0.0241±0.0013	0.8667±0.0189
DGDL	0.0139±0.0023	0.0271±0.0043	0.9828±0.0055	0.0128±0.0010	0.0266±0.0029	0.9773±0.0050	0.0067±0.0012	0.0159±0.0020	0.9459±0.0163
TGCN-S	0.0223±0.0013	0.0543±0.0018	0.9293±0.0052	0.0196±0.0016	0.0479±0.0017	0.9245±0.0057	0.0082±0.0010	0.0240±0.0006	0.8659±0.0097
UltraModel(Ours)	<b>0.0049±0.0012</b>	<b>0.0070±0.0015</b>	<b>0.9989±0.0005</b>	<b>0.0053±0.0022</b>	<b>0.0073±0.0024</b>	<b>0.9983±0.0009</b>	<b>0.0030±0.0007</b>	<b>0.0048±0.0006</b>	<b>0.9951±0.0014</b>

Table 2: The results of the three substances on the DIS-COL dataset, where the best results are highlighted in **bold**, and the second best scores are underlined.  $y_1$  represents benzene,  $y_2$  represents toluene, and  $y_3$  represents xylene.

Parameter Name	Setting
Optimizer	Adam [Kingma, 2014]
Learning rate	3e-3
Max training epochs	800
Dimension of hidden layer	128
Batchsize	4500
Dropout rate	0.1
# Graph convolution layers	2

Table 3: Parameter setting.

Model	DIS-COL			ACE-FAC		
	TT /s	IT /s	Paras	TT /s	IT /s	Paras
Transformer	<u>164</u>	<u>0.0410</u>	20003	92	0.405	<u>7921</u>
GCN	444	0.1583	17883	60	<b>0.0106</b>	9277
GAT	448	0.2157	18407	70	<u>0.0140</u>	9803
RADA	485	0.2091	18875	121	0.0231	10341
DGDL	457	0.2068	18259	74	0.0164	9481
TGCN-S	4720	1.7559	<u>17599</u>	<b>26</b>	0.0154	9023
UltraModel	<b>122</b>	<b>0.0389</b>	<b>10964</b>	<u>55</u>	0.0286	<b>2662</b>
Noise	DIS-COL			ACE-FAC		
	MAE	RMSE	R2	MAE	RMSE	R2
SNR=10	0.0150	0.0456	0.9258	0.0250	0.0745	0.9885
SNR=20	0.0086	0.0277	0.9725	0.0134	0.0349	0.9941
SNR=30	0.0060	0.0220	0.9827	0.0070	0.0136	0.9995

Table 4: **Upper part.** TT: training time on training set; IT: inference time on test set; Paras: total model parameters. All model configurations are consistent with the experiments shown in Table 1 of the main paper. **Lower part.** The performance of UltraModel on two datasets with added noise at different signal-to-noise ratios (SNR). The data presented in the table represents the mean of results from three random runs.

- RADA [Chen *et al.*, 2024] is a residual-aware deep attention graph convolutional network which uses a residual-aware connection module to reduce data uncertainty and alleviate over-smoothing.
- DGDL [Zhu and Zhao, 2022] uses a dynamic graph to realize adaptive learning and automatic inference and aggregate node feature representations by a multihop attention graph convolution network.
- TGCN-S [Kong *et al.*, 2023] emphasis lies in using a graph structure learning module to learn potential inter-variable relationships from data.

## C Additional Experimental Results

**C.1 More Experimental Results on the DIS-COL**

A twin model of the distillation system for separating benzene, toluene, and xylene mixtures is developed on the DIS-COL dataset. Given identical inputs (with operational variables having the same values), the error between the percentage composition of the three substances separated by the twin model and their real values reflects the performance of the twin model. Due to space limitations, only the visualized

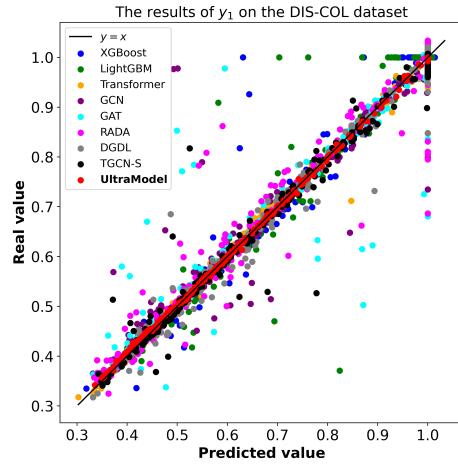
180 MAE and RMSE results of the predicted and real values for  
 181 the three substances across different models are presented in  
 182 the main text. Thus, more detailed experimental results are  
 183 provided in Table 2.

## 184 C.2 Visualization

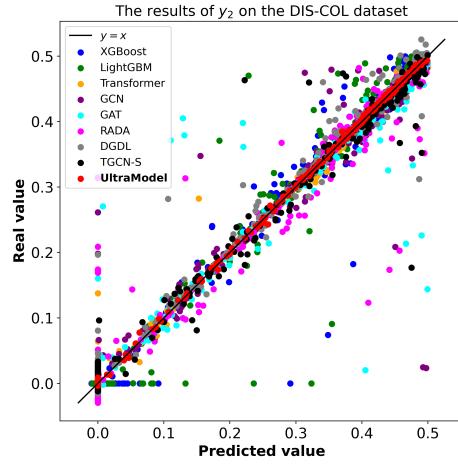
185 Due to space limitations, we only visualized the prediction  
 186 results for benzene on the DIS-COL dataset in the main text.  
 187 Here, we additionally visualize the prediction results for the  
 188 other two substances on the DIS-COL dataset, as well as  
 189 the prediction results on the ACE-FAC dataset. The visualiza-  
 190 tion results on the DIS-COL dataset are shown in Fig. 2.  
 191 In all three subplots, the x-axis represents the predicted val-  
 192 ues, while the y-axis represents the real values. Each subplot  
 193 displays the prediction results of all models for one of  
 194 the three substances. The closer the predicted values are to  
 195 the real values, the nearer the corresponding points fall to  
 196 the  $y = x$  line, indicating superior model performance. A  
 197 comparison reveals that our UltraModel consistently achieves  
 198 significantly better predictions for all three substances com-  
 199 pared to baseline models. The visualization results on the  
 200 ACE-FAC dataset are presented in Fig. 3, where the mean-  
 201 ings of the x-axis and y-axis are identical to those in Fig. 2.  
 202 Overall, apart from the benchmark mechanism model, the  
 203 Lee-Kesler method, which performs relatively poorly, other  
 204 models appear to perform reasonably well. However, when  
 205 we zoom in on a specific region, such as 0.5-1.0 (as shown  
 206 in the bottom-right corner of this figure), the differences in  
 207 model performance become evident. The predictions of our  
 208 UltraModel nearly align with the  $y = x$  line, whereas the  
 209 predictions of other baseline models exhibit noticeable devi-  
 210 ations. In summary, both the quantitative metrics results and  
 211 the qualitative visualization results demonstrate that our Ul-  
 212 traModel outperforms the baseline models and is capable of  
 213 modeling different types of industrial objects within a uni-  
 214 fied framework, echoing our initial aspiration to address such  
 215 challenges through a modeling paradigm.

## 216 C.3 Efficiency and Robustness

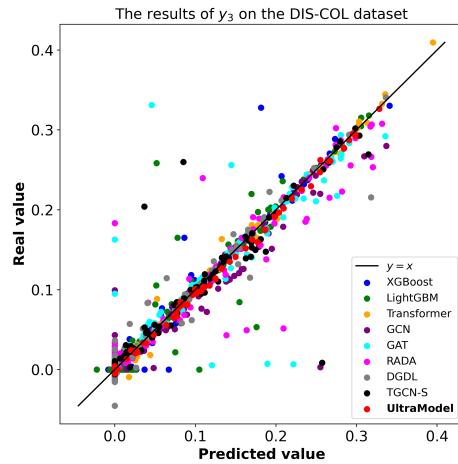
217 We further conducted experiments to evaluate the efficiency  
 218 of UltraModel and tested its robustness by adding noise to the  
 219 original signals. The experimental results are shown in Table  
 220 4. The upper part of the table compares the efficiency of Ul-  
 221 traModel with various baselines, while the lower part eval-  
 222 uates UltraModel’s performance under different levels of noise  
 223 interference. Specifically, as shown in the upper part, Ultra-  
 224 Model consistently has the fewest total parameters (Paras),  
 225 to some extent, indicating minimal resource usage; It also  
 226 achieves the shortest training time (TT) on DIS-COL and the  
 227 second shortest on ACE-FAC, demonstrating high training  
 228 efficiency and not requiring large retraining for deployment  
 229 in new environments; UltraModel has the shortest inference  
 230 time (IT) on the first dataset. Although its performance on the  
 231 second dataset is not excellent, its millisecond-level inference  
 232 speed should meet the needs of most industrial scenarios. As  
 233 for the lower part, compared with Table 1 in the main pa-  
 234 per, even with an SNR of 20, UltraModel outperforms most  
 235 baselines, demonstrating its ability to maintain high accuracy  
 236 under noisy conditions.



(a) Visualization of benzene prediction results  $y_1$  on the DIS-COL dataset.



(b) Visualization of toluene prediction results  $y_2$  on the DIS-COL dataset.



(c) Visualization of xylene prediction results  $y_3$  on the DIS-COL dataset.

Figure 2: Visualization results on DIS-COL dataset.

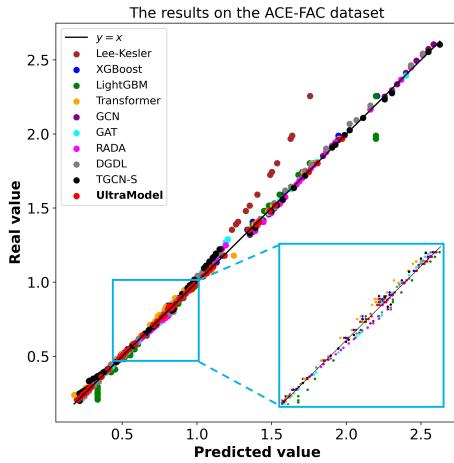


Figure 3: Visualization results on ACE-FAC dataset.

## References

- [238] [Abakporo, 2021] Obiechina I Abakporo. *General Thermo-dynamic Framework for Organic Rankine Cycles*. PhD thesis, Florida Agricultural and Mechanical University, 2021.
- [242] [Al-Malah, 2022] Kamal IM Al-Malah. *Aspen plus: chemical engineering applications*. John Wiley & Sons, 2022.
- [244] [Biswas *et al.*, 2023] Sayandep Biswas, Yunsie Chung, Josephine Ramirez, Haoyang Wu, and William H Green. Predicting critical properties and acentric factors of fluids using multitask machine learning. *Journal of Chemical Information and Modeling*, 63(15):4574–4588, 2023.
- [249] [Chen and Guestrin, 2016] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 785–794, 2016.
- [253] [Chen *et al.*, 2024] Yitao Chen, Yalin Wang, Qingkai Sui, Xiaofeng Yuan, Kai Wang, and Chenliang Liu. Residual-aware deep attention graph convolutional network via unveiling data latent interactions for product quality prediction in industrial processes. *Expert Systems with Applications*, 245:123078, 2024.
- [259] [Edmister and Lee, 1984] Wayne C Edmister and BK Lee. Applied hydrocarbon thermodynamics. vol. 1. 1984.
- [261] [Ke *et al.*, 2017] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30, 2017.
- [266] [Kingma, 2014] Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [269] [Kipf and Welling, 2016] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.
- [Kong *et al.*, 2023] Liyuan Kong, Chunjie Yang, Siwei Lou, Yu Cai, Xiaoke Huang, and Mingyang Sun. Collaborative extraction of intervariable coupling relationships and dynamics for prediction of silicon content in blast furnaces. *IEEE Transactions on Instrumentation and Measurement*, 72:1–13, 2023.
- [Lee and Kesler, 1975] Byung Ik Lee and Michael G Kesler. A generalized thermodynamic correlation based on three-parameter corresponding states. *AIChE Journal*, 21(3):510–527, 1975.
- [Liu *et al.*, 2022] Yanbing Liu, Xinglin Yang, Jiaqi Zhang, and Zongyuan Zhu. Process simulation of preparing biochar by biomass pyrolysis via aspen plus and its economic evaluation. *Waste and Biomass Valorization*, pages 1–14, 2022.
- [Nath *et al.*, 1976] Jagan Nath, SS Das, and ML Yadava. On the choice of acentric factor. *Industrial & Engineering Chemistry Fundamentals*, 15(3):223–225, 1976.
- [Riazi, 2005] MR Riazi. *Characterization and properties of petroleum fractions*, volume 50. ASTM international, 2005.
- [Vaswani, 2017] A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017.
- [Veličković *et al.*, 2017] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.
- [Zhang *et al.*, 2024] Yinan Zhang, Wenhui Wang, Qiang Yang, Xiaoyu Tang, Wei Ruan, Yingze Li, Surong Daoerji, Xiang Zhang, Yuqi Ye, Jiawei Huang, et al. Promoting digital twin technology application for process industry: A novel generation modelling platform and its implementations. *Digital Twins and Applications*, 1(1):51–74, 2024.
- [Zhu and Zhao, 2022] Kun Zhu and Chunhui Zhao. Dynamic graph-based adaptive learning for online industrial soft sensor with mutable spatial coupling relations. *IEEE Transactions on Industrial Electronics*, 70(9):9614–9622, 2022.