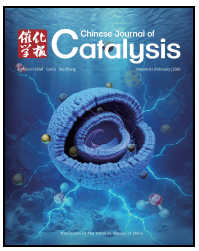
Chinese Journal of Catalysis 81 (2026) 124–135

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.sciencedirect.com/journal/chinese-journal-of-catalysis



Article

Machine-learning-aided discovery of methanol-to-olefins zeolite catalysts with ultra-high initial selectivity

Xinyi Wang [a,d,1], Chaoqi Wang [b,1], Miao Yang [a,*], Xiaoguang Wang [b,*], Yuezhong Zuo [b], Zhuangzhuang Zhang [b], Yimo Wu [a,e], Jingfeng Han [a], Bing Li [a], Wei Huang [c], Limin Ren [c], Yingxu Wei [a], Xinmei Liu [d], Peng Tian [a], Zhongmin Liu [a,e,*]

[a] *National Engineering Research Center of Lower-Carbon Catalysis Technology, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, Liaoning, China*

[b] *School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, Liaoning, China*

[c] *State Key Laboratory of Fine Chemicals, School of Chemistry, Dalian University of Technology, Dalian 116024, Liaoning, China*

[d] *State Key Laboratory of Heavy Oil and College of Chemistry and Chemical Engineering, China University of Petroleum, Qingdao 266580, Shandong, China*

[e] *University of Chinese Academy of Sciences, Beijing 100049, China*

ARTICLE INFO

Article history:
Received 15 August 2025
Accepted 22 September 2025
Available online 5 February 2026

Keywords:
Methanol to olefins
Zeolite catalyst
Machine learning
Rational design and synthesis
Prediction

ABSTRACT

With the continuous advancement of the industrialized methanol-to-olefins (MTO) process and a profound understanding of its mechanism, designing MTO catalysts to enhance light olefin yields and flexibly regulate product distribution has emerged as a significant challenge. Data-driven modeling allows chemists to anticipate reaction trends and outcomes. However, for models to be instructive for specific chemical issues, chemists must collect experimental data, encode the relevant variables and retrain specialized models. In this work, we demonstrate how to use a machine learning (ML) workflow to discover a potential MTO zeolite catalyst. An MTO database was built, on which over 20 types of ML models were trained, followed by their evaluation and experimental validation. The decision rules for high selectivity were extracted, facilitating the targeting of potential MTO catalysts and the understanding of MTO reaction mechanism. A rapid prediction of optimal MTO evaluation conditions and results for a given zeolite catalyst was also realized, greatly saving the cost of trial and error. In particular, a special MTO catalyst with high initial ethene selectivity over 60% was found, demonstrating the effectiveness and capability of ML techniques.

1. Introduction

Molecular sieve-catalyzed chemical reactions and industrial processes have created enormous economic value in industrial manufacturing and daily life [1,2], but the rational design and synthesis of molecular sieves for their catalytic applications remains a major challenge due to the lack of theoretical guidance. As a prime example, the methanol-to-olefins (MTO) reaction catalyzed by the molecular sieve catalyst represents an important catalytic pathway for the production of light olefins

* Corresponding author. E-mail: yangmiao@dicp.ac.cn (M. Yang), wangxg@dlut.edu.cn (X. Wang), liuzm@dicp.ac.cn (Z. Liu).

[1] These authors contributed equally to this work.

This work was supported by the National Key Research and Development Program of China (2024YFE0207000), National Natural Science Foundation of China (22171259, 12461050, 21991090, 21991091, 22288101, 22372020), and the AI S&T Program of Yulin Branch, Dalian National Laboratory for Clean Energy, CAS (DNL-YL A202206).
https://doi.org/10.1016/S1872-2067(25)64903-5

from non-petroleum sources [3,4]. Since the world's first commercial MTO plant was put into operation in China in 2010, the total effective capacity of the sectors has exceeded 20 million tons/year [5]. The regulation of light olefin selectivity and catalyst lifetime has reached an optimal balance that is difficult to break through [6]. However, it is still highly desirable to modify the product distribution according to the fluctuating market demand and retard the carbon deposition rate, so that the economic efficiency of MTO technology can be further improved, and the goals of carbon peaking and carbon neutralization can be realized [7].

MTO is a typical shape-selective catalytic reaction. The three classes of shape-selectivity, including selection of the reactants, transition states, and products are all strongly influenced by the catalyst used. An extensive screening of MTO catalysts with different topologies (channels/cavity networks and dimensions; window openings providing access to internal cavities), compositions (number and distribution of acidic sites, defects), and morphologies (crystal dimensions, micro- and mesoporosity) has been performed over the last four decades [8]. Interestingly, the selectivity limits are difficult to overcome by influences other than changing the zeolite topology [9–11]. As verified, medium-porous 10-membered ring (MR) MFI (ZSM-5) [12,13] and large-porous 12-MR BEA [14,15] and CON [16] catalysts can obtain high propene selectivity by modifying the acidity and increasing the reaction temperature to intensify the olefins-based cycle and cracking reaction. But the yield of ethene is always much lower than 8-MR zeolites such as SAPO-34 [17], SAPO-18 [18], SSZ-13 [19], and SSZ-39 [19] due to the insufficient shape selectivity on light olefins by 10- and 12-MR windows. The shape and size of 8-MR pore openings of microporous zeolite catalysts affect the distribution of $C_2^=$, $C_3^=$ and $C_4^=$ light olefins also [20]. In addition, a suitable reaction space is required for a zeolite catalyst to form and accommodate hydrocarbon pool intermediates (HCPs), and the HCPs types are responsible for the different MTO activity, pathway, and product selectivity [21–24]. It is the so-called autocatalysis and "hydrocarbon pool" mechanism [25]. To develop a structure-property relationship between the framework topology and the MTO light olefin product distribution, a geometric concept of cage-defining ring size has been proposed, which shows a strong correlation to the light olefin product distributions with the sizes of polymethyl aromatic intermediate species [9]. As an interesting example, a 8-MR SAPO-14 with a narrow 10-MR sized AFN cage exhibits unexpectedly highest record of propene selectivity [26,27]. The ultra-narrow 10-MR like AFN cage and weak acidity allow the MTO reaction occur in an olefins-based dominant route. Combined with the diffusion limitation of the flat round 8-MR pore openings, the propylene selectivity is maximized. Although a considerable amount of catalytic data and experience has been accumulated [25,28–30], the mechanism of MTO reaction is still insufficient to guide the design and preparation of MTO catalysts [31,32]. A quantitative principle of shape selectivity for MTO zeolite catalysis has been constructed recently [33]. The assembly of descriptors involved thermodynamics, reaction kinetics and molecular diffusion within confined zeolite framework is necessary for the MTO modelling, but the collection of these descriptors needs high acquisition cost. It is still expected to correlate the apparent characteristics of molecular sieve catalysts directly to their MTO catalytic performance.

Artificial intelligence (AI) technology can extract knowledge from big data and form independent opinions on handling practical issues [34–37]. To accomplish laborious data mining tasks, machine learning (ML), one of the most important data-driven branches of AI, has been developed, which uses advanced algorithms to analyze data and build reliable relationships. These revolutions have a great impact on traditional catalysis research and could accelerate the development of catalysts [38,39]. Using support vector machines and deep feed-forward neural networks, Zahrt *et al*. [38] demonstrate accurate predictive modeling in the chiral phosphoric acid-catalyzed thiol addition to N-acylimines. Nielsen *et al*. [40] demonstrates that a random forest (RF) model can accurately predict reaction outcomes and aid in identifying optimal conditions for new alcohols formation from deoxyfluorination reactions. In addition to organic reactions, Ding *et al.* [41] tried 16 popular AI algorithms to analyze the features and predict the performance of the membrane electrode assembly (MEA) in proton-exchange membrane fuel cells (PEMFCs), where the decision tree (DT) and the extreme gradient boosting machine classifier (XGBoost) showed good accuracy in determining the key factors for high performance MEA, and the artificial neural network (ANN) showed the best accuracy in predicting the maximum power density. A DT ML model was also constructed and used to evaluate the relationship between the characteristics and the NO_x removal efficiency of zeolite-based SCR catalysts at low temperatures [42]. The ML methods are very well suited to replace the intuition-based approaches traditionally used to guide experiments, and have also been developed to process inorganic synthesis data and guide the synthesis of new materials including oxides [43], MOFs [34,44], and zeolites [37,45–47].

The molecular sieve-catalyzed MTO reaction has been studied for more than 30 years, and a large amount of experimental data has been accumulated. It is reasonable to build a database and learn from it by using ML technology [48]. The complicated MTO reaction mechanism can be directly ignored to establish reliable models by locking the structural parameters and physicochemical properties of molecular sieve catalysts, and to predict the possible catalytic performance. Herein, we construct a zeolite-catalyzed MTO reaction database to train ML models and search for potential MTO catalysts. The MTO reaction mechanism can be better understood by using interpretable ML models. Decision rules for potential MTO zeolite catalysts were meanwhile extracted, and experiments were conducted to verify the reliability and accuracy of the ML-model predictions. In particular, a unique MTO catalyst with high initial ethene selectivity has been discovered under the guidance of ML. Our work demonstrates that ML, as an effective new tool, can help to break through traditional cognition and experience, bringing new opportunities for zeolite catalyst research and development.

2. Experimental

2.1. Materials

All reagents are commercially obtained without further purification, including sodium hydroxide (NaOH, 96 wt%, Tianjin Guangfu Reagent Co., Ltd., Tianjin, China), sodium aluminate ($NaAlO_2$, Al_2O_3 ≥ 41%, Sinopharm Chemical Reagent Co., Ltd., Shanghai, China), *N*,*N*,*N*-Trimethyl-1-adamantammonium hydroxide hydroxide (TMAdaOH, 25 wt%, Shanghai Macklin Biochemical Technology Co., Ltd., Shanghai, China) and column-layer chromatographic silica gel (SiO_2, 99 wt%, Qingdao Chengyu Chemical Co., Ltd., Shandong, China).

2.2. Syntheses

The typical synthesis procedure of SSZ-13 was as follows. Sodium hydroxide was added to an aqueous solution containing TMAdaOH and deionized water under stirring. After a clear solution is obtained (about 10 min), column-layer chromatographic silica gel was further added and stirred for 4 h. The final gel with a molar composition of 100 SiO_2: 0.33/1.25 Al_2O_3: 3.5 Na_2O: 1000 H_2O: 16 TMAdaOH was transferred to a Teflon-lined stainless-steel autoclave and crystallized at 160 °C for 2 d. The product was collected by filtration, washed three times with deionized water, and dried at 100 °C for 24 h. The proton forms of the zeolites were obtained by triple ion-exchange of the calcined zeolite with a 1 mol L^{-1} NH_4Cl solution at 80 °C for 2 h and calcination in air at 550 °C for 4 h.

2.3. Characterizations

The powder X-ray diffraction (PXRD) data used for the phase identification were collected on a PANalytical X'Pert PRO X-ray diffract meter using the Cu K_α radiation (λ = 1.54059 Å), operating at 40 kV and 40 mA. The inorganic elemental compositions of the solid samples were measured with Philips Magix-601 X-ray fluorescence (XRF) spectrometer. The crystal morphology was observed using a scanning electron microscope (Hitachi SU8020). The textural properties of the calcined samples were measured by N_2 sorption at −196 °C on a Micromeritics ASAP2020 volumetric adsorption analyzer. Prior to the measurement, the sample was degassed at 350 °C under vacuum for 4 h. The total surface area was evaluated based on the BET equation. The micropore surface area and micropore volume were calculated using the *t*-plot method.

2.4. Catalytic tests

Methanol-to-olefin (MTO) reaction was performed in a quartz tubular fixed-bed reactor at atmospheric pressure. 0.3 g of catalyst (40–60 mesh) was loaded in the quartz reactor and activated at 550 °C in a N_2 flow of 30 mL min^{-1} for 30 min before starting each reaction run, and then the temperature was adjusted to reaction temperature as required. The methanol was fed by passing the carrier gas through a saturator containing methanol and was mixed with empty carrier gas, which gave a desired WHSV. The reaction products were analyzed using an online gas chromatograph (Agilent GC 7890N), equipped with a flame ionization detector (FID) and a Plot-Q column. The conversion and selectivity were calculated on the basis of CH_2. Dimethyl ether (DME) was considered to be a reactant in the calculation.

General introduction of ML methods can be found in Supporting Information.

3. Results and discussion

3.1. Database construction and curation

Fig. 1 presents a pipeline schematic of extracting and combining zeolite-catalyzed MTO data. Parts of the structural parameters of zeolites, including the largest ring size (LRS), framework density (FD_{Si}), channel dimension (CD), maximum diameter of a sphere that can diffuse along the *a*, *b* and *c* directions (MD_a, MD_b, MD_c), and maximum diameter of a sphere that can be entrapped (MD_i), were first extracted from the International Zeolite Association (IZA) database [49]. All of these parameters are considered as important inputs since the MTO activity and selectivity varies with the cavity and channel architectures [50]. Then CrossRef searching and literature information extraction were conducted (Table S1 and Table S2).

Various physicochemical parameters of zeolite catalysts, including whether modified or not (Mod.), acid strength (AS) and density (A/T), and crystal size (CS), were collected from literatures to describe more details of molecular sieve catalysts. Since the acid strength (AS) of SAPO molecular sieves is typically weaker than that of aluminosilicate zeolites, we tentatively classified AS into two types: the aluminosilicate type (assigned a value of 1) and the SAPO type (assigned a value of 0). The acid amount is associated with both the incorporation level and environments of the respective atoms, where the relevant atoms refer to framework Al atoms for aluminosilicates and framework Si atoms for SAPOs. Herein, atomic environments are not considered for the present study, primarily due to the difficulties in their characterization and description. Acid density is quantified using the A/T ratio, which is uniformly defined as the molar ratio of Al atoms (for aluminosilicates) or Si atoms (for SAPOs) to the total number of normalized framework atoms (T), ignoring their specific positions. Reaction temperature (RT), time on stream (TOS) and weight hourly space velocity of methanol (WHSV) are also taken as inputs. Methanol conversion, selectivity towards ethene and propene were collected as outputs. Since we focus on exploring more structural possibilities in MTO molecular sieve catalysts, the catalytic data over various molecular sieves are tried to keep balance. And the redundant reports on several MTO popular molecular sieves such as CHA, AEI, and MFI zeolites, were excluded. In total, 69 papers were filtered and selected from more than 70 thousand journal articles covering 41 type IZA codes, over 80 type molecular sieve materials, and more than 6000 MTO data points (Table S3 and Fig. S1). The distributions of categorical variables and units are listed in Table 1 and Figs. S2–S10. Given the significant imbalance in the data, the data

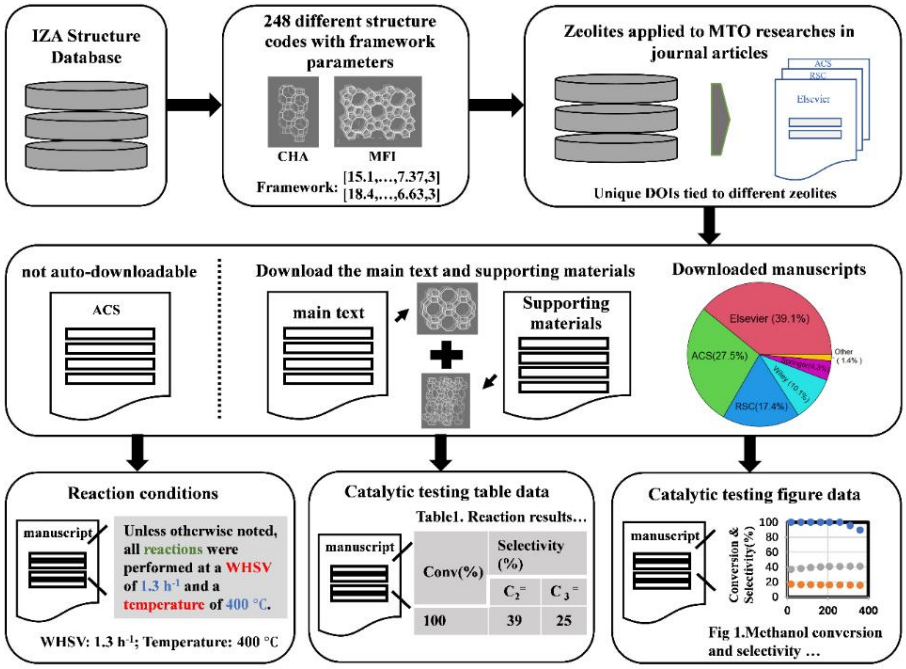


Fig. 1. Schematic for curating the datasets: extracting structural parameters of zeolites from the IZA database, searching for MTO catalytic results of different zeolites from published papers, downloading the papers, and mining text, tables and graphs for reaction conditions and catalytic results.

needs to be randomly split for a sufficient number of times during the training of ML models to avoid any potential bias. In addition, ensemble methods, such as RF and XGBoost, will be employed particularly in ML model training, as these methods exhibit robustness to imbalanced data at the algorithmic level. A schematic diagram on some parameters and the correlation heatmap of all input features are depicted in Fig. 2. The collinearity effects may be present, as the absolute values of some correlation coefficients are higher than 0.5. Therefore, some traditional statistical linear models are also considered to compare with the ML models.

Table 1
The type, name and range of the variables in the MTO database.

Variables		Name of variables/units	Range/counts
Input variables	catalyst property	Mod.	0/5463 1/1273
		AS	0 (SAPO)/1959 1 (zeolite)/4777
		A/T	0–1
		FD_{Si}/T/1000Å [3]	12.80–19.70
		LRS	8-MR/4159 10-MR/1843 12-MR/712 18-MR/22
		MD_a/Å	1.27–7.35
		MD_b/Å	1.49–7.35
		MD_c/Å	1.53–11.20
		MD_i/Å	5.25–11.74
		CD	1D/436 2D/1421 3D/4879
		CS/μm	0.02–5.00
	reaction conditions	RT/°C	270–700
		WHSV/h^{-1}	0.35–48.00
		TOS/min	0–30156.92
Output variables		methanol conversion/%	0.00–100.00
		S_{ethene}/%	0.00–68.61
		$S_{propene}$/%	0.00–68.90
		$S_{C2^=+C3^=}$/%	0.00–92.21
		P/*E*	0.00–37.37

3.2. Classification of conversion

A classification method has to be used to handle the conversion data, since the points with 100% conversion are significantly more than those with other values (Fig. S11). The value of 98% was thus used to categorize the data into two groups, resulting in 2621 and 3553 samples labeled as "High (≥ 98%)" and "Low (< 98%)", respectively. 20 kinds of ML algorithms were tried. As shown in Fig. 3, all the tree-based learning methods perform well with AUC above 0.9. In particular, the random forest (RF) classifier, the extreme gradient boosting (XGBoost), the CatBoost classifier, the extra trees (ET), and the light gradient boosting (LGB) show AUC metrics equal to or greater than 0.96. Comparatively, these ensemble methods are more robust to imbalanced data and more suitable for analyzing the MTO issue. When making decisions, decision trees mainly split data based on local feature rather than the overall sample distribution. Even if the overall data is imbalanced, the sample distribution within the local space may be relatively balanced, enabling the decision tree to better capture the characteristics of the minority class. Ensemble methods, such as RF, further enhance the ability to identify the minority class by aggregating the votes or weighted averages of multiple decision trees, thereby reducing the bias of a single tree. The blending and stacking methods are procedures designed to increase the predictive performance by blending or combining the predictions of multiple ML models, received the highest AUC score of

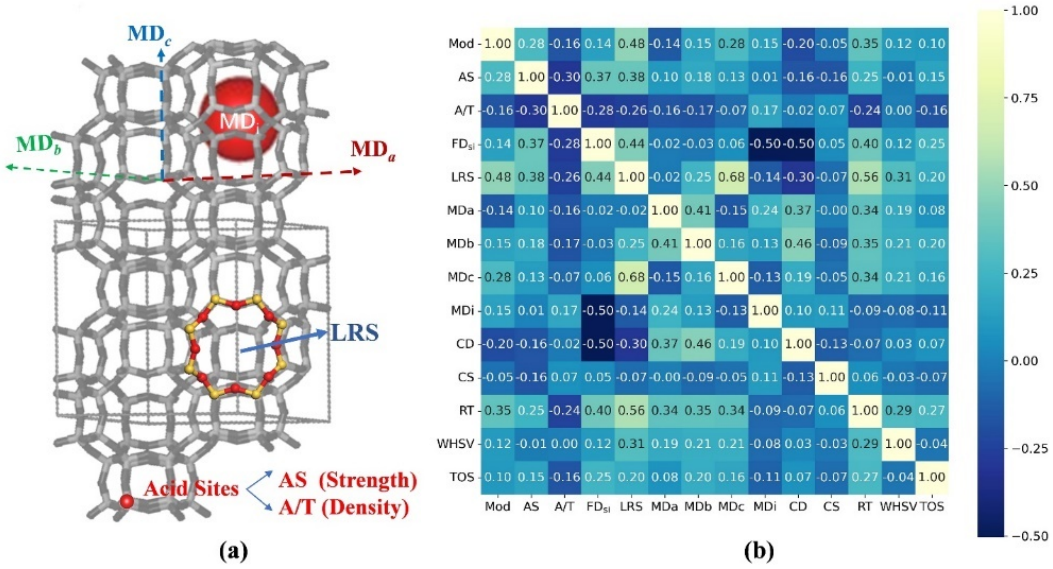


Fig. 2. (a) Illustration of some inputs including AS (acid strength): 0 for SAPO and 1 for aluminosilicate; A/T (acid density): Si/T for SAPO and Al/T for aluminosilicate (T refers to the sum of framework atoms [T atoms]); Mod.: 1 for modified and 0 for unmodified. (b) The correlation heatmap of all inputs in the MTO dataset.

0.98. The optimal tuning parameters of six representative methods are listed in Table S4. Further evaluation of these methods by the other metrics confirms the better performance of the ensemble learning methods, as shown in Table S5 and Figs. S12–S16. A confusion matrix was performed by the stacking classifier on the entire dataset to assess its quality. As seen in the inset of Fig. 3 and Table S6. An accuracy of 97.2% indicating a weighted precision of the method that would be truly reliable when faced with new samples in real-world situations. Note that the classification methods do not provide a complete S-shape conversion curve as a function of TOS, but only whether or not the conversion is greater than 98% at a given reaction time. We therefore consider the catalyst to be deactivated if there are three consecutive conversion points showing "low".

The decision tree (DT), as a strong interpretable model, presents decision rules and possible outcomes in the form of a tree graph which can clearly show the catalytic behavior of a zeolite in MTO reaction by combining different variables. Herein, the DT classification (Fig. 4 and Table S4) with a satisfactory prediction ability (AUC = 0.90) was used to comprehend the essential relationship between the inputs and conversion. As shown on the right side of Fig. 4, the root node MD_b divides the data into two parts. Some of the typical leave nodes have been circled. It is clear that the "High" results are more when follows the rule of MD_b > 4.335 Å and RT > 375 °C (red circles). High WHSV (> 5.425 h^{-1}) or long TOS (>5571.632 min) can even have opportunities to receive "High" results. On the contrary, "Low" results are more likely when MD_b ≤ 4.335 Å, and it's hard for TOS to last 192.837 min. It should be noted that 4.335 Å is the maximum sphere diameter that can diffuse

Fig. 3. Summary of the AUC for each ML algorithm for classification of methanol conversion. The inset is the performance of classification on the whole dataset given by the stacking classifier.

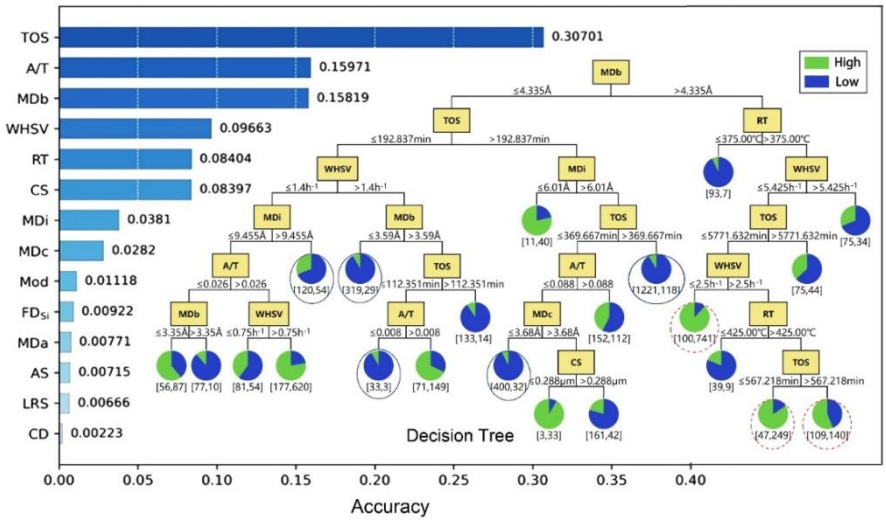


Fig. 4. Feature importance heuristic given by the DT classifier for methanol conversion and visual structure of the DT. Some of the typical leave notes mentioned have been highlighted with bule solid (with more "low" points) and red dotted line (with more "high" points) circles.

along the 8-MR channels, examples are known as LTA and ITE topologies with the 8-MR size of *ca.* 4.1 Å. The MD_b of 4.46 Å for ZSM-5 represents the smallest critical sphere diameter passing through 10-MR channels. The catalytic lifetime of ZSM-5 is significantly prolonged [51], revealing the positive contribution of the approximately 0.4 Å widened orifice on conversion. When MD_b ≤ 4.335 Å, MD_i and A/T are additional important factors which need to be modified especially for high conversion. Combined with the results of the exit nodes, MD_i should not be too large (≤ 9.455 Å), and A/T should not be too low (> 0.008). The feature importance analysis given by DT classification provides a comprehensive comparison for these impact factors. Accuracy is recognized as one of the best criteria for accomplishing feature selection. The greater the Accuracy, the stronger the discrimination ability of the corresponding features. As shown on the left side of Fig. 4, the top two important features for zeolite are A/T and MD_b followed by CS and MD_i. It suggests that acidity and channel size are the basic requirements for a zeolite to catalyze MTO reaction. It is a popular way to prolong the catalytic lifetime by decreasing the crystal size of zeolite catalyst [6,18,52–54]. The validity of this strategy is rationally confirmed from the perspective of data analysis. The importance of cage size (MD_i) for high conversion is less concerned although cavity-controlled methanol conversion over zeolite catalysts has been well discussed [55,56]. According to the previous experience, the formed hydrocarbon species (HCPs) are responsible for the activity, selectivity and deactivation of the catalyst material, which change with reaction temperature (RT) [57], acidity (A/T and AS) [55] and reaction space (MD_i) [21–24]. It is reasonable to assume that the appropriate cavity size matching the A/T and RT could determine the type of HCPs formed, and thus control the catalytic capacity for methanol conversion. An attractive method to synthesize zeolites with a molecular-recognition MD_i is to use organic structure-directing agents that mimic the transition state of preestablished reactions to be catalyzed [58]. More effective methods of synergistic regulation of channel and cavity size, crystal morphology and acidity (chemical composition) for zeolites are expected to be developed. Comparatively, the previous mechanism research using comparative experiments or theoretical simulation and calculation usually needs to simplify the variables as much as possible to facilitate the understanding of a certain impact factor [59–62]. DT classification has indeed acquired the intelligence to identify the most influential parameters on MTO conversion through data-driven, which even form a very comprehensive summary and certain clues for catalyst research.

3.3. *Regression of selectivity*

The continuous product distribution in the MTO data enables regression analysis. The selectivity of ethene plus propene are interrelated during the entire reaction process (Fig. S17). Therefore, instead of considering them individually, we took their sum ($S_{C_2^=+C_3^=}$, Fig. S18) and ratio (P/E, Fig. S19) into account to construct a joint distribution subject to the following constraint. The Box-Cox transformation [63] is conducted to improve the Gaussianity of the data distribution (Figs. S20 and S21) for enhancing their performance of a ML algorithm.

$$\begin{cases} S_{C_2^=+C_3^=} = S_{\text{ethene}}\cdot(\%) + S_{\text{propene}}\cdot(\%) \le 1, \\ P/E = \dfrac{S_{\text{propene}}\cdot(\%)}{S_{\text{ethene}}\cdot(\%)} > 0, \end{cases}$$

Twenty-three regression methods were trained and evaluated using four different metrics (MAE, RMSE, R^2 and RMSLE), to determine which model provided fits best. The R^2 value (coefficient of determination) is a widely used indicator that takes values between 0 to 1. The closer the value is to 1, the better the model fits the data. As depicted in Figs. 5(a) and 5(b), the R^2 values of all the tree-based ensemble learning methods, such as RF, ET and XGBoost, are close to 1. In contrast, there is a significant drop in R^2 values after Decision tree (DT, with R^2 of 0.88 for $S_{C_2^=+C_3^=}$ and 0.87 for P/E) and Gradient Boosting Regressor (GBR, with R^2 of 0.85 for $S_{C_2^=+C_3^=}$ and 0.88 for P/E). This indicates that the tree-based methods outperform other methods. These results are consistent with those obtained from other

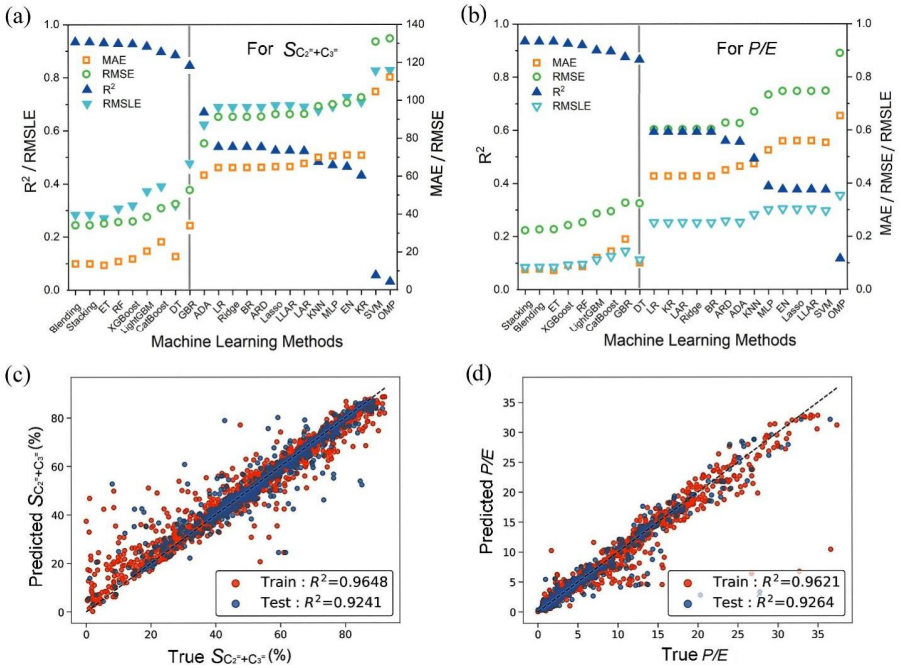


Fig. 5. The performance of 23 ML methods for $S_{C_2^=+C_3^=}$ (a) and P/E (b) respectively. Four metrics, MAE, RMSE, R^2, and RMSLE were used to evaluate these methods. The ML methods before the gray line are better. The training and test sets of $S_{C_2^=+C_3^=}$ (c) and P/E (d) by Stacking Regressor.

evaluation indicators, like MAE, RMSE and RMSLE, where the smaller values are more favorable (Figs. 5(a) and 5(b), and Tables S7–S10). The Blending and Stacking methods can harness the capabilities of the well-performing methods mentioned above, and they exhibit better performance than any single one. The predictive capabilities of the Stacking regressor for $S_{C_2^=+C_3^=}$ and P/E are shown in Figs. 5(c) and 5(d), respectively. The similar R^2 values exceeding 90% for both training and test sets confirm the high reliability of the models and indicate the minimal overfitting of the data. For the P/E ratio, certain data points in Fig. 5(d) deviate markedly from the ideal trend line, which is primarily attributed to the skewed distribution of P/E data. It is highly desirable to supplement more relevant data and explore more appropriate descriptors in the future.

The DT regression aids in identifying the most universal input features necessary for achieving a high $S_{C_2^=+C_3^=}$, which offers some interpretability and provides guidance regarding our concerns. As depicted in Fig. 6(a), all data points with an $S_{C_2^=+C_3^=}$ above 80% (marked by blue lines) correspond to cases of LRS ≤ 9 and MD_i ≤ 7.71 Å. The optimal range for MD_i is between 7.35–7.71 Å, and an A/T ratio ≤ 0.105 is also preferable. In contrast, the parts of MD_i > 7.71 Å, as well as the branch of LRS > 9 (marked by red lines) lead to more $S_{C_2^=+C_3^=}$ less than 63%. When LRS exceeds 9, an A/T ratio less than 0.01 becomes more crucial as it leads to an $S_{C_2^=+C_3^=}$ higher than 50%. In the feature importance analysis (Fig. 6(b)), LRS is the most important parameters as expected, and the A/T and the MD_i are two other significant factors influencing $S_{C_2^=+C_3^=}$. SHAP plots (Fig. 6(c)) demonstrate that a lower A/T ratio, smaller LRS and MD_i have a positive effect on $S_{C_2^=+C_3^=}$. Based on the above analysis, zeolite with parameter combination of LRS ≤ 9, A/T ≤ 0.01, and MD_i in the range of 7.35–7.71 Å might possess potential characteristics for achieving a high $S_{C_2^=+C_3^=}$. Checking the IZA database found that seven topologies including AEI [64,65], AWW [66,67], AFT [68,69], AFX [70–72], CHA [73–76], DDR [77,78], and SFW [79] meet these structural criteria roughly. The current synthetic status of these molecular sieves is listed in Table 2. It can be seen that to synthesize these materials with low acid density (0 < A/T ≤ 0.01), is a main challenge at present.

Noting that the current selectivity analysis did not correlate with conversion, a relationship analysis between olefin yields and discrete input variables was further carried out. The results are presented in the violin plots of Fig. S22. The crucial input variables for an ideal MTO catalyst are further elucidated. Based on the experimental data obtained thus far, 8-MR (LRS = 8) SAPO materials (AS = 0) with three-dimensional channels (CD = 3) will yield higher amounts of light olefins, and the modification of zeolites has minimal impact on the results. This conclusion is consistent with the fact that SAPO-34 molecular sieve with CHA structure is the optimal MTO catalyst currently.

Besides achieving a high selectivity of ethene plus propene, attaining high selectivity for either ethene or propene alone is also appealing yet challenging. There are 3274 P/E data points in the range of 1.0–2.0 at a RT lower than 437.5 °C, as seen in Fig. S23(a) marked by blue lines. It represents the most common MTO product distribution. The P/E data greater than 3.0 is the second-most common part belonging to condition of RT > 437 °C and A/T ≤ 0.01, which is marked by red lines. Comparatively, the P/E data below 1.0 is quite rare, resulting in the difficulty for decision rules extraction for enhancing ethene selectivity. RT presents as the most important factor in the feature importance analysis (Fig. S23(b)), followed by the A/T and the MD_a. However, the effect of RT on the P/E ratio appears uncer-

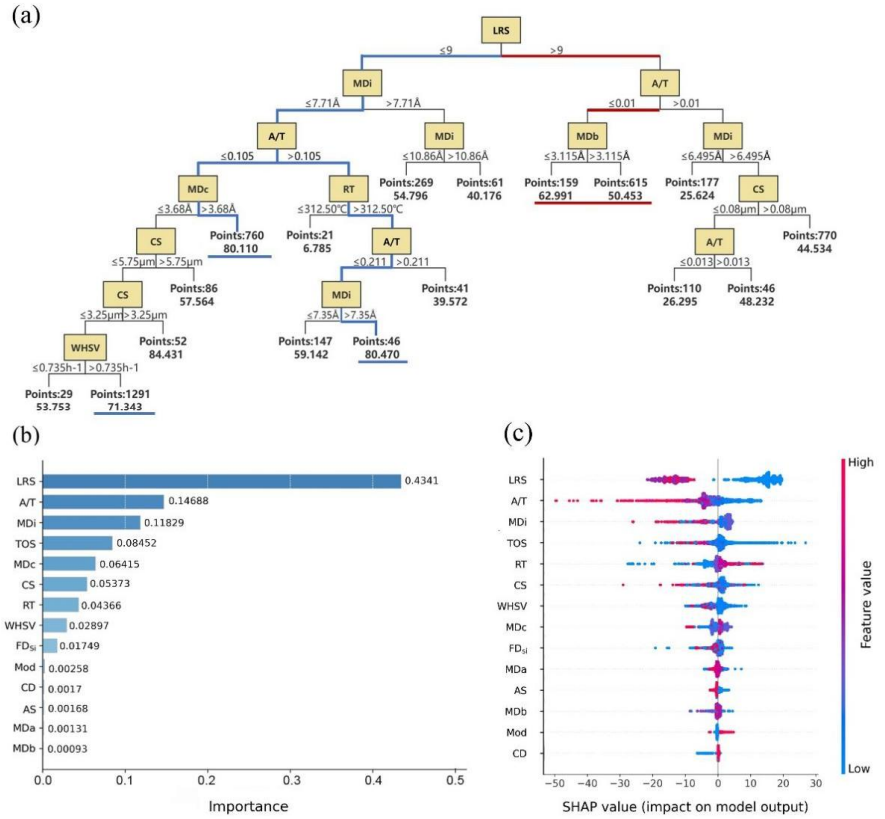


Fig. 6. (a) The visual structure of the DT. The numbers below the points are the average values of $S_{C_2^=+C_3^=}$. (b) The feature importance heuristic given by the DT Regressor for $S_{C_2^=+C_3^=}$. (c) SHAP analysis for $S_{C_2^=+C_3^=}$. A color gradient, ranging from blue solid line (low feature values) to pink/red dotted line (high feature values), helps visualize the correlation between feature values and their SHAP values.

tain in the SHAP analysis (Fig. S23(c)), whose importance rank drops to the 7th herein. This ambiguous result suggests a complex relationship between RT and the *P*/*E* ratio. In addition, it seems that the lower A/T, the higher the *P*/*E*.

The above regression analyses reflect the correlation between structural and physicochemical parameters of zeolite catalysts and the light-olefins selectivity to the greatest extent, which are basically consistent with the current recognition [80]. It is worth noting that these recognitions on MTO were derived from years of research involving numerous well-designed experiments and analyses conducted by researchers. In this work, the similar knowledge can be easily achieved through the ML analysis of the collected MTO data. When a large amount of multi-parameter data that is beyond the analysis ability of human brain comes in quickly and directly (fast and automatic access to experimental and computational data), ML will be an indispensable tool for efficiently extracting the underlying rules and correlations.

Table 2
Lists of the information on materials belonging to the seven topologies.

Topology	SAPOs	A/T	Zeolites	A/T	CD
AEI	SAPO-18 [64]	0–0.1	SSZ-39 [65]	0.1	3
AWW	AlPO-22 [66] AlPO-CJB1 [67]	0	none	—	1
AFT	AlPO-52 [68]	0	SSZ-112 [69]	~0.13	3
AFX	SAPO-56 [70,71]	0.15–0.2	SSZ-16 [72]	0.14	3
CHA	SAPO-34 [73]	0.05–0.2	SSZ-13 [74–76]	0.02–0.2, 0	3
DDR	none	—	ZSM-58 [77,78]	0.01–0.02, 0	2
SFW	none	—	SSZ-52 [79]	0.11	3

3.4. Prediction and experimental validation

Data-driven ML analysis has assimilated a substantial amount of the expertise of MTO chemists. A software Z-MTO-P V1.1 [81] was thus designed to assist the development of MTO catalyst. The logic and process of MTO catalyst design and screening are presented in Fig. S24. An extensive prediction and screening of zeolite catalysts using it indicates that CHA remains the most promising candidate for MTO catalysts, probably due to the limitation of known molecular sieve structure data. Considering the DT extracted decision rules of LRS ≤ 9, A/T ≤ 0.01, and MD_i within 7.35–7.71 Å for achieving a high $S_{C_2^=+C_3^=}$, and the searching results shown in Table 2, the synthesis of microporous high-silica SSZ-13 has become our primary objective again. Surprisingly, the MTO study on SSZ-13 with a Si/Al ratio greater than 50 are quite rare [82], and the MTO data at RT ≥ 450 °C is insufficient [83] although SSZ-13 has been well studied as a very popular MTO catalyst. This scarcity can likely be ascribed to negative presumptions about its MTO performance based on prior experience, like the higher synthesis cost and faster deactivation than SAPO-34. Much effort has been put into controlling the synthesis cost [76] and morphology of SSZ-13 [84,85].

In this work, an SSZ-13 with a Si/Al ratio of 100 (A/T = 0.01), named SSZ-13-100, was synthesized (Fig. S25) to evaluate its MTO performance. The optimal reaction conditions were forecast in advance by software Z-MTO-P V1.0 to be 450 °C with a WHSV of 1.0 h^{-1}. The catalytic results were presented in Fig. 7(a) and Table S11. Interestingly, the initial selectivity of ethene plus propene ($S_{C_2^=+C_3^=}$) is very high (82.6%) at TOS of 3 min, with 100% methanol conversion. The $SC_2^=+C_3^=$ quickly rises to 87.6% at 37 min, accompanied by an unexpectedly high ethene selectivity of 61.1%. In contrast, the SSZ-13-19 with a Si/Al ratio of 19 (A/T ratio of 0.05) show a low start (75.4%) but gradually increased $S_{C_2^=+C_3^=}$ curve reaching to its peak of 82.6% before deactivation (258 min). The highest ethene selectivity of 56.2% also appears before inactivation at 326 min. The MTO result of SSZ-13-19 is quite typical and similar to the previously reported results of SAPO-34 and SSZ-13, which is quite different from those of SSZ-13-100. Although SSZ-13-100 sacrifices part of lifespan, it receives the highest $S_{C_2^=+C_3^=}$ and the highest ethene selectivity (lowest P/E ratio) in the shortest reaction time (TOS = 37 min) among the known MTO results up to now [6,10,24,55,86].

Currently, pre-accumulation carbon technology is widely adopted to bypass the initial low selectivity stage for higher olefin yields [87]. The rapid attainment of high $S_{C_2^=+C_3^=}$ furnishes inspiration for the design of a new MTO catalyst that does not necessitate the precoke-deposition step. Additionally, the harvest of a high ethene selectivity is extremely appealing regarding to the requirements of precise product regulation. It is possible to understand the mechanism and design a methanol-to-ethene catalyst based on the findings [26]. Recently, a water-assisted shape-selective production of ethene in the MTO reaction was reported by Zhang *et al.* [88] With the aid of ML, we have identified a new catalytic material capable of producing ethene efficiently without any additional assistance.

In addition to SSZ-13, we recently synthesized a high-silica STT zeolite via a chain rearrangement process of lamellar Si-MWW zeolite [89]. The structure of STT features both 9-MR and 7-MR windows without any 8-MR one. Its STT cage (11.7 × 8.1 $Å^2$) is located at the intersection of its 7-MR and 9-MR channels. The corresponding descriptors of STT are as follows: LRS = 9, FD_{Si} = 17.0 T/1000 $Å^3$, CD = 2, MD_a = 2.76 Å, MD_b = 1.89 Å, MD_c = 2.76 Å, and MD_i = 7.04 Å. This zeolite was previously excluded from our MTO database because its MTO data was not available before. The successful synthesis of STT allows us to experimentally validate the generalization ability and accuracy of the ML models. The Z-MTO-P V1.0 software enables the prediction of optimal MTO evaluation conditions and the corresponding results, which exhibit close consistency with experimental findings, as depicted in Fig. 7(b). These results confirm the reliability and practicality of the ML models. We demonstrate ML can be an effective and reliable research tool for catalysis research.

4. Conclusions

A data-driven ML workflow has been developed for screening potential MTO catalysts. Tree-based ensemble methods perform better with prediction accuracy in both conversion (classification) and light-olefins selectivity (regression) above 90%. The reliability and practicality of the ML models has been further validated by experiments external to the database. Based on the decision rules for high selectivity extracted from ML models, zeolite catalysts should simultaneously possess the following characteristics: the largest channel size should be smaller than or equal to 9-membered ring (LRS ≤ 9), and the maximum diameter of a sphere that can be entrapped in a zeolite (MD_i) should be in the range of 7.35–7.71 Å. Combined with the desired acid density of around 1% for the zeolite catalyst (A/T ≤ 0.01), the available molecular sieves candidates that meet the above standards are very rare. We thus refocus on high-silica SSZ-13 (Si/Al = 100) and observe its unexpectedly high initial selectivity, 87.6% for ethene plus propene at 37 min time on stream, with a notable ethene selectivity of 61.1%. It represents the highest level of ethene selectivity in the MTO reaction. This finding is expected to eliminate the pre-carbon-deposition step in the MTO industrial process, which is specifically designed to bypass the initial stage characterized by low selectivity.

ML allows us to circumvent the intricate MTO reaction mechanism and construct reliable models for direct prediction of potential MTO catalyst candidates, which offers new opportunities for tackling specific catalyst-related issues. The merits

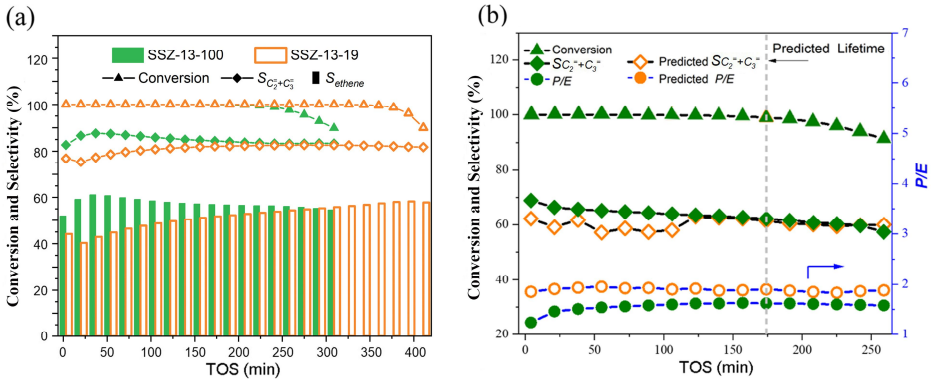


Fig. 7. (a) MTO catalytic performance of SSZ-13s with different Si/Al ratios at 450 °C, WHSV of 1.0 h^{-1}. (b) The experimental (solid symbols) and predicted (hollow symbols) MTO catalytic performance of STT-100 at 450 °C and WHSV of 1.0 h^{-1}. The predicted lifetime is labeled by the gray dotted line.

of ML technology, including efficiency, accuracy and adaptability, have been fully utilized to curtail labor-intensive laboratory R&D costs. Nonetheless, we recognize that ML methods rely heavily on the quantity and quality of data. In the context of limited data availability, it is arduous for ML to generate innovative predictions. Therefore, it is imperative to integrate the expertise and intuition of domain experts to attain innovative outcomes. The automated rapid acquisition of experimental data, along with the development of precise descriptors, represents two pivotal directions for future research. Moreover, the advancement of large language models (LLMs) to reconcile experimental data, literature data and computational data is also of substantial significance. We anticipate that artificial intelligence (AI) will serve as a potent technology, empowering chemists to overcome a wider array of intricate chemical challenges in the future, thereby revolutionizing the landscape of chemistry research.

Electronic supporting information

Supporting information is available in the online version of this article. It includes database construction details; statistics on the distribution of our database; machine learning algorithms; the performance of various machine learning methods; the hyperparameters of the representative models and the standard confusion matrix instruction and results.

Acknowledgments

The authors thank the National Key Research and Development Program of China (2024YFE0207000), National Natural Science Foundation of China (22171259, 12461050, 21991090, 21991091, 22288101, 22372020), and the AI S&T Program of Yulin Branch, Dalian National Laboratory for Clean Energy, CAS (DNL-YL A202206).

References

[1] G. W. Huber, S. Iborra, A. Corma, *Chem. Rev.*, **2006**, 106, 4044–4098.
[2] B. Yilmaz, U. Müller, *Top. Catal.*, **2009**, 52, 888–895.
[3] J. Liang, H. Li, S. Zhao, W. Guo, R. Wang, M. Ying, *Appl. Catal.*, **1990**, 64, 31–40.
[4] M. Stöcker, *Microporous Mesoporous Mater.*, **1999**, 29, 3–48.
[5] P. Tian, Y. Wei, M. Ye, Z. Liu, *ACS Catal.*, **2015**, 5, 1922–1938.
[6] M. Yang, D. Fan, Y. Wei, P. Tian, Z. Liu, *Adv. Mater.*, **2019**, 31, 1902181.
[7] X. Jiang, X. Nie, X. Guo, C. Song, J. G. Chen, *Chem. Rev.*, **2020**, 120, 7984–8034.
[8] U. Olsbye, S. Svelle, M. Bjørgen, P. Beato, T. V. W. Janssens, F. Joensen, S. Bordiga, K. P. Lillerud, *Angew. Chem. Int. Ed.*, **2012**, 51, 5810–5831.
[9] J. H. Kang, F. H. Alshafei, S. I. Zones, M. E. Davis, *ACS Catal.*, **2019**, 9, 6012–6019.
[10] P. Ferri, C. Li, C. Paris, A. Vidal-Moya, M. Moliner, M. Boronat, A. Corma, *ACS Catal.*, **2020**, 9, 11542–11551.
[11] S. Standl, O. Hinrichsen, *Catalysts*, **2018**, 8, 626.
[12] Q. Zhang, G. Chen, Y. Wang, M. Chen, G. Guo, J. Shi, J. Luo, J. Yu, *Chem. Mater.*, **2018**, 30, 2750–2758.
[13] Y. Wu, J. Han, W. Zhang, Z. Yu, K. Wang, X. Fang, Y. Wei, Z. Liu, *J. Am. Chem. Soc.*, **2024**, 146, 8086–8097.
[14] X. Zhao, L. Wang, J. Li, S. Xu, W. Zhang, Y. Wei, X. Guo, P. Tian, Z. Liu, *Catal. Sci. Technol.*, **2017**, 7, 5882–5892.
[15] X. Zhao, L. Wang, P. Guo, N. Yan, T. Sun, S. Lin, X. Guo, P. Tian, Z. Liu, *Catal. Sci. Technol.*, **2018**, 8, 2966–2974.
[16] M. Yoshioka, T. Yokoi, T. Tatsumi, *ACS Catal.*, **2015**, 5, 4268–4275.
[17] Q. Sun, Z. Xie, J. Yu, *Natl. Sci. Rev.*, **2018**, 5, 542–558.
[18] R. Martínez-Franco, Z. Li, J. Martínez-Triguero, M. Moliner, A. Corma, *Catal. Sci. Technol.*, **2016**, 6, 2796–2806.
[19] M. Dusselier, M. E. Davis, *Chem. Rev.*, **2018**, 118, 5265–5329.
[20] I. Pinilla-Herrero, U. Olsbye, C. Márquez-Álvarez, E. Sastre, *J. Catal.*,

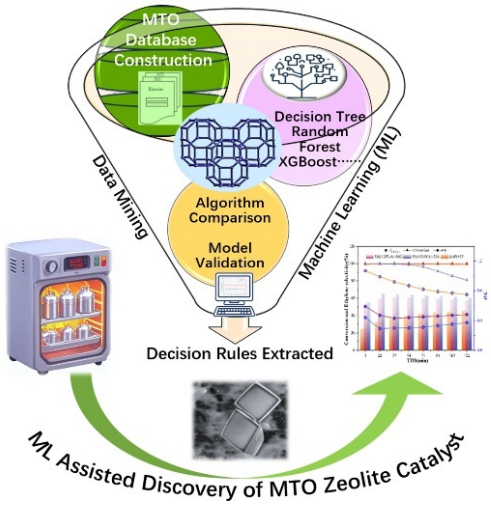
Graphical Abstract

Chin. J. Catal., 2026, 81: 124–135 doi: 10.1016/S1872-2067(25)64903-5

Machine-learning-aided discovery of methanol-to-olefins zeolite catalysts with ultra-high initial selectivity

Xinyi Wang, Chaoqi Wang, Miao Yang *, Xiaoguang Wang *, Yuezhong Zuo, Zhuangzhuang Zhang, Yimo Wu, Jingfeng Han, Bing Li, Wei Huang, Limin Ren, Yingxu Wei, Xinmei Liu, Peng Tian, Zhongmin Liu *
Dalian Institute of Chemical Physics, Chinese Academy of Sciences
Dalian University of Technology
China University of Petroleum
University of Chinese Academy of Sciences

An MTO zeolite catalyst with an ultra-high initial selectivity has been discovered at a low trial-and-error cost through a machine learning (ML)-aided workflow. Its ethene selectivity unexpectedly exceeds 60%, presenting the highest reported value in the field of MTO catalysis.



2017, 352, 191–207.
[21] W. Song, A. H. Fu, J. F. Haw, *J. Am. Chem. Soc.*, **2001**, 123, 4749–4754.
[22] J. F. Haw, W. Song, D. M. Marcus, J. B. Nicholas, *Acc. Chem. Res.*, **2003**, 36, 317–326.
[23] J. Li, Y. Wei, J. Chen, P. Tian, X. Su, S. Xu, Y. Qi, Q. Wang, Y. Zhou, Y. He, Z. Liu, *J. Am. Chem. Soc.*, **2012**, 134, 836–839.
[24] J. Li, Y. Wei, J. Chen, S. Xu, P. Tian, X. Yang, B. Li, J. Wang, Z. Liu, *ACS Catal.*, **2015**, 5, 661–665.
[25] S. Lin, Y. Zhi, W. Chen, H. Li, W. Zhang, C. Lou, X. Wu, S. Zeng, S. Xu, J. Xiao, A. Zheng, Y. Wei, Z. Liu, *J. Am. Chem. Soc.*, **2021**, 143, 12038–12052.
[26] M. Yang, B. Li, M. Gao, S. Lin, Y. Wang, S. Xu, X. Zhao, P. Guo, Y. Wei, M. Ye, P. Tian, Z. Liu, *ACS Catal.*, **2020**, 10, 3741–3749.
[27] Y. Wang, J. Han, N. Wang, B. Li, M. Yang, Y. Wu, Z. Jiang, Y. Wei, P. Tian, Z. Liu, *Chin. J. Catal.*, **2022**, 43, 2259–2269.
[28] S. Xu, Y. Zhi, J. Han, W. Zhang, X. Wu, T. Sun, Y. Wei, Z. Liu, *Adv. Catal.,* **2017**, 61, 37–122.
[29] S. Lin, Y. Zhi, Z. Liu, J. Yuan, W. Liu, W. Zhang, Z. Xu, A. Zheng, Y. Wei, Z. Liu, *Natl. Sci. Rev.*, **2022**, 9, nwac151.
[30] S. Lin, Y. Zhi, W. Zhang, X. Yuan, C. Zhang, M. Ye, S. Xu, Y. Wei, Z. Liu, *Chin. J. Catal.*, **2023**, 46, 11–27.
[31] Y. Li, H. Cao, J. Yu, *ACS Nano*, **2018**, 12, 4096–4104.
[32] S. Chen, L. Li, Y. Li, J. Yu, *Chem. J. Chin. Univ.,* **2021,** 42, 179–187.
[33] M. Gao, H. Li, J. Yu, M. Ye, Z. Liu, *AlChE. J.*, **2023**, 69, e17881.
[34] P. Raccuglia, K. C. Elbert, P. D. F. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier, A. J. Norquist, *Nature*, **2016**, 533, 73–76.
[35] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, A. Walsh, *Nature*, **2018**, 559, 547–555.
[36] I. W. Davies, *Nature*, **2019**, 570, 175–181.
[37] O. Kononova, T. He, H. Huo, A. Trewartha, E. A. Olivetti, G. Ceder, *iScience*, **2021**, 24, 102155.
[38] A. F. Zahrt, J. J. Henle, B. T. Rose, Y. Wang, W. T. Darrow, S. E. Denmark, *Science*, **2019**, 363, eaau5631.
[39] T. Toyao, Z. Maeno, S. Takakusagi, T. Kamachi, I. Takigawa, K.-I. Shimizu, *ACS Catal.*, **2020**, 10, 2260–2297.
[40] M. K. Nielsen, D. T. Ahneman, O. Riera, A. G. Doyle, *J. Am. Chem. Soc.*, **2018**, 140, 5004–5008.
[41] R. Ding, R. Wang, Y. Ding, W. Yin, Y. Liu, J. Li, J. Liu, *Angew. Chem. Int. Ed.*, **2020**, 59, 19175–19183.
[42] S. Bae, H. Lee, J. Shin, H. S. Kim, Y. Kim, D. H. Kim, J. M. Lee, *Chem. Mater.*, **2022**, 34, 7761–7773.
[43] E. Kim, K. Huang, A. Tomala, S. Matthews, E. Strubell, A. Saunders, A. McCallum, E. Olivetti, *Sci. Data*, **2017**, 4, 170127.
[44] A. Nandy, C. Duan, H. J. Kulik, *J. Am. Chem. Soc.*, **2021**, 143, 17535–17547.
[45] Z. Jensen, E. Kim, S. Kwon, T. Z. H. Gani, Y. Román-Leshkov, M. Moliner, A. Corma, E. Olivetti, *ACS Cent. Sci.*, **2019**, 5, 892–899.
[46] Z. Jensen, S. Kwon, D. Schwalbe-Koda, C. Paris, R. Gómez-Bombarelli, Y. Román-Leshkov, A. Corma, M. Moliner, E. A. Olivetti, *ACS Cent. Sci.*, **2021**, 7, 858–867.
[47] E. Pan, S. Kwon, Z. Jensen, M. Xie, R. Gómez-Bombarelli, M. Moliner, Y. Román-Leshkov, E. Olivetti, *ACS Cent. Sci.*, **2024**, 10, 729–743.
[48] Z. Hajjar, A. Khodadadi, Y. Mortazavi, S. Tayyebi, S. Soltanali, *Fuel*, **2016**, 179, 79–86.
[49] www.iza-structure.org/databases.
[50] Z.-M. Cui, Q. Liu, W. Song, L. Wan, *Angew. Chem. Int. Ed.*, **2006,** 45, 6512–6515.
[51] X. Sun, S. Mueller, Y. Liu, H. Shi, G. L. Haller, M. Sanchez-Sanchez, A. C. van Veen, J. A. Lercher, *J. Catal.*, **2014**, 317, 185–197.
[52] G. Yang, Y. Wei, S. Xu, J. Chen, J. Li, Z. Liu, J. Yu, R. Xu, *J. Phys. Chem. C*, **2013**, 117, 8214–8222.
[53] M. Yang, P. Tian, C. Wang, Y. Yuan, Y. Yang, S. Xu, Y. He, Z. Liu, *Chem. Commun.*, **2014**, 50, 1845.
[54] Z. Li, M. T. Navarro, J. Martínez-Triguero, J. Yu, A. Corma, *Catal. Sci. Technol.*, **2016**, 6, 5856–5863.
[55] P. Ferri, C. Li, R. Millán, J. Martínez-Triguero, M. Moliner, M. Boronat, A. Corma, *Angew. Chem. Int. Ed.*, **2020**, 59, 19708–19715.
[56] W. Zhang, S. Lin, Y. Wei, P. Tian, M. Ye, Z. Liu, *Natl. Sci. Rev.*, **2023**, 10, nwad120.
[57] C. Yuan, Y. Wei, J. Li, S. Xu, J. Chen, Y. Zhou, Q. Wang, L. Xu, Z. Liu, *Chin. J. Catal.*, **2012**, 33, 367–374.
[58] E. M. Gallego, M. T. Portilla, C. Paris, A. León-Escamilla, M. Boronat, M. Moliner, A. Corma, *Science*, **2017**, 355, 1051–1054.
[59] D. Chen, K. Moljord, T. Fuglerud, A. Holmen, *Microporous Mesoporous Mater.*, **1999**, 29, 191–203.
[60] D. Lesthaeghe, B. De Sterck, V. Van Speybroeck, G. Marin, M. Waroquier, *Angew. Chem. Int. Ed.*, **2007**, 46, 1311–1314.
[61] J. Chen, J. Li, Y. Wei, C. Yuan, B. Li, S. Xu, Y. Zhou, J. Wang, M. Zhang, Z. Liu, *Catal. Commun.*, **2014**, 46, 36–40.
[62] X. Li, Q. Sun, Y. Li, N. Wang, J. Lu, J. Yu, *J. Phys. Chem. C*, **2014**, 118, 24935–24940.
[63] G. E. P. Box, D. R. Cox, *J. R. Stat. Soc. B*, **1964**, 26, 211–252.
[64] J. Chen, J. M. Thomas, P. A. Wright, R. P. Townsend, *Catal. Lett.*, **1994**, 28, 241–248.
[65] M. Dusselier, M. A. Deimund, J. E. Schmidt, M. E. Davis, *ACS Catal.*, **2015**, 5, 6078–6085.
[66] N. Venkatathri, *Indian J. Chem. A,* **2002**, 41, 2223–2230.
[67] W. Yan, J. Yu, R. Xu, G. Zhu, F. Xiao, Y. Han, K. Sugiyama, O. Terasaki, *Chem. Mater.*, **2000**, 12, 2517–2519.
[68] N. K. McGuire, C. A. Bateman, C. Scott Blackwell, S. T. Wilson, R. M. Kirchner, *Zeolites*, **1995**, 15, 460–469.
[69] D. Xie, *Ind. Eng. Chem. Res.*, **2021**, 60, 15403–15415.
[70] P. Tian, L. Xu, Z. M. Liu, C. L. Sun, T. Huang, *Chem. J. Chin. Univ.,* **2001,** 22, 991–994.
[71] D. Wang, M. Yang, W. Zhang, D. Fan, P. Tian, Z. Liu, *CrystEngComm*, **2016**, 18, 1000–1008.
[72] R. F. Lobo, S. I. Zones, R. C. Medrud, *Chem. Mater.*, **1996**, 8, 2409–2411.
[73] M. Briend, R. Vomscheid, M. J. Peltre, P. P. Man, D. Barthomeuf, *J. Phys. Chem.*, **1995**, 99, 8270–8276.
[74] S. I. Zones, US 4,544,538, **1985**.
[75] M.-J. Díaz-Cabañas, P. A. Barrett, *Chem. Commun.*, **1998**, 1881–1882.
[76] L. Wang, D. Zhu, J. Wang, W. Cui, J. Han, B. Li, D. Fan, P. Tian, Z. Liu, *J. Mater. Chem. A*, **2021**, 9, 15238–15245.
[77] I. Yarulina, A. Dikhtiarenko, F. Kapteijn, J. Gascon, *Catal. Sci. Technol.*, **2017**, 7, 300–309.
[78] I. Yarulina, J. Goetze, C. Gücüyener, L. van Thiel, A. Dikhtiarenko, J. Ruiz–Martinez, B. M. Weckhuysen, J. Gascon, F. Kapteijn, *Catal. Sci. Technol.*, **2016**, 6, 2663–2678.
[79] D. Xie, L. B. McCusker, C. Baerlocher, S. I. Zones, W. Wan, X. Zou, *J. Am. Chem. Soc.*, **2013**, 135, 10519–10524.
[80] S. Lin, H. Li, P. Tian, Y. Wei, M. Ye, Z. Liu, *J. Am. Chem. Soc.*, **2025**, 147, 11585–11607.
[81] X. Wang, Y. Zuo, Z. Zhang, C. Wang, M. Yang, P. Tian, Z. Liu, software copyright 2024SR0082804, **2024**.
[82] X. Zhu, Y. Gao, M. Liu, Z. Yang, S. Li, H. Chen, B. Liu, W. Ma, E. J. M. Hensen, B. Shen, *Microporous Mesoporous Mater.*, **2022**, 345, 112260.
[83] E. Borodina, F. Meirer, I. Lezcano-González, M. Mokhtar, A. M. Asiri, S. A. Al-Thabaiti, S. N. Basahel, J. Ruiz-Martinez, B. M. Weckhuysen, *ACS Catal.*, **2015**, 5, 992–1003.

[84] X. Zhu, J. P. Hofmann, B. Mezari, N. Kosinov, L. Wu, Q. Qian, B. M. Weckhuysen, S. Asahina, J. Ruiz-Martínez, E. J. M. Hensen, *ACS Catal.*, **2016**, 6, 2163–2177.
[85] Z. Xu, H. Ma, Y. Huang, W. Qian, H. Zhang, W. Ying, *ACS Omega*, **2020**, 5, 24574–24583.
[86] F. H. Alshafei, Y. Park, S. I. Zones, M. E. Davis, *J. Catal.*, **2021**, 404, 620–633.
[87] J. Zhou, Y. Zhi, J. Zhang, Z. Liu, T. Zhang, Y. He, A. Zheng, M. Ye, Y. Wei, Z. Liu, *J. Catal.*, **2019**, 377, 153–162.
[88] C. Zhang, X. Wu, Y. Zhang, W. Zhang, S. Lin, C. Lou, S. Xu, D. He, L. Wang, Y. Wei, Z. Liu, *Chem. Catal.*, **2024**, 4, 101025.
[89] W. Huang, H. Wu, X. Yan, J. Zhang, Z. Feng, M. Yang, P. Tian, X. Guo, L. Ren, *Ind. Eng. Chem. Res.*, **2024**, 63, 15438–15447.

机器学习辅助发现具有超高初始选择性的甲醇制烯烃分子筛催化剂

王心怡[a,d,1], 王朝旗[b,1], 杨　淼[a,*], 王晓光[b,*], 左越中[b], 张壮壮[b], 吴一墨[a,e], 韩晶峰[a], 李　冰[a], 黄　玮[c], 任利敏[c], 魏迎旭[a], 刘欣梅[d], 田　鹏[a], 刘中民[a,e,*]

[a]中国科学院大连化学物理研究所, 低碳催化技术国家工程研究中心, 辽宁大连116023
[b]大连理工大学数学科学学院, 辽宁大连116024
[c]大连理工大学化工学院, 精细化工国家重点实验室, 辽宁大连116024
[d]中国石油大学(华东)化学化工学院, 重质油国家重点实验室, 山东青岛266580
[e]中国科学院大学, 北京100049

摘要: 随着甲醇制烯烃(MTO)工业化进程的持续推进和反应机理研究的不断深入, 设计合成优质的MTO催化剂以持续提升MTO工业过程的经济性和灵活调控烯烃产物分布已成为极具挑战的课题. 数据驱动模拟可以帮助研究者预测化学反应趋势和结果. 借助机器学习(ML)研究MTO催化问题, 有望打破以往经验桎梏, 带来反应机制和催化剂设计的新突破. 但要获得特定化学问题的具体指导策略, 仍要求研究者针对具体体系收集和整理实验数据、编码相关变量并重新训练专用模型.

本研究展示如何利用ML工作流程发现潜在的MTO分子筛催化剂. 首先建设分子筛催化MTO反应的数据库, 自变量包含分子筛结构与物性参数和反应条件等, 因变量包含甲醇转化率以及乙烯和丙烯选择性等. 其次构建自变量与因变量的关联模型. 基于数据特征训练了20多种机器学习模型、并对其准确性和可靠性进行评估和实验验证. 结果显示: 基于树类的集成学习方法在转化率(分类分析)和低碳烯烃选择性(回归分析)预测方面均表现出色, 准确率超过90%. 基于机器学习模型、开发出可用于预测分子筛MTO催化性能和最佳操作条件的软件, 方便分子筛催化剂研发. 为验证机器学习模型的泛化能力和准确性, 进一步合成了此前数据库中并未包含的STT分子筛, 使用软件预测最佳MTO评价条件及结果. 实验结果显示与预测值基本吻合, 证实了集成机器学习方法的可靠性和实用性. 从机器学习的可解释性出发, 利用决策树、特征重要性及SHAP分析, 提取获得高选择性的决策规则: 分子筛最大孔道不超过9元环(LRS≤9), 分子筛中可容纳球体的最大直径(MDi)在7.35–7.71 Å之间. 结合较低酸密度更为理想(A/T≤0.01)的结论, 重新聚焦高硅SSZ-13 (Si/Al = 100)分子筛, 意外发现该材料在450 ºC, 空速1 h^{-1}条件下展现出超高的初始低碳烯烃选择性(反应37 min时乙烯加丙烯选择性高达87.6%, 其中乙烯选择性高达61.1%), 这样高的单乙烯选择性在MTO反应中极为罕见.

综上, ML在分子筛催化剂研发方面展示出有效性和应用潜力, 它可以避开复杂的反应机理, 降低实验室试错成本, 并为解决特定催化剂问题提供新机遇. 然而, 我们也认识到, ML方法高度依赖数据质量和数量. 在有限数据情况下, 还必须结合领域专家的专业知识与判断, 才能获得创新性成果.

关键词: 甲醇制烯烃; 分子筛; 机器学习; 理论设计与合成; 预测

收稿日期: 2025-08-15. 接受日期: 2025-09-22. 上网时间: 2026-02-05.

*通讯联系人. 电子信箱: yangmiao@dicp.ac.cn (杨淼), wangxg@dlut.edu.cn (王晓光), liuzm@dicp.ac.cn (刘中民).
[1]共同第一作者.

基金来源: 国家重点研发计划(2024YFE0207000); 国家自然科学基金(22171259, 12461050, 21991090, 21991091, 22288101, 22372020); 榆林中科洁净能源创新研究院人工智能科技专项资助(DNL-YL A202206).