Machine learning-based prediction and optimization of green hydrogen production technologies from water industries for a circular economy

*Mohammad Mahbub Kabir[1,2,3], Sujit Kumar Roy[4], Faisal Alam[3], Sang Yong Nam[5], Kwang Seop Im[5], Leonard Tijing[1,2], Ho Kyong Shon[1,2] **

[1]*ARC Research Hub for Nutrients in a Circular Economy, School of Civil and Environmental Engineering, Faculty of Engineering and IT, University of Technology Sydney, P.O. Box 123, Broadway, NSW 2007, Australia.*

[2]*Center for Technology in Water and Wastewater, School of Civil and Environmental Engineering, Faculty of Engineering and IT, University of Technology Sydney, P.O. Box 123, Broadway, NSW 2007, Australia.*

[3]*Department of Environmental Science and Disaster Management, Noakhali Science and Technology University, Noakhali 3814, Bangladesh.*

[4]*Institute of Water and Flood Management (IWFM), Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh.*

[5] *Research Institute for Green Energy Convergence Technology, Department of Materials Engineering and Convergence Technology, Gyeongsang National University, Jinju, 52828, Republic of Korea.*

*Corresponding author

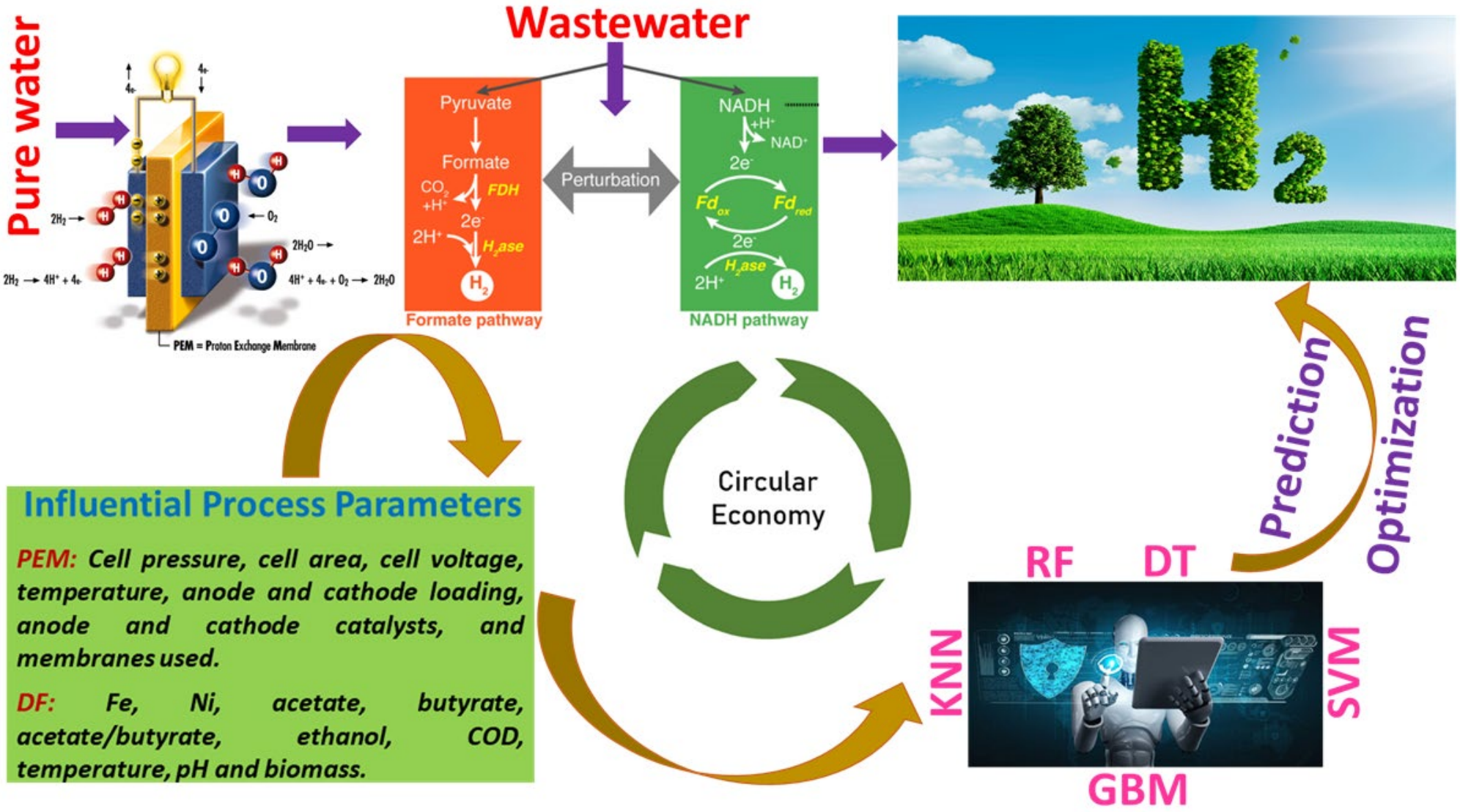
Email: Hokyong.Shon-1@uts.edu.au (H.K. Shon)

Abstract

Currently, there exists a significant number of green hydrogen production (GHP) technologies based on scaling-up issues (SCUI). Optimal prediction and process optimization could be one of the most substantial SCUI of GHP. Machine learning (ML)-based prediction and optimization of GHP technologies from water industries for a circular economy could be a plausible solution for these SCUI. We studied a detailed techno-economic and environmental feasibility study, which recommended proton exchange membrane (PEM) and dark fermentation (DF) as the most promising and environment-friendly technologies for GHP. Thus, the present investigation aims to apply different ML models to predict and optimize the GHP of DF and PEM technologies to solve the SCUI. The results revealed K-nearest neighbor and random forest are the best-fitted models to predict GHP for DF and PEM, correspondingly based on the regression co-efficient (R^2), root mean squared error and mean absolute error. The permutation variable index recommended that chemical oxygen demand (COD), butyrate, temperature, pH and acetate/butyrate ratio are the most influential process parameters in decreasing order for DF, while temperature, cell areas, pressure, voltage and catalysts loadings are the most effective process parameters for PEM in reducing order. The partial dependency analysis demonstrated GHP increases with increasing COD values up to 10 mg/L, and the optimal temperature range in the DF process is between 25 to 30 ^{0}C. On the other hand, cell temperature up to 35 ^{0}C should be considered optimum for PEM, and 40-70 cm^2 cell areas could produce a significant GHP. In summary, the present study underscores the potential of machine learning (ML) and artificial intelligence (AI) as promising techniques for optimizing GHP, ultimately addressing scaling-up challenges in large-scale industrial GHP production and ensuring a sustainable hydrogen economy (HE).

Keywords: Machine learning, green hydrogen, bibliometric analysis, partial dependency analysis, dark fermentation, proton exchange membrane.

Graphical abstract



1. Introduction

The efficient functioning of our society and economy depends heavily on energy, which is required for literally all aspects of modern life. According to Ahmad and Zhang [1], the total quantity of energy utilized globally is predicted to have substantially grown to 44 billion tons (BT) of oil equivalent in 2050, which was 13 BT in 2010. The global energy network will undergo a change in the modern era, and "renewable energy (RE) production" is being considered as the modern and sustainable solution [2]. RE refers to any form of energy derived from natural and renewable resources like sunlight, wind, water, geothermal heat, and biomass. Hydrogen (H_2), a versatile, clean, and abundant energy carrier, can be produced from various sources such as water, biomass, and natural gas. Around 115 million tons (MT) of H_2 were produced worldwide in 2020, with the majority going toward industrial activities like chemical and petroleum refining. Less than 1% of this H_2 is utilized as a fuel for transportation or energy production, and global H_2 demand could increase 10-fold by 2050, with a market reaching around 530 MT annually [3]. One of the main benefits of H_2 energy is that it emits no greenhouse gases (GHGs) when burned or utilized in fuel cells, making it an attractive alternative for reducing carbon emissions and mitigating climate change [4].

Based on the feed sources, H_2 can be produced from ultra-pure and wastewater using several different techniques. If we consider ultra-pure/demineralized water, proton exchange membrane (PEM), alkaline water electrolysis (AEWEL), and anion exchange membrane water electrolysis (AEMWEL) are the most commonly explored green H_2 production technologies. On the contrary, in terms of wastewater feed microbial electrolysis cells (MECLs), photo-electrochemical electrolysis (PECM), bio-photolysis (BP), photo fermentation PF), dark fermentation (DF), electrodialysis (ED) and reverse electrodialysis (RED) are the widely explored techniques for the production of H_2. However, to date, green H_2 production technologies are facing challenges due to the drawbacks of low purity, high cost and process energy consumption. In our previous research [5], a detailed techno-economic and environmental feasibility study (TEEFS) was conducted to select the best H_2 production technologies considering various feed sources. The results of the TEEFS recommended PEM and DF as the most promising and environment-friendly technologies to produce green H_2 with the highest production rate, purity and lowest energy consumption.

The H_2 production from PEM is typically done through the electrolysis of water, which involves passing an electric current through water to split it into H_2 and oxygen. This process

can be done using RE sources like solar, wind or hydropower, making it a potentially sustainable way to produce H_2 for different applications. Nonetheless, only 4% of total green H_2 came from PEM technology due to the lack of optimizing its process parameters [6]. Process optimization aims to increase the efficiency, durability, and performance of the PEM while reducing costs and minimizing environmental impact. The dark fermentation (DF) method for H_2 production is a biological process that utilizes specific types of bacteria to break down organic material without light. The DF method is currently considered a promising approach for sustainable H_2 production, especially for small-scale applications. However, several challenges still need to be addressed to make this method commercially viable. One of the significant challenges is the process' low H_2 yield and production rate [7], which is affected by various process parameters such as pH, temperature, substrate/biomass composition and characteristics, acetate, butyrate, ethanol, acetate to butyrate ratio, hydraulic retention time (HRT), chemical oxygen demand (COD), and some trace elements such as iron (Fe) and Nickel (Ni).

The whole water industry (WRI), which covers freshwater, wastewater and seawater, plays a significant role by providing essential services that contribute to the efficient management of water resources by producing H_2 energy for a circular economy (CRE). The CRE aims to reduce waste and pollution by extending the useful life of resources. To do this, reuse, repair, refurbishing, and recycling are all incorporated into the design of items and systems. H_2 production can have a substantial impact on the CRE [10]. Before being securely and successfully recycled, water resources in a CRE are utilized to the utmost extent possible and preserved within the economy to produce H_2 [11]. By incorporating CRE principles into water management strategies and considering H_2 as a valuable energy resource for ensuring a sustainable H_2 economy (HE), we can create a more resilient and sustainable WRI. To manage water resources sustainably, WRI increasingly needs to adopt CRE concepts. The production of H_2 is a crucial part of the CRE because it enables the storage of renewable energy in H_2, which can be utilized to power various applications [8]. By anticipating energy needs, pinpointing the ideal conditions, and minimizing energy use, machine learning (ML) can assist WRI in optimizing H_2 production processes, resulting in cost savings and improved efficiency [9, 12-21]. Implementing CRE principles and leveraging ML-WRI and H_2 production can become more sustainable, efficient, and cost-effective.

In the present study, a detailed bibliometric analysis (BA) has been conducted to examine the significant research gaps and applications that applied ML for foreseeing H_2 production

technologies from the WRIs for a CRE. It demonstrated several causal relationships between current research endeavors, the influence of research on the fields of WRI, the CRE, ML, and H_2 production technologies, the standard of research and its contributions to those fields, as well as the research accomplishments of individuals, institutions, and nations, till to date. Moreover, an ML-based prediction and optimization process of H_2 production technologies has been carried out to select the best ML models for scaling up the industrial-scale H_2 production from WRI to develop the future H_2 economy (HE). To be more precise, this study aims to develop a predictive model that can accurately identify the most efficient and sustainable H_2 production technologies for a CRE based on the data from the WRI.

2. Bibliometric analysis (BA) of ML-based prediction and optimization of H_2 production technologies from WRI for a CRE

Bibliometrics is a statistical technique used to examine patterns in bibliographic data, including publication frequency, citations, and co-authorships. It provides insights into the quantitative aspects of publications, including production, dissemination, and impact [22]. BA can be employed to assess the output and influence of specific academics, research organizations, and scientific fields, identify new areas of study, monitor the evolution of scientific disciplines over time [23], and inform decision-making regarding research prioritization and funding allocation [24]. In our study, we conducted a BA to investigate the extensive research on ML applications for predicting H_2 production technologies from WRI in the context of CRE. The aim of this analysis was to: a) understand current research activities on ML, WRI, H_2 production, and their connection to CRE, identifying areas that require further research; b) assess the impacts of research in these areas; c) evaluate the quality and trends of research and its contributions to relevant research fields; and d) compare the research performance of individuals, institutions, and countries regarding ML, H_2 production, WRI, and CRE.

2.1 Data collection

According to [19], the Scopus database is the largest academic information resource in the world, covering a wide range of disciplines. In this study, we conducted BA research on the application of ML in predicting H_2 production technologies from WRI for a CRE. To perform our search, we used the Advanced Search Statements in Scopus, with the query 'TITLE-ABS-KEY (hydrogen AND production) AND TITLE-ABS-KEY (machine AND learning) OR TITLE-ABS-KEY (circular AND economy) OR TITLE-ABS-KEY (water AND industry)' and limited the search to the period 2000-2023. The database was last updated on 22 February

2023, and we retrieved a total of 1810 papers, including 1300 articles, 251 conference papers, 220 reviews, and 39 papers of other types.

2.2 Analysis methods

Social network analysis (SNA) is a research methodology that employs network theory and analytical tools to examine relationships between individuals, groups, and organizations. Sociograms, which are visual representations of networks, are used to illustrate nodes and edges [23]. Nodes refer to the entities being studied, such as individuals, groups, or organizations, while edges represent the connections or relationships between them. SNA enables researchers to analyze the properties of social networks, such as centrality, density, and homophily. In this study, we utilized bibliometric tools such as VOSviewer [25] to conduct these analyses.

2.3 Current trends of publications

Fig. 1 (A) provides an overview of the number of publications per year from 2000 to 2023. The analysis of this data reveals that there were initially a relatively small number of papers on the topics of H_2 generation, CRE, and ML. The trend remained fairly consistent during the first five years of the study. However, starting from 2005, there was a notable exponential increase in the number of publications, with an average of 90 papers published per year up until 2023. Furthermore, Fig. 1 (C) showcases the top five trending fields based on the percentages of publications. The highest percentage (38.8%) is observed in the energy sector, followed by environmental science (30.4%), chemical engineering (12.1%), engineering (9.9%), and chemistry (8.7%). The variation in publication percentages across these fields may be influenced by factors such as funding availability and financial incentives specific to each field.

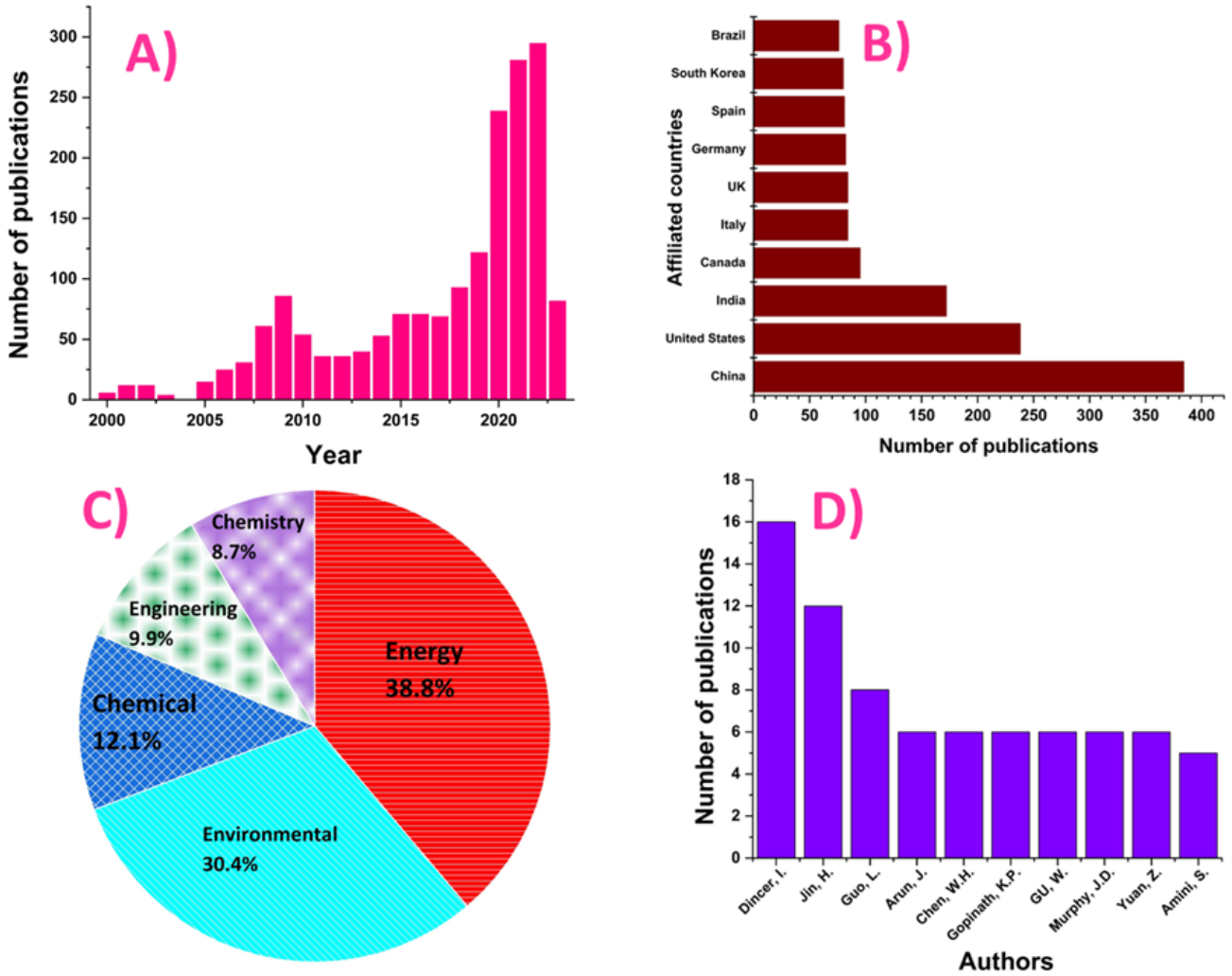


Fig. 1. A) The annual number of papers published over the period 2000-2023; B) Top 10 countries that have published the most papers during the same period; C) Top 5 subject areas according to the percentage of publications; D) Top 10 authors based on the number of publications they have produced (Scopus database: 2000-2023, with keywords hydrogen, circular economy, machine learning and water industries).

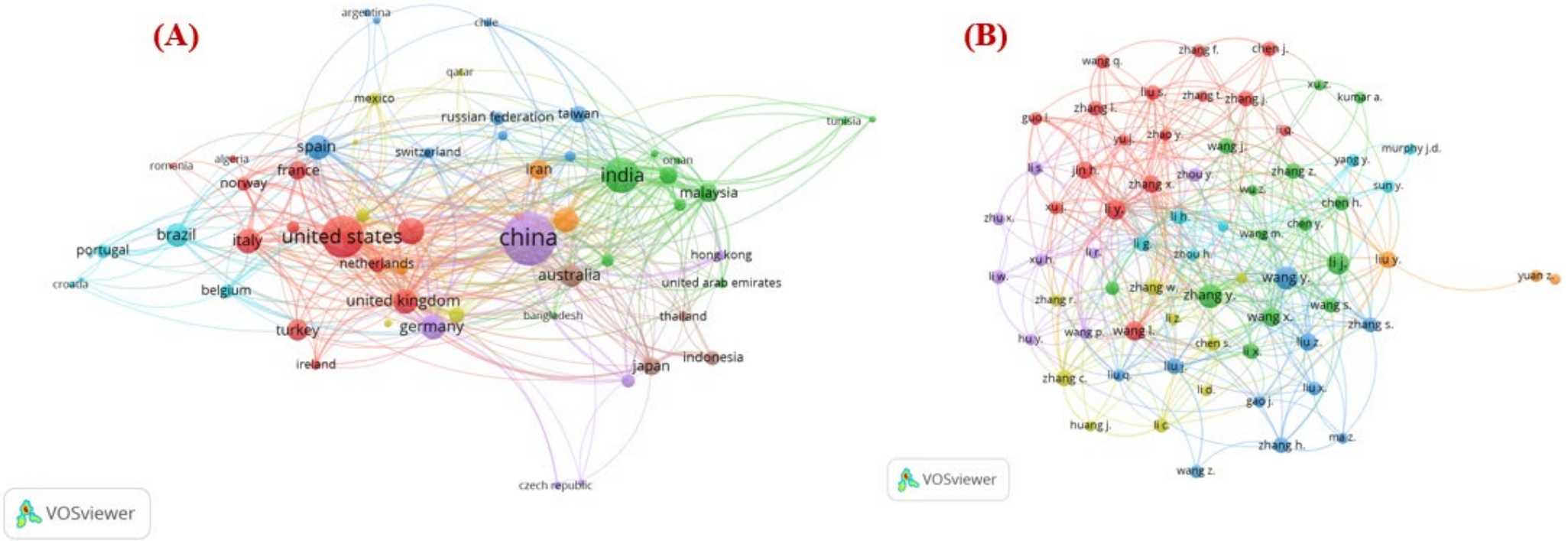


Fig. 2. Co-authorship by countries and authors A) Network mapping of countries and their publications; B) Co-authorship (Advanced search statements in Scopus, with the query 'TITLE-ABS-KEY (hydrogen AND production) AND TITLE-ABS-KEY (machine AND learning) OR TITLE-ABS-KEY (circular AND economy) OR TITLE-ABS-KEY (water AND industry)' and limited the search over the period 2000-2023).

2.4 Active nations

To analyze and visualize the data, a citation analysis was conducted using VOSviewer, resulting in the creation of a network map of nations and authors (Fig. 2). The bibliometric map consists of nodes represented by circles, highlighting nations that have produced the highest number of publications related to the studied characteristics. Central nodes indicate strong connections to other nodes and scientific articles originating from the respective country. The size of the nodes reflects the number of citations received, indicating the impact of the research. The interconnectedness observed across all studies demonstrates the global influence of research in this field. According to the cluster data presented in Fig. 1B, Chinese authors contributed the highest proportion of research articles (21.21%), followed by scholars from the United States (13.14%), India (9.5%), Canada (5.24%), and Italy (4.64%).

Table 1 The ranking of the top five journals, research areas, countries, and authors.

Attributes		**Documents**	
Scopus (2000-2023)		1810	
Articles		1300	
Conference papers		251	
Reviews		220	
Others		39	
Ranking	**Research areas**	**Number**	**Percentage (%)**
1st	Energy	1242	68.61
2nd	Environmental Science	973	53.76
3rd	Chemical Engineering	388	21.43
4th	Engineering	317	17.51
5th	Chemistry	280	15.46
Ranking	**Journals**	**Number**	**Percentage (%)**
1st	International Journal of Hydrogen Energy	219	12.09
2nd	Bioresource Technology	105	5.80
3rd	Journal of Chemical Technology & Biotechnology	58	3.2
4th	Journal of Hazardous Materials	40	2.2
5th	Water Science and Technology	39	2.15
Ranking	**Affiliations (Country)**	**Number**	**Percentage (%)**
1st	China	384	21.21
2nd	United States (USA)	238	13.14
3rd	India	172	9.5
4th	Canada	95	5.24
5th	Italy	84	4.64
Ranking	**Authors**	**Number**	**Percentage (%)**
1st	Dincer, I.	16	0.89
2nd	Jin, H.	12	0.66
3rd	Guo, L.	8	0.44
4th	Arun, J.	6	0.33
5th	Chen, W.H.	6	0.33

2.5 Distributions of journals and publications

In recent years, a comprehensive body of literature has emerged on the subjects of H_2 production technologies, CRE, ML, and WRI. This includes a total of 1300 articles, 251 conference papers, and 220 reviews (Table 1). Notably, 219 of these publications have been featured in renowned high-impact factor journals. Some notable examples of these journals include the International Journal of Hydrogen Energy, Bioresource Technology, Journal of Chemical Technology & Biotechnology, Journal of Hazardous Materials, and Water Science and Technology. These journals play a crucial role in disseminating research in this field and are widely recognized for their contribution to the advancement of knowledge in H_2 energy production and related areas.

2.6 Author analysis

Fig. 2B presents a visualization of researchers who have authored at least six articles related to ML-aligned H_2 production, WRI, and CRE. The list highlights the significant presence of Chinese researchers, who hold prominent positions in terms of publications and citations compared to researchers from other countries. The visualization incorporates citation analysis to consider the contributions of authors. Furthermore, Fig. 1D displays the top ten authors based on the number of publications. Notably, Dincer, I., a full professor of Mechanical Engineering at Ontario Tech. University stands out as the highest scorer with 16 publications.

2.7 The topic trends for the future directions

The clusters of keywords and their respective critical areas inferred from the thematic research are shown in Fig. 3. Five clusters were identified, with the first cluster featuring 13 keywords highlighted in red, relating to ML and H_2 production. The next cluster, colored green, comprises 13 keywords focused on H_2 production technologies, followed by biomass and renewable energy. The third cluster, colored blue, centers around CRE theory and contains 11 keywords, followed by bio-H_2. The fourth cluster, represented by violet-colored nodes, pertains to gasification technology and its related keywords. Lastly, the yellow color zone signifies the wastewater treatment area and includes ten items. The yellow-colored nodes are situated in the middle position among all other nodes in the keywords map, suggesting that they have the strongest interrelationship with all the research themes [26]. BA recommends that ML-based GHP technologies should get priority in future research directions. Moreover, research and

development on GHP should be based on CRE, which will ensure the circularity of resources and energy for the development of sustainable development goals (SDGs) and a future HE.

2.8 The significant findings of BA

The groundwork for conducting an in-depth study involved analyzing 1810 publications on H_2 energy production technologies related to ML and the role of WRI in CRE using the Scopus database. The research objectives were achieved through this analysis. Over the past few years, a total of 1300 journals, 220 reviews, and 251 conferences have published articles on the topic of H_2 generation with artificial intelligence (AI). The International Journal of Hydrogen Energy and Bioresource Technology stands out as the leading journal in publishing research on this subject. The annual publication count from 2000 to 2023 is presented in the study. The analysis indicates that during the first decade of the study (2000-2009, the number of publications related to energy generation showed an exponential increase. China surpassed other nations in terms of the volume of papers published on H_2 generation with computational prediction. Cluster analysis was conducted to identify trends in research keywords. The bibliometric results were synthesized and analyzed to reveal various findings, such as stakeholder participation networks, relevant theories, and the processes of technology generation. To assess publication behavior, a defined timeframe of publishing years was used. Furthermore, the results of the cluster analysis were supported by an evolutionary pathway analysis, which illustrated the changes in research on H_2 energy production over time. This analysis also provided key terms that are valuable for this type of investigation. A global analysis was subsequently conducted to explore potential research opportunities and identify research gaps that could benefit individuals in need.

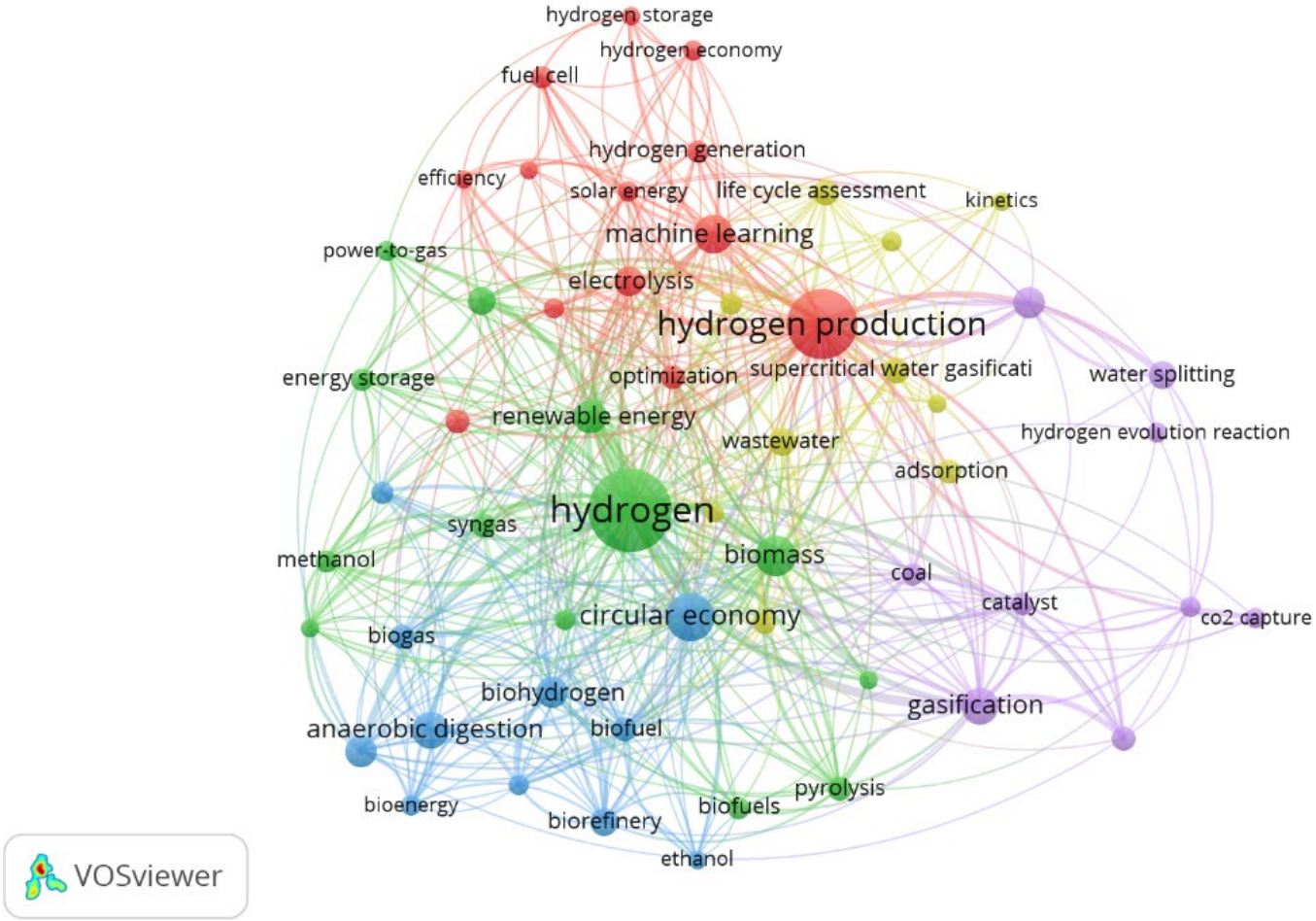


Fig. 3. Clusters of keywords mapping (Advanced search statements in Scopus, with the query 'TITLE-ABS-KEY (hydrogen AND production) AND TITLE-ABS-KEY (machine AND learning) OR TITLE-ABS-KEY (circular AND economy) OR TITLE-ABS-KEY (water AND industry)' and limited the search over the period of 2000-2023).

3. Methodology

3.1 ML data collection of DF and PEM technologies

To investigate the ML-based prediction and optimization of H_2 production technologies from WRIs for the circular economy through DF and PEM processes, an extensive literature review was conducted, encompassing various factors. The review yielded a total of 2007 data points for DF and 1699 data points for PEM, which were extracted using Plot Digitizer as well as manually. The list of selective variables investigated in this study is presented in Table 2.

Table 2 Selective process parameters of DF and PEM technologies.

DF		**PEM**	
Variables	**Units**	**Variables**	**Units**
Fe	mg/L	H_2 production rate	Nm^3/h
Ni	mg/L	Cell pressure	(Atm.)
Biomass	g VSS/L	Cell area	Cm^2
Acetate	g/L	Cell voltage	V
Butyrate	g/L	Temperature	^0C
Acetate/butyrate ratio	-	Anode loading	mg/cm^2
Ethanol	g/L	Cathode loading	mg/cm^2
pH	-	Anode catalysts	-
Temperature	^0C	Cathode catalysts	-
HRT	h	Membranes used	-
COD	g/L	-	-
H_2 production rate	mol H_2/mol glucose	-	-

Among PEM variables, the anode catalyst, cathode catalyst, and membrane used were categorical, while the remaining variables were numerical. To make the categorical data suitable for the ML models, dummy coding was applied.

3.2 Data pre-processing and features selection

Dealing with missing values is crucial to ensure the accuracy and reliability of ML models. Missing data refers to unreported features that exist in the literature. In this study, addressing the missing input data was the first step to enable a comprehensive comparison of various ML algorithms. The missing values were commonly imputed using the mean or median values, and in this case, the mean value imputation approach was adopted based on previous studies. For categorical variables, dummy encoding was utilized [27], creating a separate dummy variable for each level of the categorical variable. To handle extreme values and outliers, the boxplot method was employed to identify and remove unusual data. Then some of the input

characteristics that were deemed unimportant owing to their poor R^2 and high MSE values were deleted. However, picking features exclusively based on feature importance is not always convincing because minor features might contribute contextual power to the ML models when they come into touch with other features [28,29]. Feature choices increase data quality by reducing the size of the feature space and increasing the algorithm's speed. It also increases process speed by deleting unwanted, undesired, and redundant input characteristics that have no influence on the correctness of the ML model. We utilize the correlation matrix and Boruta model [29] approaches to tackle this problem. The finalized data were then randomly divided into training and testing datasets, with 70% allocated for training and 30% for testing. Model performance evaluation and hyperparameter tuning were conducted using cross-validation. All experiments were implemented in R-Studio, and the reported results represent the mean of five repeated experiments conducted with different random seeds.

3.3 Selection of ML procedures and modelling generality.

In accordance with Occam's Razor principle, which suggests that a model should be kept as simple as possible while being sufficiently complex [30] and considering the varying performance of different ML procedures across different applications [31], the careful selection of appropriate procedures becomes crucial. ML algorithms can be broadly categorized into supervised, unsupervised, and semi-supervised learning. Supervised ML algorithms rely on labelled data as input, whereas unsupervised ML algorithms do not require labels. As a result, supervised ML algorithms often exhibit higher accuracy compared to other ML algorithms [32]. They are commonly employed in solving regression problems. Regression problems, specifically, are frequently addressed using widely used algorithms such as random forest (RF), K nearest neighbors (KNN) regression [33], support vector machine (SVR) [34], and decision tree regression (DTR) [35]. Hence, to explore various ML approaches for the prediction and optimization of H_2 production technologies, five different classical models were employed, such as RF, DT, KNN, SVM and gradient boosting machine (GBM). The selection of these procedures was based on a thorough review of the relevant literature.

To facilitate pre-screening and in-depth modelling, the datasets were randomly split into training datasets (70%) and test datasets (30%). To ensure adequate validation and prevent overfitting, a 5-fold cross-validation technique was applied to assess the performance of the developed models. The test dataset was utilized to evaluate the generalization capability of the models [36]. To optimize the models, hyperparameters were tuned through a grid search

approach tailored for each selected procedure. Finally, the tuned hyperparameters were employed to develop and test the models. The evaluation and selection of the most suitable models from each ML procedure were based on the root mean squared error RMSE (Eq. (1)), mean absolute error (MAE) (Eq. (2), and regression coefficient (R^2) (Eq. (3)) metrics. Notably, the average values of these statistical indices were used to assess the performance of the validation phase throughout the modeling process.

$$RMSE = \sqrt{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2 \quad (1)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}| \quad (2)$$

$$R^2 = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \bar{y})^2} \quad (3)$$

Where,

$\hat{y}$ = Predicted Value of y

$\bar{y}$ = Mean value of y

3.4 ML Models

3.4.1 Support vector machine (SVM)

The support vector machine (SVM) was introduced by [37] as a supervised ML approach known for its ability to minimize structural risk and utilize statistical learning theory. SVM has been widely employed in various applications, including regression problems, text detection, troubleshooting, and image retrieval, demonstrating its effectiveness. The SVM structure consists of three layers: input, hidden, and output layers [38]. The independent and dependent variables are located in the input and output layers, respectively. Kernels, which are mathematical functions, are defined in the hidden layer to transform inputs into the required forms. SVM algorithms leverage different types of kernels and aim to find a hyperplane through nonlinear mapping to effectively train the model or classify data. By transforming the nonlinear input space into a high-dimensional feature space, SVM has shown superior performance compared to traditional statistical models in regression analysis, pattern recognition, and classification tasks. When applied for regression and function approximation, it is known as support vector regression (SVR). Commonly used kernel functions in various SVMs include linear, radial basis function (RBF), and polynomial (poly) [38]. In the context of predicting H_2 production using the DF and PEM process, the independent variables were

treated as inputs to develop an SVR model. To optimize the model's performance, a grid search approach was employed to tune the hyperparameters and select the best kernel. The grid search involved tuning hyperparameters such as C (with values ranging from 10^{-2} to 10^{2} and sigma (with values of 0.01, 0.1, and 1). The objective was to determine the optimal combination of hyperparameters for each kernel. Subsequently, the tuned hyperparameters, along with the selected kernel, were used to develop the SVR models.

3.4.2 Gradient boosting machine (*GBM)*

The gradient boosting machine (GBM) is a powerful ensemble-supervised ML approach introduced by Friedman, capable of modeling and analyzing data for both regression and classification problems [39]. GBMs consist of three key elements: weak and strong learners, a loss function, and an additive model. Weak learners serve as initial decision trees, providing predictions that are generally better than random guessing, while strong learners are created by combining multiple weak learners to achieve notable predictive performance. GBMs gradually and additively train decision trees in a sequential manner, boosting the weak learners into stronger ones to model and analyze the underlying processes. The goal is to reduce the total error or loss function, where new weak learners are added and trained to decrease the model's error while keeping the current weak learners unchanged [40,41]. For developing a GBM model in the context of DF and PEM technologies, a grid search approach was employed to identify the optimal hyperparameter settings within a defined grid [42]. The key hyperparameters considered in this process included the number of gradient-boosted trees (depth = 25), the total number of trees (600), shrinkage (0.01), and the number of nodes (5).

3.4.3 Random forest (RF)

Random forest (RF) is a supervised machine-learning strategy that models both classification and regression phenomena, and it was initially introduced by Breiman to deal with regression trees [43]. As a function of regression, RF generates a wide variety of decision trees, with the eventual proportion of the response variable being the mean of all decision trees [44]. The hyperparameters, i.e. the number of gradient-boosted trees, were 500, and the tuning grid was defined as mtry = c (2, 5, 10) for both the DF and PEM process.

3.4.4 Decision tree (DT)

Decision Tree (DT) algorithms are versatile tools that can be employed for both classification and regression tasks. They construct classification or regression models using a tree-like

structure. The process involves dividing the dataset into smaller and smaller subsets while simultaneously developing the associated decision tree. Consequently, the tree comprises decision nodes and leaf nodes [45]. Decision nodes have two or more branches, leading to further subdivisions of the data. The leaf nodes, on the other hand, represent the final decisions or classifications made by the model. The topmost node of the tree is referred to as the root node. One notable advantage of decision trees is their ability to handle both categorical and numerical data. In the case of Decision Tree Regression (DTR), the algorithm employs the concept of entropy, which represents the level of uncertainty in a system [46].

3.4.5 K-nearest neighbor (KNN)

The K-nearest neighbor (KNN) algorithm is a straightforward approach in ML that can be used for both classification and regression prediction tasks [45]. The model's structure is simple and computationally efficient [46]. KNN involves finding the K-nearest neighbors to a given data point, determining the subgroups to which these neighbors belong, and using voting to determine their categories. In regression scenarios, the algorithm outputs the average value of the K nearest data points as the prediction. The choice of K is critical, and it is typically specified as an odd number. This ensures that during the voting process, there is no possibility of a tie, which could lead to the model being unable to produce a result. In this study, the value of K was predetermined as 1:30 [46].

3.5 Permutation variable index (PVI)

The permutation variable index (PVI) was proposed by [47] as a procedure to inspect any fitted model in the tabular data. This procedure considers the errors of the developed model in predicting the output with a random permutation of the considered input; therefore, the more errors, the more importance of the feature [48]. Regarding the errors, MSE was used to measure the relative importance of the features. There are various merits for PVI procedure, e.g., fast and easy to calculate, a general method, considering both individual and interactive effects of each variable [49]. To identify the relative importance of the input variables in H_2 production through DF and PEM process, the PVI procedure was used for all the developed RF, DT, KNN, SVR and GBM models. As well as we used partial dependence analysis (PDA) and its 3D plot to get a clear conception of the variable influence for the best model.

3.6 Comparison of ML model performance

The statistical indices including RMSE, MAE and R^2 were used to compare the performances and strengths of the developed RF, DT, KNN, SVR and GBM models in predicting H_2 production by DF and PEM processes. The test datasets were used to calculate the relevant statistical parameters.

4. Results and discussion

4.1 Correlation matrix among model's variables of DF and PEM technologies

Correlation is widely used in ML for data analysis and mining, particularly in regression techniques where multiple independent variables are used to predict the dependent variable. It helps identify critical issues in feature sets that can adversely impact model fitting, while non-correlated features offer advantages such as faster algorithm learning, higher interpretability, and reduced bias. Fig. 4A and 4B elicit a detailed description of DF and PEM model outputs. The correlation coefficient is calculated for the DF model to be 0.72 for both COD and temperature, with Nickel (Ni), indicating a strong positive correlation, interpreting this as a strong positive monotonic relationship. It suggests that when the COD and temperature increase, so does the Ni. (Fig. 4A). The temperature rise in DF for H_2 production is attributed to the catalytic activity of nickel, which acts as a catalyst in breaking down organic matter into H_2 and byproducts. With increased nickel concentration, more catalytic reaction sites become available, resulting in higher reaction rates [50]. This increased activity generates more heat as a byproduct, resulting in an increase in temperature. Again, Fe and temperature have a significant connection of 0.66, and several positive moderate relationships have emerged between temperature and COD (0.5), Fe and Ni (0.44), and Fe and Acetate/Butyrate (0.44). On the other hand, the coefficient is -0.56 between temperature and butyrate, as a moderate negative monotonic relationship. It suggests an inverse relationship between them. Increasing butyrate levels in DF and H_2 production leads to a decrease in temperature due to the inhibitory effect on hydrogen-producing microbial communities, resulting in reduced H_2 production [51]. Meanwhile, pH and Fe (-0.43) and Butyrate and Ni (-0.47) express moderate negative relations. Other parameters have very weak to weak positive and negative relationships among themselves.

In the context of the PEM method, correlation coefficient results in Fig. 4B indicated that the correlation coefficient between the cathode catalyst of $Au.Al_2O_3$ and anode catalyst of Pt.Ir is calculated to be +0.91, which suggests a very strong positive correlation. The +0.91 magnitude

indicates a strong linear association between $Au.Al_2O_3$ and Pt/Ir catalysts both demonstrate excellent electrocatalytic activity in PEM technology. $Au.Al_2O_3$ efficiently facilitates the oxygen reduction reaction (ORR) at the cathode, while Pt.Ir effectively supports the H_2 oxidation reaction (HOR) at the anode. These catalysts promote rapid electron transfer, enhancing overall cell performance, and are compatible with PEM materials, remaining stable and resistant to degradation in the acidic electrolyte environment [52]. There were another three strong positive correlations found here as AcPt.Ir and MN. (0.71), $CCAu.Al_2O_3$ and MN. (0.66) and H_2 and cell areas (0.62). Furthermore, their use in combination minimizes potential issues such as catalyst poisoning or deactivation, ensuring the long-term stability and performance of the PEM electrolyzers. On the other hand, the correlation coefficient is calculated to be -0.53 and -0.55, which indicates a moderately negative correlation between the temperature and H_2 production as well as cell area. As temperature increases, a strong negative correlation is observed between the H_2 production rate and cell area (-0.53 and -0.55 magnitudes). In PEM technology, deviation from the optimum temperature range for electrochemical reactions can lead to decreased reaction rates and reduced H_2 production. Higher temperatures also increase PEM resistance, reducing proton conductivity and hindering overall H_2 production. Thermal management is crucial to prevent degradation of the membrane and catalyst materials. Additionally, higher temperatures can cause material expansion and swelling, leading to a decrease in the cell area, while efficient proton conduction in proton exchange membranes depends on the presence of water [53]. Higher temperatures can promote the evaporation of water from the membrane, leading to reduced water content and a potential decrease in cell area.

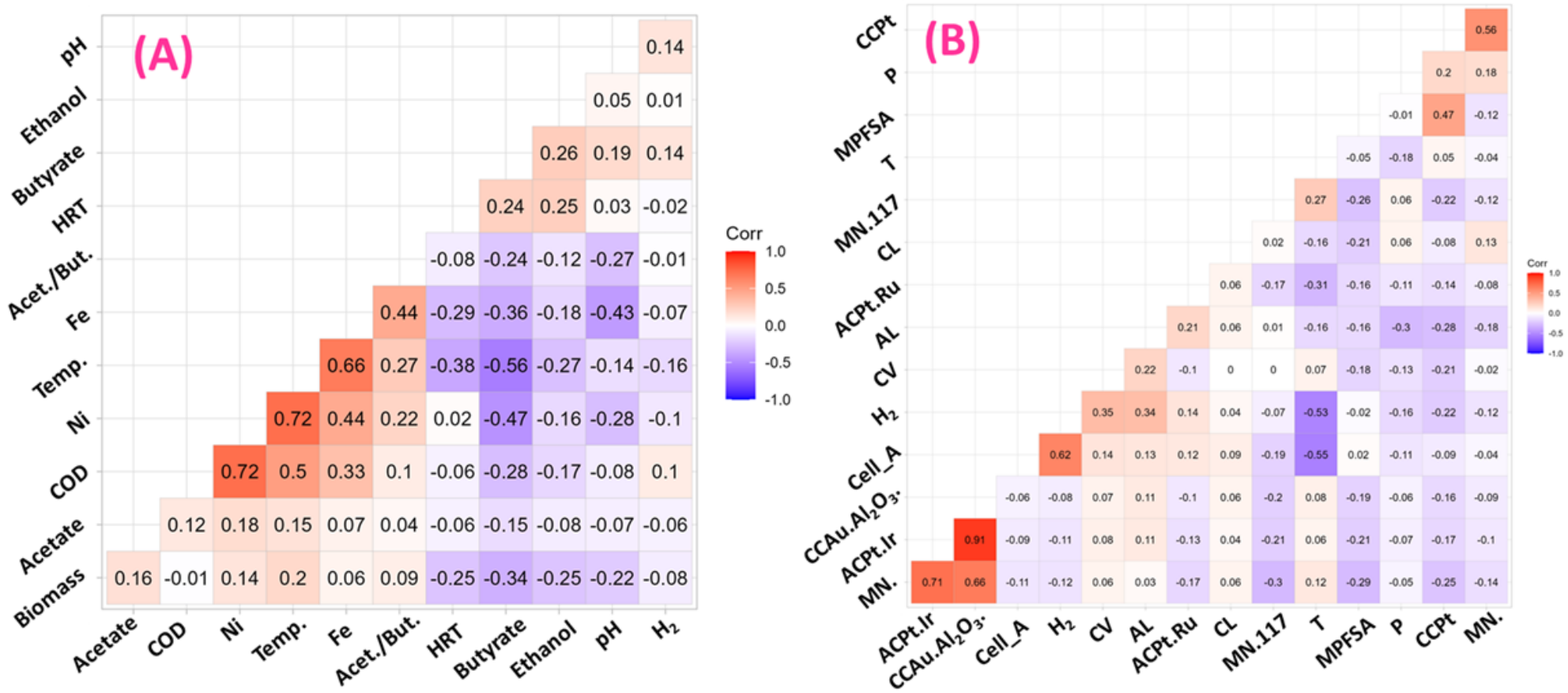


Fig. 4. Correlation matrix of variables of DF (A) and PEM (B) (HRT= Hydraulic retention time, Temp.=Temperature, Cell_A=Cell area, CV=Cell voltage, AL=Anode loading, CL=Cathode loading, T=Temperature, P= Pressure, Acet.=Acetate, But.=Butyrate, Ni=Nickel, MN=Nafion membrane, CC=Cathode catalysts, AC=Anode catalysts, Pt=Platinum, Au=Gold, Ir=Iridium, Al=Aluminum and Ru=Rubidium).

4.2 Features selection of ML models for DF and PEM technologies by Boruta model

Boruta, an ML algorithm, is utilized for feature selection, which involves reducing the number of features in a dataset by identifying those that have a significant impact on the study variable. Acting as a wrapper method, Boruta is built around the RF classifier algorithm. It operates by duplicating the original dataset's features and shuffling the values in each column to introduce randomness. These randomized features are referred to as shadow features. Fig. 5A identifies the key factors influencing H_2 production in DF. Acetate, pH, HRT, and COD are identified as the most important factors influencing H_2 production. Precise control of these parameters can lead to enhanced H_2 production. The order of their impact is as follows: COD > HRT > Acetate > pH. This significance can be attributed to the close relationship between COD and organic compound concentration, which serves as a substrate for anaerobic bacteria responsible for bio-H_2 production. Acetate plays a crucial role as an intermediate in DF, providing a substrate for microorganisms to generate H_2 gas and other valuable products. pH and HRT, on the other hand, influence microbial activity, enzyme function, and the distribution of fermentation products during the fermentation process [54]. COD plays a vital role in predicting the behavior of the DF process. However, Fe and butyrate have lower importance scores, indicating their lesser influence. Additionally, Ni, biomass, and temperature show no relative importance, suggesting they do not provide significant information for predicting the target variable in the DF method based on the training data and evaluation criteria used. [55]. It suggests that the inclusion or exclusion of Ni, biomass and temperature as a feature has minimal impact on the overall performance of the model. The primary variables affecting H_2 production in PEM technology are shown in Fig. 5B. It affirms that the most crucial variables to adjust precisely are temperature, cell area, anode loading, anode catalyst Pt.Ru, the membrane used such as N.1035 and membrane Nafion (N.) which can improve H_2 production. The chronology of the relevant importance of variables for PEM is T>CA>AL> ACPt.Ru> MN.1035> MN [56]. Higher temperatures, increased proton conductivity of the membrane, larger catalyst area, and Pt.Ru anode catalysts contribute to improved H_2 production rates. Elevated temperatures provide energy for faster reactions, while higher proton conductivity facilitates efficient proton transport.

A larger cell area allows for more electrochemical reactions and better distribution of gases. Pt.Ru catalysts enhance H_2 oxidation, and membranes with high proton conductivity, like MN.1035 and MN., facilitate efficient proton transport while restricting the crossover of other gases [57]. The membrane material needs to be chemically stable under the operating conditions of the PEM system.

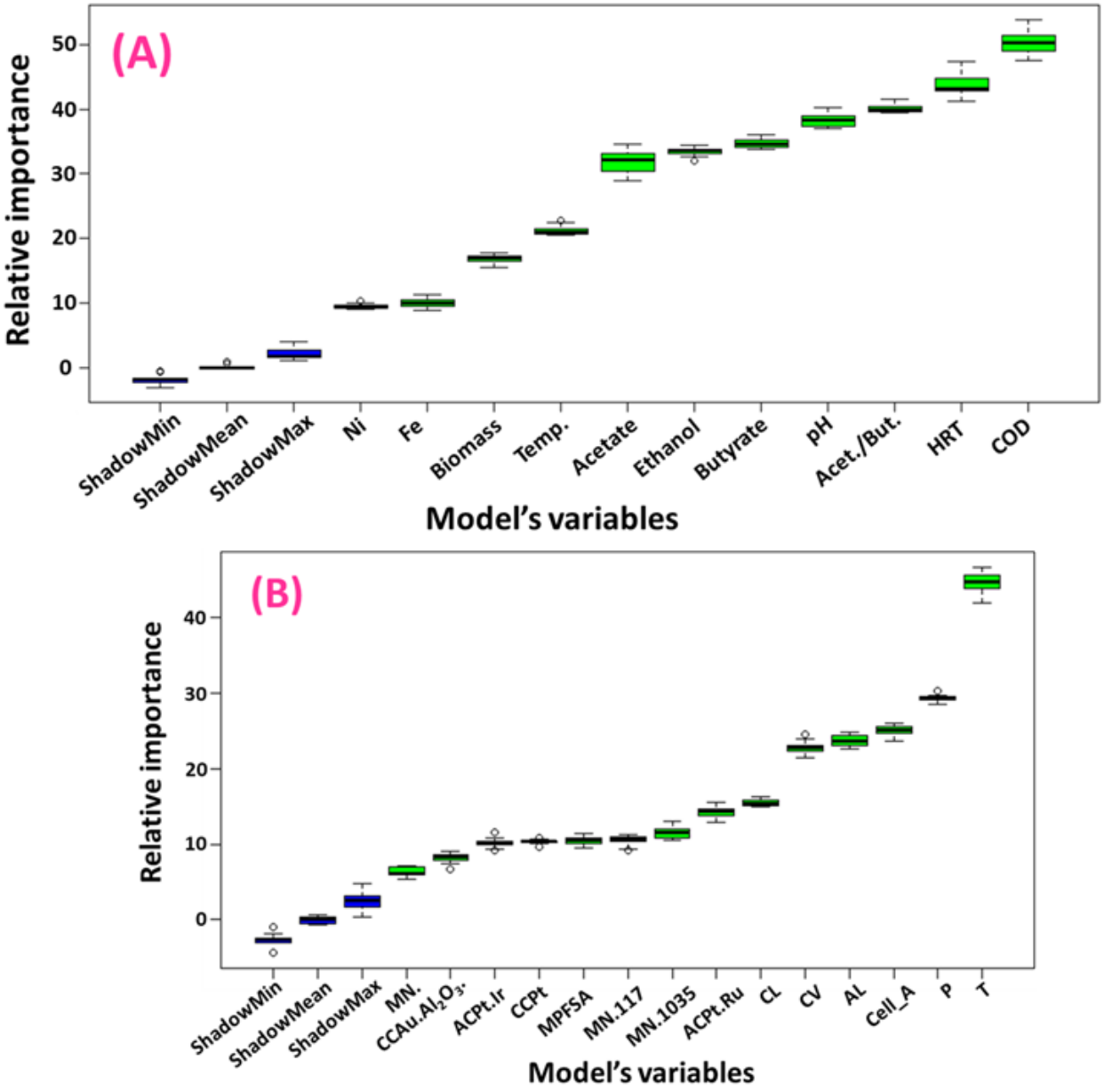


Fig. 5. Feature selection of model building, (A) DF and (B) PEM.

4.3 Comparison of ML model's capacity of H_2 by DF and PEM technology for WRI for CRE

A regression line is a straight line that represents the relationship between two variables in a linear regression model (Fig. 6). It is used to predict the value of one variable based on the value of another variable. In this study, five ML models have been studied to find the best-fitted model for H_2 production technologies. In terms of DF process, the inputs used in this study were Fe (mg/L), Ni (mg/L), biomass amount (VSS/L), acetate (g/L), butyrate(g/L), acetate to butyrate ratio, Ethanol (g/L), pH, Temperature (^{0}C), COD (g/L), and HRT (h) for H_2 production. The predicted results of KNN, DT, RF, GBM, and SVM have been shown in Fig. 6.

The model performance was determined by the correlation coefficient, root mean square error (RMSE) and mean absolute error (MAE). Among the models, the R^2, RMSE and MAE values of KNN model are 0.948, 0.038 and 0.161, respectively. Followed by 0.935, 0.114 and 0.201 for the DT model; 0.925, 0.201 and 0.115 for RF; 0.876, 0.272 and 0.191 for GBM; 0.801, 0.315 and 0.133 in terms of SVM, respectively. All the outputs revealed that the best accurate model is KNN in terms of the highest R^2 value with the lowest RMSE and MAE. As per the result, the hierarchy of the model strength would be KNN> DT> RF> GBM> SVM. The reason behind the best accuracy of KNN model could be relied on some facts; KNN performs well when the relationships between input variables and the target variable are non-linear, and another cause may be due to the localized effects. In DF, certain factors like pH, temperature, substrate composition, and microbial interactions may have localized effects on H_2 production. KNN can consider the influence of nearby observations and identify similar instances with high H_2 production [58]. Besides, SVM is the worst predictive model due to the data may not exhibit a clear linear separation or the relationship between the input features, and the target variable is highly complex and non-linear. Therefore, the DT and RF are suitable as both of the models showed a good statistical indicator nearly the KNN model which may ensure the sufficient high-quality data available in the dataset [59].

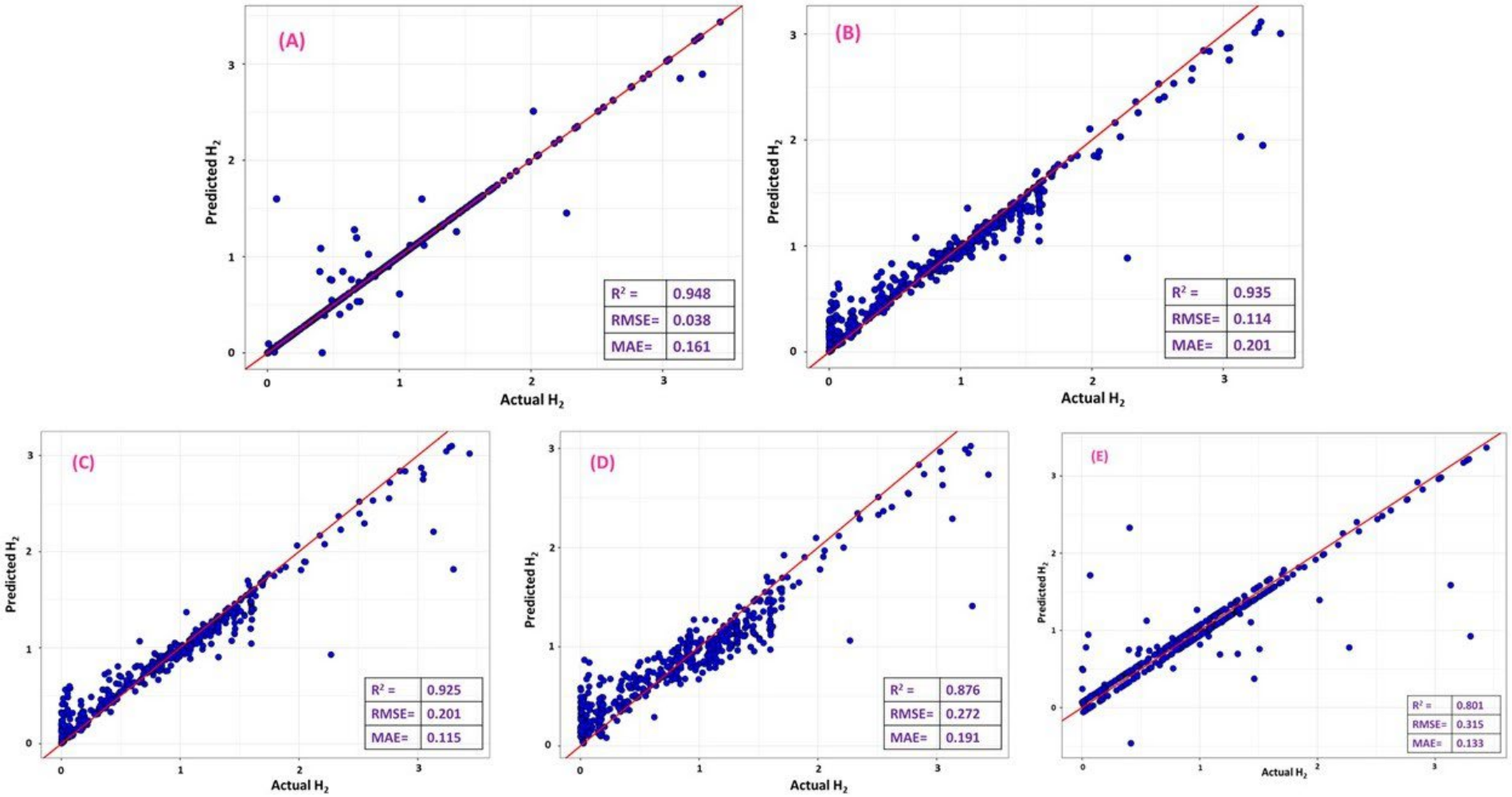


Fig. 6. Regression coefficient (R^2) of KNN (A), DT (B), RF (C), GBM (D) and SVM € for DF.

The effectiveness of the established models (KNN, DT, RF, GBM and SVM) was evaluated using three statistical indicators (R^2, RMSE and MAE) that highlight the model's superiority in the production of H_2 from pure through the (PEMWL) process. According to Fig. 7, RF (Fig 2A) is the best model regarding R^2, RMSE and MAE value as well as the highest strength carrying model exhibiting 0.901, 1.238 and 0.461, respectively. An RF (model can be a good predictive model for H_2 production in proton exchange membrane (PEM) electrolysis technology when the relationship between input variables such as current density, temperature, pressure, and H_2 production is nonlinear and complex. In this study, the RF model excels at capturing nonlinear relationships, handling interactions between variables, and providing accurate predictions even with a large number of input variables. For the DT model, the coefficient, RMSE, and MAE values are 0.899, 1.248 and 0.448, meaning that there was a sufficient amount of training data available that captures the relevant features and patterns of the system and the input variables and the target variable is relatively straightforward [60]. The KNN and GBM models resulted from 0.889 & 0.896 for R^2, 1.313 & 1.278 for RMSE, and 0.563 & 0.665 for MAE, respectively. Both KNN and GBM models showed less strength than RF and DT due to the relationship between input variables and H_2 production exhibiting non-linear behavior, reliance on distance-based similarity metrics might not be effective. Whereas the lowest strength was determined for SVM (0.860, 1.493 and 0.749) model in terms of all the statistical accuracy indicators. Due to the complex non-linear relationships or a lack of clear boundaries and overlapping between different classes or groups of PEM data points, SVM did not perform well [61].

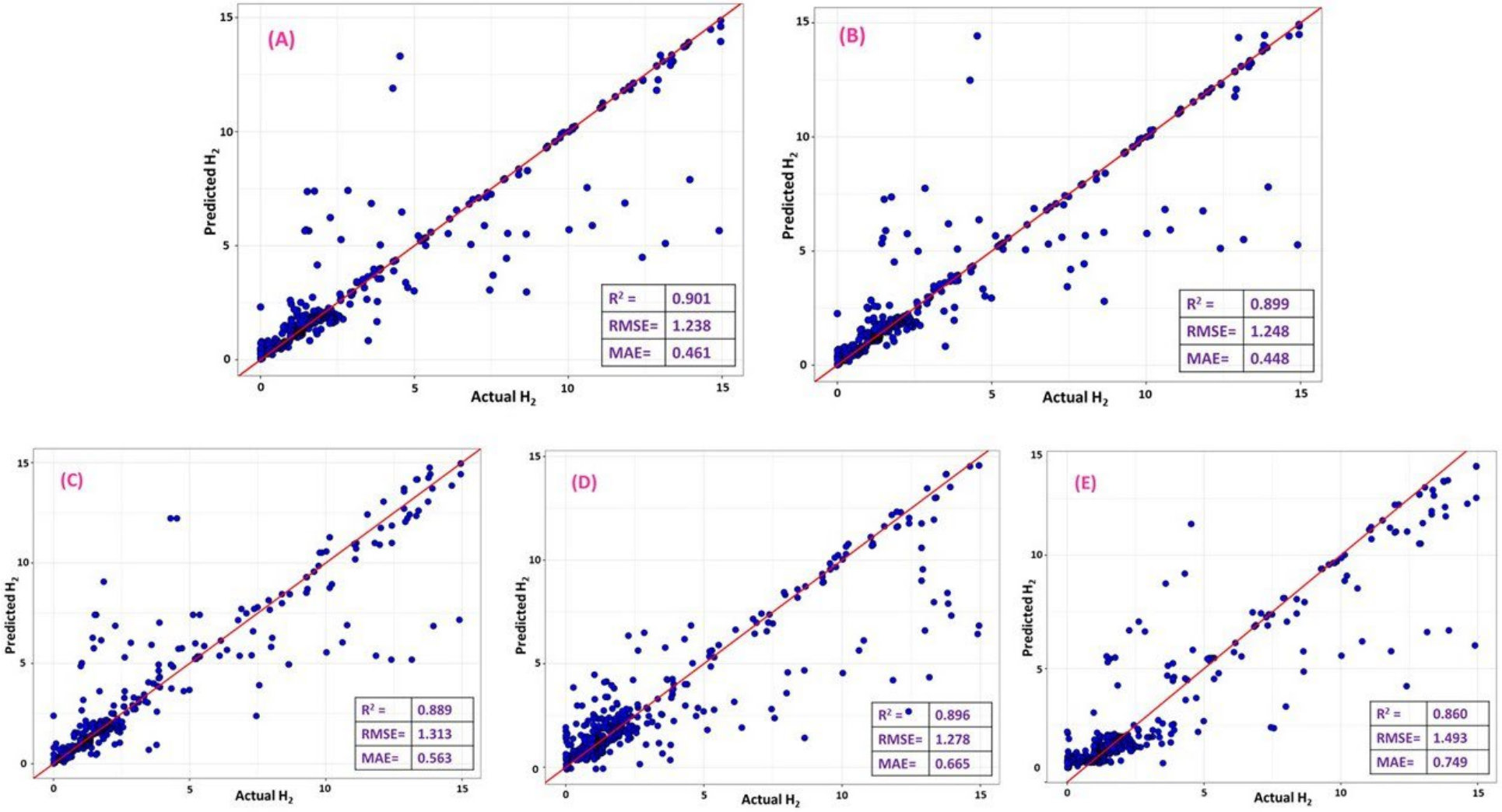


Fig. 7. Regression coefficient (R^2) of A) RF, B) DT, C) KNN, D) GBM, E) SVM for PEM.

4.4 Comparative significance of different process parameters for predicting H_2 production by DF and PEM technologies.

The relative importance of different process parameters of DF and PEM technologies provided by PVI techniques have been presented in Fig. 8 and Fig. 9, respectively. In the case of the DF process, all five models, i.e., RF, DT, KNN, GBM and SVM, revealed that COD was the highest influential variable that could control the H_2 production process in DF. DF techniques consider COD one of the most significant factors for H_2 production, where organic compounds are converted into H_2 gas by anaerobic microorganisms [62]. The organic compounds serve as the substrate for the microorganisms to produce H_2 in DF technology. The COD of a substrate indicates the amount of organic matter available for microbial degradation [63,64]. Higher COD values imply a higher concentration of organic compounds, providing a greater substrate availability for the microorganisms. This can lead to increased H_2 production. Moreover, DF relies on the metabolic activity of anaerobic microorganisms. The microorganisms require an adequate supply of organic matter to carry out the DF process efficiently. The COD value helps determine the suitable substrate concentration to maximize microbial activity and H_2 production. Thus, a higher COD generally corresponds to a higher potential for H_2 production. However, the relationship between COD and H_2 generation by DF can be complex and may depend on various factors such as microbial community composition, reactor conditions, and specific substrate characteristics [65, 66].

However, the butyrate concentration, hydraulic retention time (HRT), acetate/butyrate ratio, pH, acetate, and ethanol also played a significant role in predicting H_2 production by DF (Fig. 8) for all ML models studied, while the iron (Fe), biomass amount and the presence of nickel (Ni) did not show much considerable impacts on the overall H_2 production scenario by DF technology. The HRT determines the duration of contact between the microorganisms and the substrate, which is typically an organic waste material [67,68]. Microorganisms need time to break down complex organic compounds into simpler molecules that can be utilized for H_2 production. A longer HRT allows for more thorough substrate utilization, maximizing the conversion of the organic matter into H_2. Moreover, the microbial community in the DF process plays a crucial role in H_2 production [69].

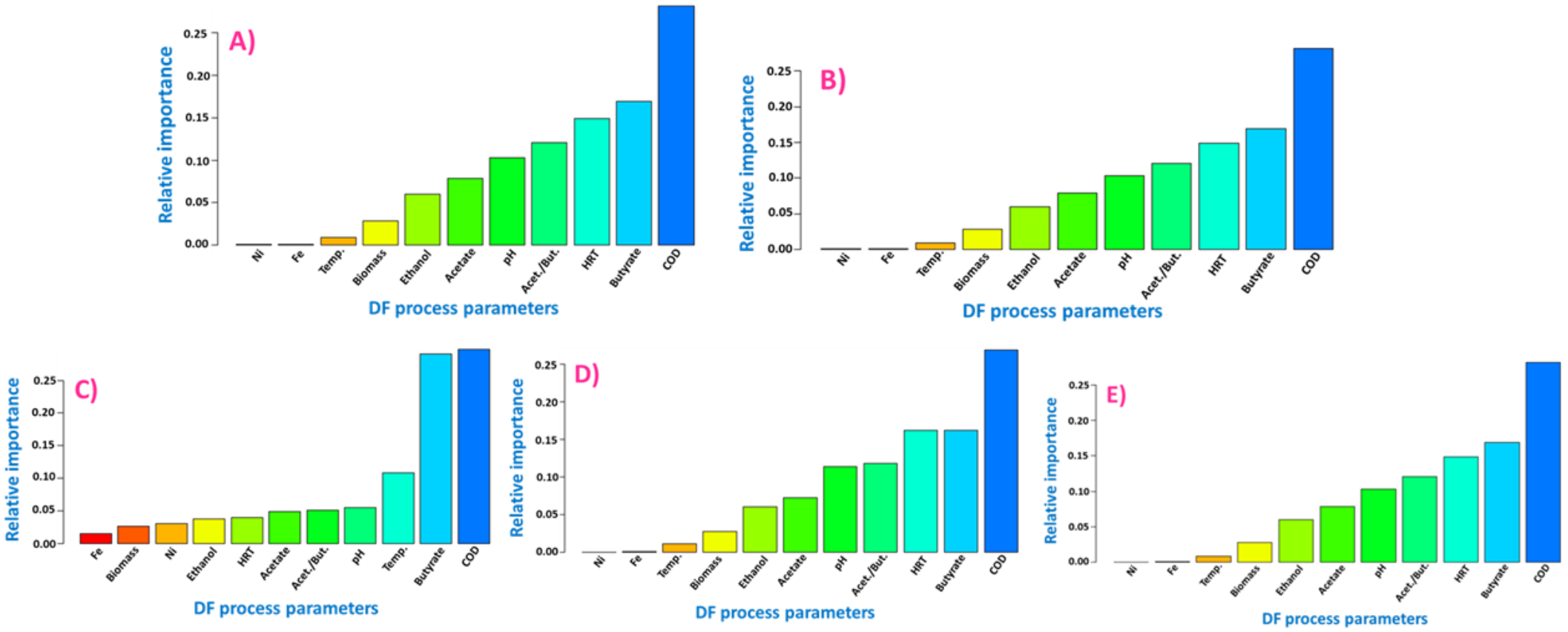


Fig. 8. Influences of different process parameters on H_2 production by DF techniques; A) RF, B) DT, C) KNN, D) GBM and E) SVM.

The HRT affects the composition and dynamics of the microbial community by favoring certain microorganisms over others. Longer HRTs can promote the growth of H_2-producing microorganisms while inhibiting the growth of competing microorganisms that consume the H_2. Controlling the HRT can shape the microbial community to enhance H_2 production in DF techniques [70]. During DF processes, organic acids such as acetate and butyrate act as electron donors. The H_2-producing microorganisms oxidize these organic acids, releasing electrons that are transferred to the electron transport chain, leading to the production of H_2 gas [71,72,73]. The ratio of acetate to butyrate is considered significant in DF for H_2 production due to its impact on the metabolic pathways and overall efficiency of H_2 production. Acetate metabolism generally yields a higher amount of H_2 per mole compared to butyrate metabolism. Acetogens can efficiently convert acetate to H_2 gas through the hydrogenase pathway, resulting in higher H_2 yields. On the other hand, butyrate metabolism typically yields less H_2 per mole of the substrate. Acetate metabolism generally generates a higher amount of H_2 per mole compared to butyrate metabolism [74]. Acetogens can efficiently convert acetate to H_2 gas through the hydrogenase pathway, resulting in higher H_2 yields. On the other hand, butyrate metabolism typically yields less H_2 per mole of the substrate. Acetogens and butyrate-producing bacteria may coexist in the DF system, and their metabolic activities can influence each other. Butyrate can inhibit acetate utilization by competing for resources and creating an unfavorable environment for acetogens. This competition can decrease the overall H_2 production efficiency [74,75]. Ethanol can play an essential role in H_2 production during DF. Ethanol can serve as a substrate for H_2-producing microorganisms in DF techniques. Some microorganisms, such as certain species of *Clostridium* and *Enterobacter*, can metabolize ethanol and convert it into H_2 gas. This provides an additional carbon source for H_2 production, expanding the range of available substrates [75]. pH plays a critical role in H_2 production. The activity and growth of hydrogen-producing microorganisms are strongly influenced by pH. Different species of H_2 -producing bacteria have specific pH ranges in which they thrive. Generally, DF for H_2 production occurs under slightly acidic to neutral pH conditions, typically around pH 5.5 to 7.5. Maintaining the appropriate pH range is crucial for supporting the growth and metabolic activity of the H_2-producing microorganisms [76,77,78]. The less significant values of four variables in the present study, i.e., biomass, temperature, Fe and Ni in Fig. 8 for all ML models, can be accredited to the fact that the optimal range of temperature and biomass in the DF process has already been known. However, biomass, temperature, Fe and Ni should also be considered significant contributing parameters to producing bio-H_2 by DF [79]. Fe and nickel Ni are trace elements that can serve as cofactors for specific enzymes involved in H_2

metabolism. They are components of hydrogenases, key enzymes responsible for H_2 production. While Fe and Ni are essential for the proper functioning of hydrogenase enzymes, their presence is often not a limiting factor in DF, as these trace elements are generally present in sufficient amounts in the natural environment or can be obtained from various sources [80]. It is important to note that the significance of these factors can vary depending on specific conditions, such as the microbial community, the particular process configuration, and the nature of the substrate. In some cases, biomass composition, temperature optimization, or adding Fe and Ni may have a more pronounced impact on H_2 production, especially when dealing with specific substrates or unique microbial consortia. Therefore, a comprehensive evaluation of these factors within the particular context of a DF system is recommended to optimize H_2 production in future studies.

However, the temperature of the PEM electrolysis cell plays a crucial role (Fig. 9) in determining the efficiency and overall performance of the process. Higher temperatures generally result in improved electrolysis kinetics and lower electrical resistances within the cell, leading to increased H_2 production rates [81,82]. The reaction kinetics at the electrodes are typically enhanced at elevated temperatures, allowing for faster ion transport and more efficient water splitting. However, it's important to note that there is an optimal temperature range for PEM electrolysis, and extreme temperatures can negatively impact the durability and stability of the materials used in the cell [83,84]. The cell area refers to the active electrode surface area within the PEM electrolysis cell. A larger cell area provides more surface area for the electrochemical reactions to occur, thereby increasing the rate of H_2 production [85]. By increasing the cell area, the available reaction sites for water splitting are increased, resulting in a higher current density and, subsequently, higher H_2 production rates. However, increasing the cell area also increases the overall system cost and requires more electrical power to maintain the desired current density [86]. In summary, optimizing the cell temperature and cell area in PEM water electrolysis is crucial for achieving higher H_2 production rates and improving the efficiency of the process. By controlling these parameters within suitable ranges, researchers and engineers can enhance the performance and viability of PEM electrolysis systems for H_2 production.

The pressure within the PEM electrolysis cell affects the diffusion of gases, including H_2 and oxygen. Higher cell pressures promote more efficient gas diffusion, allowing the gases to reach the respective electrode surfaces more quickly. This facilitates the electrochemical reactions involved in water splitting, resulting in increased H_2 production rates. The performance of the PEM in water electrolysis is influenced by the cell pressure [87]. The PEM acts as a barrier between the H_2 and oxygen electrodes, selectively allowing protons to pass through while blocking the passage of gas molecules. Adequate cell pressure helps maintain proper membrane hydration and compression, ensuring optimal proton conductivity. This promotes efficient proton transfer and overall electrolysis performance [88]. Cell pressure also affects water management within the PEM electrolysis cell. Water is an essential component for proton conduction in the PEM, and it helps maintain membrane hydration. Proper water management is crucial for the PEM to function effectively. Optimal cell pressure assists in maintaining the appropriate water content within the membrane, which ensures efficient proton transport and supports the electrochemical reactions involved in H_2 production [89]. It's important to note that there are practical limits to cell pressure due to the structural integrity of the system. Excessive pressure can lead to mechanical stresses, gas leakage, or even membrane damage. Therefore, the cell pressure must be carefully controlled within safe and operationally feasible ranges to optimize H_2 production in PEM water electrolysis.

The cell voltage, also known as the applied voltage or overpotential, represents the energy required to overcome the thermodynamic potential of the water-splitting reactions. In PEM electrolysis, water is split into H_2 and oxygen by applying an external electrical potential across the cell [90]. The cell voltage must exceed the thermodynamic potential of the reaction to drive the electrolysis process. Higher cell voltages provide the necessary driving force to accelerate the reactions and increase the rate of H_2 production. Cell voltage also influences the ohmic losses within the PEM electrolysis cell. Ohmic losses are associated with the electrical resistance of the various components in the cell, including the electrodes, electrolyte, and membrane [91]. Higher cell voltages can help mitigate ohmic losses by providing a greater driving force to overcome these resistances. Minimizing ohmic losses is essential to maximize the efficiency of the PEM electrolysis process and optimize H_2 production [92].

The amount of anode and cathode catalysts is essential in PEM water electrolysis because catalysts play a critical role in facilitating the electrochemical reactions that occur at the electrodes [93,94]. Catalysts enhance the reaction kinetics at the anode and cathode of the PEM electrolysis cell. At the anode, the catalyst facilitates the oxidation of water molecules,

generating protons, electrons, and oxygen. At the cathode, the catalyst promotes the reduction of protons and electrons to form H_2 gas [95]. The catalyst provides active sites for the reactions to occur, lowering the activation energy and increasing the reaction rates. The amount of catalyst determines the number of available active sites, directly influencing the efficiency and speed of the electrochemical reactions [96]. The amount of catalyst also affects the achievable current density in PEM water electrolysis. Current density refers to the amount of electrical current passing through a given area of the electrode surface. Higher catalyst loadings can increase the active surface area, providing more sites for the electrochemical reactions to take place [97]. This results in a higher current density, which corresponds to increased H_2 production rates. Optimizing the catalyst loading is crucial for maximizing current density and overall system performance [98].

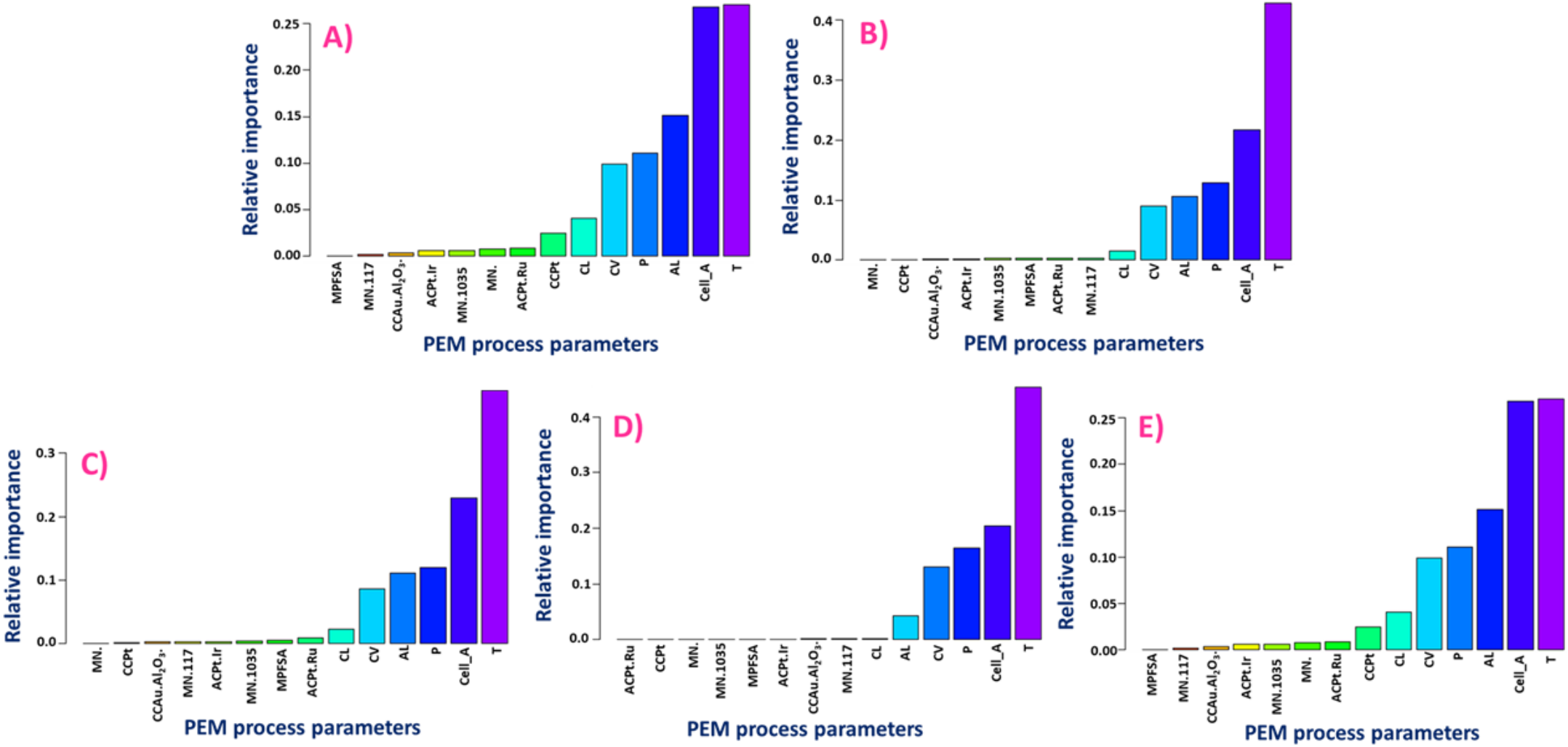


Fig. 9. Influences of different process parameters on H_2 production by PEM techniques; A) KNN, B) DT, C) RF, D) GBM, E) SVM.

4.5 Partial dependence analysis (PDA) of best ML models for DF and PEM.

This study found that the K-nearest neighbors (KNN) model is the best ML model for predicting H_2 production in the DF technology. Based on PVI analysis, a Partial Dependency Analysis (PDA) study was conducted on the top six influential process parameters in the DF process. The results of the PDA, depicted in Fig. 10, demonstrate that the H_2 production rate increases with increasing COD values up to 10 mg/L. However, beyond this point, further increases in COD values lead to a decrease in the optimal H_2 production. It is generally recognized that COD values up to 10 mg/L are beneficial for H_2 production [55,99]. This is because a certain amount of organic compounds is necessary to serve as a carbon source for the microorganisms involved in the H_2 production process [66,100]. These microorganisms utilize these organic compounds as a substrate, resulting in the production of hydrogen. At higher COD concentrations, certain organic compounds can inhibit microbial activity and be toxic to hydrogen-producing microorganisms, leading to a reduction in H_2 production rate [63,101]. Butyrate concentration variations, on the other hand, did not have a significant impact on H_2 production [74,102].

Regarding the temperature, it was found that the optimal temperature range for H_2 production in the DF process is between 25 to 30 °C. Increasing the temperature beyond 30°C can decrease H_2 production. The DF process relies on the activity of anaerobic microorganisms, particularly fermentative bacteria, which convert organic substrates into H_2 and other byproducts. These microorganisms exhibit optimal metabolic activity within the temperature range of 25 to 30 °C, ensuring their efficient production of H_2 [82,103]. The study also examined the effects of pH within the range of 4 to 7 and found no measurable impacts on H_2 production in the DF process [78,104]. As for acetate concentration, up to 2 g/L is considered excellent for H_2 production. However, further increases in acetate concentration result in a significant decrease in H_2 production. High acetate concentrations promote alternative metabolic pathways, such as acetogenesis or methanogenesis, over H_2 production. Acetate can be utilized by acetogenic bacteria to produce acetate or other byproducts instead of hydrogen. Additionally, methanogenic archaea can use acetate to generate methane. When these alternative pathways become dominant, H_2 production is suppressed [75,105].

The best H_2 predictive ML model for PEM technology is RF in this study, thus based on PVI analysis, PDA study was conducted on the top six influential process parameters in the PEM process. The results of PDA of the process parameters in PEM technology have been illustrated in Fig. 11. It is well evident that cell temperature up to 35 ^{0}C should be considered optimum

for generating H_2 in the PEM process, further increasing of temperature could diminish the H_2 production to the greatest extent [81,106]. However, increasing cell area could increase the H_2 production rate. According to PDA plots, 40-70 cm^2 could produce a significant production rate of H_2 while very low cell areas (0-19 cm^2) can be a matter of concern for H_2 production in the PEM process [85,107]. Nonetheless, the cathode catalysts loading range from 0-3 mg/ cm^2 did not have any significant impact on H_2 production by PEM. In contrast, H_2 production is quite less with a small quantity of anode (0-2 mg/ cm^2) catalysts. However, further increases in anode catalysts up to 4 mg/ cm^2 could increase the H_2 production rate to the greatest extent [94,108]. The cell pressure plays a crucial role in the production of H_2 in the PEM process. However, 0-10 atm cell pressure should be considered optimum for H_2 production and further increase of cell pressure decreased the H_2 quantity. In addition, 2-4 V cell voltage is considered excellent for H_2 generation in the PEM process. Very low voltage (0-2 V) could reduce the H_2 yield to some extent [109].

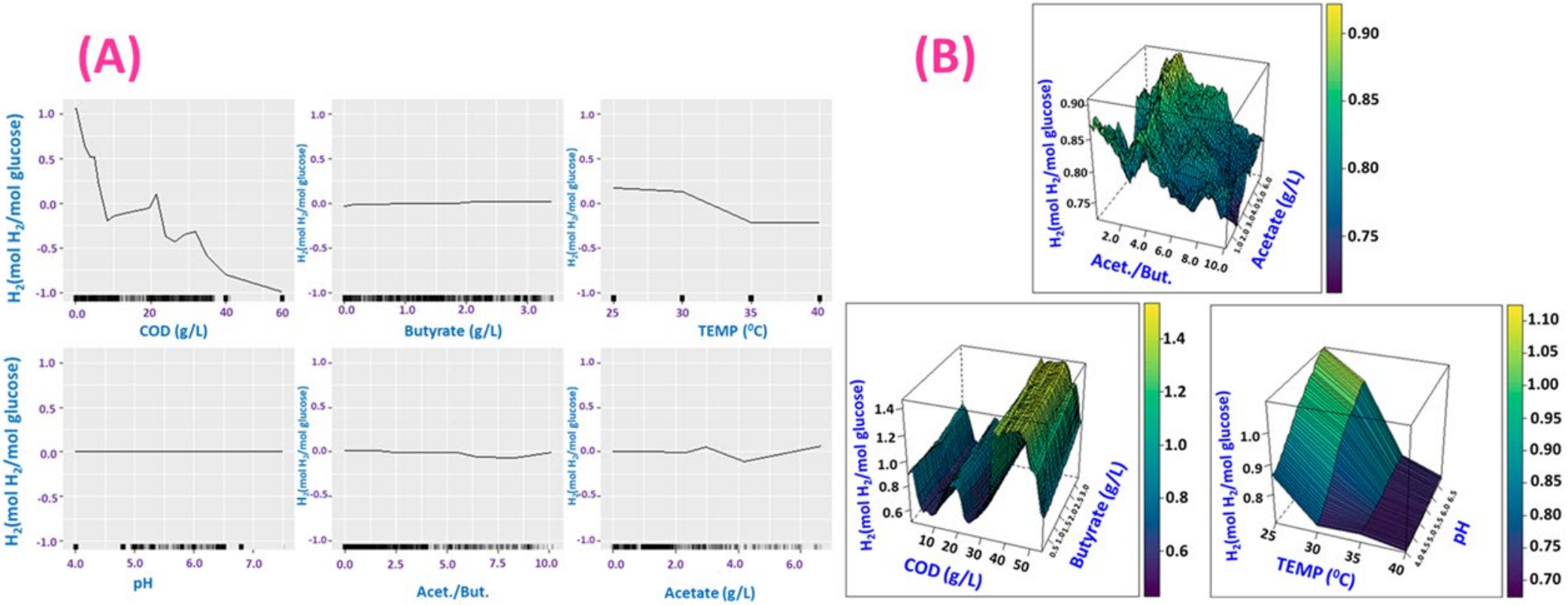


Fig. 10 Partial dependence plots of H_2 production on process parameter: (A) one-variable, (B) two-variable for DF.

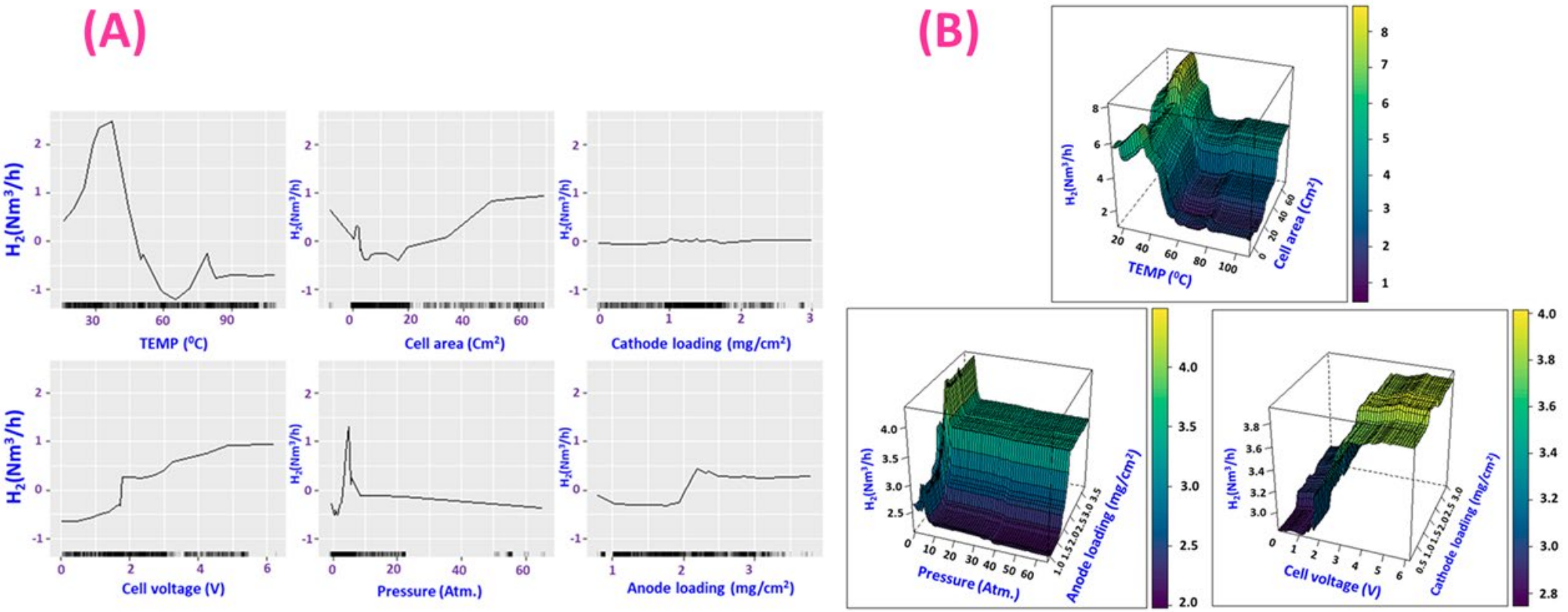


Fig. 11. Partial dependence plots of H_2 production on process parameter: (A) one-variable, (B) two-variable for PEM.

5. Conclusion

The establishment of a future H_2 economy (HE) through environmentally friendly green H_2 production (GHP) holds immense potential as the ultimate solution for ensuring global energy security in a sustainable manner. Incorporating the perspectives from water industries (WRIs) and the circular economy (CRE) becomes paramount in the context of GHP and sustainable energy management. The findings from our study indicate that K-nearest neighbors (KNN) and random forest (RF) models prove to be the most suitable for predicting GHP in the case of dark fermentation (DF) and proton exchange membrane (PEM) water electrolysis processes, respectively. Our analysis also highlights the significance of specific process parameters (PP) in GHP optimization. For DF, the permutation variable index (PVI) underscores that chemical oxygen demand (COD), followed by butyrate concentration, temperature, pH level, and the acetate/butyrate ratio, exert the most substantial influence, listed in descending order. Conversely, for the PEM process, temperature, cell area, cell pressure, cell voltage, and catalyst loadings take precedence as the most influential process parameters, presented in a diminishing order. Furthermore, partial dependency analysis (PDA) reveals key insights. GHP exhibits an upward trend with COD values up to 10 mg/L, and the DF process yields optimal results within a temperature range of 25 to 30 ^{0}C. In contrast, the PEM process benefits from a cell temperature of up to 35 ^{0}C, and a cell area ranging from 40 to 70 cm^2 proves most effective for GHP. Regarding acetate concentration, concentrations up to 2 g/L are ideal for H_2 production in DF, but further increases lead to diminishing returns. Conversely, in the PEM process, minimal quantities of anode catalysts (0-2 mg/cm^2) result in limited H_2 production, while increasing the catalyst loadings up to 4 mg/cm^2 significantly enhances H_2 production rates. Cell pressure emerges as a pivotal factor in PEM-based H_2 production, with an optimal range of 0-10 atm; higher pressures reduce H_2 output. Additionally, a cell voltage of 2-4 V is deemed optimal for H_2 generation in the PEM process, with very low voltages (0-2 V) leading to reduced H_2 yields. In summary, our study underscores the potential of machine learning (ML) and artificial intelligence (AI) as promising techniques for optimizing GHP, ultimately addressing scaling-up challenges in large-scale industrial GHP production and ensuring a sustainable HE.

Acknowledgments

This project was supported by the Australian Research Council (ARC) through the ARC Research Hub for Nutrients in a Circular Economy (NiCE) (IH210100001) and Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE) (P0017310, Human Resource Development Program for Industrial Innovation (Global). M. M. Kabir is grateful for the financial support from the International Research Scholarship (IRS) and UTS President's Scholarship (UTSP) for his PhD studies.

References

1. T. Ahmad, D. Zhang, A critical review of comparative global historical energy consumption and future demand: the story told so far. Energy Rep. 6 (2020), 1973–1991. https://doi.org/10.1016/j.egyr.2020.07.020
2. M. David, C. Ocampo-Martínez, R. Sánchez-Peña, Advances in alkaline water electrolyzers: A review, J. Energy Storage, 23 (2019), 392-403, https://doi.org/10.1016/j.est.2019.03.001
3. M. Yu, K. Wang, H. Vredenburg, Insights into low-carbon hydrogen production methods: Green, blue and aqua hydrogen, Int. J. Hydrog. Energy, 46 (2021), 21261-21273, https://doi.org/10.1016/j.ijhydene.2021.04.016
4. Q. Cui, H. B. Kuang, C. Y. Wu, Y. Li, The changing trend and influencing factors of energy efficiency: the case of nine countries. Energy, 64 (2014), 1026-1034, https://doi.org/10.1016/j.energy.2013.11.060
5. M.M. Kabir, M.M. Akter, Z. Huang, L. Tijing, H.K. Shon, Hydrogen production from water industries for a circular economy. Desalination, 554 (2023), 116448, https://doi.org/10.1016/j.desal.2023.116448
6. S. Sebbahi, N. Nabil, A. Alaoui-Belghiti, S. Laasri, S. Rachidi, A. Hajjaji, Assessment of the three most developed water electrolysis technologies: alkaline water electrolysis, proton exchange membrane and solid-oxide electrolysis, Mater. Today: Proc. 66 (2022), 140-145, https://doi.org/10.1016/j.matpr.2022.04.264
7. L. Sillero, W. G. Sganzerla, T. Forster-Carneiro, R. Solera, M. Perez, A bibliometric analysis of the hydrogen production from dark fermentation, Int. J. Hydrog. Energy 47 (2022), 27397–27420. https://doi.org/10.1016/j.ijhydene.2022.06.083

8. A. Sridhar, M. Ponnuchamy, P. S. Kumar, A. Kapoor, L. Xiao, Progress in the production of hydrogen energy from food waste: A bibliometric analysis, Int. J. Hydrog. Energy, 47 (2022), 26326-26354, https://doi.org/10.1016/j.ijhydene.2021.09.258
9. H. J. Alves, C. B. Junior, R. R. Niklevicz, E. P. Frigo, M. S. Frigo, C. H. Coimbra-Araújo, Overview of hydrogen production technologies from biogas and the applications in fuel cells, Int. J. Hydrog. Energy, 38 (2013), 5215-5225, https://doi.org/10.1016/j.ijhydene.2013.02.057
10. A. Mohanty, M. Mankoti, P. R. Rout, S. S. Meena, S. Dewan, B. Kalia, S. Varjani, J. W. C. Wong, J. R. Banu, Sustainable utilization of food waste for bioenergy production: a step towards circular bioeconomy, Int. J. Food Microbiol. 365 (2022), 109538. https://doi.org/10.1016/j.ijfoodmicro.2022.109538.
11. B. Kayal, D. Abu-Ghunmi, L. Abu-Ghunmi, A. Archenti, M. Nicolescu, C. Larkin, S. Corbet, An economic index for measuring firm's circularity: the case of water industry, J. Behav. Exp. Financ. 21(2019), 123–129. https://doi.org/10.1016/j.jbef.2018.11.007.
12. M. David, C. Ocampo-Martínez, R. Sánchez-Peña, Advances in alkaline water electrolyzers: A review, J. Energy Storage, 23 (2019), 392-403, https://doi.org/10.1016/j.est.2019.03.001
13. L. Camargo, D. Comas, Y. C. Escorcia, A. Alviz-Meza, G. Carrillo Caballero, I. Portnoy, Bibliometric Analysis of Global Trends around Hydrogen Production Based on the Scopus Database in the Period 2011–2021, Energies, 16 (2022), 87, https://doi.org/10.3390/en16010087
14. M.N.Z. Abidin, M.M. Nasef, J. Veerman, Towards the development of new generation of ion exchange membranes for reverse electrodialysis: A review, Desalination, 537 (2022), 115854, https://doi.org/10.1016/j.desal.2022.115854
15. I. Ihsanullah, J. Mustafa, A.M. Zafar, M. Obaid, M.A. Atieh, N. Ghaffour, Waste to wealth: A critical analysis of resource recovery from desalination brine, Desalination, 543 (2022), 116093, https://doi.org/10.1016/j.desal.2022.116093
16. M. Yu, K. Wang, H. Vredenburg, Insights into low-carbon hydrogen production methods: Green, blue and aqua hydrogen, Int. J. Hydrog. Energy, 46 (2021), 21261-21273, https://doi.org/10.1016/j.ijhydene.2021.04.016
17. M. Alhassan, A. A. Jalil, W. Nabgan, M. Y. S. Hamid, M. B. Bahari, M. Ikram, Bibliometric studies and impediments to valorization of dry reforming of methane for hydrogen production, Fuel. 328, (2022), 125240, https://doi.org/10.1016/j.fuel.2022.125240

18. J. Manna, P. Jha, R. Sarkhel, C. Banerjee, A. K. Tripathi, M. R. Nouni, Opportunities for green hydrogen production in petroleum refining and ammonia synthesis industries in India. Int. J. Hydrog. Energy, 46 (2021), 38212-38231, https://doi.org/10.1016/j.ijhydene.2021.09.064
19. P. Ghisellini, S. Ulgiati, 2020. Circular economy transition in Italy. Achievements, perspectives and constraints, J. Clean. Prod. 243 (2020), 118360, https://doi.org/10.1016/j.jclepro.2019.118360
20. A. Sridhar, M. Ponnuchamy, P. S. Kumar, A. Kapoor, L. Xiao, Progress in the production of hydrogen energy from food waste: A bibliometric analysis, Int. J. Hydrog. Energy, 47 (2022), 26326-26354, https://doi.org/10.1016/j.ijhydene.2021.09.258
21. H. J. Alves, C. B. Junior, R. R. Niklevicz, E. P. Frigo, M. S. Frigo, C. H. Coimbra-Araújo, Overview of hydrogen production technologies from biogas and the applications in fuel cells, Int. J. Hydrog. Energy, 38 (2013), 5215-5225, https://doi.org/10.1016/j.ijhydene.2013.02.057
22. A. Sharifi, D. Simangan, S. Kaneko, Three decades of research on climate change and peace: A bibliometrics analysis, Sustain. Sci. 16 (2021), 1079-1095, https://doi.org/10.1007/s11625-020-00853-3
23. L. Niu, X. Zhao, F. Wu, Z. Tang, H. Lv, J. Wang, M. Fang, J. P. Giesy, Hotpots and trends of covalent organic frameworks (COFs) in the environmental and energy field: Bibliometric analysis, Sci. Total Environ. 783 (2021), 146838, https://doi.org/10.1016/j.scitotenv.2021.146838
24. J.M. Lee, S.H. Lee, J.H. Baik, K. Park, Techno-economic analysis of hydrogen production electrically coupled to a hybrid desalination process, Desalination, 539 (2022), 115949, https://doi.org/10.1016/j.desal.2022.115949
25. N. Sreedhar, N. Thomas, N. Ghaffour, H.A. Arafat, The evolution of feed spacer role in membrane applications for desalination and water treatment: A critical review and future perspective, Desalination, 554 (2023), 116505, https://doi.org/10.1016/j.desal.2023.116505
26. W. Sohn, J. Jiang, S. Phuntsho, Y. Choden, H.K. Shon, Nutrients in a circular economy: Role of urine separation and treatment, Desalination, 560 (2023), 116663, https://doi.org/10.1016/j.desal.2023.116663
27. Y.P. Xu, Z.H. Lin, T.X. Ma, C. She, S.M. Xing, L.Y. Qi, S.G. Farkoush, J. Pan, Optimization of a biomass-driven Rankine cycle integrated with multi-effect desalination,

and solid oxide electrolyzer for power, hydrogen, and freshwater production, Desalination, 525 (2022), 115486, https://doi.org/10.1016/j.desal.2021.115486

28. Z. Li, M. Gong, M. Wang, A. Feng, L. Wang, P. Ma, P., 2021a. Influence of AlCl3 and oxidant catalysts on hydrogen production from the supercritical water gasification of dewatered sewage sludge and model compounds, Int. J. Hydrog. Enrg. 46 (2021a), 31262–31274, https://doi.org/10.1016/j.ijhydene.2021.07.028
29. J. Li, L. Pan, M. Suvarna, X. Wang, 2021b. Machine learning aided supercritical water gasification for H2-rich syngas production with process optimization and catalyst screening, Chem. Eng. J. 426 (2021b), 131285, https://doi.org/10.1016/j.cej.2021.131285
30. J.E. Baeten, M.C.M. van Loosdrecht, E.I.P. Volcke, Modelling aerobic granular sludge reactors through apparent half-saturation coefficients. Water Res. 146 (2018), 134–145, https://doi.org/10.1016/j.watres.2018.09.025
31. M.S. Zaghloul, O.T. Iorhemen, R.A. Hamza, J.H. Tay, G. Achari, G., Development of an ensemble of machine learning algorithms to model aerobic granular sludge reactors, Water research, 189 (2021), 116657, https://doi.org/10.1016/j.watres.2020.116657
32. T.H. Miller, M.D. Gallidabino, J.I. MacRae, S.F. Owen, N.R. Bury, L.P. Barron, L.P., Prediction of bioconcentration factors in fish and invertebrates using machine learning, Sci. of the Total Env. 648 (2019), 80-89, https://doi.org/10.1016/j.scitotenv.2018.08.122
33. G. Biau, L. Devroye, V. Dujmović, A. Krzyżak, An affine invariant k-nearest neighbor regression estimate, J. Multivar. Anal. 112 (2012), 24-34, https://doi.org/10.1016/j.jmva.2012.05.020
34. S. Ghosh, A. Roy, S. Chakraborty, Support vector regression-based metamodeling for seismic reliability analysis of structures, Appl. Math. Model. 64 (2018), 584-602, https://doi.org/10.1016/j.apm.2018.07.054
35. A. De Caigny, K. Coussement, K.W. De Bock, A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees, Eur. J. Oper. Res., 269 (2018), 760-772, https://doi.org/10.1016/j.ejor.2018.02.009
36. A.C. Serfidan, F. Uzman, M. Türkay, Optimal estimation of physical properties of the products of an atmospheric distillation column using support vector regression, Computers & Chemical Engineering, 134 (2020), 106711, https://doi.org/10.1016/j.compchemeng.2019.106711

37. X. Zhao, S. Yang, Y. Hou, H. Gao, Y. Wang, D.A. Gribble, V.G. Pol, Recent progress on key materials and technical approaches for electrochemical lithium extraction processes, Desalination, 546 (2023), 116189, https://doi.org/10.1016/j.desal.2022.116189
38. A. Zendehboudi, M.A. Baseer, R. Saidur, Application of support vector machine models for forecasting solar and wind energy resources: a review, J. Clean. Prod. 199 (2018), 272–285, https://doi.org/10.1016/j.jclepro.2018.07.164
39. J. Cai, K. Xu, Y. Zhu, F. Hu, L. Li, Prediction and analysis of net ecosystem carbon exchange based on gradient boosting regression and random forest, Appl. Energy. 262 (2020), 114566, https://doi.org/10.1016/j.apenergy.2020.114566
40. B. Grillone, S. Danov, A. Sumper, J. Cipriano, G. Mor, A review of deterministic and data-driven methods to quantify energy efficiency savings and to predict retrofitting scenarios in buildings, Renew. Sust. Energ Rev. 131 (2020), 110027, https://doi.org/10.1016/j.rser.2020.110027
41. H. Nguyen, T. Vu, T.P. Vo, H.T. Thai, Efficient machine learning models for prediction of concrete strengths, Constr. Build. Mater. 266 (2021), 120950, https://doi.org/10.1016/j.conbuildmat.2020.120950
42. L. Zhou, H. Fujita, H. Ding, R. Ma, Credit risk modeling on data with two timestamps in peer-to-peer lending by gradient boosting, Appl. Soft Comput. 110 (2021), 107672, https://doi.org/10.1016/j.asoc.2021.107672
43. J. Ma, J.C.P. Cheng, Identifying the influential features on the regional energy use intensity of residential buildings based on Random Forests, Appl. Energy, 183 (2016), 193–201, https://doi.org/10.1016/j.apenergy.2016.08.096
44. Y. Li, C. Zou, M. Berecibar, E. Nanini-Maury, J.C.W. Chan, P. van den Bossche, J. Van Mierlo, N. Omar, Random Forest regression for online capacity estimation of lithium-ion batteries, Appl. Energy, 232 (2018), 197–210, https://doi.org/10.1016/j.apenergy.2018.09.182
45. M. Zahid, N. Savla, S. Pandit, V.K. Thakur, S.P. Jung, P.K. Gupta, R. Prasad, E. Marsili, Microbial desalination cell: Desalination through conserving energy, Desalination, 521 (2022), 115381, https://doi.org/10.1016/j.desal.2021.115381

46. X. Wang, K. An, L. Tang, X. Chen, Short term prediction of freeway exiting volume based on SVM and KNN, Int. J. Transp. Sci. Technol. 4 (2015), 337-352, https://doi.org/10.1260/2046-0430.4.3.337

47. Y. He, D. Li, K. Zuo, F. Yang, T. Gao, P. Liang, Recover phosphorus as vivianite using a dual-chamber electrochemical reactor, Desalination, 550 (2023), 116385, https://doi.org/10.1016/j.desal.2023.116385

48. A. Mohammadifar, H. Gholami, J.R. Comino, A.L. Collins, Assessment of the interpretability of data mining for the spatial modelling of water erosion using game theory, Catena, 200 (2021), 105178, https://doi.org/10.1016/j.catena.2021.105178

49. B.J. Kim, H.K. Shon, D.S. Han, H. Park, H., In-situ desalination-coupled electrolysis with concurrent one-step-synthesis of value-added chemicals, Desalination, 551 (2023), 116431, https://doi.org/10.1016/j.desal.2023.116431

50. M. Alexandropoulou, G. Antonopoulou, G. Lyberatos, G., Modeling of continuous dark fermentative hydrogen production in an anaerobic up-flow column bioreactor, Chemosphere, 293 (2022), 133527, https://doi.org/10.1016/j.chemosphere.2022.133527

51. V.H. Trinh, N.A. Nguyen, O. Omelianovych, V.D. Dao, I. Yoon, H.S. Choi, M. Keidar, Sustainable desalination device capable of producing freshwater and electricity, Desalination, 535 (2022), 115820, https://doi.org/10.1016/j.desal.2022.115820

52. F. Xiao, Q. Wang, G.L. Xu, X. Qin, I. Hwang, C.J. Sun, M. Liu, W. Hua, H.W. Wu, S. Zhu, J.C. Li, Atomically dispersed Pt and Fe sites and Pt–Fe nanoparticles for durable proton exchange membrane fuel cells, Nat. Catal. 5 (2022), 503-512, https://doi.org/10.1038/s41929-022-00796-1

53. E. Ogungbemi, T. Wilberforce, O. Ijaodola, J. Thompson, A.G. Olabi, A.G., Review of operating condition, design parameters and material properties for proton exchange membrane fuel cells, Int. J. Energy Res. 45 (2021), 1227-1245, https://doi.org/10.1002/er.5810

54. A. Hosseinzadeh, J.L. Zhou, A. Altaee, D. Li, Machine learning modeling and analysis of biohydrogen production from wastewater by dark fermentation process. Bioresour. Technol. 343 (2022), 126111, https://doi.org/10.1016/j.biortech.2021.126111

55. A. Akhbari, S. Ibrahim, A.A. Zinatizadeh, H. Bonakdari, I. Ebtehaj, S. Khozani, M. Vafaeifard, B. Gharabaghi, Evolutionary prediction of biohydrogen production by dark fermentation, CLEAN–Soil, Air, Water, 47 (2019), 1700494, https://doi.org/10.1002/clen.201700494
56. S.S. Khan, H. Shareef, A.H. Mutlag, Dynamic temperature model for proton exchange membrane fuel cell using online variations in load current and ambient temperature, Int. J. Green Energy, 16 (2019), 361-370, https://doi.org/10.1080/15435075.2018.1564141
57. R. Li, D. Wu, J. Song, Y. He, W. Zhu, X. Wang, L. Wang, N.M. Dube, K. Jiang, In situ generation of reduced graphene oxide on 3D CuNi foam as high-performance electrodes for capacitive deionization, Desalination, 540 (2022), 115990, https://doi.org/10.1016/j.desal.2022.115990
58. F. Wang, S. Zhao, X. Zhang, Z. Su, Interfacial solar evaporation based on Janus films: An effective strategy to improve salt tolerance and antifouling performance, Desalination, 543 (2022), 116085, https://doi.org/10.1016/j.desal.2022.116085
59. N. Pradhan, L. Dipasquale, G. d'Ippolito, A. Fontana, A. Panico, F. Pirozzi, P.N.L. Lens, G. Esposito, Model development and experimental validation of capnophilic lactic fermentation and hydrogen synthesis by Thermotoga neapolitana, Water Res. 99 (2016), 225–234, https://doi.org/10.1016/j.watres.2016.04.063
60. R. Ding, S. Zhang, Y. Chen, Z. Rui, K. Hua, Y. Wu, X. Li, X. Duan, X. Wang, J. Li, J. Liu, Application of machine learning in optimizing proton exchange membrane fuel cells: A review, Energy and AI, (2022), 100170, https://doi.org/10.1016/j.egyai.2022.100170
61. K. Jiao, J. Xuan, Q. Du, Z. Bao, B. Xie, B. Wang, Y. Zhao, L. Fan, H. Wang, Z. Hou, S. Huo, Designing the next generation of proton-exchange membrane fuel cells, Nature, 595 (2021), 361-369, https://doi.org/10.1038/s41586-021-03482-7
62. R.N. Moussa, N. Moussa, D. Dionisi, Hydrogen production from biomass and organic waste using dark fermentation: an analysis of literature data on the effect of operating parameters on process performance, Processes, 10 (2022), 156, https://doi.org/10.3390/pr10010156
63. G. Policastro, R. Lamboglia, M. Fabbricino, F. Pirozzi, Enhancing dark fermentative hydrogen production from problematic substrates via the co-fermentation strategy, Fermentation, 8 (2022), 706, https://doi.org/10.3390/fermentation8120706
64. X. Qu, H. Zeng, Y. Gao, T. Mo, Y. Li, Bio-hydrogen production by dark anaerobic fermentation of organic wastewater, Front. Chem. 10 (2022), 978907, https://doi.org/10.3389/fchem.2022.978907.

65. X. Gu, Y. Wang, H. Li, J. Li, S. Wang, Characteristics of biohydrogen production and performance of hydrogen-producing acetogen by increasing normal molasses wastewater proportion in anaerobic baffled reactor, Archaea. 5 (2020), 8885662, https://doi.org/10.1155/2020/8885662.
66. C. Rangel, J. Sastoque, J. Calderon, J. Mosquera, P. Velasquez, I. Cabeza, P. Acevedo, Hydrogen production by dark fermentation process: effect of initial organic load, Chem. Eng. Trans. 79 (2020), 133-138, https://doi.org/10.3303/CET2079023.
67. E. L. N. Dzulkarnain, J. O. Audu, W. R. Z. W. Dagang, M. F. Abdul-Wahab, Microbiomes of biohydrogen production from dark fermentation of industrial wastes: current trends, advanced tools and future outlook, Bioresour. Bioprocess. 9 (2022), 16, https://doi.org/10.1186/s40643-022-00504-8
68. R.M. Silva, A.A. Abreu, A.F. Salvador, M. M. Alves, I. C. Neves, M. A. Pereira, Zeolite addition to improving biohydrogen production from dark fermentation of C5/C6-sugars and *Sargassum sp.* biomass. Sci Rep 11 (2021) 16350, https://doi.org/10.1038/s41598-021-95615-1.
69. I. Martins, E. Surra, M. Ventura, N. Lapa, Bio-H2 from dark fermentation of OFMSW: effect of the hydraulic retention time and organic loading rate, Appl. Sci. 12 (2022), 4240, https://doi.org/10.3390/app12094240
70. T. D. Marques, W. V. Macedo, F. S. Peiter, A. A. T. L. Bonfim, I. K. Sakamoto, R. A. C. Filho, M. H. Z. Damianovic, M. B. A. Varesche, K. R. Salomon, E. L. C. de Amorim, Influence of hydraulic retention time on hydrogen production by treating cheese whey wastewater in anaerobic fluidized bed bioreactor-an approach for developing countries, Braz. J. Chem. Eng. 36 (2019), 1109 - 1118, https://doi.org/10.1590/0104-6632.20190363s20190075
71. Y. Chai, Z. Lyu, H. Du, P. Li, S. Ding, Y. Jiang, H. Wang, Q. Min, D. Du, Y. Lin, W. Zhu, Recent progress on rational design of catalysts for fermentative hydrogen production, SusMat, 2 (2022), 392-410, https://doi.org/10.1002/sus2.75
72. A. Detman, D. Laubitz, A. Chojnacka, P.R. Kiela, A. Salamon, A. Barberán, Y. Chen, F. Yang, M.K. Błaszczyk, A. Sikora, Dynamics of dark fermentation microbial communities in the light of lactate and butyrate production, Microbiome, 9 (2021), 1-21, https://doi.org/10.1186/s40168-021-01105-x.
73. N. Mahmud, D.V.F. Alvarez, M.H. Ibrahim, M.H. El-Naas, D.V. Esposito, Magnesium recovery from desalination reject brine as pretreatment for membraneless electrolysis, Desalination, 525 (2022), 115489, https://doi.org/10.1016/j.desal.2021.115489

74. W. Kong, X. Ge, M. Yang, Q. Zhang, J. Lu, H. Wen, H. Wen, D. Kong, M. Zhang, X. Zhu, Y. Feng, Poly-p-phenylene as a novel pseudocapacitive anode or cathode material for hybrid capacitive deionization, Desalination, 553 (2023), 116452, https://doi.org/10.1016/j.desal.2023.116452
75. D. Liu, D. Liu, R.J. Zeng, I. Angelidaki, Hydrogen and methane production from household solid waste in the two-stage fermentation process, Water Res. 40 (2006), 2230–2236, https://doi.org/10.1016/j.watres.2006.03.029
76. S. V. tangkel, S. Sung, J.-J. Lay, Biohydrogen production as a function of pH and substrate concentration, Environ. Sci. Technol. 35 (2001), 4726–4730, https://doi.org/10.1021/es001979r
77. J. Wongthanate, K. Chinnacotpong, M. Khumpong, Impacts of pH, temperature and pretreatment method on biohydrogen production from organic wastes by sewage microflora. Int. J. Energy Environ. Eng. 5(2014), https://doi.org/10.1007/s40095-014-0076-6.
78. T. Tang, Y. Chen, M. Liu, Y. Du, Y. Tan, Effect of pH on the performance of hydrogen production by dark fermentation coupled denitrification, Environ. Res. 208 (2022), 112663, https://doi.org/10.1016/j.envres.2021.112663.
79. A. Hosseinzadeh, J.L. Zhou, A. Altaee, D. Li, Machine learning modeling and analysis of biohydrogen production from wastewater by dark fermentation process. Bioresour. Technol. 343 (2022), 126111, https://doi.org/10.1016/j.biortech.2021.126111
80. H.M. Ashraf, S.A. Al-Sobhi, M.H. El-Naas, Mapping the desalination journal: A systematic bibliometric study over 54 years, Desalination, 526 (2022), 115535, https://doi.org/10.1016/j.desal.2021.115535
81. F. Scheepers, M. Stähler, A. Stähler, E. Rauls, M. Muller, M. Carmo, W. Lehnert, Temperature optimization for improving polymer electrolyte membrane-water electrolysis system efficiency, Appl. Energy 283 (2021), 116270, https://doi.org/10.1016/j.apenergy.2020.116270.
82. S. S. Kumar, V. Himabindu, Hydrogen production by PEM water electrolysis- a review, Mater. Sci. Energy Technol. 2 (2019), 442-454, https://doi.org/10.1016/j.mset.2019.03.002.
83. M. Bonanno, K. Muller, B. Bensmann, R. Hanke-Rauschenbach, R. Peach, S. Thiele, Evaluation of the efficiency of an elevated temperature proton exchange membrane water electrolysis system, J. Electrochem. Soc. 168 (2021), 094504, https://doi.org/10.1149/1945-7111/ac2188

84. T. Wang, X. Cao, L. Jiao, PEM water electrolysis for hydrogen production: fundamentals, advances, and prospects, Carb. Neutrality, 1 (2022), 21, https://doi.org/10.1007/s43979-022-00022-8.
85. A. Martin, P. Trinke, M. Stahler, A. Stahler, F. Scheepers, B. Bensmann, M. Carmo, W. Lehnert, R. Hanke-Rauschenbach, The effect of cell compression and cathode pressure on hydrogen crossover in PEM water electrolysis, J. Electrochem. Soc. 169 (2022), 014502, https://doi.org/10.1149/1945-7111/ac4459.
86. J. Lim, C. Joo, J. Lee, H. Cho, J. Kim, Novel carbon-neutral hydrogen production process of steam methane reforming integrated with desalination wastewater-based CO2 utilization, Desalination, 548 (2023), 116284, https://doi.org/10.1016/j.desal.2022.116284
87. S. Li, M. Dai, I. Ali, H. Bian, C. Peng, A new idea for efficient copper recovery from wastewater by electrodeposition: Adsorption pretreatment. Desalination, 562 (2023), 116683, https://doi.org/10.1016/j.desal.2023.116683
88. R. Hancke, T. Holm, O. Ulleberg, The case for high-pressure PEM water electrolysis, Energy Convers. Manag. 261 (2022), 115642, https://doi.org/10.1016/j.enconman.2022.115642.
89. C.-W. Sun, S.-S. Hsiau, Effect of electrolyte concentration difference on hydrogen production during PEM Electrolysis, J. Electrochem. Sci. Technol. 9(2018), 99-108, https://doi.org/10.5229/JECST.2018.9.2.99
90. S. S. Lim, J.-M. Fontmorin, P. Izadi, W. R. W. Daud, K. Scott, E. H. Yu, Impact of applied cell voltage on the performance of a microbial electrolysis cell fully catalysed by microorganisms, Int. J. Hydrog. Energy, 45 (2020), 2557-2568, https://doi.org/10.1016/j.ijhydene.2019.11.142.
91. I. N. Aquigeh, M. Z. Ayissi, D. Bitondo, Multiphysical models for hydrogen production using NaOH and stainless-steel electrodes in alkaline electrolysis cell, J. Combust. (2021), 6673494, https://doi.org/10.1155/2021/6673494
92. C. E. Rustana, Sunaryo, S. J. Muchtar, I. Sugihartono, W. Sasmitaningsihhiadayah, A. D. R. Madjid, F. S. Hananto, The effect of voltage and electrode types on hydrogen production from the seawater electrolysis process, J. Phys.: Conf. Ser. (2019), 012096, https://doi.org/10.1088/1742-6596/2019/1/012096.
93. I. Vincent, A. Kruger, D. Bessarabov, Development of efficient membrane electrode assembly for low-cost hydrogen production by anion exchange membrane electrolysis, Int.

J. Hydrog. Energy, 42 (2017), 10752-10761, https://doi.org/10.1016/j.ijhydene.2017.03.069.

94. F. M. Sapountzi, M. Jose, C. J. Gracia, K.-J. Weststrate, O.A. Hans, J.W. Fredriksson, H. Niemantsverdriet, Electrocatalysts for the generation of hydrogen, oxygen and synthesis gas, Prog. Energy Combust. Sci. 58 (2017), 1-35, https://doi.org/10.1016/j.pecs.2016.09.001.
95. C. Lamy, T. Jaubert, S. Baranton, C. Coutanceau, Clean hydrogen generation through the electrocatalytic oxidation of ethanol in a proton exchange membrane electrolysis cell (PEMEC): effect of the nature and structure of the catalytic anode, J. Power Sources, 245 (2014), 927-936, https://doi.org/10.1016/j.jpowsour.2013.07.028.
96. J.-H. Leu, A. Su, J.-K. Sun, Z.-M. Huang, The catalyst loading effects on the feed rate of $NaBH_4$ solution for the hydrogen production rate and conversion efficiency, Catalysts, 10 (2020), 451, https://doi.org/10.3390/catal10040451
97. F. Zaccaria, G. Menendez Rodriguez, L. Rocchigiani, A Macchioni, Molecular catalysis in "green" hydrogen production, Front. Catal. 2 (2022), 892183, https://doi.org/10.3389/fctls.2022.892183.
98. S. Wang, A. Lu, C.-J. Zhong, Hydrogen production from water electrolysis: role of catalysts, Nano Converg. 8 (2021), 4, https://doi.org/10.1186/s40580-021-00254-x
99. Z.M. Ghazi, S.W.F. Rizvi, W.M. Shahid, A.M. Abdulhameed, H. Saleem, S.J. Zaidi, An overview of water desalination systems integrated with renewable energy sources, Desalination, 542 (2022), 116063, https://doi.org/10.1016/j.desal.2022.116063
100. N.P.B. Tan, P.M.L. Ucab, G.C. Dadol, L.M. Jabile, I.N. Talili, M.T.I. Cabaraban, A review of desalination technologies and its impact in the Philippines, Desalination, 534 (2022), 115805, https://doi.org/10.1016/j.desal.2022.115805
101. A. Kanca, H. Şenol, D. Adıgüzel, C. Medin, O.N. Ata, Boric acid recovery in dilute during the desalination process in BMED system, Desalination, 538 (2022), 115920, https://doi.org/10.1016/j.desal.2022.115920
102. P. Sivasubramanian, M. Kumar, V.S. Kirankumar, M.S. Samuel, C.D. Dong, J.H. Chang, Capacitive deionization and electrosorption techniques with different electrodes for wastewater treatment applications, Desalination, (2023), 116652, https://doi.org/10.1016/j.desal.2023.116652
103. A. Al Hinai, T. Jafary, H. Alhimali, S. Rahman, A. Al-Mamun, Desalination and acid-base recovery in a novel design of microbial desalination and chemical recovery cell, Desalination, 525 (2022), 115488, https://doi.org/10.1016/j.desal.2021.115488

104. F. Esmaeilion, M. Soltani, J. Nathwani, A. Al-Haq, Design, analysis, and optimization of a novel poly-generation system powered by solar and wind energy, Desalination, 543 (2022), 116119, https://doi.org/10.1016/j.desal.2022.116119

105. L. Hernández-Pérez, J. Carrillo-Abad, V. Pérez-Herranz, M.T. Montañés, M.C. Martí-Calatayud, Effluents from the copper electrorefining as a secondary source of antimony: Role of mass transfer on the recovery by electrodeposition, Desalination, 549 (2023), 116322, https://doi.org/10.1016/j.desal.2022.116322

106. J.H. Lee, E.T. Yun, S.Y. Ham, H.S. Kim, P.F. Sun, H.D Park, Electrically conductive carbon nanotube/graphene composite membrane for self-cleaning of biofouling via bubble generation, Desalination, 535 (2022), 115841, https://doi.org/10.1016/j.desal.2022.115841

107. T. Chen, L. Xu, S. Wei, X. Tang, H. Chen, Enhanced ammonia-rich solution production and electrode separation using magnetic nickel-loaded carbon black in flow-electrode electrochemical deionization (FEED), Desalination, 544 (2022), 116152, https://doi.org/10.1016/j.desal.2022.116152

108. M. Sedighi, M.M.B. Usefi, A.F. Ismail, M. Ghasemi, Environmental sustainability and ions removal through electrodialysis desalination: Operating conditions and process parameters, Desalination, 549 (2023), 116319, https://doi.org/10.1016/j.desal.2022.116319

109. R. Cao, S. Shi, F. Duan, Y. Xu, Y. Li, H. Cao, Y. Wang, In-situ construction of modified layer on the surface of anion exchange membrane to improve antifouling performance, Desalination, 562 (2023), 116684, https://doi.org/10.1016/j.desal.2023.116684