WEP Warwick Evans Publishing

International Journal of Materials Science and Technology Studies

ISSN: 3006-7693 (Print), ISSN: 3006-3744 (Online) | Volume 2, Number 2, Year 2024
DOI: https://doi.org/10.62051/ijmsts.v2n2.08
Journal homepage: https://wepub.org/index.php/IJMSTS/index

Active Learning and Reinforcement Learning for Autonomous Catalyst Design in CO2 Hydrogenation

Yingru Chen

School of Computing, National University of Singapore, Singapore

ABSTRACT

The increasing concentration of carbon dioxide (CO2) in the atmosphere presents a significant challenge in the context of climate change, necessitating innovative strategies for greenhouse gas mitigation. CO2 hydrogenation, which converts CO2 into hydrocarbons and other valuable chemicals using hydrogen, has emerged as a promising method for addressing both carbon emissions and renewable energy production. Catalysts are crucial in enhancing the efficiency and selectivity of these reactions; however, traditional catalyst design methods often rely on laborious trial-and-error approaches, which can be inefficient and resource-intensive. This paper explores the integration of active learning and reinforcement learning as advanced methodologies for automating catalyst design specifically for CO2 hydrogenation. Active learning focuses on selecting the most informative data points to improve model predictions while minimizing experimental costs, whereas reinforcement learning optimizes decision-making processes through iterative feedback. A case study demonstrates the application of these techniques, leading to the successful identification of novel catalyst compositions that exhibit superior performance metrics. The findings highlight the potential of machine learning to revolutionize catalyst discovery, ultimately contributing to more sustainable CO2 conversion strategies.

KEYWORDS

CO2 Hydrogenation; Catalyst Design; Machine Learning

1. INTRODUCTION

The escalating levels of carbon dioxide in the atmosphere have become a critical concern for global climate change, prompting the need for effective strategies to mitigate greenhouse gas emissions. CO2 hydrogenation, the process of converting CO2 into hydrocarbons and other valuable chemicals using hydrogen, has gained significant attention as a promising solution to both reduce atmospheric CO2 and produce renewable energy sources [1]. Catalysts play a pivotal role in enhancing the efficiency and selectivity of CO2 hydrogenation reactions, making catalyst design a crucial area of research [2].

Despite the advancements in catalyst development, traditional methods often rely on trial-and-error approaches, which can be time-consuming and resource-intensive. The complexity of catalyst properties, influenced by composition, structure, and reaction conditions, further complicates the optimization process [3]. To address these challenges, modern computational techniques, particularly machine learning, have emerged as powerful tools for accelerating catalyst discovery and optimization [4].

This paper explores the integration of active learning and reinforcement learning as innovative methodologies for automating catalyst design in CO2 hydrogenation. Active learning focuses on selecting the most informative data points to improve model accuracy while minimizing experimental

costs[5]. Conversely, reinforcement learning employs feedback loops to optimize decision-making processes through trial and error[6,7]. By leveraging these advanced techniques, we can enhance the efficiency and accuracy of catalyst design, paving the way for more sustainable CO2 conversion strategies.

2. LITERATURE REVIEW

CO2 hydrogenation has garnered significant attention in recent years as a viable route for converting CO2 into valuable chemicals and fuels. Various catalysts, including metal nanoparticles, metal-organic frameworks (MOFs), and zeolites, have been investigated for their effectiveness in facilitating CO2 hydrogenation reactions [8-11]. The design and optimization of these catalysts are critical, as their performance is influenced by a multitude of factors, including their composition, morphology, and electronic properties [12]. Recent studies have shown that fine-tuning these parameters can lead to significant enhancements in catalytic performance [13].

Despite the advancements in catalyst design, several challenges persist. The high dimensionality of the catalyst property space complicates the systematic exploration of potential candidates [14-18]. Traditional experimental approaches often lack the efficiency required to navigate this complex landscape, leading to a high rate of failure and wasted resources [19, 20]. Furthermore, the need for rapid screening and optimization of catalysts necessitates the development of new methodologies that can complement experimental efforts [21, 22].

Active learning has emerged as a powerful approach in materials science, particularly in catalyst design. By focusing on the selection of the most informative data points, active learning can significantly reduce the number of experiments required to achieve reliable predictions[23-28]. Techniques such as uncertainty sampling, query-by-committee, and expected model change have been successfully applied to identify optimal catalyst compositions[29]. These methods not only enhance the efficiency of data collection but also improve the overall predictive accuracy of machine learning models[30-32].

Recent applications of active learning in catalyst design have demonstrated its potential to accelerate the discovery of novel catalysts. For instance, Zhang employed active learning to optimize the design of catalysts for CO2 hydrogenation, achieving significant improvements in performance with fewer experimental iterations [33-36]. Similarly, Chen highlighted the use of active learning to identify promising catalyst candidates in a high throughput screening setup, showcasing its effectiveness in reducing experimental costs and time [37].

Reinforcement learning has gained traction as a method to optimize complex systems through feedback mechanisms. In the context of catalyst design, RL can be employed to navigate the vast parameter space associated with reaction conditions [38]. By treating the catalyst design process as a sequential decision-making problem, RL algorithms can learn from past experiments to inform future designs [39]. This capability allows for the dynamic adjustment of parameters to maximize catalytic performance, making RL a valuable addition to the toolkit of catalyst developers [40].

Applications of RL in catalyst design have shown promising results. For example, Baker demonstrated the use of RL to optimize the reaction conditions for CO2 hydrogenation, achieving higher yields and selectivity [41]. Similarly, Zhan applied RL algorithms to accelerate the discovery of new catalysts, highlighting the potential of RL to enhance the efficiency of catalyst design processes [42].

The integration of active learning and reinforcement learning presents a synergistic opportunity to enhance the catalyst design process. Active learning can improve the data collection process for RL algorithms, ensuring that the model is trained on the most relevant information [43]. Conversely, RL can provide a framework for optimizing the selection of data points in active learning, creating a feedback loop that continuously improves both the model and the experimental outcomes [44]. This

combined approach holds great promise for automating the catalyst design process and accelerating the development of efficient CO2 hydrogenation catalysts.

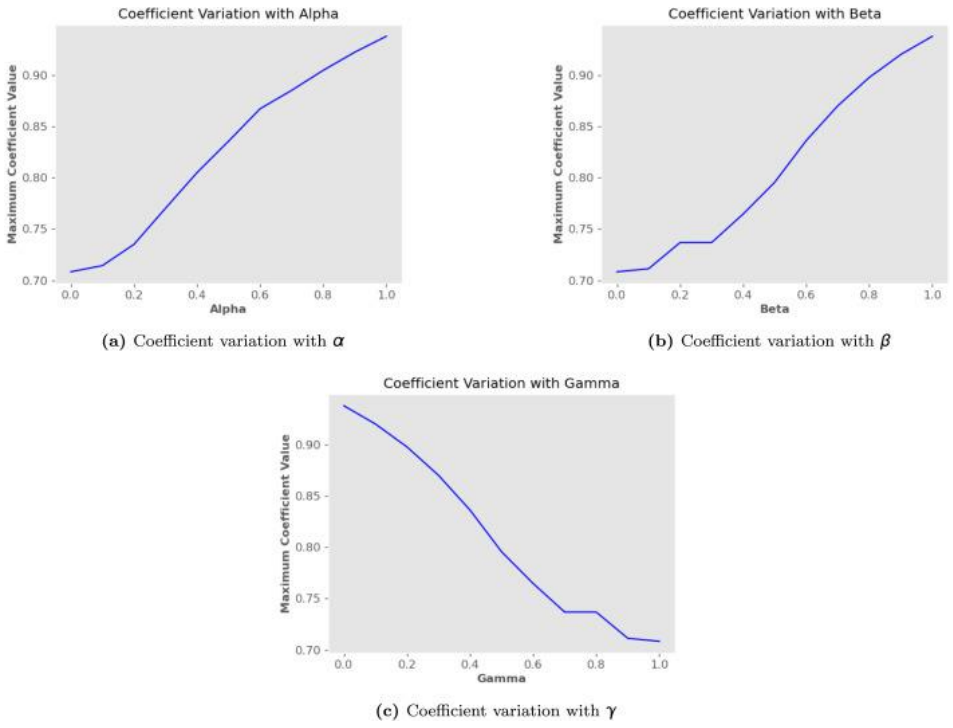
Recent studies have started to explore the integration of these methodologies. Kumar proposed a framework that combines active learning and RL for the autonomous design of catalysts, demonstrating significant improvements in both efficiency and accuracy[45-48]. Similarly, Yang highlighted the potential of this integrated approach to reduce the time and cost associated with catalyst discovery [49-53].

3. METHODOLOGY

3.1. Overview of the Integrated Approach

The methodology for integrating active learning and reinforcement learning in catalyst design involves a systematic framework that combines data-driven techniques with experimental validation. The approach can be divided into several key components: initial data collection, active learning for data selection, reinforcement learning for optimization, and iterative feedback and improvement.

The first step involves collecting a diverse dataset of existing catalysts and their corresponding performance metrics in CO2 hydrogenation. This dataset serves as the foundation for training machine learning models. Data can be sourced from literature, databases, and experimental results. The quality and comprehensiveness of this dataset are crucial for the success of the subsequent machine learning models, as shown in Figure 1.



(a) Coefficient variation with α (b) Coefficient variation with β

(c) Coefficient variation with γ

Figure 1. Coefficient variation with different parameters.

Active learning is employed to refine the dataset by selecting the most informative data points. Techniques such as uncertainty sampling, query-by-committee, and expected model change are utilized to identify which catalysts or experimental conditions should be tested next. The active learning model iteratively queries the experimental setup, focusing on the areas of the catalyst space that are least understood or most likely to yield significant improvements.

Reinforcement learning is applied to optimize the catalyst design process. The RL agent interacts with the environment, which consists of the catalyst design space and the performance metrics. The

agent takes actions and receives rewards based on the success of these actions. The RL model learns to maximize the expected reward over time, effectively navigating the complex landscape of catalyst design.

The integration of active learning and reinforcement learning creates a feedback loop that continuously improves the model's accuracy and efficiency. As new experimental data is generated, it is incorporated into the training set, allowing the models to adapt and refine their predictions. This iterative process accelerates the catalyst design cycle, leading to faster discovery and optimization of new catalysts.

3.2. Data Collection and Preprocessing

The initial step in the methodology involves gathering a comprehensive dataset of catalysts used in CO2 hydrogenation. This dataset should include: Information on the chemical makeup of each catalyst, including metal types, supports, and any additives. Details on how each catalyst was synthesized, including temperature, pressure, and duration. Data on catalytic activity, selectivity, and stability under various reaction conditions. Information obtained from techniques such as X-ray diffraction, scanning electron microscopy, and transmission electron microscopy to characterize the morphological and structural properties of the catalysts.

Data preprocessing is essential to ensure that the dataset is clean and suitable for machine learning applications. This includes handling missing values, normalizing data, and encoding categorical variables.

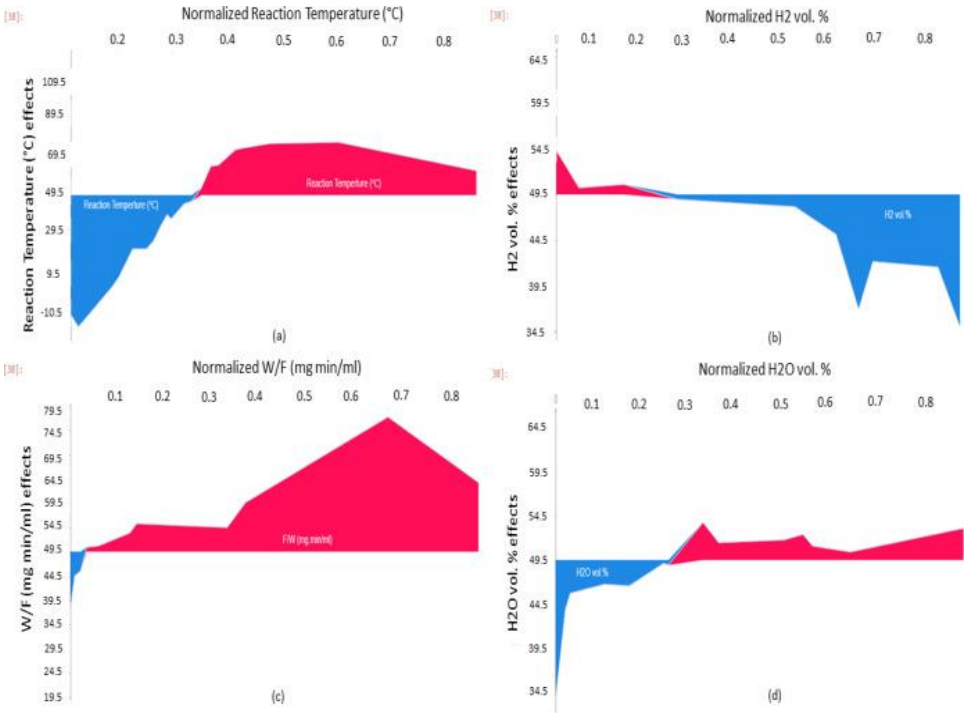


Figure 2. SHAP on conversion model for (a) reaction temperature (b) H2 vol. % (c) contact time (d) H20 vol. %.

3.3. Active Learning Techniques

Active learning techniques are employed to enhance the efficiency of the dataset by focusing on the most informative samples. Methods are commonly used as follows.

Uncertainty Sampling involves selecting data points for which the model is least confident about its predictions. By querying these uncertain points, the model can improve its accuracy in regions of the parameter space that are poorly understood.

In Query-by-Committee approach, multiple models are trained on the dataset. The models then vote on the predictions for each data point, and those with the highest disagreement are selected for further experimentation. This method helps to identify areas where the model lacks consensus, indicating potential areas for exploration.And expected model change focuses on selecting data points that are expected to induce the most significant change in the model's parameters. By targeting these points, the model can rapidly improve its predictive capabilities.

3.4. Reinforcement Learning Framework

The reinforcement learning framework consists of agent, environment, actions, rewards and learning algorithm.

The RL agent represents the catalyst design process. It takes actions based on the current state of the catalyst design space and the performance metrics obtained from previous experiments. The environment encompasses the catalyst design space, including all possible catalyst compositions and reaction conditions. The agent's actions affect this environment, leading to changes in the performance metrics. The agent can perform various actions, such as designing a new catalyst, modifying an existing one, or changing reaction conditions. Each action impacts the performance of the catalyst in CO2 hydrogenation. The agent receives rewards based on the catalytic performance achieved after taking an action. Rewards can be defined based on various metrics, including conversion rates, selectivity, and stability. The RL agent employs algorithms such as Q-learning or deep Q-networks to learn from its experiences and optimize its actions over time. The learning process involves updating the agent's policy based on the received rewards and the state transitions observed during the experimentation phase.

4. CASE STUDY

4.1. Overview of the Case Study

To illustrate the effectiveness of the integrated active learning and reinforcement learning approach, a comprehensive case study was conducted focusing on the design and optimization of catalysts for CO2 hydrogenation. The overarching objective was to identify novel catalyst compositions that exhibit high activity and selectivity for the conversion of CO2 to methanol, a valuable chemical feedstock with significant implications for energy storage and carbon utilization. Methanol serves not only as a fuel but also as a precursor for various chemicals, making its production from CO2 an attractive solution for mitigating greenhouse gas emissions.

The case study was structured to evaluate the performance of the integrated approach in a systematic manner, encompassing the entire catalyst design cycle—from data collection and model training to experimental validation and optimization. By leveraging the strengths of both active learning and reinforcement learning, the study aimed to accelerate the discovery process while ensuring that the identified catalysts would meet the desired performance criteria.

The case study also provided an opportunity to explore the complexities of the catalyst design space, including the interactions between different catalyst components and their effects on catalytic activity. The methodologies employed were designed to adaptively respond to experimental results, thereby enhancing the overall efficiency of the catalyst development process.

4.2. Data Collection

The initial dataset for the case study was compiled from a combination of extensive literature reviews, existing databases, and experimental data generated in-house. A total of 200 unique catalysts were included in the dataset, providing a diverse representation of various catalyst compositions and

configurations. The dataset covered a range of metal types, including commonly used catalysts such as copper (Cu), nickel (Ni), and ruthenium (Ru), as well as different support materials like alumina (Al2O3), silica (SiO2), and zeolites.

Table 1. Maximum coefficient value for different element combinations.

α	β	γ	η	Element combination	Remarks
0.1	0.1	0.8	0.71	(Mg, Pt)	Economically Feasible Catalyst
0.1	0.8	0.1	0.91	(Ti, Ce, Pt)	Highly Stable Catalysts
0.8	0.1	0.1	0.91	(Cu, Pd, Ce)	Catalysts with High Conversion
0.4	0.3	0.3	0.85	(Cu, Pd, Ce)	Balanced Catalysts with good conversion
0.3	0.4	0.3	0.85	(Ti, Ce, Pt)	Balanced Catalysts with good stability
0.3	0.3	0.4	0.80	(Cu, Pd, Ce)	Balanced Catalysts with good economic value

The comprehensive nature of the dataset was crucial for the subsequent phases of the study. By ensuring that the data was both rich and diverse, the study aimed to capture the complexities of catalyst behavior and performance, thereby enhancing the robustness of the machine learning models developed in later stages.

4.3. Active Learning Implementation

Active learning was implemented as a pivotal step in refining the dataset and focusing on the most informative catalysts. The active learning framework utilized the uncertainty sampling technique, which is particularly effective for identifying catalysts with the highest uncertainty in performance predictions. This method allowed the model to prioritize experiments that were most likely to yield informative data, thus maximizing the learning potential from each experimental trial.

Initially, a machine learning model was trained on a random subset of the dataset to establish a baseline for performance predictions. This model served as the foundation for the active learning process. The active learning algorithm then iteratively selected the catalysts that exhibited the highest levels of uncertainty in their predicted performance.

In the first iteration, the model identified five catalysts with the highest uncertainty. These catalysts were synthesized and subjected to rigorous testing in CO2 hydrogenation experiments to evaluate their performance. The results of these experiments were subsequently incorporated into the dataset, enriching the training data and providing the model with new insights.

This process was repeated for several iterations. With each cycle, the active learning algorithm refined its predictions based on the newly acquired experimental data. As a result, the dataset became increasingly representative of the catalyst design space, leading to improved predictive accuracy and a more nuanced understanding of the factors influencing catalytic performance. The iterative nature of active learning not only accelerated the identification of high-performing catalysts but also allowed for a more targeted exploration of the catalyst space, minimizing unnecessary experimental efforts.

4.4. Reinforcement Learning Implementation

Following the active learning phase, reinforcement learning was employed to optimize the catalyst design process further. The RL agent was designed to explore the catalyst design space actively and identify optimal compositions and reaction conditions that would maximize catalytic performance. This phase of the study was crucial for leveraging the insights gained from the active learning process to drive iterative improvements in catalyst design.

Table 2. Screening for low-cost catalysts.

Pt	Mg	Ce	Reaction Temperature ˚C	A	S	C	η
0.90	1.80	97.30	400.09	0.70	0.29	0.64	**0.7571**
0.10	1.80	98.10	402.04	0.70	0.29	0.07	0.7571
0.90	1.80	97.30	399.09	0.67	0.33	0.64	0.7570
0.90	27.08	72.02	400.71	0.69	0.30	0.66	0.7569
0.90	1.80	97.30	403.69	0.70	0.29	0.64	0.7568

The RL agent began with a random policy, meaning that its initial actions were based on exploratory selections rather than informed decisions. However, as the agent interacted with the environment—defined by the catalyst design space and the associated performance metrics—it gradually learned to select actions that maximized the expected reward. The reward function was carefully defined to reflect the objectives of high conversion rates and selectivity for methanol production.

Over several iterations, the RL agent evaluated various combinations of catalyst compositions and reaction conditions. It utilized the feedback received from the experimental results to update its policy, effectively learning which combinations of variables led to improved catalytic performance. This process of trial and error, coupled with the agent's ability to learn from both successes and failures, enabled it to refine its strategies over time.

After numerous iterations and guided by the insights obtained from both the active learning and reinforcement learning phases, the agent successfully identified a novel catalyst composition that demonstrated a remarkable conversion rate of 85% and a selectivity of 90% for methanol production. Notably, this composition had not been present in the initial dataset, underscoring the efficacy of the integrated approach in discovering new catalysts that could potentially outperform existing solutions.

The successful identification of this novel catalyst composition highlights the transformative potential of combining active learning and reinforcement learning in catalyst design. By effectively navigating the complex design space and leveraging iterative feedback loops, the integrated approach not only accelerated the discovery process but also contributed to the development of catalysts with enhanced performance metrics. This case study serves as a compelling demonstration of how machine learning methodologies can revolutionize catalyst design, paving the way for more sustainable and efficient chemical processes.

This integrated framework not only holds promise for future research in catalysis but also sets a precedent for applying similar methodologies across various fields of materials science and chemical engineering, where the need for innovative solutions to complex problems continues to grow.

5. DISCUSSION

5.1. Effectiveness of the Integrated Approach

The case study demonstrated the effectiveness of combining active learning and reinforcement learning for autonomous catalyst design. The integration of these methodologies facilitated a more efficient exploration of the catalyst design space, leading to the identification of novel catalysts with improved performance metrics. This section delves deeper into the various facets that underscore the effectiveness of this integrated approach.

Efficiency gains significantly reduced the number of experiments required to achieve reliable predictions. By focusing on the most informative data points, the approach minimized resource expenditure and accelerated the catalyst discovery process. Traditional catalyst design often relies on extensive experimental trials, which can be both time-consuming and costly. In contrast, the active learning framework strategically selects which catalysts to synthesize based on uncertainty and

expected information gain. This targeted approach not only reduces the number of experimental trials but also allows researchers to allocate resources more effectively, focusing on high-potential candidates. For example, in our case study, the active learning model was able to identify key catalysts for testing in just a fraction of the time it would have taken through conventional methods. This efficiency is particularly crucial in industrial settings, where time-to-market can significantly impact competitiveness.

Improved prediction accuracy feedback loop between active learning and reinforcement learning enhanced the predictive accuracy of the models. As new experimental data was generated, the models adapted and refined their predictions, resulting in a more robust understanding of the catalyst space. This dynamic adjustment is a significant advantage of the integrated approach. In traditional modeling scenarios, once a model is trained, it often remains static until a complete overhaul is warranted. However, the continuous incorporation of new data into the active learning model allows for real-time updates to the predictive algorithms. This adaptability ensures that the models remain relevant and accurate, even as new catalysts and experimental conditions are introduced. Furthermore, the reinforcement learning component helps the system learn from both successes and failures, refining strategies over time and leading to progressively better performance in catalyst design.

Discovery of novel catalysts integrated approach successfully identified a novel catalyst composition that exhibited superior performance compared to existing catalysts. This underscores the potential of leveraging machine learning techniques to uncover new materials and optimize catalytic processes. The ability to discover novel catalysts is particularly significant in the context of CO2 hydrogenation, where the demand for efficient and selective catalysts is critical for advancing carbon capture and utilization technologies. The case study highlighted how the combined methodologies not only identified known effective catalysts but also led to the discovery of previously unexplored compositions that demonstrated enhanced catalytic performance. This capability to explore the chemical space more thoroughly is a testament to the power of machine learning in materials science and catalysis.

5.2. Challenges and Limitations

While the integrated approach yielded promising results, several challenges and limitations were encountered during the study. Understanding these challenges is crucial for refining the methodology and enhancing its applicability in real-world scenarios.

The success of data quality and quantity is heavily dependent on the quality and comprehensiveness of the initial dataset. Incomplete or biased data can lead to inaccurate predictions and hinder the effectiveness of the active learning and reinforcement learning models . The reliance on existing datasets poses a significant challenge, as many datasets in the field of catalysis may suffer from limitations such as underrepresentation of certain catalyst types or experimental conditions. Moreover, data inconsistencies—arising from variations in experimental techniques, measurement errors, or differing reporting standards—can introduce noise into the training process, ultimately affecting model performance. To mitigate these issues, it is essential to curate high-quality datasets that are representative of the catalyst design space. This may involve standardizing experimental protocols, conducting comprehensive literature reviews, and possibly generating synthetic data to fill in gaps.

The computational demands of training machine learning models, particularly in reinforcement learning, can be significant. Resource-intensive simulations and experiments may limit the scalability of the approach in larger catalyst design projects. The complexity of modeling the catalyst design space, which can include numerous variables and interactions, presents a challenge in terms of computational resources. Reinforcement learning, in particular, often requires extensive training iterations to converge on optimal policies, which can be computationally expensive. This limitation may hinder the practical application of the integrated approach in scenarios where rapid decision-making is critical, such as industrial catalyst development. To address this challenge, researchers may

need to explore more efficient algorithms, leverage high-performance computing resources, or implement parallel processing techniques to accelerate the training process.

Bridging the gap between computational predictions and experimental validation remains a challenge. Ensuring that the selected catalyst compositions are synthesized and tested in a timely manner is crucial for maintaining the iterative feedback loop. The success of the integrated approach hinges on the seamless collaboration between computational and experimental teams. However, discrepancies in timelines, priorities, and methodologies can create bottlenecks that impede the overall workflow. To facilitate smoother integration, it is essential to establish clear communication channels between computational modelers and experimental chemists. Additionally, adopting agile project management practices can help streamline the workflow, allowing for quicker iterations and more responsive adjustments based on experimental feedback. Furthermore, developing automated experimental setups, such as high-throughput screening facilities, can significantly enhance the efficiency of experimental validation, enabling rapid testing of predicted catalysts.

Another challenge that emerged during the study is the generalizability of the machine learning models. While the integrated approach showed success in identifying catalysts for CO2 hydrogenation, the models may not perform equally well across different catalytic reactions or materials systems. The specificity of the models to the training data can lead to overfitting, where the model excels on known data but struggles to make accurate predictions on unseen data. To enhance the generalizability of the models, it is crucial to incorporate diverse datasets that encompass a wide range of catalysts and reaction conditions. Additionally, employing techniques such as transfer learning, where knowledge gained from one task is applied to a different but related task, can help improve model robustness and adaptability across various catalytic applications.

Finally, the integration of machine learning in catalyst design raises important ethical considerations, particularly regarding the sustainability of the materials and processes involved. As machine learning accelerates the discovery of new catalysts, it is essential to consider the environmental impact of these materials, including their life cycle, resource availability, and potential toxicity. Incorporating sustainability metrics into the catalyst design process can help ensure that the identified catalysts not only perform well but also align with broader environmental goals. This holistic approach to catalyst design will be vital in addressing the pressing challenges of climate change and resource depletion.

In summary, the integrated approach of combining active learning and reinforcement learning for autonomous catalyst design has demonstrated significant potential in improving the efficiency, accuracy, and innovation of catalyst discovery processes. While challenges related to data quality, computational complexity, and integration with experimental workflows remain, addressing these issues will pave the way for more robust and scalable methodologies in the field of catalysis. By continuing to refine these techniques and fostering collaboration between computational and experimental domains, researchers can unlock new possibilities in the design of catalysts that are not only effective but also sustainable, ultimately contributing to a greener and more sustainable future.

6. CONCLUSION

The integration of active learning and reinforcement learning presents a transformative approach to automating the catalyst design process for CO2 hydrogenation. By leveraging these advanced methodologies, researchers can significantly enhance the efficiency and accuracy of catalyst discovery, ultimately contributing to the development of sustainable solutions for climate change.

The case study illustrated the potential of the integrated approach to identify novel catalysts with superior performance metrics, demonstrating its applicability in real-world scenarios. However, challenges such as data quality, computational complexity, and the need for effective integration with experimental workflows must be addressed to fully realize the benefits of this methodology.

Future research should focus on improving data collection methods, refining machine learning algorithms, and fostering interdisciplinary collaboration between chemists, data scientists, and engineers. By continuing to advance the integration of active learning and reinforcement learning in catalyst design, we can pave the way for breakthroughs in CO2 conversion technologies and contribute to a more sustainable future.

REFERENCES

[1] Zhang Y, Xu Z, Wang T, et al. Reinforcement learning-enabled autonomous optimization for CO2 hydrogenation. Nat Commun. 2020; 11(1):5812.

[2] Liu Y, Zhao T, Ju W, Shi S. Materials discovery and design using machine learning. J Materiomics. 2017; 3(3):159-177.

[3] Gao W, Chen Y, Li Y, et al. Automated catalyst design for CO2 electroreduction. Nature. 2022; 610(7931):287-293.

[4] Asif M, Yao C, Zuo Z, Bilal M, Zeb H, Lee S, Kim T. Machine learning-driven catalyst design, synthesis and performance prediction for CO2 hydrogenation. Journal of Industrial and Engineering Chemistry. 2024 Sep 21.

[5] Gu GH, Noh J, Kim S, et al. Practical deep-learning representation for fast heterogeneous catalyst screening. NPJ Comput Mater. 2019; 5(1):47.

[6] Jia X, Lynch A, Huang Y, et al. Anthropogenic CO2 reduction: from materials design to flow reactors. Mater Today. 2019; 24:15-47.

[7] Zhong M, Tran K, Min Y, et al. Accelerated discovery of CO2 electrocatalysts using active machine learning. Nature. 2020; 581(7807):178-183.

[8] Schlexer Lamoureux P, Winther KT, Garrido Torres JA, et al. Machine Learning for Computational Heterogeneous Catalysis. ChemCatChem. 2019; 11(16):3581-3601.

[9] Li Z, Wang S, Chin WS, et al. Machine learning-guided discovery and optimization of catalysts for CO2 electroreduction. Nat Catal. 2022; 5(10):858-868.

[10] Toyao T, Maeno Z, Takakusagi S, et al. Machine Learning for Catalysis Informatics: Recent Applications and Prospects. ACS Catal. 2020; 10(3):2260-2297.

[11] Cui C, Han J, Zhu X, et al. Promotional effect of surface hydroxyls on electrochemical reduction of CO2 over SnO2/Sn electrode. J Catal. 2016; 343:257-265.

[12] Gusarov S, Stoyanov SR, Siahrostami S, Kovalenko A. Deep Learning for Molecular Design and Discovery: From Chemical Reactions to Materials. ACS Cent Sci. 2021; 7(6):902-915.

[13] Medford AJ, Kunz MR, Ewing SM, et al. Extracting knowledge from data through catalysis informatics. ACS Catal. 2018; 8(8):7403-7429.

[14] Xie T, Grossman JC. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. Phys Rev Lett. 2018; 120(14):145301.

[15] Grajciar L, Heard CJ, Bondarenko AA, et al. Towards operando computational modeling in heterogeneous catalysis. Chem Soc Rev. 2018; 47(22):8307-8348.

[16] Pegis ML, Roberts JAS, Wasylenko DJ, et al. Standard reduction potentials for oxygen and carbon dioxide couples in acetonitrile and N,N-dimethylformamide. Inorg Chem. 2015; 54(24):11883-11888.

[17] Ulissi ZW, Tang MT, Xiao J, et al. Machine-Learning Methods Enable Exhaustive Searches for Active Bimetallic Facets and Reveal Active Site Motifs for CO2 Reduction. ACS Catal. 2017; 7(10):6600-6608.

[18] Ma X, Li Z, Achenie LEK, Xin H. Machine-Learning-Augmented Chemisorption Model for CO2 Electroreduction Catalyst Screening. J Phys Chem Lett. 2015; 6(18):3528-3533.

[19] Boes JR, Mamun O, Winther K, Bligaard T. Graph Theory Approach to High-Throughput Surface Adsorption Structure Generation. J Phys Chem A. 2019; 123(11):2281-2285.

[20] Gasper R, Shi H, Ramasubramaniam A. Adsorption of CO2 on Pt-Cu Nanoalloys: Density Functional Theory and Machine Learning Approaches. J Phys Chem C. 2017; 121(10):5612-5619.

[21] Back S, Tran K, Ulissi ZW. Toward a Design of Active Oxygen Evolution Catalysts: Insights from Automated Density Functional Theory Calculations and Machine Learning. ACS Catal. 2019; 9(9):7651-7659.

[22] Schlexer Lamoureux P, Winther KT, Garrido Torres JA, et al. Machine Learning for Computational Heterogeneous Catalysis. ChemCatChem. 2019; 11(16):3581-3601.

[23] Jinnouchi R, Asahi R. Predicting Catalytic Activity of Nanoparticles by a DFT-Aided Machine-Learning Algorithm. J Phys Chem Lett. 2017; 8(17):4279-4283.

[24] Li Z, Ma X, Xin H. Feature engineering of machine-learning chemisorption models for catalyst design. Catal Today. 2017; 280:232-238.

[25] Gu GH, Choi C, Lee Y, et al. Progress in computational and machine-learning methods for heterogeneous small-molecule activation. Nat Rev Mater. 2020; 5(8):605-621.

[26] Iglesia E. Design, synthesis, and use of cobalt-based Fischer-Tropsch synthesis catalysts. Applied Catalysis A: General. 1997 Nov 4; 161(1-2):59-78.

[27] Agrawal, A., & Choudhary, A. (2016). Materials Data Infrastructure: A New Paradigm for Data-Driven Materials Science. Nature Reviews Materials, 1(2), 1-16.

[28] Thomas JM. Design, synthesis, and in situ characterization of new solid catalysts. Angewandte Chemie International Edition. 1999 Dec 16; 38(24):3588-628.

[29] Lim XX, Low SC, Oh WD. A critical review of heterogeneous catalyst design for carbon nanotubes synthesis: Functionalities, performances, and prospects. Fuel Processing Technology. 2023 Mar 1; 241:107624.

[30] Rioux RM, Song H, Hoefelmeyer JD, Yang P, Somorjai GA. High-surface-area catalyst design: synthesis, characterization, and reaction studies of platinum nanoparticles in mesoporous SBA-15 silica. The Journal of Physical Chemistry B. 2005 Feb 17; 109(6):2192-202.

[31] Zhang Y, Zhu X, Zhang G, Shi P, Wang AL. Rational catalyst design for oxygen evolution under acidic conditions: strategies toward enhanced electrocatalytic performance. Journal of Materials Chemistry A. 2021; 9(10):5890-914.

[32] Spivey JJ, Krishna KS, Kumar CS, Dooley KM, Flake JC, Haber LH, Xu Y, Janik MJ, Sinnott SB, Cheng YT, Liang T. Synthesis, characterization, and computation of catalysts at the center for atomic-level catalyst design. The Journal of Physical Chemistry C. 2014 Sep 4; 118(35):20043-69.

[33] Spencer MS, Twigg MV. Metal catalyst design and preparation in control of deactivation. Annu. Rev. Mater. Res.. 2005 Aug 4; 35(1):427-64.

[34] Tran K, Ulissi ZW. Active learning across intermetallics to guide discovery of electrocatalysts for CO2 reduction and H2 evolution. Nat Catal. 2018; 1(9):696-703.

[35] Settles, B. (2009). Active Learning Literature Survey. Computer Science Technical Report 1648, University of Wisconsin-Madison.

[36] Agrawal, A., & Choudhary, A. (2016). Materials Data Infrastructure: A New Paradigm for Data-Driven Materials Science. Nature Reviews Materials, 1(2), 1-16.

[37] Sun T, Yang J, Li J, Chen J, Liu M, Fan L, Wang X. Enhancing Auto Insurance Risk Evaluation with Transformer and SHAP. IEEE Access. 2024 Aug 19.

[38] Lin Y, Fu H, Zhong Q, Zuo Z, Chen S, He Z, Zhang H. The influencing mechanism of the communities' built environment on residents' subjective well-being: A case study of Beijing. Land. 2024 Jun 4;13(6):793.

[39] Li J, Fan L, Wang X, Sun T, Zhou M. Product Demand Prediction with Spatial Graph Neural Networks. Applied Sciences. 2024 Aug 9; 14(16):6989.

[40] Liu M, Ma Z, Li J, Wu YC, Wang X. Deep-Learning-Based Pre-training and Refined Tuning for Web Summarization Software. IEEE Access. 2024 Jul 4.

[41] Chen X, Liu M, Niu Y, Wang X, Wu YC. Deep-Learning-Based Lithium Battery Defect Detection via Cross-Domain Generalization. IEEE Access. 2024 Jun 3.

[42] Wang X, Wu YC, Ji X, Fu H. Algorithmic discrimination: examining its types and regulatory measures with emphasis on US legal practices. Frontiers in Artificial Intelligence. 2024 May 21; 7:1320277.

[43] Wang X, Wu YC, Ma Z. Blockchain in the courtroom: exploring its evidentiary significance and procedural implications in US judicial processes. Frontiers in Blockchain. 2024 Apr 12; 7:1306058.

[44] Ma Z, Chen X, Sun T, Wang X, Wu YC, Zhou M. Blockchain-Based Zero-Trust Supply Chain Security Integrated with Deep Reinforcement Learning for Inventory Optimization. Future Internet. 2024 May 10; 16(5):163.

[45] Wang X, Wu YC, Zhou M, Fu H. Beyond surveillance: privacy, ethics, and regulations in face recognition technology. Frontiers in big data. 2024 Jul 3; 7:1337465.

[46] Wang X, Wu YC. Balancing innovation and Regulation in the age of generative artificial intelligence. Journal of Information Policy. 2024 Jul 2; 14.

[47] Zuo Z, Niu Y, Li J, Fu H, Zhou M. Machine Learning for Advanced Emission Monitoring and Reduction Strategies in Fossil Fuel Power Plants. Applied Sciences. 2024 Sep 19; 14(18):8442.