Journal of Materials Chemistry A

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: J. Mater. Chem. A, 2024, 12, 5748

Analysis of photocatalytic CO_2 reduction over MOFs using machine learning†

Simay Özsoysal, Burcu Oral and Ramazan Yıldırım *

Photocatalytic CO_2 reduction over metal–organic frameworks (MOFs) is investigated by constructing a database from published articles and analyzed using machine learning tools to predict the total gas product yield (random forest regression) and predominant product types under various conditions (decision tree classification). Hyperparameters of the random forest model, *ntree* (120) and *mtry* (14) are optimized by 5-fold cross validation leading to R^2 values of 0.96, 0.94 and 0.60 for *training*, *validation* and *testing*, respectively indicating the predictive power of the model developed. Reactor volume, sacrificial agent and amount of catalyst per reaction volume were the most important variables for total gas production rate prediction. Decision tree models, developed for gas phase and liquid phase systems separately, to determine the predominant product types (CO or CH_4 in the gas phase, and one of CH_3OH, CO, H_2 and HCOOH in the liquid phase) depending on the photocatalyst properties and reaction conditions, were also successful with an overall testing accuracy of 87% and 77% for gas-phase and liquid-phase processes, respectively.

Received 13th November 2023
Accepted 31st January 2024

DOI: 10.1039/d3ta07001h

rsc.li/materials-a

Introduction

Global warming and environmental pollution are the major challenges of our times; in order to address these problems effectively, new, clean, renewable, and economically feasible technologies for energy generation have to be innovated. Meanwhile, effective capture or conversion technologies for CO_2 have to be also developed as CO_2 emission will likely continue to be the major cause of air pollution and global warming in the near future. The conversion of CO_2 to valuable products is a potentially attractive route to eliminate CO_2 if it can be achieved in a commercially feasible way. Carbon monoxide as syngas, methane as a chemical feedstock, methanol as a commodity product, and many other hydrocarbons can be synthesized from CO_2 through (photo) catalytic/electrocatalytic processes. Catalytic processes use noble metal catalysts, and operate at high temperatures, making them costly,[1] while the photocatalytic CO_2 conversion route offers a significant cost reduction since the energy for the reaction is obtained from the sun. However, the photocatalytic/photo-electrocatalytic CO_2 conversion efficiencies and selectivity towards the desired products are still too low for commercial feasibility.[2] There have been various attempts to overcome these challenges and commercialize photocatalytic CO_2 conversion; new materials are being searched for widely and experimental conditions are being optimized for desired results. For instance, sub-nanomaterials, double/single-atom catalysts, and 2D nanomaterials are used as novel catalysts due to their tunable characteristics and quantum effects.[3]

Metal–organic frameworks (MOFs) are porous materials that are created using metal centers and organic linkers; they have favorable properties such as a homogenous metallic distribution and the excitonic effect.[4,5] Additionally, their structure may prevent the recombination of electrons and holes, thus making them suitable photocatalysts.[6] Indeed, MOFs are investigated as promising materials for both CO_2 capture and conversion; they can be used as MOF-semiconductor composites, as photosensitizers or as conductive photo-electrocatalysts.[7] Organic linkers or metal centers are both capable of light harvesting, and according to oxidation states the photon transport can be either from organic linkers to the metal center (LMCT) or the metal center to organic linkers (MLCT); hence the separation of photogenerated charge carriers is facilitated and efficiency increases.[8] Since many products can be obtained from CO_2 conversion, it is important to have a selective photocatalyst for maximum yield. Optimizing synthesis routes, engineering the band gap, coupling with semiconductors and sensitization using a suitable sensitizer are some of the methods that are employed to increase product selectivity of MOF-based photocatalysts.[9] Stability is another challenge for MOF-based applications as various researchers are searching for a way to develop stable MOFs, which are also suitable for the function to be performed. For example, Almazan *et al.* synthesized an effective UiO-66 MOF-derived Ru@ZrO_2 catalyst for catalytic photothermal CO_2 reduction; they obtained highly dispersed Ru

Department of Chemical Engineering, Boğaziçi University, 34342 Bebek Istanbul, Turkiye. E-mail: yildirra@boun.edu.tr

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d3ta07001h

particles without stability loss and achieved high yield and selectivity for CH_4.[4] In another study, Ding *et al.* created a BiOCl/Bi-MOF catalyst *via* co-shared Bi–Cl bonding which showed high CO_2 capture and high CO production rate due to increased oxygen vacancies and decreased electron–hole recombination.[10]

Machine learning (ML) is used to analyse the patterns in large datasets and to develop models for descriptor (input variables) and performance relations; this way one can identify the most suitable materials and conditions, make predictions for the new cases or deduce heuristic rules and decision for future studies. The recent improvements in data availability through scientific publications and online databases as well as the developments in computational power, including ML algorithms, allowed ML to be widely employed in materials research, reaction modelling, and yield prediction for many photocatalytic applications. For instance, our group has previously studied the performance of various photocatalytic systems such as water splitting[11,12] and CO_2 reduction[13] as well as other catalytic systems related to CO_2 capture and conversion[14] using machine learning tools to develop models and heuristics from the studies published in the literature to enhance the photocatalytic performance in future studies. There are also studies in the literature that aim to design or discover new photocatalytic materials, as reviewed by Ge *et al.* and Mai *et al.*[15,16]

Most of the MOF-based ML studies focus on gas uptake on MOFs[17] and screening of materials for CO_2 capture[18–21] and separation.[22] There are also some studies focusing on synthesis conditions to predict stability as well as inverse material design using ML.[23] Additionally, Lin *et al.* reviewed new MOF databases developed from the experimental and DFT-based data.[24]

In this work, a literature-based dataset is constructed for gas and liquid phase CO_2 photo-reduction on MOF photocatalysts and analysed using machine learning tools. The aim was to use the experience in the literature to identify the major patterns and significant factors related to MOF properties and reaction conditions for high photocatalytic performance and develop models and heuristics to show the direction and variables to be manipulated for future studies. Predictive random forest models correlating the total gas production rate with the MOF properties and process variables were developed for the gas and liquid phase processes separately. Similarly, decision tree classification models were developed to show the effects of MOF properties and process conditions on the predominant product type as the indicator of product distribution. To best of our knowledge, there is no such work on the ML analysis of photocatalytic CO_2 reduction on MOFs.

Computational methodology

Dataset construction and pre-processing

The dataset containing 605 data points is constructed by extracting instances from 197 individual experiments reported in 57 published articles; 186 of the data points are recorded in gas-phase processes while the remaining 419 instances are from liquid-phase processes. We collected data starting from the research papers reviewed by Luo *et al.* (2023),[8] and formed the initial draft of our dataset by extracting data from these papers. Then, we added the data from the research papers such as the review articles by Dhakshinamoorthy *et al.* (2021)[25] and Zhoa *et al.* (2021)[26] if they were not already covered by Luo *et al.* (2019). Finally, we performed a Web of Science search with the keywords *CO_2 photoreduction or photocatalytic CO_2 reduction and metal organic framework* or *MOF* in the article title for the results published in 2022–2023 (on January 15, 2024).

Influential parameters related to MOFs (photocatalyst properties) and reaction conditions are selected as descriptors while the output (performance) variable is chosen as the total gas production rate (used in the random forest-regression model), indicating the sum of production rates of the resulting products and major gas type (used in the decision tree-classification model), representing the predominantly generated product as a consequence of CO_2 photoreduction. The final dataset is recorded manually on an Excel worksheet and organized in a tabular format. It comprises product types and production rates against a set of features (descriptors) in six fundamental categories (process phase, MOF-characteristics, feed composition, light source, reactor parameters and MOF synthesis conditions). Descriptors and the corresponding range of variables are listed in Table 1 while the complete dataset is provided in the ESI.

In the MOF category of *functionalized and/or metal-added MOFs* (FUNC), ligands undergo modifications *via* organic functionalization (*i.e.*, amine functionalization) or inorganic material incorporation (*i.e.*, metal addition); on the other hand, MOFs undergo metal-ion exchange rather than direct addition of metals in the category of *metal-ion exchanged and/or doped MOFs* (EXCH), which also includes metal doping.[27]

We think that we should also clarify our approach for selecting some of the descriptors. MOFs are usually used together with some semiconductors and other materials that may also act as co-catalysts. There were many different types of MOFs combined with different materials, and it would be inefficient to use photocatalysts and co-catalysts as descriptors as usually done in similar studies; instead, the pre-analysis of data indicated that it would be more practical to use different parts of the photocatalyst as different descriptors. Hence, we identified five descriptors to distinguish the photocatalysts from each other: MOF type, center metal, NH_2 presence (related to functionalized type MOFs), type of additional material and presence of a noble metal in the structure. The type of MOF, the center material and the presence of NH_2 are directly related to the characteristics of MOFs since they can be identified using the center metal (for instance Zr) and type (for instance PURE). The presence of NH_2 is related to being functionalized or not; most of the functionalized MOF types include NH_2 groups. However, there are some cases where the MOF is identified as functionalized without having an NH_2 group (they have an additional metal or –OH group). Additionally, there were also some data sets that have both NH_2 and metal for functionalization; so the presence of NH_2 was described separately to avoid confusion. The additional material and noble metal descriptors involve the type of addition performed to the MOF. Noble metals such as Pt and Au were used as co-catalysts, but the number of noble metal containing photocatalysts is too small,

Table 1 Variable list of the dataset

Variable category	Variable name	Range
Process phase	Phase	Liquid, gas
MOF characteristics	MOF type	PURE[a], COMP[b], EXCH[c], and FUNC[d]
	Center metal	Zn, Zr, Co, other, none
	Noble metal	None, present
	NH_2 functional group	None, present
	Additional functional group	Adenine, CdS, CN, CNNs, $CsPbBr_3$, Cu_2O, metal, nonmetal, OH, Ti, Ti_3C_2, TiO_2, Zn_2GeO_4, *etc.*
	Band gap (eV)	1.52–5.00
Feed composition	Catalyst per volume (mg mL^{-1})	0.037–375
	Sacrificial agent (TEOA) (%)	0–100
Light source	Electromagnetic spectrum	UV[e], UV-vis[f], and VIS[g]
Reactor parameters	Time (h)	0.5–24
	Volume (mL)	1–2700
	Pressure (atm)	0.1–1.5
	Temperature (°C)	0–230
Synthesis conditions	Time (h)	0.5–240
	Temperature (°C)	40–250
Gas production	Produced gas type	H_2, CO, CH_4, HCOOH, and CH_3OH
	Total gas production rate (μmol g_{cat}^{-1} h^{-1})	0.031–2463.2

[a] PURE: pure (unmodified) MOFs. [b] COMP: MOF composites. [c] EXCH: metal-ion exchanged and/or doped MOFs. [d] FUNC: functionalized and/or metal-added MOFs. [e] UV: 280–380 nm. [f] UV-vis: 380–420 nm. [g] VIS: 420–700 nm.

while the types and concentrations of noble metals used are rather diverse; for this reason, the number of data points was not sufficient to describe the noble metals with their names and concentrations. Hence, we decided to represent them with a separate descriptor as a *noble metal* having two values as *present* or *absent*; even this simple representation improved our models (as also evident from Fig. 5). The additional material descriptor (labeled as an additional functional group) consists of every other type of addition performed to the MOF structure (they may also act as a co-catalyst). For instance, the type of semiconductor added to the MOFs in a composite MOF or any functional group other than NH_2 was included under this descriptor.

Before the ML analysis, the log (10 based) transformation of the output variable (total gas production rate) is performed to stabilize the variance and improve the normality of data; for this, each value of the output variable is incremented by 1 before taking the logarithm as a safeguard against zero values of output. With this transformation, the total gas production rates are constrained between 0 and 3.5 while the distribution of data approached normal distribution as shown in Fig. 1b (unless stated otherwise, all subsequent numeric output values correspond to data after undergoing logarithmic transformation). Following the transformation process, the dataset is randomly divided into training and testing sets, with approximate proportions of 0.7 and 0.3, respectively. The splitting is done based on experiments (*i.e.*, all data taken in one experiment are placed in either the training or test data set only); otherwise, it may cause data leakage (changing only some variables during the experiment may not change the results, or the results may change in a predictable manner and may create false impression of high generalization ability). Subsequently, the presence of missing values in both datasets is examined and filled with the average values of the related input variables in the corresponding dataset (training or testing). An important aspect related to handling missing values is that, for the reaction pressure of liquid-phase processes, an exception is made: the missing values are not imputed using the mean value; instead,

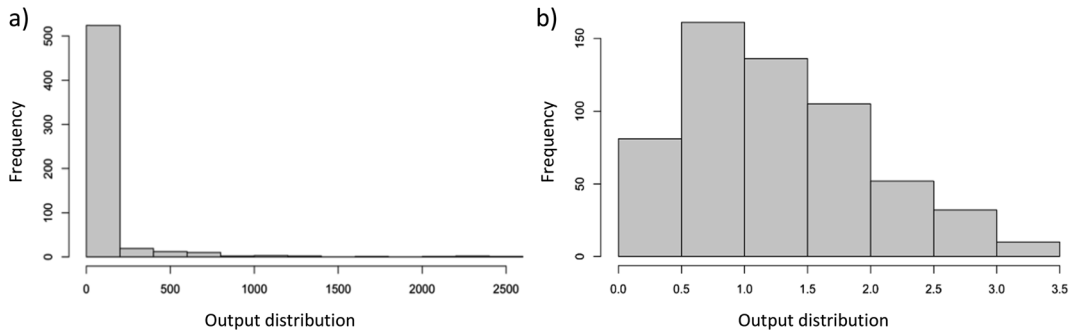

Fig. 1 Output displays of the total gas production rate (a) before and (b) after logarithmic transformation.

they are filled with a value of 1 atm. This organization is significant since it aligns with a common practice detected in most liquid-phase studies within the literature, where experiments are usually conducted under atmospheric conditions. At the end, the pre-processed (including log transformation) dataset is found to be suitable to machine learning (ML) applications. In essence, pre-processing boosted the predictive power of our new dataset in comparison to the preliminary models constructed with the raw dataset.

Implementing machine learning tools

Random forest regression model for total gas production rate. The random forest regression model is constructed by using the *randomForest*[28] package in *RStudio*, and optimized by employing *k*-fold cross validation (*k* was 5 here). Other regression models, including multiple linear regression, artificial neural network, gradient boosting and support vector machine regression models are also tested, yet none of them could reach the performance achieved by the random forest model. Therefore, no further details for those models are provided in the manuscript. The 5-fold cross-validation technique is implemented *via* dividing the training set into five subsets. In each iterative step, the data are trained with four subsets, and the remaining subset is utilized for validation. Two hyperparameters, namely *ntree* (the number of decision trees utilized for the growth of random forest) and *mtry* (the number of variables considered for splitting each node) are tuned during these iterations in order to improve performance of the random forest model.[29] To obtain the best combination of these parameters, nested loops are formed in *RStudio*, and the sequence of *ntree* and *mtry* parameters is scanned. In the iterative procedure, *ntree* values are alternated from 50 to 200, with increments of ten at each step. Similarly, the *mtry* sequence is changed within the range of five to 15 by increments of three at each step. Optimum model hyperparameters are established as 120 for *ntree* and 14 for *mtry*. Root mean square error (RMSE) values for training, validation and testing dataset models are used as the measure of the fitness of the model.

Decision tree models for classification of major gas products in liquid and gaseous mediums. By employing the *rpart* package[30] in *RStudio*, decision tree models are created, with the primary objective of classifying the predominant gas type. The term *predominant* implies the type of gas displaying the highest production rate (among H_2, CO, CH_4, HCOOH, and CH_3OH). Given that there are differentiations between gas phase and liquid phase photocatalytic CO_2 reductions, two separate models were developed. The optimum hyperparameters of both models are determined as 20 for *minsplit*, 7 for *minbucket* and 0.01 for *cp*. The seed value is arbitrarily chosen as 461 for the gas phase model and 161 for the model concerning the liquid medium.

Results and discussion

Pre-analysis of MOF types in the dataset

As shown in Table 1, MOF types are categorized into four groups namely, PURE, FUNC, EXCH and COMP; representing pure (unmodified) MOFs, functionalized and/or metal-added MOFs, metal-ion exchanged and/or doped MOFs, and MOF composites,[27] respectively. The distribution of data points, after preprocessing steps, is visualized in Fig. 2; the bubble plot in Fig. 2a shows the average gas production rate (as well as the data size indicated by the size of spheres), and Fig. 2b presents the distribution of data as a swarm plot. These representations indicate that most of the data points belong to the COMP type even though it does not represent the highest average performance. In contrast, the number of instances in the class of EXCH is the lowest, yet it has the highest gas generation rate. This outcome is anticipated given that ion exchange or doping applications modify the redox potentials and improve photocatalytic efficiency.[27]

The average effect of the MOF type is presented in Fig. 3 for both liquid and gas phases, separately. It is observable from the

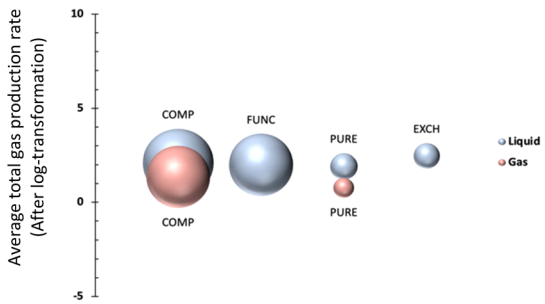

Fig. 3 Liquid and gas phase data distribution of MOF types.

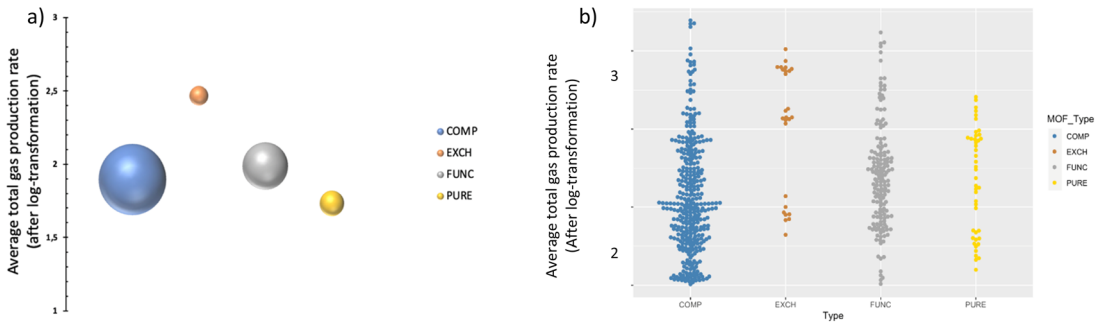

Fig. 2 MOF type distribution by using a (a) bubble graph and (b) swarm plot.

figure that average total gas production rates for the gas phase are generally lower than that of the liquid phase while higher CO_2 conversion was attained in the liquid phase. One plausible reason could be the effect of stirring, as turbulence is a critical parameter to overcome the resistance of boundary layers on the photocatalyst surface in the liquid phase.[31] Without stirring, the liquid layer on the top of the photocatalyst would act as a barrier to CO_2 molecules and hinder the efficiency of the photoreduction reaction by reducing mass transfer and therefore CO_2 conversion rates.[31]

Random forest regression of the total gas production rate

As mentioned in the *Computational methodology*, the random forest regression model, with an optimized *ntree* and *mtry* of 120 and 14 respectively, represented the dataset best; the actual *vs.* predicted total prediction plots for training, validation and *testing* sets are presented in Fig. 4. As can be seen, the model predictions closely align with the actual data points, showing that the fitting is satisfactory. In quantifiable terms, it is reported that the RMSE, which is the main indicator of model fitness, is 0.15 for training, 0.16 for validation, and 0.48 for testing while R^2, as the other measure of fitness, is 0.96 for training, 0.94 for validation and 0.60 for testing (recent data are slightly underpredicted probably due to the fact that relatively higher performances were obtained over the same materials with improved experimental conditions and experience gained). However, it is important to remember that these results are obtained after the logarithmic transformation of the data; hence, the model is successful within the order of magnitude predictions. Nevertheless, better predictions, at this stage, are not realistic with the current data structure anyway; the measurement errors associated with low product concentration and non-standard testing conditions, especially those related to irradiation (variations in frequency, density, irradiation angle and distance, transmittance of glassware used in the system, adsorption/transmittance characteristics of solution, *etc.*) and the possibility of products generated from carbon contamination in the reaction medium instead of CO_2 (ref. 32) create serious levels of uncertainty in the data.

We also examined the relative importance of descriptors for the predictive performance of our random forest model as shown in Fig. 5 displaying the percentage change in mean squared error with the change in the values of a specific

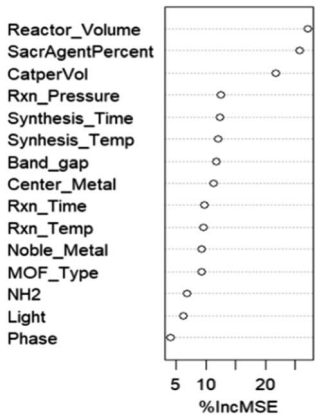

Fig. 5 Descriptors that represent the percentage increase in the mean squared error (MSE).

descriptor, and this provides insight about the influence of that descriptor on the model's predictive accuracy. It is apparent that *the reactor volume, the sacrificial agent and amount of catalyst used per reaction volume* stand out as the most influential parameters; the result involving the volume and amount of photocatalyst is highly foreseeable, given that the catalyst formulation exerts a notable influence on the photocatalytic activity.[33] The association can be further clarified by referring to the study conducted by Zhao *et al.*,[33] who examined the influence of catalyst and solvent addition on the product yield. Their findings revealed that the CO production rate displays a continual increase as the amount of catalyst (NH_2-IUO-66 and CdS in this particular study) is increased, until a maximum value is reached. Beyond the peak value, a decrease in yield is detected and the reduction of catalytic activity is attributed to the agglomeration through excess addition of particles. Accordingly, outcomes of the study prove the existence of an optimal quantity of catalyst addition, in which the highest gas production rate is achieved. The reactor volume, especially in the liquid phase, on the other hand, influences the mixing and transfer of the products from the surface of the particulate catalyst to the gas phase. The significance of the sacrificial agent indicates the significance of concentration, which is an expected result because only TEOA (triethanolamine) is used for this purpose. It is interesting to note that the *process phase* emerges as the least influential factor as shown in Fig. 5. Low

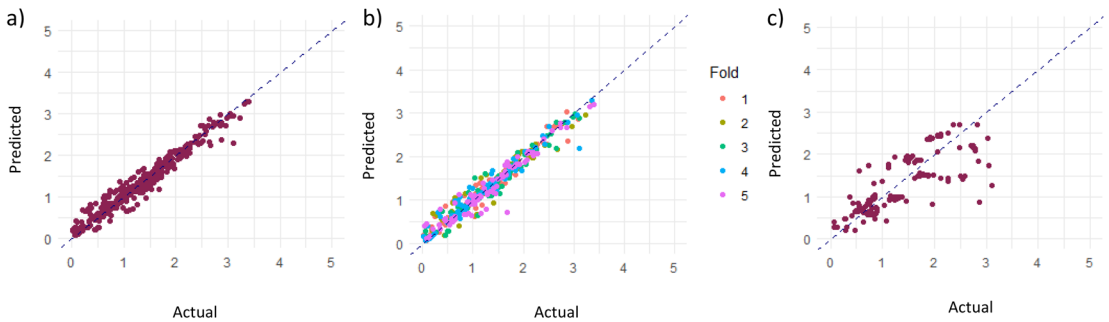

Fig. 4 Random forest regression model of (a) training, (b) validation, and (c) testing datasets.

influence of the process phase (liquid or gas) on the outcome could be attributed to the relative importance of other influential factors. In essence, other factors such as catalyst loading, feed composition, MOF properties, and reactor parameters (temperature, pressure, *etc.*) could be directly or indirectly related to the phase and these phase-specific factors could overshadow the influence of the process phase. As a result, the predictive models lack the ability to assign more importance to them.

Decision tree classification of the predominant gas type

Decision tree classification models are also constructed for prominently generated gas types in photocatalytic CO_2 reduction to provide relationships between descriptors and gas output categories, and to identify possible routes (or heuristic rules) to achieve the desired product type. Naturally, there are usually multiple gaseous products with different ratios; however, we only considered the product having the highest production rate in each case. This approach provides significant advantages for model building: (1) an efficient and simplified classification can be achieved by avoiding the categorization of every potential gas type, especially those present in trace amounts, (2) the interpretability of the model would be much higher providing more insight for future studies. Subsequently, two individual models, denoting gas and liquid phase photocatalytic CO_2 reductions separately, are provided below.

Photocatalytic CO_2 reduction in the gas phase. A decision tree classification model for gas phase photocatalytic CO_2 reduction is presented in Fig. 6; the predominant products in 184 cases were used to develop the model. Normally, our data set contains 186 cases, but we excluded two cases in which H_2 is the major product, because only two cases are not sufficient to make a generalization (compared to 130 and 54 cases having CO and CH_4 as the major products, respectively). Consequently, the model presented in Fig. 6 does not contain the route for hydrogen production; it provides potential routes for CO and CH_4 production (nodes with shades of green signify CO and blue tones denote CH_4 as the dominant gas type). It should be remembered that hydrogen, or other gases, may be produced significantly in these cases as well (especially with CO) but it does not have the highest concentration.

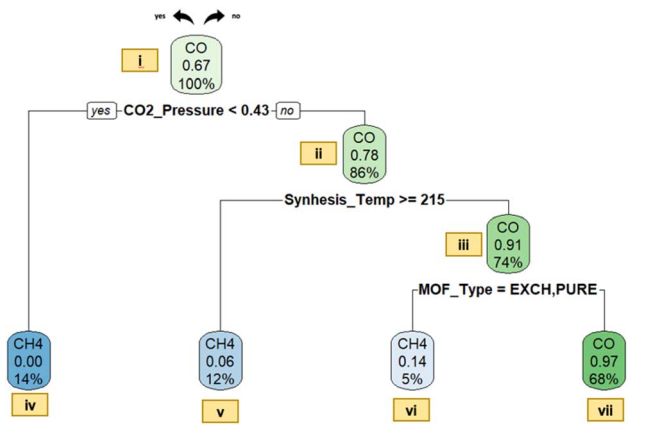

Fig. 6 Decision tree model for the predominant gas type in gas-phase photocatalytic CO_2 reduction.

Shortly, the tree starts with root node (at the top) and splits the data *via* applying some criteria (values of descriptors that cause a change in the major gas type); when the criterion is fulfilled under each splitting condition (in nodes), the decision tree proceeds along the left branch; in the opposite case, it continues along the right branch. The gas type (either CO or CH_4) specified in the first row at each node represents the prevailing class while the percentage in the middle row expresses the predicted probability of CO production. The percentage specified in the bottom row signifies the proportion of the total data stored in that particular node. The proportion at the first node (root node) is specified to be 100%, as it contains the entire dataset; where CO dominates with 71%, while CH_4 constitutes only 29%.

In order to elucidate the dominant occurrence of CO, an initial investigation into the reaction mechanism of CO_2 photoreduction is conducted. In the context of CO_2 reduction with H_2O, a simultaneous reaction of water splitting occurs that results in H_2 generation.[34] However, since the primary focus of this article does not include the details of water splitting, the particular mechanism of H_2O reduction to H_2 is intentionally omitted. The potential reactions of the CO_2 photoreduction process with H_2O vapor to generate CO and CH_4 in the context of thermodynamic reduction potentials relative to the normal hydrogen electrode (NHE) are given below in eqn (1)–(4).[35,36]

$$CO_2 + 2H^+ + 2e^- \rightarrow CO + H_2O,\ E^0 = -0.48\ V \quad (1)$$

$$2H^+ + 2e^- \rightarrow H_2,\ E^0 = -0.41\ V \quad (2)$$

$$CO_2 + 8H^+ + 8e^- \rightarrow CH_4 + 2H_2O,\ E^0 = -0.24\ V \quad (3)$$

$$2H_2O + 4h^+ \rightarrow 4H^+ + O_2,\ E^0 = +0.82\ V \quad (4)$$

The lower reduction potential difference for the reduction of CO_2 to CH_4 ($E^0(CO_2/CH_4) = -0.24$ V) compared to that of CO_2 to CO ($E^0(CO_2/CO) = -0.48$ V) facilitates the conversion to methane. Nevertheless, it is worth emphasizing that eight electrons are required for CH_4 production, yet only two are required for CO. Accordingly, this leads to a substantial yield rate of CO in the resultant gaseous products.[35] For this reason, designating CO as the initial product of CO_2 photocatalysis is not an unexpected outcome, considering the fact that photocatalytic CO_2 reduction entails a complex mechanism and results in a range of products.[34] Although CH_3OH (methanol) is reported as a potential product, a literature review reveals that CO_2 is more likely to undergo reduction to CH_4 rather than CH_3OH in most cases.[34] Our findings align with the existing literature, since CH_3OH is not detected as a major gas type in our data. Consequently, the frequent presence of CO (dominant share of 71%) in the model can be rationalized based on the aforementioned factors.

Upon examining Fig. 6, it can be seen that the first splitting condition relies on the CO_2 pressure level. A specific threshold of 0.43 atm pressure was established within this splitting

condition (node i). For pressure levels below this threshold, the branch leads to a node dominated by CH_4, where the predicted probability of CO production is zero. The choice of the splitting condition (division according to the CO_2 pressure level) is well-founded, since the products formed during the photoreduction process are a direct result of the CO_2 available for the reaction.[35] This clarifies the critical role played by CO_2 concentration as a determining factor in the photoreduction process. Accordingly, increasing (or decreasing) the concentration of CO_2 within an enclosed system correlates with a subsequent rise (or fall) in the pressure of CO_2. This adjustment can be achieved by changing the CO_2/H_2O ratio at a given reactor volume and reaction temperature. In other words, CO_2 pressure, CO_2 concentration and CO_2/H_2O ratio are interconnected factors; hence, modifying one could correspond to changes in another, ultimately affecting the resulting predominant gas type. Following is a detailed exposition on the influence of the CO_2-to-H_2O ratio:

The CO_2/H_2O ratio stands out as a significant parameter influencing both the reaction rate and product yield.[37] Tahir *et al.*[35] conducted a study to observe the changes in the H_2O/CO_2 ratio through modifying the flow rates of CO_2. Their studies indicate that increasing the H_2O/CO_2 ratio (or decreasing the CO_2/H_2O ratio) resulted in an increase in CH_4 yield. The rationale behind this enhancement is elucidated as follows: an elevated H_2O/CO_2 ratio (signifying a decrease in CO_2 concentration) promoted the adsorption of H_2O molecules on the catalyst surface and enhanced the ability to reduce CO_2 in order to generate CH_4. The alignment between Fig. 6 and findings of Tahir *et al.* becomes evident when taking into account that a decrease in CO_2 concentration (and a reduction in CO_2 pressure accordingly) may result in boosted CH_4 yield in both cases. This pattern highlights the reliability of our results and consistency of our model with prior research.

On the other hand, it can be seen that a significant portion of the data assembles at a node where CO is identified as the predominant gas (node iii). The splitting criterion of *synthesis temperature* may not be used for practical purposes; the synthesis temperature and time are usually coupled with the MOF structure (*i.e.* they are actually referring to certain MOF structures), which is the true reason for the branching but labelled with temperature or time instead. For examples, the MOF types, such as CO-ZIF-9/TiO_2, TiO_2/NH_2-UiO-66, $CsPbBr_3$ quantum dots/UiO-66(NH_2), and CPO-27-Mg/TiO_2, are accumulated on the right hand side of the branch (low synthesis temperature and usually high synthesis time) promoting CO production. The following assertions can be made concerning these photocatalysts: Co-ZIF-9 serves as an effective co-catalyst that could endorse CO_2 photoreduction over TiO_2, resulting in the production of CO, CH_4 and H_2 gases.[38] Notably, upon analyzing the CO_2 photoreduction activity data, it is apparent that CO emerges as the most predominantly produced gas among these. Besides, in the gas phase experiment involving the use of TiO_2/NH_2-UiO-66 as a photocatalyst at ambient temperature and H_2, the only detected product is CO.[39] Similarly, an extensive amount of CO was generated when $CsPbBr_3$ quantum dots/UiO-66(NH_2) was used as the photocatalyst; this notable photocatalytic performance was attributed to fast charge separation and transfer occurring at the interface between $CsPbBr_3$ quantum dots and UiO-66(NH_2) nanocomposites.[40] Lastly, in the context of using CPO-27-Mg/TiO_2 as a photocatalyst, the product mixture contained H_2, O_2, CO, and CH_4, yet CO stands out as the major product. All these photocatalysts have a common characteristic that they accumulate under a branch related to a synthesis time larger than 7.5 hours. Indeed, this criterion is highly associated with photocatalyst properties. In other words, categorizing MOFs by using their synthesis duration fundamentally equates to their classification based on photocatalyst properties. Regarding this matter, when the aforementioned photocatalysts are employed under appropriately configured experimental conditions, the likelihood of CO production can be notably boosted. Node iii is then further purified with MOF types resulting in node vi, indicating the conditions for CO production as the predominant gas product.

It is generally known that the position of band edges, magnitude of band gaps and mobility of charge carriers serve as fundamental factors in photocatalytic performance.[41] Therefore, it is imperative to conduct a comprehensive examination of photocatalyst characteristics in this context as well. Indeed, lower bandgaps tend to lead to CO while the higher values are usually linked to CH_4 as the primary product. For example, $FAPbBr_3/Ti_3C_2$ (with a band gap of about 2.0 eV) exhibits a notable CO_2 reduction performance, leading to a significant generation of CO, accompanied by minimal production of H_2 and CH_4.[42] In contrast, in certain cases where photocatalysts named $CsPbBr_3$ QD ($E_g = 2.4$ eV, absorption band edge of 530 nm) and ZIF-8 ($E_g = 3.3$ eV, absorption band edge of 380 nm) are employed, CH_4 could become the predominant gas as a result of CO_2 photoreduction.[43] However, the bandgap does not appear as a decision criterion in the decision tree in Fig. 6, even though it does in some of our (less optimum) models; similar to the discussion presented above for the coupling of synthesis time, temperature and the type of photocatalyst synthesized under specific conditions, the bandgap is also coupled with most of the structural descriptors and does not always appear directly in the tree even though it is highly influential (such cases are more frequent in small datasets in which the same data point can be represented by different descriptors).

The confusion matrix containing the statistical measures for the fitness of the decision tree is shown in Table 2. Overall

Table 2 Confusion matrix for the predominant gas type in gas-phase processes

	Accuracy	# Of data points	Actual class	Predicted class: CH_4	Predicted class: CO	Recall
Training	0.96	41	CH_4	39	2	0.95
		88	CO	3	85	0.97
			Precision	**0.93**	**0.98**	
Testing	0.87	13	CH_4	9	4	0.69
		42	CO	3	39	0.92
			Precision	**0.75**	**0.90**	

accuracies of 96% for training and 87% for testing sets indicate that the model's performance is quite satisfactory. It is reasonable to conclude that correlations and patterns are recognized successfully, and hence, a good classification model is obtained. The precision and recall values of classes CH_4 and CO are also reported in the table to assess the effectiveness of the model for each category; again, the values are satisfactory for both of these classes. Here, the precision measures how accurate the model's positive predictions are (*i.e.*, what percent of the model's positive identifications is really positive), whilst the recall signifies how proficient the model is when capturing all the actual positive cases (*i.e.*, what is the percent of positive that the model identified as positive). The high precision is especially quite critical because we usually need the paths suggested to be really correct (precision) rather than the identification of all correct paths (recall). Considering the overall confusion matrix, the binary model seems to yield notable success.

Photocatalytic CO_2 reduction in the liquid phase. A decision tree model is also constructed to categorize the predominant gas type in the liquid phase as illustrated in Fig. 7. In contrast to the gas phase model, this displays a higher degree of gas diversity, including methanol (CH_3OH), carbon monoxide (CO), hydrogen (H_2) and formic acid (HCOOH), with a total of 413 data points. The liquid dataset originally covers a total of 419 cases; the predominant gas is CH_4 in seven cases, CH_3OH, in 34, CO in 177, H_2 in 90 and HCOOH in 111 cases. Given the infrequent occurrence of CH_4, the corresponding six recordings were eliminated, and the analysis focused on the remaining data points to be able to make statistically reliable generalizations. In agreement with Fig. 6, the gas type specified in the initial row denotes the product that was produced at the highest rate under the conditions described by the branch, and the percentage at the bottom line signifies the distribution of total data contained within that node as in the case of the previous model for gas phase processes; the number in the middle row represents, from left to right, the fraction of CH_3OH, CO, H_2 and HCOOH in that node, respectively (format is slightly different than that in the gas phase because this time there are more than 2 products). A point worth mentioning is that, CO is the most commonly observed predominant gas in both liquid and gas phases, which may be attributed to its position as the first product of photocatalysis.[34] Although this is the primary outcome of the CO_2 reduction process, the general mechanism afterwards is undeniably complicated, especially for liquid phase processes.[34] It has been argued that there is no exact conversion route due to simultaneous production of varied products.[44] Yin *et al.*[45] supported this notion by describing the route of CO_2 reduction as *unclear*.

It has been inferred that conversion routes highly depend on reaction conditions,[44] and amendment of photocatalytic properties (noble metal installation, doping with non-metals, photocatalyst coupling processes, photosensitization, *etc.*) could have an impact on the variety of products.[34] Correspondingly, a detailed explanation of our model concerning photocatalyst properties and reaction conditions is provided as follows for liquid-phase processes:

The data points linked with nodes xvi and xvii are separated by the specific MOF type used, indicating that when ion exchanged or doped MOFs are selected for CO_2 reduction, the generation of H_2 could potentially exceed that of HCOOH. The notion emphasizes the importance of a case study of Sun *et al.*,[46] who investigated the effect of noble metal doping on the catalytic performance over M/NH_2-MIL-125(Ti) (M = Au and Pt) under irradiation of visible light in the presence of TEOA as a sacrificial agent. The research findings demonstrate that the

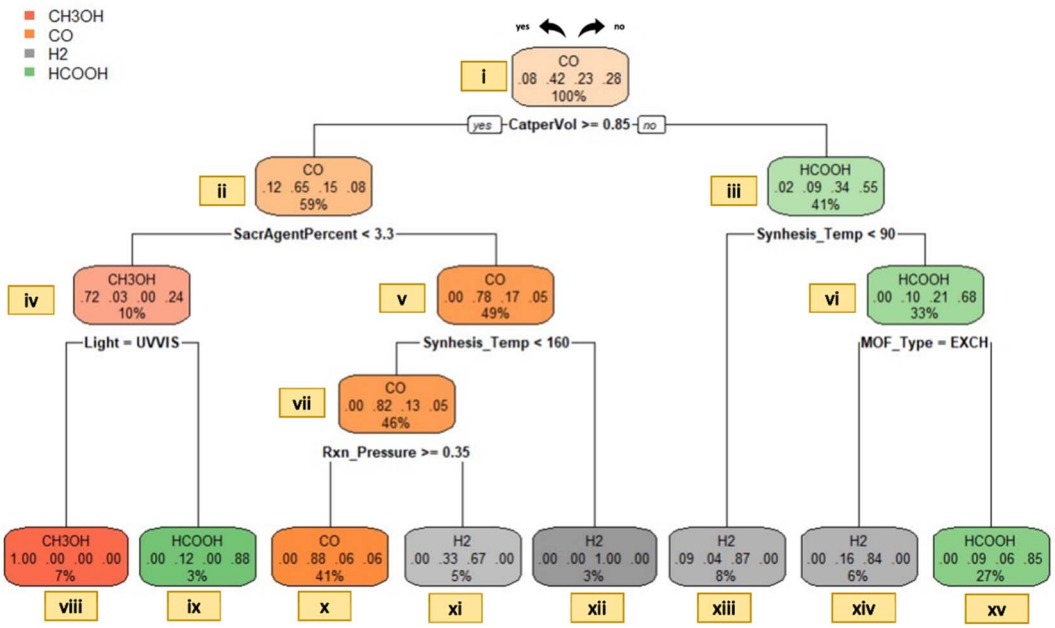

Fig. 7 Decision tree model for the predominant gas type in liquid-phase photocatalytic CO_2 reduction.

use of pure NH_2-MIL-125(Ti) results in only HCOOH generation. On the other hand, both HCOOH and H_2 production was detected when doped MOFs (Au/NH_2-MIL-125(Ti) and Pt/NH_2-MIL-125(Ti)) are used. Moreover, it is established that hydrogen production becomes so substantial that it emerged as the main product in CO_2 reduction.[46] This underscores the impact of noble metal doping on changing catalytic selectivity.[27] In connection with node xvii, it is visualized that the choice of MOF types such as PURE, FUNC or COMP can result in the generation of HCOOH as the prominent gas. According to our data, this category (node xvii) includes several photocatalyst types such as Zr-SDCA-NH_2, NH_2-MIL-101(Fe), NH_2-MIL-88(Fe), NH_2-MIL-53(Fe), NH_2-UiO-66(Zr), $(NH_2)/(NH_2)_2$-UiO-66(Zr), NH_2-MIL-125(Ti), PCN-222, AD-MOF-1, N-CNDs/NH_2-UiO-66, Ru-MOF, Ir-CP, *etc.* The existence of varied photocatalysts that favor HCOOH production can be attributed to the choice and usage amount of solvent (typically high amounts of acetonitrile) in the reaction mixture. In the context of CO_2 photoreduction in the liquid phase, it has been revealed that the polarity of the solvent plays a crucial role in stating the resulting outcome.[34] Concerning this issue, selection of a solvent with a low dielectric-constant is expected to form CO and H_2O. On the other hand, preference of a polar solvent (or high dielectric-constant solvent) favors HCOOH formation. Given that acetonitrile (MeCN) is an aprotic polar solvent with a notably high dielectric-constant, utilization in higher amounts relative to a sacrificial agent TEOA (*i.e.*, 50 : 1) is expected to favor the production of HCOOH. The rationale lies in the fact that, when a highly polar solvent is used, the anionic radical of CO_2 experiences more stabilization by solvent molecules, leading to a weaker interaction within the surface of the catalyst and resulting in the formation of HCOOH.[34] Therefore, regarding the splitting condition at node vi, it can be concluded that accumulation of data within node xvi can be credited to the use of noble metal doping, while aggregation in node xvii is connected with the choice and amount of polar solvent. In other words, the former classification is based on photocatalyst properties, whereas the latter is related to reaction conditions. (It is imperative to emphasize that the visualization does not preclude or prohibit the possibility of CO generation from the EXCH type or HCOOH production from the COMP type as dominant outcomes; it serves as a method for classification.)

As an added point, data points associated with nodes xi, xii and xiii are separated based on differences in *synthesis pressure and temperature*. Simply put, the node separation related to synthesis temperature stems from the utilization of different types of photocatalysts as we discussed in the gas phase. The division based on the pressure on the other hand, can be clarified with an experimental study of Wang *et al.*,[47] in which low-pressure conditions (0.1 atm CO_2) in their experiment resulted in the production of two different types of gases, H_2 and CO. In their study, five different photocatalysts, namely MAF-X27-OH, MAF-X27l-OH, MAF-X27-Cl, MAF-X27l-Cl, and MOF74-Co are utilized. MAF-X27-OH and MAF-X27l-OH exhibited significantly higher CO production, whereas MAF-X27-Cl, MAF-X27l-Cl, and MOF74-Co displayed predominant production of H_2 at 0.1 atm pressure. It can be justified that the use of hydrogen ligands enhanced the photocatalytic activity towards the production of CO. In other words, this study underscores the significance of using a specific ligand type and its influence on the product type in CO_2 photoreduction. An essential point concerning the study of Wang *et al.*[47] is that, when the experiments were conducted under 1 atm pressure, all photocatalysts predominantly generated CO. More precisely, an increase of pressure from 0.1 atm to 1 atm causes a shift for MAF-X27-Cl, MAF-X27l-Cl, and MOF74-Co, resulting in higher CO production compared to H_2 (data outputs of this study are stored in node xi). The significance of the increased production of CO is not an unforeseen finding: it has been reported that an elevation in CO_2 pressure in aqueous media leads to improved product selectivity.[48]

It is essential to address another aspect concerning production of CO in node xi and node xii. Data points accumulated in these nodes share a common characteristic: employing a limited quantity of MeCN solvent relative to the TEOA sacrificial agent (usually 3 : 1 or 4 : 1). On the other hand, data points gathered in node xvii exhibit a vast amount of MeCN utilization compared to TEOA (there are instances of ratios such as 30 : 1, 50 : 1, and 20 : 1). An interesting point to underline is that; not only the amount of MeCN, but also the ratio (MECN/TEOA) is remarkably high. These findings further reveal that reaction conditions (MeCN/TEOA ratio in the feed and total solvent amount) can have a substantial influence on the type of product formed.

Data points accumulated within node x characterize a study of Wang *et al.*,[49] who examined photocatalytic activity of iron-based MOFs. As is evident from Fig. 7, the region of interest does not include the entire UV-vis spectrum; rather, only the visible region is utilized for photocatalytic experiments. This is an anticipated pathway given that MOFs containing iron could be directly activated by visible light, owing to the presence of oxo-ion (Fe–O) clusters.[27] In light of the findings of Wang *et al.*,[49] it is proved that a series of MOFs containing Fe (MIL-101(Fe), MIL-53(Fe), MIL-88(Fe), NH_2-MIL-101(Fe), NH_2-MIL-53(Fe), and NH_2-MIL-88(Fe)) can lead to prevalent generation of HCOOH. Therefore, the particular classification in node x holds significance for illustrating the effect of the type of base metal (photocatalyst properties) on the resulting product type.

In contrast to node x, node ix contains the entire UV-vis spectrum and indicates the category where CH_3OH is the predominant gas product. Relevant data points for this category are linked with a study conducted by Liu *et al.*[50] According to their research findings, the use of a composite type of MOF (such as incorporation of ZIF-8 into Zn_2GeO_4) could lead to a shift in the spectrum and the overall modification could enhance the catalytic performance. In their study, they also mention that there could be the presence of methane, ethanol, and other types of hydrocarbons in trace amounts; however, CH_3OH remains the major gas output both before and after the incorporation process. Accordingly, it can be established that through adjusting reaction conditions and selecting appropriate MOF characteristics, it is possible to promote the production of CH_3OH.

Generation of CH_3OH is important because it represents a forefront in future green chemistry and a valuable compound

Table 3 Confusion matrix for the predominant gas type in liquid-phase processes

	Accuracy	# Of data points	Actual class	Predicted class				Recall
				CH_3OH	CO	H_2	HCOOH	
Training	0.94	24	CH_3OH	21	3	0	0	0.87
		127	CO	0	126	0	1	0.99
		75	H_2	2	5	63	5	0.84
		71	HCOOH	0	0	0	71	1.00
			Precision	**0.91**	**0.94**	**1.00**	**0.92**	
Testing	0.77	12	CH_3OH	11	1	0	0	0.91
		42	CO	0	37	1	4	0.88
		36	H_2	0	5	21	10	0.58
		24	HCOOH	0	0	5	19	0.79
			Precision	**1.00**	**0.86**	**0.77**	**0.58**	

for industry.[51] Unfortunately, it has not been observed in our gas phase dataset. The absence can be explained by the fact that in a gaseous medium, CO_2 is more likely to undergo reduction to CH_4 rather than CH_3OH.[34] In contrast, CO_2 photoreduction in the liquid phase, CH_3OH production is foreseen to be higher than CH_4; due to the growth of carbonic acid, carbonate, and bicarbonate products which result in HCOOH generation.[34] Accordingly, one can explore pathways, leading to widespread production of CH_3OH (or any type of desired output) *via* examining decision tree models.

Lastly, it is imperative to underscore the notable prevalence of H_2 formation. As mentioned earlier, in CO_2 photoreduction by H_2O, there exists a simultaneous water splitting process that results in the generation of H_2.[34] It has been reported that the formation of H_2 as the prevalent product is highly common.[52] The reason can be ascribed to several factors: (1) the transfer process of H_2O to H_2 requires 2-electrons; while the reduction of CO_2 to hydrogenated compounds usually necessitates 4–8 electrons.[34] (2) the utilization of H_2O as a hydrogen source is widespread, given that it is a cost-effective and abundant substance,[34] (3) H_2O can be reduced to H_2 without the necessitation of any supplementary reagents[52] and (4) the standard reduction potential of H_2O to H_2 is zero, whereas that of CO_2 to $^{\bullet}CO_2^-$ is −1.9 V, indicating that water splitting is thermodynamically more encouraging.[34] In essence, the dominance of H_2 aligns with our research findings; H_2 is the leading gas in 90 instances in the liquid dataset.

The confusion matrix related to liquid phase processes is provided in Table 3. Significant patterns are spotted for classification of CH_3OH as it is able to accumulate in a single leaf node with 100% accuracy and precision. The accuracy for the other products is also quite good indicating that the change in predominant product type as a result of changing the photocatalyst properties and operational conditions is also predictable. Compared to the gas phase, it is noticeable that the accuracy values for both training and testing are slightly decreased in the liquid phase; nevertheless, the model still displays a high degree of effectiveness. Although the number of data points utilized for the creation of a model for liquid medium processes is greater, reductions in accuracy and precision values are totally expectable because the classification is performed on four distinct classes (instead of two in the gas phase). It is justifiable to suggest that a robust classification model is achieved for the liquid phase dataset as well.

Conclusions

Through implementing the optimized hyperparameters of *ntree* (as 120) and *mtry* (as 14) resulting from the 5-fold cross validation strategy, our regression model was able to demonstrate its ability to make accurate and successful predictions for the total gas production rate as also supported by the low RMSE scores of 0.15, 0.16 and 0.48 for training, validation and testing, respectively with the corresponding R^2 values of 0.96, 0.94 and 0.60. The reactor volume and amount of catalyst per reaction volume are discovered to be the most important variables for total gas production rate prediction. Additionally, our decision tree models, which are developed to classify the predominant product type in the gas and liquid phase separately (because the products are different), were quite successful in predicting the major product that can be obtained under specific photocatalyst and reaction conditions. The overall testing accuracy was 87% for the gas phase and 77% for the liquid phase. For the model concerning the gas phase, CO and CH_4 were the major products in most of the cases while CO_2 pressure, band gap and MOF types (show itself as synthesis time) were used as the decision criteria; the relation of CO_2 pressure, CO_2 concentration, and CO_2/H_2O ratio was also examined with the help of related literature studies. In the liquid phase on the other hand, four product types (CH_2OH, CO, H_2 and HCOOH) were the major type in a sufficiently large number of cases with an apparently more complex reaction network.

One of the most important implications of this work for future studies is that it shows the predictability of the performance using various variables and relative contribution of these variables, providing an opportunity for improvement. The predictive power of the models is expected to be much higher in the future with the use of larger datasets curated continuously using the increasing number of experimental studies published in the literature. Such models will also provide more information for the effects of individual variables so that both photocatalyst and reaction conditions can be optimized further for better performance. The results also showed that the heuristic rules and pathways to determine the predominant product type

can be identified. The models developed with larger datasets in the future not only allow more accurate determination of the major product type, but also provide more information about the feature combinations that can serve as pathways for product distribution. By examining routes leading to desired or undesired (if there are any) product types, one can pinpoint particular paths that should be followed or bypassed. These routes could characterize potential strategies to reach distinct outcomes and are highly beneficial for optimization purposes.

Author contributions

Simay Özsoysal: data curation, methodology, formal analysis, original draft preparation; Burcu Oral: methodology, original draft preparation; Ramazan Yıldırım: conceptualization, supervision, writing – review & editing.

Conflicts of interest

There are no conflicts to declare.

Notes and references

1 L. H. Vieira, L. F. Rasteiro, C. S. Santana, G. L. Catuzo, A. H. M. da Silva, J. M. Assaf and E. M. Assaf, *ChemCatChem*, 2023, 15.
2 D. Lu, X. Li, B. Liu, Y.-H. Zhu, G. Liu, K. Wang, J. Zhou and N. Li, *J. Photonics Energy*, 2023, **13**, 1–17.
3 X. Wang, Z. Zhao, K. Zahra, J. Li and Z. Zhang, *Chem. Res. Chin. Univ.*, 2023, **39**, 580–598.
4 F. Almazán, M. Lafuente, A. Echarte, M. Imizcoz, I. Pellejero and L. M. Gandía, *Chemistry (Switzerland)*, 2023, **5**, 720–729.
5 W. Zhang, W. Huang, B. Wu, J. Yang, J. Jin and S. Zhang, *Coord. Chem. Rev.*, 2023, **491**, 215235.
6 K. Poonia, S. Patial, P. Raizada, T. Ahamad, A. A. Parwaz Khan, Q. Van Le, V. H. Nguyen, C. M. Hussain and P. Singh, *Environ. Res.*, 2023, **222**, 115349.
7 M. Yaseen, R. N. Ali, C. Maouche, H. Jiang, L. Sun, Z. Lu, L. Wang, H. Tang, J. Yang and Q. Liu, *Ceram. Int.*, 2022, **48**, 32677–32695.
8 T. Luo, L. Gilmanova and S. Kaskel, *Coord. Chem. Rev.*, 2023, **490**, 215210.
9 M. Tahir, B. Ajiwokewu, A. A. Bankole, O. Ismail, H. Al-Amodi and N. Kumar, *J. Environ. Chem. Eng.*, 2023, **11**, 109408.
10 L. Ding, Y. Li, Y. Ding, F. Bai, B. Jia, H. Li and X. Wang, *Appl. Surf. Sci.*, 2023, **624**, 157100.
11 B. Oral, E. Can and R. Yildirim, *Int. J. Hydrogen Energy*, 2022, **47**, 19633–19654.
12 E. Can and R. Yildirim, *Appl. Catal., B*, 2019, **242**, 267–283.
13 D. Saadetnejad, B. Oral, E. Can and R. Yıldırım, *Int. J. Hydrogen Energy*, 2022, **47**, 19655–19668.
14 M. G. Yildiz, T. Davran-Candan, M. E. Günay and R. Yildirim, *J. CO2 Util.*, 2019, **31**, 27–42.
15 L. Ge, Y. Ke and X. Li, *Chem. Commun.*, 2023, 59.
16 H. Mai, T. C. Le, D. Chen, D. A. Winkler and R. A. Caruso, *Chem. Rev.*, 2022, 122.
17 T. Bailey, A. Jackson, R. A. Berbece, K. Wu, N. Hondow and E. Martin, *J. Chem. Inf. Model.*, 2023, **63**, 4545–4551.
18 W. Wang, L. Zhang, C. Cai, S. Li, H. Liang, Y. Wu, H. Zheng and Z. Qiao, *Sep. Purif. Technol.*, 2023, **325**, 124546.
19 J. Guan, T. Huang, W. Liu, F. Feng, S. Japip, J. Li, X. Wang and S. Zhang, *Cell Rep. Phys. Sci.*, 2022, **3**, 100864.
20 M. Alizamir, A. Keshavarz, F. Abdollahi, A. Khosravi and S. Karagöz, *Sep. Purif. Technol.*, 2023, **325**, 124689.
21 S. Kancharlapalli and R. Q. Snurr, *ACS Appl. Mater. Interfaces*, 2023, **15**, 28084–28092.
22 W. Gao, W. Zheng, K. Yan, W. Sun and L. Zhao, *Fuel*, 2023, **350**, 128757.
23 H. Demir, H. Daglar, H. C. Gulbalkan, G. O. Aksu and S. Keskin, *Coord. Chem. Rev.*, 2023, **484**, 215112.
24 J. Lin, Z. Liu, Y. Guo, S. Wang, Z. Tao, X. Xue, R. Li, S. Feng, L. Wang, J. Liu, H. Gao, G. Wang and Y. Su, *Nano Today*, 2023, **49**, 101802.
25 A. Dhakshinamoorthy, A. M. Asiri and H. García, *Angew. Chem., Int. Ed.*, 2016, 55.
26 X. Zhao, J. Li, X. Li, P. Huo and W. Shi, *Chin. J. Catal.*, 2021, 42.
27 R. R. Ikreedeegh and M. Tahir, *J. CO_2 Util.*, 2021, **43**, 101381.
28 A. Liaw and M. Wiener, *R News*, 2002, **2**, 18–22.
29 H. T. T. Nguyen, T. M. Doan and V. Radeloff, in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 2018, vol. 42.
30 T. Therneau and E. Atkinson, *Comprehensive R Archive Network*, 2019, pp. 1–60.
31 N. Shehzad, M. Tahir, K. Johari, T. Murugesan and M. Hussain, *J. CO2 Util.*, 2018, 26.
32 Y. Zhang, D. Yao, B. Xia, M. Jaroniec, J. Ran and S. Z. Qiao, *ACS Energy Lett.*, 2022, **7**, 1611–1617.
33 H. Zhao, X. Yang, R. Xu, J. Li, S. Gao and R. Cao, *J. Mater. Chem. A*, 2018, **6**, 20152–20160.
34 E. Karamian and S. Sharifnia, *J. CO_2 Util.*, 2016, **6**, 194–203.
35 M. Tahir and N. A. S. Amin, *Appl. Catal., B*, 2013, **142**, 512–522.
36 C. Zhao, A. Krall, H. Zhao, Q. Zhang and Y. Li, *Int. J. Hydrogen Energy*, 2012, **37**, 9967–9976.
37 S. Ali, M. C. Flores, A. Razzaq, S. Sorcar, C. B. Hiragond, H. R. Kim, Y. H. Park, Y. Hwang, H. S. Kim, H. Kim, E. H. Gong, J. Lee, D. Kim and S. Il, *Catalysts*, 2019, **9**, 727.
38 S. Yan, S. Ouyang, H. Xu, M. Zhao, X. Zhang and J. Ye, *J. Mater. Chem. A*, 2016, **4**, 15126–15133.
39 A. Crake, K. C. Christoforidis, A. Kafizas, S. Zafeiratos and C. Petit, *Appl. Catal., B*, 2017, **210**, 131–140.
40 S. Wan, M. Ou, Q. Zhong and X. Wang, *Chem. Eng. J.*, 2019, **358**, 1287–1295.
41 H. Yan, X. Wang, M. Yao and X. Yao, *Prog. Nat. Sci.: Mater. Int.*, 2013, **23**, 402–407.
42 M. Que, Y. Zhao, Y. Yang, L. Pan, W. Lei, W. Cai, H. Yuan, J. Chen and G. Zhu, *ACS Appl. Mater. Interfaces*, 2021, **13**, 6180–6187.
43 Z. C. Kong, J. F. Liao, Y. J. Dong, Y. F. Xu, H. Y. Chen, D. Bin Kuang and C. Y. Su, *ACS Energy Lett.*, 2018, **3**, 2656–2662.
44 N. Sadeghi, S. Sharifnia and M. Sheikh Arabi, *J. CO_2 Util.*, 2016, **16**, 450–457.

45 W. J. Yin, B. Wen, Q. Ge, X. B. Li, G. Teobaldi and L. M. Liu, *Dalton Trans.*, 2020, **49**, 12918–12928.
46 D. Sun, W. Liu, Y. Fu, Z. Fang, F. Sun, X. Fu, Y. Zhang and Z. Li, *Chem.–Eur. J.*, 2014, **20**, 4780–4788.
47 Y. Wang, N. Y. Huang, J. Q. Shen, P. Q. Liao, X. M. Chen and J. P. Zhang, *J. Am. Chem. Soc.*, 2018, **140**, 38–41.
48 O. Ola and M. M. Maroto-Valer, *J. Photochem. Photobiol., C*, 2015, 24.
49 D. Wang, R. Huang, W. Liu, D. Sun and Z. Li, *ACS Catal.*, 2014, **4**, 4254–4260.
50 Q. Liu, Z. X. Low, L. Li, A. Razmjou, K. Wang, J. Yao and H. Wang, *J. Mater. Chem. A*, 2013, **1**, 11563–11569.
51 F. Dalena, A. Senatore, A. Marino, A. Gordano, M. Basile and A. Basile, in,*Methanol Science and Engineering*, 2018, pp. 3–28.
52 A. Corma and H. Garcia, *J. Catal.*, 2013, **308**, 168–175.