



Cite this: *New J. Chem.*, 2025, **49**, 6888

Artificial intelligence-driven advances in photocatalytic hydrogen production

Leandro Goulart de Araujo * and David Farrusseng

This perspective provides an overview of recent studies on the use of photocatalysis for hydrogen production, with a particular focus on water splitting. It examines the developments in this field that have been facilitated by artificial intelligence tools, especially machine learning algorithms. The photocatalytic generation of hydrogen has been the subject of extensive study in recent years, as the necessity for higher efficiency and hydrogen production rates represents a crucial step in the development of this technology for its mass deployment. The known difficulties of these systems pertain to the complexities associated with the photocatalysts, including the effect of reactants, the synthesis process, and their efficiency, particularly in harvesting sunlight. Moreover, the design of the reactor is a challenging undertaking, particularly in light of the dynamic behavior and the interaction between photons, solutions, photocatalysts, and co-catalysts that must be considered in the photocatalytic production of hydrogen. Research on this subject must consider the use of green materials and processes for synthesis, avoid extensive experimentation to reduce the carbon footprint, and seek efficient and less resource-intensive computational resources. To surmount the challenges inherent to the synthesis and development process, while simultaneously enabling the establishment of structure–performance relationships for knowledge acquisition, high-performing computational methods are key. This article concludes by potential avenues for improvement by highlighting strategies that have been successfully implemented in other related fields and could be beneficial for this field.

Received 6th February 2025,
Accepted 7th April 2025

DOI: 10.1039/d5nj00505a

rsc.li/njc

IRCELYON, Institut de Recherches sur la Catalyse et l'Environnement de Lyon, UMR5256 CNRS-Université de Lyon, 69626 Villeurbanne, France.
E-mail: lgoulart@alumni.usp.br, david.farrusseng@ircelyon.univ-lyon1.fr



Leandro Goulart de Araujo

Leandro Goulart de Araujo holds a PhD in Chemical Engineering from the University of São Paulo. He was a postdoctoral researcher at the Nuclear and Energy Research Institute (2019–2021) and Université de Lorraine/Institut Jean Lamour (2022–2023), working on machine learning for radioactive waste management and photocatalysis. Currently, he is a postdoc at IRCELYON, focusing on CFD modeling of biogas cracking for turquoise hydrogen production. His research explores kinetic modeling and AI in (photo)catalysis. He has co-authored 35 articles and holds an MSc in Nuclear Engineering and MBAs in Project Management and Data Science & Analytics.



David Farrusseng

David Farrusseng is currently a group leader at IRCELYON. His research activities are focused on the design of materials for innovative catalytic & separation processes. He has co-authored over 200 research articles and holds 17 patents, covering various aspects of adsorption and catalytic processes. In recognition of his contributions to porous materials, particularly MOFs, he received the prestigious IACS Award in 2016. He also serves as IACS President, coordinates large EU-funded R&D programs, and acts as a consultant for major energy, petrochemical, and materials companies.

1. Introduction

In recent years, considerable attention has been devoted to the advancement of efficient and environmentally friendly energy sources, with the objective of addressing the emerging challenges posed by climate change and the need for renewable energy. While solar energy is considered an attractive option due to its abundance and cost-free nature, its low conversion efficiency and challenges associated with storing the generated energy present significant obstacles to its implementation on a larger scale.¹ Nevertheless, the production of hydrogen *via* photocatalysis represents a formidable approach to the exploitation of solar energy.² The production of hydrogen *via* water splitting with photocatalysis offers numerous advantages, including (i) the reasonable distribution of water resources around the planet; and (ii) its status as a clean energy source, particularly given its high calorific value, which eliminates the generation of pollutants during combustion.^{3,4} Additionally, utilizing alternative substrates, such as wastewater or organic waste, presents an attractive waste-to-energy approach.^{5,6} This method not only treats contaminated water but also generates hydrogen alongside carbon dioxide and water, providing a dual benefit of waste treatment and clean energy generation.⁷ Nevertheless, the primary challenge remains the development of highly efficient and cost-effective photocatalysts for sustainable hydrogen production.³

In this context, considerable effort has been directed towards the development of novel photocatalysts and the enhancement of existing ones that demonstrate the capacity to efficiently harvest visible light and facilitate photocatalytic hydrogen production. The anticipated attributes of photocatalysts include superior photochemical activity, a broad spectrum of light absorption, and the provision of sufficient charge separation.⁸

The growing studies on known photocatalysts and the continuous discovery of new ones make the screening of these materials a critical focus in this field.⁹ The complex nature of these materials and photocatalysis processes, which presents a significant challenge to understanding the systems, necessitates the use of mathematical approaches to advance scientific understanding and develop practical systems for industrial applications.

The complexity of this system arises from its highly non-linear behavior, which complicates the identification of relationships and the assessment of variable effects in reactor and system design. Additionally, the use of advanced and intricate materials as photocatalysts further amplifies the challenges in understanding these processes. Key examples of this complexity include: (i) light absorption and charge recombination, which are inherently quantum-based phenomena;^{10,11} (ii) the influence of surface defects in solids, often the most catalytically active sites due to their low coordination;^{12–14} and (iii) the presence of non-Euclidean structures in certain photocatalysts, such as metal–organic frameworks (MOFs), adding further intricacy to their study and optimization.¹⁵ For instance, the heteroepitaxial structure of $\text{PbTiO}_3/\text{SrTiO}_3$ exemplifies this complexity: its interface demonstrates quantum-based phenomena, such as light

absorption and charge recombination, while utilizing surface defects as active catalytic sites.¹⁶ These intricate material designs underscore the challenges in optimizing photocatalytic systems for hydrogen production *via* water splitting.

Given these multifaceted challenges, advanced computational tools become indispensable. In this context, the application of artificial intelligence (AI) tools, especially those related to machine learning (ML) methods, provides tools that may allow the advancement of knowledge and provide unanticipated directions for discovery of new photocatalysts systems.

In recent years, there has been spectacular progress in the utilization of ML models across a range of scientific disciplines. This has led to the implementation of numerous applications of these models, which have been met with interest from the scientific community. Among the various developments in photocatalysis that have been facilitated by the use of ML techniques, we seek to highlight the deployment of large language models (LLMs) and the more specific implementations derived therefrom in the context of photocatalysis. This advancement has enabled the development of automated and highly efficient data collection tools for obtaining secondary data, a topic that will be discussed in this perspective. This has been addressed by combining ML algorithms with state-of-the-art characterization equipment, which has enhanced our comprehension of the photocatalytic processes for hydrogen production from water and wastewater, in addition to organic compounds.

In this perspective, we examine the implementation of ML in photocatalytic processes for hydrogen production, with a particular focus on water splitting, while limiting the discussion on oxygen evolution. We explore the acquisition of secondary data through automated tools, the use of ML for photocatalyst design and properties prediction, and the identification of relationships between operating variables. The complex interplay between physicochemical factors, including materials features, the use of sacrificial agents, pH, photocatalyst dosage, and their interactions, presents a significant challenge in developing empirical models that consider both individual and interaction influences. Failing to consider interactions may result in an insufficient amount of information and potentially unsatisfactory predictions. Consequently, this work also examines the methods that have employed ML to gain a deeper understanding of the process, enhance degradation rates, study kinetics, and propose mathematical models. Additionally, we emphasize the potential for cross-field applications from materials science and general catalysis, including the utilization of expert-curated datasets and the application of advanced techniques from deep learning to gather information from analytical data.

2. Setting the scene of photocatalysis in hydrogen production

2.1 Bibliometric analysis of the utilization of photocatalysis for hydrogen production

The potential benefits of hydrogen production *via* photocatalysis-based processes have garnered significant interest

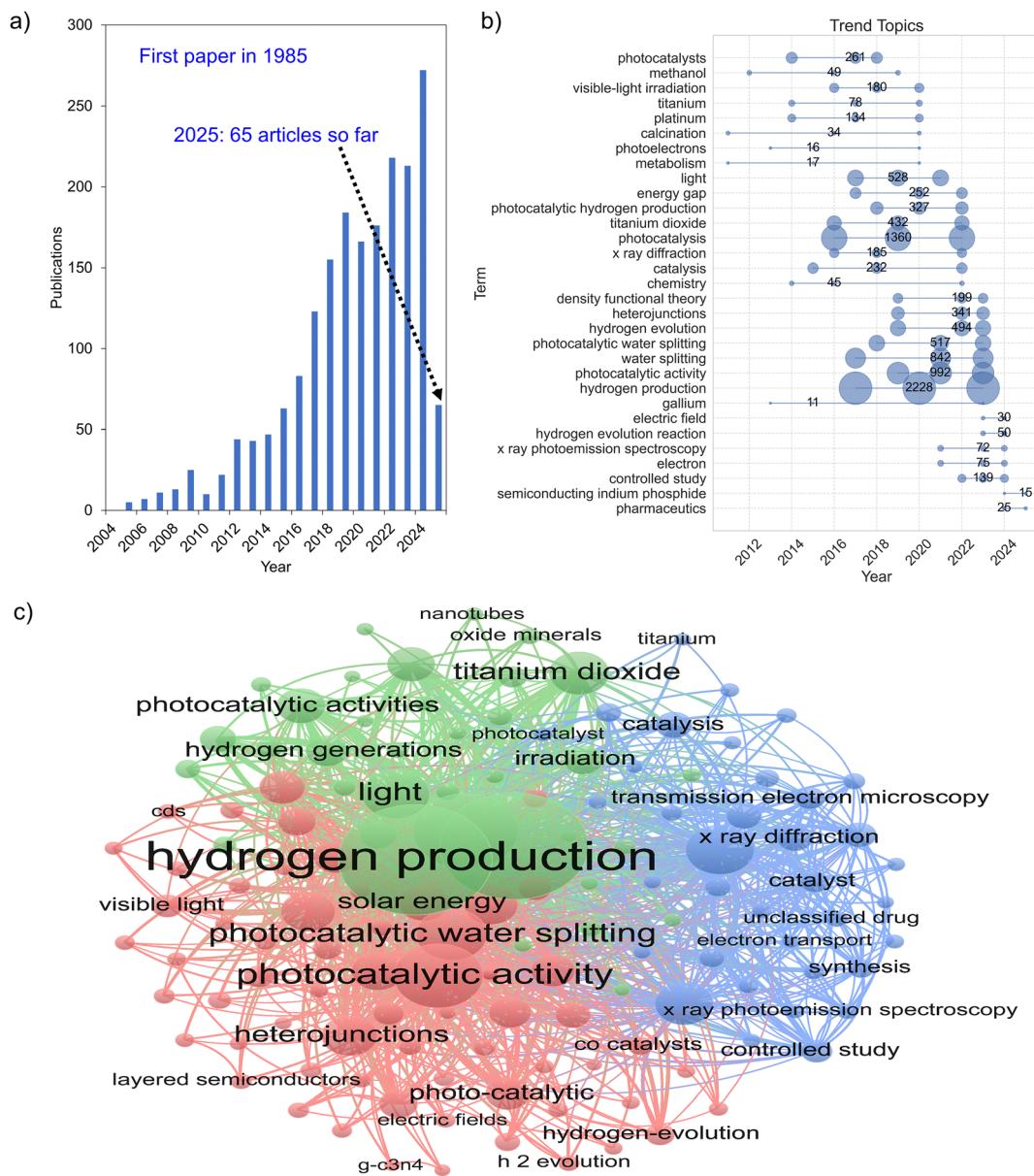


Fig. 1 Bibliometric data for “photocatalysis” AND (“hydrogen production”) OR (“hydrogen generation”) AND “water splitting”. (a) Annual scientific production, (b) recent trend topics from the last decade. The results were obtained using the bibliometrix package in RStudio. (c) Co-occurrence network with all keywords (clustering algorithm: Louvain-based, normalization: association, number of nodes: 137, minimum cluster size: 30, minimum number of occurrences of a keyword: 50), obtained by using the VOSviewer software.¹⁷ Data were collected via the Scopus document search in February 2025 (language: English, document type: articles).

within the scientific community. As illustrated in Fig. 1a, there has been a notable surge in the number of scientific publications exploring the use of photocatalysis for hydrogen production through water splitting in the last decade (2014–2024), with 1700 publications from a total of 1949 papers, and 65 papers already published in 2025 (as of February 2025).

As shown in Fig. 1b, the utilization of titanium as a photocatalyst for hydrogen generation through water splitting has declined over time. Despite its longstanding use and extensive research, titanium has lost its initial appeal. In contrast, gallium, although appearing slightly before than titanium, is

still a subject of active research. Notably, semiconducting indium phosphide has emerged as a recent development, with 15 documented appearances in the past year. Some of the recent trend topics are “controlled study” and “density functional theory” (DFT), which may indicate efforts to understand the effects of operating parameters and to develop novel photocatalysts using theory-driven approaches. Another term that warrants attention is “heterojunctions,” which denotes the utilization of two or more semiconductors, exhibiting disparate characteristics such as band gaps and electronic properties, to yield materials with distinct electrical and optical attributes

compared to their isolated counterparts. In fact, DFT investigations have played a role in the development of materials utilizing heterojunction interfaces.^{18,19} In a further section will examine recent developments in the synthesis of photocatalysts, including studies on heterojunction interfaces and the use of ML to enhance the synthesis process or to provide a deeper understanding of it.

As demonstrated in Fig. 2, three primary clusters can be identified. The main node of the green cluster is “hydrogen production”, the main one for the red cluster is “photocatalytic activity”, and the main one of the blue cluster is “X-ray diffraction”, representing other nodes related to analytical techniques. It can be observed that “photocatalytic activity” is associated with “solar energy” and “heterojunctions”. This observation underscores contemporary initiatives aimed at leveraging solar energy and heterojunction-based materials. The presence of materials such as titanium dioxide (TiO_2), carbon nitride ($\text{g-C}_3\text{N}_4$), and II–VI semiconductors, including cadmium sulfide (CdS), as well as nanotubes and oxide minerals, is also noted. Titanium dioxide is a logical inclusion, given its established status as a powerful semiconductor and its high efficiency under UVA irradiation. Nevertheless, the majority of research has been directed towards the development of materials capable of efficiently harvesting visible light irradiation, thereby enhancing their performance under solar light.

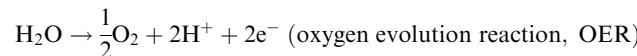
This has led to the utilization of materials such as $\text{g-C}_3\text{N}_4$ and II–VI semiconductors, such as CdS , as they possess distinctive features, including adjustable forbidden bands, which potentially enhance light absorption in the visible and ultraviolet (UV) ranges, higher electronic mobility, and ease of synthesis and structure modification. Despite the recent emergence of semiconducting indium phosphide, III–V semiconductors were

not observed in the co-occurrence network, likely due to the novel applicability of the latter. Other keywords pertain to traditional materials characterization techniques, such as X-ray diffraction (XRD) and scanning electron microscopy (SEM), and there is also the presence of DFT, as previously observed in the trend topics.

2.2 Mechanistic insights and modeling of photocatalytic water splitting

2.2.1 Fundamental mechanisms of photocatalytic water splitting

Photocatalytic water splitting is a promising method for converting solar energy into chemical energy, enabling the decomposition of water into hydrogen (H_2) and oxygen (O_2) without the need for external electrical input. The two fundamental half-reactions in water splitting are:



This process relies on a photocatalyst capable of absorbing photons with energy equal to or greater than its band gap. This promotes the excitation of electrons from the valence band (VB) to the conduction band (CB), leaving behind positively charged holes (h^+). The photogenerated electrons in the CB drive HER, while the holes in the VB participate in OER. This mechanism is analogous to electrolysis but relies on light energy rather than an external electrical bias.²¹ The electron–hole pairs mediate the redox reactions involved (Fig. 2a).

Significant research efforts have focused on understanding the overall process of water splitting and the structure–property

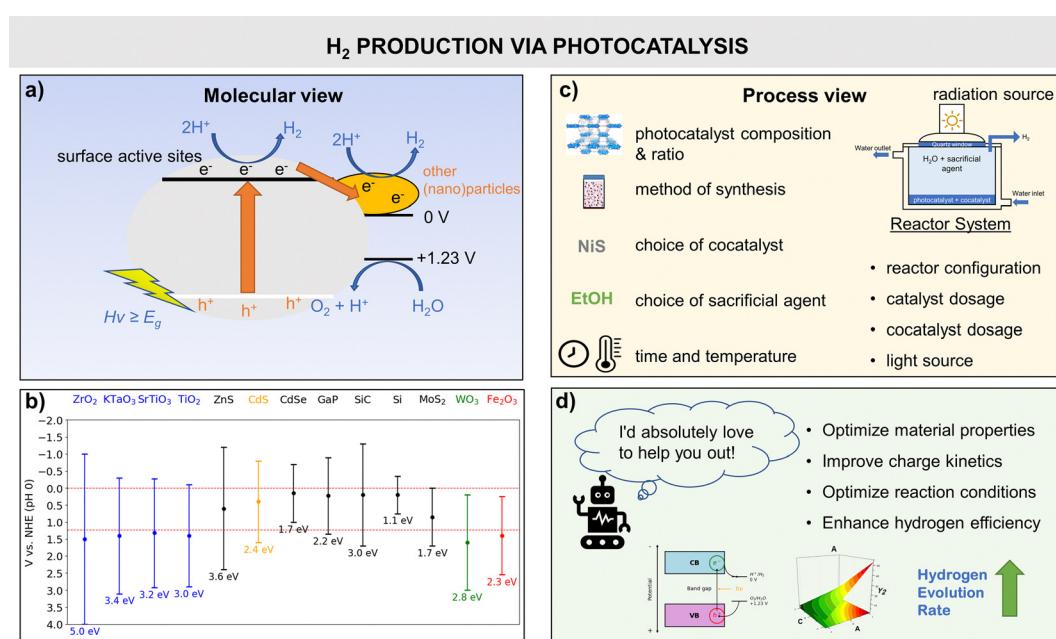


Fig. 2 A schematic diagram showing the molecular and process views for the hydrogen production cycle via photocatalysis. (a) Principle of water splitting using semiconductor photocatalysts; (b) relationship between band structure of semiconductor and redox potentials of water splitting (adapted from ref. 20); (c) commonly parameters studied in this process; (d) opportunities for ML intervention.

relationships of photocatalysts.²² One strategy to enhance efficiency is the use of sacrificial reagents, which act as electron donors or acceptors, suppressing recombination losses and increasing reaction rates.²³ However, a catalyst alone, even in the presence of a sacrificial reagent, does not always ensure the desired 2:1 H₂:O₂ ratio.

The overall water splitting (OWS) process transforms light energy into chemical energy, accumulating Gibbs free energy.²⁴ In OWS, a direct redox reaction with the pair electron/hole is the driving force. Although OWS typically operates under pure water conditions, it can also involve sensitizers, such as organic dyes, to enhance light absorption and charge transfer. In this case, water remains the primary reactant, distinguishing it from photocatalytic reforming of organics (PRO), where organics actively participate in the reaction as electron donors. In this process, organic species act as electron donors, facilitating charge separation and reducing recombination losses.²⁵ Unlike OWS, where charge recombination competes with water oxidation and proton reduction, PRO benefits from sequential electron transfer, limiting back-electron transfer and stabilizing reaction intermediates, thereby improving hydrogen yield.^{25,26} The key distinction between these two approaches lies in the electron transfer mechanism, which significantly influences reaction kinetics and selectivity.

While OWS follows a stoichiometric pathway, PRO enhances hydrogen yield by supplying external electrons, facilitating proton reduction, and mitigating recombination effects. However, PRO often requires sacrificial agents that are consumed in the reaction,²⁷ whereas OWS aims for long-term sustainability by utilizing only water as the reactant. In contrast to OWS, which requires precise band alignment to drive both HER and OER, PRO relaxes this constraint, making it a more viable alternative under certain conditions.²⁸ This flexibility allows PRO to operate effectively even when the photocatalyst does not perfectly match the redox potentials of water splitting. Nonetheless, the development of advanced photocatalysts and co-catalysts has significantly improved charge separation and efficiency in OWS, narrowing the difference in performance between these two approaches.²⁹

Despite its advantages, PRO can lead to the formation of undesirable intermediate organic compounds when oxidation is incomplete, including CO, CO₂, CH₄, C₂H₄ (greenhouse gases), as well as aldehydes and carboxylic acids.²⁵ These intermediates may not only lower efficiency but also complicate product purification and reactor operation. The selection between these methods depends on factors such as electron transfer efficiency, photocatalyst stability, and overall reaction kinetics.³⁰ Since pure water splitting is thermodynamically demanding due to the high energy required to break O-H bonds and charge carrier recombination,³¹ organic reforming provides a more favorable pathway for hydrogen evolution. Yet, OWS is often regarded as a promising long-term strategy for green hydrogen production, as it eliminates reliance on organic species and offers a potentially renewable and self-sustaining process.

The efficiency of water splitting is influenced by several factors, including the choice of co-catalysts, heterostructures,

and dopants, which improve charge separation and transfer efficiency.^{29,32,33} While diverse light sources can be used, visible-light-driven systems are preferred due to the limited availability of UV light in the solar spectrum.³⁴

Despite significant advancements, charge separation and efficiency remain major obstacles to the practical application of photocatalytic materials. Several strategies have been explored to improve these aspects:

(1) Heterojunction engineering: constructing type-II,^{35,36} Z-scheme,^{37,38} or Schottky junctions^{39,40} to enhance directional charge transfer.

(2) Surface modification: introducing co-catalysts (*e.g.*, Pt, Ni, RuO₂) to facilitate charge transfer and reduce overpotentials.^{41–43}

(3) Defect engineering: creating oxygen vacancies⁴⁴ or introducing dopants to generate mid-gap states, promoting charge separation.^{45,46}

To ensure optimal photocatalytic performance, thermodynamic and band gap considerations are essential. For effective water splitting, the conduction band minimum (CBM) must be more negative than the H⁺/H₂ reduction potential (0 V *vs.* NHE), while the valence band maximum (VBM) must be more positive than the O₂/H₂O oxidation potential (+1.23 V *vs.* NHE).⁴⁷ Consequently, the minimum theoretical band gap required is 1.23 eV, corresponding to light with a wavelength of approximately 1100 nm.²⁰ However, in practical application, photocatalysts with band gaps near this value exhibit poor reaction kinetics and high recombination rates, necessitating additional kinetic overpotentials.^{48–50}

To optimize performance under visible light, band gaps in the range of 1.8–3.0 eV are preferred, as displayed in Fig. 2b. This range balances efficient photon absorption with sufficient charge separation potential, ensuring both effective light harvesting and reasonable reaction kinetics. Photocatalysts with band gaps below 1.8 eV often suffer from high charge carrier recombination, reducing efficiency, whereas those above 3.0 eV require high-energy UV photons, limiting their practical use under sunlight-driven applications. This trade-off between charge separation efficiency and light absorption spectrum defines the key challenge in developing effective photocatalysts. The relationship between the band gap energy and the photon wavelength is given by:

$$E_g \text{ (eV)} = \frac{1240}{\lambda \text{ (nm)}}$$

where E_g is the band gap energy and λ is the incident photon wavelength. Although these principles guide the development of efficient photocatalysts, several intrinsic challenges still hinder large-scale implementation, as discussed in the following section.

2.2.2 Challenges and limitations in photocatalytic water splitting. Photocatalytic water splitting faces multiple fundamental challenges that limit its large-scale application. These challenges include thermodynamic constraints, charge carrier dynamics, material stability, and fabrication scalability.⁵¹ One of the most widely studied photocatalysts, TiO₂, serves as an example of these limitations.⁵² TiO₂-based photocatalysis has been extensively employed in diverse fields, such as water

treatment,⁵³ and hydrogen generation,⁵⁴ yet it presents several inherent limitations.

The VB of anatase-phase TiO₂ is positioned below the required potential for efficient water oxidation (+1.23 V vs. NHE), leading to insufficient oxidative power for OER. Additionally, TiO₂ exhibits a large band gap (~3.2 eV), which confines its light absorption to the UV spectrum. Given that UV accounts for only 4% of the solar spectrum,⁵⁰ this severely restricts its solar-to-hydrogen efficiency (STHE) when operating under natural sunlight. Conversely, alternative photocatalysts with narrower band gaps often suffer from stability issues, resulting in a dilemma between efficiency and durability. Alternatively, these catalysts may be too narrow, leading to a high recombination rate.⁵⁵

Despite decades of research, photocatalytic water splitting still faces several critical limitations that hinder its large-scale implementation. These challenges stem from both intrinsic material properties—such as band edge positioning and charge carrier recombination—and practical constraints, including stability under reaction conditions and economic feasibility. The most critical limitations are outlined below.⁵⁶

(1) Redox potential and band edge limitations: a suitable conduction band for the photocatalyst and a more positive VB than the potential to oxidize water to oxygen. However, only a limited number of materials satisfy these thermodynamic conditions while maintaining stability.

(2) Limited light absorption and charge carrier recombination: the majority of highly efficient photocatalysts (*e.g.*, TiO₂, SrTiO₃, GaN) are primarily UV-responsive, limiting their practical solar utilization. While there are visible-light responsive semiconductors, they tend to exhibit instability and photocorrosion. Photogenerated electron-hole pairs have short lifetimes, leading to recombination before the reaction can occur. This results in low reaction rates and poor STHE values. While a few pioneering studies have demonstrated STHE values close to 10%,^{57,58} these results typically require optimized laboratory conditions, including concentrated solar irradiation and carefully engineered catalyst systems. In contrast, most recent studies under standard sunlight exposure report an STHE of only about 2%,⁵⁸ highlighting the gap between experimental breakthroughs and practical scalability.

(3) Chemical stability and photocorrosion: many narrow-band-gap semiconductors (*e.g.*, CdS, BiVO₄) exhibit photocorrosion due to self-oxidation or self-reduction under illumination. This degradation pathway leads to the dissolution of the photocatalyst into the reaction medium, significantly reducing its long-term performance.^{59–61} Oxidative/reductive environments during water splitting may alter the photocatalyst structure, leading to deactivation.³¹

(4) Toxicity and environmental considerations: several high-performance photocatalysts rely on toxic or scarce elements, such as cadmium (Cd) and platinum (Pt).^{62,63} The environmental impact of these elements raises concerns regarding sustainable large-scale production.

(5) Cost and scalability limitations: efficient co-catalysts (*e.g.*, RuO₂, Pt) are expensive and rare, making large-scale deployment so far economically unfeasible.^{64,65} Many photocatalysts

exist as fine powders, which hinder reactor integration – dispersion and recovery – and require additional processing to enhance light absorption and reactant adsorption.⁵⁶

A comprehensive account of these challenges can be found in the referenced literature.⁵⁶ Note that these challenges are interconnected. As aforementioned and as examples, reducing the band gap to improve visible-light absorption often compromises stability due to increased photocorrosion susceptibility. Similarly, low-cost materials may exhibit limited efficiency, while high-performance photocatalysts may rely on toxic or rare elements.

Recent studies have sought to balance efficiency and scalability. For instance, heterostructure designs such as SiC/WS₂ have demonstrated promising STHE values (~17%),⁶⁶ yet achieving theoretical maximum efficiency (~18%) remains a challenge due to suboptimal reaction kinetics and material degradation. Addressing these challenges requires a multi-faceted approach, combining computational modeling techniques to predict and optimize photocatalyst performance and experimental validation to confirm theoretical insights. The next section explores how conventional methodologies contribute to overcoming these barriers.

2.2.3 Conventional strategies: experimental validation and computational modeling. To overcome these primary obstacles, a multitude of photocatalysts have been subjected to rigorous experimental scrutiny for their potential in photocatalytic water splitting. Prominent examples include MoS₂-based materials,⁶⁷ GaSe/YAlS₃,⁶⁸ MXene-based materials,⁶⁹

The synthesis of novel photocatalytic materials has been explored and analyzed through different schemes of charge transfer, with a particular focus on the electronic and optical properties of these materials. Moreover, the diverse array of advanced characterization techniques, including X-ray photoelectron spectroscopy (XPS), electron paramagnetic resonance spectroscopy (EPR) and atomic force microscopy (AFM), has been utilized to enhance comprehension of the materials' properties.

Recent studies have demonstrated the application of advanced characterization techniques in elucidating key aspects of photocatalyst performance. XPS has been widely employed to investigate electron transfer pathways and band structure modifications, such as in S-scheme heterojunctions, where it provides insights into charge migration at interfaces and the impact of cation doping on electronic states.^{70,71} EPR has been instrumental in identifying defect states, particularly oxygen vacancies, which influence charge carrier dynamics and photocatalytic activity, as observed in N-doped TiO₂.⁷² Meanwhile, AFM has been used to analyze the surface morphology of photocatalysts, with advanced techniques like Kelvin probe force microscopy (KPFM) enabling the mapping of charge separation under illumination, offering a deeper understanding of the electronic properties of nanostructured semiconductors.⁵⁰ These techniques collectively offer key insights into the structural, electronic, and charge transport properties of photocatalysts, guiding the rational design of more efficient water-splitting systems.

Advances in material science and nanotechnology are also of particular interest with regard to the next generation of

photocatalysts. The use of hybrid materials,⁷³ quantum dots,⁷⁴ and plasmonic nanoparticles⁷⁵ is expected to result in improved light absorption and catalytic efficiency. Furthermore, combined technologies have been the subject of study with a view to overcoming current limitations, including the combination of photocatalysis with other renewable technologies, such as solar concentrators⁷⁶ and electrocatalysis.⁷⁷

While the selection of the photocatalyst, including data regarding its characteristics and capacity for performance under sunlight, is of paramount importance, it is also evident that other experimental conditions exert a significant influence on the efficiency of STHE.

The primary variables under consideration in experimental studies, as illustrated in Fig. 2c, present researchers with a crucial decision: to limit their investigation to a select few parameters and examine them individually, or to employ experimental design methods and screen multiple variables across diverse experimental setups. This decision is influenced by the potential for numerous variables to impact the outputs of interest, such as STHE.

The decision-making process is not straightforward due to the lack of knowledge regarding the effects of these variables on efficiency. The influence of these variables on the target responses is often unknown, making it challenging to determine the optimal approach.

Although it would be advantageous to screen the maximum number of potential variables, the high dimensionality of the problem makes it challenging to gain an understanding of the process. Moreover, the financial burden associated with the acquisition of equipment, personnel, and other resources may prove to be a significant obstacle to feasibility.

Three of the most widely used computational modeling techniques in photocatalytic water splitting are DFT, kinetic Monte Carlo (KMC), and computational fluid dynamics (CFD) (Table 1).

DFT is a quantum mechanical modeling method used to study the electronic structure, mainly the ground state, of materials at the atomic scale,¹³⁶ and to determine optimized structures and preferred locations of metal nanoparticles on catalyst supports, such as Pt on $\gamma\text{-Al}_2\text{O}_3$.¹³⁷ In the context of photocatalysis, it helps predict band structures,^{138,139} charge transfer mechanisms,¹⁴⁰ surface adsorption properties,¹⁴¹ and reaction energetics (*e.g.*, energy barriers, reaction energy profiles, and Gibbs free energy)^{142–144} in photocatalytic materials. DFT is widely used to determine the band edge positions of semiconductors and evaluate the effects of doping or heterostructure formation on visible light absorption.

KMC is a stochastic simulation method used to model reaction kinetics, charge carrier dynamics, and surface interactions over extended time scales.⁸⁶ Unlike traditional quantum mechanical methods, KMC employs random sampling to simulate state-to-state transitions in dynamic systems based on pre-defined reaction rate constants.⁸⁷ In photocatalysis, it is primarily used to simulate charge transport, recombination, diffusion, and reaction kinetics on semiconductor surfaces, providing insights into efficiency and degradation mechanisms.⁸⁷ For instance,

KMC has been used to model reaction rates and pathways,⁸⁷ and electron–hole migration in BiVO_4 , revealing how defects and doping influence photocatalytic performance.¹⁴⁵

CFD is a numerical modeling technique used to simulate fluid flow, heat transfer, and mass transport in photocatalytic reactors.⁸⁸ It is widely employed in reactor design optimization, predicting light distribution, and modeling gas evolution.^{89,90} In photocatalysis, CFD helps enhance reactor efficiency by optimizing light penetration and catalyst dispersion in slurry reactors,⁹⁴ ensuring optimal photon absorption and reaction kinetics. Additionally, CFD aids in troubleshooting and system redesign, providing a cost-effective tool for scaling up processes.

A comparative discussion of these techniques in the context of conventional catalysis and photocatalysis is provided in Section 2.3, which also draws attention to the distinctions between these techniques when employed in the contexts of conventional catalysis and photocatalysis, where applicable.

While conventional computational techniques such as DFT, KMC, and CFD have been instrumental in modeling photocatalytic processes, they often require extensive computational resources and may struggle with complex, multivariable interactions. To address these challenges, AI-driven approaches, particularly ML-based models, are emerging as powerful complementary tools, offering data-driven insights that accelerate the discovery and optimization of photocatalysts and reaction conditions.

2.2.4 Bridging the gaps: AI and machine learning for advancing photocatalytic water splitting. From band gap engineering to reaction kinetics and stability, ML enables data-driven insights that facilitate the discovery, design, and optimization of photocatalysts and reaction conditions, ultimately enhancing hydrogen production efficiency (Fig. 2d). Common inefficiencies in photocatalytic processes, which can be mitigated through ML-driven techniques, are summarized in Fig. 3.

ML models accelerate materials discovery, reduce experimental costs, and enable real-time optimization of reaction conditions. In photocatalytic water splitting, ML techniques have been applied across various stages, from material screening to performance prediction. The most commonly used ML approaches include:

- Artificial neural networks (ANNs)-based
- Gaussian process regression (GPR)
- Support vector regression (SVR)
- Graph convolutional neural networks (GCNNs)
- Random forest (RF)
- Gradient boosting regression (GBR)

Detailed information about these methodologies can be found elsewhere.^{146–151} A concise overview of their applications in photocatalysis for water splitting is provided below, with further discussions on their implementation found in Sections 2.3 and 3.

ANNs are computational frameworks inspired by the neural structure of the human brain, consisting of multiple interconnected layers that process input data to recognize patterns and make predictions.¹⁵² In photocatalysis, ANN-based methods have been extensively used to predict band gap energies of

Table 1 A non-exhaustive overview of computational and machine learning modeling techniques in photocatalytic water splitting

| Technique | Purpose | Strengths | Limitations | Use cases in photocatalysis |
|--|---|---|--|---|
| DFT | Quantum mechanical modeling of electronic, optical, and photocatalyst properties, ⁷⁸ in addition to reaction mechanisms. ⁷⁹ | Provides fundamental insights into electronic structures, ⁸⁰ and reaction energetics. ⁸¹ | Computationally expensive for large-scale systems, e.g., large-diameter nanotubes. ⁸² | Used to study electronic band structures, ⁸³ doping and defect states, ⁸⁴ and reaction mechanisms in water splitting catalysts. |
| KMC | Simulating surface reactions and charge transfer kinetics over extended time scales. ⁸⁵ | Captures reaction dynamics over extended time scales (discrete space and continuous time) with stochastic accuracy. ⁸⁶ | Requires predefined reaction pathways, rate constants, and can be computationally demanding. ⁸⁶ | Models the kinetics of charge transport and recombination in semiconductor surfaces. ⁸⁷ |
| CFD | Simulating fluid dynamics, heat transfer, and light penetration in photocatalytic reactors. ⁸⁸⁻⁹⁰ | Accounts for fluid flow, temperature gradients, and light-matter interactions in reactors. ^{91,92} | High computational cost for complex reactor geometries. ⁹³ | Simulates the impact of reactor geometry and light intensity on hydrogen production. ⁹⁴ |
| ANN-Based Models (e.g., MLPNN, CNN, RNN) | Predicting nonlinear relationships between experimental conditions (e.g. catalyst size, temperature) and conversions; ⁹⁵ image-based analysis of photocatalytic materials. ⁹⁶ | Learns complex patterns from experimental data with high predictive power; ⁹⁷ excels at extracting spatial features from imaging and spectroscopy data. ⁹⁸ | Requires large, high-quality datasets for training; ⁹⁹ computationally expensive for deep architectures. ¹⁰⁰ | Predicts photocatalytic activity based on material descriptors (band gap, surface area, crystal structure); ¹⁰¹ analyzes SEM/TEM images to correlate material morphology with efficiency; ¹⁰² extracts spectral features from XRD, XPS, and UV-VIS data. ¹⁰³ |
| GPR | Probabilistic regression for physicochemical catalyst property predictions with uncertainty estimates. ¹⁰⁴ | Provides probabilistic predictions with confidence intervals; good for small datasets. ¹⁰⁵ | Scales poorly with large datasets, ¹⁰⁶ computationally expensive for large models. ¹⁰⁷ | Prediction of photocatalyst properties such as band gaps with uncertainty estimates. ¹⁰⁸ |
| SVR | Nonlinear regression modeling for catalyst efficiency and reaction rates. ^{109,110} | Efficient for small datasets and generalizes well. ¹⁰⁹ | Limited performance with complex, nonlinear datasets due to kernel mapping issues. ¹¹¹ | Identifies links between dopant characteristics and the performance of doped photocatalytic materials. ¹¹² |
| GNN | Predicting electronic structure and adsorption properties of materials using graph-based models. ^{113,114} | Well-suited for modeling atomic connectivity; captures spatial relationships in material networks, improving catalyst screening. ¹¹⁵ | Needs large training datasets and significant computational power due to high-dimensional node relationships. ¹¹⁶ | Used for high-throughput screening of novel possible photocatalysts with complex atomic structures. ¹¹⁷ |
| RF | Decision-tree-based modeling for identifying key features affecting catalytic performance. ¹¹⁸ | Handles high-dimensional data and identifies important variables in large datasets. ¹⁰⁹ | Prone to overfitting ¹¹⁹ and requires careful feature selection. ¹²⁰ | Identifies influential parameters affecting water splitting efficiency and predicts hydrogen production rates. ¹⁰¹ |
| GBR-based models | Ensemble of weak learners, e.g. DFT, RF, etc., for higher prediction power; ¹²¹ predictive modeling for catalytic activity and material properties. ¹²² | High accuracy, robust against overfitting, and effective for structured tabular data. ¹²¹ | Requires careful hyperparameter tuning, ¹²³ and may lack interpretability. ¹²⁴ | Predicting photocatalytic hydrogen evolution rates from reaction predictors. ¹²⁵ |
| Hybrid DFT+ML | Reducing the computational cost of DFT by using ML to predict band gaps, adsorption energies, and reaction barriers. ^{126,127} | Reduces computational cost of DFT by replacing expensive calculations with ML predictions; ¹²⁸ assists in accurately computing rate constants for catalytic events, including water chemisorption on potential materials. ¹²⁹ | Depends on the quality and quantity of training data from DFT calculations. ¹³⁰ | Accelerates material discovery by predicting electronic properties of new photocatalysts. ³¹ |
| Hybrid KMC+ML | Improving kinetic models by using ML to predict reaction rate constants, diffusion behavior, and charge transport. ¹³² | Improves the accuracy of kinetic models, and decreases computational cost. ¹³² | Limited by the availability of high-quality training data for kinetic modeling. | Can replicate experimental trends and provide insights into atomic orderings, as shown for electrochemical degradation. ¹³³ |
| Hybrid CFD+ML | Enhancing CFD reactor simulations by using ML to optimize continuous flow, radiation transfer, reaction kinetics, and transport properties. ¹³⁴ | Optimizes reactor design and operating conditions efficiently by integrating ML into CFD simulations. ¹³⁴ | Requires experimental validation for model accuracy and reliability. ¹³⁵ | Enhances photoreactor modeling by integrating ML for dynamic process optimization. ^{134,135} |

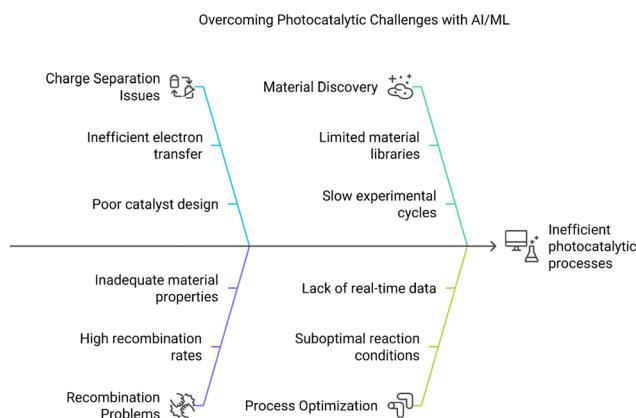


Fig. 3 Main photocatalytic challenges which may result in inefficient processes and suboptimal hydrogen production.

oxide photocatalysts based on structural properties,¹⁰⁹ as well as for modeling reaction kinetics to optimize hydrogen evolution rates.¹⁵³

Advanced ANN architectures, such as convolutional neural networks (CNNs), specialize in analyzing visual and spatial data through learnable filters that extract meaningful features.¹⁵⁴ These models have been used to identify structural defects in photocatalysts from electron microscopy images (*e.g.*, TEM, SEM),¹⁰² to assess porosity variations and spatial heterogeneity in catalyst supports using SEM image segmentation,¹⁵⁵ to classify photocatalyst morphologies using scanning probe microscopy (AFM, STM),¹⁵⁶ and to extract features from XRD and spectroscopy data to predict material properties.^{103,157}

GPR is a probabilistic ML method that provides predictions while also quantifying uncertainty.¹⁵⁸ It has been applied as a surrogate model to approximate high-fidelity CFD simulations for optimizing hydrogen production efficiency.¹³⁵ Additionally, GPR has been used as a proxy model for HER rate prediction and to identify critical data points for model training, improving overall predictive accuracy.¹⁵⁹

SVR, an extension of support vector machines (SVMs) for regression tasks, effectively captures nonlinear relationships in data.¹¹⁰ Recent applications of SVR are often hybrid approaches, such as the genetic algorithm-based support vector regression (GSVR) for band gap energy estimation,¹⁶⁰ and the hybrid genetic algorithm-based support vector regression (GABSVR) model for predicting the energy gap of strontium titanate.¹⁶¹

GCNNs are deep learning models designed for graph-structured data, commonly used in materials informatics.¹⁶² They can predict material properties by representing molecules and crystals as graphs, where atoms serve as nodes and bonds as edges.¹¹³ Examples include crystal GCNN (CGCNN) to encode molecular structures of metal oxide photocatalysts,¹⁶³ and general graph neural networks (GNNs) for predicting the bandgap of 2D-generated materials.¹⁵⁴

RF is an ensemble learning method that combines multiple decision trees to enhance predictive accuracy and remains robust even in the presence of missing values or multicollinearity.¹⁰⁹ A recent application of RF involved its integration into a stacking

model alongside LightGBM and XGBoost, improving predictive performance for TiO₂ photocatalytic water splitting.¹⁶⁴ Another study applied association rule mining to identify key factors influencing photocatalytic performance, using decision trees to develop selection rules and heuristics for high-performance catalysts. RF was then used to predict hydrogen production rates, demonstrating its effectiveness in data-driven material discovery.¹⁰¹

GBR builds strong predictive models by iteratively combining weaker models such as decision trees and RF.¹²¹ It minimizes errors sequentially, offering high accuracy in predictive tasks.¹⁶⁵ In photocatalysis, GBR-based methods such as LightGBM and XGBoost have been applied in stacking models for TiO₂ water splitting,¹⁶⁴ as well as in modeling energy levels of metal-oxide catalysts for hydrogen production.¹²⁶

Hybrid approaches, such as combining ML with DFT, have been explored to accelerate reaction rate calculations and improve the accuracy of reaction energy profiles, particularly in heterogeneous catalysis. An example is the use of adaptive multilevel splitting (AMS) with ML-based reaction coordinates to compute rate constants more accurately than conventional static methods.¹²⁹ This approach has been demonstrated for water chemisorption on γ -alumina, highlighting its potential in modeling catalytic reactions relevant to photocatalysis. More details on hybrid approaches, including the integration of DFT with ML, KMC with ML, and CFD with ML, are summarized in Table 1.

In fact, the integration of AI/ML with physics-driven models has emerged as a promising approach to address computational bottlenecks in photocatalysis simulations. One particularly relevant method is physics-informed neural networks (PINNs), which combine data-driven learning with physical laws to solve partial differential equations (PDEs) more efficiently.¹⁶⁶ PINNs have shown immense potential in modeling charge-carrier dynamics and solar energy collection, offering insights into complex multiscale phenomena.^{98,167}

In reactor-scale modeling, PINNs have been coupled with CFD simulations, enabling more efficient reactor optimization and system identification by incorporating reaction kinetics into the modeling framework.¹⁶⁸ This integration has significantly reduced computational costs while improving the predictive accuracy of fluid dynamics models.¹⁶⁹ A recent review details advances in AI-assisted CFD simulations, including CFD-PINN hybrid approaches for modeling chemical reactors.¹⁷⁰

At the atomic scale, physics-informed machine learning (PIML) complements DFT-based simulations by providing rapid approximations of band structures, adsorption energies, and reaction mechanisms.^{171,172} Instead of explicitly solving Schrödinger's equation, PIML can approximate quantum interactions with substantially lower computational cost.^{173,174}

Beyond PINNs, other physics-informed ML approaches have been explored for photocatalysis. Symbolic regression, for instance, has been used in perovskite design to derive mathematical relationships between input features (*e.g.*, oxidation state, ionic radii) and material properties such as formation energy and ionic mobility.¹⁷⁵ Additionally, symbolic regression

has enabled the automated design of hybrid descriptors for halide perovskites, improving the prediction of their photoelectrochemical properties.¹⁷⁶

These methods are particularly appealing because they reduce computational costs while preserving physical interpretability. However, several challenges remain, including the enforcement of physical constraints, handling of sparse experimental datasets, inference of dynamic systems (*e.g.*, chaotic, turbulent), and, in some cases, the computational demands of hybrid models.¹⁷⁷

2.3 Photocatalysis *versus* general catalysts

The use of light as an energy source to promote electron–hole pairs that initiate redox reactions distinguishes photocatalysis as a unique catalysis process. This leads to significant differences between photocatalysis and general catalysis in terms of underlying mechanisms, material design, and applications. For example, conventional thermal or enzymatic catalysis does not require the consideration of parameters related to light interaction and its effects; therefore, these parameters are not considered in these traditional catalytic methods.

There are mainly six key differences between photocatalysis and other catalytic processes: energy source, catalyst activation mechanism, material design, reactor design, reaction pathways and selectivity, and environmental impact (Fig. 4). In addition, we briefly describe the use of the traditional mathematical approaches and compare them to ML methods for both photocatalysis and other catalytic processes, highlighting similarities and differences.

2.3.1 Energy source. This is a critical factor in catalyst activation. Photocatalysis relies predominantly on UV or visible light, contrasting with conventional catalysis, which uses thermal energy or specific chemical activators, such as metal complexes

or acid–base catalysts. This distinction makes photocatalysis appealing due to its ability to operate at ambient temperatures and pressures,¹⁷⁸ reducing energy consumption and improving sustainability.

In the context of mathematical modeling, photocatalysis introduces unique challenges due to the dependence on light as the energy source. Traditional mathematical modeling approaches in photocatalysis range from simple methods, such as using fixed light intensities and assuming homogeneous light distribution, to more complex models like those used to solve the radiative transfer equations (RTE).^{179–181} More rigorous methods include CFD,^{94,182} which simulates detailed light distribution and reaction kinetics but at a moderate to high computational cost, depending on the settings.

Similarly, in general catalysis, simple approaches often use Arrhenius-based kinetics or empirical rate constants, while advanced models incorporate reaction-diffusion equations or dynamic simulations under varying conditions.^{183–185} CFD is also widely used to predict reaction kinetics and energy transfer in general catalysis.¹⁸⁶ By leveraging advanced CFD software, it is feasible to consider energy sources such as electric heating and microwave heating, facilitating the establishment of a connection between the heat source, the reaction site, and the distribution of heat over the reactor volume. However, as previously mentioned, its computational cost remains a significant constraint.

ML, however, offers similar advantages for both photocatalysis and general catalysis as regards the energy sources. In the context of photocatalysis, a promising approach entails the integration of ML with tools such as CFD and ray-tracing simulation to generate precise predictions.¹³⁴ This integration facilitates the convergence of theory-driven methodologies with ML, thereby leading to a substantial reduction in computational

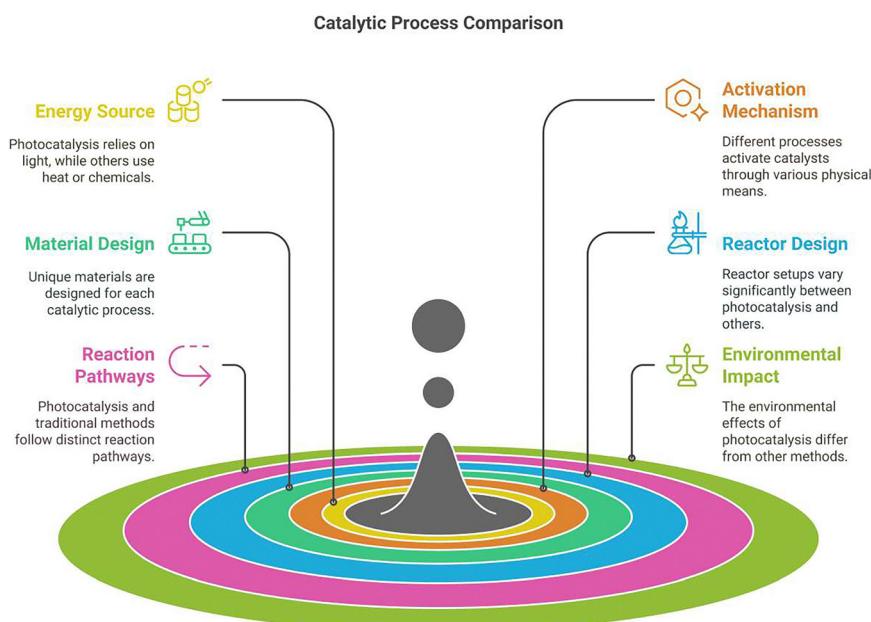


Fig. 4 Six key differences between photocatalytic and general catalytic processes.

expenses. Additionally, ML possesses the capacity to assimilate spectral data and photon flux distributions with other features, enabling the prediction of rates.¹⁵⁹ This capability extends to the assessment of the significance of photon flux in conjunction with other features of relevance. This level of complexity is difficult to achieve with theory-driven mathematical methods alone.

Similarly, ML can be used in general catalysis,¹⁸⁷ by leveraging CFD simulations and consider reactions conditions such as temperature and pressure, and also in assisting the identification of activation energy relationships. A notable application of ML in this context involves the prediction of activation energies in gas-phase reactions. This is achieved by integrating thermodynamic quantities, topological indices, and molecular fingerprints.¹⁸⁸ Another example is the utilization of ML to model the activation energy of biomass pyrolysis, considering biomass properties and reaction degree.¹⁸⁹

For both photocatalysis and general catalysis, ML algorithms can use large datasets of experimental results to uncover patterns in (photo)catalytic efficiency under varying experimental conditions, significantly reducing the need for exhaustive CFD simulations. The energy sources directly impact the catalyst activation mechanism.

2.3.2 Catalyst activation mechanism. In the photocatalytic process, photons generated by lamps or sunlight are absorbed by a semiconductor material, generating excited electron–hole pairs. These charge carriers migrate to the surface of the catalyst, where they promote reduction and oxidation reactions. In contrast, thermal catalysis entails the transfer of energy through collisions and the activation of reactants by thermal motion.

In the field of photocatalysis, traditional mathematical models are frequently used to simulate the dynamics of charge carriers. These models are based on quantum mechanics and include approaches such as DFT.¹⁹⁰ DFT is also a widely applied in the study of thermal catalysis. However, collision-based kinetic models are often employed as a less complex alternative, frequently relying on techniques such as transition-state theory.¹⁹¹ ML has been demonstrated to enhance photocatalytic studies by reducing computational demand and predicting charge recombination phenomena,¹⁹² as well as simulating strategies to enhance carrier lifetimes across diverse material sets.¹⁹³

In general catalysis, ML is frequently employed in conjunction with DFT, such as in taking intrinsic electronic characteristics of single metal atoms and the substrates as input to accurately predict the catalytic activity of single-atom catalysts for nitrogen reduction reaction.¹⁹⁴ This approach is validated by available experimental data and independent DFT computations. The understanding of the catalyst activation mechanism serves as basis for the design of the photocatalyst for given application, such as hydrogen production.

2.3.3 Material design. Photocatalytic materials must exhibit specific optical and electronic properties to efficiently absorb and utilize light energy. These properties include an appropriate band gap, typically ranging from 1.5 to 3.0 eV,^{195,196} depending on the desired application. Band gaps of less than

1.23 eV can also be used under applied potential, as in the context of photoelectrochemical water splitting.¹⁹⁶ While photocatalysis and thermal catalysis share common material design principles—such as high surface area, the presence of active sites, and thermal stability—photocatalysts must also optimize charge carrier generation, separation, and transport to minimize recombination losses.

Unlike thermal catalysts, where heat transfer properties dominate performance considerations, photocatalysts require precise control over their optical absorption and electronic structure to efficiently drive photoinduced reactions. Additionally, emerging fields like plasmonic photocatalysis¹⁹⁷ and photothermal catalysis¹⁹⁸ are hybrid and complex mechanisms that combine light absorption and thermal effects, demonstrating enhanced activity.

In the fields of catalysis and photocatalysis, material design is typically informed by prior investigations, followed by a trial-and-error approach to problem-solving. These processes are often time-consuming and costly, necessitating the selection of specific conditions based on factors such as availability, cost, and the novelty of the approach. Typically, this involves the improvement of existing (photo)catalysts. Traditional simulations, such as DFT, are widely employed to provide insights into tailored materials, which can then be validated. The primary applications of ML in material design include its use as surrogate models,^{199,200} which has the advantage of reducing computational cost and enabling the use of more realistic models by theory-driven approaches. Additionally, ML can predict properties such as band gap energies, charge carrier properties, and light absorption from structural data.^{109,192,201}

The analysis of extensive datasets using ML has been shown to unveil latent patterns and correlations that can inform the design of efficient photocatalysts. Additionally, these findings can guide the design of the entire system reaction, including the reactor.

2.3.4 Reactor design. Most of the literature has focused predominantly on catalyst design, with comparatively less attention devoted to the design of reactors. However, given the necessity of scaling up systems in the face of photocatalytic systems, which require efficient light penetration and uniform illumination, this level of design is of crucial importance. Recent publications have emphasized the significance of system design and performance, reviewing notable contributions from researchers in the field.²⁰² The application of these systems at large scales has also been explored.²⁰³ The most prevalent system types include flat-plate reactors,^{204,205} cylindrical with compound parabolic concentrators (CPCs),^{53,181,206} slurry systems,²⁰⁷ and optical fiber reactors.²⁰⁸

A significant distinction between thermal catalytic systems and photocatalysis pertains to the emphasis placed on temperature and pressure control in the former, in contrast to the latter's prioritization of light distribution. Conventionally, mathematical models in photochemistry employ radiative transfer equations to incorporate the radiation field into the modeling process,¹⁷⁹ while fluid dynamics is utilized to optimize light distribution.⁹⁴ ML-driven simulations augment these

models by predicting optimal geometries and mixing patterns based on experimental data,^{97,134} thereby reducing the necessity for trial-and-error iterations.

In the context of conventional catalytic reactors, where light does not constitute a factor in reactor design, the primary focus of ML applications is on optimizing flow dynamics and reaction kinetics.^{209–211} By integrating the reactor configuration into the mathematical modeling framework, the rate constants can be estimated independently of the system and can be extrapolated to other scenarios. For both general catalysis and photocatalysis, studies on reaction pathways and selectivity are important aspects in research.

2.3.5 Reaction pathways and selectivity. In the domain of photocatalysis, reaction pathways are influenced by the interaction of charge carriers with adsorbed species on the catalyst surface, a phenomenon referred to as the photoinduced pathway. This interaction frequently results in distinctive selectivity patterns and the formation of radical intermediates, which are not typically observed in thermal catalytic processes.

Photocatalysis functions independently of thermal activation, exhibiting distinct reaction steps: (1) light absorption, (2) charge separation, (3) surface reactions – electrons (e^-) that reduce reactants and holes (h^+) that oxidize reactants, (4) product formation and charge recombination. Reaction pathways in photocatalysis must then consider all these distinct reaction steps, the media in which photocatalysis will occur (e.g., water, organic compounds), and rigorous consideration of the radiation field.

In contrast, general catalysis is governed by adsorption/desorption and the formation of intermediates on the catalyst surface. The reaction mechanisms are typically elucidated through the application of Langmuir–Hinshelwood-derived models or Eley–Rideal models,²¹² which are widely recognized as the prevailing mechanisms in heterogeneous catalysis. The choice of model depends primarily on the interaction of reactants with the catalyst surface, including the density and availability of active surface sites.²¹³ To facilitate catalysis, the catalyst must possess active sites, where bond breaking and bond formation reactions occur. In both cases, pseudo-order equations are frequently used,^{214–218} and DFT is commonly employed when reaction pathways in atomic scale are needed.^{219,220} ML has been used in reaction pathway studies to uncover hidden reaction mechanisms as a result of its ability in identifying key descriptors.²²¹ In photocatalysis, ML is especially important in the prediction of charge recombination rates and optimizing co-catalysts for selectivity control.^{193,222}

Regarding selectivity, in photocatalysis it is influenced by band gap energy, electron–hole recombination rates and surface modifications.^{223,224} Photon energy is vital in determining which reactions will occur, because it directly affects the redox potential of electron–hole pairs. The radical-driven pathways, such as hydroxyl radicals, introduce non-selective oxidation, especially in the presence of organic compounds.²²⁵ Selectivity towards water splitting mainly depends on electron density, co-catalyst, and light intensity.²²⁶ The control over radical generation and selective photocatalytic transformation are

crucial aspects in the photocatalytic biomass conversion,²²⁷ such as the solar hydrogen generation from lignocellulose.²²⁸

In general catalysis, selectivity is controlled by catalyst composition, active site geometry, reaction temperature, and pressure.^{229,230} Reaction conditions such as temperature and reactant ratio influence the adsorption energy and intermediate stability, which determine the final product distribution.²³¹ An example is CO₂ hydrogenation, where different temperatures, pressures, and catalysts favour the production of several products, including C1 chemicals (CH₄, CO, CH₃OH), gasolines, aromatics, and olefins.^{232,233} On one hand, general catalysis has well-established mechanisms, allowing to better predict selectivity. On the other hand, photocatalysis allows unique selectivity patterns due to excited-state chemistry.²³⁴

Conventional mathematical models for selectivity are rooted in analysis of variance,²³⁵ reaction network analysis,²³⁶ KMC simulations,²³⁷ and molecular dynamics.²³⁸ ML has been demonstrated to effectively identify optimal reaction conditions to enhance selectivity and photocatalytic activity. This can be achieved by systematically gathering data from experimental parameters and reaction results. Additionally, it can facilitate the interpretation of relationships such as vibrational spectral signals and catalytic performances, leading to the discovery of hidden physical data.²³⁹ This, in turn, aids in the design of more targeted (photo)catalysts and multifunctional catalysts for both thermal and photocatalytic reactions.

In addition to the search for high-performance and optimized photocatalysts and system conditions, the environmental impact is a high priority when choosing materials, from photocatalyst synthesis to real application scenarios.

2.3.6 Environmental impact. Photocatalysis offers a green alternative for many reactions, as it often uses sunlight as an abundant and renewable energy source. Thermal catalysis, particularly in industrial applications, frequently relies on fossil fuels, contributing to carbon emissions. This sustainability aspect has driven interest in photocatalysis for hydrogen generation and environmental remediation. Mathematical models for environmental impact typically rely on life cycle assessments (LCA).^{240–242}

Recently, ML-based life cycle optimization studies have been proposed.^{243,244} These investigations have yielded significant insights into the most substantial environmental impacts. For example, they have identified catalyst synthesis as the most impactful activity due to its substantial energy demand.²⁴³ Additionally, the production of nickel nitrate for the carbon dioxide methanation process has been identified as a high-impact activity.

The environmental risks associated with these activities, such as climate change, fossil depletion, and photochemical oxidant formation, can be systematically enumerated. These risks can then be integrated into an ANN with a genetic algorithm, which can optimize the process performance. This approach involves a decision-making process that utilizes the insights derived from the environmental risk analysis. Another example of the use of ML is in techno-economic assessments, considering bio-oil yield from a pyrolysis process, coupled with tools such as Aspen Plus to include material and energy

balances, Monte Carlo simulation for uncertainty calculations, and life cycle assessment.²⁴⁴

ML has been demonstrated to enhance the process of life cycle analysis by integrating performance metrics, including energy efficiency and carbon footprint, with life cycle data. This integration enables the prediction of overall sustainability. In conventional systems, environmental impacts are often calculated based on standardized emission factors. This approach has limited ability to adapt to specific conditions or materials. In the subsequent sections, the advances of AI and ML in photocatalysis, with a particular emphasis on hydrogen production, are examined in greater detail.

3. Advances of artificial intelligence in photocatalysis

3.1 Bibliometric analysis of the utilization of artificial intelligence in photocatalysis for hydrogen production

While AI and its subgroup ML have been extensively employed in various fields, including photocatalysis of organic pollutants, their use in studies of photocatalysis for hydrogen production remains in its infancy.

Fig. 5a illustrates a notable surge in the publication of articles in this field, with 23 articles published in 2024. This figure stands in stark contrast to the nine publications recorded in 2023, signifying a notable surge in research activity. The 23 articles published in 2024 constitute a substantial proportion of the total published articles in this search, amounting to 74 articles. This suggests that researchers have recently begun utilizing these tools to facilitate their research on this subject.

As shown in Fig. 5b, the term “machine learning” emerged as a trend topic in mid-2022, with a frequency of 49. However, the concept of ANN predates this, having appeared as early as 2009, with a frequency of 38.

As depicted in Fig. 5c, the keywords have been divided into three clusters. The first cluster (blue) consists of keywords associated with analytical equipment and characteristics of materials, including “Fourier transform infrared spectroscopy” and “particle size”.

The second cluster (red) is predominantly associated with experimental conditions and the system in general, encompassing “ultraviolet radiation”, “pH”, “initial concentration”, and also words related to experimental design, such as “response surface methodology” and “optimization”. This is logical, given that experimental data are used as inputs for experimental designs and traditional modeling. It is noteworthy that studies have combined the photocatalytic degradation of pollutants with hydrogen production, indicating that hydrogen production does not necessitate the use of clean water.

The third cluster (green) encompasses “machine learning”, “hydrogen production”, “controlled study”, and associated terminology. The integration of “machine learning” with “synthesis”, “hydrogen production” and “hydrogen evolution” is emphasized, underscoring the significance of ML in deriving enhanced insights from the synthesis of novel materials, as well as

in understanding outputs of interest, such as the capability of the systems in generating hydrogen.

3.2 Deployment of the utilization of artificial intelligence in photocatalysis for hydrogen production

As demonstrated in Fig. 6, a proposed sequence involves the application of AI tools for data generation, with the objective of identifying novel photocatalytic materials or enhanced photocatalytic systems for water splitting. Elaborated flowcharts are available in recent applications of ML in this domain.^{97,109,130,245}

The initial step in virtually all research is to obtain data. It is widely acknowledged that a number of high-performance ML algorithms have a voracious appetite for data,²⁴⁶ necessitating not only a vast quantity but also a high level of data quality. For this reason, in-house data may prove insufficient when dealing with highly complex systems and data-driven approaches. To address this challenge, research has been conducted on combining in-house experiments—which may yield a limited number of data points but of high quality—with databases and the use of scraping techniques combined with natural language processing (NLP) to acquire a sufficient number of data points and achieve more accurate predictions. We start by highlighting the new approaches researchers have been producing regarding the obtainment of data.

3.2.1 Data collection, feature collection, and selection. A variety of methods have been employed to obtain data from the literature. As illustrated in Fig. 7, these methods can be classified into the following categories: secondary data collection (literature-based data and existing online databases) and high-throughput experimentation platforms. Common challenges include financial barriers, such as extensive experimentation and synthesis of expensive photocatalysts, and skill shortages when using tools such as ML for faster advancement of the field.

One avenue for data acquisition is through online sources such as the Royal Society of Chemistry (RSC), Elsevier, Springer, and many others. By scanning scientific papers on photoelectrochemical water splitting, researchers were able to obtain 10 560 data points from almost six hundred experiments from the consulted references, which were then subjected to ML analysis.²⁴⁵ The authors obtained several data points for these systems, *e.g.*, regarding the materials (BiVO_4 , Bi_2WO_6 , and others), layers of doping (B, C, Cl, Co, Ga, *etc.*), co-catalysts (Au, Cds, CePi, *etc.*), synthesis methods, calcination times and temperatures, among others, were identified as the variables to be considered. As the output variables, the band gap (eV) of the working electrode and the photocurrent density (mA cm^{-2}) were selected. It is unclear from the methodology section how the authors extracted the data, whether manually or automatically.

Similarly, Yan *et al.*¹³⁰ consulted 106 research papers by searching for the photocatalyst: elemental doped graphitic carbon nitride ($\text{D-g-C}_3\text{H}_4$) for photocatalytic hydrogen generation. Again, papers from known publishers were sought, resulting in 767 observations for further ML application. The data was extracted manually, with a particular focus on the experimental

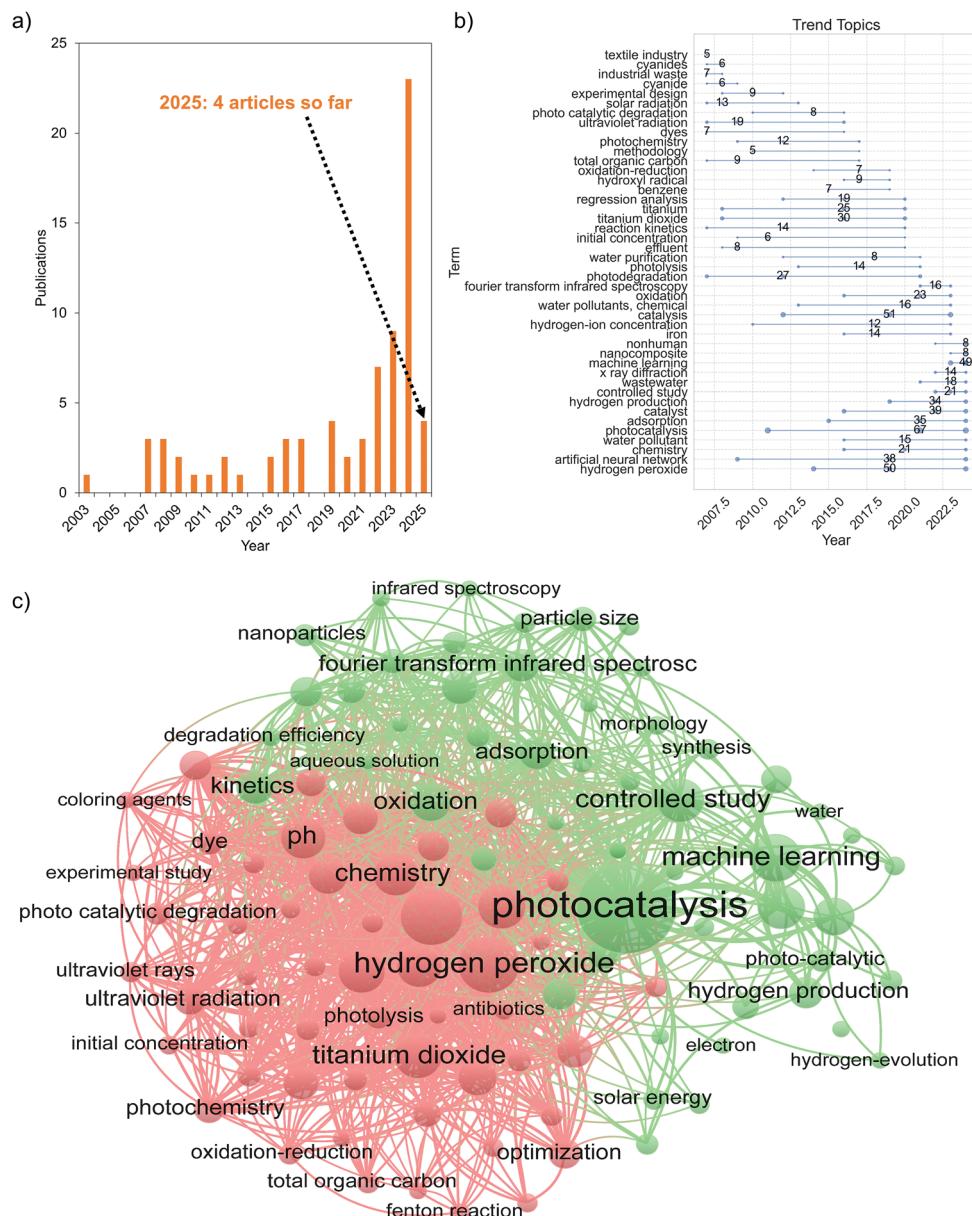


Fig. 5 Bibliometric data for “photocatalysis” AND (“machine learning” OR “artificial intelligence” OR “neural network” OR “regression”) AND “hydrogen”. (a) Annual scientific production, (b) recent trend topics from the first publication. The results were obtained using the bibliometrix package in RStudio. (c) Co-occurrence network with all keywords (clustering algorithm: Louvain-based, normalization: association, number of nodes: 105, minimum cluster size: 25, minimum number of occurrences of a keyword: 5), obtained by using the VOSviewer software.¹⁷ Data were collected via the Scopus document search in February 2025 (language: English, document type: articles). Data were collected via the Scopus document search in February 2025 (language: English, document type: articles).

conditions related to hydrogen production, including the light source and catalyst concentration. Parameters related to the synthesis of the material, such as the precursors, dopants and their mass ratios, were also considered. The authors identified the band gap and specific surface area as properties of interest and evaluated the performance of the systems by measuring the hydrogen generation rates.

However, manually constructing datasets can be time-consuming, tedious, and prone to errors such as typos. In addition, much data is presented in static figures, requiring

the use of tools such as the Plot Digitizer (<https://apps.automeris.io/wpd/>). As a result, researchers are typically limited to collections of a few hundred articles at most. The automated construction of datasets has emerged as a topic of interest in the scientific community across a range of fields.²⁴⁷ The high-throughput extraction of chemical information has demonstrated effectiveness when applied to texts and figures, with the use of LLMs,²⁴⁸ including generative pre-trained transformers (GPTs),²⁴⁹ and optical character recognition (OCR)-based technologies, such as the optical chemical structure recognition (OCSR).²⁴⁷

AI-Driven Photocatalysis Laboratory Process

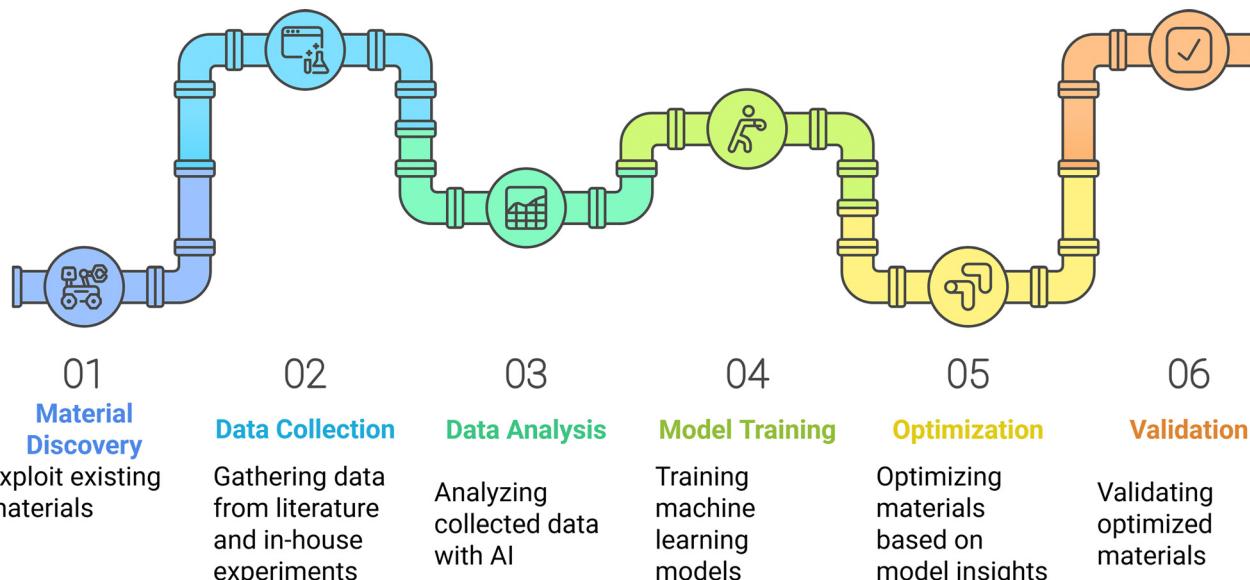


Fig. 6 A possible sequence to advance photocatalysis research with machine learning approaches- from data collection to enhanced photocatalysts and systems for further implementation.

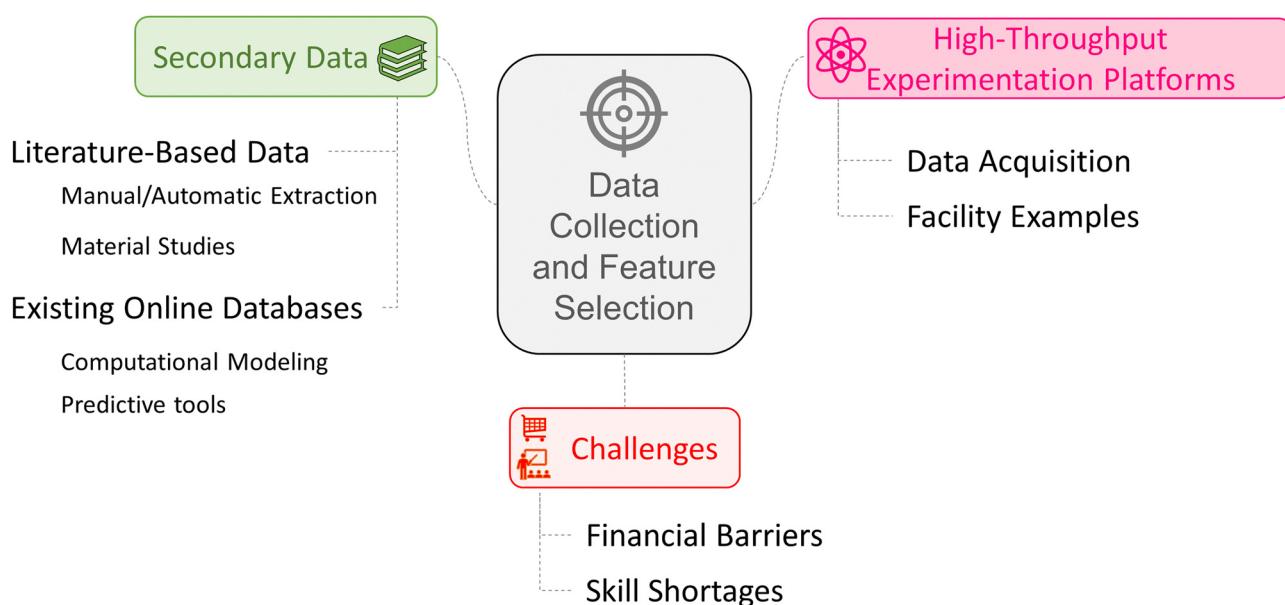


Fig. 7 Actual characteristics of data collection and feature selection in the generation of hydrogen via water splitting through photocatalysis.

Isazawa and Cole²⁵⁰ in a recent publication demonstrated the step-by-step process of developing an extraction tool related to the use of photocatalysis for water splitting applications. The authors presented the newer version of the ChemDataExtractor toolkit.²⁵¹ In contrast to the construction of datasets comprising a considerable number of DFT results, as observed in various available datasets such as the Materials Project (<https://next-gen.materialsproject.org/>), the authors focused on the collection of pivotal photocatalytic reaction attributes

and the measurement of catalytic efficiency. Their work presents detailed information regarding the attributes and relations of photocatalytic activity, the attributes of the compounds themselves, and also information about the light sources. To update the aforementioned toolkit, the authors collected information from several documents in XML/HTML format and used a document reader integrated with the ChemDataExtractor. The workflow included filtering, tokenization, and clause extraction to represent photocatalytic knowledge. They introduced the

dependency-based information extraction (DepIE) algorithm to generate dependency graphs with root phrases and performed a merging step to compile the photocatalyst record. The authors provide a detailed account of the clause extraction process and the use of DepIE.

The advent of automated systems, particularly those employed in the design of catalysts, represents a significant technological advancement. These automated tools are optimized for high-throughput *in silico* exploration, amassing a greater quantity of data and enabling the implementation of data-driven methodologies for predictive purposes.^{252,253} This approach combines the acquisition of in-house experiments with ease in organizing and collecting the target parameters of interest in a greater amount of data, facilitating the application of data-driven approaches. Notable examples of high-throughput experimentation include the UPenn HTE Center (<https://web.sas.upenn.edu/hte-center/>) at the University of Pennsylvania, the CEA-SCBM HTE platform at the CEA Joillet (https://joliot.cea.fr/drj/joliot/en/Pages/research_entities/medicines_healthcare_technologies/scbm/lcb/HTE.aspx), and REALCAT (<https://www.realcat.fr/en/home/>). Nevertheless, the financial investment required to implement and maintain automated systems may prove prohibitive for the majority of laboratories and research centers at this time. Additionally, concerns may arise regarding the increased use of materials and the potential shortage of qualified professionals. To equip professionals with the requisite skills to utilize sophisticated tools for the advancement of (photo)catalysis, it is necessary to combine chemical-related backgrounds with data science and analytics expertise.

One straightforward method for incorporating data into experiments is to obtain data from existing online datasets. A plethora of information sources are available, including the <https://catalysis-hub.org> (<https://www.catalysis-hub.org/>), Materials Project (<https://next-gen.materialsproject.org/>), NOMAD (<https://nomad-lab.eu/prod/v1/gui/about/information>), JARVIS (<https://jarvis.nist.gov>), and AFLOW (<https://aflowlib.org>). The Catalysis Hub offers data and software for computational catalysis research, including thousands of reaction energies and barriers derived from DFT calculations on surface systems. The Materials Project provides computed information on known and predicted materials, containing information on composition, electronic and optical properties, structural properties, and thermodynamics. It is widely used for studying photocatalytic materials due to its extensive band structure and absorption spectra data. The NOMAD has comprehensive datasets on materials, offering divisions in theory, experiment, and AI-driven applications. It contains specific information on solar cells (especially perovskites), MOFs, and charge transport properties. Additionally, NOMAD provides insights into charge carrier dynamics, adsorption properties, and optical responses.

JARVIS and AFLOW are also prominent databases that provide electronic structure and materials property datasets, including band gap predictions and charge carrier mobilities. Additionally, RDKit (<https://www.rdkit.org>), Open Babel (<https://openbabel.org/index.html>), and the Chemistry Development Kit (<https://cdk.github.io>) are of considerable utility. These three

tools exhibit numerous similarities, as they are employed for cheminformatics, enabling computational chemistry and predictive modeling. They possess a multitude of features and are predominantly utilized due to their comprehensive molecular functionalities.

Further, resources such as SpringerMaterials (<https://materials.springer.com>) and the Computational Materials Repository (<https://cmr.fysik.dtu.dk>) contain experimental and computational data on absorption properties and optical responses. The Cambridge Structural Database (<https://www.ccdc.cam.ac.uk>) and ICSD (<https://icsd.fiz-karlsruhe.de>) provide crystallographic information that can be used to model band alignments and charge transport mechanisms.

The fulfilment of data requirements can be accomplished through the utilization of application programming interfaces (APIs). While these tools are not specific for photocatalysis, they can be very useful when the target properties are adequately searched for and the main goal is to increase efficiency in hydrogen production, including *via* water splitting.

The necessity of high data quality and sufficient diversity for optimal prediction performance across various catalytic systems can also be addressed by leveraging the utilization of handbooks.²⁵⁴ The implementation of best practice procedures for data collection and measurement is imperative for maximizing the potential of ML techniques and can be particularly advantageous in in-house and high-throughput experimentation settings.²⁵⁵

3.2.2 Model – complex relationships, and selection. Following the collection and organization of data in a manner suitable to ML applications, algorithms trained on experimental and theoretical data can identify correlations between structural features and photocatalytic performance, thereby guiding the synthesis of efficient and stable materials. ML algorithms, such as RF or ANN, can model complex relationships between these parameters to optimize reaction conditions for maximum hydrogen yield. Optimizing reaction conditions represents another area of interest when applying ML in these systems.

Tables 2 and 3 provide up-to-date overviews of the current applications of ML in the field of photocatalysis, demonstrating the great prospective for exploiting these tools in promoting advancement in this field. It reveals that literature has focused on two key areas: (1) discovery of materials – finding high-performing photocatalysts by predicting their properties (Tables 2), (2) discovery of knowledge – identifying descriptors that influence the system (Table 3).

ML has been widely used for screening materials relevant to photocatalytic water splitting for hydrogen generation.^{15,126,131,256,257} Recently, multi-fidelity ML (MFML) was applied to identify halide perovskites with STHEs exceeding 20%.²⁵⁶ The study built on a prior database of ~800 ABX₃ compounds, combining band gaps, photovoltaic efficiencies, and decomposition energies from DFT and literature.²⁶² Regularized greedy forest (RGF) was used to train predictive models for decomposition energy and band gap, with a 90-10 train-test split, 5-fold cross-validation, and GridSearchCV for hyperparameter tuning. ML-DFT integration identified thousands of potential photocatalysts,

Table 2 A summary of ML applications in the discovery of materials in photocatalysis for water splitting and hydrogen production. Batch size = BS, learning rate = LR, number of hidden neuron unit = NHNU, number of hidden layers = NHL, number of estimators = nest

| Goal | Specific ML application | Size of source data | Source | ML techniques | Validation approach | Optimization tools | Hyperparameter optimized values | Performance metrics | Uncertainty quantification | Main results from ML application | Ref. |
|--|--|---------------------|--|---|---|--------------------|--|---|---|--|------|
| Screen novel halide perovskites for photocatalytic water splitting | To identify novel lead-free halide perovskites with high STH efficiency ($\eta_{\text{STH}} > 20\%$) | ~1000 | Computational (DFT-based, multi-fidelity dataset) | MFM1, RGF | 5-fold cross-validation; 9:1 train-test split | GridSearchCV | — | RMSE | Prediction uncertainty quantified via standard deviation over 4000 individual runs | Identified hundreds of stable, band gap-appropriate halide perovskites for photocatalytic water treatment | 256 |
| Screen efficient metal-oxide catalysts for photocatalytic water splitting | To rapid screen promising metal oxide photocatalysts for hydrogen generation | ~860 | Computational (DFT-based) – Material Explore website | XGBoost, LSTM, lightGBM, GBR | 8:2 train-test split | — | — | MSE, RMSE, R^2 | Selected 10 photocatalysts over 860 metal oxides within seconds based on the established screening model | Selected 10 photocatalysts over 860 metal oxides within seconds based on the established screening model | 126 |
| Screen 2D materials for photocatalytic water splitting | To ascertain the degree of uncertainty associated with G_0W_0 band gap | 316 505 | Computational (V2DB database) | ANN, RF, XGBoost | 10-fold cross-validation | GridSearchCV | ANN: ReLU activation, NHL (400, 10), $\alpha = 0.1$, max iterations = 100; RF: 1000 estimators, default settings; XGBoost: max depth = 10, LR = 0.3, min child weight = 1, gamma = 0, colsample by tree = 0.7 | MAE, RMSE | ML predicted band gap uncertainty ($E_g^{\text{unc(GW)}}$) used as a filtering criterion to select high-confidence candidates | Assisted in finding 11 promising 2D photocatalysts | 131 |
| Reveal the potential of MOFs for the overall water splitting | ML with high-throughput computations: identify MOFs with OWS capability | 20 000+ | Computational (quantum-MOF database) | CGCNN | 8:1:1 train-validation-test split | — | Convergence around 200th step (≈ 0.05 eV) | MAE, MSE | Not explicitly mentioned | Identified 14 MOFs from the QMOF database ($> 20\ 000$ MOFs) for visible-light-driven OWS | 15 |
| Identify 2D MoSi_2N_4 that exhibit enhanced piezoelectricity, high thermal conductivity, and stiffness | Examine the flexoelectric and piezoelectric properties of complex structures | Not specified | Computational (DFT-based) | MLIP | — | — | — | — | Statistical error of predictions based on AIMD-derived data variation and strain-dependent polarization modeling | Enabled the examination of flexoelectric and piezoelectric properties of complex structures | 257 |
| Screen surface functionalized monolayer SiC for photocatalytic water splitting and UV applications | Predict the stability, band gap, and effective mass of 4624 functionalized monolayer SiC | ~4900 | Computational (DFT-based) | PLSR, LASSO, DT, RF, SVR, GPR, SVM, NuSVM, ET, RidgeCV, LassoCV, XGBoost, Bagging, RF, AdaBoost, MLFF | 8:2 train-test split | GridSearchCV | 1000 iterations | MAE, RMSE, R^2 , AUC, ROC, confusion matrix | Errors were kept within < 0.25 eV | Predicted the properties (E_f , E_g , CBM, VBM, and m^*) of over 4000 surface-functionalized monolayer SiC materials | 258 |

Table 2 (continued)

| Goal | Specific ML application | Size of source data | Source | ML techniques | Validation approach | Optimization tools | Hyperparameter optimized values | Performance metrics | Uncertainty quantification | Main results from ML application | Ref. |
|--|---|--------------------------|--|-----------------------------|--|--------------------|---------------------------------|--------------------------|--|----------------------------------|------|
| Investigate WS ₂ N, WG ₂ N ₄ , and WSiGeN ₄ monolayers as photocatalysts for hydrogen production | To assist AIMD simulations in validating the thermal stability of the monolayers at 300 K | Not explicitly mentioned | DFT calculations (VASP) + AIMD + ML force fields | — | — | — | Bayesian error + RMSE | Bayesian error | AIMD-MLFF calculations indicate that these mono-layers exhibit thermal, mechanical and dynamic stabilities | AIMD-MLFF | 126 |
| Predict the type and band gap for photocatalytic g-GaN-based vdW heterojunction applied in the field of photocatalysis | To screen for g-GaN based vdW heterojunction applied in the field of photocatalysis | ~1000 | Computational (2D material database) | SVM, Ada-boost, RF, XGBoost | 10-Fold cross-validation; 90:10 train-test split | — | MAE, RSME, AUC, ROC | Not explicitly mentioned | Predicted types and energy bands of the g-GaN heterostructures | g-GaN | 260 |

primarily FA-based iodides or Cs-based bromides with divalent cations (groups II or IV) at the B-site.

In their study, Zhou *et al.*¹²⁶ employed a range of ML algorithms, including extreme gradient boosting (XGBoost), long short-term memory (LSTM), LightGBM, and GBR, to facilitate and accelerate the screening of efficient metal oxide photocatalysts. By utilizing the Material Explore website,²⁶³ the researchers collected data pertaining to the composition, space group, and theoretically calculated the VBM and CBM. Through the application of various ML models, they were able to identify ten theoretical metal oxide candidates with promising characteristics. Of these, CsYO₂ was selected as the most appropriate, and its effectiveness was further validated *via* DFT calculations. By employing the XGBoost regression algorithm, the authors developed a high-performing energy level prediction model.

Wang *et al.*¹³¹ screened 316 505 2D materials for photocatalytic water splitting using a workflow combining ML predictions and DFT calculations. Initial screening based on band gaps, valence and conduction band positions, and uncertainty yielded 11 top candidates after filtering through structure optimization, phonon dispersion, *ab initio* molecular dynamics (AIMD), and G₀W₀ calculations.

The band positions are depicted in Fig. 8a, where the black lines represent the reduction potential of H⁺/H₂ and O₂/H₂O, and the white lines represent the band gaps from 1.23 to 3 eV. ML models (ANN, RF, XGBoost) guided the process, followed by hybrid-DFT (HSE06) and G₀W₀ evaluations. All 11 candidates had G₀W₀ band gaps exceeding 1.23 eV, with three (SnSSe, TiFCl, WTeO) under 3 eV, indicating visible light photocatalysis potential. Nine materials exhibited band edges suitable for water redox reactions, while two were limited to oxygen evolution.

Wang *et al.*¹⁵ evaluated 20 375 MOFs for photocatalytic water splitting using data from the quantum-MOF (QMOF) database. A crystal graph convolutional neural networks (CGCNN)-based ML model trained on 10 810 MOFs with known band gaps (HSE06) was used to predict band gaps for 9565 additional MOFs. Filters applied to band gap range (1.23–3.0 eV), stability, LPD (>5.0 Å), and electronic/optical properties identified 14 promising candidates for visible-light-driven water splitting. As shown in Fig. 8b, five MOFs showed superior electronic and optical performance compared to well-known MOF photocatalysts like UIO66(Zr)-NH₂ and MIL125(Ti)-NH₂.

Mortazavi *et al.*²⁵⁷ used ML interatomic potentials (MLIP) trained on 2000 AIMD time steps to study the flexoelectric and piezoelectric properties of MA₂Z₄ (M = Cr, Mo, W; A = Si, Ge; Z = N, P) monolayers. Structural, electronic, and energetic properties were calculated for 2H- and 1T-phase nanosheets, identifying MoSi₂N₄, WG₂N₄, and WS₂N₄ as promising candidates for photocatalytic water splitting due to suitable band edges, high carrier mobilities, and visible light absorption. Additionally, MSi₂N₄ lattices showed exceptional mechanical properties and thermal conductivities (~440–500 W m K⁻¹). Notably, WS₂N₄, CrSi₂N₄, and MoSi₂N₄ exhibited the highest piezoelectric coefficients, surpassing all known 2D materials.

Table 3 A summary of ML applications in the discovery of knowledge in photocatalysis for water splitting and hydrogen production. Batch size = BS, learning rate = LR, number of hidden neuron unit = NHNU, number of hidden layers = NHL, number of estimators = nest

| Goal | Specific ML application | Size of source data | Source | ML techniques | Validation approach | Optimization tools | Hyperparameter optimized values | Performance metrics | Uncertainty quantification | Main results from ML application | Ref. |
|---|--|--------------------------|--|---|---|--------------------|---|---------------------|----------------------------|---|------|
| Predict photocatalytic water splitting for hydrogen production under natural light | Enhance the precision of hydrogen output forecasts by discerning intricate features | 21 experimental datasets | Experimental (hydrogen production dataset) | GRU-NN, RFR, XGBoost, GBDT | <i>k</i> -Fold, train-test split (specifications not given) | GridSearchCV | 682 epochs; BS = 400, LR = 0.001, NHNU = 512, and NHL = 1 | MSE | Not explicitly mentioned | Estimated the long-term H ₂ production – large-scale application of this technology | 97 |
| Identify the factors that contribute to the photoelectric conversion efficiency of hematite photoanodes under natural light | To reduce the number of features, generate predictive models, classify the sample performance, and select the most appropriate descriptors | 101 samples | Experimental (photoanode dataset) | PLSR, LASSO, stepwise regression, DT regression, RF regression, SVR, GPR, DT regression, PCA, KNN | 5-Fold cross-validation | — | — | RSD | Not explicitly mentioned | Predicted the photocurrent density and PEC activities of hematite photoanodes. Identified the dominant factors responsible for sample variations and inactivity | 261 |

Ma *et al.*²⁵⁸ integrated ML and DFT to design and screen 4624 surface-functionalized SiC monolayers for photocatalytic water splitting. ML models with $R^2 > 0.9$ predicted key properties (E_f , E_g , CBM, VBM, m^*), reducing computational costs. Shapley additive explanations (SHAP) identified influential features, enabling the selection of three promising photocatalysts ($\text{Si}_4\text{C}_4\text{HF}_3\text{Br}_2\text{I}_2$, $\text{Si}_4\text{C}_4\text{F}_3\text{ClBr}_3\text{I}$, $\text{Si}_4\text{C}_4\text{F}_3\text{ClBr}_2\text{I}_2$) with suitable band properties, high carrier mobility, and good optical performance. One candidate supports OER without external voltage, while others require 0.2 V. Additionally, six materials were identified for UV detection with high mobility ($> 10^3 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$) and strong UV absorption. Such workflows underscore the efficacy of ML in expediting materials discovery.

Himmet *et al.*²⁵⁹ investigated WSi_2N_4 , WGe_2N_4 , and Janus WSiGeN_4 monolayers as photocatalysts for water splitting. Using DFT and AIMD simulations, they demonstrated the thermodynamic, mechanical, and dynamic stability of these materials, including their thermal stability at 300 K. To reduce the computational cost of AIMD calculations, a machine-learned force field (MLFF) was employed, leveraging descriptors based on the Gaussian representation of atomic distribution. Bayesian linear regression was used to predict energy, forces, and stress tensor components. The Janus WSiGeN_4 monolayer exhibited optimized band gaps and superior band alignment for redox reactions across a wide pH range, highlighting its excellent photocatalytic activity. All monolayers showed strong visible light absorption, supporting solar energy utilization.

Zhao *et al.*²⁶⁰ developed a band offset precondition and ML model to predict band types in g-GaN-based van der Waals (vdW) heterostructures, achieving an area under the curve (AUC) above 0.9. They compared traditional first-principles methods—requiring convergence tests, structural relaxation, and energy band calculations—with ML approaches that rapidly predict heterojunction properties using data-driven methods (Fig. 8c). A number of ML techniques were evaluated, including SVM, Adaboost, RF, and XGBoost. The latter showed the best discrimination ability according to the confusion matrix. Using HSE06 functionals, they verified type II vdW heterostructures suitable for photocatalytic hydrogen evolution and water splitting. A regression model for band gap prediction achieved a mean absolute error (MAE) of 0.24 eV with 10-fold cross-validation. This study highlights an automated process for early screening of 2D g-GaN based vdW heterostructures and provides valuable guidance for theoretical and experimental research.

The identification of critical factors in photocatalytic systems is contingent upon the availability of analytical devices and reaction apparatuses, as well as the specific objectives of the research endeavor. However, certain parameters are often considered readily quantifiable or of paramount importance, and thus, are generally considered during the evaluation of the various components of the photocatalytic system.

Yang *et al.*⁹⁷ evaluated hydrogen production rates (HPR) under varying radiation intensities ($77.28\text{--}1093.23 \text{ W m}^{-2}$), reaction temperatures ($7.8\text{--}47.1 \text{ }^\circ\text{C}$), reactant concentration

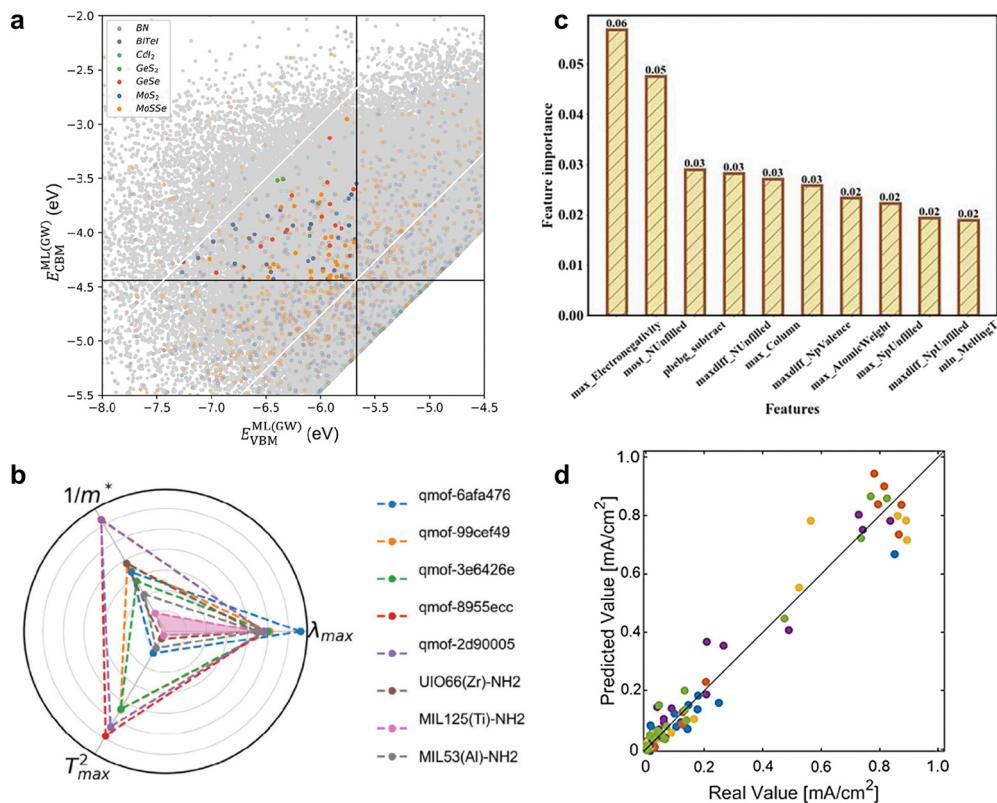


Fig. 8 Two of key areas in the use of machine learning in hydrogen generation by photocatalysis are in finding high-performing photocatalysts by predicting their properties and to identify individual variables and the descriptors influencing the system. (a) The distribution of ML-predicted band edges in the virtual 2D materials database – the light-colored dots represent seven candidates among the 81 candidates obtained after the first-stage screening;¹³¹ (b) five new ML-aided predicted MOF structures theoretically surpassed the performance of known MOFs;¹⁵ (c) identification of the most relevant features in the feature space via a classification model;²⁶⁰ (d) prediction results of photocurrent density predictions for a test data set ($R^2 = 0.95$) – the five colors represent each cross validation fold.²⁶¹

ratios (single, double, triple), solar concentrators (CPC or SUC), with a constant catalyst concentration (0.5 g L^{-1}), and sacrificial materials (0.35 mol L^{-1} Na₂S and 0.25 mol L^{-1} Na₂SO₃). Using a gated recurrent unit neural network (GRU-NN) with the Adam optimizer, they identified radiation intensity as the most critical factor, followed by reaction temperature, both positively influencing HPR. Solar concentrators were also shown to enhance HPR. The GRU-NN achieved a test loss of 2.61×10^{-3} over 682 epochs.

Idei *et al.*²⁶¹ employed ML to predict the photocurrent values at 1.23 V for hematite-based photoanodes (Fig. 8d). Due to limited understanding of factors affecting photoelectric conversion efficiency, the study combined analytical data to identify key operating factors. Linear methods including partial least-squares regression (PLSR), least absolute shrinkage and selection operator (LASSO), and stepwise regression, in addition to nonlinear algorithms, including decision trees (DT) and RF regression, were used to reduce features. Using seven selected descriptors—related to surface facets, band gap energy, and electrical properties at the hematite/fluorine doped tin oxide interface and bulk—the authors built predictive models with SVR, GPR, DT, and RF regression. Emphasis was placed on resistance at the hematite/interface, an underexplored descriptor. The study highlighted the impact of a TiO₂ underlayer on

improving electrical properties at the hematite/interface and film quality.

4. Perspectives in AI-assisted photocatalysis for hydrogen production

The integration of AI and ML methodologies has already demonstrated significant advancements in photocatalysis for hydrogen production. However, their application remains limited in the current literature, leaving much of their potential untapped. Chemistry, particularly catalysis science, has seen pioneering applications of these tools, providing a foundation for extending knowledge into photocatalysis. Although cross-disciplinary knowledge is beneficial, especially in closely related fields, photocatalysis has its particular challenges.

Despite the noteworthy advancements in hydrogen production rates over recent years, AI and ML tools remain underutilized. These tools have the potential to expedite progress in this field, particularly given the intricacy of the involved systems, encompassing the mechanisms of solar energy harvesting, reactor design for scale-up, and environmental impact assessment.

The catalysis community in general has expressed skepticism regarding the utilization of these techniques, primarily due to the black box algorithms that hinder the capacity for reliable interpretability. This has led to a degree of suspicion concerning the possibility that the observed discoveries are merely a consequence of chance, the selection of extrapolated values that might have been identified within a more extensive experimental domain, or whether the observed effect is truly attributable to the efficacy of a robust extrapolation technique in preserving the multidimensionality inherent to the complexity of these systems.

Our perspective is that AI and ML tools could be utilized in more realistic scenario applications. The rationale for this hypothesis is that one powerful attribute of ML algorithms, such as reinforcement learning, is its use in real-time applications. This attribute has been shown to be of great importance in other fields, allowing for in-time changes of operating parameters. Furthermore, we find still scarce the use and the availability of expert-curated datasets, which could present an additional avenue for accelerating progress. Datasets encompassing chemical structures, elemental compositions, physico-chemical properties, and pre-trained DFT data can augment in-house data collection and streamline initial calculations. These resources highlight the potential for leveraging data-driven approaches to enhance photocatalyst design and system optimization. With the growing performance of ML in this domain, further developments are anticipated as datasets and computational tools continue to evolve.

Advanced techniques, such as deep learning, are also underutilized, often due to the complexities of hyperparameter optimization and the scarcity of large, high-quality datasets for model validation. Moreover, data collection practices—whether manual or automated—face regulatory ambiguities under frameworks like the general data protection regulation (GDPR). Current methods for accessing scientific data, such as automated downloads or formal requests to publishers, have limitations and raise questions about data ownership and sharing. Initiatives like the “findable, accessible, interoperable, and reusable” (FAIR) principles aim to address these issues, emphasizing the need for open, structured, and reusable data to advance the field.

Future efforts should focus on overcoming these barriers by fostering collaborations, improving data accessibility, and advancing ML techniques to unlock the full potential of AI-assisted photocatalysis for sustainable hydrogen production.

Author contributions

Leandro Goulart de Araujo: conceptualization, investigation, writing – original draft; David Farrusseng: supervision, writing – review & editing.

Data availability

No primary research results, software or code have been included and no new data were generated or analyzed as part of this review.

Conflicts of interest

There are no conflicts to declare.

References

- X. Gu, P. Liu, L. Peng, Z. Zhang, L. Bian and B. Wang, *Sol. Energy Mater. Sol. Cells*, 2021, **232**, 111343.
- G. Colón, *Appl. Catal., A*, 2016, **518**, 48–59.
- S. Liu, C. Zhang, Y. Sun, Q. Chen, L. He, K. Zhang, J. Zhang, B. Liu and L.-F. Chen, *Coord. Chem. Rev.*, 2020, **413**, 213266.
- Y. Gong, J. Yao, P. Wang, Z. Li, H. Zhou and C. Xu, *Chin. J. Chem. Eng.*, 2022, **43**, 282–296.
- Y. Sari, P. L. Gareso, B. Ardynah and D. Tahir, *Int. J. Hydrogen Energy*, 2024, **55**, 984–996.
- A. Kumar, P. Sharma, G. Sharma, P. Dhiman, G. T. Mola, M. Farghali, A. K. Rashwan, M. Nasr, A. I. Osman and T. Wang, *Environ. Chem. Lett.*, 2024, **22**, 2405–2424.
- X. Zhang, Z. Cheng, C. Bo, Y. Sun and L. Piao, *J. Colloid Interface Sci.*, 2025, **677**, 189–197.
- Y. Liu, D. Huang, M. Cheng, Z. Liu, C. Lai, C. Zhang, C. Zhou, W. Xiong, L. Qin, B. Shao and Q. Liang, *Coord. Chem. Rev.*, 2020, **409**, 213220.
- H. Masood, C. Y. Toe, W. Y. Teoh, V. Sethu and R. Amal, *ACS Catal.*, 2019, **9**, 11774–11787.
- N. O. Eddy, A. O. Odiongenyi, R. Garg, R. A. Ukpe, R. Garg, A. E. Nemr, C. M. Ngwu and I. J. Okop, *Environ. Sci. Pollut. Res.*, 2023, **30**, 64036–64057.
- N. A. Deskins, P. M. Rao and M. Dupuis, in *Springer Handbook of Inorganic Photochemistry*, Springer International Publishing, Cham, 2022, pp. 365–398.
- Y. Zhang, N. Afzal, L. Pan, X. Zhang and J. Zou, *Adv. Sci.*, 2019, **6**, 1900053.
- J. Xue, M. Fujitsuka and T. Majima, *Chem. Commun.*, 2021, **57**, 3532–3542.
- D. Maarisetty and S. S. Baral, *J. Mater. Chem. A*, 2020, **8**, 18560–18604.
- C. Wang, Y. Wan, S. Yang, Y. Xie, S. Chu, Y. Chen and X. Yan, *Adv. Funct. Mater.*, 2024, **34**, 2313596.
- P. Du, J. Huang, Z. Ye, C. Zhen, J. Liu, Y. Yang, Z. Ren, L. Yin, L. Wang and G. Liu, *Adv. Funct. Mater.*, 2024, **34**, 2312888.
- N. J. Van Eck and L. Waltman, *Leiden: Universiteit Leiden*, 2013, **1**, 1–53.
- X. Yu, J. Xie, Q. Liu, H. Dong and Y. Li, *J. Colloid Interface Sci.*, 2021, **593**, 133–141.
- J. Luo, L. Wu, D. Liu, Y. Chen, Q. Lv and H. Deng, *J. Alloys Compd.*, 2025, **1010**, 177888.
- A. Kudo and Y. Miseki, *Chem. Soc. Rev.*, 2009, **38**, 253–278.
- M. El Ouardi, A. El Idrissi, M. Arab, M. Zbair, H. Haspel, M. Saadi and H. Ait Ahsaine, *Int. J. Hydrogen Energy*, 2024, **51**, 1044–1067.
- T. Jafari, E. Moharreri, A. Amin, R. Miao, W. Song and S. Suib, *Molecules*, 2016, **21**, 900.

- 23 S. Sahani, K. Malika Tripathi, T. Il Lee, D. P. Dubal, C.-P. Wong, Y. Chandra Sharma and T. Young Kim, *Energy Convers. Manage.*, 2022, **252**, 115133.
- 24 N. Fajrina and M. Tahir, *Int. J. Hydrogen Energy*, 2019, **44**, 540–577.
- 25 C. Y. Toe, C. Tsounis, J. Zhang, H. Masood, D. Gunawan, J. Scott and R. Amal, *Energy Environ. Sci.*, 2021, **14**, 1140–1175.
- 26 N. Serpone, A. V. Emeline, V. K. Ryabchuk, V. N. Kuznetsov, Y. M. Artem'ev and S. Horikoshi, *ACS Energy Lett.*, 2016, **1**, 931–948.
- 27 V. Kumaravel, M. Imam, A. Badreldin, R. Chava, J. Do, M. Kang and A. Abdel-Wahab, *Catalysts*, 2019, **9**, 276.
- 28 M. Yusuf, P. Rosha, F. Qureshi, F. M. Ali and H. Ibrahim, *Sustainable Mater. Technol.*, 2025, **43**, e01332.
- 29 N. N. Rosman, R. M. Yunus, N. R. A. M. Shah, R. M. Shah, K. Arifin, L. J. Minggu and N. A. Ludin, *Int. J. Energy Res.*, 2022, **46**, 11596–11619.
- 30 L. Sun, W. Wang, T. Kong, H. Jiang, H. Tang and Q. Liu, *J. Mater. Chem. A*, 2022, **10**, 22531–22539.
- 31 B. Abhishek, A. Jayarama, A. S. Rao, S. S. Nagarkar, A. Dutta, S. P. Duttagupta, S. S. Prabhu and R. Pinto, *Int. J. Hydrogen Energy*, 2024, **81**, 1442–1466.
- 32 G. Wang, S. Lv, Y. Shen, W. Li, L. Lin and Z. Li, *J. Materomics*, 2024, **10**, 315–338.
- 33 X. Cai, L. Mao, M. Fujitsuka, T. Majima, S. Kasani, N. Wu and J. Zhang, *Nano Res.*, 2022, **15**, 438–445.
- 34 W. Zhang, H. He, H. Li, L. Duan, L. Zu, Y. Zhai, W. Li, L. Wang, H. Fu and D. Zhao, *Adv. Energy Mater.*, 2021, **11**, 2003303.
- 35 R. Kumar, D. Das and A. K. Singh, *J. Catal.*, 2018, **359**, 143–150.
- 36 J. Zhang, X. Wang, X. Wang and C. Li, *Acc. Chem. Res.*, 2025, acs.accounts.4c00582.
- 37 X. Zhou, J. Liang, L. Xu, S. Wu, M. Xie, Q. Liang, J. Luo, X. Fan, X. Zhou and X. Zhou, *J. Colloid Interface Sci.*, 2025, **689**, 137240.
- 38 Y. Zhu, J. Ren, G. Huang, C. Dong, Y. Huang, P. Lu, H. Tang, Y. Liu, S. Shen and D. Yang, *Adv. Funct. Mater.*, 2024, **34**, 2311623.
- 39 X. Zheng, L. Feng, Y. Dou, H. Guo, Y. Liang, G. Li, J. He, P. Liu and J. He, *ACS Nano*, 2021, **15**, 13209–13219.
- 40 M. Gao, F. Tian, X. Zhang, Y. Liu, Z. Chen, Y. Yu, W. Yang and Y. Hou, *Nano Energy*, 2022, **103**, 107767.
- 41 M. Zhai, Z. Li, R. Li, Z. Feng, Q. Wang, C. Zhang, H. Chi, N. Ta and C. Li, *J. Catal.*, 2025, **443**, 115989.
- 42 L. Xie, J.-G. Hao, H.-Q. Chen, Z.-X. Li, S.-Y. Ge, Y. Mi, K. Yang and K.-Q. Lu, *Catal. Commun.*, 2022, **162**, 106371.
- 43 H. Jiang, Y. Shi and S. Zang, *Int. J. Hydrogen Energy*, 2023, **48**, 17827–17837.
- 44 Z. Su, F. Fang, X. Li, W. Han, X. Liu and K. Chang, *J. Colloid Interface Sci.*, 2022, **626**, 662–673.
- 45 Y. Jiang, J. Zhang, X. Zhang and Y. Ma, *Appl. Surf. Sci.*, 2025, **694**, 162832.
- 46 A. Yamakata, J. J. M. Vequizo, T. Ogawa, K. Kato, S. Tsuboi, N. Furutani, M. Ohtsuka, S. Muto, A. Kuwabara and Y. Sakata, *ACS Catal.*, 2021, **11**, 1911–1919.
- 47 S. Cao, T. Sun, Y. Peng, X. Yu, Q. Li, F. L. Meng, F. Yang, H. Wang, Y. Xie, C. Hou and Q. Xu, *Small*, 2024, **20**, 2404285.
- 48 Q. Wang and K. Domen, *Chem. Rev.*, 2020, **120**, 919–985.
- 49 L. Miao, W. Jia, X. Cao and L. Jiao, *Chem. Soc. Rev.*, 2024, **53**, 2771–2807.
- 50 C. Kranz and M. Wächtler, *Chem. Soc. Rev.*, 2021, **50**, 1407–1437.
- 51 *Towards Sustainable and Green Hydrogen Production by Photocatalysis: Scalability Opportunities and Challenges (Volume 1)*, ed. A. Kumar, American Chemical Society, Washington, DC, 2024, vol. 1467.
- 52 R. Li, J. Luan, Y. Zhang, L. Jiang, H. Yan, Q. Chi and Z. Yan, *Renewable Sustainable Energy Rev.*, 2024, **206**, 114863.
- 53 M. T. M. Moya, L. G. D. Araujo, F. S. Lopes and A. C. S. C. Teixeira, *Int. J. Chem. React. Eng.*, 2023, **21**, 1211–1223.
- 54 O. F. Aldosari and I. Hussain, *Int. J. Hydrogen Energy*, 2024, **59**, 958–981.
- 55 K. Lee, T.-H. Kim, T.-H. Kim, J. Lee and S. Yu, *Sep. Purif. Technol.*, 2023, **312**, 123390.
- 56 K. Villa, J. R. Galán-Mascarós, N. López and E. Palomares, *Sustainable Energy Fuels*, 2021, **5**, 4560–4569.
- 57 P. Zhou, I. A. Navid, Y. Ma, Y. Xiao, P. Wang, Z. Ye, B. Zhou, K. Sun and Z. Mi, *Nature*, 2023, **613**, 66–70.
- 58 Y. Li, L. Li, H. Yuan, K. He, H. Chen, J. Xie, B. Wang and X. Wang, *Appl. Energy*, 2025, **381**, 125179.
- 59 Z. Lu, H. Yan, B. Li, M. Song, Y. Hang, G. Zhou, Y. Xu, C. Ma, S. Han and X. Liu, *Appl. Surf. Sci.*, 2022, **605**, 154694.
- 60 H. Hu, Y. He, X. Wang, T. Zhang, D. Li, M. Sun and C. Deng, *Mater. Sci. Eng., B*, 2024, **305**, 117431.
- 61 F. Tezcan, M. Kahya Düdükcü and G. Kardaş, *J. Electroanal. Chem.*, 2022, **920**, 116595.
- 62 A. E. Charkiewicz, W. J. Omeljaniuk, K. Nowak, M. Garley and J. Nikliński, *Molecules*, 2023, **28**, 6620.
- 63 F. Gao, Z. Xu and H. Zhao, *Proc. Combust. Inst.*, 2021, **38**, 6503–6511.
- 64 S. M. Thabet, H. N. Abdelhamid, S. A. Ibrahim and H. M. El-Bery, *Sci. Rep.*, 2024, **14**, 10115.
- 65 S. B. Potdar, C.-M. Huang, B. Praveen, S. Manickam and S. H. Sonawane, *Catalysts*, 2022, **12**, 78.
- 66 J. Ren, J. Zhang, B. Tian, Z. Pan, S. Wang, H. Chen, K. He, H. Zhong and Q. Wang, *Int. J. Hydrogen Energy*, 2024, **87**, 554–565.
- 67 I. Alfa, H. Y. Hafeez, J. Mohammed, S. Abdu, A. B. Suleiman and C. E. Ndikilar, *Int. J. Hydrogen Energy*, 2024, **71**, 1006–1025.
- 68 P.-Y. Li, J.-H. Yuan, J. Wang, Y. Wang and P. Zhang, *Int. J. Hydrogen Energy*, 2024, **55**, 1254–1264.
- 69 A. David Gaima Kafadi, H. Yusuf Hafeez, J. Mohammed, C. E. Ndikilar, A. B. Suleiman and A. T. Isah, *Int. J. Hydrogen Energy*, 2024, **53**, 1242–1258.
- 70 J. Zhang, L. Zhang, W. Wang and J. Yu, *J. Phys. Chem. Lett.*, 2022, **13**, 8462–8469.
- 71 Z. Zhao, R. V. Goncalves, S. K. Barman, E. J. Willard, E. Byle, R. Perry, Z. Wu, M. N. Huda, A. J. Moulé and F. E. Osterloh, *Energy Environ. Sci.*, 2019, **12**, 1385–1395.

- 72 Y. Li, Z. Wang, Y. Wang, A. Kovács, C. Foo, R. E. Dunin-Borkowski, Y. Lu, R. A. Taylor, C. Wu and S. C. E. Tsang, *Energy Environ. Sci.*, 2022, **15**, 265–277.
- 73 H. Li, W. Chen, Y. Sun, Z. Yuan, W. Lei, M. Yang, P. Ma, J. Wang and J. Niu, *Fuel*, 2024, **356**, 129647.
- 74 S. Xu, M. Li, Y. Wang and Z. Jin, *Int. J. Hydrogen Energy*, 2024, **51**, 16–30.
- 75 T. Kong, A. Liao, Y. Xu, X. Qiao, H. Zhang, L. Zhang and C. Zhang, *RSC Adv.*, 2024, **14**, 17041–17050.
- 76 D. Gunawan, J. Zhang, Q. Li, C. Y. Toe, J. Scott, M. Antonietti, J. Guo and R. Amal, *Adv. Mater.*, 2024, 2404618.
- 77 X.-L. Song, L. Chen, J.-T. Ren, L.-J. Gao and Z.-Y. Yuan, *Coord. Chem. Rev.*, 2024, **507**, 215752.
- 78 M. Aktary, M. Kamruzzaman and R. Afrose, *RSC Adv.*, 2022, **12**, 23704–23717.
- 79 S. Gusanov, *Materials*, 2024, **17**, 2119.
- 80 Z.-Y. Song, Y.-Y. Li, W. Duan, X.-Y. Xiao, Z.-W. Gao, Y.-H. Zhao, B. Liang, S.-H. Chen, P.-H. Li, M. Yang and X.-J. Huang, *TrAC, Trends Anal. Chem.*, 2023, **160**, 116977.
- 81 M. Bursch, J. Mewes, A. Hansen and S. Grimme, *Angew. Chem.*, 2022, **134**, e202205735.
- 82 H. Li, Z. Wang, N. Zou, M. Ye, R. Xu, X. Gong, W. Duan and Y. Xu, *Nat. Comput. Sci.*, 2022, **2**, 367–377.
- 83 H. M. N. Ullah, M. Rizwan, S. S. Ali, Z. Usman and C. Cao, *Mater. Sci. Eng., B*, 2022, **286**, 116041.
- 84 Z. Zhu, X. Tang, T. Wang, W. Fan, Z. Liu, C. Li, P. Huo and Y. Yan, *Appl. Catal., B*, 2019, **241**, 319–328.
- 85 M. Pineda and M. Stamatakis, *J. Chem. Phys.*, 2022, **156**, 120902.
- 86 M. Pineda and M. Stamatakis, *J. Chem. Phys.*, 2017, **147**, 024105.
- 87 G. Pahlevanpour and H. Bashiri, *Int. J. Hydrogen Energy*, 2022, **47**, 12975–12987.
- 88 T. Y. Ahmed, M. M. Ahmad, S. Yusup, A. Inayat and Z. Khan, *Renewable Sustainable Energy Rev.*, 2012, **16**, 2304–2315.
- 89 Y. Yang, D. Jing and L. Zhao, *Appl. Therm. Eng.*, 2020, **173**, 115220.
- 90 Q. Jamil, K. B. Rana and L. Matoh, *Water*, 2024, **16**, 1828.
- 91 A. Yusuf and G. Palmisano, *Chem. Eng. Sci.*, 2021, **229**, 116051.
- 92 C. M. Venier, L. O. Conte, M. Pérez-Moya, M. Graells, N. M. Nigro and O. M. Alfano, *Chem. Eng. J.*, 2021, **410**, 128246.
- 93 S. Schwarz and M. Grünwald, *Chem. Ing. Tech.*, 2024, **96**, 734–740.
- 94 P. Pandey, A. Krishna, S. Mohanan and A. Surenjan, *Process Integr. Optim. Sustainability*, 2025, **9**, 471–485.
- 95 R. Kanthasamy, I. Ali, B. V. Ayodele and H. A. Maddah, *Fuel*, 2023, **344**, 128026.
- 96 Z. Jiang, J. Hu, A. Samia and X. (Bill) Yu, *Catalysts*, 2022, **12**, 746.
- 97 Y. Yang, Y. Zheng, S. Liu, M. Shan, J. Guo, R. Yang, L. Zhao and D. Jing, *Energy Convers. Manage.*, 2024, **301**, 118007.
- 98 J. Liu, L. Liang, B. Su, D. Wu, Y. Zhang, J. Wu and C. Fu, *J. Mater. Inf.*, 2024, **4**, 1–33.
- 99 F. Hashimoto, H. Ohba, K. Ote, A. Kakimoto, H. Tsukada and Y. Ouchi, *Phys. Med. Biol.*, 2021, **66**, 015006.
- 100 E. Wang, J. J. Davis, R. Zhao, H.-C. Ng, X. Niu, W. Luk, P. Y. K. Cheung and G. A. Constantinides, *ACM Comput. Surv.*, 2020, **52**, 1–39.
- 101 E. Can and R. Yıldırım, *Appl. Catal., B*, 2019, **242**, 267–283.
- 102 Y. Hayashi, Y. Nagai, Z. Pan and K. Katayama, *J. Mater. Chem. A*, 2023, **11**, 22522–22532.
- 103 P. A. Koyale, S. V. Mulik, J. L. Gunjakar, T. D. Dongale, V. B. Koli, N. B. Mullani, S. S. Sutar, Y. G. Kapdi, S. S. Soni and S. D. Delekar, *Langmuir*, 2024, **40**, 13657–13668.
- 104 V. Chow, R. C.-W. Phan, A. C. L. Ngo, G. Krishnasamy and S.-P. Chai, *Process Saf. Environ. Prot.*, 2022, **161**, 848–859.
- 105 Y. Zhang, X. Yang, C. Zhang, Z. Zhang, A. Su and Y.-B. She, *Processes*, 2023, **11**, 2614.
- 106 H. Liu, Y.-S. Ong, X. Shen and J. Cai, *IEEE Trans. Neural Networks Learn. Syst.*, 2020, **31**, 4405–4423.
- 107 A. Banerjee, D. B. Dunson and S. T. Tokdar, *Biometrika*, 2013, **100**, 75–89.
- 108 H. Lamouadene, M. El Kassaoui, M. El Yadari, A. El Kenz and A. Benyoussef, *Chem. Phys.*, 2025, **591**, 112603.
- 109 Q. Tao, T. Lu, Y. Sheng, L. Li, W. Lu and M. Li, *J. Energy Chem.*, 2021, **60**, 351–359.
- 110 O. A. Montesinos López, A. Montesinos López and J. Crossa, *Multivariate Statistical Machine Learning Methods for Genomic Prediction*, Springer International Publishing, Cham, 2022, pp. 337–378.
- 111 Y. Cheng and J. Hu, *The 2012 International Joint Conference on Neural Networks (IJCNN)*, IEEE, Brisbane, Australia, 2012, pp. 1–8.
- 112 Z. Wang, Y. Gu, L. Zheng, J. Hou, H. Zheng, S. Sun and L. Wang, *Adv. Mater.*, 2022, **34**, 2106776.
- 113 R. Kojima, S. Ishida, M. Ohta, H. Iwata, T. Honma and Y. Okuno, *J. Cheminf.*, 2020, **12**, 32.
- 114 X. Li, W. Huang, Y. Lian and S. Tao, *J. Phys. Chem. Lett.*, 2024, **15**, 10725–10733.
- 115 K. Choudhary and K. F. Garrity, *Digital Discovery*, 2024, **3**, 1365–1377.
- 116 D. Zheng, C. Ma, M. Wang, J. Zhou, Q. Su, X. Song, Q. Gan, Z. Zhang and G. Karypis, *2020 IEEE/ACM 10th Workshop on Irregular Applications: Architectures and Algorithms (IA3)*, IEEE, GA, USA, 2020, pp. 36–44.
- 117 A. S. Rosen, V. Fung, P. Huck, C. T. O'Donnell, M. K. Horton, D. G. Truhlar, K. A. Persson, J. M. Notestein and R. Q. Snurr, *npj Comput. Mater.*, 2022, **8**, 112.
- 118 Y. Guan, D. Chaffart, G. Liu, Z. Tan, D. Zhang, Y. Wang, J. Li and L. Ricardez-Sandoval, *Chem. Eng. Sci.*, 2022, **248**, 117224.
- 119 L. Barreñada, P. Dhiman, D. Timmerman, A.-L. Boulesteix and B. Van Calster, *Diagn. Prognostic Res.*, 2024, **8**, 14.
- 120 M. F. Javed, M. Z. Shahab, U. Asif, T. Najeh, F. Aslam, M. Ali and I. Khan, *Sci. Rep.*, 2024, **14**, 13688.
- 121 J. Cai, K. Xu, Y. Zhu, F. Hu and L. Li, *Appl. Energy*, 2020, **262**, 114566.
- 122 J. Zhang, Y. Wang, X. Zhou, C. Zhong, K. Zhang, J. Liu, K. Hu and X. Lin, *Nanoscale*, 2023, **15**, 11072–11082.
- 123 C. Bentéjac, A. Csörgő and G. Martínez-Muñoz, *Artif. Intell. Rev.*, 2021, **54**, 1937–1967.

- 124 A. V. Konstantinov and L. V. Utkin, *Knowledge-Based Systems*, 2021, **vol. 222**, p. 106993.
- 125 R. Bakir, C. Orak and A. Yüksel, *Int. J. Hydrogen Energy*, 2024, **67**, 101–110.
- 126 P. Zhou, M. Wang, F. Tang, L. Ling, H. Yu and X. Chen, *Mater. Res. Bull.*, 2024, **179**, 112956.
- 127 K. Curtis and S. O. Odoh, *J. Comput. Chem.*, 2025, **46**, e70006.
- 128 M. Ceriotti, C. Clementi and O. Anatole Von Lilienfeld, *Chem. Rev.*, 2021, **121**, 9719–9721.
- 129 T. Pigeon, G. Stoltz, M. Corral-Valero, A. Anciaux-Sedrakian, M. Moreaud, T. Lelièvre and P. Raybaud, *J. Chem. Theory Comput.*, 2023, **19**, 3538–3550.
- 130 L. Yan, S. Zhong, T. Igou, H. Gao, J. Li and Y. Chen, *Int. J. Hydrogen Energy*, 2022, **47**, 34075–34089.
- 131 Y. Wang, M. C. Sorkun, G. Brocks and S. Er, *J. Phys. Chem. Lett.*, 2024, **15**, 4983–4991.
- 132 C. A. Kim, N. D. Ricke and T. Van Voorhis, *J. Chem. Phys.*, 2021, **155**, 144107.
- 133 J. Jung, S. Ju, P. Kim, D. Hong, W. Jeong, J. Lee, S. Han and S. Kang, *ACS Catal.*, 2023, **13**, 16078–16087.
- 134 G. X. De Oliveira, S. Kuhn, H. G. Riella, C. Soares and N. Padoin, *React. Chem. Eng.*, 2023, **8**, 2119–2133.
- 135 T. Ren, L. Wang, C. Chang and X. Li, *Energy Convers. Manage.*, 2020, **216**, 112935.
- 136 S. Sharma, P. Kumar and R. Chandra, *Molecular Dynamics Simulation of Nanocomposites Using BIOVIA Materials Studio, Lammps and Gromacs*, Elsevier, 2019, pp. 1–38.
- 137 A. T. F. Batista, W. Baaziz, A.-L. Taleb, J. Chaniot, M. Moreaud, C. Legens, A. Aguilar-Tapia, O. Proux, J.-L. Hazemann, F. Diehl, C. Chizallet, A.-S. Gay, O. Ersen and P. Raybaud, *ACS Catal.*, 2020, **10**, 4193–4204.
- 138 R. Wang, R. Liu, S. Luo, J. Wu, D. Zhang, T. Yue, J. Sun, C. Zhang, L. Zhu and J. Wang, *Chem. Eng. J.*, 2021, **421**, 129596.
- 139 J. Guo, X. Liao, M.-H. Lee, G. Hyett, C.-C. Huang, D. W. Hewak, S. Mailis, W. Zhou and Z. Jiang, *Appl. Catal., B*, 2019, **243**, 502–512.
- 140 X. Zhou and H. Dong, *ChemCatChem*, 2019, **11**, 3688–3715.
- 141 J. Wirth, R. Neumann, M. Antonietti and P. Saalfrank, *Phys. Chem. Chem. Phys.*, 2014, **16**, 15917–15926.
- 142 W.-N. Zhao and Z.-P. Liu, *Chem. Sci.*, 2014, **5**, 2256–2264.
- 143 B. Samanta, Á. Morales-García, F. Illas, N. Goga, J. A. Anta, S. Calero, A. Bieberle-Hütter, F. Libisch, A. B. Muñoz-García, M. Pavone and M. Caspary Toroker, *Chem. Soc. Rev.*, 2022, **51**, 3794–3818.
- 144 L. Ju, J. Shang, X. Tang and L. Kou, *J. Am. Chem. Soc.*, 2020, **142**, 1492–1500.
- 145 A. Ngoipala, L. Ngamwongwan, I. Fongkaew, S. Jungthawan, P. Hirunsit, S. Limpijumnong and S. Suthirakun, *J. Phys. Chem. C*, 2020, **124**, 4352–4362.
- 146 C. M. Bishop, *Pattern recognition and machine learning*, Springer, New York, 2006.
- 147 I. Goodfellow, A. Courville and Y. Bengio, *Deep learning*, The MIT Press, Cambridge, Massachusetts, 2016.
- 148 C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, The MIT Press, 2005.
- 149 T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning*, Springer New York, New York, NY, 2009.
- 150 *Automated Machine Learning: Methods, Systems, Challenges*, ed. F. Hutter, L. Kotthoff and J. Vanschoren, Springer International Publishing, Cham, 2019.
- 151 *Graph Neural Networks: Foundations, Frontiers, and Applications*, ed. L. Wu, P. Cui, J. Pei and L. Zhao, Springer Nature Singapore, Singapore, 2022.
- 152 S. Agatonovic-Kustrin and R. Beresford, *J. Pharm. Biomed. Anal.*, 2000, **22**, 717–727.
- 153 Y. Zheng, R. Yang and Y. Yang, *Int. J. Hydrogen Energy*, 2025, **111**, 205–219.
- 154 A. Agarwal, S. Goverapet Srinivasan and B. Rai, *Front. Mater.*, 2021, **8**, 679269.
- 155 A. Główka, E. Jolimaitre, A. Hammoumi, M. Moreaud, L. Sorbier, C. De Faria Barros, V. Lefebvre and M.-O. Coppens, *J. Phys. Chem. C*, 2024, **128**, 8395–8407.
- 156 D. Mondal, S. Datta and D. Jana, *Phys. Chem. Chem. Phys.*, 2025, **27**, 4531–4566.
- 157 V. Y. Yurova, K. O. Potapenko, T. A. Aliev, E. A. Kozlova and E. V. Skorb, *Int. J. Hydrogen Energy*, 2024, **81**, 193–203.
- 158 J. Zhang, C. Xiao, W. Yang, X. Liang, L. Zhang and X. Wang, *Geothermics*, 2023, **114**, 102787.
- 159 Y. Haghshenas, W. P. Wong, D. Gunawan, A. Khataee, R. Keyikoğlu, A. Razmjou, P. V. Kumar, C. Y. Toe, H. Masood, R. Amal, V. Sethu and W. Y. Teoh, *EES Catal.*, 2024, **2**, 612–623.
- 160 T. O. Owolabi and M. A. Abd Rahman, *Symmetry*, 2021, **13**, 411.
- 161 S. O. Olatunji, *Comput. Mater. Sci.*, 2021, **200**, 110797.
- 162 S. Zhang, H. Tong, J. Xu and R. Maciejewski, *Comput. Soc. Networks*, 2019, **6**, 11.
- 163 Z. Jiang, J. Hu, M. Tong, A. C. Samia, H. (Judy) Zhang and X. (Bill) Yu, *Catalysts*, 2021, **11**, 1107.
- 164 Q. Liu, K. Pan, L. Zhu, Y. Zhou, Y. Lu, S. Wang, Z. Ding, W. Du and Y. Zhou, *Green Chem.*, 2023, **25**, 8778–8790.
- 165 J. H. Friedman, *Ann. Stat.*, 2001, **29**, 1189–1232.
- 166 M. Raissi, P. Perdikaris and G. E. Karniadakis, *J. Comput. Phys.*, 2019, **378**, 686–707.
- 167 J. Y. Y. Loh, A. Wang, A. Mohan, A. A. Tountas, A. M. Gouda, A. Tavasoli and G. A. Ozin, *Adv. Sci.*, 2024, **11**, 2306604.
- 168 S. I. Ngo and Y.-I. Lim, *Catalysts*, 2021, **11**, 1304.
- 169 P. Sousa, C. V. Rodrigues and A. Afonso, *Comput. Methods Appl. Mech. Eng.*, 2024, **429**, 117133.
- 170 X. Chen, Z. Wang, L. Deng, J. Yan, C. Gong, B. Yang, Q. Wang, Q. Zhang, L. Yang, Y. Pang and J. Liu, *Eng. Appl. Comput. Fluid Mech.*, 2024, **18**, 2407005.
- 171 V. Fung, P. Ganesh and B. G. Sumpter, *Chem. Mater.*, 2022, **34**, 4848–4855.
- 172 F. Sorourifar, Y. Peng, I. Castillo, L. Bui, J. Venegas and J. A. Paulson, *Ind. Eng. Chem. Res.*, 2023, **62**, 15563–15577.
- 173 J. Li and B. Li, *Chaos, Solitons Fractals*, 2022, **164**, 112712.
- 174 Q. Zhang, C. Qiu, J. Hou and W. Yan, *Commun. Nonlinear Sci. Numer. Simul.*, 2024, **138**, 108229.

- 175 Z. Zhang, Y. Zhang and S. Liu, *Comput. Mater. Sci.*, 2024, **236**, 112889.
- 176 Z. Pan, Y. Zhou and L. Zhang, *ACS Appl. Mater. Interfaces*, 2022, **14**, 9933–9943.
- 177 J. D. Toscano, V. Oommen, A. J. Varghese, Z. Zou, N. Ahmadi Daryakenari, C. Wu and G. E. Karniadakis, *Mach. Learn. Comput. Sci. Eng.*, 2025, **1**, 15.
- 178 J. Ran, A. Talebian-Kiakalaieh, S. Zhang, E. M. Hashem, M. Guo and S.-Z. Qiao, *Chem. Sci.*, 2024, **15**, 1611–1637.
- 179 L. G. de Araujo, L. Oscar Conte, A. Violeta Schenone, O. M. Alfano, A. C. S. C. Teixeira, L. O. Conte, A. V. Schenone, O. M. Alfano and A. C. S. C. Teixeira, *Environ. Sci. Pollut. Res.*, 2020, **27**, 7299–7308.
- 180 C. Nchikou, *Chem. Eng. Commun.*, 2025, **212**, 422–440.
- 181 J. Geng, Q. Wei, B. Luo, S. Zong, L. Ma, Y. Luo, C. Zhou and T. Deng, *Catalysts*, 2024, **14**, 237.
- 182 R. Peralta Muniz Moreira and G. Li Puma, *Catal. Today*, 2021, **361**, 77–84.
- 183 P. Srilatha, K. Karthik, K. V. Prasad, A. Abdulrahman, R. S. V. Kumar, R. J. P. Gowda and R. N. Kumar, *J. Comput. Sci.*, 2024, **82**, 102428.
- 184 P. A. Bello, K. S. Adegbie and A. Adewole, *Results Eng.*, 2024, **21**, 101855.
- 185 L. Guo and J. N. Harvey, *Catal. Sci. Technol.*, 2024, **14**, 961–972.
- 186 H. Maleki, M. Fulton and V. Bertola, *Chem. Eng. J.*, 2021, **411**, 128595.
- 187 L. G. De Araujo, L. Vilcocq, P. Fongarland and Y. Schuurman, *Chem. Eng. J.*, 2025, **508**, 160872.
- 188 S. Choi, Y. Kim, J. W. Kim, Z. Kim and W. Y. Kim, *Chem. – Eur. J.*, 2018, **24**, 12354–12358.
- 189 J. Liu, H. Jia, K. Mairaj Deen, Z. Xu, C. Cheng and W. Zhang, *Fuel*, 2023, **343**, 128005.
- 190 Y. Xiang, X. Wang, L. Rao, P. Wang, D. Huang, X. Ding, X. Zhang, S. Wang, H. Chen and Y. Zhu, *ACS Energy Lett.*, 2018, **3**, 2544–2549.
- 191 G. B. Park, T. N. Kitsopoulos, D. Borodin, K. Golibrzuch, J. Neugebohren, D. J. Auerbach, C. T. Campbell and A. M. Wodtke, *Nat. Rev. Chem.*, 2019, **3**, 723–732.
- 192 R. Kumar and A. K. Singh, *npj Comput. Mater.*, 2021, **7**, 197.
- 193 S. Agrawal, B. Wang, Y. Wu, D. Casanova and O. V. Prezhdo, *Nanoscale*, 2024, **16**, 8986–8995.
- 194 Z. W. Chen, Z. Lu, L. X. Chen, M. Jiang, D. Chen and C. V. Singh, *Chem Catal.*, 2021, **1**, 183–195.
- 195 J. Wang, M. Zhang, J. Meng, Q. Li and J. Yang, *RSC Adv.*, 2017, **7**, 24446–24452.
- 196 H. Zhang, J. Liu, T. Xu, W. Ji and X. Zong, *Catalysts*, 2023, **13**, 728.
- 197 M. M. Abouelela, G. Kawamura and A. Matsuda, *J. Cleaner Prod.*, 2021, **294**, 126200.
- 198 H. Han and X. Meng, *J. Colloid Interface Sci.*, 2023, **650**, 846–856.
- 199 M. Tamtaji, S. Chen, Z. Hu, W. A. Goddard Iii and G. Chen, *J. Phys. Chem. C*, 2023, **127**, 9992–10000.
- 200 O. Allam, M. Maghsoudi, S. S. Jang and S. D. Snow, *ACS Appl. Mater. Interfaces*, 2024, **16**, 36215–36223.
- 201 G. Ramkumar, M. Tamilselvi, S. D. Sundarsingh Jebaseelan, V. Mohanavel, H. Kamyab, G. Anitha, R. Thandaiah Prabu and M. Rajasimman, *Int. J. Hydrogen Energy*, 2023, S0360319923035656.
- 202 R. Binjhade, R. Mondal and S. Mondal, *J. Environ. Chem. Eng.*, 2022, **10**, 107746.
- 203 H. Nishiyama, T. Yamada, M. Nakabayashi, Y. Maehara, M. Yamaguchi, Y. Kuromiya, Y. Nagatsuma, H. Tokudome, S. Akiyama, T. Watanabe, R. Narushima, S. Okunaka, N. Shibata, T. Takata, T. Hisatomi and K. Domen, *Nature*, 2021, **598**, 304–307.
- 204 H. L. Otálvaro-Marín, M. A. Mueses and F. Machuca-Martínez, *Int. J. Photoenergy*, 2014, **2014**, 1–8.
- 205 A. Núñez F., R. Vial, C. Evensen, C. Muñoz-Herrera, L. Díaz R., V. Jovicic, A. Delgado and M. Toledo, *Chem. Eng. Sci.*, 2023, **282**, 119298.
- 206 D. Jing, H. Liu, X. Zhang, L. Zhao and L. Guo, *Energy Convers. Manage.*, 2009, **50**, 2919–2926.
- 207 V. Chausse and J. Llorca, *Catal. Today*, 2022, **383**, 156–163.
- 208 P. S. Walko and R. N. Devi, *Int. J. Hydrogen Energy*, 2023, **48**, 17086–17096.
- 209 L.-T. Zhu, X.-Z. Chen, B. Ouyang, W.-C. Yan, H. Lei, Z. Chen and Z.-H. Luo, *Ind. Eng. Chem. Res.*, 2022, **61**, 9901–9949.
- 210 M. R. Kunz, A. Yonge, Z. Fang, R. Batchu, A. J. Medford, D. Constales, G. Yablonsky and R. Fushimi, *Chem. Eng. J.*, 2021, **420**, 129610.
- 211 E. Rakić, M. Grilc and B. Likozar, *Chem. Eng. J.*, 2023, **472**, 144836.
- 212 R. Michiels, N. Gerrits, E. Neyts and A. Bogaerts, *J. Phys. Chem. C*, 2024, **128**, 11196–11209.
- 213 D. Kiani and I. E. Wachs, *ACS Catal.*, 2024, **14**, 10260–10270.
- 214 D. Aouf, Y. Khane, F. Fenniche, S. Albukhaty, G. M. Sulaiman, S. Khane, A. Henni, A. Zoukel, N. Dizge, H. A. Mohammed and M. M. Abomughaid, *Nanotechnol. Rev.*, 2024, **13**, 20240002.
- 215 M. Ikram, S. Moeen, A. Shahzadi, M. M. Al-Anazy and M. Jeridi, *Mater. Sci. Semicond. Process.*, 2024, **181**, 108633.
- 216 H. Abdollahzadeh, Y. Rashtbari, J. H. P. Américo-Pinheiro, A. Azari, S. Afshin, M. Fazlzadeh and Y. Poureshgh, *Biomass Convers. Biorefin.*, 2024, **14**, 27307–27316.
- 217 H. Esfandian, M. Rostamnejad Cherati and M. Khatirian, *Inorg. Chem. Commun.*, 2024, **159**, 111750.
- 218 S. Bimli, S. R. Mulani, E. Choudhary, V. Manjunath, P. Shinde, S. R. Jadkar and R. S. Devan, *Int. J. Hydrogen Energy*, 2024, **51**, 1497–1507.
- 219 J. Ma, S. Qi, G. Yan, A. M. Kirillov, L. Yang and R. Fang, *Mol. Catal.*, 2024, **560**, 114126.
- 220 Y. Liu, Y. Yao, Z. Liang, Z. Gong, J. Li, Z. Tang and X. Wei, *J. Phys. Chem. C*, 2024, **128**, 9894–9903.
- 221 P. Sinha, M. V. Jyothirmai, B. M. Abraham and J. K. Singh, *Mater. Chem. Phys.*, 2024, **326**, 129805.
- 222 S. Özsoysal, B. Oral and R. Yıldırım, *J. Mater. Chem. A*, 2024, **12**, 5748–5759.
- 223 G. Lu, F. Chu, X. Huang, Y. Li, K. Liang and G. Wang, *Coord. Chem. Rev.*, 2022, **450**, 214240.

- 224 H. Zhang, Y. Gao, S. Meng, Z. Wang, P. Wang, Z. Wang, C. Qiu, S. Chen, B. Weng and Y. Zheng, *Adv. Sci.*, 2024, **11**, 2400099.
- 225 L. Xiong and J. Tang, *Adv. Energy Mater.*, 2021, **11**, 2003216.
- 226 S. Wei, S. Chang, J. Qian and X. Xu, *Small*, 2021, **17**, 2100084.
- 227 Z. Huang, N. Luo, C. Zhang and F. Wang, *Nat. Rev. Chem.*, 2022, **6**, 197–214.
- 228 M. F. Kuehnel and E. Reisner, *Angew. Chem., Int. Ed.*, 2018, **57**, 3290–3296.
- 229 V. L'hospital, L. G. De Araujo, Y. Schuurman, N. Guilhaume and D. Farrusseng, *New J. Chem.*, 2024, **48**, 9656–9662.
- 230 C. Vogt and B. M. Weckhuysen, *Nat. Rev. Chem.*, 2022, **6**, 89–111.
- 231 Y. Lei, Z. Wang, A. Bao, X. Tang, X. Huang, H. Yi, S. Zhao, T. Sun, J. Wang and F. Gao, *Chem. Eng. J.*, 2023, **453**, 139663.
- 232 Y. Xie, J. Wen, Z. Li, J. Chen, Q. Zhang, P. Ning and J. Hao, *ACS Mater. Lett.*, 2023, **5**, 2629–2647.
- 233 V. L'hospital, S. Heyte, S. Paul, K. Parkhomenko and A.-C. Roger, *Fuel*, 2022, **319**, 123689.
- 234 S. Baburaj, L. K. Valloli, J. Parthiban, D. Garg and J. Sivaguru, *Chem. Commun.*, 2022, **58**, 1871–1880.
- 235 P. Tang, H. Li, X. Zhang and X. Sun, *RSC Adv.*, 2023, **13**, 10703–10714.
- 236 P. Shambhawi, J. M. Weber and A. A. Lapkin, *Chem. Eng. J.*, 2023, **466**, 143212.
- 237 P. Lozano-Reis, P. Gamallo, R. Sayós and F. Illas, *ACS Catal.*, 2024, **14**, 2284–2299.
- 238 H. Chang, K. Tu, X. Zhang, J. Zhao, X. Zhou and H. Zhang, *J. Phys. Chem. C*, 2019, **123**, 25091–25101.
- 239 Y. Wang, Y. Sun, X. Liu and F. Dong, *PNAS Nexus*, 2024, **3**, pgae339.
- 240 I. Ioannou, S. C. D'Angelo, Á. Galán-Martín, C. Pozo, J. Pérez-Ramírez and G. Guillén-Gosálbez, *React. Chem. Eng.*, 2021, **6**, 1179–1194.
- 241 Y. Luo and R. Su, *Toxics*, 2024, **12**, 652.
- 242 T. Cordero-Lanzac, A. Ramirez, A. Navajas, L. Gevers, S. Brunialti, L. M. Gandía, A. T. Aguayo, S. Mani Sarathy and J. Gascon, *J. Energy Chem.*, 2022, **68**, 255–266.
- 243 A. Sayyah, M. Ahangari, J. Mostafaei, S. R. Nabavi and A. Niaezi, *J. Cleaner Prod.*, 2023, **426**, 139120.
- 244 A. Omidkar, A. Alagumalai, Z. Li and H. Song, *Appl. Energy*, 2024, **355**, 122321.
- 245 B. Oral, E. Can and R. Yildirim, *Int. J. Hydrogen Energy*, 2022, **47**, 19633–19654.
- 246 A. Bhattacharya, J. A. Benavides, L. F. Gerlein and S. G. Cloutier, *Sci. Rep.*, 2022, **12**, 21874.
- 247 Y. Su, X. Wang, Y. Ye, Y. Xie, Y. Xu, Y. Jiang and C. Wang, *Chem. Sci.*, 2024, **15**, 12200–12233.
- 248 A. M. Bran, S. Cox, O. Schilter, C. Baldassari, A. D. White and P. Schwaller, *Nat. Mach. Intell.*, 2024, **6**, 525–535.
- 249 Z. Zheng, O. Zhang, C. Borgs, J. T. Chayes and O. M. Yaghi, *J. Am. Chem. Soc.*, 2023, **145**, 18048–18062.
- 250 T. Isazawa and J. M. Cole, *Sci. Data*, 2023, **10**, 651.
- 251 M. C. Swain and J. M. Cole, *J. Chem. Inf. Model.*, 2016, **56**, 1894–1904.
- 252 A. V. Kalikadien, A. Mirza, A. N. Hossaini, A. Sreenithya and E. A. Pidko, *ChemPlusChem*, 2024, **89**, e202300702.
- 253 S. Mace, Y. Xu and B. N. Nguyen, *ChemCatChem*, 2024, **16**, e202301475.
- 254 A. Trunschke, G. Bellini, M. Boniface, S. J. Carey, J. Dong, E. Erdem, L. Foppa, W. Frandsen, M. Geske, L. M. Ghiringhelli, F. Girgsdies, R. Hanna, M. Hashagen, M. Hävecker, G. Huff, A. Knop-Gericke, G. Koch, P. Kraus, J. Kröhnert, P. Kube, S. Lohr, T. Lunkenbein, L. Masliuk, R. Naumann d'Alnoncourt, T. Omojola, C. Pratsch, S. Richter, C. Rohner, F. Rosowski, F. Rüther, M. Scheffler, R. Schlögl, A. Tarasov, D. Teschner, O. Timpe, P. Trunschke, Y. Wang and S. Wrabetz, *Top. Catal.*, 2020, **63**, 1683–1699.
- 255 M. Bonchio, J. Bonin, O. Ishitani, T.-B. Lu, T. Morikawa, A. J. Morris, E. Reisner, D. Sarkar, F. M. Toma and M. Robert, *Nat. Catal.*, 2023, **6**, 657–665.
- 256 M. Biswas, R. Desai and A. Mannodi-Kanakkithodi, *Phys. Chem. Chem. Phys.*, 2024, **26**, 23177–23188.
- 257 B. Mortazavi, B. Javvaji, F. Shojaei, T. Rabczuk, A. V. Shapeev and X. Zhuang, *Nano Energy*, 2021, **82**, 105716.
- 258 X. Ma, J. Yuan and Y. Mao, *Mater. Today Chem.*, 2024, **41**, 102302.
- 259 F. Himmet, G. Surucu, S. B. Lisesivdin, O. Surucu, G. Altuntas, B. Bostan and A. Gencer, *Int. J. Hydrogen Energy*, 2024, **78**, 761–772.
- 260 Z. Zhao, Y. Shen, H. Zhu, Q. Zhang, Y. Zhang, X. Yang, P. Liang and L. Chen, *Appl. Surf. Sci.*, 2023, **640**, 158400.
- 261 T. Idei, Y. Nagai, Z. Pan and K. Katayama, *ACS Appl. Mater. Interfaces*, 2023, **15**, 55644–55651.
- 262 J. Yang, P. Manganaris and A. Mannodi-Kanakkithodi, *J. Chem. Phys.*, 2024, **160**, 064114.
- 263 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder and K. A. Persson, *APL Mater.*, 2013, **1**, 011002.