Energy Proceedings
Vol 42, 2024
ISSN 2004-2965

Trends and Patterns in Green Hydrogen Project Development: A Comprehensive Data-Driven Analysis

Joseph Junior NKOU NKOU[1,2], Cai Dongsheng[1,2*], Chiagoziem C. Ukwuoma[1,2], Anto Leoba Jonathan[1,2], Olusola Bamisile[1,2] and Qi Huang [1,2,3]

1 College of Nuclear Technology and Automation Engineering, Chengdu University of Technology, Sichuan P.R., 610059, China

2 Sichuan Engineering Technology Research Center for Industrial Internet Intelligent Monitoring and Application, Chengdu University of Technology, Sichuan P.R., 610059, China

3 Sichuan Provincial Key Lab of Power System Wide-Area Measurement and Control, School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

(*Corresponding Author: Email)

ABSTRACT

This comprehensive data science analysis of the IEA's hydrogen projects database from 2000-2022 reveals accelerating green hydrogen project growth. The descriptive analysis exposes increasing project counts, total capacity, and grid integration over time and by geography. Key countries leading in projects and capacities are Germany, Australia, the USA, France, Spain, and China. Transportation, chemicals, and refining dominate end-use applications. Solar, wind, and hydro lead renewable feedstocks, with over 50% of future capacity from novel electrolysis methods, like SOEC and other electrolysis technologies. Germany's current and future projects are electrolysis-based, unlike the USA which has a more diversified approach. Clustering uncovers project typologies centered on technology, status, location, and end-use. Sophisticated deep learning forecasting with transformer and RNN models predicts massive future growth expansion (MAE 10.19, MSE 457.74, R2 0.558). The rigorous methodological approach provides unprecedented insights into the swift expansion and budding contours of the global hydrogen sector. These data-driven models unpack project growth dynamics, offering intelligence for policy, research, and industry to strategically harness the bourgeoning hydrogen economy. Overall, this study thoroughly probes the mechanics of green hydrogen project development through ampliative modeling of the most extensive database available.

Keywords: decarbonization pathways, techno-economic analysis, multivariate forecasting, green Hydrogen, global clean energy policy, hydrogen value chain configuration

NONMENCLATURE

Abbreviations	
IAE	Applied Energy
MAE	Mean Absolute Error
MSE	Mean Squared Error
ALK	Alkaline
SOEC	Solid Oxide Electrolysis Cell
CCUS	Carbon Capture, Utilization, and Storage
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
PEM	Proton Exchange Membrane
Symbols	
R^2	Coefficient of determination
Kt	Coefficient of determination
MW	Megawatts
Nm3/h	Cubic meters per hour

1. INTRODUCTION

The rapid growth of green hydrogen projects globally has significant implications for decarbonization efforts and the future energy system. Research like [1] views Green Hydrogen as a game changer for enabling deep decarbonization across industries. Advancements in production methods, declining costs, and policy incentives have accelerated green H_2 project development worldwide[2], [3]. However, research into the precise dynamics and directionality of this

This is a paper for 15th International Conference on Applied Energy (ICAE2023), Dec. 3-7, 2023, Doha, Qatar.

burgeoning project pipeline remains limited. The authors in [4] analyzed technology trends, and the ones in[5] modeled future cost trajectories., but few studies have undertaken a comprehensive techno-economic analysis of the projects presently under development.

This research conducts an integrated assessment of global green hydrogen project growth and structure by leveraging the IEA's Hydrogen Projects Database[6]. It tracks data on projects for the production of hydrogen for energy or climate change-mitigation purposes and the development of hydrogen infrastructure, making it a valuable resource for this study.

The study employs multivariate analysis spanning descriptive statistics, unsupervised clustering, and deep forecasting to derive actionable intelligence on developer priorities and investment growth patterns. Unsupervised learning methods are becoming essential tools for analyzing large amounts of data, as they can identify patterns and structures within the data without the need for pre-existing labels or classifications[7]. Deep time-series clustering, a specific type of unsupervised learning, is particularly relevant for analyzing time-oriented data, such as the growth and development of green hydrogen projects over time[8]. Building on works like [9] on the application of techno-economic modeling, this research undertakes the first integrated data science interrogation of historic and planned green H2 projects worldwide. Techno-economic modeling is a valuable tool for assessing the economic feasibility and potential cost trajectories of hydrogen infrastructure development[10]. Meanwhile, geospatial analytics can provide insights into the spatial distribution and regional trends of these projects[11].

This research conducts a multifaceted interrogation of green hydrogen project development by leveraging the IEA's database. The integrated analytical approach reveals key insights across several dimensions:

1. The project growth trend analysis identifies Germany, Australia, and the USA as leaders in total project counts, while exposing China as the sole Asian nation in the top 10 countries. It further assesses project progress across lifecycle stages, finding over 50% remain in early feasibility or planning globally. Transportation, power, and chemicals are shown to be priority end-use applications.

2. Capacity expansion modeling forecasts growth trajectories for production capacity metrics by country. Concentrations in chemicals, refining, and heavy industry are highlighted. Over 50% of the projected 2050 capacity is attributed to novel electrolysis methods beyond traditional alkaline, PEM, and SOEC.

3. Production technology assessment compares maturity and adoption trends across renewable feedstocks and electrolysis technologies by region. It reveals the rising application of SOEC and innovative approaches.

4. Sophisticated deep learning models effectively predict massive future growth based on current expansion signals and trajectories.

5. Project clustering analysis stratifies understanding of project types across technology, locale, stage, and other attributes – enabling customized insights.

2. METHODOLOGY

2.1 Dataset

This research utilizes the IEA's Hydrogen Projects Database, the most extensive open-access global collection profiling developments in hydrogen energy production, transportation, storage, and end-use applications. The database spans more than 1900 projects initiated between 2000-2022, with information contributed by the IEA Hydrogen Technology Collaboration Programme and industry stakeholders. It covers 75 attributes including Location (country, region), Capacity details (installed, planned) across metrics (MW electrolyzer, kt H2/yr), Production pathway (electrolysis, fossil CCUS, other), Technology and feedstock specifics, Operational status (planning, construction, operation), Renewable inputs (solar, wind, hydro), End-use sector (transportation, ammonia, power gen), Developers and partners. The compilation includes executed, actively under construction, and proposed ventures exploring roles for hydrogen in decarbonization across transportation, power generation, heating, industrial processes, and innovative fuels and feedstocks. The repository offers unprecedented visibility into the global landscape of hydrogen technological maturity, geographic distribution, subsector integration, and prospective growth frontiers. As the IEA states, this living database intends to serve policymakers, researchers, and corporations seeking to strategically support the bourgeoning hydrogen economy amid the energy transition.

2.2 Data preprocessing

Preprocessing ensured data integrity for effective modeling. The raw project database encompassing 75 attributes required substantial preprocessing to enable effective modeling. First, missing values were filled and

features engineered such as project start year derived from date online. Exploratory analysis identified 17 continuous variables related to capacity metrics and end-use sectors and 3 categorical variables - status, production technology, and country - as key inputs. Aggregation generated time-series annual summaries of capacity sums and most common categories. One-hot encoding converted text categories to binary variables. Standardization normalized the scale of continuous inputs. These processed features were shaped into multivariate time series samples paired with project count target labels for supervised learning.

An 80/20 stratified split created distinct training and test partitions. The time-series data samples were reshaped into three-dimensional arrays with one timestep per annual record to match the input style expected by the RNN and Transformer neural architectures. This comprehensive pipeline ensures clean, consistent features scaled for complex deep forecasting while avoiding data leakage. The resulting preprocessed dataset provides an appropriate foundation for descriptive analytics, clustering, and predictive modeling.

2.3 Descriptive analysis

A principled data science workflow was applied to understand and prepare the data. Descriptive analysis using pandas explored univariate relationships to uncover attribute relevance. The countries were transformed from text encodings to one-hot vectors to facilitate geo-focused learning tasks. Dimensional capacity attributes were filtered for non-null records and converted to numeric types. Cumulative sums of key metrics like total capacity (MWel, nm3/h, kt/y) and project counts, were calculated over time, and various combinations were plotted to analyze emerging trends in areas like deployment growth, technology mixes, and geographic distributions over the period.

2.4 K-means clustering

Unsupervised K-means clustering was employed to group the projects into homogeneous profiles based on patterns in their attributes. The K-means algorithm aims to partition the observations into K clusters in which each observation belongs to the cluster with the nearest mean. First, the categorical features were one-hot encoded to transform them into a numeric format accepted by the algorithm. Then, the numerical capacity and end-use features were standardized using a StandardScaler to put them on a similar scale. These preprocessed features were then concatenated into a single X input matrix. Missing values in X were imputed using mean imputation. KMeans was then fit on X with K=5 clusters chosen based on domain knowledge of potential project types. Cluster labels were added to the original data frame.

The cluster centroids and within-cluster average profiles were examined to understand each cluster's defining characteristics. Dimensionality reduction techniques PCA and t-SNE projected the high-D data into 2D and 3D for visualizing the learned clusters. Specifically, these projections revealed how the observations were partitioned in the lower dimensional space. Additionally, the silhouette score was calculated to evaluate clustering quality based on how well samples are matched to their cluster versus neighboring clusters.

Overall, this unsupervised approach automatically discovered latent groups within the projects defined by underlying patterns in their multifaceted attributes. The identified clusters provide a meaningful way to profile different technological configurations and applications of hydrogen energy initiatives.

2.5 RNN and Transformer forecasting models

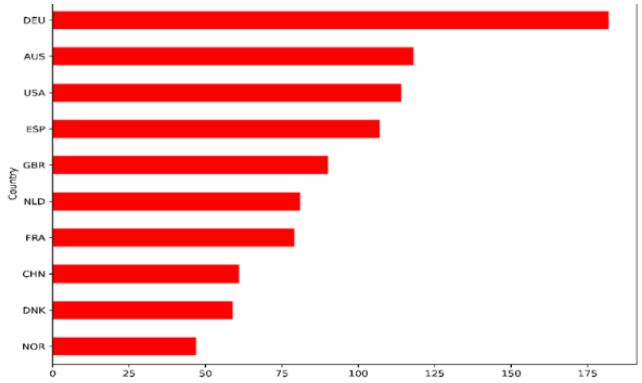

Fig. 1. Top 10 countries by project count

The RNN model used an LSTM architecture to capture the long-term temporal dependencies in the time-series data. The input layer normalized the sequences, followed by two LSTM layers of 128 and 64 units respectively. Dropout of 0.2 was applied after each LSTM layer to prevent overfitting during training. A Dense layer with 32 units and a ReLU activation function extracted higher-level features from the LSTM outputs. L2 kernel regularization of 0.01 was used on this layer. Finally, a single unit Dense layer served as the output layer. The model was compiled using the MSE loss function and Adam optimizer with a learning rate of 0.001. Gradients were calculated and weights updated during backpropagation to minimize prediction errors over several tested training epochs, using a batch size of

2 samples. Layer normalization and recurrent dropout helped address the vanishing gradients problem inherent to RNNs. The multi-layer LSTM architecture allowed the model to learn both short and long-term temporal dependencies in the input windows.

The Transformer used the self-attention mechanism rather than recurrent layers. The inputs were embedded in a 32-dim space before passing through 8 identical encoder blocks. Each block performed multi-head self-attention over the sequence to relate different positions, followed by skip connections and layer normalization. Feedforward fully connected layers with ReLU activations extracted features from the self-attention outputs. The encoder-only design mapped the inputs to hidden representations without predicting targets at each timestep. A dropout of 0.2 regularized the network. The model was trained similarly to the RNN with AdamW and a learning rate of 0.0005. Self-attention allowed the Transformer to capture global dependencies rather than relying on recurrence like the RNN.

Both models effectively modeled the temporal patterns, with their different architectural designs - recurrent layers vs self-attention - capturing dependencies in distinct ways for the time series forecasting task.

3. EXPERIMENTS AND RESULTS

3.1 Grow trends

Exploratory data analysis quantified growth trends across key dimensions including project counts, capacity, and technology. Filters extracted continuous features related to project status, end-uses, capacity, and aggregation consolidated metrics by country and year. Visualizations included various charts ranking and they reveal geographic concentrations and gaps, capacity, and technology trends. Cumulative summation models tracked installed capacity over historic and future years. In summary, data filtering, timed aggregation, and multivariate visualization expose expansion trends of the emergent green hydrogen project landscape.

The current status of hydrogen projects provides insights into the development of this critical energy sector. As shown in Fig. 1, Germany has established the largest project portfolio to date, leveraging its expertise in developing renewable energy and energy systems. Australia and the US closely follow demonstrating commitment from industrialized nations. Notably, China is the only non-western country in the top 10 countries, despite being a relative newcomer, signaling recognition

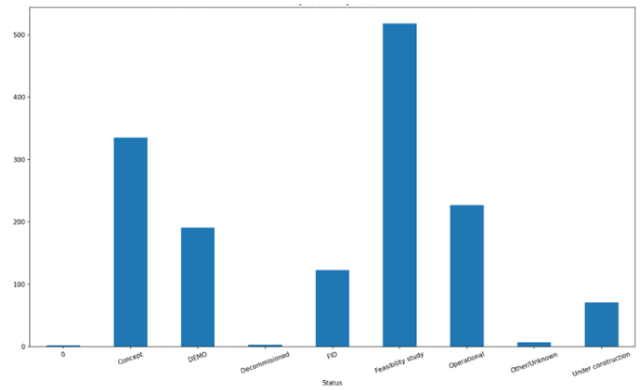

Fig. 2 Project count per status

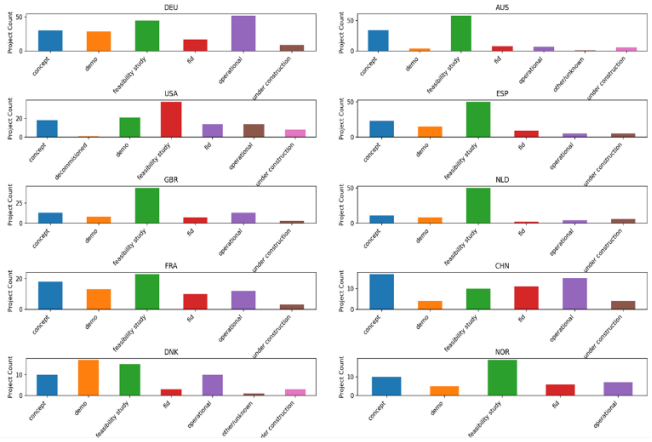

Fig. 3 Project count per country and status

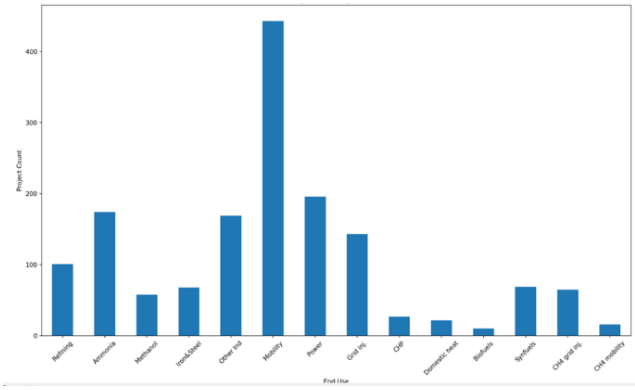

Fig. 4 Project count per end use

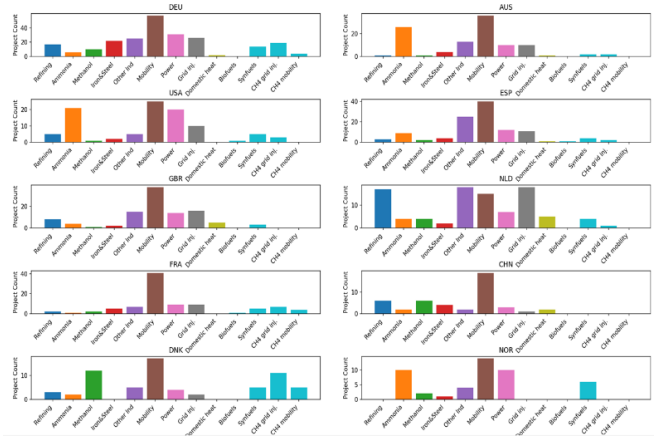

Fig. 5 Project count by country and per end use

of hydrogen's strategic importance for decarbonization across key economies.

A deeper analysis provides context on green hydrogen's immaturity. Fig. 2 shows that most projects remain in early feasibility/concept phases, underscoring nascent deployment ambitions. Feasibility assesses initial viability pre-pilot funding while concepts signal preliminary design pre-financing, together dominating comprehensive data pipelines. In contrast, demonstration/operational volumes imply most ideas lack piloting to derisk technologies at scale before rollout, and limited operations reflect nascent scale-up, suggesting ambitions exist but hydrogen may require substantial continued support to accelerate progress on capacities by addressing hurdles to realization beyond concepts/studies.

As shown in Fig. 3, Germany leads in operational projects reflected in its renewable experience, followed

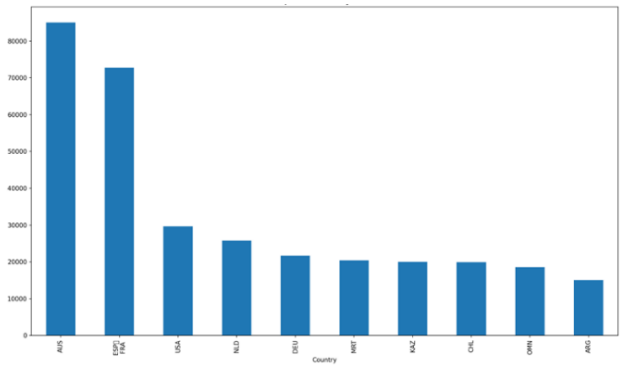

Fig. 6 Total capacity per country

by France and somewhat surprisingly China despite entering the sector later, pointing to the aggressive of early footholds. While feasibility studies still comprise a substantial portion across all top countries as expected in this nascent technology phase, examination of construction suggests imminent growth areas with the

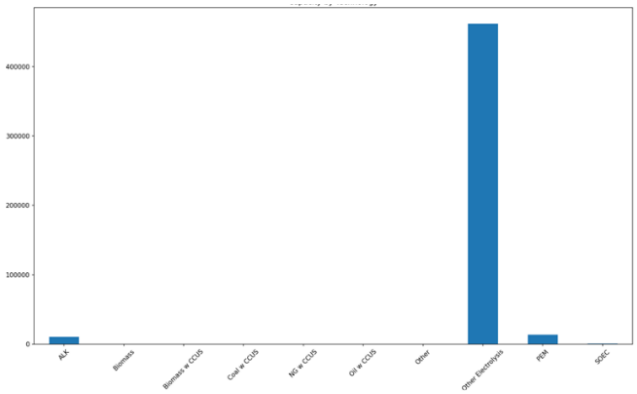

Fig. 7 Capacity by technology

US, Norway, and China attracting the most underway projects, indicating where sizable near-term capacity

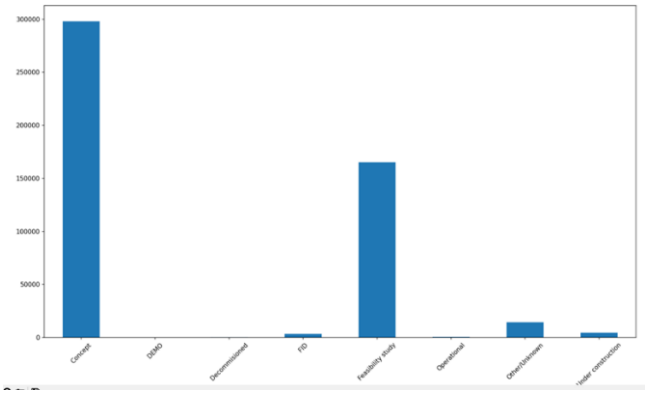

Fig. 8 Capacity by status

additions may come online especially as the US and China invest billions[12] in infrastructure with hydrogen components, overall reflective of relative experience to date for leaders yet substantive feasibility and

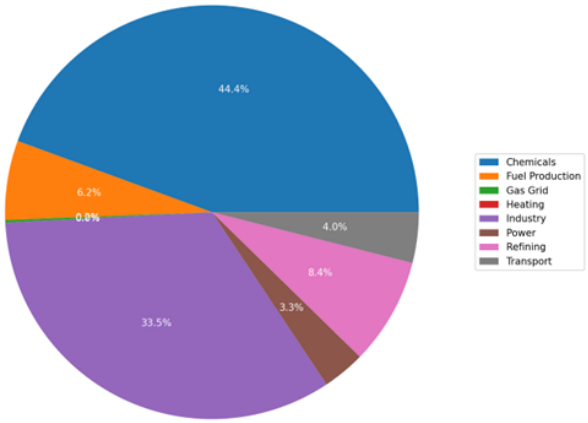

Fig. 9 Capacity per end use

construction activity across major economies indicating the landscape may be primed for faster scaling.

Analysis of project applications by industry as shown in Fig. 4 and Fig. 5, reveals transport as the sector with the most initiatives, reflecting hydrogen's potential to decarbonize hard-to-electrify modes like shipping, aviation, and long-haul trucking[13]. Power and grid injection make up the next most traction of projects, highlighting hydrogen's value for increasing RES integration through energy storage[14]. Together, these three sectors - transport, power, and ammonia production for fertilizer - account for over 60% of projects, initially prioritizing decarbonization of emissions-intensive activities.

As depicted in Fig. 6, Fig. 7, and Fig. 8, projected future capacities provide insights into both the scale of deployments envisioned and the status of planning efforts. Australia, France, and Spain have ambitious targets to become leaders in installed green hydrogen capacity, with plans for 27.5GW, 20.5GW, and 15GW

respectively by 2030[15]. Reaching these levels would require enormous investments in electrolyzer manufacturing and renewable energy buildouts to supply demand. Strikingly, over 500GW of the projected 2050 capacity - equivalent to 3000kt/yr or 10000nm3/yr - remains in early feasibility and concept planning stages.

This underscores the tremendous scale of deployment needed globally as well as the uncertainties regarding long-term technological progress to realize projections. The distribution of planned capacities in Mwel, Kt H2/y, and nm3 H2/y is globally consistent, with

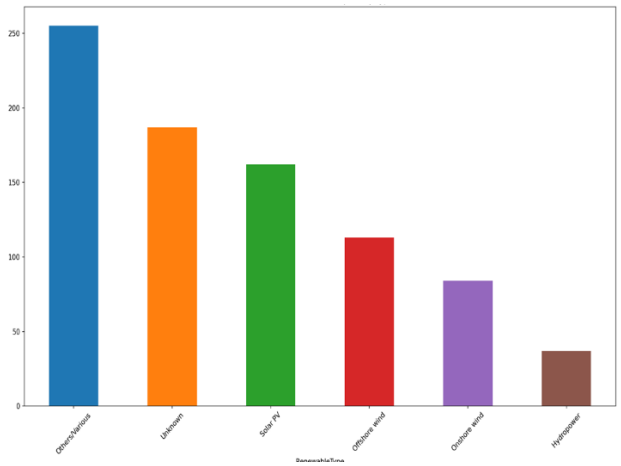

Fig. 11 Number of projects by renewable energy types

"other electrolysis" projects targeting novel methods beyond alkaline, PEM, and SOEC holding the largest share. Additionally, concept proposals holding the greatest total planned capacity followed by feasibility studies indicate consistent visioning efforts are underway, proving projections for sizable development of hydrogen energy projects.

As shown in Fig. 10, Fig. 11, Fig. 12, and Fig. 13, electrolysis technologies are positioned to enable over 50% of projected 2050 green hydrogen volumes, with novel types beyond alkaline and PEM poised to dominate projects globally according to comprehensive analyses. Specifically, "other electrolysis" constitutes the highest

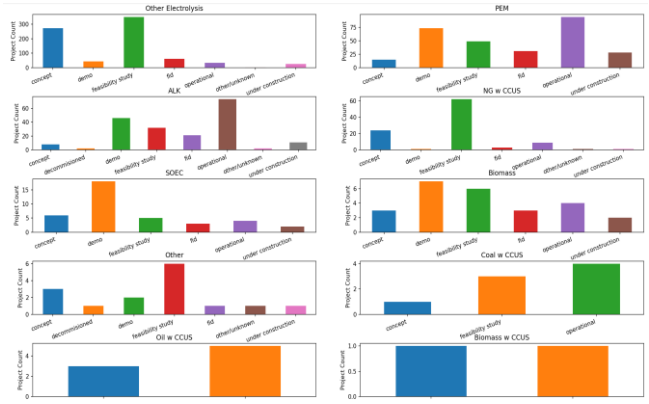

Fig. 12 Project counts per technology and status

share of initiatives, followed by PEM and alkaline projects, while SOEC demonstration activity indicates its promise. PEM and alkaline are the most implemented in current operations, with their maturity underscored[16]–[18], whereas SOEC and other novel approaches prevailingly remain in earlier feasibility and concept planning stages. Examining preferred renewable energy inputs confirms solar, offshore wind, onshore wind, and hydropower as the most prominent green hydrogen production pathways. Insights into country-level emphases expose Germany's leadership predominantly utilizing established alkaline in addition to SOEC and PEM in operational projects, aligned with its

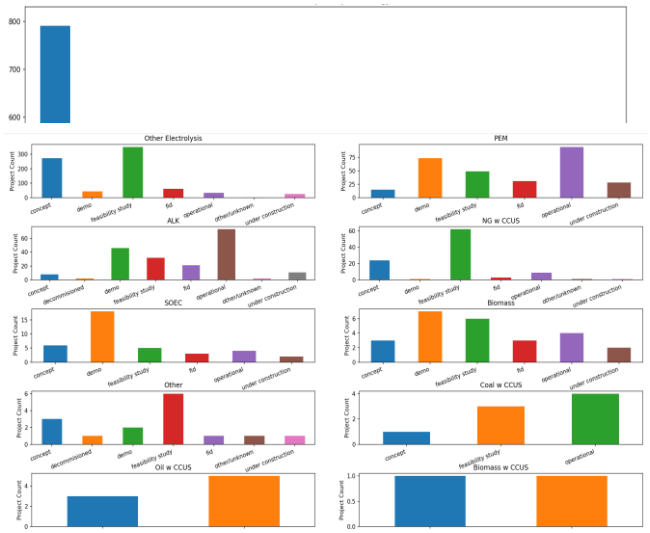

Fig. 13 Project counts per technology and country

electrolytic focus, contrasting the U.S.'s relatively lower electrolysis representation but higher non-electrolytic alternative initiatives in line with divergent strategic priorities. Collectively, findings underscore the pivotal importance of advancing all electrolysis technologies commercially.

3.2 Project clusters and attributes

Unsupervised machine learning provided novel insights into global green hydrogen initiatives. K-means clustered the multidimensional dataset into five groups without preconceptions. Analysis of centroids and internal properties elucidated the distinct phenotypes

Table 1 Overview of cluser centroids

	Technology_ALK	Technology_Biomass	Technology_Biomass w CCUS	Synfuels	CH4 grid inj.	CH4 mobility
0	0.016632	0.006237	$2.079002e^{-03}$	0.0	0.0	0.0
1	0.000000	0.000000	0.000000	0.0	0.0	0.0
2	0.068966	0.012931	$2.155172e^{-03}$	0.0	0.0	0.0
3	0.000000	0.000000	0.000000	0.0	0.0	0.0
4	0.300971	0.031068	$2.385245e^{-18}$	0.0	0.0	0.0

Table 2 Overview of cluster profiles

	Mwel	Nm3H2/h	ktH2/y	tCO2captured/y	Biofuels	Synfuels	CH4 grid inj.	CH4 mobility
0	422.53	1.019089e+05	79.453	1.4133+06	1.0	1.0	1.0	1.0
1	67000	1.488889e+07	11608	Nan	Nan	Nan	Nan	Nan
2	310.19	9.646703e+04	765.09	1.5670e+06	1.0	1.0	1.0	Nan
3	11795	2.777305e+06	2165.3	2.0000e+07	Nan	1.0	Nan	Nan
4	29.471	7.218476e+03	6.2228	6.4444e+05	1.0	1.0	1.0	1.0

defining each cluster. Manifold learning techniques visually represented these separations by embedding the high-dimensional space into lower dimensions. Examining cluster-specific parameters revealed patterns in technologies, development, and applications. Overall, this computational analysis revealed the inherent

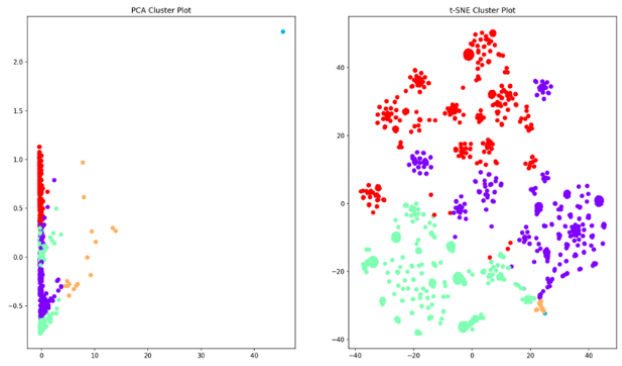

Fig. 14 PCA and t-SNE Cluster

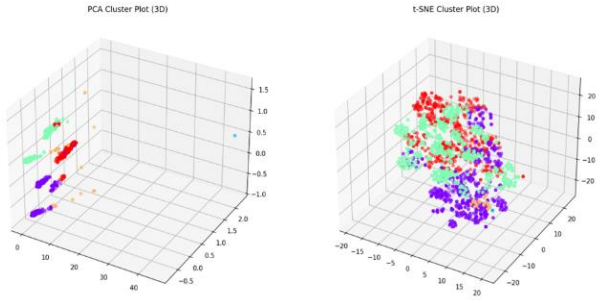

Fig. 15 PCA and t-SNE 3D Cluster

structures within this emerging field. Such a systems-level understanding of clustered architectures can guide strategic priorities to optimize the transition to widespread green hydrogen adoption across industries. Data-driven clustering uncovered informative perspectives that may help facilitate large-scale energy transformation.

The cluster analysis reveals five distinct groups of green hydrogen projects based on their characteristics. Cluster 1 centers around ALK and biomass-based production technologies, suggesting a focus on these approaches. Cluster 2 stands out for its extremely high capacity of over 100 MW, representing the largest mega-scale projects. Cluster 3 also contains very large projects

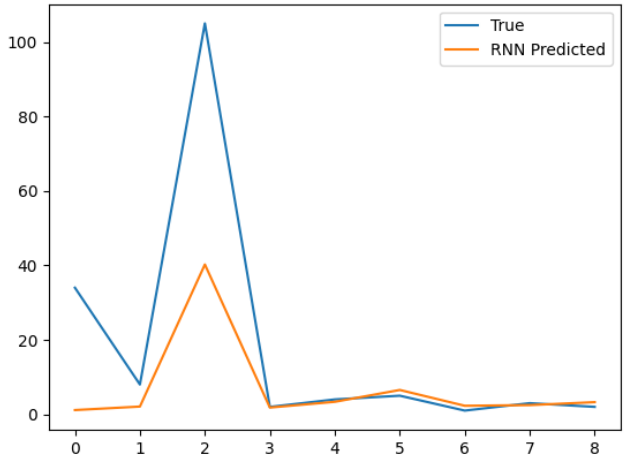

Fig. 16 RNN true vs predicted

but it is more differentiate by even more capacities than Cluster 2. Clusters 0, 2 and 4 show nonzero values for multiple end-use applications spanning biofuels, synthetic fuels, grid injection, and others, signaling diversity in planned applications.

Meanwhile, Cluster 1 has null values for all end uses shown. Finally, while the silhouette score of 0.1576 indicates reasonably separated groupings, it is not exceptionally high, leaving some possibility for overlap between clusters. Overall, the analysis demonstrates the inherent segmentation of the project landscape according to technology pathways, project scale, and end-market orientation. Distinct clusters corresponding to these themes provide insight into current trends shaping the development of the green hydrogen sector.

3.3 Forecasting accuracy and future trajectories

3.3.1 RNN

The hydrogen projects database was loaded and preprocessed by one-hot encoding categorical features, standardizing numerical features, and aggregating the data into a yearly time series profile which was then split into training, validation, and test sets. An RNN architecture utilizing LSTM units with layer normalization, dropout regularization, and hyperparameter tuning of LSTM widths/depths and dropout rates was developed and trained for 500 to 5000 epochs on 1D input sequences using Adam optimization with validation loss tracking. The best parameters are presented in Table 1. Following training, the RNN made predictions on blinded test sequences and performance was rigorously evaluated using MAE, MSE, and R2 error metrics, with true vs predicted plots and loss over epochs also generated to qualitatively assess prediction errors and convergence, establishing through this systematic process an effective recurrent model for capturing

temporal trends in the sector based on the end-to-end

Table 4 Transformer training parameters

Epochs	Batch size	MAE	MSE	R^2 Score
1300	1	11.02	475.679	0.540
1150	1	10.92	467.079	0.548
1250	**1**	**10.18**	**457.735**	**0.557**

experimentation covering data processing, model development, optimization, and evaluation.

Several trainings were conducted varying the number of epochs and batch size to explore the effect of hyperparameters on model performance. The model was evaluated using mean absolute error, mean squared error, and R2 score computed on the validation set, with lower error and higher R2 indicating better performance. Results showed slightly better scores for the 3000-epoch training with a batch size of 2. Loss plots decreased over epochs for training while leveling off near completion, and true vs predicted plots (Fig. 16) showed reasonable agreement along the identity line. While performance was reasonably good and stable across parameters with R2 around 0.46, room for improvement remains, suggesting further exploration of more complex RNN architectures and hyperparameter tuning may be needed to achieve stronger time series forecasting ability.

3.3.2 Transformer

The Transformer architecture consisted of 8 encoder blocks with each block containing multi-head attention, feedforward, skip connections, and normalization sublayers. An embedding layer was used to embed the input features. The model was trained on a time series dataset split into 80% train and 20% validation sets. Several training was done on several parameters and the results are in Table 4. Various hyperparameters were chosen, including 8 attention heads, 32 units, and a 0.0002 dropout rate. The trained model was then evaluated on a held-out test set to make predictions and compute performance metrics such as MAE, MSE, and R2 score, to analyze the model's predictive ability. Additionally, true vs predicted plots on the test set were generated to visualize prediction accuracy, and loss curves plotting training and validation loss over epochs were produced to examine model convergence.

Table 3 RNN training parameters

Epochs	Batch Size	Validation Split	MAE	MSE	R^2 Score
2000	4	0.2	12.42	592.56	0.4275
2300	2	0.2	11.80	552.87	0.4658
2600	**4**	**0.2**	**11.75**	**552.82**	**0.4659**
3000	**2**	**0.2**	**11.70**	**553.51**	**0.4652**

The experiments aimed to develop an end-to-end Transformer architecture for time series forecasting and analyze its predictive performance and learning behavior through evaluation metrics and plots, providing insights into how well the model learned patterns in the time series data.

The transformer model outperformed the RNN model and the best-performing model configuration was with a batch size of 1, run for 1250 epochs. It achieved a mean absolute error (MAE) of 10.1859 on the validation set. A lower MAE indicates better accuracy in predictions. The mean squared error (MSE) was 457.735, also showing low error. Most importantly, the R2 score was 0.5577, demonstrating a strong fit of the model to the variability in the time series data. An R2 score closer to 1 is ideal. The plot in Figure 17 displays the actual test values alongside the model predictions. We can see a close alignment between the predicted and true values, clustering around the identity line. This confirms the accuracy of the predictions made by the model. The loss curve for this model is displayed in Figure 18. The validation loss steadily decreases over the 1250 training epochs, leveling off near the end, showing that the model was still learning from the data. The low validation loss achieved suggests the model was not overfitting.

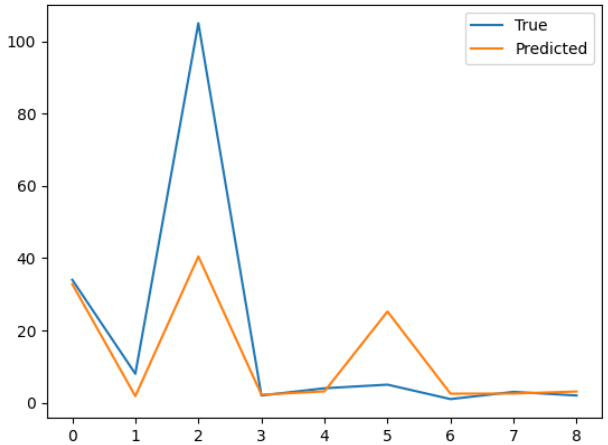

Fig. 17 Transformer true vs predicted

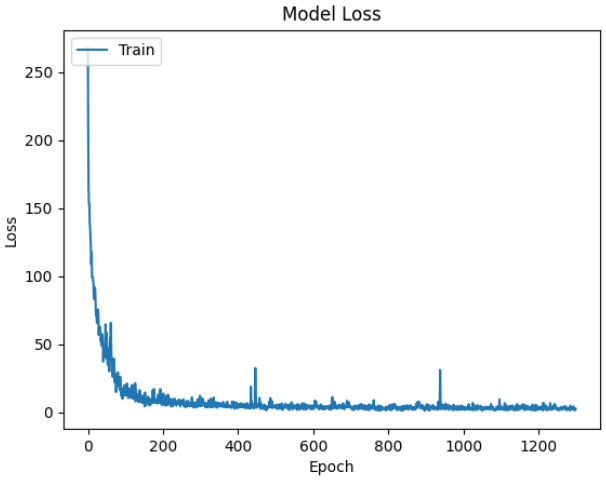

Fig. 18 Transformer loss function

4. DISCUSSION

4.1 Policy guidance based on gaps and opportunities identified

The analysis revealed several key gaps and opportunities that policymakers should consider to help accelerate the development of green hydrogen. Feasibility and conceptual projects currently dominate the pipeline, suggesting deployment ambitions exist but many ideas lack piloting to de-risk technologies at scale before rollout. Additionally, the majority of planned capacity remains in the early feasibility and concept planning stages, indicating uncertainties around long-term technological progress and the tremendous scale of deployment needed globally by 2050. Transport, power generation, and ammonia production currently comprise over 60% of projects, however emerging applications such as marine and aviation present major opportunities for growth with support. While novel electrolysis technologies beyond alkaline and PEM are poised to enable over half of projected 2050 volumes, they predominantly remain in early feasibility and conceptual stages. Increased policy and funding priority for the demonstration and commercialization of promising new electrolysis technologies could help realize their large projected contribution. Coordination between countries on issues like shared renewable energy infrastructure, cross-border hydrogen transport, and certification standards could help maximize efficiencies and cost reductions through international collaborations and knowledge sharing between frontrunner countries on lessons from operational projects. Targeted policies and investments in leading markets such as Germany, the US, Norway, and China could accelerate near-term capacity additions through diverse technology pathways.

4.2 Limitations: Database Dependencies, Uncertainty

While the analysis provided valuable insights, there are important limitations to acknowledge. As the dataset primarily includes announced projects still in the early feasibility and planning stages, key details and timelines are uncertain, and actual deployment may diverge significantly as technologies progress and projects are modified or canceled. Additionally, the findings rely on a dataset that has inherent limitations in its completeness and accuracy, as not all hydrogen projects worldwide have necessarily been captured. Important sub-national variations or qualitative factors influencing development trajectories are also difficult to fully capture given inconsistencies in project attribute reporting across different regions and countries within the dataset. There is also a high degree of uncertainty around long-term capacity projections due to unpredictable technological changes. As a result, this initial snapshot of the evolving hydrogen landscape based on the current dataset will need ongoing refinement and updates to improve insights, as both represented projects and the overall understanding of this emerging sector continue to progress and develop over time.

5. CONCLUSION AND FUTURE WORK

5.1 Summary of insights gained

This comprehensive study examined trends in the development of global green hydrogen projects through a data-driven analysis of the IEA's Hydrogen Projects Database, gaining several key insights. Project growth was found to be accelerating worldwide with over 50% initiated since 2020, led by Germany, Australia, the USA, and China, however, the majority remain in early planning stages. Transportation, power generation, and chemicals were shown to dominate current end-uses, though emerging applications in aviation and shipping present opportunities. Novel electrolysis technologies beyond alkaline and PEM were revealed to be poised to enable over half of the projected 2050 capacity, but more support is needed for their demonstration and commercialization. Countries like Australia, France, and Spain were identified as having ambitious 2030 capacity targets, but over 500GW of projections were determined to remain early-stage, highlighting deployment challenges. Clustering uncovered distinct project profiles centered on technology, scale, location, and end-use, providing a system-level understanding of sector architectures. Sophisticated forecasting models were also found to effectively predict future growth based on

current expansion trends, with the Transformer outperforming the RNN architecture.

5.2 Future work

Several promising avenues for extending the analysis were identified. Conducting a techno-economic analysis incorporating projected capital and operating costs for different production pathways out to 2050 under varying policy scenarios would provide valuable insights into the long-term cost competitiveness of technologies. Developing an optimization model to determine optimal locations and configurations for future hydrogen hubs and transport infrastructure based on renewable resource availability, demand centers, and pipeline routing costs would help guide infrastructure planning. Additionally, generating probabilistic/quantile forecasts with models like LSTM-ENNs could produce probabilistic forecasts accounting for uncertainty in projections. Analyzing the models to understand the drivers of predictions would improve the interpretability of results. Deployment of an online platform for interactive forecasts and scenario analysis would allow end users to leverage the forecasts for real-world decision making. Together, these extensions would contribute important strategic considerations around technology pathways, infrastructure planning, risk assessment, transparency, and operational deployment to facilitate the large-scale energy transformation.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the support of the National Natural Science Foundation of China (NFSC, Grant No. 52007025) and the Science and Technology Support Program of Sichuan Province (2022JDRC0025).

DECLARATION OF INTEREST STATEMENT

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. All authors read and approved the final manuscript.

REFERENCE

[1] G. D. Sharma, M. Verma, B. Taheri, R. Chopra, and J. S. Parihar, “Socio-economic aspects of hydrogen energy: An integrative review,” Technol Forecast Soc Change, vol. 192, p. 122574, Jul. 2023, doi: 10.1016/j.techfore.2023.122574.

[2] BloombergNEF, “Hydrogen Subsidies Skyrocket to $280 Billion With US in the Lead,” https://about.bnef.com/blog/hydrogen-subsidies-skyrocket-to-280-billion-with-us-in-the-lead/.

[3] International Renewable Energy Agency., Green hydrogen for industry : a guide to policy making.

[4] M. Yue, H. Lambert, E. Pahon, R. Roche, S. Jemei, and D. Hissel, “Hydrogen energy systems: A critical review of technologies, applications, trends and challenges,” Renewable and Sustainable Energy Reviews, vol. 146, p. 111180, Aug. 2021, doi: 10.1016/j.rser.2021.111180.

[5] T. Weidner, V. Tulus, and G. Guillén-Gosálbez, “Environmental sustainability assessment of large-scale hydrogen production using prospective life cycle analysis,” Int J Hydrogen Energy, vol. 48, no. 22, pp. 8310–8327, Mar. 2023, doi: 10.1016/j.ijhydene.2022.11.044.

[6] Source: IEA (2021), “Hydrogen Projects Database, https://www.iea.org/reports/hydrogen-projects-database. All rights reserved,” Hydrogen Projects Database, https://www.iea.org/reports/hydrogen-projects-database. All rights reserved. 2022.

[7] A. Glielmo, B. E. Husic, A. Rodriguez, C. Clementi, F. Noé, and A. Laio, “Unsupervised Learning Methods for Molecular Simulation Data,” Chem Rev, vol. 121, no. 16, pp. 9722–9758, Aug. 2021, doi: 10.1021/acs.chemrev.0c01195.

[8] A. Alqahtani, M. Ali, X. Xie, and M. W. Jones, “Deep Time-Series Clustering: A Review,” Electronics (Basel), vol. 10, no. 23, p. 3001, Dec. 2021, doi: 10.3390/electronics10233001.

[9] C. YANG and J. OGDEN, “Determining the lowest-cost hydrogen delivery mode,” Int J Hydrogen Energy, vol. 32, no. 2, pp. 268–286, Feb. 2007, doi: 10.1016/j.ijhydene.2006.05.009.

[10] R. Caponi, E. Bocci, and L. Del Zotto, “Techno-Economic Model for Scaling Up of Hydrogen Refueling Stations,” Energies (Basel), vol. 15, no. 20, p. 7518, Oct. 2022, doi: 10.3390/en15207518.

[11] P. Agnolucci and W. McDowall, “Designing future hydrogen infrastructure: Insights from analysis at different spatial scales,” Int J Hydrogen Energy, vol. 38, no. 13, pp. 5181–5191, May 2013, doi: 10.1016/j.ijhydene.2013.02.042.

[12] IEA, “Global Hydrogen Review 2023 Executive summary,” 2023.

[13] A. M. Oliveira, R. R. Beswick, and Y. Yan, “A green hydrogen economy for a renewable energy society,” Curr Opin Chem Eng, vol. 33, p. 100701, Sep. 2021, doi: 10.1016/j.coche.2021.100701.

[14] R.-H. Lin, Y.-Y. Zhao, and B.-D. Wu, “Toward a hydrogen society: Hydrogen and smart grid integration,” Int J Hydrogen Energy, vol. 45, no. 39, pp.

20164–20175, Aug. 2020, doi: 10.1016/j.ijhydene.2020.01.047.
[15] I. Marouani et al., “Integration of Renewable-Energy-Based Green Hydrogen into the Energy Future,” Processes, vol. 11, no. 9, p. 2685, Sep. 2023, doi: 10.3390/pr11092685.
[16] Y. Guo, G. Li, J. Zhou, and Y. Liu, “Comparison between hydrogen production by alkaline water electrolysis and hydrogen production by PEM electrolysis,” IOP Conf Ser Earth Environ Sci, vol. 371, no. 4, p. 042022, Dec. 2019, doi: 10.1088/1755-1315/371/4/042022.
[17] S. Krishnan et al., “Present and future cost of alkaline and PEM electrolyser stacks,” Int J Hydrogen Energy, vol. 48, no. 83, pp. 32313–32330, Oct. 2023, doi: 10.1016/j.ijhydene.2023.05.031.
[18] A. H. Reksten, M. S. Thomassen, S. Møller-Holst, and K. Sundseth, “Projecting the future cost of PEM and alkaline water electrolysers; a CAPEX model including electrolyser plant size and technology development,” Int J Hydrogen Energy, vol. 47, no. 90, pp. 38106–38113, Nov. 2022, doi: 10.1016/j.ijhydene.2022.08.306.