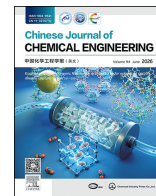




Contents lists available at ScienceDirect

Chinese Journal of Chemical Engineering

journal homepage: www.elsevier.com/locate/CJChE

Full Length Article

Data-efficient and interpretable machine learning for iridium-catalyzed methanol carbonylation: Ru-enabled operating windows and process-ready guidance

Xiaokun Zhou, Tao Yang, Yao Qin, Yue Li, Qiyue Wei, Jinhui Xie, Yanxun Wu, Yuchang Liu, Zengxi Wei^{*}, Shuangliang Zhao^{*}

University Engineering Research Center of Green Chemical New Materials, School of Chemistry and Chemical Engineering, Guangxi University, Nanning 530004, China

ARTICLE INFO

Article history:

Received 7 November 2025

Received in revised form

6 January 2026

Accepted 6 January 2026

Available online 26 February 2026

Keywords:

Methanol carbonylation

Iridium catalysis

Ruthenium cocatalyst

Acetic acid

Data augmentation

ABSTRACT

Methanol carbonylation is the primary method for producing acetic acid, but the formulation screening and process control for homogeneous iridium systems are hindered by limited and heterogeneous datasets. We compile literature and patent records and construct stratified models based on the presence of a ruthenium cocatalyst, integrating data augmentation with Bayesian model selection to achieve high-accuracy, interpretable predictors of formation rate. Multi-scale interpretation, (including rank correlations and SHAP analysis, reveals that the addition of ruthenium (Ru) alters rate control: the *in-situ* water level emerges as the primary driver through its influence on acidity/iodide activity and methyl iodide (MeI) regeneration. Consequently, the marginal effectiveness of iridium (Ir) increases, whereas the independent impact of initial acetic acid (AcOH)/MeI is diminished due to the swift establishment of solvent and MeI pools. Guided by these findings, we define Ru-enabled operating windows for Ru-containing systems, prioritizing *in-situ* concentration management by keeping H₂O at a relatively low but non-zero level and maintaining MeOAc in a MeOAc-rich region where the SHAP curves begin to flatten, indicating diminishing returns. On this basis, fine-tune Ir loading and pressure in small increments to enhance per-Ir effective turnover without disturbing the established *in-situ* balance. © 2026 The Chemical Industry and Engineering Society of China, and Chemical Industry Press Co., Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

1. Introduction

1.1. Industrial methanol carbonylation

Acetic acid (CH₃COOH) is among the highest-volume and most widely used basic chemicals in modern industry [1–3]. At present, the methanol carbonylation route is the dominant production technology, accounting for over 85% of global output [4]. Rhodium-based catalytic systems have been widely employed in this process [5,6]. However, these catalysts are expensive and prone to precipitation-induced deactivation during operation, leading to substantial economic losses [7,8]. In recent years, iridium-based catalysts have exhibited excellent catalytic performance at lower

cost and thus hold promise as viable alternatives to rhodium-based systems [9,10].

In industrial practice, iridium-based methanol carbonylation systems comprise three principal components. (i) The catalytically active center is an iodide-ligated iridium complex that mediates the key elementary steps of methylation, CO insertion, and product-forming reductive elimination [11]. (ii) The iodine cycle, consisting primarily of methyl iodide (MeI) and hydrogen iodide/iodide (HI/I[−]), supplies the methylating agent and helps maintain the iridium species in its active form [12]. (iii) The solvent medium and promoters: reactions are conducted in acetic acid/methyl acetate under low-water conditions, with small amounts of ruthenium (Ru) optionally added as a cocatalyst to enhance turnover [13].

In terms of performance, the catalytic activity is governed mainly by the iridium loading, MeI concentration, CO partial pressure, and the effective activity of iodide; a small amount of water facilitates hydrolysis of acetyl iodide to acetic acid [4]. Selectivity is dictated by the coordination environment of the

^{*} Corresponding authors.

E-mail addresses: zengxiwei@gxu.edu.cn (Z. Wei), szhao@gxu.edu.cn (S. Zhao).

catalyst: the electronic and steric properties of the ligands modulate binding/release of $I^-/MeI/CO$ and the availability of vacant sites at the metal center, thereby reshaping the relative ease of oxidative addition, CO insertion, and reductive elimination and determining whether the main reaction or side reactions prevail [11,12]. Catalyst stability and speciation, in turn, require a proper balance of water and iodine: insufficient levels promote aggregation and deactivation of iridium species, whereas excessive water can cause corrosion and increase side reactions [14].

1.2. Challenges in Ir-based systems

Nevertheless, formulation screening and process optimization for homogeneous iridium systems remain constrained by three interlocking bottlenecks: (i) incomplete understanding of the structure-activity relationships of key active species and their deactivation pathways, compounded by sparse and highly heterogeneous supporting data [15]; (ii) reliance on empirical trial-and-error coupled with low-throughput experimentation, whereby optimizing a single formulation can require thousands of experiments over multiple years at prohibitive cost [16–18]; and (iii) intrinsic trade-offs among activity, selectivity, and stability, which hinder simultaneous multi-objective optimization and robustness across operating windows. These challenges are especially acute in complex catalyst formulations where cocatalysts, solvents, iodides, water, and esters are strongly coupled variables.

1.3. Data-driven methods and Mixup

In recent years, catalyst formulation discovery has been shifting from experience-driven practice to an integrated paradigm that combines high-throughput experimentation, computational chemistry, and data-driven modeling. Automation and micro-channel platforms have markedly increased the efficiency of data acquisition [19,20]; density functional theory (DFT) and molecular simulations continue to deepen mechanistic insight into electronic structure, transition states, and solvent effects [21]; and machine-learning and materials-genomics approaches offer a path to overcome the combinatorial explosion from composition and structure to performance, enabling rapid predictive modeling and the inverse design of actionable process-control strategies [22]. However, for catalytic systems where data are expensive to obtain, limited in quantity, and unevenly distributed, constructing high-accuracy, interpretable models under small-sample constraints remains a central challenge for practical deployment [23,24].

To address this challenge, the machine learning community has advanced strategies such as transfer learning, semi-supervised learning, and data augmentation to enhance model generalization and robustness [25–27]. Among these, Mixup generates synthetic samples by linearly interpolating both features and labels—effectively creating examples “between” real samples. This increases the density of learnable decision boundaries without altering the overall data distribution, thereby mitigating overfitting to idiosyncratic noise and chance artifacts [28]. Mixup has shown strong performance across diverse domains: Smucny *et al.* [29] applied Mixup to fMRI-based prediction of treatment response in psychiatric disorders, raising accuracy from 76.5% to 80.1% while reducing overfitting and improving generalization; Dong *et al.* [30] proposed MixCode, a Mixup-based approach for source-code classification, achieving a 6.24% accuracy gain and a 26.06% improvement in robustness; and Wang *et al.* [31] incorporated Mixup into a finger-vein recognition system, reaching 99.51% detection accuracy.

1.4. Scope and contributions

We compiled 380 iridium-based methanol carbonylation entries from the literature and patents and stratified them by the presence of a ruthenium cocatalyst (Ru): 144 with Ru and 236 without Ru. Targeting the acetic-acid formation rate, we built parallel prediction-and-interpretation workflows for the two subsets. After systematically comparing six candidate algorithms (integrating Mixup augmentation and Bayesian optimization, and benchmarking by cross-validated R^2/MAE and robustness), the optimal configurations were identified as follows: support vector regression (SVR) for the Ru-containing subset and a recurrent neural network (RNN) for the Ru-free subset. We then quantified main effects and interactions using Kendall rank correlations and SHAP analyses to test whether the rate-determining factors were reordered upon Ru introduction and to elucidate the underlying mechanistic shifts. Based on the above results, we provide process adjustment suggestions for both the Ru-containing and Ru-absent scenarios to support the stable operation and formulation optimization of the continuous production facility under a wider range of operating conditions.

2. Results and Discussion

2.1. Dataset and feature design

This study compiles 380 iridium-based carbonylation records from the open literature and patents, comprising 144 entries with Ru and 236 without Ru. The prediction target is the acetic-acid formation rate ($\text{mol} \cdot \text{L}^{-1} \cdot \text{h}^{-1}$). To align with mechanistic and process characterization, we model the two subsets separately, using an identical set of input features while distinguishing by the presence or absence of Ru: Pressure, AcOH, MeI, MeOAc(1), $\text{H}_2\text{O}(1)$, Ir, (Ru, for the Ru-containing subset only), MeOAc(2), and $\text{H}_2\text{O}(2)$. Here, (1) denotes the initial charge mass, (2) denotes the concentration at the time the rate is measured, and variables without a superscript refer to initial charges.

The overall machine learning workflow of this study is shown in Fig. 1. We first integrate iridium-based methanol carbonylation data from the literature and patents (partitioned into Ru-containing and Ru-free subsets) and perform data cleaning and standardization. Within this grouped framework, we apply Mixup augmentation ($\alpha = 0.1\text{--}0.4$, expansion factor 2–6) and construct six candidate models. Bayesian optimization is then used to efficiently search key hyperparameters, with model selection based on cross-validation and external testing. After identifying the optimal model for each subset (with Ru: SVR; without Ru: RNN), we conduct multi-scale interpretability analyses using Kendall rank correlation and SHAP. We quantify the relative magnitudes of the positive effect of Ir on the rate and the negative effects of $\text{H}_2\text{O}(2)$ and MeOAc(2). The results show that, upon introducing Ru, these dependencies become weaker, yielding a system that is more stable and easier to control. Finally, we perform residual diagnostics and package the optimal models into a web-based rate-prediction tool to support process decision-making.

2.2. Mixup augmentation and robustness evaluation

Given the limited sample size and strong variable coupling, we adopt Mixup, which performs linear interpolation in both feature and label spaces [32,33]. Specifically, given two samples x_i and x_j with corresponding labels y_i and y_j , Mixup generates a new sample $\tilde{x}$ and label $\tilde{y}$:

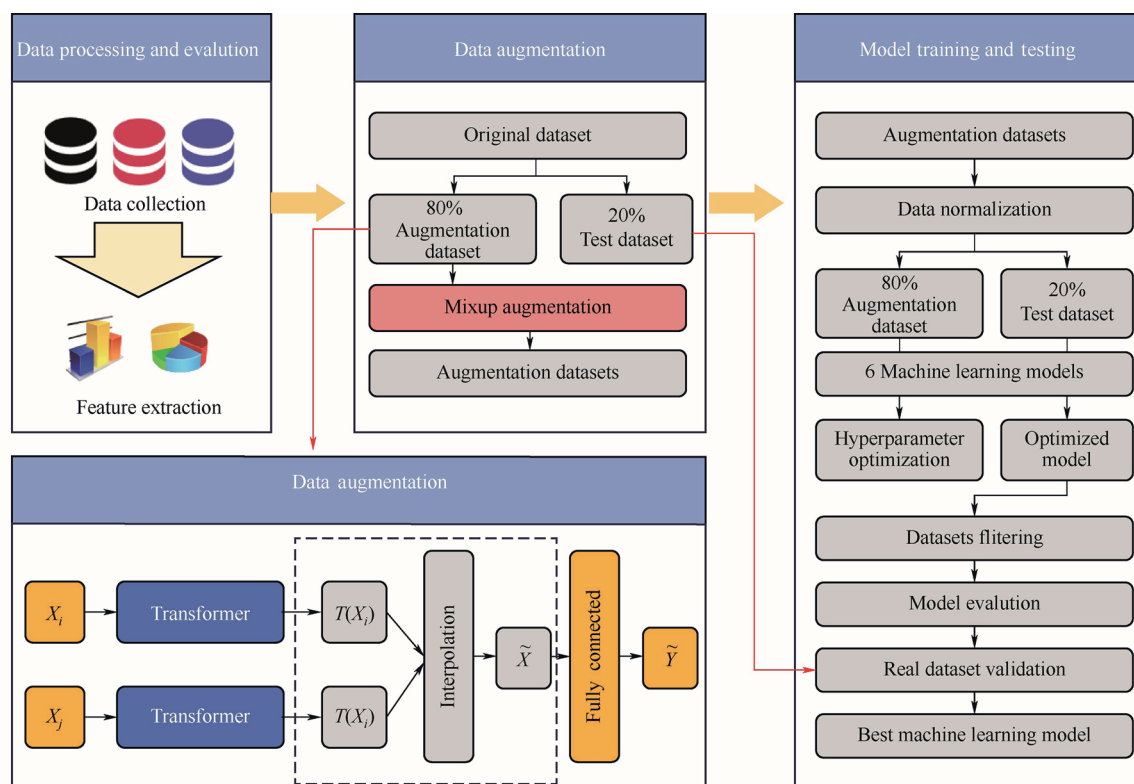


Fig. 1. Overall machine learning workflow.

$$\tilde{\mathbf{x}} = \lambda \mathbf{x}_i + (1 - \lambda) \mathbf{x}_j$$

$$\tilde{\mathbf{y}} = \lambda \mathbf{y}_i + (1 - \lambda) \mathbf{y}_j$$

$$\lambda = \text{Beta}(\alpha, \alpha)$$

We set $\alpha = 0.1$ – 0.4 and an expansion factor of 2 – 6 , generating 20 augmented datasets for each of the Ru-containing and Ru-free subsets (40 in total) to assess model robustness (see Fig. 2). To avoid physically infeasible samples, Mixup pairing and weight sampling were restricted within the same subset, and Mixup-generated samples were constrained to remain within the observed minimum and maximum values of each feature in the original dataset, ensuring that synthetic formulations stayed within the experimentally explored operating domain (thus avoiding unrealistically high MeI or vanishing water contents).

A comprehensive comparison shows that Mixup delivers consistent performance gains: most models exhibit higher R^2 , lower MAE, and substantially reduced variance. In the Ru-containing subset, SVR improves markedly from ($R^2 = 0.950$, MAE = 1.47) to ($R^2 = 0.985$, MAE = 0.363), with notable error contraction also observed for random forest (RF) and AdaBoost. For CatBoost, R^2 decreases slightly while MAE declines, a typical pattern indicating reduced model variance accompanied by a minor increase in bias after augmentation. On the Ru-free data, improvements are smaller but directionally consistent; the best-performing model shifts from kernel/tree methods to an RNN ($R^2 = 0.958$, MAE = 0.73). This suggests that in the absence of Ru, stronger inter-variable couplings, error magnification with rate, and greater relational complexity favor an RNN, which is better suited to capturing complex nonlinearities.

2.3. Model comparison and Bayesian optimization

We employed Bayesian optimization for hyperparameter search, targeting maximization of validation R^2 and minimization of MAE, with performance assessed via 5-fold cross-validation [34–36]. A fixed search budget was used (an initial set of random/uniform samples, followed by a total of 1000 evaluations), and all models were compared under identical data splits and random seeds. The search was terminated once metric improvements plateaued. Apart from the search space and budget, all other implementation details followed standard defaults.

In the Ru-containing subset, SVR with Mixup ($\alpha = 0.1$, $4\times$ expansion) achieved the best performance ($R^2 = 0.985$, MAE = 0.363), with optimal hyperparameters $C = 43.91$, $\epsilon = 0.00164$, $\gamma = 0.0642$, and an RBF kernel. In the Ru-free subset, an RNN under the same augmentation strategy reached $R^2 = 0.958$ and MAE = 0.73 , with an optimal architecture of hidden units $[106, 116, 99]$ and training hyperparameters: learning rate = 3.19×10^{-4} , batch size = 16 , dropout = 0.115 , and weight decay = 6.65×10^{-5} . In both subsets, the Mixup settings were chosen based on preliminary trials rather than an exhaustive hyperparameter search. We examined α values in the range 0.1 – 0.4 and found, as expected, that smaller α (i.e., synthetic samples closer to the original data) generally led to better performance; $\alpha = 0.1$ gave the highest cross-validated R^2 without signs of overfitting, whereas larger α values did not provide additional benefit and in some cases slightly degraded accuracy. For the expansion factor, augmentation levels from $2\times$ to $6\times$ were tested; the average performance improved as the expansion factor increased and reached its maximum at $4\times$, but started to decline beyond this point, suggesting that overly aggressive augmentation begins to harm generalization.

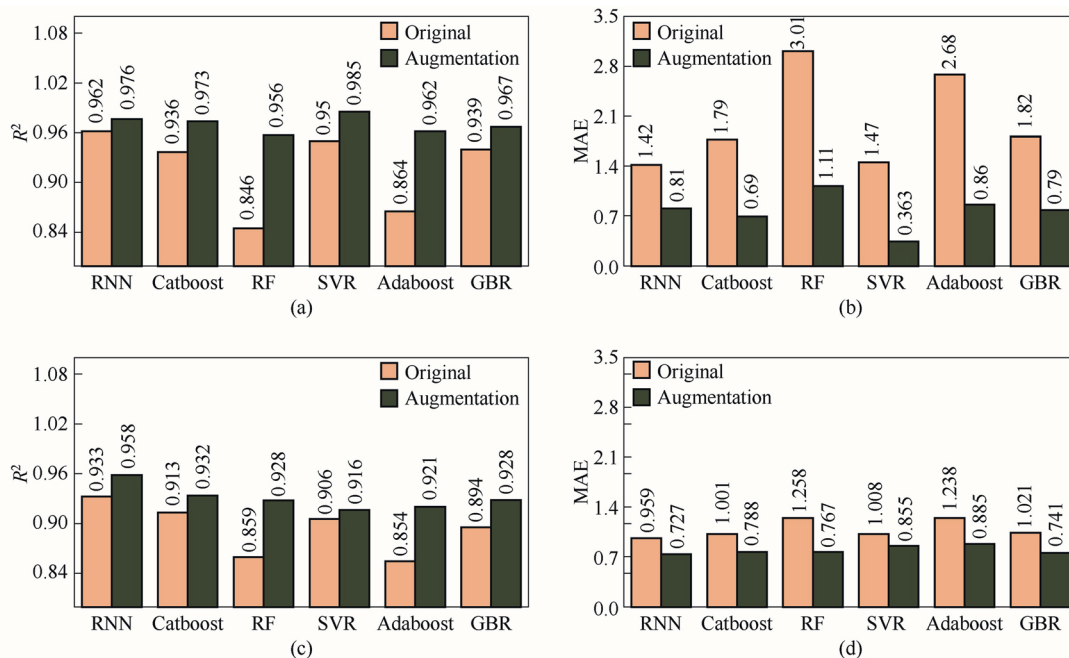


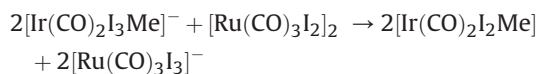
Fig. 2. Performance comparison of six models on the methanol carbonylation dataset before and after Mixup augmentation: (a), (b) Ru-containing subset. (c), (d) Ru-free subset. Orange: trained on original data; green: trained on augmented data.

Accordingly, we adopted $\alpha = 0.1$ and a $4\times$ expansion factor in the final reported models as a practical compromise, and did not observe any change in the ranking of the candidate algorithms within this parameter window. From a bias-variance perspective, SVR in the Ru-containing case exhibits low variance and strong decision-boundary characterization, whereas the RNN in the Ru-free case better accommodates nonlinearity and heteroscedasticity—findings that are corroborated by the residual diagnostics presented later.

2.4. Feature interaction analysis

To statistically elucidate which variables dominate the rate variations under different formulation scenarios and how they interact. We combine two complementary tools: Kendall's τ to capture rank correlations and robustly characterize monotonic relationships among variables, and SHAP, a game-theoretic approach that quantifies each feature's marginal contribution to both individual predictions and the global model [37–40] (see Fig. 3).

Upon introduction of Ru, the primary rate-determining factors are reordered because Ru mainly acts on steps $2 \rightarrow 4$ of the catalytic cycle (Fig. 4). Water rises to the top in global importance, the importance of Ir increases markedly, while the initial charges of acetic acid and methyl iodide diminish in importance. Mechanistically, Ru serves as an iodide acceptor and promotes CO coordination, facilitating conversion of the anionic precursor $[\text{Ir}(\text{CO})_2\text{I}_3\text{Me}]^-$ into the neutral, rapidly inserting tricarbonyl species $[\text{Ir}(\text{CO})_3\text{I}_2\text{Me}]$. In other words, the consecutive steps “deiodination $\rightarrow$ CO coordination $\rightarrow$ migratory insertion” (Fig. 4) are accelerated. The key iodide-abstraction step can be written as:



Owing to a reduction in idle or inhibited Ir sites and a concomitant increase in per-Ir effective turnover, Ir regains a leading rank in feature importance.

According to Sunley & Watson (2000), within the low-to-moderate water regime relevant to this study, water content is positively correlated with the steady-state HI level (decreasing water lowers HI) [41]. At the extreme low-water limit, insufficient HI diverts the system into the neutral cycle and depresses the rate; thus, the dependence is not strictly monotonic at the boundary. Accordingly, our discussion is confined to the positive-correlation range within the operational window. Because HI participates as a Brønsted acid in the regeneration loop $\text{HI} + \text{MeOAc} \rightleftharpoons \text{MeI} + \text{AcOH}$ changes in water content shift acid–base and solvation equilibria to raise $a(\text{H}^+)$ and the steady-state HI level, thereby increasing the MeI regeneration flux. In the presence of Ru, this effect is amplified: Ru's iodide-accepting role and promotion of CO coordination lower the barrier for the $2 \rightarrow 4 \rightarrow$ insertion sequence, while Ru in acidic media helps sequester free I^- and accelerates the acid-catalyzed loop above. Consequently, the sensitivity of the overall rate to $a(\text{H}^+)/\text{HI}$ is further enhanced, and water—by governing $a(\text{H}^+)$ and HI—assumes a markedly greater marginal influence on the rate.

It is important to note that the MeI variable in the diagram represents the initial charge, not the equilibrium concentration during the reaction. Under conditions with ruthenium, MeI is quickly regenerated and sustained through the acid-catalyzed loop $\text{HI} + \text{MeOAc} \rightleftharpoons \text{MeI} + \text{AcOH}$. Furthermore, oxidative addition is not the rate-determining step; consequently, once the initial MeI surpasses the threshold of sufficiency, its additional impact on the reaction rate becomes negligible, manifesting as a low global importance. This does not reduce the mechanistic significance of MeI. Instead, it suggests that its efficacy is dictated by the regeneration rate rather than the initial concentration.

AcOH functions as both a solvent medium and a product. Upon the introduction of Ru, the flux through the main cycle significantly increases, enhancing the production rate of AcOH and rapidly establishing a steady-state solvent environment composed of $\text{AcOH}/\text{H}_2\text{O}/\text{MeOAc}$. Consequently, the marginal influence of the initial AcOH charge on the subsequent rate is diminished. Simultaneously, system acidity is primarily governed by the dynamic

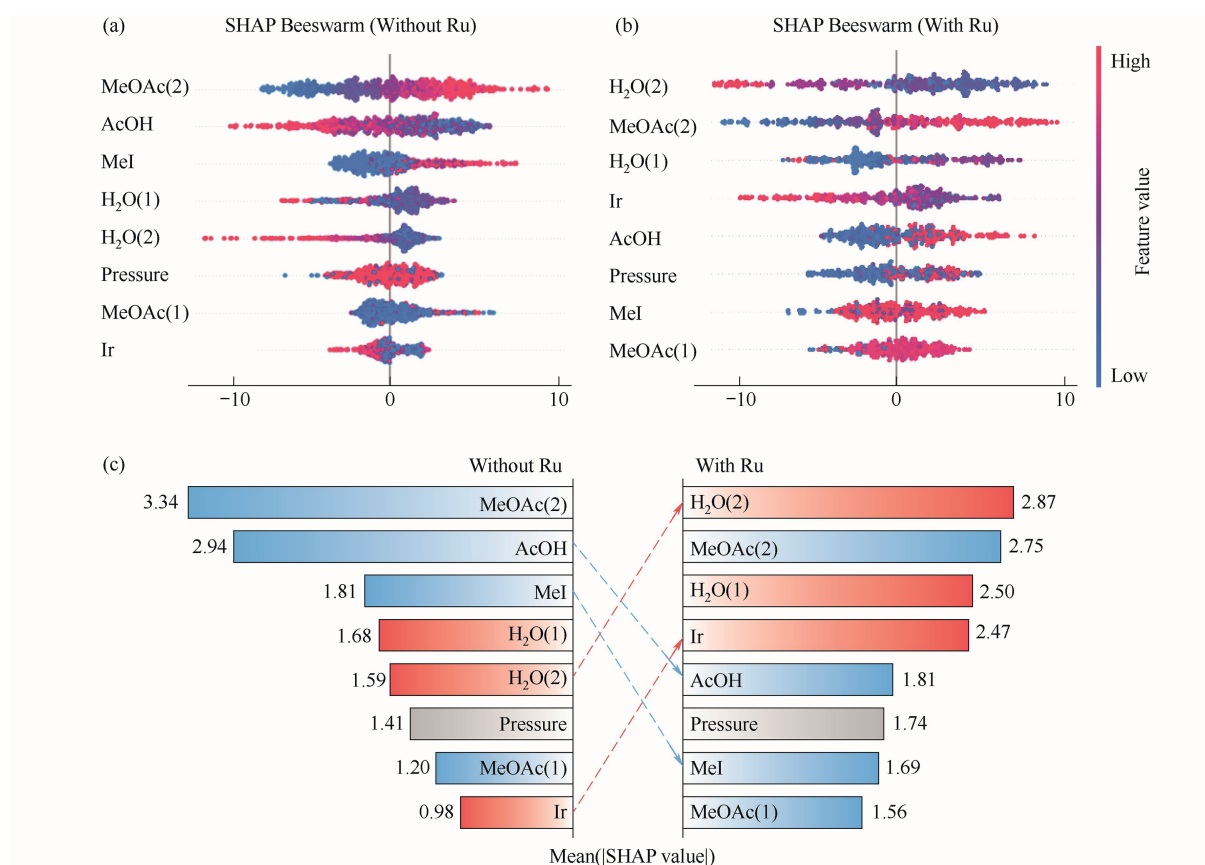


Fig. 3. (a) SHAP beeswarm plot (without Ru); (b) SHAP beeswarm plot (with Ru). (c) Global feature importance (Mean |SHAP|): values denote the mean absolute contribution of each feature to the predicted shift in the acetic-acid formation rate; dashed arrows indicate changes in ranking after adding Ru. MeOAc(1) and H₂O(1) are initial charge masses, while MeOAc(2) and H₂O(2) are *in-situ* concentrations at the time of rate measurement.

pool of strong acids (HI/H₃O⁺), rather than being directly set by the initial AcOH. Collectively, these factors reduce the global importance of AcOH (initial charge) to a moderate level under Ru-containing conditions.

Overall, the two correlation heatmaps (left: without Ru; right: with Ru) differ only slightly: most cells are near zero, indicating that feed variables were largely tuned independently in the experiments (see Fig. 5). The addition of Ru did not result in variables becoming “locked” into co-varying in either an upward or downward direction. A few pairwise relationships have been revised. The negative association between MeI and AcOH has been weakened (with Ru, faster MeI regeneration means AcOH no longer pulls the equilibrium back as strongly). The negative correlation between Ir and the downstream water level H₂O(2) has been slightly strengthened (when water levels are higher, the marginal gain from adding more Ir is smaller). The coupling between water and MeOAc has been weakened overall (water more directly sets acidity/iodide activity, diminishing MeOAc’s influence). And a mild positive correlation has emerged between Pressure and H₂O(1) (a weak operational linkage of modest pressurization within the low-water window).

In essence, Ru alters which factors are significant—and under which conditions—rather than compelling variables to uniformly increase or decrease, aligning with the SHAP results that indicate “water becomes more critical, Ir more impactful, and the independent role of AcOH diminishes.”

In summary, with the introduction of Ru, rate control transitions from the MeI/AcOH material ratios to a predominantly single-axis control governed by water (through its influence on a(H⁺) and the effective concentrations of I[−] and MeOAc (2). In practice, operations should focus on coordinated adjustments of water and MeOAc (2) within a low-but-nonzero water regime, followed by incremental fine-tuning of the Ir dosage to achieve optimal yield. MeI should be maintained at a sufficient level to prevent starvation in the 1 → 2 step of Fig. 4. However, an excess of MeI could increase HI/iodide activity and partially negate the advantages.

2.5. Predictive performance and residual diagnostics

Parity Plot for both the Ru-containing and Ru-free datasets cluster tightly along the diagonal, with narrow 95% confidence and prediction bands. Calibration is stronger for the Ru-containing case ($R^2 = 0.985$, MAE = 0.363), while the Ru-free case also shows good agreement ($R^2 = 0.958$, MAE = 0.735). Insets indicate slight conservatism (modest underestimation) in the low-rate regime for both models, but the magnitude is limited. Residual plots show residuals centered around zero with no systematic drift across the prediction range, indicating no functional-form bias; the broader spread in the Ru-free residuals suggests mild heteroscedasticity and higher noise. Holdout-set validation remains highly consistent: With Ru, $R^2 = 0.975$ and

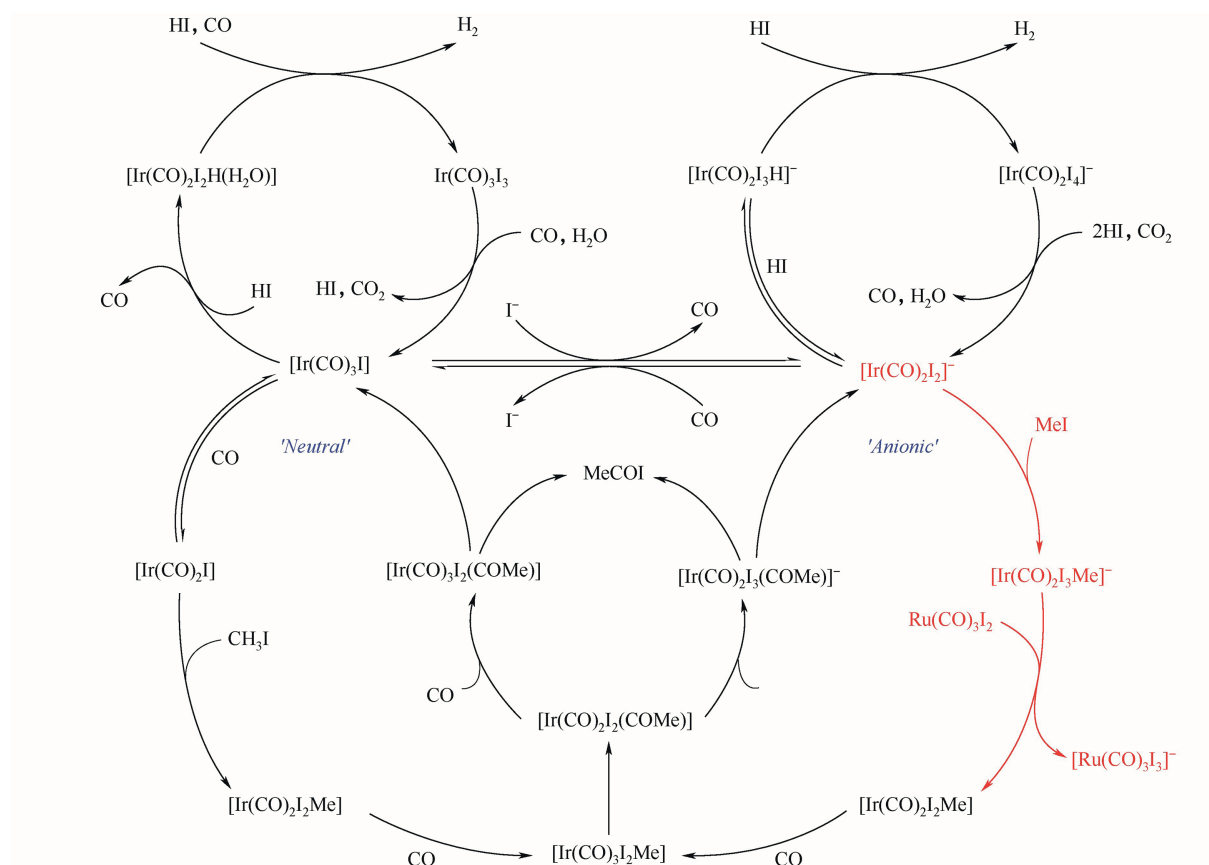


Fig. 4. Förster's mechanism for iridium-catalyzed methanol carbonylation.

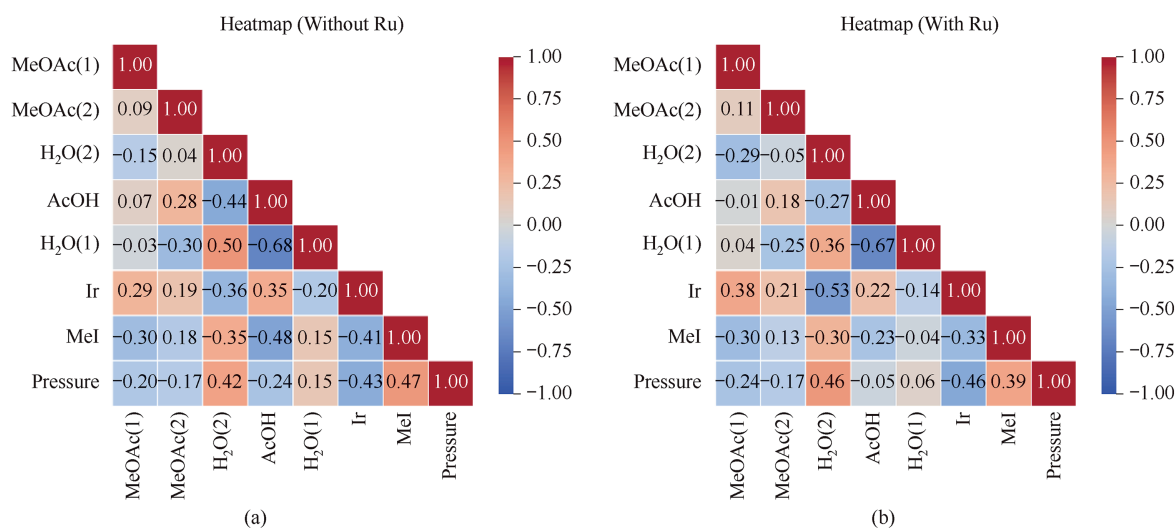


Fig. 5. Feature correlation heatmaps (Kendall's τ). (a) Without Ru; (b) With Ru. Color scale: red = positive correlation, blue = negative correlation, gray ≈ 0 ; numbers denote Kendall's τ .

MAE = 1.464; Without Ru, $R^2 = 0.958$ and MAE = 0.895. Overall, the models exhibit sound calibration and robustness across training/validation and external data, providing a reliable foundation for SHAP based mechanistic interpretation and process optimization (see Fig. 6).

2.6. Guidelines for process decision-making

The combined evidence from Kendall's τ correlation analysis, SHAP-based feature attributions, and residual structure examination highlights several regulatory priorities under Ru-promoted

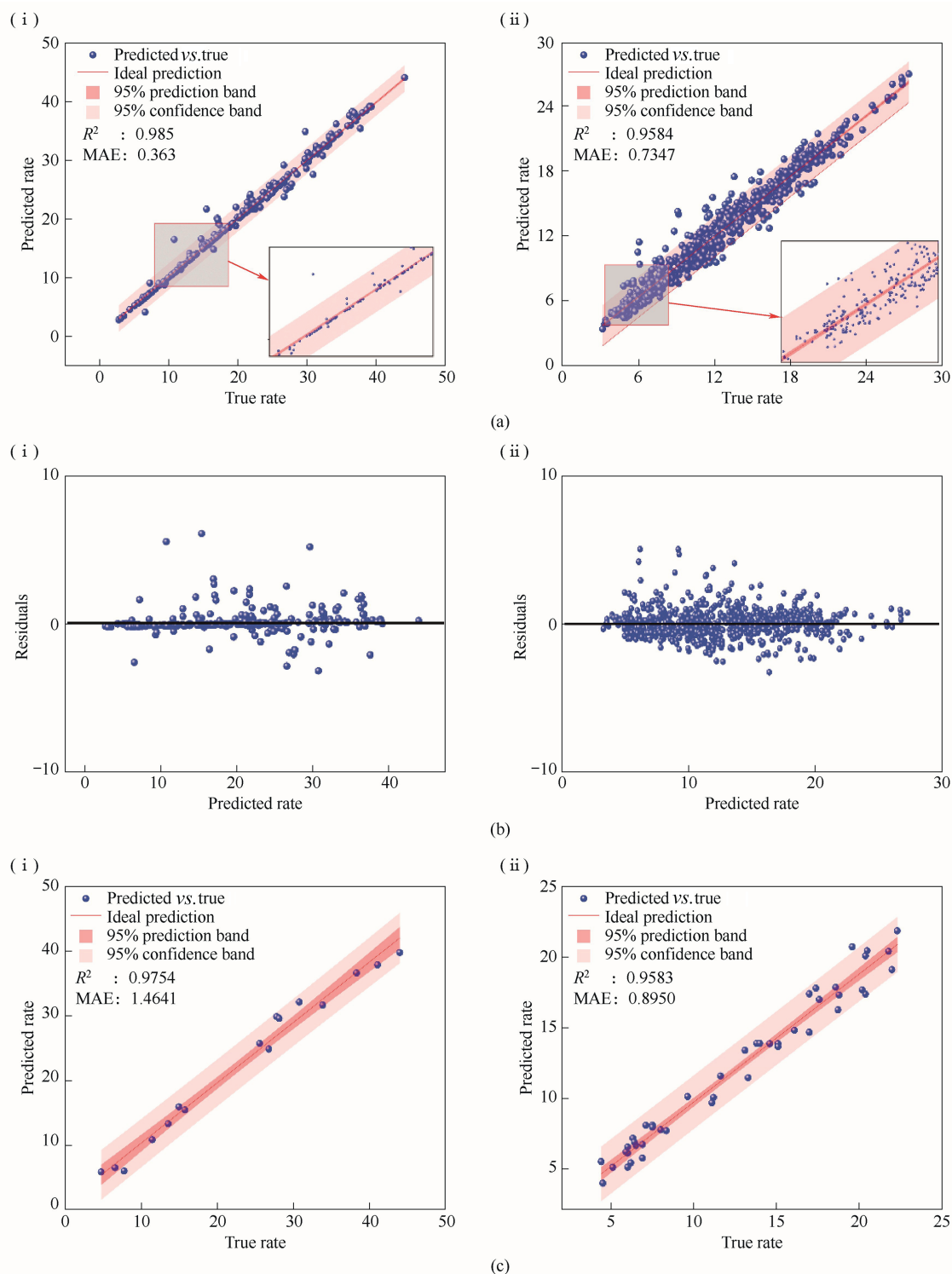


Fig. 6. Model predictive performance and residual diagnostics. (a) Predicted vs. observed scatter with 95% confidence and prediction bands; (b) Residuals vs. predictions; (c) External test-set scatter. Left column (i): with Ru; right column (ii): without Ru.

conditions. The introduction of Ru facilitates iodide abstraction, CO coordination, and neutral migratory insertion, thereby lowering the barrier of the main catalytic cycle. Under these circumstances, process optimization should primarily focus on the fine regulation of *in-situ* acidity and iodide activity. Maintaining $\text{H}_2\text{O}(2)$ at a low

yet non-zero concentration is essential to sustain the $\text{HI}/\text{a}(\text{H}^+)$ balance and ensure sufficient MeI regeneration flux. $\text{MeOAc}(2)$ should be adjusted to the threshold beyond which additional increases no longer yield appreciable benefits, thus avoiding unnecessary dilution or side reactions. In contrast, Ir loading and

system pressure should only be modified incrementally to improve the turnover efficiency per Ir center, as excessive pressure adjustments may perturb the established steady-state equilibrium. Finally, the initial MeI concentration should be set at the minimum level required to avoid feed limitation, while the initial AcOH concentration can remain within the conventional operating window, consistent with the Ru-free baseline.

Furthermore, acknowledging the significance of managing uncertainty in industrial applications, we have encapsulated the two optimal models into a web-based rate-prediction tool. This tool includes cross-validation-based interval outputs and input-compliance checks, which enable a swift evaluation of rates and uncertainties across various formulations and operating conditions. It also supports plant engineers in making informed adjustments to the operating window.

3. Conclusions

Based on 380 records of iridium-catalyzed methanol carbonylation, we constructed stratified models considering the presence or absence of a Ru cocatalyst. Within a combined framework of Mixup augmentation and Bayesian optimization, we established a high-accuracy, interpretable, and practically deployable predictor for the acetic acid formation rate. Support Vector Regression (SVR) proved optimal for the Ru-containing subset ($R^2 = 0.985$, MAE = 0.363), while a Recurrent Neural Network (RNN) was optimal for the Ru-free subset ($R^2 = 0.958$, MAE = 0.73).

Leveraging multi-scale interpretation through SHAP and Kendall's τ , we offer statistical evidence and mechanistic insights into the reordering of rate-controlling factors following the introduction of Ru: the *in situ* water level emerges as the primary feature (by modulating $a(\text{H}^+)$ /effective I^- and enhancing the MeI regeneration flux); the marginal return of Ir increases (Ru abstracts iodide, facilitates CO coordination, and activates a low-barrier pathway consisting of sequential iodide abstraction, CO coordination, and migratory insertion); whereas the independent effects of initial AcOH and initial MeI diminish (the system quickly establishes steady solvent and MeI pools with continuous replenishment). Correlation heatmaps suggest that the overall covariation structure largely remains intact, although several pairs associated with the acidity/iodide network have been restructured (for example, a weaker MeI-AcOH anticorrelation, partial decoupling of H_2O -MeOAc, and a slightly more negative Ir- $\text{H}_2\text{O}(2)$), which aligns with the SHAP findings.

Accordingly, we propose two actionable process guidelines. Utilizing Ru, optimization is shifted from front-end ratios and total pressure towards Ru-enabled operating windows in which *in-situ* $\text{H}_2\text{O}(2)$ and MeOAc(2) act as the primary control levers: $\text{H}_2\text{O}(2)$ should be kept in a relatively low but non-zero region of its observed range, while MeOAc(2) should be maintained in a relatively high region where the SHAP profiles begin to flatten and further increases provide only modest additional benefit, and Ir loading and pressure then serve mainly as secondary, fine-tuning knobs within conventional industrial ranges. Without Ru, moderately increase the availability of AcOH/MeI/MeOAc to accelerate the rate while preventing short-lived spikes in H_2O and MeOAc(2). Finally, we encapsulated the optimal models in a web-based rate-prediction tool, enabling plant engineers to rapidly evaluate rates under varying formulations and operating conditions and to adjust the operating window accordingly.

CRedit Authorship Contribution Statement

Xiaokun Zhou: Writing – original draft. Tao Yang: Conceptualization. Yao Qin: Data curation. Yue Li: Data curation. Qiyue Wei:

Conceptualization. Jinhui Xie: Investigation. Yanxun Wu: Formal analysis. Yuchang Liu: Methodology. Zengxi Wei: Writing – review & editing. Shuangliang Zhao: Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

S. Zhao and Z. Wei thanks the support from Guangxi Science and Technology Major Program (AA23062019, AA23073019, AA23062020). This work is also supported by the National Natural Science Foundation of China (22308063, 22178072). Z. Wei also thanks the support from the Dean Project of Guangxi Key Laboratory of Petrochemical Resource Processing and Process Intensification Technology (2023Z007). This work is also supported by the Guangxi Major Talent Project.

Supplementary Material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cjche.2026.01.005>.

References

- [1] A. Haynes, Catalytic methanol carbonylation, *Advances in Catalysis*, Elsevier (2010).
- [2] A.C. Dimian, A.A. Kiss, Enhancing the separation efficiency in acetic acid manufacturing by methanol carbonylation, *Chem. Eng. Technol.* 44 (10) (2021) 1792–1802.
- [3] A. Haynes, Acetic acid synthesis by catalytic carbonylation of methanol, *Catalytic Carbonylation Reactions*, Springer, (2006).
- [4] P. Kalck, C. Le Berre, P. Serp, Recent advances in the methanol carbonylation reaction into acetic acid, *Coord. Chem. Rev.* 402 (2020) 213078.
- [5] J.H. Jones, The cativa™ process for the manufacture of acetic acid, *Platin. Met. Rev.* 44 (2000) 94–105.
- [6] P.I.P. Elliott, S. Haak, A.J.H.M. Meijer, G.J. Sunley, A. Haynes, Reactivity of Ir(III) carbonyl complexes with water: alternative by-product formation pathways in catalytic methanol carbonylation, *Dalton Trans.* 42 (47) (2013) 16538–16546.
- [7] H. Cheung, M.E. Huckman, G.P. Torrence, Low water methanol carbonylation process for high acetic acid production and for water balance control, US Patent 7005541B2 (2006).
- [8] S.S. Zhang, W.X. Ji, N. Feng, L.P. Lan, Y.Y. Li, Y.L. Ma, Study on Rh(I)/Ru(III) bimetallic catalyst catalyzed carbonylation of methanol to acetic acid, *Materials* 13 (18) (2020) 4026.
- [9] N. Yoneda, S. Kusano, M. Yasui, P. Pujado, S. Wilcher, Recent advances in processes and catalysts for the production of acetic acid, *Appl. Catal. Gen.* 221 (1–2) (2001) 253–265.
- [10] M. Volpe, G. Wu, A. Iretskii, P.C. Ford, Photochemical and time resolved spectroscopic studies of intermediates relevant to iridium-catalyzed methanol carbonylation: photoinduced CO migratory insertion, *Inorg. Chem.* 45 (4) (2006) 1861–1870.
- [11] A. Haynes, P.M. Maitlis, G.E. Morris, G.J. Sunley, H. Adams, P.W. Badger, C.M. Bowers, D.B. Cook, P.I.P. Elliott, T. Ghaffar, H. Green, T.R. Griffin, M. Payne, J. M. Pearson, M.J. Taylor, P.W. Vickers, R.J. Watt, Promotion of iridium-catalyzed methanol carbonylation: mechanistic studies of the cativa process, *J. Am. Chem. Soc.* 126 (9) (2004) 2847–2861.
- [12] V. Hosseinpour, M. Kazemeini, A. Mohammadrezaee, Optimisation of Ru-promoted Ir-catalysed methanol carbonylation utilising response surface methodology, *Appl. Catal. Gen.* 394 (1–2) (2011) 166–175.
- [13] L.A. Key, M.J. Payne, A.D. Poole, Process for the production of acetic acid, Pat WO2004026805A1 (2004).
- [14] M.J.R.C.W. Baker, M.F. Giles, C.S. Garland, M.J. Muskett, Process for the production of acetic acid by carbonylation, Pat.EP0752406A1 (1997).
- [15] M. Kazemeini, V. Hosseinpour, Minimizing CO_2 formation in Ir-catalyzed methanol carbonylation process, *Procedia Eng.* 42 (2012) 1179–1188.
- [16] G. Lyall-Brookes, A.C. Padgham, A.G. Slater, Flow chemistry as a tool for high throughput experimentation, *Digit. Discov.* 4 (9) (2025) 2364–2400.
- [17] G. Tom, S.P. Schmid, S.G. Baird, Y. Cao, K. Darvish, H. Hao, S. Lo, S. Pablo-García, E.M. Rajaonson, M. Skreta, N. Yoshikawa, S. Corapi, G.D. Akkoc, F. Strieth-Kalthoff, M. Seifrid, A. Aspuru-Guzik, Self-driving laboratories for chemistry and materials science, *Chem. Rev.* 124 (16) (2024) 9633–9732.

- [18] R. Nsouli, G. Galiyan, L.K.G. Ackerman-Biegasiewicz, Advancing organic chemistry using high-throughput experimentation, *Angew. Chem. Int. Ed.* 64 (40) (2025) e202506588.
- [19] Z.J. Lin, T.Y. Wang, Q. Li, Designing active and stable Ir-based catalysts for the acidic oxygen evolution reaction, *Ind. Chem. Mater.* 1 (3) (2023) 299–311.
- [20] 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, Machine-learning-assisted materials discovery using failed experiments, *Nature* 533 (7601) (2016) 73–76.
- [21] Y.K. Lee, Density functional theory (DFT) calculations and catalysis, *Catalysts* 11 (4) (2021) 454.
- [22] B.M. Abraham, M.V. Jyothirmai, P. Sinha, F. Viñes, J.K. Singh, F. Illas, Catalysis in the digital age: unlocking the power of data with machine learning, *Wires Comput. Mol. Sci.* 14 (5) (2024) e1730.
- [23] B.Z. Dou, Z.L. Zhu, E. Merkurjev, L. Ke, L. Chen, J. Jiang, Y.Y. Zhu, J. Liu, B.G. Zhang, G.W. Wei, Machine learning methods for small data challenges in molecular science, *Chem. Rev.* 123 (13) (2023) 8736–8780.
- [24] E. Shim, A. Tewari, T. Cernak, P.M. Zimmerman, Machine learning strategies for reaction development: toward the low-data limit, *J. Chem. Inf. Model.* 63 (12) (2023) 3659–3668.
- [25] S.J. Pan, Q. Yang, A survey on transfer learning, *IEEE Trans. Knowl. Data Eng.* 22 (10) (2010) 1345–1359.
- [26] J.E. van Engelen, H.H. Hoos, A survey on semi-supervised learning, *Mach. Learn.* 109 (2) (2020) 373–440.
- [27] C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning, *J. Big Data* 6 (1) (2019) 60.
- [28] X. Jin, H. Zhu, S. Li, Z. Wang, Z. Liu, J. Tian, C. Yu, H. Qin, S.Z. Li, A Survey on Mixup Augmentations and Beyond, 2025, <https://doi.org/10.48550/arXiv.2409.05202>.
- [29] J. Smucny, G. Shi, T.A. Lesh, C.S. Carter, I. Davidson, Data augmentation with mixup: enhancing performance of a functional neuroimaging-based prognostic deep learning classifier in recent onset psychosis, *Neuroimage Clin* 36 (2022) 103214.
- [30] Z.M. Dong, Q. Hu, Y.J. Guo, M. Cordy, M. Papadakis, Z.Y. Zhang, Y. Le Traon, J.J. Zhao, MixCode: enhancing code classification by mixup-based data augmentation, in: 2023 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER). Taipa, Macao, China, IEEE, 2023, pp. 379–390.
- [31] Z.D. Wang, S.Q. Yang, H.F. Qin, Y.K. Liu, J.Q. Wang, MixCFormer: a CNN–transformer hybrid with mixup augmentation for enhanced finger vein attack detection, *Electronics* 14 (2) (2025) 362.
- [32] L.J. Zhang, Z. Deng, K. Kawaguchi, A. Ghorbani, J. Zou, How Does Mixup Help with Robustness and Generalization?, 2020, <https://doi.org/10.48550/arXiv.2010.04819> arXiv: 2010.04819.
- [33] L. Carratino, M. Cissé, R. Jenatton, J.P. Vert, On mixup regularization 23 (1) (2022) 14632–14662.
- [34] B. Shahriari, K. Swersky, Z.Y. Wang, R.P. Adams, N. de Freitas, Taking the human out of the loop: a review of Bayesian optimization, *Proc. IEEE* 104 (1) (2016) 148–175.
- [35] T. Yang, J. Liang, Y. Qin, X.K. Zhou, J. He, Y. Wang, Y.Q. Zhu, Z.X. Wei, S.L. Zhao, Machine learning-driven prediction of ethylene homopolymerization reaction performance: FI catalyst activity, polyethylene molecular weight, and molecular weight distribution, *Chem. Eng. J.* 511 (2025) 162198.
- [36] Q. Wang, Y. Qin, J.C. Xie, Y.F. Kong, Q. Sun, Z.X. Wei, S.L. Zhao, Size-controllable high-entropy alloys toward stable hydrogen production at industrial-scale current densities, *Adv. Mater.* 37 (10) (2025) e2420173.
- [37] E. Strumbelj, I. Kononenko, Explaining prediction models and individual predictions with feature contributions, *Knowl. Inf. Syst.* 41 (3) (2014) 647–665.
- [38] S.M. Lundberg, G. Erion, H. Chen, A. DeGrave, J.M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.I. Lee, From local explanations to global understanding with explainable AI for trees, *Nat. Mach. Intell.* 2 (1) (2020) 56–67.
- [39] E.F. El-Hashash, R.H. Ali Shiekh, A comparison of the Pearson, spearman rank and Kendall tau correlation coefficients using quantitative variables, *Asian J. Probab. Stat* 20 (3) (2022) 36–48.
- [40] I. Couso, O. Strauss, H. Saulnier, Kendall's rank correlation on quantized data: an interval-valued approach, *Fuzzy Sets Syst* 343 (2018) 50–64.
- [41] G.J. Sunley, D.J. Watson, High productivity methanol carbonylation catalysis using iridium, *Catal. Today* 58 (4) (2000) 293–307.