

Insights and analysis of machine learning for benzene hydrogenation to cyclohexene

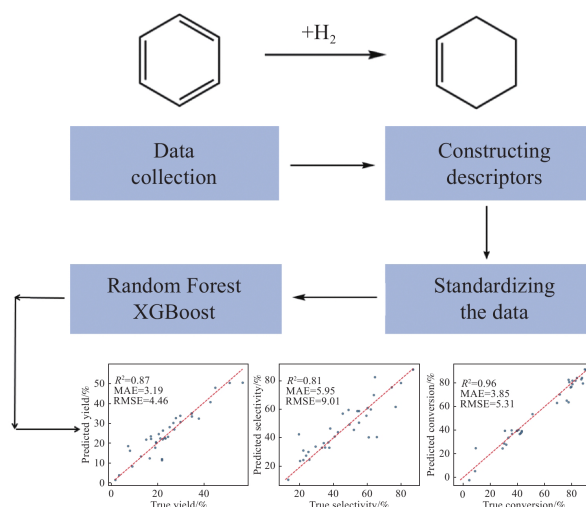
SUN Chao^{1,2}, ZHANG Bin^{1,*}

¹ State Key Laboratory of Coal Conversion, Institute of Coal Chemistry, Chinese Academy of Sciences, Taiyuan 030001, China;

² Center of Materials Science and Optoelectronics Engineering, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: Cyclohexene is an important raw material in the production of nylon. Selective hydrogenation of benzene is a key method for preparing cyclohexene. However, the Ru catalysts used in current industrial processes still face challenges, including high metal usage, high process costs, and low cyclohexene yield. This study utilizes existing literature data combined with machine learning methods to analyze the factors influencing benzene conversion, cyclohexene selectivity, and yield in the benzene hydrogenation to cyclohexene reaction. It constructs predictive models based on XGBoost and Random Forest algorithms. After analysis, it was found that reaction time, Ru content, and space velocity are key factors influencing cyclohexene yield, selectivity, and benzene conversion. Shapley Additive Explanations (SHAP) analysis and feature importance analysis further revealed the contribution of each variable to the reaction outcomes. Additionally, we randomly generated one million variable combinations using the Dirichlet distribution to attempt to predict high-yield catalyst formulations. This paper provides new insights into the application of machine learning in heterogeneous catalysis and offers some reference for further research.

Key words: machine learning; heterogeneous catalysis; hydrogenation of benzene; XGBoost



Cyclohexene is an important raw material for nylon production. Selective hydrogenation of benzene is a key method for preparing cyclohexene^[1]. Compared to the partial hydrogenation of benzene to cyclohexane, the complete hydrogenation of benzene is a thermodynamically more favorable process (Figure 1). However, the challenge in this field lies in controlling the kinetics to improve the yield of cyclohexene. Researchers have gradually enhanced the selectivity for cyclohexene by modifying catalyst structures and reaction processes. Although this technology was industrialized over half a century ago, current industrial processes still rely on metal salts in aqueous systems, high Ru loading in the catalysts, and low cyclohexene yield.

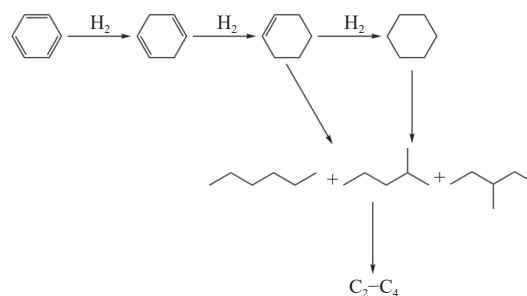


Fig. 1 Reaction pathway of benzene hydrogenation over metal catalysts showing stepwise hydrogenation to cyclohexane and possible ring-opening/fragmentation products

Received: 2025-09-01; Revised: 2025-09-25

Foundation items: Supported by CAS Basic and Interdisciplinary Frontier Scientific Research Pilot Project (XDB1190300, XDB1190302), Youth Innovation Promotion Association CAS (Y2021056), Joint Fund of the Yulin University and the Dalian National Laboratory for Clean Energy (YLU-DNL Fund 2022007) and The special fund for Science and Technology Innovation Teams of Shanxi Province (202304051001007).

* Corresponding author. E-mail: zhangbin2009@sxicc.ac.cn.

DOI: 10.1016/S1872-5813(25)60608-6 CSTR: 32157.14.2097-213X.2025.JFCT.0153 Associated Data DOI: 10.57760/sciencedb.27385

本文的英文电子版由 Elsevier 出版社在 ScienceDirect 上出版 (<http://www.sciencedirect.com/science/journal/18725813>)

Combined with extensive single-factor experiments, traditional methods typically reveal patterns by controlling additives and supports. For benzene hydrogenation, the yield of cyclohexene increased through continuous attempts and adjustments to the mechanistic investigations and catalyst compositions. In 1934, Horiuti and colleagues proposed the stepwise hydrogenation mechanism of partial benzene hydrogenation by Group VIII metals^[2]. In this mechanism, adsorbed activated benzene and dissociatively adsorbed hydrogen proceed through intermediates such as cyclohexa-1,4-diene and cyclohexene, ultimately producing cyclohexane. Subsequently, Horiuti et al.^[2-3] proposed that the selective hydrogenation of benzene occurs through two distinct processes. One is the stepwise hydrogenation, where the π electrons of benzene form a σ bond with the unbonded d electrons of the metal, and hydrogen is dissociatively adsorbed. The adsorbed benzene and dissociatively adsorbed hydrogen undergo a stepwise hydrogenation process, ultimately producing cyclohexane. The other process involves the direct hydrogenation of benzene to cyclohexane, where planar benzene, adsorbed via its large π bond, interacts with six hydrogen atoms to form a van der Waals-activated complex, leading to the direct formation of cyclohexane^[3-5]. Generally, under low-pressure conditions, the reaction follows the first process, while under high-pressure reaction conditions, the reaction favors the second process. Over the past century, various methods have been attempted to increase the yield of cyclohexene in benzene hydrogenation reactions, including experimenting with different supports, adding multiple additives, altering the size and shape of catalysts, and using DFT calculations to predict performance^[1,6-12]. Through years of research, the system, which utilizes an aqueous zinc sulfate additive in combination with a Ru catalyst, has been shown to improve the selectivity for cyclohexene occasionally. However, the yield of cyclohexene remains limited to around 60%.

Machine learning, a crucial tool in the development of artificial intelligence, has been increasingly applied to various scientific fields in recent years due to its ability to uncover patterns and correlations between different datasets efficiently. In recent years, machine learning has also attracted the attention of chemical researchers. For instance, Liu et al.^[13] utilized neural networks based on experimental data to elucidate the crucial roles of reaction conditions, pretreatment methods, and post-reaction iron phase compositions. They demonstrated that controlling the iron phase ratio and adding appropriate promoters can effectively suppress methane selectivity. Shi et al.^[14] utilized extensive literature data and combined machine learning algorithms, including random forests and extreme gradient boosting, to successfully build a model for predicting the yield and selectivity of 1-octene hydrogenation, which was experimentally validated. Han et al.^[15] utilized machine learning to aid in designing electrocatalytic materials for high-performance lithium-sulfur batteries. The review by Jaison et al.^[16] discusses the

application of machine learning in the field of photocatalysis, with a particular focus on its role in environmental science. In these fields, machine learning has demonstrated certain potential and has been applied to solve problems in catalysis. However, in traditional thermal catalysis, predicting the conversion, selectivity, and yield by using machine learning algorithms presents significant challenges due to the difficulty of quantifying data from experimental processes and achieving consistency across datasets. Although there is a wealth of data left by previous researchers, most of this data is challenging to use as standard datasets for model building. Additionally, thermal catalysis reactions tend to have complex conditions and generally long reaction time, making it difficult to generate a large amount of data through self-conducted experiments.

Over the years, considerable data have been accumulated on the partial hydrogenation of benzene to cyclohexane. Given this context, we aim to utilize the existing experimental data and employ innovative methods to analyze the factors influencing the benzene conversion, selectivity, and yield of cyclohexene in benzene hydrogenation. Specifically, we selected high-quality data with a balanced distribution from numerous literature sources on the hydrogenation of benzene. After organizing the data into a dataset, we built models using Random Forest and XGBoost algorithms to predict the conversion, selectivity, and yield. Additionally, we seek to predict high-performance catalyst compositions and reaction conditions.

1 Results and discussion

1.1 Construction of databases

Before collecting data, it is essential to identify which descriptors are easy to standardize and are important for the reaction. The hydrogenation of benzene has been extensively studied by numerous researchers over the years, with some efforts focusing on the addition of less common additives or supports to regulate the yield of a specific product. For our purposes, these uncommon factors could potentially act as interfering variables in the model. Therefore, by reviewing nearly 50 relevant papers, we selected those that fully meet the established descriptors and have a substantial number of data points. This approach ensures the reliability of the original dataset while minimizing interference^[8,17-25].

In this study, benzene conversion (x), cyclohexene selectivity (s) and cyclohexene yield (Y) are defined as follows:

$$x_{C_6H_6} (\%) = \frac{n_{C_6H_6,in} - n_{C_6H_6,out}}{n_{C_6H_6,in}} \times 100\% \quad (1)$$

$$s_{C_6H_8} (\%) = \frac{n_{C_6H_8,formed}}{\sum n_{products\ from\ C_6H_6}} \times 100\% \quad (2)$$

$$Y_{C_6H_8}(\%) = \frac{n_{C_6H_8, \text{formed}}}{n_{C_6H_6, \text{in}}} \times 100\% \quad (3)$$

The descriptors encompass 25 factors (Table 1), including reaction temperature, reaction pressure, reaction time, preparation method, and catalyst composition, among others. It should be noted that some catalysts were not subjected to a reduction process. For these data points, we set their reduction temperature to 298 K to ensure the integrity of the dataset. Detailed information on the dataset is presented in the Supporting Information.

The “Preparation” column in the descriptors, which records three preparation methods, was one-hot encoded. The remaining descriptors were standardized and normalized based on their value ranges and numerical characteristics. Subsequently, XGBoost^[26] and Random Forest^[27] models were trained to generate prediction results. The specific code is documented in the Supporting Information.

In brief, we first preprocess the dataset as described, then split it into training and testing sets, with 80% for training and 20% for testing, using a random seed of 42. We then perform feature scaling and combine all features.

Table 1 Variable summary of the dataset

Variable	Possible values
Input	
Benzene/water (mL/mL)	0.175–0.5
Benzene/catalyst (mL/mg)	0.05–0.3846
Pretreatment temperature/K	298–873
Reduction temperature/K	298–573
Hydrogen concentration/mol%	0/5/99.9
Reduction time/h	0–3
Reaction temperature/K	373–453
Reaction time/min	1–720
Reaction pressure/MPa	2–7
$S_{\text{BET}}/(\text{m}^2 \cdot \text{g}^{-1})$	8.4–318
$d_{\text{pore}}/\text{nm}$	2.63–42.2
$v_{\text{pore}}/(\text{cm}^3 \cdot \text{g}^{-1})$	0.03–1.05
VHSV/(mL·mg ⁻¹ ·h ⁻¹)	0.00058–11.538
Ru/%	0.18–20
Cu/%	0–1.26
ZnO/%	0–97.93
NaOH/%	0–93.26
ZnSO ₄ /%	0–90.99
AlOOH/%	0–43.68
SiO ₂ /%	0–43.68
TiO ₂ /%	0–97.1
Output	
Selectivity/%	0.7–89.3
Yield/%	0.3–60.64
Conversion/%	4.04–98.58

Next, we define the XGBoost regression model using GridSearchCV for hyperparameter tuning, cross-validation, and model training, with a separate model trained for each target variable. Finally, we perform model prediction and evaluation.

For model evaluation, we use R-squared (R^2), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE).

R^2 measures the extent to which the model explains the variability in the data, representing the correlation between the predicted and actual values. Its value ranges between [0, 1], with values closer to 1 indicating a better fit. The calculation formula is as follows:

$$R^2 = 1 - \frac{\sum (y_{\text{true}} - y_{\text{pred}})^2}{\sum (y_{\text{true}} - \bar{y})^2} \quad (4)$$

MAE measures the average absolute difference between predicted and actual values. It is a linear metric that is not influenced by outliers. A smaller MAE indicates a lower model's prediction error, and the model performs better. The units of MAE are the same as those of the target variable, which helps in understanding the practical significance of the error. MAE is less sensitive to outliers because it does not square the errors. However, it may not accurately reflect the distribution of errors, as it does not impose a more significant penalty for larger errors. The calculation formula is as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_{\text{true},i} - y_{\text{pred},i}| \quad (5)$$

RMSE measures the square root of the average of the squared differences between the predicted values and the actual values. RMSE is more sensitive to larger errors because it squares the differences, meaning larger errors receive a more significant penalty. A lower RMSE indicates a more minor prediction error in the model, and the model performs better. Because of the squaring process, RMSE is particularly sensitive to significant prediction errors (such as outliers). The calculation formula is as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{\text{true},i} - y_{\text{pred},i})^2} \quad (6)$$

1.2 Model building and analysis

The advantage of Random Forest lies in its simplicity, high robustness, and low risk of overfitting, making it suitable for rapid model development with small to medium-sized datasets^[27]. XGBoost, on the other hand, offers high prediction accuracy, strong regularization capabilities, and excellent support for handling missing values and coefficient data^[26,28]. Often, a Random Forest can be quickly trained as a baseline, and then XGBoost can be tried to achieve even higher performance. We used both algorithms to predict the target reaction's conversion, selectivity, and yield, which were evaluated using R^2 , MAE and RMSE (Figure 2).

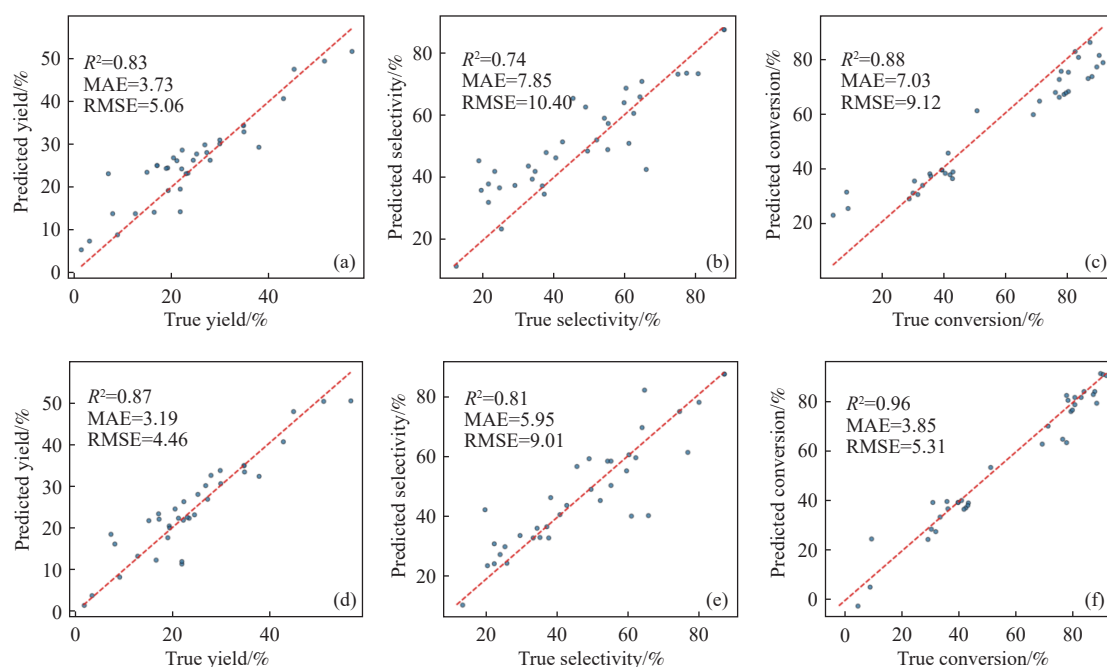


Fig. 2 Scatter plot of the prediction performance on the Random Forest and XGBoost test set: RF model performance in predicting the yield of cyclohexene (a), the selectivity of cyclohexene (b), and the conversion of benzene (c), XGBoost performance for cyclohexene yield (d), selectivity (e), and conversion (f) for benzene hydrogenation over the Ru catalysts

The results show that the overall prediction performance of XGBoost is better than that of Random Forest (the closer the blue dots are to the red line, the better the prediction performance). For the yield of cyclohexene, XGBoost achieved an R^2 of 0.87, MAE of 3.19, and RMSE of 4.46; for selectivity, $R^2 = 0.81$, MAE = 5.95, and RMSE = 9.01; and for conversion, $R^2 = 0.96$, MAE = 3.85, and RMSE = 5.31, indicating that the model has relatively high accuracy and stability for this task.

In the prediction of selectivity, the prediction errors of both models are larger than those for conversion and yield, which may be due to this variable being more influenced by other hidden factors that the model finds difficult to capture fully. Conversion is the easiest variable to predict, indicating that decision trees learn their variation pattern more readily. XGBoost's superior performance compared to Random Forest may be attributed to its ability to continuously adjust weights and optimize residuals during training, thereby uncovering deeper relationships in the dataset and reducing errors.

We use a heatmap to examine which variables have strong linear relationships and which do not. Since the heatmap of more than twenty descriptors is quite complex, it is included in the Supporting Information. Some of the strong correlations may require explanation. For example, the strong negative correlation between zinc sulfate and hydrogen reduction concentration and reduction time may be due to the fact that systems with a large amount of zinc sulfate do not require the reduction step in the catalyst preparation process. The hydrogen reduction time and concentration in the dataset were recorded as 0 for reactions

that do not require reduction, leading to the high correlation observed in the heatmap.

Figure 3 presents the SHAP plot analysis and feature importance analysis for each target variable. These three target variables are yield, selectivity, and conversion, respectively. The SHAP plot shows the average contribution of each feature to the model output (including both positive and negative effects)^[28]. It captures the interaction effects between features, while the feature importance plot indicates the frequency or other simple metrics on how often a feature is used in decision trees. The influence of features on the target variable in this model decreases from top to bottom. The features at the top typically have the most significant impact on the prediction outcome. The larger the SHAP value (on the horizontal axis), the greater the positive contribution of the feature to predicting the target variable for this data point. Conversely, a smaller SHAP value indicates a negative influence.

Figures 3(a) and 3(b) show that the Ru content, in the range of 0.18% to 20%, significantly impacts both yield and selectivity. As the Ru content increases, it negatively affects both the yield and selectivity of cyclohexene in benzene hydrogenation while having a minor impact on the conversion rate. The reaction time has a significant effect on all three target variables. However, the relationship is not linear and straightforward, because the selective hydrogenation of benzene to cyclohexene typically increases initially with time and then gradually decreases, which is consistent with findings reported in other literature^[1,9–10,23–24].

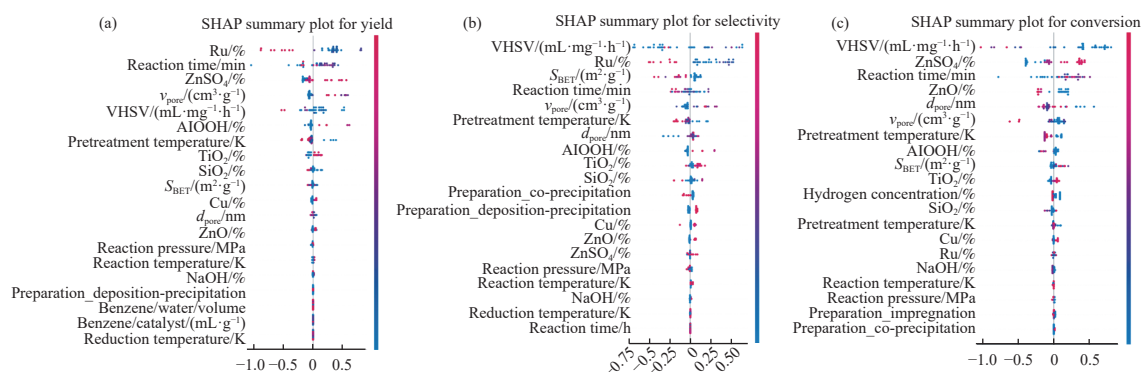


Fig. 3 SHAP analysis plot for target variables obtained using XGBoost: SHAP summary plot for yield (a), selectivity (b) and conversion (c) (In each row, blue dots represent samples with lower feature values, while red dots represent higher ones)

The content of ZnSO_4 in the reaction has a positive impact on all three target variables, meaning that within the range of our dataset, the addition of ZnSO_4 can increase the yield, selectivity, and conversion rate of the reaction.

VHSV reflects the relationship between the feed flow rate and the catalyst volume. It is commonly used to describe the rate at which fluid passes through a catalyst, especially in gas-solid or liquid-solid catalytic reactions. A higher VHSV value indicates a larger flow rate per unit time and a shorter contact time with the catalyst, while a lower VHSV value means a longer contact time between the reactants and the catalyst. From the overall trend, low space velocity has a positive impact. However, the relationship with selectivity is not simple, although space velocity significantly influences the reaction's selectivity. This situation may be because space velocity is linked to reaction time.

We can infer several insights from the feature importance analysis plot provided by XGBoost (Figure 4). For example, reaction time is the most critical factor for the yield of cyclohexene, while variations in space velocity, reaction time, and Ru content are relatively crucial for the selectivity of cyclohexene. Incorporating insights from the SHAP plot, it can be concluded that controlling an appropriate reaction time, maintaining a lower space velocity, limiting Ru content, and managing the type and ratio of supports and additives can help improve the cyclohexene selectivity. A higher pore volume also contributes positively to the reaction.

The above analysis can guide the development of new catalysts. We randomly generated variables, setting the variation ranges for each variable based on the conclusions obtained earlier. The Dirichlet distribution method was used to ensure that the total sum of the catalyst components remained 100%. A total of 1000000 sets of variables were randomly sampled within the defined range, and these were input into the trained model to predict several sets of high-yield values. The input values corresponding to these high-yield predictions are also presented. In addition, we have included a section of code (supplementary information) at the end to allow for manual input of values, which can then be used to predict the yield. When the key parameters are

set to Ru content = 1.6%, TiO_2 content = 13.5%, ZnSO_4 content = 84.9%, reaction temperature = 413 K, reaction time = 20 min and reaction pressure = 5.0 MPa, the model predicts a cyclohexene yield of 60.6%.

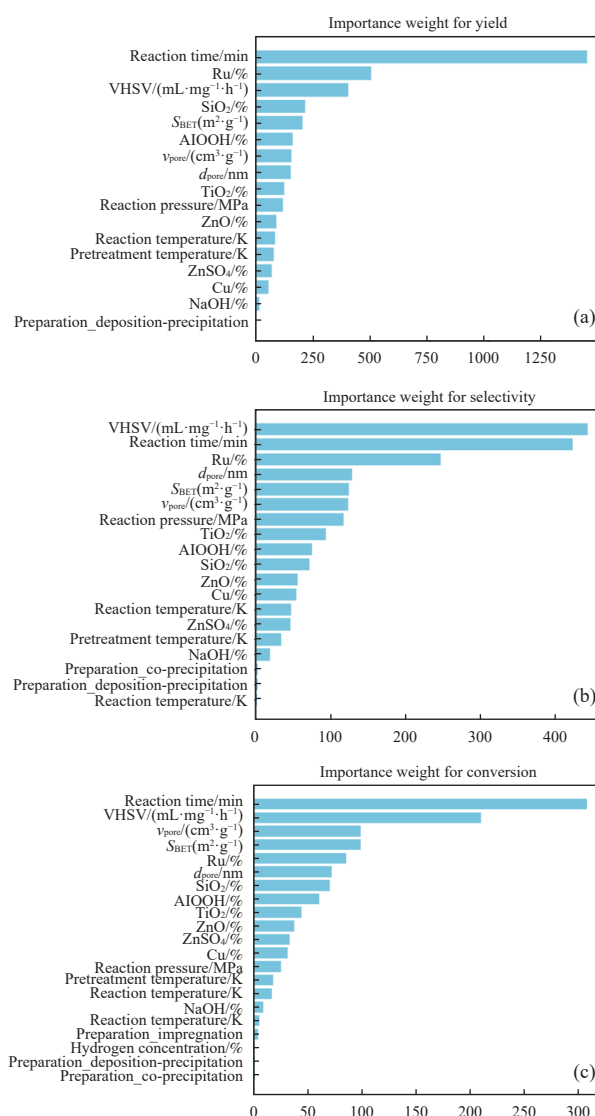


Fig. 4 Feature importance analysis plots for the three target variables using XGBoost: (feature Importance for yield (a), selectivity (b) and conversion (c))

2 Conclusions

This paper collects data from existing literature and uses XGBoost and Random Forest algorithms to predict and analyze the yield, selectivity, and conversion of selective benzene hydrogenation to cyclohexene. SHAP plots and feature importance analysis indicate that controlling an appropriate reaction time, maintaining a lower space velocity, limiting Ru content, and selecting suitable supports and additives can increase the yield of cyclohexene. These models are used to predict unexplored catalyst combinations and reaction conditions. This work can also serve as a reference for the application of machine learning in heterogeneous thermal catalysis. At the same time, it is worth noting that machine learning faces numerous challenges when predicting complex systems, such as heterogeneous thermal catalysis. Achieving high-precision predictions and predicting a reasonable, practical catalyst system is challenging.

Abbreviations

SHAP Shapley additive explanations
 DFT Density functional theory
 XGBoost Extreme gradient boosting
 S_{BET} Specific surface area determined by the Brunauer-Emmett-Teller method
 d_{pore} Average pore diameter
 V_{pore} Total pore volume determined
 VHSV Volumetric hourly space velocity
 RF Random forest
 R^2 R-squared
 MAE Mean absolute error
 RMSE Root mean square error

Declarations

Ethics and Consent to Participate

Not applicable.

Consent for Publication

Not applicable.

Declaration of Competing Interest

The authors declare no conflict of interest.

Author Contribution

C. Sun wrote the main manuscript text and prepared figures. B. Zhang reviewed and modified the manuscript.

Availability of data and materials

Supplementary data associated with this article can be found in the online version at <https://github.com/ttoono->

[takaki/Code-and-dataset.git](https://github.com/ttoono-takaki/Code-and-dataset.git). and <https://doi.org/10.57760/sciencedb.27385>.

References

- [1] CHEN Z, SUN H, PENG Z, et al. Selective hydrogenation of benzene: Progress of understanding for the Ru-based catalytic system design[J]. *Ind Eng Chem Res*, 2019, 58: 13794–13803.
- [2] HORIUTI I, POLANYI M. Exchange reactions of hydrogen on metallic catalysts[J]. *Trans Faraday Soc*, 1934, 30: 1164–1172.
- [3] OTSUKA K, HATAN M. The catalysts for the synthesis of formaldehyde by partial oxidation of methane[J]. *J Catal*, 1987, 108: 252–255.
- [4] JENKINS G I, RIDEAL E. The catalytic hydrogenation of ethylene at a nickel surface. Part I. The chemisorption of ethylene[J]. *J Chem Soc*, 1955, 2490–2496.
- [5] JENKINS G I, RIDEAL E. The catalytic hydrogenation of ethylene at a nickel surface. Part II. The reaction mechanism[J]. *J Chem Soc*, 1955, 2496–2500.
- [6] FAN C, ZHU Y-A, ZHOU X-G, et al. Catalytic hydrogenation of benzene to cyclohexene on Ru(0001) from density functional theory investigations[J]. *Catal Today*, 2011, 160: 234–241.
- [7] JIANG L, DONG Y, ZHOU G, et al. Promoting the performances of TiO_2 submicrosphere-embedded Ru nanoparticles in benzene selective hydrogenation by morphology manipulation[J]. *Ind Eng Chem Res*, 2019, 59: 1083–1092.
- [8] LIU H, LIANG S, WANG W, et al. The partial hydrogenation of benzene to cyclohexene over Ru-Cu catalyst supported on ZnO [J]. *J Mol Catal A: Chem*, 2011, 341: 35–41.
- [9] SILVEIRA E T, UMPIERRE ALEXANDRE P, ROSSI LIANE M, et al. The partial hydrogenation of benzene to cyclohexene by nanoscale ruthenium catalysts in imidazolium ionic liquids[J]. *Chem*, 2004, 10: 3734–3740.
- [10] SUN H, JIANG H, LI S, et al. Effect of alcohols as additives on the performance of a nano-sized Ru-Zn(2.8%) catalyst for selective hydrogenation of benzene to cyclohexene[J]. *Chem Eng J*, 2013, 218: 415–424.
- [11] UTELBAEVA A B, ERMAKHANOV M N, ZHANABAI N Z, et al. Hydrogenation of benzene in the presence of ruthenium on a modified montmorillonite support[J]. *Russ J Phys Chem A*, 2013, 87: 1478–1481.
- [12] YU X L, LI Y, XIN S M, et al. Partial hydrogenation of benzene to cyclohexene on Ru@XO_2 ($x = \text{Ti, Zr, or Si}$)[J]. *Ind Eng Chem Res*, 2018, 57: 1961–1967.
- [13] LIU Y, ZHANG X, LI L, et al. Machine learning insights into catalyst composition and structural effects on CH_4 selectivity in iron-based Fischer Tropsch synthesis[J]. *Artif Intell Chem*, 2024, 2: 100062–100070.
- [14] SHI H, SHEN C, HUANG Z, et al. Machine learning-guided prediction of hydroformylation[J]. *Chemphyschem*, 2025, 26: e202400773–e202400778.
- [15] HAN Z, CHEN A, LI Z, et al. Machine learning-based design of electrocatalytic materials towards high-energy

- lithium||sulfur batteries development[J]. Nat Commun, 2024, 15, 8433.
- [16] JAISON A, MOHAN A, LEE Y C. Machine learning-enhanced photocatalysis for environmental sustainability: Integration and applications[J]. Mater Sci Eng: R: Reports, 2024, 161.
- [17] ZHOU G, DONG Y, JIANG L, et al. Effect of support composition on the structural and catalytic properties of Ru/AlOOH-SiO₂ catalysts for benzene selective hydrogenation[J]. Catal Sci Technol, 2018, 8: 1435–1446.
- [18] WU T, ZHANG P, JIANG T, et al. Enhancing the selective hydrogenation of benzene to cyclohexene over Ru/TiO₂ catalyst in the presence of a very small amount of ZnO[J]. Sci China: Chem, 2014, 58: 93–100.
- [19] HAO F, ZHENG J S, OUYANG D H, et al. Selective hydrogenation of benzene over Ru supported on surface modified TiO₂[J]. Korean J Chem Eng, 2021, 38: 736–746.
- [20] ZHOU G, WANG F, SHI R. Nanoparticulate Ru on morphology-manipulated and Ti³⁺ defect-riched TiO₂ nanosheets for benzene semi-hydrogenation[J]. J Catal, 2021, 398: 148–160.
- [21] ZHOU G B, DOU R F, BI H Z, et al. Ru nanoparticles on rutile/anatase junction of P25 TiO₂: Controlled deposition and synergy in partial hydrogenation of benzene to cyclohexene[J]. J Catal, 2015, 332: 119–126.
- [22] YAN X, ZHANG Q, ZHU M, et al. Selective hydrogenation of benzene to cyclohexene over Ru-Zn/ZrO₂ catalysts prepared by a two-step impregnation method[J]. J Mol Catal A: Chem, 2016, 413: 85–93.
- [23] WANG Z, ZHANG Q, LU X, et al. Ru-Zn catalysts for selective hydrogenation of benzene using coprecipitation in low alkalinity[J]. Chin J Catal, 2015, 36: 400–407.
- [24] SUN H J, WANG H X, JIANG H B, et al. Effect of (Zn(OH)₂)₃(ZnSO₄)(H₂O)₅ on the performance of Ru-Zn catalyst for benzene selective hydrogenation to cyclohexene[J]. Appl Catal A: Gen, 2013, 450: 160–168.
- [25] XUE X G, LIU J, RAO D M, et al. Double-active site synergistic catalysis in Ru-TiO₂ toward benzene hydrogenation to cyclohexene with largely enhanced selectivity[J]. Catal Sci Technol, 2017, 7: 650–657.
- [26] SHERIDAN R P, WANG W M, LIAW A, et al. Extreme gradient boosting as a method for quantitative structure-activity relationships[J]. J Chem Inf Model, 2016, 56: 2353–2360.
- [27] BREIMAN L. Random forests[J]. Mach Learn, 2001, 45: 5–32.
- [28] LUNDBERG S M, LEE S I. In proceedings of the 31st international conference on neural information processing systems[C]. 2017, 4768–4777.

苯加氢生成环己烯的机器学习研究与分析

孙 超^{1,2}, 张 斌^{1,*}

(1. 中国科学院山西煤炭化学研究所 煤炭高效低碳利用全国重点实验室, 山西 太原 030001;

2. 中国科学院大学 材料科学与光电工程中心, 北京 100049)

摘 要: 环己烯是尼龙生产中的重要原料, 而苯的选择性加氢是制备环己烯的关键途径。当前工业过程中使用的钌基催化剂仍面临金属用量高、工艺成本高以及环己烯产率低等问题。本研究利用已有文献数据结合机器学习方法, 对苯加氢生成环己烯反应中影响转化率、选择性和产率的因素进行了分析, 并基于 XGBoost 与随机森林算法构建了预测模型。分析结果表明, 反应时间、Ru 含量及空速是影响反应产率、选择性和转化率的关键因素。通过 SHAP(Shapley 加性解释)分析及特征重要性分析, 进一步揭示了各变量对反应结果的贡献。研究利用 Dirichlet 分布随机生成了 100 万组变量组合, 用于预测高产催化剂配方。本研究为机器学习在多相催化领域的应用提供了新的视角, 并为后续研究提供了参考。

关键词: 机器学习; 多相催化; 苯加氢; XGBoost

中图分类号: O643; TP181.1

文献标识码: A

作者简介:



孙超(2001), 硕士研究生, 研究方向为多相催化, sunchao23@mails.ucas.ac.cn。



张斌(1986), 博士, 研究员, 研究方向为原子层沉积及能源催化, zhangbin2009@sxicc.ac.cn。