Machine learning for hydrogen technologies: a comprehensive review of challenges, opportunities, and emerging trends

Laag, R.P. van der, Rizzato, A.; Bäck, T.H.W.; Fan, Y.

International Journal of Hydrogen Energy 197 (2026) 152556

Contents lists available at ScienceDirect

International Journal of Hydrogen Energy

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

Review Article

Machine learning for hydrogen technologies: A comprehensive review of challenges, opportunities, and emerging trends

Robin van der Laag *, Agnese Rizzato, Thomas Bäck, Yingjie Fan

Leiden Institute of Advanced Computer Science, Einsteinweg 55, Leiden, Netherlands

ARTICLE INFO

Keywords:
Hydrogen
Sustainability
Machine learning
Artificial intelligence
Optimization

ABSTRACT

Hydrogen technologies are central to the transition toward more sustainable energy systems. Machine learning (ML) and artificial intelligence (AI) are emerging as key enablers across the hydrogen value chain, from production and transport to storage and utilization. This review synthesizes ML applications across these domains, mapping hydrogen technologies to ML task types and method families and assessing reported benefits and limitations. A systematic review of 314 peer-reviewed studies (2019–2024) shows rapid growth in ML-driven research, concentrated in production and material characterization, with transport, underground storage, and several utilization topics still underexplored. Predictive modeling dominates, typically using neural networks and tree-based ensembles, while optimization and hybrid physics-informed approaches are less common. The review highlights gaps in data availability, transparency, and explainability, and outlines priorities for standardized data reporting, adoption of physics-informed and explainable ML, and more rigorous cross-domain benchmarking.

Contents

* Corresponding author.
E-mail address: r.p.van.der.laag@liacs.leidenuniv.nl (R. van der Laag).

https://doi.org/10.1016/j.ijhydene.2025.152556
Received 23 June 2025; Received in revised form 10 November 2025; Accepted 11 November 2025
Available online 19 November 2025

1. Introduction

1.1. Background

Hydrogen is emerging as a key enabler in the global energy transition, particularly in carbon-intensive industries such as heavy manufacturing and long-distance transportation. As countries aim to meet their net-zero carbon emission targets, hydrogen offers a versatile, cleaner solution for a range of energy systems. According to the International Energy Agency, low-emission hydrogen production could reach up to 49 MTPA by 2030 [1], helping to reduce reliance on fossil fuels and limit global CO_2 emissions. This versatility is due to hydrogen's ability to be produced from diverse sources, including renewable sources like solar and wind, as well as its applications in a wide variety of industries.

Hydrogen's potential lies not only in its role as a cleaner fuel but also as a long-term energy storage medium that can support grid stability by balancing intermittent renewable energy generation. Unlike conventional fossil fuels, hydrogen combustion does not produce any carbon emissions and has a higher calorific value on a mass basis, making it a sustainable energy carrier for industrial sectors requiring intensive heat generation. Moreover, green hydrogen, produced via water electrolysis using renewable electricity, represents a key pathway to decarbonize energy systems.

Governments worldwide are recognizing hydrogen's potential, with countries such as the Netherlands [2], the US [3], and Japan [4] developing national strategies to accelerate the adoption of hydrogen technologies. These strategies reflect the broader trend toward a hydrogen economy, in which hydrogen plays a pivotal role in reducing reliance on carbon-intensive energy systems while boosting economic growth through the development of new infrastructures and industries. However, realizing hydrogen's potential requires addressing technical, economic, and logistical challenges across its entire value chain, including production, transportation, storage, and end use.

1.2. The hydrogen value chain: From production to utilization

Each of these four stages presents unique optimization challenges. Hydrogen production can be divided into several methods, including fossil fuel-based processes such as steam methane reforming (SMR), and cleaner alternatives like electrolysis. SMR remains the most widely used process [1], but it is carbon-intensive unless combined with carbon capture, utilization, and storage (CCUS) technologies.

Green hydrogen from renewable-powered electrolysis is growing, but high cost and energy use create limitations. Once produced, hydrogen must be transported efficiently to end users via compression, liquefaction, or conversion into liquid carriers for easier handling. Each method comes with specific energy demands and infrastructure challenges, particularly concerning pipeline networks and refueling stations. Pipeline transport is feasible, but requires overcoming material degradation and leakages from hydrogen embrittlement, where prolonged hydrogen exposure makes steels more brittle and crack-prone.

Hydrogen storage is critical for balancing supply and demand and can be done as compressed gas, liquid hydrogen, or solid-state (e.g., metal hydrides). Large underground options like salt caverns or depleted gas fields are promising but still limited by cost and site availability.

On the end-use side, hydrogen is being adopted in fuel cells for transport and stationary power, and is a candidate for hard-to-abate industries (e.g., steel, aviation). Yet its wide flammability range (4%–75% vol. in air) and very low ignition energy (0.02 mJ) raise safety requirements, so leak detection and strict protocols are essential. Across all stages, materials discovery, such as low–noble metal electrolysis catalysts and advanced storage materials like MOFs, aims to improve safety, efficiency, and cost [5,6].

1.3. Challenges in scaling hydrogen technologies

Despite its potential, the hydrogen sector faces significant challenges that hinder its widespread adoption. One of the primary barriers is the high cost of hydrogen production, particularly green hydrogen. Electrolysis, the most promising process for producing carbon-neutral hydrogen, remains expensive due to the cost of renewable electricity and the capital investments required to manufacture electrolyzer systems. Currently, the cost of hydrogen produced via renewable-powered electrolysis is in the range of 3–12 USD/kg H_2, which is significantly higher than the cost of hydrogen produced via SMR, at 0.8–5.7 USD/kg [1].

Another major hurdle is the limited infrastructure for hydrogen transport, storage, and distribution. In many regions, hydrogen refueling stations and pipelines are either underdeveloped or nonexistent, significantly slowing the uptake of hydrogen as a viable fuel source. Moreover, existing natural gas transport infrastructure is not always compatible with hydrogen due to embrittlement and leakage risks.

From a technical standpoint, material durability and efficiency are critical concerns. For example, current materials employed in electrolyzers and fuel cells suffer degradation, resulting in a consequent drop in performance. Improving the safety of hydrogen technologies is equally important. Public concerns over the flammability and handling of hydrogen, reinforced by incidents involving hydrogen leaks and fires, have created barriers to its adoption. These issues, combined with challenges around public acceptance and awareness, must be addressed to ensure that hydrogen plays a significant role in the future energy landscape.

1.4. Machine learning for advancing hydrogen technologies

Artificial intelligence (AI) and machine learning (ML) have significant potential to address many of the challenges associated with hydrogen technologies. These data-driven techniques can optimize hydrogen production processes by improving electrolyzer efficiency and identifying optimal operating conditions through predictive modeling.

Nomenclature

MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
RMSE	Root Mean Square Error

Hydrogen

AEM	Anion Exchange Membrane
ATR	Autothermal Reforming
BoP	Balance of Plant
CCUS	Carbon Capture, Utilization, and Storage
DFT	Density Functional Theory
FCV	Fuel Cell Vehicles
HER	Hydrogen Evolution Reaction
MEC	Microbial Electrolysis Cell
MOF	Metal-Organic Framework
PEM	Proton Exchange Membrane
PEMFC	Proton Exchange Membrane Fuel Cell
PEMWE	Proton Exchange Membrane Water Electrolysis
POX	Partial Oxidation
PV	Photovoltaic
SCWG	Supercritical Water Gasification
SE-SMR	Sorption-enhanced SMR
SMR	Steam Methane Reforming
SO	Solid-Oxide
UHS	Underground Hydrogen Storage

Machine Learning

AE	Autoencoder
AI	Artificial Intelligence
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Network
AR	Autoregressive
CNN	Convolutional Neural Network
DT	Decision Tree
ELM	Extreme Learning Machine
ERT	Extremely Randomized Trees
GB	Gradient Boosting
GPR	Guassian Process Regression
k-NN	k-Nearest Neighbors
KRR	Kernel Ridge Regression
LR	Linear Regression
LSTM	Long Short-Term Memory
ML	Machine Learning
MLR	Multiple Linear Regression
PCA	Principle Component Analysis
PIML	Physics-Informed Machine Learning
RBF	Radial Basis Function
RF	Random Forest
RL	Reinforcement Learning
RNN	Reccurent Neural Network
RR	Ridge Regression
SR	Symbolic Regression
SVM	Support Vector Machine

Optimization

BO	Bayesian Optimization
DE	Differential Evolution
GA	Genetic Algorithm
NSGA	Non-dominated Sorting Genetic Algorithm
PSO	Particle Swarm Optimization
WOA	Whale Optimization Algorithm

In the transport and storage sectors, ML can be used for predictive maintenance to prevent equipment failures and ensure safe operation by mapping the risk of leaks and fractures along transport pipelines.

ML algorithms are also increasingly applied to model complex hydrogen systems, such as cryogenic storage systems or electrolyzers, where they can improve performance by predicting degradation patterns and optimizing usage scenarios. Additionally, ML can assist in the integration of hydrogen with broader energy infrastructure, such as balancing hydrogen production with renewable energy availability.

To structure this rapidly developing field, ML applications in hydrogen technologies can be grouped by both hydrogen value-chain stage and ML task. In this review, we focus on four broad stages — production, transport, storage, and utilization — and on three main classes of ML tasks: prediction and forecasting, optimization and control, and discovery and design. Within each stage, we analyze how specific ML methods (e.g., tree-based ensembles, neural networks, deep learning, and hybrid models) are used to address concrete technical challenges such as efficiency, degradation, safety, and system integration.

1.5. Research questions

Given the rapidly evolving landscape of hydrogen technologies and the huge potential of data-driven techniques, understanding the role of ML (and more broadly AI) is crucial. This review seeks to assess the current state of ML applications in the hydrogen sector, focusing on trends, limitations, and opportunities for improvement. The objective is to explore how these technologies are being utilized across the hydrogen value chain and to identify emerging techniques that could unlock new efficiencies and safety measures. Hence, this review aims to highlight key gaps in the literature and provide recommendations for future research, based on the following research questions:

- RQ1: What are the current ML techniques being applied in the hydrogen sector, and which areas do they primarily target?
- RQ2: What benefits do ML applications bring to the hydrogen sector in terms of cost reduction, energy efficiency, scalability, reliability, and safety, as reported in the existing literature?
- RQ3: What are the main technical challenges and limitations of applying ML in the hydrogen sector?
- RQ4: Which emerging data-driven approaches show promise for future development?

To guide our analysis, we formulate the following working hypotheses. First, we expect ML adoption to be concentrated in data-rich subdomains, particularly hydrogen production and certain storage applications. Second, we hypothesize that explainability, hybrid modeling, and physics-informed ML are underutilized relative to the safety-critical nature of many hydrogen technologies. Third, we anticipate that performance reporting and benchmarking are fragmented, with limited use of standardized metrics and external validation.

The remainder of this review article is organized as follows: Section 2 explains the methodology used to conduct the literature review; Section 3 presents the results with detailed analysis of findings within each stage of the hydrogen value chain. Lastly, key benefits and challenges of ML applications as well as emerging trends and future research directions are discussed in Section 4.

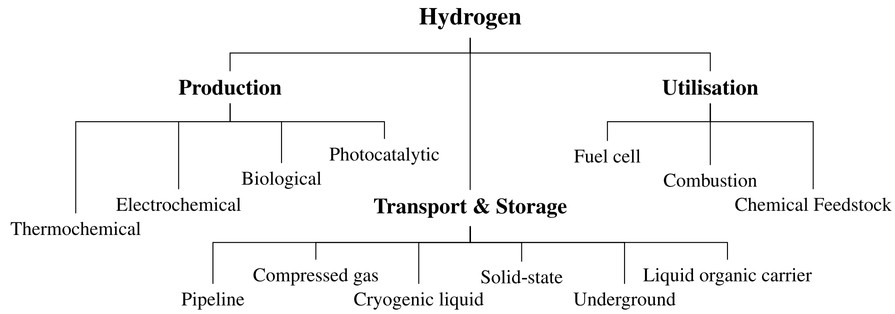

Fig. 1. Schematic representation of the hydrogen value chain.

2. Methodology

To answer RQ1–RQ4 and test the working hypotheses formulated in Section 1.5, we conducted a structured literature review in four steps. First, we defined the scope of hydrogen domains and performed a targeted database search. Second, we coded the selected studies by hydrogen domain, ML method family, and application type. Third, we appraised basic methodological quality indicators. Finally, we used this coded corpus to perform both qualitative synthesis and simple quantitative analyses. The following subsections provide a detailed description of each of these steps.

2.1. Literature search strategy

The literature search for this review was conducted using two major academic databases, namely ScienceDirect and Web of Science. These platforms were chosen for their extensive coverage of peer-reviewed articles across multiple disciplines, including the hydrogen and artificial intelligence sectors, and for their indexing of key journals such as the *International Journal of Hydrogen Energy*. The search targeted research papers published in English within the last five years (from 2019 to September 2024), reflecting the most recent developments in this area.

To capture relevant studies, various search strings and keywords were employed. The foundational search query was formulated using Boolean operators as follows: `‘‘hydrogen {domain}’’ AND ‘‘machine learning’’ OR ‘‘hydrogen {domain}’’ AND ‘‘artificial intelligence’’`, where `{domain}` is a placeholder for: `production`, `transport`, `storage`, and `usage`. Variations in terminology, such as `utilization` for usage, were included to ensure completeness.

While other databases such as Scopus or IEEE Xplore also index relevant AI- and hydrogen-related studies, a set of trial searches indicated substantial overlap with the records obtained from ScienceDirect and Web of Science. For transparency and tractability, the systematic search was therefore restricted to these two databases, acknowledging that some studies indexed exclusively elsewhere may not have been captured.

2.2. Definition of hydrogen domains

For clarity and consistency, the hydrogen value chain was divided into four key domains: production, transport, storage, and usage. Each domain represents a distinct phase in the hydrogen life cycle and is subject to different challenges that require the application of ML. A schematic representation is shown in Fig. 1.

For hydrogen production, we include four main methods, namely thermochemical, electrochemical, biological, and photocatalytic processes [7]. In this context, ML applications include prediction and optimization of reaction parameters, process efficiency, and reduction of energy consumption and overall cost. Papers that treated hydrogen only as a minor by-product in broader energy systems, without explicit modeling of hydrogen production performance, were not assigned to this domain.

For hydrogen transportation, we primarily focus on pipelines for gaseous hydrogen transit, since other forms of transportation are addressed in the storage domain. In this case, ML applications focus on leak detection, crack propagation, and optimization of compression requirements.

For hydrogen transportation, we primarily focus on pipelines for gaseous hydrogen transit, since other forms of transportation are addressed in the storage domain. In this case, ML applications focus on leak detection, crack propagation, and optimization of compression requirements.

Storage includes gaseous, cryogenic liquid, liquid carrier, solid-state, and underground storage methods. ML can be used to optimize material selection, predict material degradation, and model the most efficient storage conditions. Studies that only mentioned hydrogen storage tangentially, or focused exclusively on non-hydrogen energy carriers, were excluded from this domain.

The use of hydrogen covers applications in fuel cells, combustion, and as a chemical feedstock for processes like ammonia and methanol synthesis. ML applications include monitoring of fuel cell degradation, prediction of laminar burning velocity for combustion systems, and process optimization for chemical syntheses.

Safety and material characterization were considered as overarching themes applicable to any of the above-mentioned domains. For instance, ML applications for hydrogen production catalysts screening and design, as well as for hydrogen storage materials optimization, were considered in the review. Papers were tagged as safety- or material-focused only when hydrogen-specific properties, degradation mechanisms, or hazards formed a central part of the modeling task, rather than when hydrogen was one of many generic gases or materials.

2.3. Core machine learning methods for modeling and optimization in hydrogen technologies

Machine learning, a foundational area within artificial intelligence, encompasses a suite of algorithms that enable data-driven modeling and decision-making without requiring explicit rule-based programming. In the context of hydrogen technologies, ML methods facilitate the modeling of complex, nonlinear systems and the extraction of insights from high-dimensional or heterogeneous data sources. These methods can be broadly classified according to their learning paradigm and data requirements [8].

Supervised learning involves training algorithms on labeled datasets, where the model learns to map inputs to known outputs. This approach is particularly relevant in applications such as fault detection in hydrogen fuel cells or performance prediction in electrolyzers. Common supervised algorithms include decision trees (DT), support vector machines (SVM), artificial neural networks (ANN), and ensemble methods such as random forests (RF) and gradient boosting (GB), which are used for both classification and regression tasks.

Unsupervised learning, in contrast, deals with unlabeled data, aiming to uncover latent patterns or structures. Techniques such as principal component analysis (PCA) are used for dimensionality reduction and noise filtering, whereas clustering algorithms like k-means assist in exploratory analysis, such as identifying

Semi-supervised learning combines a limited set of labeled data with a larger corpus of unlabeled data to improve learning performance—an advantage in domains like sensor-based monitoring, where data labeling can be costly or impractical.

Reinforcement learning (RL) is suited for dynamic environments where an agent learns optimal strategies through trial-and-error interactions, guided by a reward function. RL is gaining relevance in control problems, for instance, in optimal operation of electrolyzers or hybrid energy systems.

Beyond these traditional paradigms, physics-informed and hybrid models represent a growing class of methods that integrate data-driven ML with domain knowledge. These approaches embed physical constraints or mechanistic models into the learning process, enhancing generalizability and interpretability, which is particularly valuable in safety-critical applications. In this review, we designated a study as using a hybrid or physics-informed approach when it explicitly combined an ML model with a first-principles or reduced-order physical model (e.g., through gray-box modeling, surrogate modeling of a mechanistic simulator, or inclusion of physics-based constraints and regularizers), or when it employed physics-informed neural networks.

ML is also intrinsically linked to optimization, a critical aspect of engineering system design and control. Optimization problems are typically formulated as either single- or multi-objective tasks, the latter being especially relevant in hydrogen systems where trade-offs between efficiency, costs, and environmental impact must be navigated. Multi-objective optimization yields a Pareto front of non-dominated solutions, each representing a different balance of competing objectives [9].

Building on this, solving such optimization problems — particularly in the multi-objective context — relies on advanced strategies capable of efficiently navigating complex, high-dimensional design spaces. In cases where experimental evaluations are time-consuming or resource-intensive, data-efficient optimization approaches are especially beneficial. Many of these strategies draw inspiration from natural or social systems, emulating adaptive behaviors to explore trade-offs between competing objectives. They are well-suited for generating diverse sets of Pareto-optimal solutions and are often integrated with surrogate models or used for fine-tuning system parameters. This makes them highly applicable not only in conjunction with machine learning workflows but also as stand-alone tools.

In the context of this review, each included study was assigned one or more ML method labels (e.g., ANN, convolutional neural network, random forest, support vector machine, Gaussian process, reinforcement learning, physics-informed neural network) based on the models explicitly reported in the paper or, where necessary, inferred from the mathematical formulation and implementation details. Studies that solely used simple linear regression or basic curve-fitting without any form of learning or model generalization were not tagged as ML-based.

These methodological paradigms are reflected in three broad application domains within hydrogen technologies: predictive modeling, clustering and feature learning, and optimization and decision-making. Predictive modeling tasks typically leverage supervised learning to model system behaviors, forecast performance metrics, or detect faults in system components. Clustering and feature learning applications often employ unsupervised or semi-supervised methods to explore patterns in operational data, reduce dimensionality, or enhance interpretability—crucial for sensor-rich environments or early-stage diagnostic tasks. Finally, optimization and decision-making tasks encompass both classical and ML-driven approaches, including reinforcement learning and surrogate-assisted optimization, to guide system configurations, control strategies, or resource allocation under technical and economic constraints. This tripartite framework provides a practical lens through which to assess the role of machine learning across the hydrogen value chain.

2.4. Selection criteria and scope of the review

The inclusion criteria for this review were designed to ensure that the selected papers focused on the intersection of ML techniques and hydrogen technologies. Only papers that applied ML methods to one or more stages of the hydrogen value chain (namely, production, transportation, storage, and usage) were considered. These techniques included supervised learning, unsupervised learning, reinforcement learning, multi-objective optimization, and hybrid models used to optimize various aspects across the hydrogen production, transport, storage, and usage stages.

In addition, papers were required to be directly relevant to at least one stage of the hydrogen value chain outlined above. Papers exploring cross-cutting themes such as safety and material characterization were also included, provided they had direct implications for hydrogen applications. The review focused on peer-reviewed publications from the past five years (2019–2024) to capture the most recent advancements and trends. Studies outside the direct scope of hydrogen technologies or those lacking a clear focus on ML applications were excluded. Specifically, studies related to oil refining, wastewater treatment, and nuclear energy (domains where hydrogen plays a role, but beyond the central focus of this review) were omitted.

Moreover, studies that lacked sufficient technical depth or clarity in describing the ML methods used were excluded, even if they mentioned hydrogen-related processes. Non-peer-reviewed papers, such as white papers, reports, or preprints without formal evaluation, were also excluded. Finally, papers that focused on hydrogen in isolation without discussing its intersection with ML technologies were excluded to maintain the review's focused scope.

During full-text screening, we also assessed a small set of quality indicators for each study, including whether the dataset size and provenance were reported, whether appropriate performance metrics and validation strategies (e.g., cross-validation, train/test split, external or temporal validation) were used, and whether any comparison with baseline or benchmark methods was provided. These indicators were recorded to support a transparent appraisal of the robustness of the reported ML applications and are used descriptively in the results and discussion.

2.5. Classification framework and paper selection

The classification framework for this review was structured around two main categories: the hydrogen value chain domains and the ML techniques applied to address specific challenges within each domain. Papers were categorized by their relevance to one or more stages of the hydrogen value chain, namely production, transport, storage, or usage. This allowed for a clear understanding of how ML techniques are being applied across the hydrogen value chain.

Subsequently, these papers were further classified according to the application domains (see Fig. 2) of the ML methods, as well as the ML paradigms used, such as supervised learning, unsupervised learning, reinforcement learning, (multi-objective) optimization, or hybrid methods. For each paper, we recorded the primary hydrogen domain(s), the ML task type (prediction/forecasting, optimization/control, discovery/design), and the specific algorithm family or families employed. The resulting cross-tabulation of hydrogen domains, ML methods, and application types is presented and discussed in the Results section (see Figs. 4–6).

This cross-categorization enabled the identification of which ML techniques are most prevalent and effective in each domain. This dual classification framework ensured that all relevant studies were systematically analyzed for trends, gaps, and emerging techniques.

Data extraction was conducted by reviewing each paper for detailed information on the ML technique used, the hydrogen domain targeted, and the impact of the application. A thematic analysis was conducted to

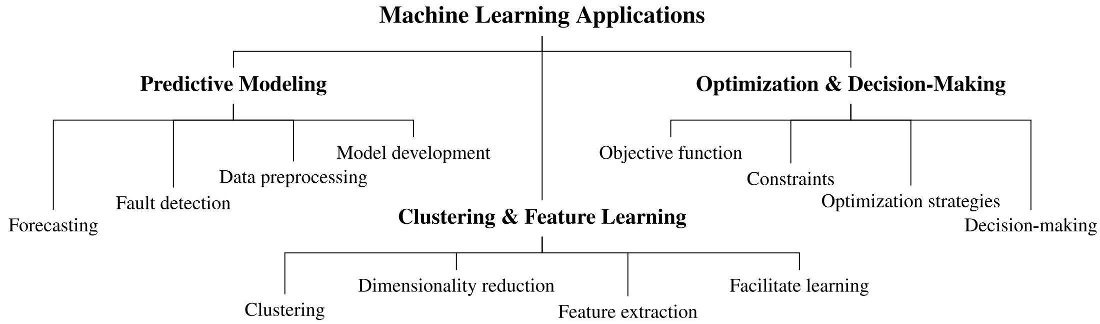

Fig. 2. Schematic representation of the different ML/AI application domains.

identify common trends across papers, with a focus on recurring challenges and benefits, such as cost reduction, efficiency improvement, or system scalability. Quantitative analysis was performed to assess the measurable impact of ML on each hydrogen domain.

Data extraction was conducted by reviewing each paper for detailed information on the ML technique used, the hydrogen domain targeted, and the impact of the application. A thematic analysis was conducted to identify common trends across papers, with a focus on recurring challenges and benefits, such as cost reduction, efficiency improvement, or system scalability. In addition to the qualitative synthesis, we performed quantitative analyses of the coded dataset, including yearly publication trends, distributions of studies across hydrogen domains, ML paradigms and application types, and simple co-occurrence statistics between domains and methods.

Basic bibliometric indicators (e.g., number of studies per journal and per country) and co-occurrence matrices for author keywords were also derived to help identify underexplored areas, such as specific method–domain combinations with few reported applications. All quantitative analyses and visualizations were carried out using standard scientific computing libraries in Python (pandas, NumPy, SciPy, and matplotlib).

The initial database search yielded over 590 papers, which were then subjected to title and abstract screening to exclude those that did not meet the inclusion criteria. After this preliminary screening, a final subset of 314 papers was selected as the basis for the review.

3. Results

The following section presents the publication search results, beginning with an overview of their distribution over the past five years and across the hydrogen value chain. This is followed by a detailed analysis of ML applications within the key domains of production, transportation, storage, and utilization.

3.1. Overview of results

The results of the publication search indicate a marked increase in ML applications in the hydrogen sector over the past five years, rising from 5 studies in 2019 to 138 by the end of September 2024.

To show this trend more clearly, a bar-chart is presented in Fig. 3 showing the number of publications per year across different segments of the hydrogen value chain. From Fig. 3, we see a sharp increase in the total volume of research papers in recent years. This growth is dominated by two segments: material characterization and hydrogen production. These areas benefit from relatively mature technological development and data availability, particularly in thermochemical (e.g., SMR) and electrochemical (e.g., electrolysis) methods, as shown in the research pathway mapping (Fig. 4).

Building on the classification framework defined in Section 2, Figs. 4–6 provide a quantitative mapping between hydrogen domains, ML application types, and method families across the 314 reviewed studies.

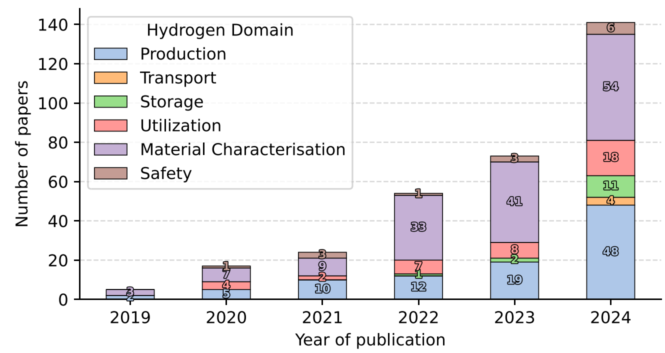

Fig. 3. Growth of ML/AI publications across the hydrogen value chain (2019–2024), color-coded by hydrogen domain. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

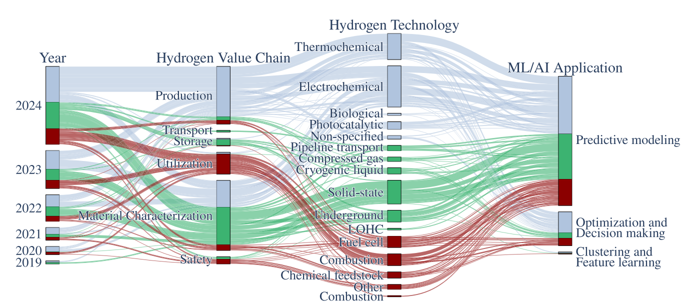

Fig. 4. Mapping research pathways: Hydrogen value chain segments, hydrogen technologies, and ML/AI applications.

In Fig. 4, we illustrate the connections between the different hydrogen domains and the ML applications. Despite the overall growth in publication volumes, hydrogen transport, storage, and utilization remain relatively underexplored, although there is a growing interest in storage, especially in solid-state and cryogenic liquid storage technologies. In contrast, pipeline transport and underground storage receive limited attention, highlighting potential research gaps.

Across all hydrogen domains, predictive modeling emerges as the most dominant ML application, primarily used to forecast system behavior, material performance, or process outcomes (Fig. 5). Fig. 5 shows publication volumes among reviewed papers using each of the ML/AI methods, including predictive modeling, optimization & decision making, and clustering & feature learning, across the hydrogen value chain. We can clearly see that the majority of the studies (78.4%) focus on predictive modeling, particularly in the production and material characterization domains, with a specific emphasis on production catalysts and solid-state storage materials. In contrast, optimization & decision-making, as well as clustering & feature learning, are less

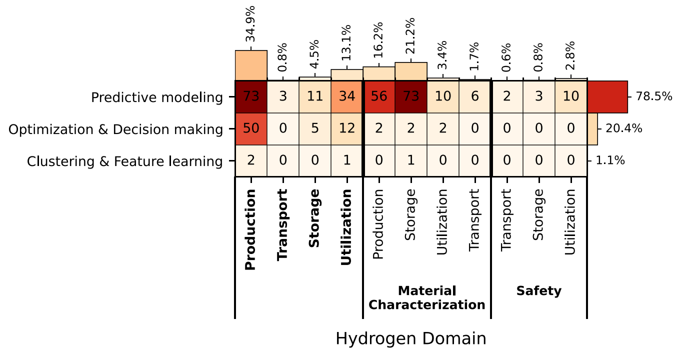

Fig. 5. Classification of the reviewed studies by hydrogen domain and primary ML application category.

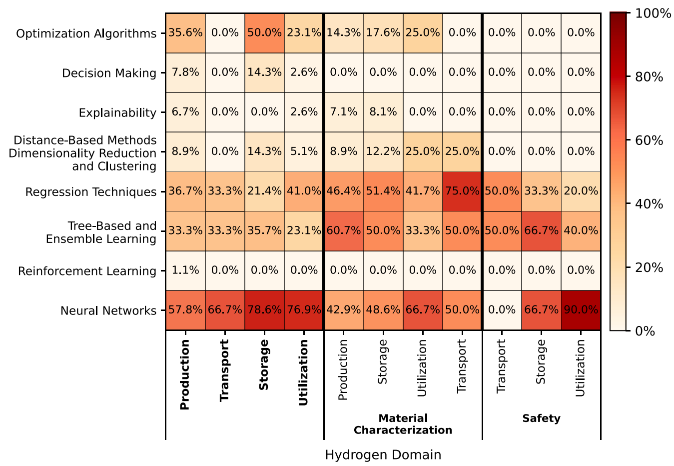

Fig. 6. Mapping of hydrogen domains to ML method families, expressed as the percentage of studies in each domain.

frequently applied, although they are gaining visibility, particularly in production and utilization research (Fig. 4).

A review of the commonly used ML techniques applied across different hydrogen domains is shown in Fig. 6, highlighting their relative usage frequencies. Fig. 6 shows that neural networks are the most widely used ML technique across all hydrogen domains, particularly dominating in production, storage, utilization, and material characterization. Regression techniques and tree-based and ensemble learning methods are also commonly applied, though less dominant than neural networks. Optimization algorithms are most frequently applied in material characterization and production, but are rarely used in transport or storage. Reinforcement learning, decision-making models, and explainability methods are noticeably underrepresented across all domains, highlighting a gap in model interpretability and real-time control applications. Overall, the figure indicates a strong methodological bias toward prediction-focused, data-intensive models, with limited adoption of interpretable or decision-support frameworks. Consistent with this, most studies report point prediction metrics such as RMSE, MAE, MAPE, or R^2, while only a small fraction of works explicitly discuss predictive uncertainty or the use of post-hoc explainability tools.

In the remainder of this section, we provide a detailed analysis and summary of the applications of specific ML/AI methods in each segment of the hydrogen value chain.

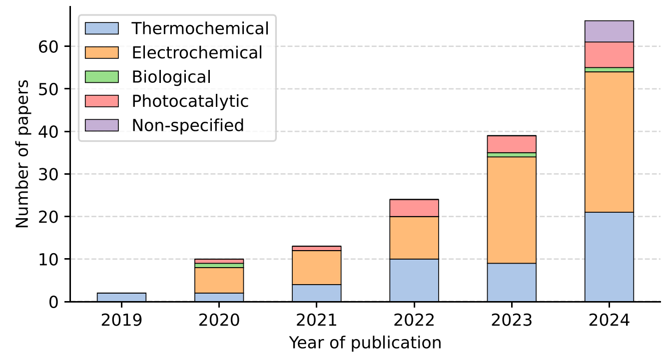

Fig. 7. Distribution of publications related to the production domain within the last five years, color-coded by production method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.2. Hydrogen production

Hydrogen production is one of the most studied segments of the hydrogen value chain. As described in Section 2.2, the hydrogen production methods considered in this review include thermochemical, electrochemical, biological, and photocatalytic production processes (Fig. 1), each characterized by different levels of operational complexity and stages of technological maturity.

Thermochemical processes include SMR, partial oxidation (POX), autothermal reforming (ATR), biomass gasification, supercritical water gasification (SCWG), chemical looping reforming, sorption-enhanced SMR (SE-SMR), and plasma reforming. These technologies are the most mature and widely implemented. SMR, in particular, dominates global hydrogen production today but faces increasing pressure to reduce its carbon intensity through the integration of CCUS technologies [1].

Electrochemical methods, particularly water electrolysis technologies such as alkaline, proton exchange membrane (PEM), solid-oxide (SO), anion exchange membrane (AEM), bipolar electrolysis, and microbial electrolysis cells (MECs), are rapidly scaling up from laboratory research to pilot-scale applications. These methods are promising for renewable-powered hydrogen production but still face challenges in improving efficiency, enhancing scalability, and reducing costs simultaneously.

Biological methods, including dark fermentation and anaerobic digestion, as well as photocatalytic water splitting, remain in the early stages of development. These methods offer the potential for low-energy and carbon-neutral hydrogen production; however, improving their yields, stability, and scalability remains a challenge.

A common challenge across all methods is the discovery of efficient, durable, and cost-effective catalysts. Catalysts are critical to improving reaction efficiency and reducing operational costs, making catalyst optimization a key research priority.

3.2.1. Trends in ML/AI applications for hydrogen production

To tackle these challenges, ML and optimization have been implemented in studies as valuable tools for improving the hydrogen production process. Fig. 7 shows the yearly distribution of ML/AI-related publications in hydrogen production, categorized by production method from 2019 to 2024. The results reveal a steady increase in research output, with electrochemical methods dominating the recent years. Thermochemical methods also show significant growth. Biological and photocatalytic methods remain less explored but show gradual increases in publication volume. Some studies do not specify the production method in detail, as indicated by the "Non-specified" category.

Table 1
Application of ML/AI methods in hydrogen production (Section 3.2).

Ref.	Year	Application	ML/AI methods	H_2 technology
[10]	2024	Process optimization	NN, GA	Thermochemical
[11]	2024	Anomaly detection	Graph AE, Graph CNN	Thermochemical
[12]	2024	Prediction of heat and H_2 production	LSTM	Thermochemical
[13]	2024	Prediction and optimization of load model	RF, LSTM	Thermochemical
[14]	2024	Process optimization	XGBoost, NSGA-II	Thermochemical
[15]	2024	Prediction and optimization of H_2 yield	XGBoost, WOA	Thermochemical
[16]	2024	Prediction of system performance	–	Electrochemical
[17]	2024	Prediction of PV power generation and H_2 production volume	RF, SVM, LSTM, NN, GA	Electrochemical
[18]	2024	Forecast of H_2 production potential	LSTM	Electrochemical
[19]	2024	Prediction of microbial electrolysis cell performance	RF	Electrochemical
[20]	2024	Prediction of H_2 production rates	k-NN, DT, RF, XGBoost	Biological
[21]	2023	Prediction of micro algal H_2 production	–	Biological
[22]	2023	Prediction of H_2 production rate	RF, XGBoost, LGBM	Photocatalytic
[23]	2022	Prediction of H_2 production	ELM, GWO	Photocatalytic
[24]	2024	Prediction of deactivation of Ni catalyst	RF, NN	Thermochemical
[25]	2022	Prediction of catalyst performance	NN	Thermochemical
[26]	2020	Prediction of catalyst deactivation	NN, RBF	Thermochemical
[27]	2024	Prediction of HER activity	RF, SVR, GPR, KRR, XGBoost, GB	Electrochemical
[28]	2024	Prediction of HER performance	NN, RF, SVM, GPR, k-NN, LR	Electrochemical
[29]	2020	Prediction of Gibbs free energy of H_2 adsorption	NN, RF, KRR	Electrochemical
[30]	2024	Prediction of HER	RF, DT, SVM, XGBoost, AdaBoost	Photocatalytic
[31]	2024	Prediction of photocatalytic H_2 yield	RF, DT, XGBoost, GB, ERT, Bagging	Photocatalytic
[32]	2022	Prediction of H_2 production rate	NN, DE	Photocatalytic
[33]	2020	Prediction of Gibbs free energy of H_2 adsorption	NN, RF, SVR, KRR	Electrochemical
[34]	2022	Prediction of stability and electrocatalytic activity of SACs	NN, RF, k-NN, GPR, SVR, CatBoost, GB	Electrochemical
[35]	2024	Screening of HER catalysts	Graph CNN	Electrochemical

Predictive modeling is the dominant application of ML/AL in hydrogen production research, with studies primarily focusing on performance prediction, process optimization, and catalyst discovery across thermochemical, electrochemical, biological, and photocatalytic methods. Across these works, model performance is typically reported in terms of regression metrics such as RMSE, MAE, MAPE, and coefficient of determination (R^2); several studies also compare multiple algorithm families on the same dataset, allowing a first, albeit heterogeneous, view of how different ML methods trade off accuracy and complexity.

3.2.2. ML/AI applications by production method

Table 1 shows a summary of recent literature on ML and AI applications in hydrogen production mentioned in this subsection.

Thermochemical Hydrogen Production. Thermochemical hydrogen production benefits from predictive models and optimization methods for hydrogen yield prediction and production process optimization. For instance, supervised learning models, such as gradient boosting (GB) algorithms, random forests (RF), and recurrent neural networks (RNN), have been employed to optimize catalyst utilization rates [10], identify system anomalies [11], and predict heat production in SMR systems [12]. Where multiple methods are benchmarked on the same thermochemical dataset, ensemble models such as XGBoost and RF generally achieve equal or lower prediction errors than simpler linear or single-tree baselines, while recurrent and other neural-network architectures are particularly useful when temporal or high-dimensional process features are available.

Optimization algorithms play a crucial role in thermochemical process optimization, helping balance production cost, energy efficiency, and emissions reduction. Algorithms such as NSGA-II and Whale Optimization Algorithms (WOA) have been used to identify optimal operational conditions, promoting cost-effective and sustainable hydrogen production in industrial-scale thermochemical processes [13–15,36]. A combination of multiple ML/AI methods, such as predictive modeling and multi-objective optimization, helps maximize hydrogen yield and operational efficiency while minimizing cost and emissions.

Electrochemical Hydrogen Production. Electrochemical hydrogen production, particularly through proton exchange membrane water electrolysis (PEMWE), is increasingly recognized as a method compatible with renewable energy sources for sustainable hydrogen generation. Current ML applications in electrochemical processes focus on system optimization and predictive modeling to facilitate integration with renewable power sources like solar and wind energy. For example, ML models have been developed to predict hydrogen production rates in PEMWE systems, enabling the dynamic optimization of wind-to-hydrogen [16] and photovoltaic (PV)-PEMWE systems [17,18]. Additionally, ML/AI techniques have been applied to analyze the performance of MECs, further enhancing electrochemical hydrogen production [19].

Commonly used ML methods in electrochemical system modeling and optimization include supervised learning models such as neural networks (NN), random forests (RF), support vector machines (SVM), and long short-term memory (LSTM) networks. Optimization of electrolysis processes is further enhanced by multi-objective optimization and genetic algorithms, which support the tuning of operational parameters for increased efficiency, lower costs, and reduced emissions. By integrating ML techniques, electrochemical hydrogen production demonstrates significant advantages in improving electrolyzer efficiency, reducing costs, and enhancing the adaptability to fluctuating renewable energy sources, supporting the realization of scalable and economically viable green hydrogen production.

Commonly used ML methods in electrochemical system modeling and optimization include supervised learning models such as neural networks (NN), random forests (RF), support vector machines (SVM), and long short-term memory (LSTM) networks. Optimization of electrolysis processes is further enhanced by multi-objective optimization and genetic algorithms, which support the tuning of operational parameters for increased efficiency, lower costs, and reduced emissions. By integrating ML techniques, electrochemical hydrogen production demonstrates significant advantages in improving electrolyzer efficiency, reducing costs, and enhancing the adaptability to fluctuating renewable energy sources, supporting the realization of scalable and economically viable green hydrogen production. A small but growing subset of studies also couples data-driven surrogates with physics-based electrochemical or power-system models, illustrating the potential of hybrid ML and physics-informed approaches to improve stability and extrapolation.

Biological and Photocatalytic Processes. Biological methods of hydrogen production, while still in the early developmental stages, offer unique opportunities for low-energy, potentially carbon-neutral hydrogen generation. However, these methods rely on complex biological systems, often governed by nonlinear kinetic relationships.

ML applications in biological hydrogen production have focused on predicting hydrogen production rates, particularly in dark fermentation and anaerobic biomass fermentation-based hydrogen production systems. By analyzing the correlations between operational parameters and hydrogen production, ML models contribute to a more comprehensive understanding of the kinetic behaviors that govern these complex biological systems.

Supervised learning techniques such as DTs, RFs, k-nearest neighbors (k-NN), and SVMs have been applied to establish relationships between biological parameters and hydrogen yield [20,21]. Through ML-driven analysis, researchers gain valuable insights into how production rates respond to varying conditions, thereby accelerating advancements in biological hydrogen production. Given that most biological datasets remain relatively small and noisy, cross-validation and careful regularization are commonly employed to mitigate overfitting, although external validation on independent experimental campaigns is still rare.

Photocatalytic hydrogen production involves the use of light-sensitive materials to split water molecules, a promising pathway for low-cost, renewable hydrogen generation. While still in an experimental phase, ML applications in this domain focus on predicting hydrogen production rates and analyzing the effects of different photocatalyst compositions. Gradient boosting techniques and RFs have been applied to identify optimal conditions and accelerate the development of photocatalytic systems by elucidating the complex interactions between system parameters [22,23]. Similar to biological hydrogen production, the integration of ML in photocatalytic production methods offers accelerated development and the opportunity to gain insights into parameter correlations that are difficult to model analytically; reported error metrics (e.g., RMSE, MAE, R^2) indicate that these models can capture the main trends in production rate, but systematic cross-study benchmarking is still lacking.

Catalyst Discovery. Catalyst development is central to all hydrogen production methods, as catalysts play a pivotal role in process efficiency and yield. ML-driven material characterization focuses on accelerating catalyst discovery, design, and optimization. Techniques such as NNs, RFs, Gaussian process regression (GPR), and gradient boosting have been used to predict catalyst properties such as activity, selectivity, and durability across SMR [24–26], hydrogen evolution reaction (HER) [27–29] and photocatalytic [30–32] processes. Integration of ML with density functional theory (DFT) allows for more accurate and rapid predictions, advancing catalyst research by enabling the exploration of new materials.

ML applications in catalyst characterization also involve analyzing existing catalyst databases to uncover correlations between structural and functional properties, aiding in the identification of promising catalyst combinations [33–35]. The main benefit of ML in this area is the accelerated development of efficient and durable catalysts, a critical factor in the scalability of sustainable hydrogen production, although more systematic reporting of dataset provenance, representativeness, and validation strategies would further strengthen confidence in the reported models.

3.2.3. Challenges and research gaps in hydrogen production

While ML applications in hydrogen production offer substantial benefits, several challenges hinder widespread adoption. For established thermochemical processes like SMR, the primary challenge is justifying the investment required to upgrade existing hardware to support plant-wide ML integration. Although ML can enhance efficiency and reduce costs, demonstrating a clear return on investment in capital-intensive industrial settings remains a challenge.

In electrochemical processes, the limited availability of quality data poses a significant challenge. Since many electrochemical processes are still transitioning from laboratory to pilot scales, data scarcity and model uncertainty constrain model accuracy and generalizability. As the availability of experimental data grows, integration of ML models

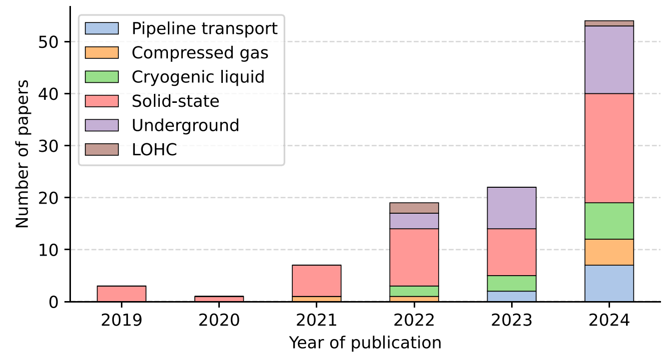

Fig. 8. Distribution of publications related to the transport and storage domains within the last five years, color-coded by transport and storage method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

with physics-based models and simulations could mitigate this issue, allowing for more reliable models and an understandable process.

For biological and photocatalytic methods, the early-stage nature of these technologies presents obstacles in both data quality and model scalability. The lack of large, validated datasets and the high variability of experimental results limit the robustness of ML models. However, as hydrogen production methods advance, the increasing volume of data will support more robust and reliable ML models. In addition, combining data-driven and physics-based approaches could yield valuable insights, enhancing the interpretability and utility of ML in hydrogen production systems.

Finally, uncertainty quantification and interpretability remain cross-cutting challenges across all production methods. Data-driven predictions are often “black-box” models, meaning their outputs lack transparency, which can reduce user trust, especially in safety-critical industrial environments. The development of “gray-box” or “glass-box” ML models, incorporating both physics-based and ML techniques, offers a promising pathway to improve model interpretability and trustworthiness, as well as industrial adoption. Finally, uncertainty quantification and interpretability remain cross-cutting challenges across all production methods. Data-driven predictions are often “black-box” models, meaning their outputs lack transparency, which can reduce user trust, especially in safety-critical industrial environments. Only a small subset of the reviewed studies explicitly report predictive uncertainty or use model-agnostic explainability tools, which makes it difficult to assess how robust the reported performance metrics would be under distribution shifts or in industrial deployment. The development of “gray-box” or “glass-box” ML models, incorporating both physics-based and ML techniques, offers a promising pathway to improve model interpretability and trustworthiness, as well as industrial adoption.

To summarize, the application of ML has transformative potential for optimizing hydrogen production processes, from well-established thermochemical processes to experimental biological and photocatalytic methods. ML can assist in the optimization of the production system as well as catalyst discovery, providing valuable insights into complex material dynamics. While challenges remain, particularly in data availability, model scalability, and interpretability, the ongoing development of hybrid ML, physics-informed approaches, combined with closer collaboration between domain experts and data scientists, could help overcome these barriers and accelerate ML/AI adoption in industrial hydrogen production and support its role in a low-carbon energy future.

3.3. Hydrogen transport and storage

Transporting and storing hydrogen present a unique set of challenges due to the distinct properties of the element. Hydrogen, the

Table 2
Descriptions of the tasks and ML methods used in the papers referenced in Section 3.3.

Ref.	Year	Task	ML methods	H_2 technology
[37]	2024	Prediction of fatigue degradation	NN, RF	Pipeline transport
[38]	2024	Prediction of temperature distribution in pipelines	GPR	Pipeline transport
[39]	2024	Leak detection	AR, LASSO	Pipeline transport
[40]	2024	Leak detection	CNN, ANFIS	Pipeline transport
[41]	2024	Prediction of heat transfer coefficient of subcooled liquid hydrogen under various flow conditions	NN	Cryogenic liquid
[42]	2022	Prediction of hydrogen flow boiling heat transfer coefficient	RF, DT, k-NN, ERT, AdaBoost, XGBoost, Bagging	Cryogenic liquid
[43]	2024	Prediction of hydrogen adsorption of functionalized carbonaceous nanomaterials	NN, DT, XGBoost, GPR	Solid-state
[44]	2024	Optimization of hydrogen absorption thermodynamics	BO	Solid-state
[45]	2024	Prediction of hydrogen storage capacity and enthalpy of hydride formation	RF, LR, RR, KRR, GPR, ERT, LASSO	Solid-state
[46]	2024	Prediction of hydrogen storage capacity	CatBoost	Solid-state
[47]	2024	Prediction of hydrogen solubility	DT, SVR, k-NN, AdaBoost	Underground
[48]	2023	Prediction of hydrogen adsorption	NN, RF, DT, SR	Underground
[49]	2024	Prediction of UHS optimal location	NN, SVM, k-NN, MLR, XGBoost, LightBoost, CatBoost	Underground
[50]	2024	Prediction of UHS performance metrics	NN	Underground
[51]	2024	Optimization of hydrogen liquefaction process	NN, PSO	Cryogenic liquid
[52]	2024	Multi-criteria investigation and optimization framework	NN, NSGA-II, TOPSIS	Cryogenic liquid
[53]	2024	Optimization of operational variables in the hydrogen liquefaction process	NSGA-II, PSO, GWO, LINMAP, TOPSIS	Cryogenic liquid

smallest and lightest molecule, exhibits a very low volumetric energy density compared to conventional fuels. This means it necessitates very high pressures or cryogenic conditions to be transported and stored efficiently, thereby posing specific engineering requirements. Hydrogen's high diffusivity, which allows it to penetrate and diffuse through materials, contributes to embrittlement and leakage risks. This characteristic, combined with its high flammability, means even minimal concentrations of hydrogen can create a flammable environment. The need for extreme caution becomes evident as leaks in pipelines or storage tanks could potentially ignite in the presence of a spark, with the risk of escalation to explosive reactions if the concentration is within the explosion limits (18.3% to 59% in air at atmospheric pressure [54]). Therefore, the practical requirements of hydrogen transport and storage systems include highly sensitive sensors capable of detecting hydrogen at very low concentrations, embrittlement-resistant materials to ensure high longevity and safety, as well as sophisticated control systems that can optimize the operation of transport and storage networks. Contemporary ML-based research in this area, therefore, focuses on quantifying and mitigating these modern, system-level risks in pipelines, cryogenic facilities, and subsurface formations.

3.3.1. ML applications in hydrogen transport and storage

In response to these technical demands, ML applications have gained traction in both hydrogen transport and storage applications (see Table 2). As shown in Fig. 8, there has been a steep increase in ML applications in hydrogen transport and storage applications during the last year, with solid-state hydrogen storage reporting the highest interest within the five years.

Pipeline Transport. In pipeline transportation, common issues such as leak detection and fatigue degradation of materials benefit from ML's predictive capabilities. For instance, supervised learning techniques like linear regression (LR), NNs, and RFs have been compared for predicting fatigue degradation, with RF models demonstrating superior accuracy [37]. Additionally, ML models such as GPR have proven effective for predicting temperature distributions in hydrogen pipelines, a critical factor for avoiding temperature-induced stress and embrittlement [38]. Other applications include the use of gradient boosting techniques for predicting crack growth in pipelines, which helps in maintenance planning and early intervention.

Leak detection, crucial in hydrogen transport due to safety concerns, also significantly benefits from ML innovations. Hybrid approaches combining physics-based and data-driven models have shown promise in real-time detection. For example, with a combination of LASSO and autoregressive (AR) models, [39] have accurately identified leak points and estimated leak rates. Similarly, the integration of a real-time transient model with Convolutional Neural Networks (CNN) and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) has enabled the development of a sophisticated leak detection system in pipelines transporting methane-hydrogen blends of varying compositions [40]. These advancements demonstrate that ML-based leak detection and maintenance forecasting can provide a more responsive approach to hydrogen transport management; in the reviewed studies, model performance is typically summarized using classification metrics (e.g., accuracy, precision, recall or F1-score) or regression errors on leak rate and location, and ML-based approaches generally outperform simple threshold- or rule-based schemes on the considered test scenarios.

Hydrogen Storage. Regarding hydrogen storage, ML techniques have been utilized across various storage types, including cryogenic liquid, solid-state, and underground storage.

Cryogenic storage of hydrogen, which requires extremely low temperatures to liquefy the gas, profits from ML models that optimize the liquefaction process by predicting heat transfer coefficient and other physical properties in subcooled conditions [41,42]. This optimization not only improves energy efficiency but also reduces operational costs, a primary goal in large-scale hydrogen applications.

Solid-state storage, which relies on materials capable of absorbing hydrogen, such as metal hydrides and MOFs, presents unique challenges that ML can address. In this domain, predictive models focusing on the adsorption behavior of hydrogen within storage media have become valuable tools [55]. Techniques like NNs, RFs, SVMs, and DTs aid in modeling and predicting adsorption capacity [43–46], an essential metric for optimizing solid-state storage materials.

Underground hydrogen storage (UHS), another viable option for large-scale seasonal storage, can also take advantage of ML integration, particularly for predicting hydrogen solubility and adsorption in various underground formations [47,48]. These parameters inform decisions on the optimal location of UHS systems [49] and help model storage performance metrics essential for reliable system operation [50].

Across these different storage tasks, reported metrics (e.g., RMSE, MAPE, R^2) consistently indicate that ML models can capture the dominant trends in heat-transfer coefficients, adsorption capacities, and solubilities, even though differences in data sources and targets currently limit direct cross-study comparisons.

3.3.2. System-level optimization and uncertainty analysis

System integration and optimization represent another area where ML has demonstrated its value across hydrogen storage solutions. Approaches like multi-objective optimization and genetic algorithms have facilitated the optimization of the entire storage system. Techniques such as Particle Swarm Optimization (PSO), Gray Wolf Optimizer (GWO), and Non-dominated Sorting Genetic Algorithm (NSGA) have been utilized to balance various operational objectives, such as minimizing cost and emissions, while maximizing efficiency [51,52]. The resulting Pareto fronts provide explicit trade-off curves between economic and thermodynamic indicators (e.g., specific energy consumption, exergy efficiency, and capital cost), enabling operators to select operating points that reflect their preferred balance between cost and performance. ML models can also be employed for uncertainty analysis to understand which variables most affect performance and consequently prioritize system improvements accordingly [53].

3.3.3. Material characterization for hydrogen storage

Accurate characterization of materials used in hydrogen transport and storage is crucial for developing embrittlement-resistant and high-capacity storage media. ML techniques play a pivotal role in predicting hydrogen adsorption behavior and assessing material performance metrics. Supervised learning methods, including SVMs, DTs, RFs, and gradient boosting, are commonly used for predicting material performance, offering a data-driven approach to optimize the design and functionality of solid-state storage materials [55]. Neural network-based models, including deep learning variants, have gained popularity due to their effectiveness in capturing non-linear relationships between structural and functional properties in storage media, thus accelerating the discovery and optimization of suitable materials for hydrogen storage.

3.3.4. Benefits of ML integration

The integration of ML in hydrogen transport and storage systems has several benefits. In transport applications, ML models offer predictive capabilities to improve pipeline operation and maintenance, enabling operators to anticipate and address issues like fatigue degradation and leak risks. Given the limited existing hydrogen pipeline transport infrastructure, ML tools provide a proactive approach to design and maintenance, with the ultimate goal of lowering operations costs, improving efficiency, and safety. Moreover, as more data on hydrogen transport becomes available, ML methods will be instrumental in refining models, generating insights from historical data, and adjusting operational protocols to ensure safe hydrogen handling. In storage applications, ML techniques contribute to faster system optimization, with the potential to improve efficiency and safety, while reducing costs. With multi-objective optimization, ML tools facilitate integrated system analysis that considers complex interdependencies between different components of the storage system and varying operating conditions. These tools enhance decision-making in areas such as network management and energy use, allowing for greater flexibility and resilience in hydrogen infrastructure.

Material characterization for hydrogen storage further exemplifies ML's impact, as these models could significantly accelerate material discovery, design, and optimization. The ability to identify correlations between structural and functional properties aids in material selection for specific storage needs, promoting faster development cycles and reducing experimental costs.

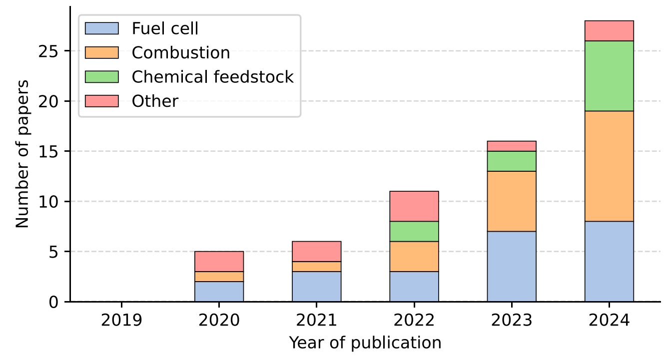

Fig. 9. Distribution of publications related to the utilization domain within the last five years, color-coded by utilization method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.3.5. Challenges and research gaps in hydrogen transport and storage

Despite the benefits, several challenges impede the full realization of ML's potential in hydrogen transport and storage. One primary issue is the lack of quality data, which ML models rely on for training. The hydrogen sector's relative immaturity means historical data are limited, while available data from literature often vary in terms of operating conditions and quality. Consistent data collection and standardization, alongside the integration of physics-based models, can help mitigate this challenge by supplementing scarce real-life data with simulated scenarios.

Another challenge lies in scalability. To date, ML techniques have been predominantly applied to lab- or pilot-scale hydrogen systems, raising questions about their accuracy and reliability at larger scales. As the sector expands, integration of data-driven models with physics-based models and existing domain knowledge will be crucial for maintaining accuracy and managing complexities throughout the scale-up of these technologies. Uncertainty in both physics-based and data-driven models is also a concern, as these uncertainties can significantly impact the reliability of predictions. Developing robust uncertainty quantification methods is essential for establishing the boundaries within which ML models can operate confidently.

Finally, the interpretability of ML models poses a barrier to wider adoption. Many ML techniques are perceived as "black boxes", limiting user trusts and hindering practical implementation. The move toward explainable ML models and hybrid approaches combining ML with physics-based insights, may offer a solution to improve the interpretability and transparency, thus boosting confidence among practitioners. Across the transport and storage corpus, however, only a limited number of studies explicitly report uncertainty estimates or use explainability tools, indicating a clear opportunity for more systematic treatment of these issues in future work.

In summary, ML applications hold considerable promise for advancing hydrogen transport and storage technologies. From predictive maintenance in pipelines to optimization of storage conditions and material characterization, these techniques address key challenges and enhance system performance. Nonetheless, the successful deployment of data-driven solutions within the hydrogen sector requires overcoming data scarcity, scalability issues, and model uncertainty, as well as the need of interpretable models. These advances are expected to play a pivotal role in meeting the demands of a growing hydrogen sector, supporting a sustainable transition to hydrogen as a viable energy carrier.

3.4. Hydrogen utilization

Hydrogen has multiple applications, each with unique operational and technological requirements. Table 3 contains recent literature in

Table 3
Descriptions of the tasks and ML methods used in the papers referenced in Section 3.4.

Ref.	Year	Task	ML methods	H_2 technology
[56]	2024	Prediction of laminar burning velocity	NN, GPR, XGBoost	Combustion
[57]	2024	Prediction of flame behavior of hydrogen-rich gas combustion	NN, RF, GPR	Combustion
[58]	2024	Prediction of turbulent flame speed	NN, RF, MLR, SVR	Combustion
[59]	2024	Prediction of emissions	NN, RF, DT, XGBoost, GB, Bagging	Combustion
[60]	2024	Analysis of dataset performance and prediction of combustion engine performance parameters	NN, GPR	Combustion
[61]	2024	Prediction of performance and emissions of dual-fuel hydrogen-diesel combustion engine	NN	Combustion
[62]	2024	Prediction of fuel cell degradation and battery aging	NN	Fuel cell
[63]	2022	Prediction of fuel cell degradation and battery aging	NN	Fuel cell
[64]	2024	System optimization	NN, NSGA-II, PSO, TOPSIS	Fuel cell
[65]	2024	Economic and efficiency evaluation/optimization	NN, GA	Fuel cell
[66]	2024	Prediction of exergy losses and methanol molar flow rate, Process optimization	NN, GPR, GA	Chemical feedstock
[67]	2024	Process optimization	RF, GPR, NSGA-II	Chemical feedstock
[68]	2023	Optimization of kinetic mechanisms (reaction rate constants)	PSO	Combustion
[69]	2024	Catalyst optimization	BO	Chemical feedstock
[70]	2024	Characterization of degree of metal-support interactions	NN, RF, k-NN, SVM, MLR	Chemical feedstock
[71]	2024	Prediction of thermal conductivity	NN, RBF	Combustion

the utilization domain, summarized in the subsequent subsections. Hydrogen is utilized in direct combustion, either as pure hydrogen or in mixtures with other fuels (e.g., methane or methane-air blends), which serves industrial heating and, potentially, internal combustion engines for transport applications. Another prominent use is in power generation through fuel cells, particularly PEM fuel cells (PEMFCs), which can be found in both stationary and transport applications. Fuel cell vehicles (FCVs) and refueling stations represent key areas of focus in this domain. Hydrogen also serves as a critical chemical feedstock for producing building-block chemicals, such as ammonia and methanol, central to the chemical industry.

Practical needs across these applications vary, but they share certain core objectives. In combustion applications, efficient and clean-burning systems for hydrogen-rich mixtures and pure hydrogen are essential. Optimizing combustion equipment to handle hydrogen's specific characteristics, like flame speed, is a crucial challenge. For fuel cells, prolonging lifetime by minimizing degradations, enhancing efficiency, and managing system components are vital priorities, particularly for PEMFCs in transport applications. In the chemical industry, hydrogen-based processes like CO_2 hydrogenation and ammonia synthesis require process optimization and advanced, poison-resistant catalysts to improve durability and reduce costs. ML plays an instrumental role in addressing these needs by optimizing combustion and fuel cell systems, improving established chemical processes, and supporting catalyst design.

3.4.1. ML in hydrogen combustion systems

Fig. 9 reports the distribution of publications within the last five years, showing a similar trend to the other domains previously analyzed, with growing ML applications in hydrogen combustion systems. Here, ML applications center around predicting complex combustion properties and optimizing system performance. For example, ML models can be used to predict the laminar burning velocity and flame behavior of hydrogen-enriched mixtures, which in turn are essential for understanding combustion dynamics [56–58]. Such predictions are critical for designing efficient and safe combustion systems. Moreover, ML models have been applied to optimize the performance and emissions of combustion engines operating on hydrogen-enriched fuels [59–61].

Supervised learning techniques, such as NN, RF, gradient boosting, RNN, and CNN, are commonly applied in hydrogen combustion studies used to model flame characteristics and emission profiles. Additionally, multi-objective optimization algorithms, such as GA, help balance multiple operational criteria, enhancing combustion efficiency while minimizing emissions. Correspondingly, the underlying studies usually report RMSE, MAE, R^2, or percentage error on targets such as laminar burning velocity, flame speed, and emission indices, and multi-objective results are often summarized via Pareto fronts that explicitly show trade-offs between efficiency and pollutants (e.g., NO_x, CO, unburned hydrocarbons). The benefits of integrating ML techniques in hydrogen combustion research include the ability to accelerate the modeling of complex system dynamics. This is particularly useful for integrating ML with CFD models, enabling more accurate and efficient simulations. By predicting flame behavior, emissions, and combustion characteristics, ML has the potential to improve the design and operation of hydrogen combustion systems, supporting their adoption in industries seeking low-carbon fuel alternatives.

3.4.2. ML in fuel cell systems

Fuel cells, especially PEMFCs, represent one of the most promising hydrogen applications in transportation and stationary power generation. ML models have been used to predict PEMFC degradation, allowing researchers and engineers to develop strategies to extend the operational life of fuel cells [62]. Accurate lifetime prediction is essential for FCVs, as it influences the cost-effectiveness and adoption of fuel cells in transportation. ML models also facilitate system optimization, improving efficiency and reducing costs through the optimization of the associated balance of plant (BoP) and system components. In fuel cell applications, neural networks and multi-objective optimization algorithms — including GWO, PSO, and NSGA-II — are found in the literature. These models have been applied to predict fuel cell degradation patterns [63], optimize the system operation [64], and identify the most efficient operational configurations [65]. Across these works, degradation and performance models typically use time-series operational features (e.g., current, voltage, temperature, humidity) and are evaluated with metrics such as RMSE, MAE, and MAPE over prediction horizons ranging from short-term dynamics to hundreds or thousands of operating hours, while system-level optimization studies report improvements in efficiency and cost indicators relative to reference operating points.

The benefits of ML integration in fuel cell applications include improved lifetime prediction and more efficient system optimization. In turn, this enables the development of better predictive maintenance schedules, performance optimization, and limits the need for costly and time-consuming physical testing. Overall, it enhances the viability of fuel cells for both stationary and mobile power applications, particularly in sectors like transportation, where system durability is paramount.

3.4.3. ML in hydrogen-based chemical production

As a chemical feedstock, hydrogen is integral to producing essential chemicals like methanol and ammonia. These processes require precise control over operating conditions to optimize yield and efficiency, making them ideal candidates for ML-based optimization. ML is used to model and optimize key processes, such as CO_2 hydrogenation and ammonia synthesis, providing rapid feedback and fine-tuning traditional simulation techniques.

ANNs and multi-objective genetic algorithms have been applied to these processes, integrating data analysis with optimization routines [66–68]. Compared to conventional simulation software, ML approaches can provide quicker, more efficient optimization, identifying optimal operating parameters faster and with greater adaptability to varying input conditions. In multi-objective settings, the resulting Pareto fronts highlight process conditions that jointly improve energy or exergy efficiency and reduce greenhouse gas emissions or undesired by-products compared to baseline operation.

In this domain, ML's primary benefit lies in the rapid optimization it enables, which is often faster than conventional simulation methods. Moreover, ML's ability to integrate data analysis with process optimization facilitates more responsive and accurate adjustments to process conditions, improving the efficiency and yield of hydrogen-based chemical feedstock processes.

3.4.4. ML in material characterization for hydrogen usage

Material characterization for hydrogen usage application involves designing and optimizing hydrogenation catalysts, predicting material properties in combustion systems, and assessing membrane degradation rates in PEMFCs. For instance, Bayesian optimization has been applied to support CO_2 hydrogenation catalyst optimization [69]. By integrating ML with physics-based methods like DFT, it is possible to accelerate catalyst design and material optimization, enabling rapid exploration of possible catalyst structures and identifying those with the most desirable properties [70].

In addition to catalyst development, ML models can help improve the understanding of hydrogen-based mixtures, assisting in designing systems with tailored thermal properties. For example, different neural networks have been used and compared to predict the thermal conductivity of hydrogen mixtures, which are critical for a safe and efficient design of combustion systems [71]. As in the production and storage domains, model performance is generally reported using regression metrics, with most studies demonstrating good agreement with experimental or high-fidelity simulation data within the range of conditions covered by the training set.

The benefits of ML in material characterization are significant, particularly in the speed and accuracy of catalyst development. ML allows for quick exploration of material databases, uncovering correlations between properties and predicting catalyst performance. This reduces the need for extensive and expensive experiments, accelerating innovation in catalyst design.

3.4.5. Challenges and research gaps in hydrogen utilization

Despite its benefits, ML still faces challenges in hydrogen use applications. A primary challenge is the scarcity of high-quality data. For emerging technologies, such as fuel cells and hydrogen combustion engines, datasets are often limited, constraining model training and limiting predictive accuracy. Established processes like CO_2 hydrogenation, instead, often lack accessible, structured, and consistent datasets suitable for ML applications.

Moreover, the complex data-driven nature of ML models can lead to interpretability issues, which may hinder their acceptance in highly regulated industries that require transparent operational models. Moreover, ML for hydrogen applications must balance between predictive accuracy and computational efficiency, especially when used in real-time settings. Lastly, the integration of ML with physics-based models presents both opportunities and challenges. On the one hand, hybrid ML-physics models offer a promising solution combining the advantages of both methods. On the other hand, developing robust hybrid models requires expertise in both ML and domain-specific knowledge, underscoring the need for interdisciplinary collaboration to address the challenges of ML application in the hydrogen sector. At the same time, only a small number of utilization-focused studies explicitly report uncertainty estimates or use explainability techniques, making it difficult to compare models across technologies and operating regimes.

In summary, ML applications are transforming hydrogen's role in the energy and chemical industry by accelerating system optimization, predicting performance and lifetime of hydrogen technologies, and supporting data-driven material characterization. Despite data availability and interpretability challenges, ongoing advancements in hybrid modeling and interdisciplinary approaches hold the potential to unlock ML's full potential in hydrogen applications.

4. Discussion

Building on the domain-specific results in Section 3, this section synthesizes the benefits and challenges of applying ML across the hydrogen value chain and consolidates the main findings with respect to RQ1–RQ4. We first summarize where ML is currently most and least established, then discuss reported benefits, cross-cutting challenges, and emerging methodological trends, and finally outline near-term priorities for future work.

4.1. Synthesis with respect to the research questions

RQ1: What are the current ML techniques being applied in the hydrogen sector, and which areas do they primarily target? The findings indicate a significant increase in the application of ML and AI methods, particularly in hydrogen production and material characterization, with storage and utilization following, and transport and underground storage remaining comparatively underexplored. Across domains, predictive modeling is the dominant application type, accounting for nearly four-fifths of the reviewed studies, while optimization and clustering/feature-learning tasks are less frequent. Method-wise, neural networks and tree-based ensembles are the most widely used techniques, complemented by support vector machines, Gaussian processes, and, more rarely, reinforcement learning and physics-informed approaches. This distribution reflects both the maturity of available data in production and materials domains and a methodological bias toward black-box, prediction-focused models.

RQ2: What benefits do ML applications bring to the hydrogen sector in terms of cost reduction, energy efficiency, scalability, reliability, and safety, as reported in the existing literature? The benefits of applying ML and AI techniques are evident across multiple domains, including enhanced process efficiency, safer operations, and accelerated materials discovery. In production, ML models support yield prediction, process control, and catalyst screening, enabling improved efficiency and reduced operating costs relative to baseline conditions. In transport and storage, ML-based fatigue and leak detection, as well as improved prediction of heat-transfer and adsorption properties, contribute to safer and more reliable infrastructures. In utilization, ML-driven combustion modeling, fuel cell degradation prediction, and optimization of hydrogen-based chemical processes help to balance efficiency, emissions, and system durability. Collectively, these advancements not only contribute to technological progress but also support broader global goals for a sustainable energy transition.

RQ3: What are the main technical challenges and limitations of applying ML in the hydrogen sector? Several technical challenges remain and recur across domains. Data scarcity and limited representativeness are prominent, especially in emerging areas such as underground storage, hydrogen pipelines, and fuel cell degradation, where long-term, high-quality datasets are rare. Scalability from lab- and pilot-scale studies to industrial deployments is not yet systematically demonstrated, and validation on independent data or under shifted operating conditions is often limited. The interpretability of ML models also remains a concern: only a minority of studies explicitly report predictive uncertainty or employ explainability tools, which hinders trust and adoption in safety-critical and regulated environments. Furthermore, a lack of interdisciplinary expertise and fragmented reporting practices (e.g., incomplete documentation of datasets, features, metrics, and baselines) pose critical barriers to integrating ML consistently across the full hydrogen value chain.

RQ4: Which emerging data-driven approaches show promise for future development? Emerging trends, such as physics-informed machine learning (PIML), digital twins, and hybrid models combining ML with physics-based approaches, represent promising directions for future research. These approaches aim to improve model accuracy, robustness, and extrapolation by embedding physical constraints, and they offer a principled route to bridge the gap between simulation and real-world application. In parallel, advances in explainable AI, uncertainty quantification, and surrogate-assisted optimization can help turn ML models from stand-alone predictors into trustworthy decision-support tools. Moving forward, realizing the full potential of ML in the hydrogen sector will require more systematically reported and shared benchmark datasets, hybrid ML–physics pipelines that are evaluated against transparent baselines, and community practices that encourage the publication of code, models, and interpretability analyses alongside performance metrics.

4.2. Cross-cutting challenges and emerging trends

The synthesis above aligns with the domain-level patterns observed in Section 3. Transport, underground storage, and utilization account for only a small fraction of the 314 reviewed studies (Figs. 3, 8, and 9), and fewer than 10% of studies in any domain explicitly report using explainability tools or uncertainty quantification (see the "Explainability" row in Fig. 6). Four cross-cutting themes emerge.

Data availability and standardization. A foundational barrier to effective ML adoption is the limited availability and heterogeneous quality of data across hydrogen-related applications. Biological and photocatalytic production, pipeline transport, and underground storage are particularly affected, with relatively few studies and small, noisy datasets. Even for more mature technologies, operational data are often proprietary and inconsistently documented, limiting robust model comparison and transferability. Addressing this will require more systematic reporting of dataset size and provenance, greater use of shared or synthetic benchmark datasets, and coordinated data-sharing initiatives between academia and industry.

Interpretability, uncertainty, and hybrid modeling. Many of the most widely used models in the corpus (e.g., deep neural networks and ensemble methods) are treated as black boxes, with limited insight into which inputs drive predictions or how robust these predictions are under shifted conditions. At the same time, only a minority of studies employ hybrid or physics-informed approaches, despite the availability of mature mechanistic models in several hydrogen domains. Expanding the use of explainability tools, predictive uncertainty estimates, and physics-informed architectures — particularly in safety-critical applications such as fuel cells, combustion, and subsurface storage — is therefore a key avenue for future work.

Scalability, generalizability, and rapid technological change. Most studies focus on lab- or pilot-scale systems, often at a single site, with limited demonstration of model robustness at industrial scale or under evolving regulatory and operating conditions. As hydrogen technologies and standards develop rapidly, models must be regularly updated, recalibrated, and revalidated to remain reliable. This highlights the need for adaptive and transferable ML approaches, as well as clearer lifecycle management practices for deployed models.

Underexplored domains and system-level integration. While production and material characterization dominate the current literature, other parts of the value chain — notably pipeline transport, underground storage, and integrated system operation — remain underexplored from a data-driven perspective. Digital twins for hydrogen plants and networks, system-level optimization frameworks, and ML-based safety monitoring represent promising but still nascent directions. Expanding ML applications in these areas, with a focus on interpretable and physics-consistent models, will be essential for the safe and effective integration of hydrogen into broader energy systems.

4.3. Future research directions

Based on the above synthesis, we highlight the following priorities for researchers and practitioners working at the intersection of ML and hydrogen technologies:

1. *Strengthen reporting and transparency*: systematically document dataset provenance, feature sets, baselines, metrics, and validation protocols, and, where possible, share minimal reproducible datasets and code.
2. *Quantify uncertainty and improve interpretability*: routinely accompany point predictions with uncertainty estimates and apply explainability tools to clarify which inputs drive model behavior, especially in safety-critical applications.
3. *Adopt and benchmark hybrid and physics-informed models*: treat hybrid ML–physics approaches as candidate baselines, particularly in domains where physical constraints are well understood but data remain scarce.
4. *Target underexplored domains and cross-domain integration*: prioritize ML studies in transport, underground storage, and system-level hydrogen integration, where our review identifies clear gaps but high potential impact.

Together, these actions provide a concrete, data-grounded agenda for advancing ML-enabled hydrogen technologies in the near term.

5. Conclusion

This review has examined the rapidly growing integration of ML and AI techniques within the hydrogen sector by systematically analyzing 314 peer-reviewed studies published between 2019 and 2024. By classifying these works across the hydrogen value chain, ML task types, and method families, it provides a consolidated picture of where ML is currently most impactful (production and material characterization) and where significant gaps remain (pipeline transport, underground storage, and system-level integration).

The review contributes three main elements to the literature on hydrogen and ML. First, it offers a cross-value-chain synthesis that maps hydrogen domains to ML tasks and methods, complementing earlier reviews focused on individual technologies. Second, it combines this taxonomy with a quantitative characterization of recent work, highlighting the dominance of predictive modeling and black-box methods, and the relative scarcity of explainable, hybrid, and physics-informed approaches. Third, it identifies concrete methodological and domain gaps — in data availability, transparency, interpretability, and hybrid modeling — and translates them into a set of prioritized recommendations for future research.

Taken together, these findings underscore that ML already plays an important role in improving the efficiency, safety, and reliability of hydrogen technologies, but that its full potential will only be realized through more robust data practices, stronger integration with physical and engineering knowledge, and targeted efforts in underrepresented parts of the hydrogen value chain.

CRediT authorship contribution statement

Robin van der Laag: Writing – review & editing, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Agnese Rizzato:** Writing – original draft, Conceptualization. **Thomas Bäck:** Supervision. **Yingjie Fan:** Writing – review & editing, Supervision.

Declaration of competing interest

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

Acknowledgments

This research is supported by the HyTROS project, funded out of the Dutch Growth Fund Program GroenvermogenNL (Green capacity for the Dutch economy and society) via the NWO call "NGF: Transport en opslag van waterstof Groenvermogen NL - Werkpakket 2". This funding is further complemented with in-kind and cash funding from several of the HyTROS participants.

References

[1] IEA. Global hydrogen review 2024. Tech. rep., Paris: IEA; 2024, URL https://www.iea.org/reports/global-hydrogen-review-2024.

[2] Government strategy on hydrogen. Tech. rep., Ministry of Economic Affairs; 2024, URL https://www.government.nl/documents/publications/2020/04/06/government-strategy-on-hydrogen.

[3] U.S. National clean hydrogen strategy and roadmap. Tech. rep., U.S. Departement of Energy; 2023, URL https://www.hydrogen.energy.gov/library/roadmaps-vision/clean-hydrogen-strategy-roadmap.

[4] Basic hydrogen strategy. Tech. rep., The Ministrial Council on Renewable Energy, Hydrogen and Related Issues; 2023, URL https://www.meti.go.jp/shingikai/enecho/shoene_shinene/suiso_seisaku/pdf/20230606_5.pdf.

[5] Ahmed A, Siegel DJ. Predicting hydrogen storage in MOFs via machine learning. Patterns 2021;2(7):100291. http://dx.doi.org/10.1016/j.patter.2021.100291, URL https://www.sciencedirect.com/science/article/pii/S2666389921001240.

[6] Sun X, Zheng J, Gao Y, Qiu C, Yan Y, Yao Z, Deng S, Wang J. Machine-learning-accelerated screening of hydrogen evolution catalysts in MBenes materials. Appl Surf Sci 2020;526:146522. http://dx.doi.org/10.1016/j.apsusc.2020.146522, URL https://www.sciencedirect.com/science/article/pii/S0169433220312794.

[7] Zhang L, Jia C, Bai F, Wang W, An S, Zhao K, Li Z, Li J, Sun H. A comprehensive review of the promising clean energy carrier: Hydrogen production, transportation, storage, and utilization (HPTSU) technologies. Fuel 2024;355:129455. http://dx.doi.org/10.1016/j.fuel.2023.129455, URL https://www.sciencedirect.com/science/article/pii/S0016236123020690.

[8] Sarker IH. Machine learning: Algorithms, real-world applications and research directions. SN Comput Sci 2021;2(3):160. http://dx.doi.org/10.1007/s42979-021-00592-x.

[9] Feng Z, Huang J, Jin S, Wang G, Chen Y. Artificial intelligence-based multi-objective optimisation for proton exchange membrane fuel cell: A literature review. J Power Sources 2022;520:230808. http://dx.doi.org/10.1016/j.jpowsour.2021.230808, URL https://www.sciencedirect.com/science/article/pii/S0378775321012994.

[10] Alotaibi FN, Berrouk AS, Salim IM. Scaling up dry methane reforming: Integrating computational fluid dynamics and machine learning for enhanced hydrogen production in industrial-scale fluidized bed reactors. Fuel 2024;376:132673. http://dx.doi.org/10.1016/j.fuel.2024.132673, URL https://www.sciencedirect.com/science/article/pii/S0016236124018222.

[11] Seo J, Noh Y, Kang Y-J, Lim J, Ahn S, Song I, Kim KC. Graph neural networks for anomaly detection and diagnosis in hydrogen extraction systems. Eng Appl Artif Intell 2024;135:108846. http://dx.doi.org/10.1016/j.engappai.2024.108846, URL https://www.sciencedirect.com/science/article/pii/S0952197624010042.

[12] Liu Z, Shi M, Mohammadi MH, Luo H, Yang X, Babaei M. An improved capacitance–resistance model for analysing hydrogen production with geothermal energy utilisation. Int J Hydrog Energy 2024. http://dx.doi.org/10.1016/j.ijhydene.2024.08.205, URL https://www.sciencedirect.com/science/article/pii/S0360319924033639.

[13] Naveena K, Krishnamoorthy M, Karuppiah N, Gouda PK, Hariharan S, Saravanan K, Kumar A. Elevating sustainability with a multi-renewable hydrogen generation system empowered by machine learning and multi-objective optimization. Meas Sens 2024;33:101192. http://dx.doi.org/10.1016/j.measen.2024.101192, URL https://www.sciencedirect.com/science/article/pii/S2665917424001685.

[14] Liu Z, Cui Z, Wang M, Liu B, Tian W. A machine learning proxy based multi-objective optimization method for low-carbon hydrogen production. J Clean Prod 2024;445:141377. http://dx.doi.org/10.1016/j.jclepro.2024.141377, URL https://www.sciencedirect.com/science/article/pii/S0959652624008242.

[15] Wang L, Li H, Du C, Hong W. Optimizing hydrogen yield in sorption-enhanced steam methane reforming: A novel framework integrating chemical reaction model, ensemble learning method, and whale optimization algorithm. J Energy Inst 2024;114:101649. http://dx.doi.org/10.1016/j.joei.2024.101649, URL https://www.sciencedirect.com/science/article/pii/S1743967124001272.

[16] Hosseini Dehshiri SS, Firoozabadi B. A multidisciplinary approach to select wind turbines for power-hydrogen production: Energy, exergy, economic, environmental under uncertainty prediction by artificial intelligence. Energy Convers Manage 2024;310:118489. http://dx.doi.org/10.1016/j.enconman.2024.118489, URL https://www.sciencedirect.com/science/article/pii/S0196890424004308.

[17] Yang Q, Ma Z, Bai L, Yuan Q, Gou F, Li Y, Du Z, Chen Y, Liu X, Yu J, Zhou X, Qian C, Liu Z, Tian Z, Zhang A, Sun K, Ding L, Tang C, Meng T, Min F, Zhou Y. Machine learning assisted prediction for hydrogen production of advanced photovoltaic technologies. DeCarbon 2024;4:100050. http://dx.doi.org/10.1016/j.decarb.2024.100050, URL https://www.sciencedirect.com/science/article/pii/S2949881324000167.

[18] Sareen K, Panigrahi BK, Shikhola T, Nagdeve R. Deep learning solar forecasting for green hydrogen production in India: A case study. Int J Hydrog Energy 2024;50:334–51. http://dx.doi.org/10.1016/j.ijhydene.2023.08.323, URL https://www.sciencedirect.com/science/article/pii/S0360319923045111.

[19] Yoon J, Cheong D-Y, Baek G. Predicting current and hydrogen productions from microbial electrolysis cells using random forest model. Appl Energy 2024;371:123641. http://dx.doi.org/10.1016/j.apenergy.2024.123641, URL https://www.sciencedirect.com/science/article/pii/S0306261924010249.

[20] Wang X, Wang Y, Tseng Y-Y, Gao Y, Li K, Wang M-H, Wu J, Lay C-H, Huo M. Integration of the grey relational analysis with machine learning for sucrose anaerobic hydrogen production prediction. Int J Hydrog Energy 2024;68:388–97. http://dx.doi.org/10.1016/j.ijhydene.2024.04.242, URL https://www.sciencedirect.com/science/article/pii/S0360319924015404.

[21] Ahmad Sobri MZ, Khoo KS, Sahrin NT, Ardo FM, Ansar S, Hossain MS, Kiatkittipong W, Lin C, Ng H-S, Zaini J, Bilad MR, Lam MK, Lim JW. Kinetic model derived from machine learning for accurate prediction of microalgal hydrogen production via conversion from low thermally pre-treated palm kernel expeller waste. Chemosphere 2023;338:139526. http://dx.doi.org/10.1016/j.chemosphere.2023.139526, URL https://www.sciencedirect.com/science/article/pii/S0045653523017939.

[22] Liu Q, Pan K, Zhu L, Zhou Y, Lu Y, Wang S, Ding Z, Du W, Zhou Y. Ensemble learning to predict solar-to-hydrogen energy conversion based on photocatalytic water splitting over doped TiO2. Green Chem 2023;25:8778–90. http://dx.doi.org/10.1039/D3GC02644B.

[23] Ren T, Ma T, Liu S, Li X. Bi-level optimization for the energy conversion efficiency improvement in a photocatalytic-hydrogen-production system. Energy 2022;253:124138. http://dx.doi.org/10.1016/j.energy.2022.124138, URL https://www.sciencedirect.com/science/article/pii/S0360544222010416.

[24] Kumbhat A, Madaan A, Goel R, Appari S, Al-Fatesh AS, Osman AI. Predicting nickel catalyst deactivation in biogas steam and dry reforming for hydrogen production using machine learning. Process Saf Environ Prot 2024;191:1833–46. http://dx.doi.org/10.1016/j.psep.2024.09.020, URL https://www.sciencedirect.com/science/article/pii/S095758202401139X.

[25] Kim C, Won W, Kim J. Early-stage evaluation of catalyst using machine learning based modeling and simulation of catalytic systems: Hydrogen production via water–gas shift over pt catalysts. ACS Sustain Chem Eng 2022;10(44):14417–32. http://dx.doi.org/10.1021/acssuschemeng.2c03136, arXiv:https://doi.org/10.1021/acssuschemeng.2c03136.

[26] Alsaffar MA, Ayodele BV, Mustapa SI. Scavenging carbon deposition on alumina supported cobalt catalyst during renewable hydrogen-rich syngas production by methane dry reforming using artificial intelligence modeling technique. J Clean Prod 2020;247:119168. http://dx.doi.org/10.1016/j.jclepro.2019.119168, URL https://www.sciencedirect.com/science/article/pii/S0959652619340387.

[27] Yang Y, Zhao X, Liu T, Zhang Y, Hu Y, Liu X, Wang G, Wang D, Bi J, Luo Z, Wang W, Yang S, Xiao W. The rational co-doping strategy of transition metal and non-metal atoms on g-CN for highly efficient hydrogen evolution by DFT and machine learning. Int J Hydrog Energy 2024;56:949–58. http://dx.doi.org/10.1016/j.ijhydene.2023.12.229, URL https://www.sciencedirect.com/science/article/pii/S0360319923065308.

[28] Wang J-N, Xu W, He J, Ma H, Cen W-L, Shen Y. Screening of transition metal dual-atom catalysts for hydrogen evolution reaction based on high-throughput calculation and searching surrogate prediction model using simple features. Appl Surf Sci 2024;659:159942. http://dx.doi.org/10.1016/j.apsusc.2024.159942, URL https://www.sciencedirect.com/science/article/pii/S016943322400655X.

[29] Fung V, Hu G, Wu Z, Jiang D-e. Descriptors for hydrogen evolution on single atom catalysts in nitrogen-doped graphene. J Phys Chem C 2020;124(36):19571–8. http://dx.doi.org/10.1021/acs.jpcc.0c04432, arXiv:https://doi.org/10.1021/acs.jpcc.0c04432.

[30] Bakır R, Orak C, Yüksel A. A machine learning ensemble approach for predicting solar-sensitive hybrid photocatalysts on hydrogen evolution. Phys Scr 2024;99(7):076015. http://dx.doi.org/10.1088/1402-4896/ad562a.

[31] Bakır R, Orak C, Yüksel A. Optimizing hydrogen evolution prediction: A unified approach using random forests, lightGBM, and bagging regressor ensemble model. Int J Hydrog Energy 2024;67:101–10. http://dx.doi.org/10.1016/j.ijhydene.2024.04.173, URL https://www.sciencedirect.com/science/article/pii/S0360319924014460.

[32] Altuner EE, El Houda Tiri RN, Aygun A, Gulbagca F, Sen F, Iranbakhsh A, Karimi F, Vasseghian Y, Dragoi E-N. Hydrogen production and photocatalytic activities from NaBH4 using trimetallic biogenic PdPtCo nanoparticles: Development of machine learning model. Chem Eng Res Des 2022;184:180–90. http://dx.doi.org/10.1016/j.cherd.2022.05.021, URL https://www.sciencedirect.com/science/article/pii/S0263876222002477.

[33] Zheng J, Sun X, Qiu C, Yan Y, Yao Z, Deng S, Zhong X, Zhuang G, Wei Z, Wang J. High-throughput screening of hydrogen evolution reaction catalysts in MXene materials. J Phys Chem C 2020;124(25):13695–705. http://dx.doi.org/10.1021/acs.jpcc.0c02265, arXiv:https://doi.org/10.1021/acs.jpcc.0c02265.

[34] Wang Y, Huang X, Fu H, Shang J. Theoretically revealing the activity origin of the hydrogen evolution reaction on carbon-based single-atom catalysts and finding ideal catalysts for water splitting. J Mater Chem A 2022;10:24362–72. http://dx.doi.org/10.1039/D2TA07167C.

[35] Zheng J, Wang S, Deng S, Yao Z, Hu J, Wang J. Accelerating the screening of modified MAZ catalysts for hydrogen evolution reaction by deep learning-based local geometric analysis. Energy Environ Mater 2024;7(6):e12743. http://dx.doi.org/10.1002/eem2.12743, e12743 EEM-2023-0573.R2. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/eem2.12743. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/eem2.12743.

[36] Mirjalili S, Lewis A. The whale optimization algorithm. Adv Eng Softw 2016;95:51–67.

[37] Campari A, Ustolin F, Alvaro A, Paltrinieri N. Inspection of hydrogen transport equipment: A data-driven approach to predict fatigue degradation. Reliab Eng Syst Saf 2024;251:110342. http://dx.doi.org/10.1016/j.ress.2024.110342, URL https://www.sciencedirect.com/science/article/pii/S0951832024004149.

[38] Yang D, Oh J, Lee G, Lee S, Choi S. Detection of hydrogen gas leak using distributed temperature sensor in green hydrogen system. Int J Hydrog Energy 2024;82:910–22. http://dx.doi.org/10.1016/j.ijhydene.2024.07.450, URL https://www.sciencedirect.com/science/article/pii/S0360319924031380.

[39] Suzuki Y, Nakayama J, Suzuki T, Soma T, Izato Y-I, Miyake A. Consequence analysis of a small-scale hydrogen leakage from the overhead hydrogen piping based on machine learning and physical modeling. J Loss Prev Process Ind 2024;90:105328. http://dx.doi.org/10.1016/j.jlp.2024.105328, URL https://www.sciencedirect.com/science/article/pii/S095042302400086X.

[40] Cristello J, Dang Z, Hugo R, Park SS. Artificial intelligence based leak detection in blended hydrogen and natural gas pipelines. Int J Hydrog Energy 2024;91:744–64. http://dx.doi.org/10.1016/j.ijhydene.2024.10.146, URL https://www.sciencedirect.com/science/article/pii/S036031992404343X.

[41] Alipour Bonab S, Yazdani-Asrami M. Investigation on the heat transfer estimation of subcooled liquid hydrogen for transportation applications using intelligent technique. Int J Hydrog Energy 2024;84:468–79. http://dx.doi.org/10.1016/j.ijhydene.2024.08.257, URL https://www.sciencedirect.com/science/article/pii/S0360319924034190.

[42] He Y, Hu C, Jiang B, Sun Z, Ma J, Li H, Tang D. Data-driven approach to predict the flow boiling heat transfer coefficient of liquid hydrogen aviation fuel. Fuel 2022;324:124778. http://dx.doi.org/10.1016/j.fuel.2022.124778, URL https://www.sciencedirect.com/science/article/pii/S0016236122016234.

[43] Wang Y, Shahbeik H, Moradi A, Rafiee S, Shafizadeh A, Khoshnevisan B, Ghafarian Nia SA, Nadian MH, Li M, Pan J, Tabatabaei M, Aghbashlo M. Predictive modeling for hydrogen storage in functionalized carbonaceous nanomaterials using machine learning. J Energy Storage 2024;97:112914. http://dx.doi.org/10.1016/j.est.2024.112914, URL https://www.sciencedirect.com/science/article/pii/S2352152X24025003.

[44] Halpren E, Yao X, Chen ZW, Singh CV. Machine learning assisted design of BCC high entropy alloys for room temperature hydrogen storage. Acta Mater 2024;270:119841. http://dx.doi.org/10.1016/j.actamat.2024.119841, URL https://www.sciencedirect.com/science/article/pii/S1359645424001940.

[45] Verma A, Wilson N, Joshi K. Solid state hydrogen storage: Decoding the path through machine learning. Int J Hydrog Energy 2024;50:1518–28. http://dx.doi.org/10.1016/j.ijhydene.2023.10.056, URL https://www.sciencedirect.com/science/article/pii/S0360319923051224.

[46] Qiu H, Xia Y, Xiang C, Xu F, Sun L, Zou Y. Prediction of hydrogen storage in metal-organic frameworks using CatBoost-based approach. Int J Hydrog Energy 2024;79:952–61. http://dx.doi.org/10.1016/j.ijhydene.2024.07.078, URL https://www.sciencedirect.com/science/article/pii/S0360319924027447.

[47] Lv Q, Zhou T, Zheng H, Amiri-Ramsheh B, Hadavimoghaddam F, Hemmati-Sarapardeh A, Li X, Li L. Modeling hydrogen solubility in water: Comparison of adaptive boosting support vector regression, gene expression programming, and cubic equations of state. Int J Hydrog Energy 2024;57:637–50. http://dx.doi.org/10.1016/j.ijhydene.2023.12.227, URL https://www.sciencedirect.com/science/article/pii/S036031992306528X.

[48] Kalam S, Arif M, Raza A, Lashari N, Mahmoud M. Data-driven modeling to predict adsorption of hydrogen on shale kerogen: Implication for underground hydrogen storage. Int J Coal Geol 2023;280:104386. http://dx.doi.org/10.1016/j.coal.2023.104386, URL https://www.sciencedirect.com/science/article/pii/S0166516223002045.

[49] Derakhshani R, Lankof L, GhasemiNejad A, Zaresefat M. Artificial intelligence-driven assessment of salt caverns for underground hydrogen storage in Poland. Sci Rep 2024;14(1):14246. http://dx.doi.org/10.1038/s41598-024-64020-9.

[50] Mao S, Chen B, Malki M, Chen F, Morales M, Ma Z, Mehana M. Efficient prediction of hydrogen storage performance in depleted gas reservoirs using machine learning. Appl Energy 2024;361:122914. http://dx.doi.org/10.1016/j.apenergy.2024.122914, URL https://www.sciencedirect.com/science/article/pii/S0306261924002976.

[51] Rehman A, Zhang B, Riaz A, Qadeer K, Min S, Ahmad A, Zakir F, Ismail MA, Lee M. Artificial intelligence-based surrogate modeling for computational cost-effective optimization of hydrogen liquefaction process. Int J Hydrog Energy 2024. http://dx.doi.org/10.1016/j.ijhydene.2024.04.331, URL https://www.sciencedirect.com/science/article/pii/S0360319924016434.

[52] Nutakki TUK, Alghassab MA, Dutta AK, Abdullaeva BS, Alkhalaf S, Alharbi FS, Ghandour R, Al Barakeh Z, Knani S. Thermo-environmental multi- investigation and ANN-based optimization of a novel heat integration criteria system integrated with a marine engine generating liquefied hydrogen. Case Stud Therm Eng 2024;56:104240. http://dx.doi.org/10.1016/j.csite.2024.104240, URL https://www.sciencedirect.com/science/article/pii/S2214157X24002715.

[53] Ghorbani B, Zendehboudi S, Alizadeh Afrouzi Z, Lohi A, Khan F. Efficient design of the hydrogen liquefaction system: Thermodynamic, economic, environmental, and uncertainty perspectives. Ind Eng Chem Res 2024;63(33):14668–99. http://dx.doi.org/10.1021/acs.iecr.3c04286, arXiv:https://doi.org/10.1021/acs.iecr.3c04286.

[54] Dagdougui H, Sacile R, Bersani C, Ouammi A. Chapter 7 - hydrogen logistics: Safety and risks issues. In: Dagdougui H, Sacile R, Bersani C, Ouammi A, editors. Hydrogen infrastructure for energy applications. Academic Press; 2018, p. 127–48. http://dx.doi.org/10.1016/B978-0-12-812036-1.00007-X, URL https://www.sciencedirect.com/science/article/pii/B978012812036100007X.

[55] Osman AI, Abd-Elaziem W, Nasr M, Farghali M, Rashwan AK, Hamada A, Wang YM, Darwish MA, Sebaey TA, Khatab A, Elsheikh AH. Enhanced hydrogen storage efficiency with sorbents and machine learning: A review. Environ Chem Lett 2024;22(4):1703–40. http://dx.doi.org/10.1007/s10311-024-01741-3.

[56] Üstün CE, Eckart S, Valera-Medina A, Paykani A. Data-driven prediction of laminar burning velocity for ternary ammonia/hydrogen/methane/air premixed flames. Fuel 2024;368:131581. http://dx.doi.org/10.1016/j.fuel.2024.131581, URL https://www.sciencedirect.com/science/article/pii/S0016236124007294.

[57] Shen Y, Morgans AS. Predicting the effect of hydrogen enrichment on the flame describing function using machine learning. Int J Hydrog Energy 2024;79:267–76. http://dx.doi.org/10.1016/j.ijhydene.2024.06.282, URL https://www.sciencedirect.com/science/article/pii/S0360319924025175.

[58] Issondj Banta NJ, Patrick N, Offole F, Mouangue R. Machine learning models for the prediction of turbulent combustion speed for hydrogen-natural gas spark ignition engines. Heliyon 2024;10(9):e30497. http://dx.doi.org/10.1016/j.heliyon.2024.e30497, URL https://www.sciencedirect.com/science/article/pii/S2405844024065289.

[59] Tian E, Lv G, Li Z. Evaluation of emission of the hydrogen-enriched diesel engine through machine learning. Energy 2024;307:132303. http://dx.doi.org/10.1016/j.energy.2024.132303, URL https://www.sciencedirect.com/science/article/pii/S0360544224020772.

[60] Farhan M, Chen T, Rao A, Shahid MI, Xiao Q, Liu Y, Ma F. Performance, emissions and combustion analysis of hydrogen-enriched compressed natural gas spark ignition engine by optimized Gaussian process regression and neural network at low speed on different loads. Energy 2024;302:131857. http://dx.doi.org/10.1016/j.energy.2024.131857, URL https://www.sciencedirect.com/science/article/pii/S036054422401630X.

[61] Khandal S, Razak A, Veza I, Afzal A, Alwetaishi M, Shaik S, Ağbulut Ü, Rashedi A. Hydrogen and dual fuel mode performing in engine with different combustion chamber shapes: Modelling and analysis using RSM-ANN technique. Int J Hydrog Energy 2024;52:973–1005. http://dx.doi.org/10.1016/j.ijhydene.2022.09.193, URL https://www.sciencedirect.com/science/article/pii/S0360319922043567.

[62] Mousavi SB, Ahmadi P, Raeesi M. Performance evaluation of a hybrid hydrogen fuel cell/battery bus with fuel cell degradation and battery aging. Renew Energy 2024;227:120456. http://dx.doi.org/10.1016/j.renene.2024.120456, URL https://www.sciencedirect.com/science/article/pii/S0960148124005214.

[63] Ahmadi P, Raeesi M, Changizian S, Teimouri A, Khoshnevisan A. Lifecycle assessment of diesel, diesel-electric and hydrogen fuel cell transit buses with fuel cell degradation and battery aging using machine learning techniques. Energy 2022;259:125003. http://dx.doi.org/10.1016/j.energy.2022.125003, URL https://www.sciencedirect.com/science/article/pii/S0360544222019004.

[64] Ghorbani B, Zendehboudi S, Alizadeh Afrouzi Z, Mohammadzadeh O. Efficient hydrogen production via electro-thermochemical process and solid oxide fuel cell: Thermodynamics, economics, optimization, and uncertainty analyses. Energy Convers Manage 2024;307:118175. http://dx.doi.org/10.1016/

j.enconman.2024.118175, URL https://www.sciencedirect.com/science/article/pii/S019689042400116X.

[65] Chen H, Ahmed OA, Singh PK, Abdullaeva BS, Alhadrawi M, Elmasry Y, Safi MS, Mahariq I. Coupling a thermoelectric-based heat recovery and hydrogen production unit with a SOFC-powered multi-generation structure; an in-depth economic machine learning-driven analysis. Case Stud Therm Eng 2024;61:105046. http://dx.doi.org/10.1016/j.csite.2024.105046, URL https://www.sciencedirect.com/science/article/pii/S2214157X24010773.

[66] Samad A, Saghir H, Musawwir A, Ahmad I, Caliskan H. A data-driven multi-objective optimization approach for enhanced methanol yield and exergy loss minimization in direct hydrogenation of CO2. Appl Therm Eng 2024;251:123517. http://dx.doi.org/10.1016/j.applthermaleng.2024.123517, URL https://www.sciencedirect.com/science/article/pii/S1359431124011852.

[67] Hara N, Taniguchi S, Yamaki T, Nguyen TT, Kataoka S. Impacts of hydrogen price and carbon dioxide emission factor on bi-objective optimizations of absorption and subsequent methanation processes of carbon dioxide capture, utilization, and storage. J Clean Prod 2024;456:142358. http://dx.doi.org/10.1016/j.jclepro.2024.142358, URL https://www.sciencedirect.com/science/article/pii/S0959652624018067.

[68] Wang H, Sun C, Haidn O, Aliya A, Manfletti C, Slavinskaya N. A joint hydrogen and syngas chemical kinetic model optimized by particle swarm optimization. Fuel 2023;332:125945. http://dx.doi.org/10.1016/j.fuel.2022.125945, URL https://www.sciencedirect.com/science/article/pii/S0016236122027697.

[69] Ramirez A, Lam E, Gutierrez DP, Hou Y, Tribukait H, Roch LM, Copéret C, Laveille P. Accelerated exploration of heterogeneous CO2 hydrogenation catalysts by Bayesian-optimized high-throughput and automated experimentation. Chem Catal 2024;4(2):100888. http://dx.doi.org/10.1016/j.checat.2023.100888, URL https://www.sciencedirect.com/science/article/pii/S266710932300489X.

[70] Qin H, Zhang H, Wang X, Fan W. Theoretical and experimental studies of CO2 hydrogenation to methanol on TM/In2O3-ZrO2 catalysts (TM=Cu, Ni and Pd): Mechanism analysis and quantitative description of MSI. J Catalysis 2024;436:115586. http://dx.doi.org/10.1016/j.jcat.2024.115586, URL https://www.sciencedirect.com/science/article/pii/S0021951724002999.

[71] Naghizadeh A, Amiri-Ramsheh B, Atashrouz S, Abuswer MA, Abedi A, Mohaddespour A, Hemmati-Sarapardeh A. Modeling thermal conductivity of hydrogen-based binary gaseous mixtures using generalized regression neural network. Int J Hydrog Energy 2024;59:242–50. http://dx.doi.org/10.1016/j.ijhydene.2024.01.216, URL https://www.sciencedirect.com/science/article/pii/S0360319924002362.