Academic Journal of Earth Sciences
ISSN: 2960-0154
DOI: 10.61784/ajes3002

MACHINE LEARNING-DRIVEN HIGH-THROUGHPUT SCREENING OF CATALYST CANDIDATES FOR CO2 HYDROGENATION

Emma Bucher, Christopher Rau, Quan Phuong*
School of Environmental and Life Sciences, The University of Newcastle, Australia.
Corresponding Author: Quan Phuong, Email: quanphuong29@newcastle.edu.au

Abstract: The escalating levels of carbon dioxide (CO2) in the atmosphere, primarily due to human activities, pose significant challenges to global climate stability. CO2 hydrogenation emerges as a promising technology to mitigate these emissions by converting CO2 into valuable hydrocarbons and fuels, thus providing a sustainable energy pathway. Catalysts are essential for enhancing the efficiency and selectivity of this process; however, the discovery of effective catalysts is complicated by the vast array of potential materials and the intricate interactions that influence catalytic performance. Traditional methods of catalyst screening are often laborious and resource-intensive, necessitating innovative approaches to expedite the discovery process. This paper explores the integration of machine learning (ML) techniques into high-throughput screening (HTS) methodologies to facilitate the rapid evaluation of thousands of catalyst candidates for CO2 hydrogenation. By leveraging ML algorithms, we can analyze extensive datasets, identify performance patterns, and prioritize promising candidates for experimental validation. The findings demonstrate that ML-driven HTS not only accelerates catalyst discovery but also optimizes resource utilization, paving the way for more efficient solutions to combat CO2 emissions.
Keywords: Machine learning; High-throughput screening; CO2 hydrogenation

1 INTRODUCTION

The increasing levels of carbon dioxide in the atmosphere due to anthropogenic activities have become a significant contributor to climate change and global warming. As a result, the development of technologies that can effectively reduce CO2 emissions is of paramount importance. One promising approach to mitigate CO2 emissions is CO2 hydrogenation, a chemical process that converts CO2 into valuable hydrocarbons and fuels using hydrogen. This process not only helps in reducing atmospheric CO2 levels but also provides a sustainable pathway for producing energy-dense fuels and chemicals[1-10].

Catalysts play a critical role in CO2 hydrogenation by significantly enhancing the reaction rates and selectivity towards desired products. Various materials, including metals and metal oxides, have been investigated as potential catalysts for this reaction[11-13]. However, the discovery of effective catalysts remains a challenge due to the vast number of possible materials and the complex interactions that govern catalytic performance[14]. Traditional methods of catalyst screening are often time-consuming and resource-intensive, necessitating the development of more efficient approaches.

High-throughput screening is a systematic approach that allows researchers to rapidly evaluate and compare a large number of catalyst candidates. By using automated experimental setups, HTS enables the efficient screening of thousands of materials in a relatively short time frame[15]. Despite its advantages, traditional HTS methods often face limitations, such as high costs, limited data analysis capabilities, and challenges in interpreting results[16].

The integration of machine learning techniques into the HTS process has the potential to transform catalyst discovery. ML algorithms can analyze large datasets, identify patterns, and make predictions about catalyst performance based on historical data[17-20]. By leveraging ML, researchers can prioritize promising candidates for experimental validation, thereby accelerating the discovery process and reducing resource consumption[21,22]. This paper aims to investigate the use of machine learning models to screen and evaluate thousands of potential catalyst materials efficiently for CO2 hydrogenation.

2 LITERATURE REVIEW

The application of machine learning in materials science, particularly in catalyst discovery, has gained significant attention in recent years. The need for more efficient and effective methods to identify and optimize catalytic materials has led researchers to explore the potential of ML techniques. These methods promise to enhance the efficiency, accuracy, and speed of catalyst screening processes, ultimately contributing to the development of more sustainable energy solutions.

The integration of machine learning into catalyst discovery has transformed traditional methodologies. Xie and Grossman demonstrated the use of graph convolutional networks (GCNs) for predicting the formation energies of materials[23-25]. This innovative approach allows for the modeling of complex relationships within materials data by treating the materials as graphs, where atoms are nodes and bonds are edges. This representation enables GCNs to capture intricate structural features that are critical for predicting material properties.

Similarly, Schütt introduced a deep learning method designed to predict molecular properties, which emphasizes the versatility of neural networks in handling various types of data in materials science[26-28]. Their work illustrates how deep learning can uncover hidden patterns in large datasets, providing insights that are often missed by conventional methods.

Huang provided a comprehensive overview of high-throughput screening methods, discussing how automation and advanced data analysis techniques can significantly enhance catalyst discovery[29]. They highlighted the role of robotic systems in conducting experiments at an unprecedented scale, allowing for the rapid testing of thousands of candidates. This automation, coupled with ML algorithms, facilitates the identification of promising materials more efficiently than traditional manual approaches.

Takahashi critically examined the limitations of traditional HTS approaches, noting issues such as high costs and long timeframes associated with experimental validation[30-35]. They proposed the integration of ML to improve data interpretation and candidate selection, suggesting that ML could serve as a powerful tool to prioritize the most promising candidates for further investigation.

The search for effective catalysts for CO2 hydrogenation has been a focal point of research in recent years [36]. Guo conducted a comprehensive review of various catalytic materials, discussing the mechanisms underlying the hydrogenation process and the challenges faced in different catalyst systems[37]. They highlighted the importance of understanding catalyst stability, selectivity, and activity in the context of CO2 conversion, emphasizing that the optimal catalyst must balance these factors to achieve high performance.

Zhang investigated the performance of metal-based catalysts, providing insights into the structure-activity relationships that govern catalytic efficiency[38]. Their findings underscore the significance of metal particle size, support effects, and electronic properties in influencing the catalytic activity. This research contributes to the ongoing efforts to develop tailored catalysts that can operate efficiently under realistic reaction conditions.

Data-driven methodologies have emerged as a critical component of modern catalyst discovery. Previous works emphasized the importance of leveraging historical data to accelerate the discovery process[39,40]. By employing machine learning models trained on extensive datasets, researchers can predict the performance of new catalyst candidates, significantly reducing the time and effort required for experimental validation[41].

Agrawal and Choudhary discussed the role of materials databases in facilitating ML-driven discovery[42]. They highlighted the significance of curated datasets, which provide the necessary high-quality data for training predictive models. The availability of comprehensive databases enables researchers to harness the power of ML to uncover new materials and optimize existing ones.

Despite the promising advancements, several challenges remain in applying machine learning to materials science. Bartel addressed the issues of data quality and availability, emphasizing that the success of ML applications hinges on the quality of the data used for training models[4-45]. They called for standardized data formats to facilitate model training and ensure reproducibility across studies.

Wang explored the interpretability of machine learning models in materials science, proposing techniques to enhance the understanding of model predictions[46]. Interpretability is crucial for gaining trust in ML models, particularly in fields like catalysis, where the implications of predictions can be significant. Their work advocates for the development of methods that can elucidate the reasoning behind model outputs, making it easier for researchers to validate and apply ML predictions in practical scenarios.

Looking ahead, the integration of quantum computing with machine learning has been proposed as a means to further accelerate materials discovery[47]. Quantum computing holds the potential to solve complex problems that are currently intractable for classical computers, enabling researchers to explore vast chemical spaces and optimize materials at unprecedented speeds.

The potential for transfer learning to improve model performance across different materials systems has also been highlighted. Transfer learning allows models trained on one dataset to be adapted for another, which can significantly reduce the need for extensive training data in new materials domains. This approach could be particularly beneficial in catalyst discovery, where diverse materials are being explored.

The application of ML-driven HTS is not limited to CO2 hydrogenation but extends to other catalytic processes, such as methane reforming and biomass conversion. The versatility of machine learning techniques allows for their application across various domains, potentially leading to breakthroughs in multiple areas of catalysis and materials science.

Lastly, the importance of collaboration between academia and industry in advancing catalyst discovery has been emphasized. Collaborative efforts can facilitate the sharing of resources, expertise, and data, ultimately leading to more rapid advancements in the field. Open-access data sharing initiatives are crucial for fostering innovation and accelerating research, as highlighted. By making data readily available, researchers can build upon existing knowledge and drive the field forward more effectively.

3 METHODOLOGY

3.1 Overview of HTS Process

High-throughput screening is an automated process that allows researchers to rapidly evaluate a large number of catalyst candidates under controlled conditions. The primary goal of HTS is to identify promising materials quickly,

thereby accelerating the discovery and optimization of catalysts for various chemical reactions, including CO2 hydrogenation. The HTS process can be broken down into several key steps:

The first step involves selecting a diverse set of catalyst candidates based on theoretical predictions, previous experimental data, or computational modeling. This selection aims to cover a wide range of materials that may exhibit varying catalytic properties. Once candidates are selected, they are synthesized using methods such as sol-gel processes, co-precipitation, or physical vapor deposition. Automation plays a crucial role in this step, as robotic systems can synthesize multiple samples simultaneously, significantly reducing time and labor.

After synthesis, the physical and chemical properties of the catalysts must be characterized. Techniques such as X-ray diffraction, scanning electron microscopy, and transmission electron microscopy are commonly used to determine the structural properties of the catalysts. The catalysts are then subjected to catalytic testing under controlled reaction conditions. Parameters such as temperature, pressure, and reactant concentrations are carefully monitored to evaluate the performance of each catalyst in the CO2 hydrogenation reaction.

The results from the catalytic tests are collected and analyzed to identify the most promising candidates. This phase often involves statistical analysis and data visualization techniques to interpret the performance metrics effectively.

Based on the results obtained, researchers can refine their candidate selection and testing processes. This iterative approach allows for continuous improvement and optimization of catalysts. The integration of machine learning with HTS represents a paradigm shift in catalyst discovery. Machine learning models can analyze vast datasets generated by HTS and predict the performance of catalyst candidates more accurately than traditional methods. The workflow for screening candidates using ML can be outlined as follows:

The first step involves gathering data from previous HTS experiments, including both successful and unsuccessful catalyst candidates. This dataset serves as the foundation for training the ML models. Relevant features that influence catalytic performance must be identified and extracted from the dataset. These features can include electronic properties, structural characteristics, and thermodynamic data. Effective feature engineering is crucial for enhancing the predictive power of ML models. The collected dataset is then used to train various ML models. Common algorithms employed include regression models for predicting quantitative performance metrics and classification models for identifying suitable candidates based on categorical outcomes.

Once trained, the ML models can predict the performance of new catalyst candidates that have not been tested experimentally. This predictive capability allows researchers to prioritize candidates for further experimental validation. The predictions made by the ML models can be validated through subsequent experimental testing. The results of these tests can then be fed back into the model to refine its predictions, creating a continuous learning loop that improves the model's accuracy over time.

3.2 Case Studies

Several case studies have demonstrated the successful application of ML-driven HTS in catalyst discovery:

One notable example is where a machine learning model was developed to predict the activity of various metal catalysts for CO2 hydrogenation. The model utilized a dataset of previously tested catalysts and achieved a significant improvement in prediction accuracy compared to traditional methods.

In another study, researchers employed ML to screen a library of MOFs for their catalytic activity in CO2 conversion processes. The integration of ML with HTS allowed for the rapid identification of promising candidates that were later validated experimentally, demonstrating the efficacy of this approach as in Figure 1.

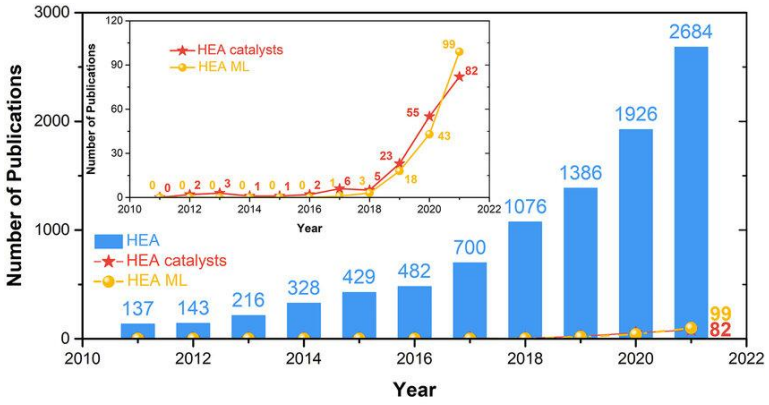

Figure 1 Number of Publications on HEAs, HEA Catalysts and HEA Machine Learning (ML)

A comparative analysis of traditional HTS methods versus ML-driven approaches revealed that the latter can reduce the time required for catalyst discovery by up to 50%. This reduction is primarily attributed to the ability of ML models to predict performance accurately, allowing researchers to focus on the most promising candidates.

3.3 Machine Learning Models for Catalyst Screening

Various machine learning models can be employed for catalyst screening, each with its strengths and weaknesses:
Regression models are commonly used to predict quantitative outcomes, such as catalytic activity or selectivity. Techniques like linear regression, support vector regression, and Gaussian processes can provide insights into how specific features influence performance. Classification models are used to categorize candidates into distinct classes based on their suitability for a given reaction. Common algorithms include decision trees, random forests, and logistic regression. These models can help identify which candidates are likely to be successful based on historical data.
Ensemble methods, such as gradient boosting and bagging, combine multiple models to improve prediction accuracy. These methods are particularly useful in materials science, where the relationships between features can be complex and nonlinear.
Deep learning models, particularly neural networks, have gained popularity in recent years due to their ability to capture intricate patterns in large datasets. Convolutional neural networks and recurrent neural networks can be applied to various types of data, including images and sequences, making them versatile tools for catalyst screening.
The effectiveness of machine learning models relies heavily on the quality of the training datasets and the validation techniques employed:
A well-curated training dataset is essential for building robust ML models. The dataset should include a diverse range of catalyst candidates with varying properties to ensure that the model can generalize effectively to new data. Cross-validation is a widely used technique to assess the performance of ML models. By dividing the dataset into training and validation sets, researchers can evaluate how well the model performs on unseen data. Hyperparameter tuning is also crucial, as it involves optimizing model parameters to enhance performance.
Evaluating the performance of machine learning models is critical to ensure their reliability in predicting catalyst performance:
Common metrics include accuracy, precision, recall, and F1-score. These metrics provide insights into the model's ability to make correct predictions and identify suitable candidates. One of the primary challenges in machine learning applications is interpreting the predictions made by models. Understanding how features contribute to the final predictions is crucial for gaining trust in the model and guiding experimental efforts.

4 DATA ANALYSIS

4.1 Data Collection

The first step in utilizing machine learning models for screening and evaluating potential catalyst materials is the systematic collection of relevant data. This data can be sourced from a variety of avenues, each contributing unique insights into the performance characteristics of catalyst materials.
High-throughput screening experiments are pivotal in generating empirical data on the performance of various catalyst candidates. This data encompasses a range of metrics, including: Catalytic Activity: Measured in terms of turnover frequency or reaction rates, this metric indicates how effectively a catalyst facilitates a chemical reaction. Selectivity refers to the ability of a catalyst to preferentially promote a specific reaction pathway over others, which is crucial for producing desired products with minimal by-products. Stability is Long-term performance data under operational conditions helps assess the durability of catalysts, providing insights into their practical viability in industrial applications, as in Figure 2.

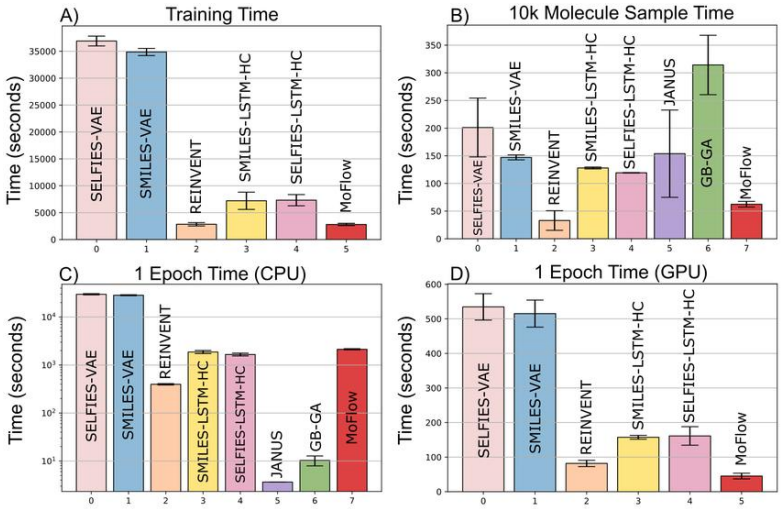

Figure 2 Results of Model Timing Comparisons

Advances in computational chemistry allow researchers to predict the properties of catalyst materials using methods such as density functional theory. These computational approaches can generate valuable data, including:

Electronic Structures is information about the distribution of electrons in a material, which affects its reactivity and catalytic properties. Formation Energies is the energy required to form a material from its constituent elements, which can indicate its stability and feasibility for synthesis. Reaction Pathways insights into the mechanisms by which reactants are converted into products, helping to identify rate-determining steps and potential bottlenecks.
Materials Databases established materials databases such as the Materials Project, AFLOW, and the Cambridge Structural Database serve as rich repositories of information. These databases provide extensive data on a wide array of materials, including: Structural Properties: Crystal structures, lattice parameters, and symmetry operations that define the arrangement of atoms in a material. Thermodynamic Characteristics: Data related to phase stability, enthalpy changes, and Gibbs free energy, which are essential for predicting reaction feasibility and efficiency.
In cases where experimental data is sparse or unavailable, synthetic data can be generated through various computational techniques. For instance:
Generative adversarial networks deep learning models can create realistic synthetic data that mimics the statistical properties of real datasets, thereby augmenting training sets for ML models. Monte Carlo simulations probabilistic method can be used to model the behavior of catalysts under different conditions, generating synthetic datasets that explore a wide range of scenarios.

4.2 Data Preprocessing

After collecting the data, it is imperative to preprocess it to ensure that it is clean, consistent, and suitable for training machine learning models. This preprocessing phase typically involves several key steps:
The raw data is often messy and may contain missing values, duplicates, or erroneous entries. Data cleaning involves handling missing values and removing duplicates. Techniques such as imputation or exclusion can be employed based on the extent and nature of the missing data as in Table 1.

Model		Activity	Stability	
	RMSE Loss	R2 Score	RMSE Loss	R2 Score
LGBM	8.65	0.98	0.07	0.88
LGBM + Hyperparameter Tuning	6.63	0.98	0.04	0.91
LGBM + Hyperparameter Tuning + Cross Validation	**3.52**	**0.99**	**0.004**	**0.94**
Chattoraj et al. (2022) ANN + XGBoost Ensemble	6.87	0.95	–	–

Table 1 Comparison of Different Applied Models for Conversion Modelling

Normalization and Scaling is to facilitate effective model training, numerical features should be normalized or scaled to a common range. This is essential because: Features such as catalytic activity and surface area may have vastly different ranges, leading to biases in model training. Techniques like Min-Max scaling (rescaling features to a range of [0, 1]) or Z-score normalization (standardizing features to have a mean of 0 and a standard deviation of 1) can help mitigate this issue. Many datasets contain categorical variables (such as material types or synthesis methods), which need to be converted into numerical formats for ML algorithms to process them effectively. Common techniques include:
One-Hot encoding creates binary columns for each category, allowing the model to interpret categorical data without imposing an ordinal relationship. Label encoding technique assigns a unique integer to each category, which can be useful for ordinal data but should be used cautiously to avoid misinterpretation by the model.

4.3 Feature Selection and Engineering

Feature selection and engineering are critical steps in the data analysis process, as they directly influence the performance of machine learning models. This involves identifying which features are most relevant for predicting catalytic performance and creating new features that may enhance model accuracy.
Identifying Relevant Features: The selection of features is fundamental to the success of ML models. Key electronic characteristics such as band gap, electronegativity, and density of states can significantly influence catalytic behavior. For instance, the band gap can provide insights into the electronic transitions that facilitate catalytic reactions. Information about crystal structure, surface area, and porosity is crucial, as these factors can affect the accessibility of active sites and the overall reactivity of the catalyst. Features such as formation energy, enthalpy of reaction, and Gibbs free energy are essential for understanding the stability and driving forces behind catalytic processes.
As datasets grow in complexity, dimensionality reduction techniques can be employed to simplify the feature space while retaining essential information. This step can help mitigate the “curse of dimensionality” and improve model performance. Common techniques include:
Principal Component Analysis (PCA): A statistical method that transforms the data into a new coordinate system, where the greatest variance by any projection lies on the first coordinate (principal component). PCA helps in reducing the number of features while preserving the most informative aspects of the data.

T-distributed Stochastic Neighbor Embedding: This technique is particularly useful for visualizing high-dimensional data in lower dimensions, making it easier to identify patterns and clusters among catalyst candidates.
Deriving new metrics that capture interactions between electronic and structural properties can yield insights into how these factors collectively influence catalytic performance. For example, combining surface area and electronic density may provide a more holistic view of a material's catalytic potential. Engaging with domain experts can help identify additional features that may not be immediately apparent but are critical for understanding catalyst behavior. This collaboration can lead to the discovery of novel features that improve model performance.

4.4 Model Training

With a well-prepared dataset, the next step is to train machine learning models to predict catalyst performance. Several algorithms can be utilized, each with its strengths and weaknesses.
If the task involves predicting continuous outcomes (e.g., catalytic activity), regression models such as linear regression, support vector regression, or random forests can be employed. These models can capture linear and nonlinear relationships between features and outcomes.
For tasks involving categorical outcomes (e.g., classifying catalysts as high-performing or low-performing), classification algorithms such as decision trees, logistic regression, or gradient boosting can be utilized. These models can help identify which candidate materials are likely to succeed based on historical data.
Techniques that combine the predictions of multiple models, such as bagging or boosting, can improve accuracy and robustness. Random forests, for example, utilize an ensemble of decision trees to enhance predictive performance and reduce overfitting.
For large and complex datasets, deep learning models such as neural networks may be employed. These models can automatically learn hierarchical representations of data, making them particularly effective for capturing intricate patterns in catalyst behavior. Convolutional neural networks can be particularly useful for analyzing structural data, while recurrent neural networks can be applied to sequential data.

4.5 Model Evaluation

After training the models, it is essential to evaluate their performance using appropriate metrics to ensure that they generalize well to unseen data.
Depending on the nature of the task, various metrics can be used to assess model performance. Metrics such as Mean Absolute Error, Mean Squared Error, and R-squared values provide insights into how well the model predicts continuous outcomes. Metrics such as accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve can be employed to evaluate the model's ability to classify catalyst candidates correctly.
To ensure that the model generalizes well to unseen data, k-fold cross-validation can be employed. This technique involves dividing the dataset into k subsets, training the model on k-1 subsets, and validating it on the remaining subset. This process is repeated k times, allowing for a more robust evaluation of model performance.
Optimizing hyperparameters is crucial for improving model performance. Techniques such as grid search or randomized search can be used to identify the best hyperparameter settings. This step is essential for fine-tuning the model to achieve optimal performance.

4.6 Insights and Interpretation

Once the models have been trained and evaluated, the next step is to derive insights and interpret the results effectively.
The trained models can be used to predict the performance of thousands of potential catalyst materials, allowing researchers to efficiently identify the most promising candidates for further experimental validation. This capability significantly accelerates the discovery process, enabling researchers to focus their efforts on the most viable options.
Analyzing feature importance scores can provide valuable insights into which properties most significantly influence catalytic performance. This information can guide further research efforts and help researchers focus on optimizing specific features that enhance catalyst efficacy.
Data visualization techniques, such as scatter plots, heatmaps, and feature importance plots, can be employed to present the results effectively. Visualizations can help communicate findings to stakeholders and facilitate decision-making, making complex data more accessible and understandable.
The insights gained from the analysis can inform the next iteration of data collection and model training. As new experimental data becomes available, it can be incorporated into the dataset, allowing for continuous improvement of the models and further refinement of catalyst screening. This feedback loop is crucial for maintaining the relevance and accuracy of the predictive models.
The insights derived from ML models can also help guide experimental efforts by suggesting specific synthesis methods or conditions that are likely to yield successful catalysts. This targeted approach can save time and resources, ultimately leading to more efficient catalyst development.

5 CONCLUSION

The integration of machine learning with high-throughput screening represents a transformative approach to catalyst discovery for CO2 hydrogenation. ML-driven HTS enables researchers to efficiently identify and optimize catalysts, significantly reducing the time and resources required for experimental validation. By leveraging vast datasets and advanced predictive models, researchers can make informed decisions about which candidates to prioritize for testing.
The implications of this research extend beyond CO2 hydrogenation. The methodologies developed can be applied to a wide range of catalytic processes, contributing to the advancement of sustainable energy solutions. As the demand for cleaner energy sources continues to grow, the ability to rapidly discover and optimize catalysts will be crucial in addressing global challenges related to climate change and energy production.
To fully realize the potential of machine learning-driven high-throughput screening, further research and collaboration within the scientific community are essential. Encouraging open-access data sharing and fostering partnerships between academia and industry will facilitate innovation and accelerate progress in catalyst discovery. By working together, researchers can harness the power of machine learning to create a more sustainable future.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

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] 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.
[3] Gao W, Chen Y, Li Y, et al. Automated catalyst design for CO2 electroreduction. Nature, 2022, 610(7931): 287-293.
[4] Sun T, Yang J, Li J, et al. Enhancing Auto Insurance Risk Evaluation with Transformer and SHAP. IEEE Access, 2024.
[5] Lin Y, Fu H, Zhong Q, et al. The influencing mechanism of the communities' built environment on residents' subjective well-being: A case study of Beijing. Land, 2024, 13(6):793.
[6] Li J, Fan L, Wang X, et al. Product Demand Prediction with Spatial Graph Neural Networks. Applied Sciences, 2024, 14(16): 6989.
[7] Liu M, Ma Z, Li J, et al. Deep-Learning-Based Pre-training and Refined Tuning for Web Summarization Software. IEEE Access, 2024.
[8] Chen X, Liu M, Niu Y, et al. Deep-Learning-Based Lithium Battery Defect Detection via Cross-Domain Generalization. IEEE Access, 2024.
[9] Wang X, Wu YC, Ji X, et al. Algorithmic discrimination: examining its types and regulatory measures with emphasis on US legal practices. Frontiers in Artificial Intelligence, 2024, 7: 1320277.
[10] 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, 7:1306058.
[11] Ma Z, Chen X, Sun T, et al. Blockchain-Based Zero-Trust Supply Chain Security Integrated with Deep Reinforcement Learning for Inventory Optimization. Future Internet, 2024, 16(5): 163.
[12] Wang X, Wu YC, Zhou M, et al. Beyond surveillance: privacy, ethics, and regulations in face recognition technology. Frontiers in big data, 2024, 7: 1337465.
[13] Wang X, Wu YC. Balancing innovation and Regulation in the age of generative artificial intelligence. Journal of Information Policy, 2024, 14.
[14] Zuo Z, Niu Y, Li J, et al. Machine Learning for Advanced Emission Monitoring and Reduction Strategies in Fossil Fuel Power Plants. Applied Sciences, 2024, 14(18): 8442.
[15] Toyao T, Maeno Z, Takakusagi S, et al. Machine Learning for Catalysis Informatics: Recent Applications and Prospects. ACS Catal, 2020, 10(3): 2260-2297.
[16] Zhong M, Tran K, Min Y, et al. Accelerated discovery of CO2 electrocatalysts using active machine learning. Nature, 2020, 581(7807): 178-183.
[17] Schlexer Lamoureux P, Winther KT, Garrido Torres JA, et al. Machine Learning for Computational Heterogeneous Catalysis. ChemCatChem, 2019, 11(16): 3581-3601.
[18] 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.
[19] Gu GH, Noh J, Kim S, et al. Practical deep-learning representation for fast heterogeneous catalyst screening. NPJ Comput Mater, 2019, 5(1): 47.
[20] Jia X, Lynch A, Huang Y, et al. Anthropogenic CO2 reduction: from materials design to flow reactors. Mater Today, 2019, 24: 15-47.
[21] Medford AJ, Kunz MR, Ewing SM, et al. Extracting knowledge from data through catalysis informatics. ACS Catal, 2018, 8(8): 7403-7429.
[22] 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.

[23] Grajciar L, Heard CJ, Bondarenko AA, et al. Towards operando computational modeling in heterogeneous catalysis. Chem Soc Rev, 2018, 47(22): 8307-8348.
[24] 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.
[25] Boes JR, Mamun O, Winther K, et al. Graph Theory Approach to High-Throughput Surface Adsorption Structure Generation. J Phys Chem A, 2019, 123(11): 2281-2285.
[26] 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.
[27] 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.
[28] Li Z, Ma X, Xin H. Feature engineering of machine-learning chemisorption models for catalyst design. Catal Today, 2017, 280: 232-238.
[29] 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.
[30] Gusarov S, Stoyanov SR, Siahrostami S, et al. Deep Learning for Molecular Design and Discovery: From Chemical Reactions to Materials. ACS Cent Sci, 2021, 7(6): 902-915.
[31] 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.
[32] Kitchin JR. Machine learning in catalysis. Nat Catal, 2018, 1(4): 230-232.
[33] Wan Y, Zhang Z, Xu X, et al. Engineering active sites of polyoxometalate-based metal-organic frameworks for boosting CO2 electroreduction. Nat Commun, 2021, 12(1): 2870.
[34] Gómez-Bombarelli R, Wei JN, Duvenaud D, et al. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. ACS Cent Sci, 2018, 4(2): 268-276.
[35] Andersen M, Levchenko SV, Scheffler M, et al. Beyond Scaling Relations for the Description of Catalytic Materials. ACS Catal, 2019, 9(4): 2752-2759.
[36] Rosen AS, Iyer SM, Ray D, et al. Machine learning the quantum-chemical properties of metal–organic frameworks for accelerated materials discovery. Matter, 2021, 4(5): 1578-1597.
[37] Artrith N, Butler KT, Coudert FX, et al. Best practices in machine learning for chemistry. Nat Chem, 2021, 13(6): 505-508.
[38] Schleder GR, Padilha ACM, Acosta CM, et al. From DFT to machine learning: recent approaches to materials science–a review. J Phys Mater, 2019, 2(3): 032001.
[39] Melander M, Kuisma MJ, Christensen TE, et al. Grand-canonical approach to density functional theory of electrocatalytic systems: Thermodynamics of solid-liquid interfaces at constant ion and electrode potentials. J Chem Phys, 2019, 150(4): 041706.
[40] Janet JP, Kulik HJ. Machine Learning in Inorganic Chemistry. J Phys Chem A, 2021, 125(39): 8561-8568.
[41] Asif M, Yao C, Zuo Z, et al. Machine learning-driven catalyst design, synthesis and performance prediction for CO2 hydrogenation. Journal of Industrial and Engineering Chemistry, 2024.
[42] Xiong Z, Wang Y, Cheng T, et al. Revealing the active phase of copper during the electroreduction of CO2 in aqueous electrolyte by correlating in situ X-ray spectroscopy and in situ electron microscopy. Nat Catal, 2020, 3(8): 636-642.
[43] Palkovits R, Palkovits S. Efficient catalyst discovery for sustainable chemical synthesis via high-throughput experimentation and machine learning. Nat Commun, 2022, 13(1): 3675.
[44] Chen C, Zhu X, Wen X, et al. Coupling N2 and CO2 in H2O to synthesize urea under ambient conditions. Nat Chem, 2020,12(8): 717-724.
[45] Nørskov JK, Bligaard T, Rossmeisl J, et al. Towards the computational design of solid catalysts. Nat Chem, 2009, 1(1): 37-46.
[46] Wang Y, Liu J, Wang Y, et al. Synergy of electron-phonon coupling and lattice distortion in crystal structure-dependent CO2 activation. Nat Commun, 2021, 12(1): 1013.
[47] Tao H, Stach EA, Goodman ED, et al. Accelerating Catalysts Discovery through Machine Learning and High-Throughput Experimentation. ACS Catal, 2022, 12(5): 2962-2968.