Scientific Reports

https://doi.org/10.1038/s41598-026-36838-y

Article in Press

Bridging photocatalysis and artificial intelligence to maximize CH_4 and CO production from CO_2 reduction using synthesized g-C_3N_4/TNTAs photocatalysts

Received: 6 October 2025

Accepted: 16 January 2026

Published online: 06 March 2026

Cite this article as: Hossen M.A., Prima M., Aziz A.A. *et al.* Bridging photocatalysis and artificial intelligence to maximize CH_4 and CO production from CO_2 reduction using synthesized g-C_3N_4/TNTAs photocatalysts. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-36838-y

Md. Arif Hossen, Meherunnesa Prima, Azrina Abd Aziz, Nadeem A. Khan & Yunus Ahmed

Bridging photocatalysis and artificial intelligence to maximize CH_4 and CO production from CO_2 reduction using synthesized g-C_3N_4/TNTAs photocatalysts

Md. Arif Hossen[1*], Meherunnesa Prima[2], Azrina Abd Aziz[3,4], Nadeem A Khan[5], Yunus Ahmed[2*]

[1] Institute of River, Harbor and Environmental Science (IRHES), Chittagong University of Engineering & Technology, Chattogram-4349, Bangladesh.
[2] Department of Chemistry, Chittagong University of Engineering & Technology, Chattogram-4349, Bangladesh.
[3] Faculty of Civil Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, 26300 Gambang, Pahang, Malaysia.
[4] Center for Advanced Intelligent Materials, Universiti Malaysia Pahang Al-Sultan Abdullah, 26300 Gambang, Pahang, Malaysia.
[5] Civil Engineering Department, College of Engineering, King Khalid University, Abha, 61421, Saudi Arabia
*Corresponding author: yunus.acctiu@gmail.com

Abstract

The photocatalytic CO_2 conversion into value-added chemicals such as CH_4 and CO is a highly promising sustainable approach to meet rising energy demands while mitigating atmospheric CO_2 levels. This study investigates the prediction and optimization of CH_4 and CO production using tree-based machine learning (ML) models applied to g-C_3N_4/TiO_2 nanotube arrays (TNTAs) photocatalysts for gas-phase CO_2 conversion. To predict the photoconversion rates of CO_2 into CH_4 and CO, several tree-based ML algorithms, including AdaBoost, Bagging, CatBoost, Decision Tree, Extra Trees, Gradient Boosting, HistGradientBoosting, LightGBM, RandomForest, and XGBoost were employed. The production of CH_4 and CO (μmol/cm^2) were designated as the target outputs, while input parameters included catalyst exposed surface area, initial concentration of CO_2, feed pressure, light power,

and irradiation time. The performance of ML algorithms was appraised using five statistical metrics. Bayesian optimization was employed to fine-tune the hyperparameters of the machine learning model algorithms. Among the evaluated models, CatBoost (CB) performed the most accurately, with R^2 = 0.9887 (training) and R^2 = 0.9883 (test) for CH_4 production and R^2 = 0.9885 (training) and R^2 = 0.9874 (test) for CO production. Feature importance analysis and SHAP plots highlighted the significant influence of irradiation time and catalyst exposed surface area on the production efficiency of both products. Additionally, the input parameters were systematically optimized using CB model predictions, which were validated against experimental data and achieved almost similar prediction. These results support the effectiveness of ML-directed modeling in maximizing CO_2 conversion efficiency and revealing the potential of data-driven strategies in steering photocatalytic technology.

Keywords: Photocatalytic CO_2 conversion; g-C_3N_4/TiO_2 nanotube arrays (TNTAs); Machine learning (ML); CatBoost (CB).

1. Introduction

The dependency on fossil fuels is progressively increasing worldwide for energy production, which results in escalating releases of greenhouse gases (GHGs), including CO_2 thus striking a serious and inevitable threat to the environment and sustainable future. The production of solar fuels, which harness the plentiful sunlight, is one of the most auspicious routes for renewable energy production. A sustainable and realistic solution to the present energy crisis and environmental issues can be a photocatalytic CO_2 reduction (PC-CO_2R) approach. During the reduction process, CO_2 is usually converted into C_1 (one-carbon) and C_{1+} (multi-carbon) compounds solar fuels such as CO, CH_4, CH_3OH, C_2H_4 and C_2H_5OH [1–3]. The utilization of reliable and effective photocatalysts as well as appropriate reaction conditions are crucial for enhancing the production of solar fuels.

Semiconductor photocatalysts are the most auspicious candidates for the photoinduced CO_2 reduction [4,5]. The most investigated UV light-active photocatalyst is TiO_2 owing to its non-toxicity, photo-chemical stability, and availability [6,7]. Anodic TiO_2 in the form of nanotube arrays (NTAs) have drawn an ample amount of attention as prospective photocatalysts to overcome its inherent wide bandgap constraint [8–11]. Still, TNTAs require modification to reduce the rate of recombination of photoexcited charges and enhance performance under the exposure of visible light. Graphitic carbon nitride (g-C_3N_4) synthesized from earth-abundant elements, is another widely used semiconductor photocatalyst owing to its diverse hierarchical structures, outstanding photochemical stability and charge transfer efficiency [12]. Despite these benefits, the rapid charge carrier recombination rate and relatively broad bandgap usually prevent it from being effective as a standalone photocatalyst. The development of a robust heterojunction between TNTAs and g-C_3N_4 is one of the most practical ways to lessen the recombination of photogenerated charge carriers and enhance the visible light sensitivity of pure TNTAs and g-C_3N_4 [13,14]. Besides the properties of photocatalytic materials, the photocatalytic reaction conditions also significantly affect the efficiency of PC-CO_2R. Therefore, to achieve practical applications of photocatalytic CO_2 reduction, it is essential to optimize photocatalytic settings and find out the most influential parameters.

Across a wide range of scientific fields, machine learning (ML) has become a potent tool that facilitates precise predictions directed by leveraging data-driven different algorithms, effective optimization, and a deeper understanding of complicated phenomena [15–17]. In recent times, many ML techniques, such as regression, neural network, and decision tree-based models, have often been used to design photocatalyst materials, optimize process parameters and predict photocatalytic reduction efficiency. For example, Zhang and Xu [18] employed a gaussian process regression (GPR) model to optimize the bandgaps of doped-TiO_2 photocatalysts

considering various structural and morphological attributes as input parameters. In another study, the partial least-squares (PLS) ML algorithm was used for the virtual design of rare-earth-metal-doped TiO_2 photocatalyst materials [19]. Li and coworkers [20] combined ML and rigorous experimental investigation to discover photo-active molecular organic photocatalysts. In a similar type of study, Tao et al. [21] utilized different tree-based and neural network-based ML algorithms to discover new visible light-active perovskite photocatalyst materials. Recently, Yurova and colleagues [22] optimized the synthesis conditions to achieve optimal properties of g-C_3N_4 photocatalysts applying multiple tree-based ML algorithms. To forecast bandgap and photo-adsorption characteristics of emerging metal-organic frameworks (MOFs) materials, Jose and co-workers [23] employed diverse regression tree-based models. More straightforward and low-dimensional descriptors were found to be more effective in the low data regime than higher-dimensional descriptors, which were appropriate for handling large amounts of datasets. The scholarly community increasingly employs tree-based ML models to build photocatalyst materials because of their precision in predictions and capacity to handle substantial amounts of data.

Beyond the design and discovery of photocatalytic materials, applying different ML algorithms to predict photocatalytic efficiency is well supported by extensive literature. Researchers are increasingly applying machine learning techniques to improve the accuracy of predicting photocatalytic pollutants removal, detection, and energy production. For instance, Liu and coworkers [24] applied four ML models including linear regression (LR), random forest (RF), XGBoost (XGB) and LightGBM (LGBM) to predict the degradation efficiency of different organic dyes over doped-TiO_2 photocatalysts. The LGBM model with the best credibility (R^2 = 0.898) suggested irradiation time as the most crucial parameter to enhance degradation efficiency. Most recently, PC removal efficiency of methylene blue (MB) dye was optimized using ten diverse tree-based ML algorithms over

synthesized $CuWO_4$@TiO_2 composite [25]. Among all utilized models, they reported HistGradientBoosting (HGB) as the best-suited model with R^2 value of 0.992 at testing stage to achieve 98.99 % prediction efficiency of MB. To predict the photocatalytic (PC) degradation efficiency of tetracycline (TC) antibiotics in wastewater using metal-organic framework (MOF)-based photocatalysts, Salahshoori et al. [26] implemented an ensemble of ML models. Among these, the hybrid GAPSO-LSSVM model demonstrated superior accuracy and reliability in forecasting TC photodegradation performance. In a related study, Javed et al. [27] utilized a comprehensive dataset derived from TiO_2-based photocatalysts to predict the photocatalytic degradation activity of airborne pollutants, evaluating a total of 13 distinct ML algorithms for this purpose. The XGBoost (XGB) model outperformed other models as robust and effective in predicting degradation efficiency. The dual-functional performance including detection and degradation of organic pollutants of Ag nanoparticles (NPs)-decorated ZnO nanorods (NRs) coated by SiO_2 nanofibers (NFs) was predicted to be using different ML-assisted algorithms [28]. The integrated k-nearest neighbors (KNN) model demonstrated 90.8 % accuracy in quantitative categorization and 92.3 % precision in qualitative detection. To design the synthesis of carbon quantum dots (CQDs) photocatalysts and predict their pollutant detection performance, Kakhki and Mohammadpoor [29] utilized multiple regression algorithms. The most favorable values of performance indicator features were found in the XGB model.

Compared to the PC degradation rate and the prediction of pollutants in contaminated sources, the application of ML algorithm techniques to forecast photocatalytic energy production is insufficient. Most studies focused on using secondary data to predict the energy generation efficiency of PCs. Tao et al. [21] collected 160 perovskite-based visible photocatalysts data points from 43 experimental literature to predict H_2 evolution rate (HER). They employed four distinct ML algorithms namely support vector regression

(SVR), gradient boosting regression (GBR), backpropagation artificial neural network (BPANN), and random forest (RF). The BPANN model demonstrated the highest R^2 (0.9740 for the testing data. Recently, Haghshenas and his team [30] utilized 971 secondary data points to forecast the PC HER using modified-TiO_2 photocatalysts. Among the five reaction conditions studied, photon flux emerged as the influential factor to control HER. Saadetnejad and colleagues [31] utilized decision trees (DT) and random forest (RF) to predict the PC-CO_2R rate. They collected 281 liquid-phase and 268 gas-phase data points from 80 experimental literature and found that the DT ML algorithm was robust in forecasting CO_2 reduction efficiency with 80 % accuracy. While the RF model showed superior performance in predicting the liquid-phase CO_2 reduction rate with 79 % precision. Similar findings were also observed to predict PC-CO_2R over MOFs-based photocatalysts [32]. In a separate study, Liu and coworkers [15] applied an RF model with various pretreatment strategies to prepare the surfaces of TiO_2 for predicting the liquid-phase CO_2 reduction through the photocatalytic process. The light irradiation method demonstrated the highest effectiveness in enhancing the HCOOH production rate compared to all other pretreatment strategies evaluated. However, there are very limited ML studies predicting photocatalytic CO_2 reduction efficiency, especially for the gas-phase, which is still at a rudimentary stage level. The intricate system of interdependent variables that drive the gas-phase PC-CO_2R process has an impact on several performance indicators. Different machine learning algorithms can handle large, complex and uncertain data sets to investigate comprehensive implicit relationships among various parameters, apply relevant details from the experimental data points, and produce reliable forecasting models through data extraction and evaluation.

This work is a groundbreaking addition to the gas-phase PC CO_2 reduction and production of CH_4 and CO products, and optimizes the reaction conditions to get maximized CO_2 conversion by applying the ML approach over g-C_3N_4/TNTAs composite. Despite the growing body of literature on g-

C_3N_4/TNTAs composites, most studies focus on pollutant degradation, where mass transfer limitations, solvent effects, and scavenger dominate photocatalytic efficiency. The gas-phase photocatalytic CO_2 reduction remains significantly underexplored for this heterostructure by using the ML-driven optimization framework, which representing a key knowledge gap this work tries to fill. A total of ten tree-based ML algorithms were employed with precisely optimized hyperparameters to confirm the robust predictions. Five significant factors were taken into consideration as input parameters to forecast the rates of CH_4 and CO production: catalyst exposed surface area, initial concentration of CO_2, feed (CO_2 + H_2O) pressure, light power and irradiation time. The features importance and robustness of the best model were assessed by using the SHapley Additive exPlanations (SHAP) and Gaussian noise technique. The determination of optimal input parameters by differential evolution (DE) optimization considering real-life scenarios and experimental validation provides insightful information for the practical implementation of photocatalytic CO_2 reduction.

2. Materials and methods

2.1. Synthesis and characterization of g-C_3N_4/TNTAs photocatalyst

The two-step anodization method was used to synthesize the TiO_2 nanotube arrays (TNTAs) [11]. The detailed synthesis procedure along with the required chemicals and materials is described in **Texts S1** and **S2**. The overall procedure employed to synthesize TNTAs photocatalysts is shown in **Fig. S1**a. The g-C_3N_4 nanosheets was prepared from melamine powder using the calcination technique and used as a co-catalyst with TNTAs [33]. The g-C_3N_4/TNTAs composite was constructed by a facile dispersion method, as mentioned in a previous study [34]. In this process, pure TNTAs foils were submerged in 250 mL beakers containing a separately prepared g-C_3N_4 methanol solution (0.1 g of g-C_3N_4 dispersed in 50 mL methanol), followed by sonication for 15 minutes to ensure uniform deposition and subsequent drying at 100 °C for 12 hours. The g-C_3N_4 loading was standardized across

all samples by depositing 0.1 g of g-C_3N_4 onto 4 cm^2 of TNTAs, yielding a nominal surface loading of 25 mg/cm^2. Although the exact weight percentage of g-C_3N_4 was not measured due to the absence of pre- and post-deposition mass measurements, an estimated value of ~18 wt.% was derived based on Ti foil thickness of 0.25 mm. The physical, chemical and optical properties of the synthesized photocatalysts were examined by Field Emission Scanning Electron Microscopy (FESEM), X-ray diffraction (XRD), Photoluminescence (PL) and Ultraviolet-visible Diffuse Reflectance Spectroscopy (UV-Vis DRS). The detailed specifications of the characterization tools are provided in **Text S3**.

2.2. Photocatalytic experiments

The photocatalytic (PC) CO_2 reduction performance of the synthesized materials was systematically evaluated in a gas-solid photoreactor (**Fig. S1b**). Specifically, the g-C_3N_4/TNTAs heterostructured photocatalysts were assessed under a range of operational parameters to determine their influence on CO_2 conversion efficiency. These parameters included photocatalyst exposed surface area (4, 8, 12, and 16 cm^2), initial CO_2 concentration (10, 30, 50, and 100 vol%), total feed pressure of the CO_2 + H_2O vapor mixture (100, 200, 300, and 400 mbar), incident light intensity corresponding to lamp powers of 100, 200, 300, and 400 W, and irradiation duration (1, 2, 3, and 4 h). To simulate realistic gas compositions, varying CO_2 concentrations were balanced with inert (Ar, N_2) and oxidizing (O_2) gases. Irradiation was provided by a 500 W adjustable xenon (Xe) arc lamp equipped with a bandpass filter to deliver a wavelength range of 365–720 nm, simulating visible–near-UV solar irradiation. The light source was positioned directly above the reactor to ensure uniform illumination. Prior to each experiment, the reactor was purged with high-purity N_2 for 30 minutes to eliminate residual atmospheric gases and establish an inert baseline environment. During the photocatalytic runs, the reactor chamber was maintained at a controlled temperature of 298–303 K via a closed-loop water

cooling system to minimize thermal effects on reaction kinetics. A series of control experiments were also performed to confirm that the gaseous products is mainly contributed by the photocatalytic reduction of CO_2 rather than through the photodecomposition of organic residues on the surface of photocatalysts. The control experiments were conducted under three conditions: (i) in the absence of light but with photocatalyst and CO_2/H_2O flow, (ii) without photocatalyst but under light irradiation with CO_2/H_2O flow, and (iii) without H_2O vapour but with photocatalyst, CO_2 flow, and light irradiation. At predetermined time intervals, gaseous products generated from the photocatalytic reduction were extracted from the reactor outlet using a calibrated suction pump and stored in pre-evacuated gas sampling bags for subsequent analysis. Quantitative determination of the gaseous products was performed using a Shimadzu GC-2014 gas chromatograph (GC) equipped with both a flame ionization detector (FID) and a thermal conductivity detector (TCD), enabling simultaneous detection of hydrocarbons and permanent gases. The primary reduction products identified were methane (CH_4) and carbon monoxide (CO), with trace amounts of hydrogen (H_2) occasionally observed. Product concentrations were quantified in micromoles (μmol) by correlating peak areas from the chromatograms with calibration curves generated from standard gas mixtures of known composition. Finally, the product formation rate (yield rate) was calculated using Eq. (1), providing a standardised metric for comparing photocatalytic activity across different experimental conditions.

$$\text{Rate of product yield } (\mu\text{mol/cm}^2) = \frac{\text{Amount of products in } \mu\text{mol}}{\text{Exposed surface area of photocatalysts } (\text{cm}^2)} \quad (1)$$

2.3. Data collection and analysis

Data is vital for constructing any machine learning models, because the size, quality, and diversity of datasets greatly affect the accuracy and reliability with which models can forecast performance [35]. This work is

based on two broad datasets, each with 1024 reproducible experimental data points, on predicting CH_4 and CO production efficiencies using g-C_3N_4/TNTAs composite as photocatalysts. These two datasets were developed very carefully by controlled experimental setup and thus are consistent and reliable. The five most crucial above-mentioned photocatalytic factors were employed to investigate the product yield rate during gas-phase PC-CO_2R. The experimental datasets were preprocessed to address inconsistencies, normalize the variables, and prepare the dataset for machine learning algorithms. The normalization of the data was conducted to reduce the overfitting and simplify complex computations according to Eq. (2) [36].

$$\text{Normalized value } X_{norm} = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad (2)$$

In the above equation, X_i represents any data, X_{min} and X_{max} correspond to minimum and maximum values.

2.4. Machine learning models

In the present study, ten distinct tree-based machine learning (ML) algorithms were employed to predict the photocatalytic CO_2 reduction performance of the synthesized materials. These models encompassed AdaBoost (AB), Bagging, CatBoost (CB), Decision Tree (DT), Extra Trees (ET), Gradient Boosting (GB), HistGradientBoosting (HGB), LightGBM (LGB), Random Forest (RF), and Extreme Gradient Boosting (XGB). All ML implementations were carried out using the Scikit-learn library within the Python programming environment, executed on the Google Colab computational platform [37]. A comprehensive schematic illustrating the overall ML modeling workflow is provided in **Fig. 1**, while detailed descriptions of each algorithm's theoretical underpinnings and implementation specifics are elaborated in Supplementary **Text S4**. The dataset was partitioned into training and testing subsets, with approximately 80% of the data allocated for model training and the remaining 20% reserved for testing performance evaluation. This stratified split ensured a robust

assessment of each model's generalization capability on previously unseen data. Model performance was rigorously evaluated using a suite of quantitative metrics, as summarized in **Table 1**. These metrics were formulated based on standard regression evaluation criteria, where y_i denotes the experimentally measured (actual) value, $\hat{y}_i$ represents the corresponding model-predicted (estimated) value, $\bar{y}$ signifies the mean of the observed values, and n corresponds to the total number of data points in the test set. This systematic approach facilitated a comparative analysis of predictive accuracy, robustness, and reliability across all ten tree-based ML models.

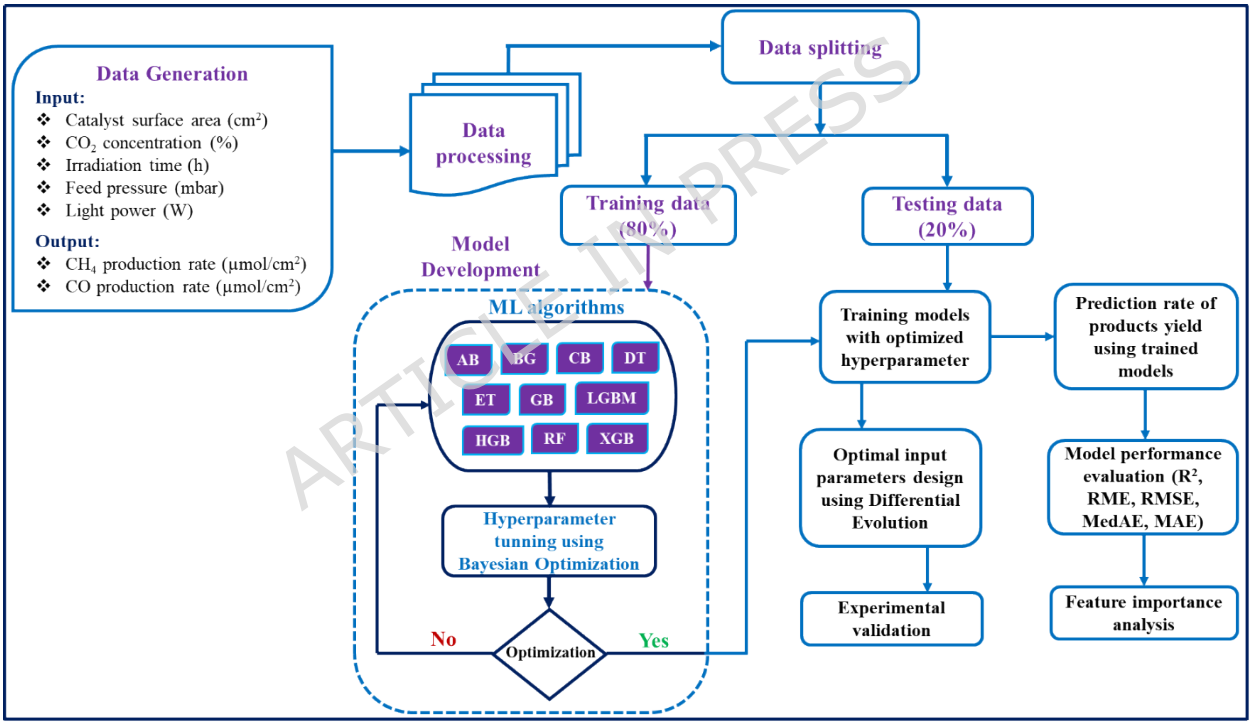


Fig. 1. The illustration of the comprehensive process involved in machine learning (ML) modeling.

Table 1. Evaluation matrices to evaluate the performance of given ML models.

Evaluation Metric	Description	Formula	Best Result
R^2	Coefficient of determination	$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$	1 (perfect prediction)
MSE	Mean squared error	$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	0 (perfect prediction)
RMSE	Root mean squared error	$MSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$	0 (perfect prediction)
MAE	Mean absolute error	$MAE = \frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$	0 (perfect prediction)
MedAE	Median Absolute Error	$MedAE = median(\lvert y_i - \hat{y}_i\rvert)$	0 (perfect prediction)

2.5. Hyperparameter tuning and models interpretation

The hyperparameter setting in machine learning models has a direct and considerable impact on the model's prediction performance [38]. Thus, hyperparameter adjustment of utilized models is an essential step toward ensuring model resilience and prediction accuracy. Bayesian optimization was applied in the tuning of each model's hyperparameters using efficient probabilistic methods of searching in hyperparameter space, such as the Gaussian process (GP). This results in faster convergence, reducing computational costs compared to conventional search methods [39]. The optimal hyperparameter values for all 10 models of CH_4 and CO production rates prediction were computed, and the resulting values are presented in

Tables S1 and S2. Subsequently, all the Bayesian optimization models were trained in the training and evaluated using the testing dataset.

When predicting a target variable, the analysis of feature importance is a study of uncertainty that interprets the model performance by determining the relative value of each feature in a dataset. This analysis ensured that the selected features effectively captured all the necessary elements influencing photocatalytic efficiency and enhanced model reliability [27,40]. The relative importance of all the input parameters on photocatalytic CO_2 reduction were also determined using feature importance analysis. The most widely used SHapley Additive exPlanations (SHAP) analysis was also performed to examine the impact of input features on the predictive performance of the top-performing models. The Differential Evolution (DE) algorithm was used to optimize the input parameters to get maximum production rates of CH_4 and CO using the predictive values of outperformed models. This approach has been instructive not only for prediction efficiency but also for improving experimental settings to attain real-life applications of photocatalytic CO_2 reduction. Experimental validation was performed after getting the optimal conditions for input features. The effectiveness of the top-performing model was assessed using a Gaussian noise method [41].

3. Results and discussion

3.1. Physicochemical analysis of g-C_3N_4/TNTAs

The morphological characteristics of both pristine TiO_2 nanotube arrays (TNTAs) and the g-C_3N_4TNTAs heterostructured photocatalysts were investigated using FESEM. As shown in **Fig. 2a**, the TNTAs exhibited a vertically aligned and highly ordered architecture, with an average inner diameter of 102.57 ± 12.61 nm and a length of 5327 ± 677 nm. Such dimensional enlargement of the nanotubular structure contributes to an expanded specific surface area, which is conducive to enhanced adsorption of reactants and improved photocatalytic efficiency [42,43]. FESEM imaging

of the g-C_3N_4/TNTAs composite (**Fig. 2b**) revealed that ultrathin g-C_3N_4 nanosheets (NSs) were uniformly and robustly anchored onto the external and internal surfaces of the TNTAs, without inducing any noticeable structural distortion. This preservation of the underlying nanotubular morphology confirms the successful and conformal deposition of g-C_3N_4 onto the TNTAs framework. Further structural validation was provided by X-ray diffraction (XRD) analysis (**Fig. 2c**), which displayed the characteristic diffraction peaks of g-C_3N_4 at 2θ values of 12.88° and 27.64°, corresponding to the (100) and (002) crystallographic planes, respectively. These peaks were clearly discernible in the g-C_3N_4/TNTAs composite spectrum, while the anatase (101) and rutile (110) reflections of TiO_2, originating from the TNTAs substrate remained unaltered, indicating that the integration of g-C_3N_4 did not compromise the crystalline integrity of the TiO_2 nanotubes [44,45].

The optical properties of the synthesized photocatalysts were assessed via UV–Vis diffuse reflectance spectroscopy (DRS). As illustrated in **Fig. 2d**, the incorporation of g-C_3N_4 nanosheets induced a marked redshift in the absorption edge of TNTAs, extending light absorption into the visible region (>400 nm). The corresponding bandgap energies were derived from the DRS data using the methodology detailed in Supplementary **Text S5**, employing Tauc plot analysis. As shown in **Fig. 2e**, the bandgap of pristine TNTAs (3.08 eV) was reduced to 2.92 eV upon hybridization with g-C_3N_4. This narrowing of the bandgap enhances the material's ability to harness visible light, thereby promoting more efficient generation of electron–hole pairs and accelerating photocatalytic reaction kinetics [46,47]. Photoluminescence (PL) spectroscopy was further employed to probe the dynamics of charge carrier recombination. Among all modified samples, the g-C_3N_4/TNTAs heterostructure exhibited the lowest PL emission intensity, signifying effective suppression of photogenerated electron-hole recombination. This behavior is attributed to the formation of a well-defined heterojunction at the g-C_3N_4–TNTAs interface, which facilitates spatial separation of charge

carriers under photoexcitation. Such interfacial engineering has been widely reported to enhance charge transfer efficiency and prolong carrier lifetimes, thereby significantly boosting overall photocatalytic performance [48–50].

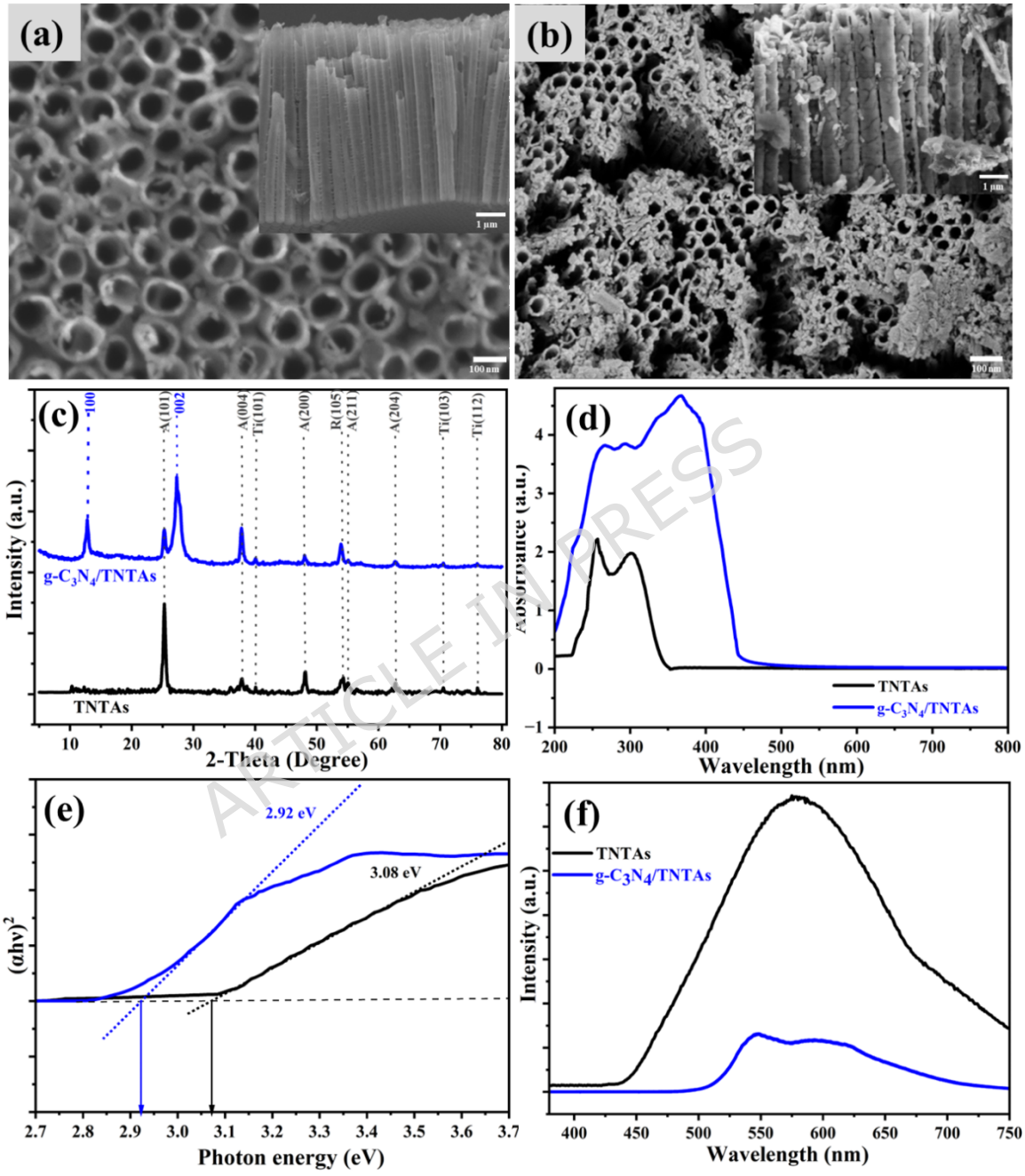


Fig. 2. Characteristics of the synthesized photocatalysts (a, b) FESEM images of TNTAs and g-C_3N_4/TNTAs; (c) XRD; (d) UV-DRS; (e) Tauc-plot; (f) PL spectra.

3.2. Descriptive statistics

In the present study, a comprehensive series of experiments was conducted to evaluate the photocatalytic efficacy of the synthesized g-C_3N_4/TNTAs composite in converting CO_2 into value-added chemical products, specifically methane (CH_4) and carbon monoxide (CO). The pristine TNTAs has been modified using g-C_3N_4 to improve the PC CO_2 reduction activity. As depicted in **Fig. S2**, the rate of CH_4 and CO production improved notably after the modification of TNTAs compared to pure TNTAs under identical experimental conditions. The overall experiments were carried out under systematically varied operational and photoreactor conditions, as detailed in **Table 2**. To facilitate a rigorous statistical interpretation of the experimental outcomes, a multivariate descriptive analysis was performed based on seven key parameters: catalyst surface area, initial CO_2 concentration, feed pressure (CO_2 + H_2O vapor), incident light power, irradiation duration, and the production rates of CH_4 and CO.

Descriptive statistics were computed to characterize the dataset and elucidate the typical experimental conditions. The mean values represent the central tendency of each variable for example, a mean catalyst exposed surface area of 10 cm^2, CO_2 concentration of 47 vol%, feed pressure of 250 mbar, light power of 250 W, irradiation time of 2.5 h, CH_4 production rate of 6.52 µmol/cm^2 and CO production rate of 75.95 µmol/cm^2. These averages provide a benchmark for representative operating conditions within the experimental design. The standard deviation (SD) was calculated to quantify the dispersion or variability of each parameter around its mean, thereby indicating the degree of data spread and experimental consistency. Additionally, the range defined by the minimum and maximum observed values for each variable delineates the boundaries of the experimental domain and informs the spacing and coverage of the sampled conditions.

To further assess the distributional characteristics of the data, percentile values (25th, 50th, and 75th) were determined. These quartiles partition the

dataset into four equal segments, enabling the identification of potential skewness, clustering, or the presence of outliers that may influence model performance or interpretation. Skewness was employed to evaluate the symmetry of each variable's distribution; values near zero suggest approximate symmetry, whereas significant positive or negative skewness indicates a tail extending toward higher or lower values, respectively. Concurrently, kurtosis was used to describe the tailedness and peakedness of the distribution relative to a normal (Gaussian) distribution; negative kurtosis implies a flatter, more uniformly distributed profile with lighter tails. Collectively, these statistical descriptors encompassing measures of central tendency (mean, median), dispersion (standard deviation, range), and shape (skewness, kurtosis) offer a holistic characterization of the experimental dataset. This robust statistical foundation not only enhances the interpretability of the photocatalytic performance trends but also supports subsequent modeling, optimization, and generalization of the CO_2 photoreduction process under diverse operational regimes.

Table 2 Descriptive statistics of the input and output data utilized in this research.

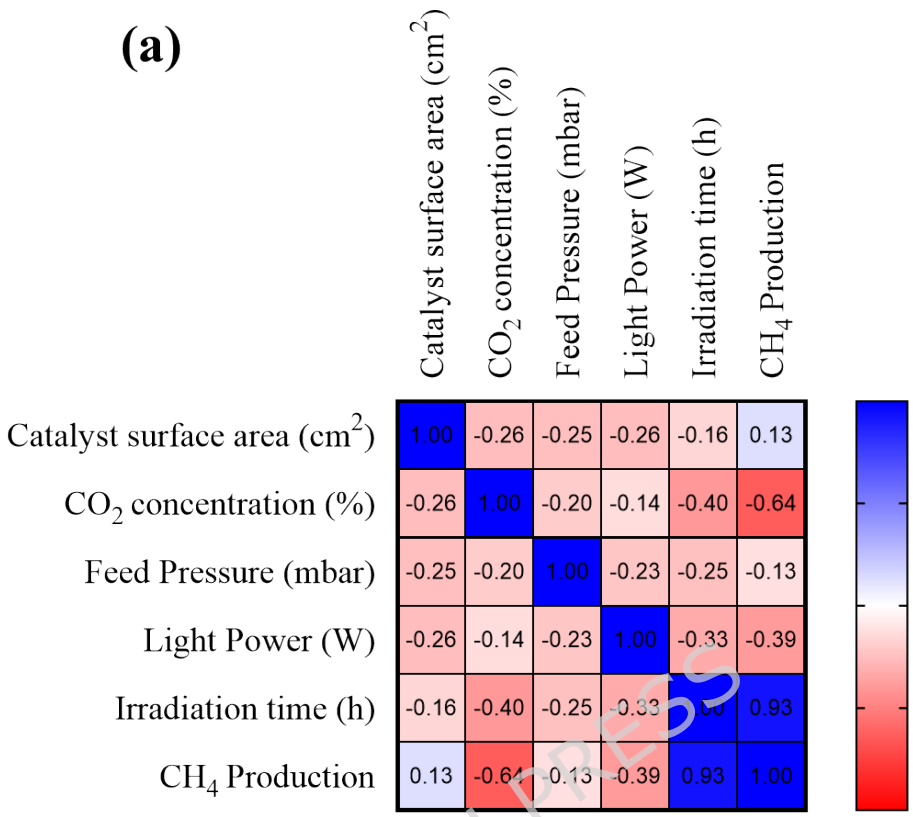
Statistics	Catalyst surface area (cm^2)	Initial concentration of CO_2 (%)	Feed pressure (mbar)	Light power (W)	Irradiation time (h)	CH_4 production ($\mu mol/cm^2$)	CO production ($\mu mol/cm^2$)
Mean	10	47	250	250	2.5	6.52	75.95
Standard deviation	4.47	33.46	111.86	111.86	1.11	2.39	16.51
Minimum	4	10	100	100	1	0.52	34.21
25 %	7	25	175	175	1.75	4.72	63.99

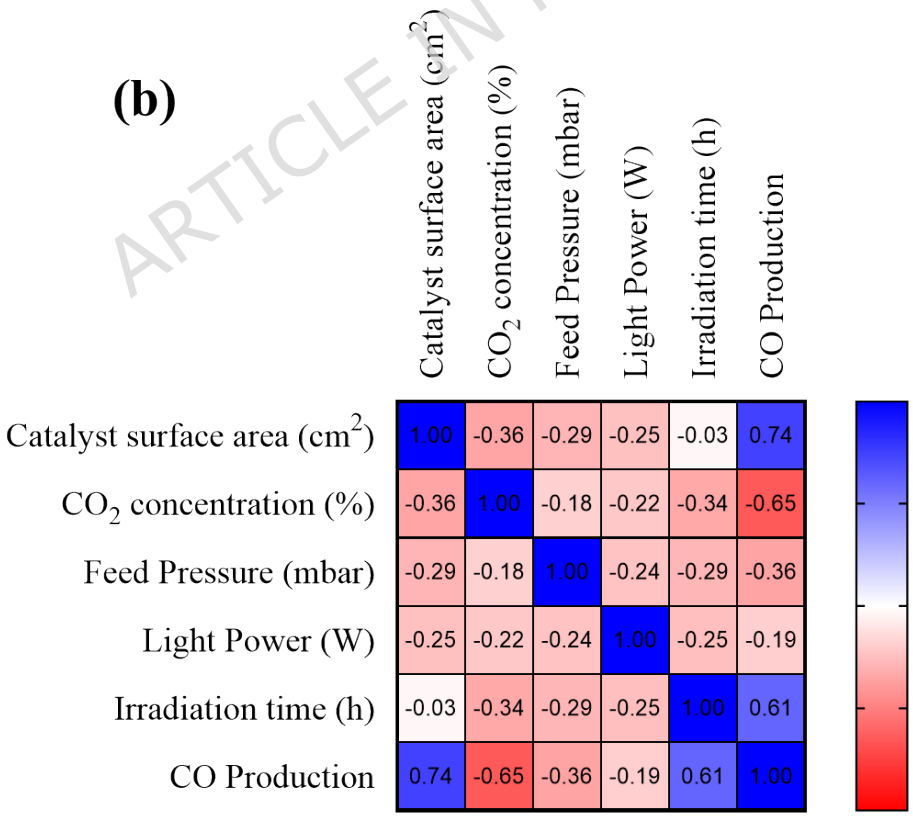
50 %	10	40	250	250	2.5	6.66	76.10
75 %	13	62	325	325	3.25	8.32	87.27
Maximum	16	100	400	400	4	12.12	118.27
Skewness	0	0.58	0	0	0	-0.13	-0.13
Kurtosis	-1.36	-1.07	-1.36	-1.36	-1.36	-0.74	-0.48

To elucidate the interdependencies between the experimental input variables and the photocatalytic response metrics, a Pearson correlation analysis was conducted. The resulting correlation matrix is visualized as a heatmap in **Fig. 3**, which illustrates the pairwise relationships among five input parameters: catalyst surface area, initial CO_2 concentration, feed pressure, light power, and irradiation time, as well as the two primary output variables: CH_4 and CO production rates. In this heatmap, the intensity and hue of the color encoding convey both the magnitude and direction of the linear associations. Specifically, darker shades denote stronger correlations, while lighter tones indicate weaker relationships. Blue hues signify positive correlations (i.e., as one variable increases, the other tends to increase), whereas brown or reddish tones represent negative correlations (i.e., an increase in one variable corresponds to a decrease in the other), with reference to photocatalytic conversion efficiency. Notably, the analysis revealed minimal intercorrelation among the input variables themselves, suggesting that these parameters are largely independent and non-collinear a desirable characteristic for robust multivariate modeling. The most pronounced correlation was observed between irradiation time and CH_4 production rate, with a Pearson correlation coefficient of $r = 0.93$ ($p < 0.01$) (**Fig. 3a**), indicating a very strong positive linear relationship. This implies that prolonged exposure to light irradiation significantly enhances methane generation, likely due to extended photoexcitation duration facilitating

greater charge carrier utilization. A similarly positive, albeit moderate, correlation was identified between irradiation time and CO production rate (r = 0.64, p < 0.01) (**Fig. 3b**), underscoring the beneficial role of longer irradiation periods in promoting CO evolution as well.

In contrast, the initial CO_2 concentration exhibited substantial negative correlations with both CH_4 (r = –0.64, p < 0.01) and CO (r = –0.64, p < 0.01) production rates. This inverse relationship suggests that higher CO_2 concentrations in the feed gas mixture may impede photocatalytic efficiency potentially due to mass transfer limitations, competitive adsorption effects, or reduced light penetration at elevated gas densities. Furthermore, catalyst exposed surface area demonstrated a significant positive correlation with CO production (r = 0.74, p < 0.01), highlighting the importance of accessible active sites in driving CO formation. In comparison, the remaining input parameters (e.g., feed pressure and light power) showed relatively weak or negligible correlations with the product yields, indicating a less dominant influence under the tested conditions. Overall, the correlation heatmap provides an intuitive and informative visual summary of the directional and quantitative relationships between operational parameters and photocatalytic performance. This insight not only aids in identifying key drivers of CO_2 photoreduction efficiency but also informs the selection and prioritization of variables for subsequent predictive modeling and process optimization.


(a)
Catalyst surface area (cm2)
CO2 concentration (%)
Feed Pressure (mbar)
Light Power (W)
Irradiation time (h)
CH4 Production
Catalyst surface area (cm2) 1.00 -0.26 -0.25 -0.26 -0.16 0.13
CO2 concentration (%) -0.26 1.00 -0.20 -0.14 -0.40 -0.64
Feed Pressure (mbar) -0.25 -0.20 1.00 -0.23 -0.25 -0.13
Light Power (W) -0.26 -0.14 -0.23 1.00 -0.33 -0.39
Irradiation time (h) -0.16 -0.40 -0.25 -0.33 1.00 0.93
CH4 Production 0.13 -0.64 -0.13 -0.39 0.93 1.00


(b)
Catalyst surface area (cm2)
CO2 concentration (%)
Feed Pressure (mbar)
Light Power (W)
Irradiation time (h)
CO Production
Catalyst surface area (cm2) 1.00 -0.36 -0.29 -0.25 -0.03 0.74
CO2 concentration (%) -0.36 1.00 -0.18 -0.22 -0.34 -0.65
Feed Pressure (mbar) -0.29 -0.18 1.00 -0.24 -0.29 -0.36
Light Power (W) -0.25 -0.22 -0.24 1.00 -0.25 -0.19
Irradiation time (h) -0.03 -0.34 -0.29 -0.25 1.00 0.61
CO Production 0.74 -0.65 -0.36 -0.19 0.61 1.00

492

Fig. 3. Heatmap representation of the experimental dataset of (a) CH_4 and (b) CO production.

3.3. Performance evaluation of ML models

3.3.1. Prediction of CH_4 production

Photocatalytic CO_2 reduction is an appealing route to sustainable mitigation of greenhouse gas emissions while instantaneously producing value-added products and fuels like CH_4 and CO. In this process, CO_2 is converted to CH_4, a green fuel, and CO, an important industrial feedstock, through the action of photoactivated catalysts. Precise quantitation of CH_4 and CO production is paramount to evaluate the efficiency of the photocatalytic system and its environmental footprint. Tracking these products not only ensures the highest efficiency of the photocatalytic reaction but also restricts the formation of unwanted byproducts, paving the way towards a cleaner environment. By accurately determining CH_4 and CO yield, researchers can optimize reaction conditions, enhance selectivity, and support the development of eco-friendly technologies for renewable energy and carbon-neutral applications. In addition to experimental measurements, ML algorithms have emerged as an essential tool for predicting CH_4 and CO production in photocatalytic CO_2 reduction. By utilizing extensive datasets produced from experiments, ML models have been analyzed through various statistical performance metrics and visual comparison plots. The production of CH_4 via CO_2 photoreduction by the g-C_3N_4/TNTAs photocatalysts was predicted using ten tree-based ML algorithms: AB, BG, CB, DT, ET, GB, HGB, LGBM, RF, and XGB. To assess the prediction accuracy of all listed ML models, scatter and residual plots were presented between observed and predicted CH_4 and CO production on the training and testing datasets. Each scatter plot showed a strong concordance of experimental and predicted values, clustering around the 45-degree line for high accuracy (**Fig. 5**). In each relationship, the ideal regression line denoted by y = x is represented by red dotted lines, while the values of the training and testing data sets are

indicated by black and red circles, respectively. Points close to the red diagonal line indicate accurate predictions, while deviation from the line suggests prediction errors, which may indicate limitations in the model or insufficient data. The residual plots in **Fig. S3** show the error values between the expected and experimental datasets. The residuals were plotted along the x-axis in ascending order of their respective experimental values to observe the distribution patterns better. A model is deemed more dependable when its residual points are distributed around the zero line, indicating good model performance. In contrast, systematic patterns may suggest underfitting or overfitting [25,51,52]. Radar plots serve as an effective tool for summarizing complex datasets and guiding decision-making in multifactorial systems. Each axis of the radar plot represents a distinct parameter, normalized for comparison, while the enclosed area indicates overall performance, with larger areas signifying better outcomes. Deviations along specific axes reveal influential parameters, guiding targeted optimization.

The predictive performance of the ten employed machine learning (ML) algorithms in forecasting CH_4 production during photocatalytic CO_2 reduction was rigorously evaluated using a comprehensive suite of regression metrics, including the coefficient of determination (R^2), mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and median absolute error (MedAE), as detailed in **Table S3**. To facilitate a holistic comparison, the distribution of these performance indicators across both training and testing datasets for all models was visualized using radar plots (**Fig. 4**).

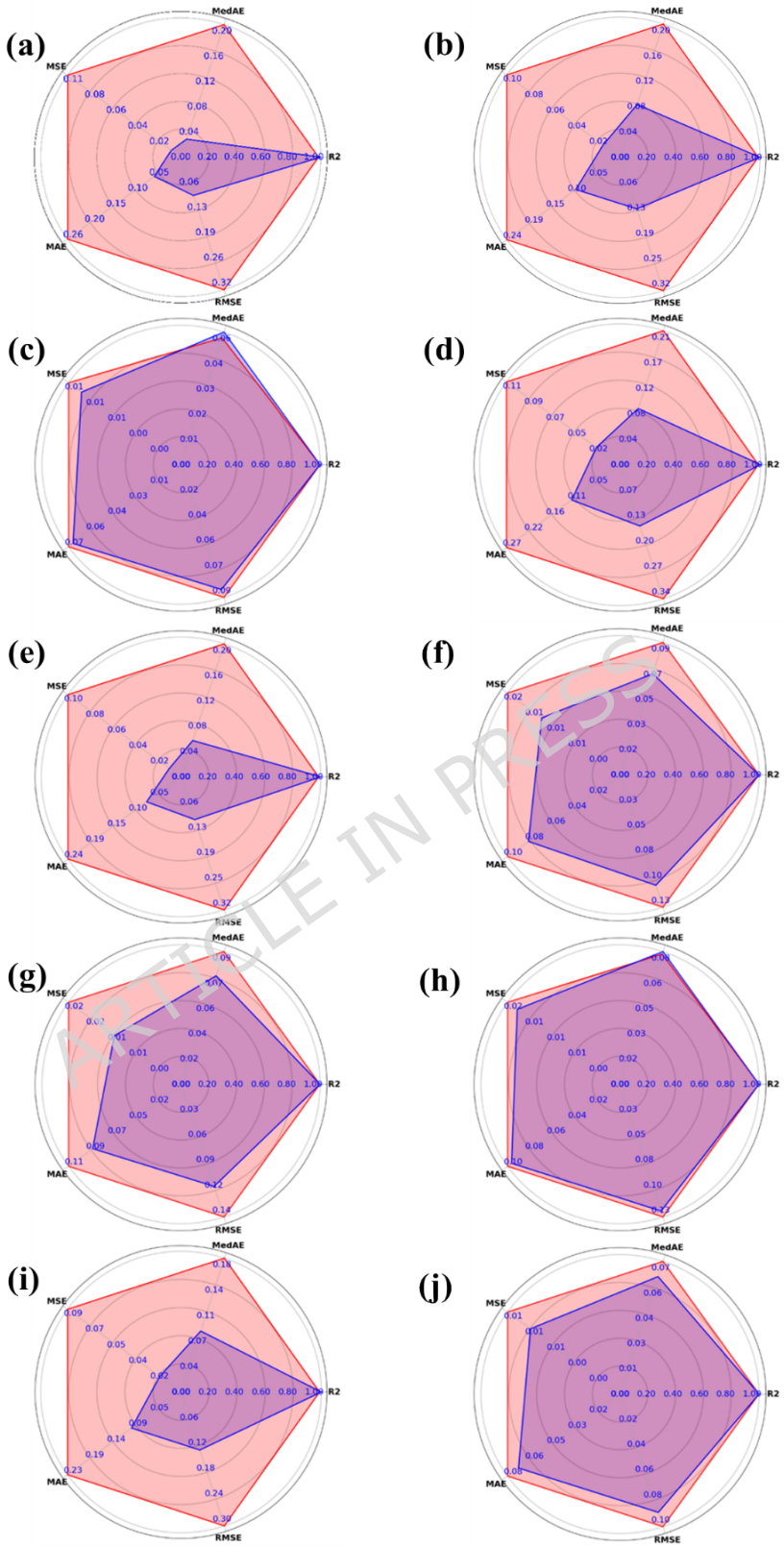
The R^2 metric quantifies the proportion of variance in the observed CH_4 production rates that is explained by the model predictions; values approaching unity indicate an excellent fit and strong alignment between predicted and actual outcomes. The training phase yielded exceptionally high R^2 scores ranging from 0.9172 (AdaBoost) to 0.9887 (Extra Trees) reflecting effective hyperparameter tuning via Bayesian optimization and robust model

fitting. Specifically, the R^2 values for AdaBoost (AB), Bagging (BG), CatBoost (CB), Decision Tree (DT), Extra Trees (ET), Gradient Boosting (GB), HistGradientBoosting (HGB), LightGBM (LGBM), Random Forest (RF), and XGBoost (XGB) were 0.9172, 0.9877, 0.9887, 0.9784, 0.9864, 0.9883, 0.9823, 0.9772, 0.9769, and 0.9786, respectively. Complementing R^2, error-based metrics provide nuanced insights into prediction accuracy. MSE emphasizes larger errors by squaring deviations between predicted and actual values, thereby penalizing substantial outliers more severely. In contrast, MAE offers a linear measure of average absolute prediction errors, treating all deviations uniformly, while MedAE-based on the median of absolute errors, is more robust to extreme values and better reflects typical prediction performance in the presence of outliers. Lower values across MSE, MAE, and MedAE, along with reduced RMSE (the square root of MSE), collectively signify higher predictive fidelity. The models exhibited notably low MAE values during training: 0.5747 (AB), 0.0871 (BG), 0.0667 (CB), 0.0525 (DT), 0.1117 (ET), 0.0762 (GB), 0.1646 (HGB), 0.2731 (LGBM), 0.1033 (RF), and 0.0701 (XGB), underscoring their capacity to closely approximate true CH_4 yields.

Scatter plots in **Fig. 5** further illustrate model performance by comparing predicted versus actual CH_4 production rates. Points closely aligned with the ideal diagonal (black dashed line) denote high predictive accuracy. Among all models, CatBoost (**Fig. 5c**), Gradient Boosting (**Fig. 5f),** and XGBoost (**Fig. 5j**) demonstrated the tightest clustering around this reference line, indicating superior consistency and minimal deviation across both training and testing phases. However, residual analysis depicted in supplementary residual plots (**Fig. S3**) revealed critical differences in model behavior. Residuals, defined as the difference between observed and predicted values, span a range of –2 to +2 $\mu mol/cm^2$. A pronounced dispersion of residuals, particularly at higher CH_4 production levels, suggests increased prediction uncertainty for elevated CO_2 conversion rates. Notably, the LightGBM model exhibited residuals tightly clustered near zero but failed to capture the full spectrum of data

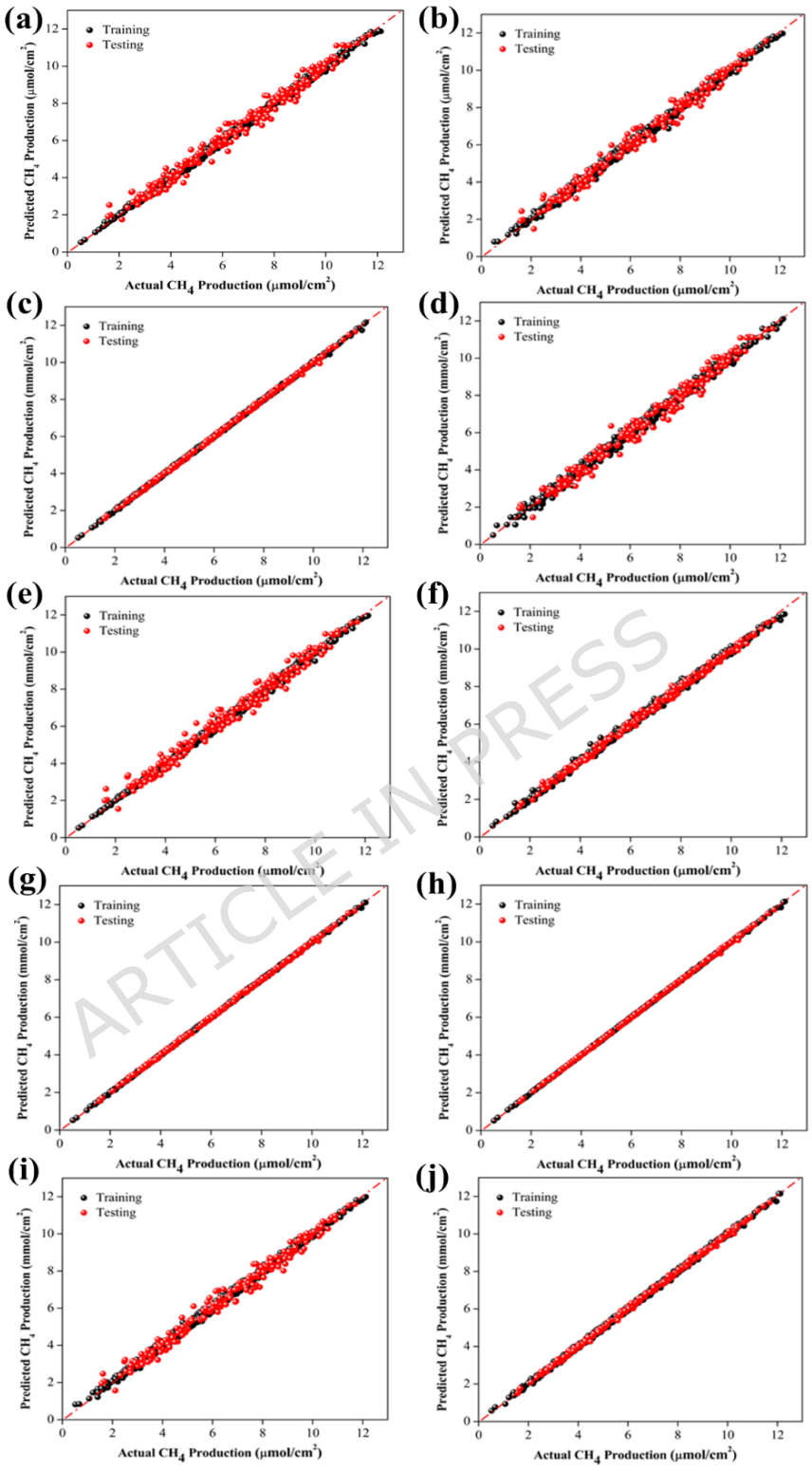
variability, a pattern indicative of overfitting where the model memorizes training patterns rather than learning generalizable relationships. In contrast, the CatBoost (CB) model displayed a more uniform residual distribution confined within a narrower band (–1 to +1 μmol/cm^2), with no systematic bias across the range of predicted outputs. This balanced error profile implies effective generalization and reliable performance on unseen data.

This conclusion is further corroborated by the radar plots in **Fig. 4**, which enable a multi-metric comparison of model efficacy. CatBoost consistently outperformed its counterparts, achieving the lowest values of MSE, RMSE, and MAE while maintaining a high R^2 (>0.98) on both training and testing sets. Crucially, the performance gap between training (pink) and testing (brown) phases for CB was minimal, particularly in error metrics, suggesting limited overfitting and strong robustness. The uniformity observed across datasets underscores CatBoost's suitability for practical deployment in predicting CH_4 yields from photocatalytic CO_2 reduction. This consistency highlights CatBoost model emerges as the most effective among the ten evaluated algorithms, distinguished by its exceptional predictive accuracy, low error rates, balanced generalization capability, and resilience to overfitting. These attributes position it as a robust and reliable tool for modeling and optimizing solar-driven CO_2-to-fuel conversion processes.



604

Fig. 4. The radar plots on training and testing data for CH_4 production (a) AB, (b) BG, (c) CB, (d) DT, (e) ET, (f) GB, (g) HGB, (h) LGBM, (i) RF, and (j) XGB models.



608

Fig. 5. The scatter plot representing the association between the experimental and anticipated CH_4 production for (a) AB, (b) BG, (c) CB, (d) DT, (e) ET, (f) GB, (g) HGB, (h) LGBM, (i) RF, and (j) XGB models.

3.3.2. Prediction of CO production

In the context of photocatalytic CO_2 reduction, accurate prediction of carbon monoxide (CO) production is pivotal for optimizing reaction parameters and maximizing process efficiency. To rigorously assess the predictive capabilities of the ten tree-based machine learning (ML) models under investigation, a suite of visualization and statistical diagnostic tools, including scatter plots, residual plots, and radar plots was employed. As illustrated in **Fig. 6**, the scatter plots comparing predicted versus actual CO production rates reveal that CB, HGB, LGBM, and XGB achieved the highest predictive accuracy, with their data points closely aligned along the ideal 1:1 diagonal line. The remaining models GB, ET, RF, BG, AB, and DT also demonstrated commendable performance, albeit with marginally higher deviations from the reference line, indicating only minor discrepancies in their predictive outputs.

To further scrutinize model fidelity and uncover potential systematic biases, residual analysis was conducted using plots presented in **Fig. S4.** Residual plots serve as a fundamental diagnostic instrument in regression modeling, as they elucidate the distribution of prediction errors and help identify patterns that may indicate overfitting, underfitting, or heteroscedasticity [53]. In this study, residuals were plotted against the sorted experimental (actual) CO production values along the x-axis to facilitate a clearer interpretation of error distribution trends. The residual patterns for the CB and XGB models were notably well-behaved, exhibiting random scatter around the zero-error line without discernible structure, thereby affirming their reliability and robustness in capturing the underlying data-generating process. Quantitative model comparison was further supported by a comprehensive set of statistical performance metrics

compiled in **Table S4**. During the training phase, the coefficient of determination (R^2) values for AB, BG, CB, DT, ET, GB, HGB, LGB, RF, and XGB were 0.9838, 0.9874, 0.9885, 0.9789, 0.9777, 0.9844, 0.9817, 0.9724, 0.9755, and 0.9792, respectively. These high R^2 scores, coupled with consistently strong performance on the independent testing dataset, underscore the efficacy of Bayesian optimization in fine-tuning hyperparameters to enhance predictive accuracy across all models.

Corresponding MAE values for the training set were recorded as 0.0382 (AB), 0.6985 (BG), 0.0846 (CB), 1.8967 (DT), 0.5972 (ET), 0.5311 (GB), 0.4769 (HGB), 2.1519 (LGBM), 0.8797 (RF), and 0.3762 (XGB). Notably, the CatBoost model not only achieved the highest R^2 but also exhibited the lowest error metrics across the board specifically, the smallest MAE, RMSE, and MedAE, thereby establishing its superiority among all evaluated algorithms.This conclusion is visually reinforced by the radar plots in **Fig. 7**, which juxtapose training (blue) and testing (brown) performance across multiple evaluation criteria. For the CatBoost model, the near-perfect overlap between training and testing metric profiles indicates minimal overfitting and exceptional generalization capability. As quantified in Fig. 10c, CatBoost attained an outstanding training R^2 of 0.9885, accompanied by remarkably low error values: MSE = 0.0114, RMSE = 0.1066, MAE = 0.0846, and MedAE = 0.0707. These results reflect a strong concordance between predicted and observed CO yields, with negligible prediction bias. Moreover, on the testing dataset, CatBoost maintained high predictive fidelity, achieving an R^2 of 0.9874 and correspondingly low errors (MSE = 0.0358, RMSE = 0.1891, MAE = 0.1484, MedAE = 0.1187), further validating its robustness and suitability for real-world application in forecasting CO production from photocatalytic CO_2 conversion.This model exhibited higher performance matrix values than other models, especially when predicting larger, more complex, and diverse datasets [54–56]. HistGradientBoosting (**Fig. 7**g) and XGBoost (**Fig. 7**j) show remarkable performance, achieving a high coefficient of determination

(R^2) and demonstrating relatively low error metrics. The Bagging (**Fig. 7**b), ET model (**Fig. 7**e), GB model (**Fig. 7**f), and LightGBM model (**Fig. 7**h), demonstrate superior training performance ($R^2 > 0.982$). Nevertheless, there is a less severe drop in performance on the testing dataset, with an $R^2 > 0.985$. The AdaBoost, and Random Forest, as shown in **Figs. 7**a and **7**i, respectively, display high precision in training with an R^2 value of more than 0.98. On the other hand, they encounter difficulties when trying to generalize their performance. On the other hand, the DT (**Fig. 7**d) demonstrates the lowest robustness compared to other models utilized.

In order to circumvent potential leakage issue, we split the data according to experimental batch/nearest grid cell in groups. This ensures that similar points never end up in both training and test sets. The hyperparameter tuning was done inside the nested cross-validation to prevent information leakage, and the learning curves were checked for overfitting. The permutation testing of CH_4 and CO production gave a mean R^2 of –0.367 and –0.374 ($p < 0.01$) respectively, indicating that our model indeed captures real patterns, not spurious correlations. Outer fold metrics with 95% confidence intervals are $R^2 = 0.9816$ and 0.9818, RMSE = 0.0462 and 0.2175, which shows that the predictive performance is good but generalized realistically (**Table S5**). These results confirm that high accuracy in the model is real and not an artifact of data leakage.

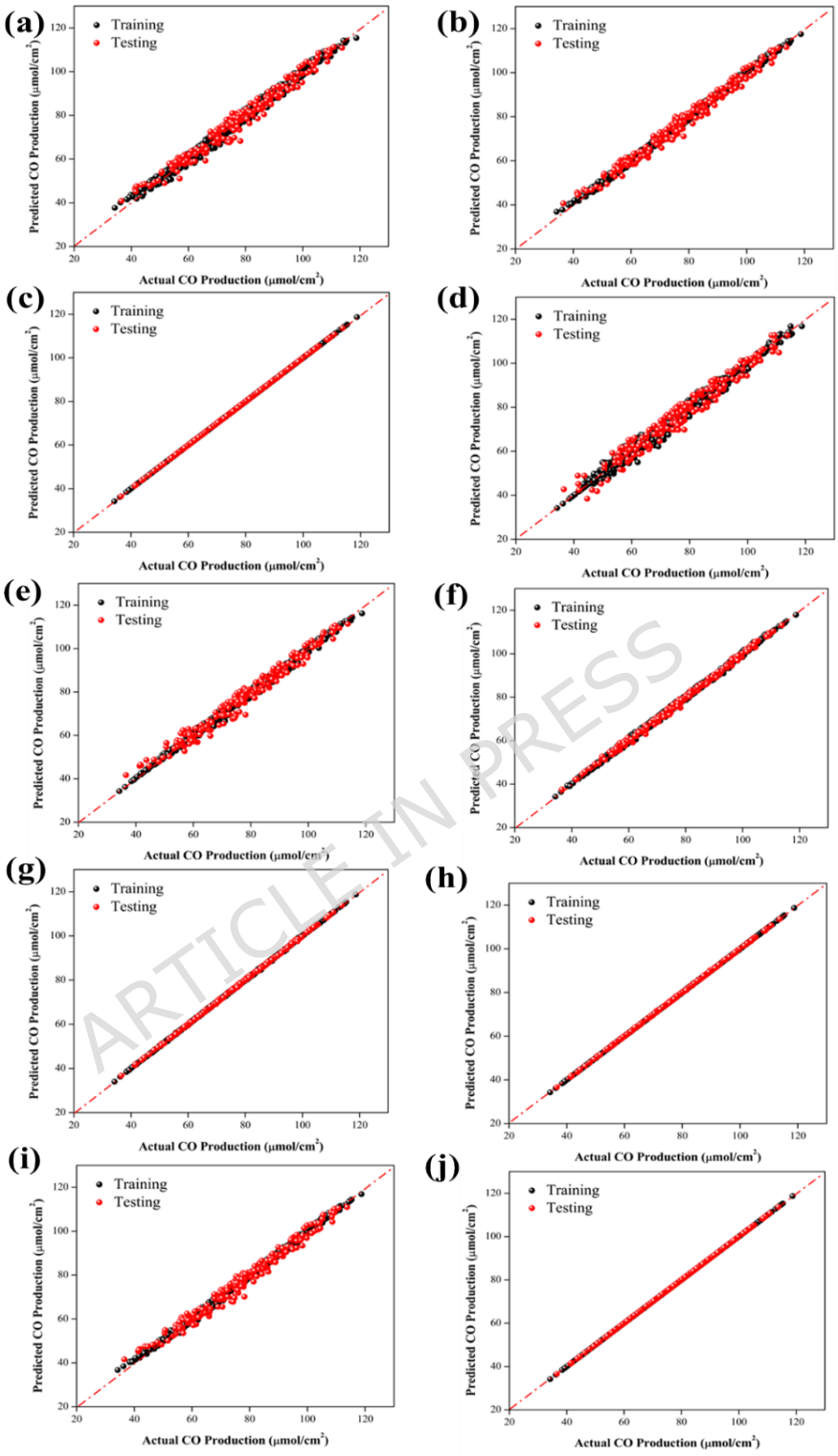
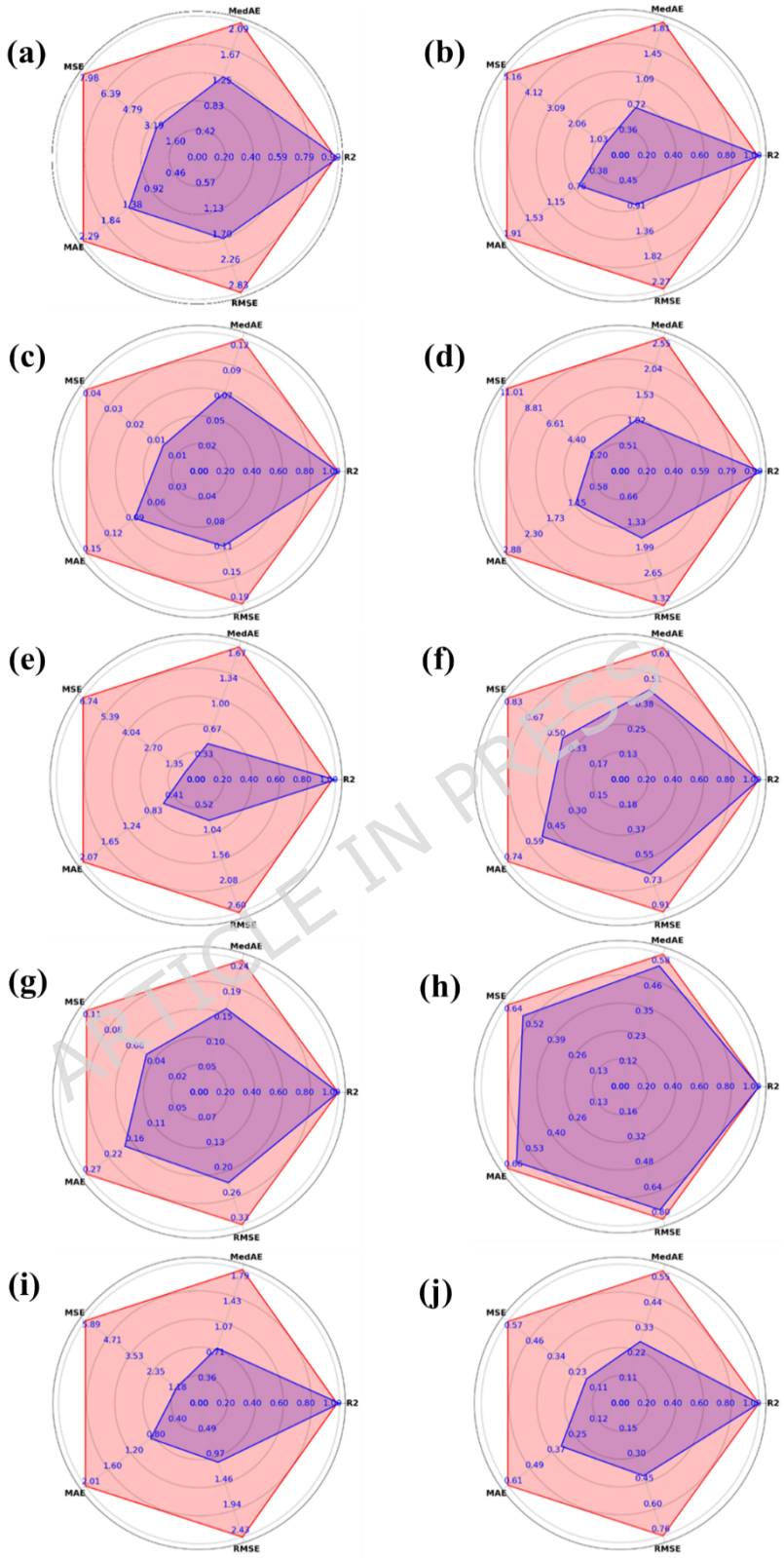


Fig. 6. The scatter plot representing the association between the experimental and anticipated CO production for (a) AB, (b) BG, (c) CB, (d) DT, (e) ET, (f) GB, (g) HGB, (h) LGBM, (i) RF, and (j) XGB models.



694

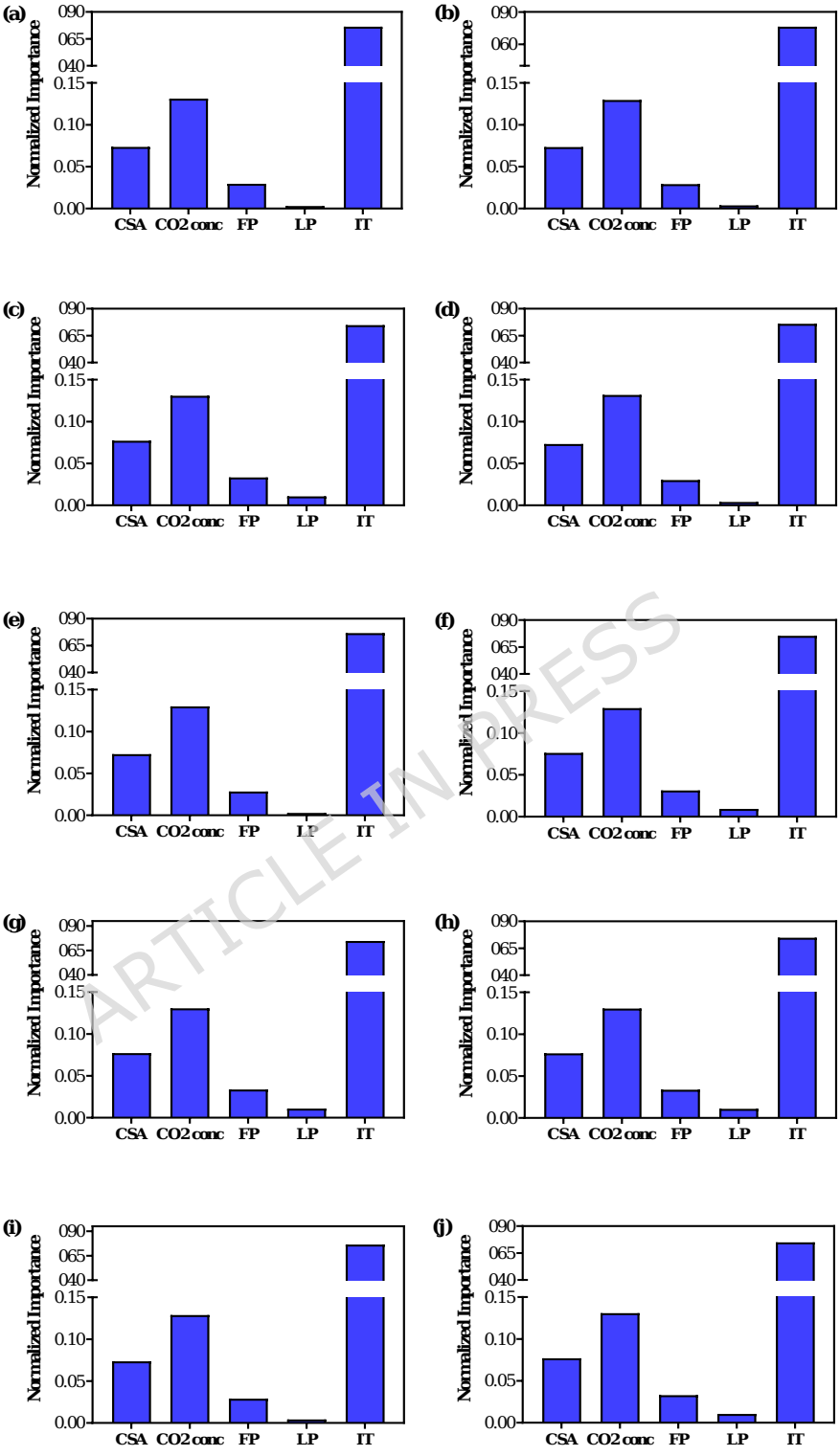
Fig. 7. The radar plots on training and testing data for CO production (a) AB, (b) BG, (c) CB, (d) DT, (e) ET, (f) GB, (g) HGB, (h) LGBM, (i) RF, and (j) XGB models.

The learning curves also show that both CH_4 and CO production models have high training R^2 consistently at ~1.0, showing an excellent fit on the training data (**Fig. S5**). Cross-validation scores are low for small training sizes due to limited samples but increase with the use of more data until reaching a point of convergence near the training score. This trend indicates a scenario in which the model generalizes extremely well when sufficient data becomes available, with very low overfitting for large proportions of training.

3.3.3. Feature importance and robustness of the models

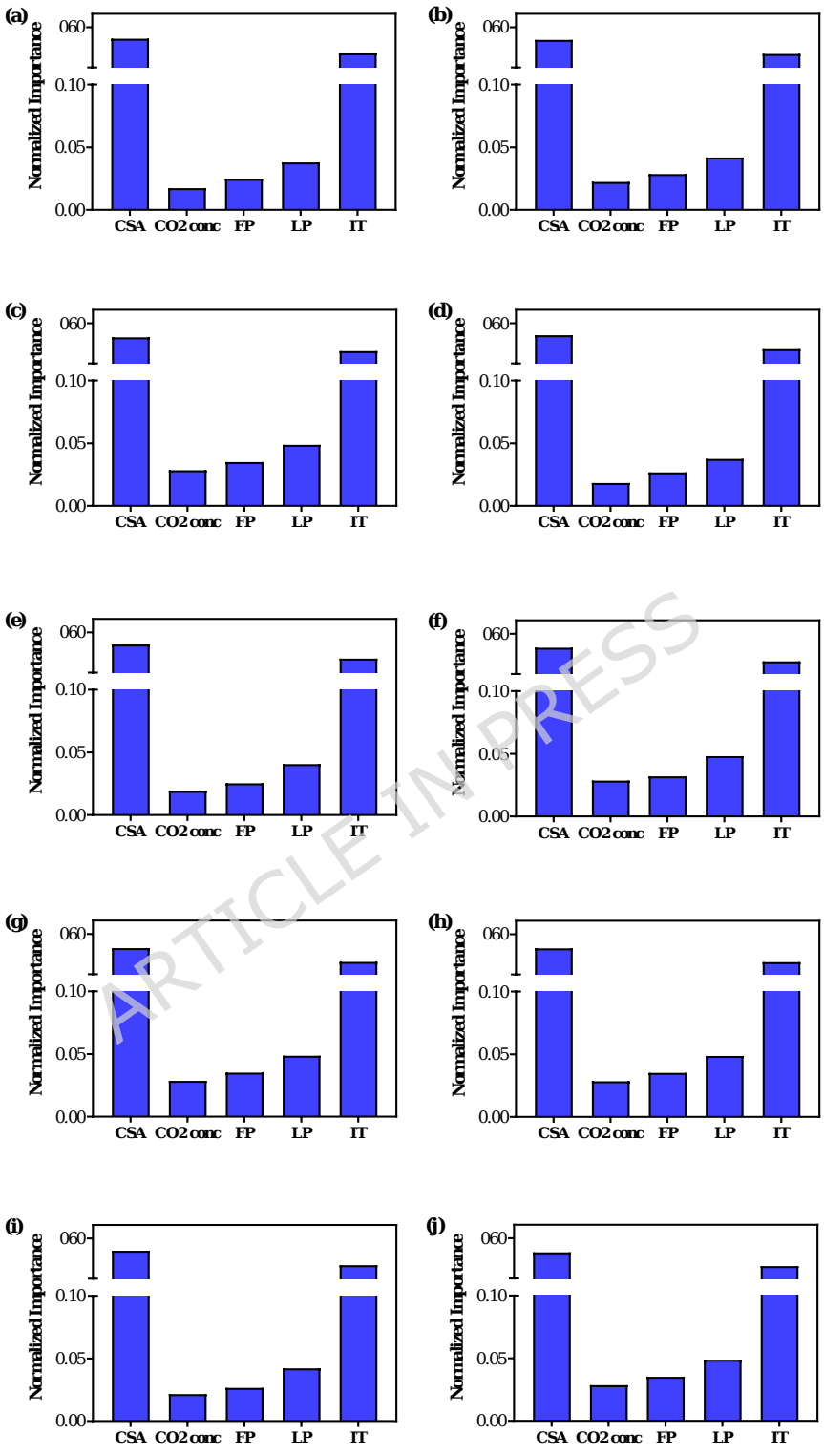
Photocatalytic CO_2 reduction is a complex process influenced by various experimental and photocatalytic parameters, each of them affecting the selectivity toward CH_4 and CO production. It is simple to visualize the impact of input parameters on the targeted variables by using the measure of feature importance. As shown in **Figs. 8** and **9**, the normalized importance of input attributes affecting the competence of PC conversion of CO_2 into CH_4 and CO products had been retrieved for each utilized predictive model. To provide a better overview, the input parameters were abbreviated as CSA for catalyst surface area, CO_2 conc. for initial concentration of CO_2, FP for feed pressure, LP for light power, and IT for irradiation time. For CH_4 production, the IT was identified as the most influential factor by all the models due to the fact that more prolonged exposure allows more chances for better interaction between CO_2 molecules with the catalyst, which enhanced photoconversion efficiency [57–59]. This is followed by CO_2 conc., CSA, FP, and LP in descending order (**Fig. 8**). In CO production, CSA was the most influential parameter, emphasizing the need for ample active sites, followed by IT, LP, FP, and CO_2 conc., which had the least effect (**Fig. 9**). Based on the rating of importance it is apparent that these parameters play a significant role in CH_4 and CO production. Irradiation time is a crucial parameter in both, showing its

importance in maintaining reaction rates for effective conversion. However, for CH_4, CO_2 conc. plays a more vital role in ensuring the sufficiency of the reactant, while catalyst surface area is less critical. The reverse is the case for the production of CO, with high dependence on CSA and negligible CO_2 conc. These differences indicate different reaction pathways and dependencies, which can be of significant value for fine-tuning g-C_3N_4/TNTAs photocatalysts and photocatalytic conditions toward optimum selectivity and efficiency in the CO_2 reduction process. **Tables S6 and S7** demonstrate the statistical robustness of the input features concerning the CH_4 and CO production during photocatalytic CO_2 reduction.



735

736 **Fig. 8.** The feature importance matrix of input parameters for CH_4
737 production evaluated by (a) AB, (b) BG, (c) CB, (d) DT, (e) ET, (f) GB, (g) HGB,
738 (h) LGBM, (i) RF, and (j) XGB models.



739

Fig. 9. The feature importance matrix of input parameters for CO production evaluated by (a) AB, (b) BG, (c) CB, (d) DT, (e) ET, (f) GB, (g) HGB, (h) LGBM, (i) RF, and (j) XGB models.

Understanding the factors that influence photocatalytic processes is critical to optimizing the conversion of CO_2 into valuable chemicals such as CH_4 and CO. In this work, the contribution of key experimental parameters was quantified using SHapley Additive exPlanations (SHAP) to determine the global feature importance rankings of the outperformed CB model using g-C_3N_4/TNTAs photocatalysts. For CH_4 production, **Fig. 10**a represents that the irradiation time showed the most impactful feature in magnitudes of a mean SHAP value higher than +1.75. Such a feature enabled increasing interactions between CO_2 molecules with the catalyst. Further, CO_2 concentration is around +0.73 for enough reactant supply, followed by a catalyst surface area at around +0.57. The corresponding average SHAP values, which are close to +0.35 and +0.18, represent these less pronounced changes in the reaction medium associated with light power and feed pressure. For CO production, as explained in **Fig. 10**b, the catalyst surface area became the most crucial factor, where the average SHAP value surpassed +10.6, showing that the maximization of active sites is essential. It was followed very closely by irradiation time, with a value above +8.7, showing it plays a role in sustaining the reaction. On the other hand, light power, CO_2 concentration, and feed pressure had SHAP values of approximately +2.4, +2.1, and +1.9, respectively, which describes a lesser impact. Global feature importance ranks in the CB model have elicited similar insights: established variables reflect robust evidence for hierarchical effects and reinforce such findings. These provide a variety of routes and dependencies on CH_4 and CO production and could well direct efforts at enhancing the selectivity and efficiency of CO_2 photoreduction via g-C_3N_4/TNTAs composite.

The SHAP summary plots of CH_4 and CO production give further insight into how different experimental parameters influence model predictions for both processes. For CH_4 production, the most influential factor is identified to be the irradiation time (**Fig. 10**c). The high values of the irradiation time, depicted in red, positively influence CH_4 production since it prolongs the interaction of the CO_2 molecules with the catalyst, enhancing its catalytic efficiency. The lower values of the same, coloured in blue, affect production negatively and reduce the overall conversion. The second most critical parameter is the initial concentration of CO_2, whereby higher concentrations in red enhance CH_4 production by ensuring sufficient reaction availability, while lower concentrations presented in blue decrease efficiency. The catalyst surface area also has a significant effect: the larger the active surface area, red, the more it enhances production due to higher catalytic activity, while smaller surface areas blue hinder the reaction. Among all of these, feed pressure and light power were found to be less sensitive, while their higher values also support CH_4 production, though at a lower level. The distribution of SHAP values for CO production appeared in different hierarchy of features (**Fig. 10**d). In that case, catalyst surface area was the most important, and its high value in red signifies enhanced CO production by providing active sites where the reaction can take place; smaller areas in blue showed reduced efficiency. The irradiation time is the second crucial factor; the longer the exposure to irradiation, the more contact and conversion can occur, but the shorter exposure inhibits the reaction. The generation of CO was more significantly impacted by light power than CH_4 because CO is favorably formed by higher light intensities [60–62]. The CO_2 concentration and feed pressure for CO generation exhibited the least impact. The SHAP analysis of relative importance underlines these differences in the parameter dependencies of CH_4 and CO production. Although the irradiation time is most relevant in both prediction models, the higher CO_2 concentration for CH_4 production is outweighed by the dominance of the catalyst surface area in the case of CO production.

From the model-agnostic permutation importance analysis, it can be seen that the key drivers for CH_4 and CO production are quite different (**Fig. S6**). The most influential feature for CH_4 production is irradiation time, since it produces the largest drop in R^2 if shuffled, making it the primary driver of the output. On the other hand, CO_2 concentration and catalyst surface area have a moderate impact, whereas feed pressure and light power contribute very little. In contrast, the dominant factor for CO production is catalyst surface area. Irradiation time remains important but less so than the former. light power contributes moderately, feed pressure minimally, whereas CO_2 concentration has the least effect in view of its minimal role in CO prediction.

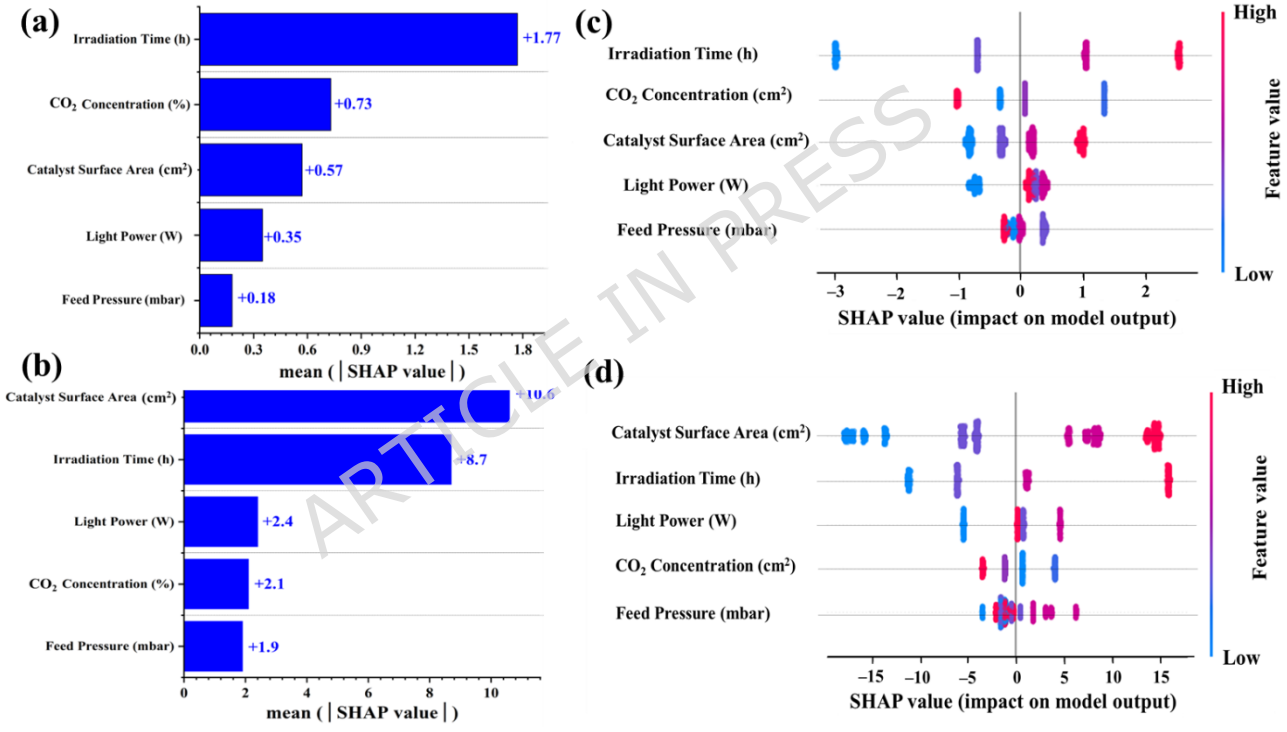


Fig. 10. Global feature importance of CB model using SHAP values (a) The importance rankings for CH_4 prediction and (b) CO prediction; (c) Distribution of SHAP values for CH_4 prediction and (b) CO prediction.

3.3.4. Robustness of the best-predicted model

In the analysis of PC CO_2 conversion to CH_4 and CO, scatter plots and residual plots were utilized to evaluate the predictive accuracy and error

distribution of the models, revealing that the CB model demonstrated superior alignment between predicted and actual values with minimal systematic bias. Radar plots provided a comparative visualization of experimental conditions, highlighting the CB model's ability to consistently outperform other models across diverse parameter spaces. Feature importance charts and SHAP plots further underscored the interpretability of the CB model, identifying key influencing factors such as irradiation time and catalyst surface area, while offering insights into their contributions to the observed outcomes. Collectively, these analyses confirm the CB model as the most robust and reliable choice for the tested datasets in predicting and optimizing CO_2 conversion performance. The performance of the CB model for training and testing datasets was further investigated using box plot, violin plot and Taylor diagram (**Fig. S7**). For CH_4 production, the box plots showed that both the training and testing datasets had a median error close to zero, which could be taken to mean this model's predictions were reasonably accurate and relatively unbiased (**Fig. S7**a). The interquartile range (IQR) is a measure of variability for the central 50 % of data. A small IQR suggests reliable predictions that are mostly stable, with only a handful of outliers present. The violin plots visualized the distribution of errors for the CH_4 production prediction of the CB model. As presented in **Fig. S7**b, the training errors were roughly within the range from –0.1 to 0.2, and the test errors extended a little bit more, within the range from –0.1 to 0.3. The Taylor diagram is supportive, enabling the analysis of performance by comparing model output and observed data for some crucial metrics: STD, R^2, and centered root mean square error. The Taylor diagram showed CH_4 production with excellent proficiency, with a predicted point lying close to the observed one (**Fig. S7**c). The standard deviation (STD) was well predicted for the training dataset with an estimated value of 2.373 compared with the observed one of 2.381 and an almost perfect correlation coefficient of 0.9836. For the test dataset, a similarly high-quality estimation was achieved for STD with a

predicted value of 2.420 in comparison with the observed 2.435 and a corresponding correlation coefficient of 0.9825.

The robustness of the CB model to predict CO generation efficiency was further investigated using box plot, violin plot and Taylor diagram (**Fig. S8**). The CO production errors showed a slightly wider IQR in the testing dataset, with several outliers as presented in **Fig. S8**a. This suggests that the model struggles a bit harder to predict the higher values of CO production. Although overall, the performance of the models is good, the higher variability of CO production errors shows that further refinement is necessary in feature selection methods or addressing data imbalances for improving the accuracy of the predictions. The test errors (–0.75 to 1.25) in violin plots exhibited considerable variability, where the border range indicated higher complexity in the proper forecast of its values (**Fig. S8**b). These plots showed the differences in model performance between the training and testing phases and highlighted the need to work on the variability in error distributions in order to achieve better predictive reliability. These results, when plotted on a Taylor diagram, showed that during training and testing phases the model bears a strong potential to replicate the variability with a strong linear relationship. With a correlation coefficient of 0.9818, the training dataset's predicted STD of 16.504 was nearly identical to the observed one of 16.509, which was practically desirable. The test dataset similarly showed strong alignment, with a high correlation coefficient of 0.9831 supporting the observed STD of 16.521 and the anticipated STD of 16.440. The strong performance indicates the model's accuracy and dependability in predicting photocatalytic CO_2 reduction efficiency.

3.3.5. Optimization of input variables

Table 3 summarizes the optimized operational parameters for photocatalytic CO_2 reduction, alongside corresponding experimental and model-predicted yields of methane (CH_4) and carbon monoxide (CO), thereby

validating the robustness and predictive fidelity of the proposed machine learning driven optimization framework. The identified conditions represent the most favorable configuration for maximizing catalytic efficiency in the context of TiO_2 nanotube array (TNTAs) - based photocatalysis. For CH_4 production, the optimal setup comprises a catalyst exposed surface area of 16 cm^2, an initial CO_2 concentration of 35 vol%, a feed (CO_2 + H_2O) pressure of 270 mbar, a light irradiance of 190 W, and an irradiation duration of 4 hours. Under these conditions, pristine TNTAs yielded an experimental CH_4 production rate of 1.47 µmol/cm^2 with a selectivity of 15.28%. In stark contrast, the g-C_3N_4-modified TNTAs exhibited a markedly enhanced CH_4 output of 12.15 µmol/cm^2, representing an approximately 8.25-fold increase relative to the unmodified catalyst. This substantial improvement is attributed to the favorable electronic and structural modifications induced by g-C_3N_4 integration. Critically, this experimental value aligns almost perfectly with the model's prediction of 12.17 µmol/cm^2, differing by a mere 0.02 µmol/cm^2, thereby underscoring the model's exceptional accuracy in forecasting photocatalytic performance under optimized conditions. The enhanced activity of the g-C_3N_4/TNTAs heterostructure can be rationalized through the formation of a Type-II (staggered) band alignment at the interface. In this configuration, photogenerated electrons in the conduction band (CB) of g-C_3N_4 migrate to the CB of TiO_2, while holes in the valence band (VB) of TiO_2 transfer to the VB of g-C_3N_4 (as described in **Eqs. 3** and **4**). This spatial charge separation effectively suppresses electron-hole recombination, a key bottleneck in photocatalysis, thereby significantly augmenting quantum efficiency and overall reaction kinetics.

For CO production, the optimization algorithm recommended a distinct set of conditions: a catalyst surface area of 16 cm^2, CO_2 concentration of 20 vol%, feed pressure of 280 mbar, light power of 270 W, and irradiation time of 4 hours. Experimental validation under these parameters yielded CO production rates of 44.32 µmol/cm^2 for bare TNTAs and 118.50 µmol/cm^2 for

g-C_3N_4/TNTAs, with the latter achieving a high selectivity of 84.72% toward CO. The close agreement between these empirical results and the model-predicted values, evidenced by negligible deviations, further corroborates the framework's capacity to accurately simulate and anticipate catalytic behavior across diverse product channels. Quantitatively, the final predictive models exhibited outstanding statistical performance. For CH_4, the model achieved an R^2 of 0.9864, MSE of 0.0033, RMSE of 0.0578, MAE of 0.0427, and MedAE of 0.0341. Similarly, for CO prediction, the metrics were R^2 = 0.9758, MSE = 0.0541, RMSE = 0.2326, MAE = 0.1742, and MedAE = 0.1424. These results collectively affirm that the Bayesian-optimized machine learning model not only captures the complex, nonlinear relationships governing photocatalytic CO_2 conversion but also provides a highly reliable tool for process design, performance forecasting, and catalyst evaluation in solar fuel synthesis applications.

$$e_{CB}^{-}\ (g\text{-}C_3N_4) \rightarrow e_{CB}^{-}\ (TiO_2) \quad (3)$$

$$h_{VB}^{+}\ (TiO_2) \rightarrow h_{VB}^{+}\ (g\text{-}C_3N_4) \quad (4)$$

The differential evolution (DE) optimization was explicitly constrained to the experimental range for all features; no extrapolation beyond the tested values was allowed. **Table S8** summarizes the withheld test performance metrics of CH_4 and CO, obtained using the CB model, which showed the best predictive performance among all tested algorithms. Since new laboratory experiments could not be conducted, the model's reliability was established by testing its performance with a withheld test set of 205 samples that were excluded both during training and tuning of hyperparameters, ensuring true out-of-sample performance of the prediction. Therefore, this model has very high predictive accuracy, as reflected in the R^2 value of 0.9827 for CH_4 and 0.9903 for CO, suggesting almost perfect agreement between the predicted and withheld experimental values. Low RMSE values of 0.0427 and 0.0567 for CH_4 and CO, respectively, along with negligible MAE and MedAE, further

establish the model as reliable. From the results, it emerges that the CatBoost model can predict photocatalytic product yields with high accuracy under novel experimental conditions.

The Pareto plots in **Figures S9 and 10** were used to find optimal experimental conditions by visualizing Pareto fronts for multi-objective optimization. These plots show that the high CO or CH_4 production in large measure occurs at or near boundaries of the key experimental variables, with respect to catalyst exposed surface area, CO_2 concentration, and irradiation time. They result from a multi-objective optimization that, in parallel, was performed to maximize product yields (CH_4 or CO) at minimum light power and thus defines conditions balancing high photocatalytic efficiency with energy consumption. Such Pareto fronts give, for each product, solutions that yield the best compromise between maximum production and minimum light intensity of the corresponding product, showing that the conditions selected using DE-optimization correspond to true or near-optimum solutions for both target products. This not only confirms the reliability of the CatBoost-predicted yields but also yields practical guidelines for experiments under energy-efficient conditions.

As part of the uncertainty assessment, the reliability of the highest CO and CH_4 production values was interrogated with respect to the five nearest experimental neighbors in **Table S9**. For CO, the distance-weighted mean and the very minimal standard deviation (±3.32%) indicate that the surrounding experiments display minimal variation in CO production, suggesting that the peak value of CO is both locally consistent and supportive of the dataset's underlying trends. In contrast, the uncertainty for CH_4 is relatively higher (±8.08%), reflecting greater variability among the proximal experiments and suggesting that the maximum CH_4 point is somewhat less stable in its local experimental neighborhood. Using five neighbors increases the robustness of the analysis, as it captures a more representative and reliable snapshot of the local data distribution. Overall, these findings show

that uncertainty quantification provides a stringent and data-driven means to assess the credibility of the maximum production values measured.

These findings highlight the effectiveness of the optimization framework and support its ability to lessen trial-and-error in experimental setups. The framework will enable the exploration of optimal reaction parameters by accurately predicting production values, thereby conserving resources and saving time. The high accuracy achieved for various reactions, such as CH_4 and CO, underscores the framework's adaptability and utility across a broad range of catalytic processes. This advancement is crucial for developing greener and economically feasible methods for chemical production, positioning this framework as a valuable asset in the field of catalysis, especially for the photocatalytic conversion process utilizing TNTA's.

Table 3. Optimized input variables to achieve maximum products.

Catalysts	Products	Catalyst surface area (cm^2)	CO_2 concentration (%)	Feed Pressure (mbar)	Light power (W)	Irradiation time (h)	Predicted production (µmol/cm²)	Experimental production (µmol/cm^2)
TNTAs	CH_4	16	35	270	190	4	1.23	1.47
	CO	16	20	280	270	4	46.25	44.32
g-C_3N_4/TNTAs	CH_4	16	35	270	190	4	12.17	12.15
	CO	16	20	280	270	4	118.72	118.50

4. Conclusion

The gas-phase photocatalytic (PC) reduction of CO_2 into value-added fuels, specifically methane (CH_4) and carbon monoxide (CO), using graphitic carbon nitride (g-C_3N_4)-modified TiO_2 nanotube arrays (TNTAs) constitutes a

pivotal advancement toward sustainable energy generation and atmospheric CO_2 mitigation. This study demonstrates the successful integration of tree-based machine learning (ML) methodologies to not only predict but also strategically optimize the yields of CH_4 and CO in a gas-phase photocatalytic system. Comprehensive model evaluation, grounded in a suite of regression performance indicators including the R^2, MSE, RMSE, MAE, and MedAE-unequivocally established CatBoost (CB) as the superior predictive algorithm across all assessed metrics. Specifically, the CatBoost model exhibited exceptional performance in forecasting CH_4 production, achieving R^2 scores of 0.9887 on the training set and 0.9883 on the testing set. Similarly, for CO prediction, it attained R^2 values of 0.9885 (training) and 0.9874 (testing), underscoring its high fidelity, minimal bias, and robust generalization capability. The consistency between two complementary interpretability techniques, normalized feature importance and SHAP values, further reinforces the credibility of the model's insights and provides a coherent, physically interpretable understanding of how input variables jointly influence photocatalytic outcomes. According to the optimized CatBoost model, the relative influence of operational parameters on product formation follows a clear hierarchy: catalyst exposed surface area emerged as the most dominant factor, followed by irradiation time, light power, initial CO_2 concentration, and, to a significantly lesser extent, feed pressure. This ranking aligns well with established photocatalytic principles, where surface area governs active site availability, irradiation duration dictates charge carrier utilization, and light intensity modulates excitation efficiency, thereby validating both the model's internal logic and the underlying reaction mechanisms. These data-driven insights offer actionable guidance for experimental design, enabling researchers to rationally engineer photocatalytic systems with enhanced efficiency for CO_2-to-fuel conversion. By leveraging ML-guided optimization, the need for labor-intensive, iterative trial-and-error experimentation is substantially reduced, leading to significant savings in time, resources, and material costs. Moreover, the high

predictive accuracy and interpretability of the CatBoost framework facilitate the rapid identification of optimal operational windows, accelerating the development of next-generation photocatalysts. Critically, the study also highlights the exceptional promise of g-C_3N_4/TNTAs heterostructures as high-performance platforms for solar-driven CO_2 valorization, capable of achieving markedly elevated production rates of CH_4 and CO under optimized conditions. Taken together, these findings underscore the transformative potential of integrating advanced machine learning with materials science to advance sustainable photocatalytic technologies for renewable fuel synthesis and carbon neutrality.

CRediT authorship contribution statement

Md. Arif Hossen: Conceptualization, Methodology, Data collection, Data analysis, Writing–original draft, Writing – review & editing. **Meherunnesa Prima:** Data curation, Methodology, Software, Visualization, Writing – review & drafting. **Azrina Abd Aziz**: Funding acquisition, Project administration, Writing– review & editing. **Nadeem A Khan**: Conceptualization, Software, Writing – review & editing. **Yunus Ahmed**: Conceptualization, Software, Supervision, Visualization, Writing – review & editing.

Conflicts of interest

The authors affirm that none of the findings in this research may have been impacted by any known competing financial interests or personal relationships.

Data availability

Data will be made available on request to corresponding authors.

Acknowledgment

The authors highly acknowledged the support from the Chittagong University of Engineering and Technology (CUET), Bangladesh and Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA), Malaysia.

Funding

No funding information is available for this research.

References

[1] A. Crake, K.C. Christoforidis, R. Godin, B. Moss, A. Kafizas, S. Zafeiratos, J.R. Durrant, C. Petit, Titanium dioxide/carbon nitride nanosheet nanocomposites for gas phase CO_2 photoreduction under UV-visible irradiation, Appl. Catal. B Environ. 242 (2019) 369–378. https://doi.org/10.1016/j.apcatb.2018.10.023.

[2] Q. Xu, Z. Xia, J. Zhang, Z. Wei, Q. Guo, H. Jin, H. Tang, S. Li, X. Pan, Z. Su, S. Wang, Recent advances in solar-driven CO_2 reduction over g-C_3N_4-based photocatalysts, Carbon Energy 5 (2023). https://doi.org/10.1002/cey2.205.

[3] L.C. Sim, P.L. Yee, K.H. Leong, A.A. Aziz, M.A. Hossen, CO_2 photoreduction to hydrocarbons and oxygenated hydrocarbons, INC, 2024. https://doi.org/10.1016/B978-0-443-19235-7.00017-8.

[4] L. Liu, S. Wang, H. Huang, Y. Zhang, T. Ma, Surface sites engineering on semiconductors to boost photocatalytic CO_2 reduction, Nano Energy 75 (2020) 104959. https://doi.org/10.1016/j.nanoen.2020.104959.

[5] Y. Lu, X. Li, X. Jing, Y. Huang, Y. Qiang, G. Cao, R. Qin, Q. Cao, J. Cao, S.C. Lee, Advances and roles of oxygen vacancies in semiconductor photocatalysts for solar-driven CO_2 reduction, Surfaces and Interfaces 53 (2024) 104957. https://doi.org/10.1016/j.surfin.2024.104957.

[6] M.A. Hossen, H.M. Solayman, K.H. Leong, L.C. Sim, N. Yaacof, A. Abd Aziz, L. Wu, M.U. Monir, Recent progress in TiO_2-Based photocatalysts for conversion of CO_2 to hydrocarbon fuels: A systematic review, Results Eng. 16 (2022) 100795. https://doi.org/10.1016/j.rineng.2022.100795.

[7] N. Farooq, P. Kallem, Z. ur Rehman, M. Imran Khan, R. Kumar Gupta, T. Tahseen, Z. Mushtaq, N. Ejaz, A. Shanableh, Recent trends of titania (TiO_2) based materials: A review on synthetic approaches and potential applications, J. King Saud Univ. - Sci. 36 (2024) 103210.

https://doi.org/10.1016/j.jksus.2024.103210.

[8] M.A. Hossen, H.M. Solayman, K.H. Leong, L.C. Sim, N. Yaacof, A. Abd Aziz, W. Lihua, M.U. Monir, A Comprehensive Review on Advances in TiO_2 Nanotube (TNT)-Based Photocatalytic CO_2 Reduction to Value-Added Products, Energies 15 (2022). https://doi.org/10.3390/en15228751.

[9] H. Yao, R. Zhang, Y. Wen, Y. Liu, G. Yu, Z.H. Xie, Metal-organic framework [NH_2-MIL-53(Al)] functionalized TiO_2 nanotube photoanodes for highly stable and efficient photoelectrochemical cathodic protection of nickel-coated Mg alloy, J. Mater. Sci. Technol. 182 (2024) 67–78. https://doi.org/10.1016/j.jmst.2023.09.038.

[10] R.R. Ikreedeegh, M.A. Hossen, M. Tahir, A.A. Aziz, A comprehensive review on anodic TiO_2 nanotube arrays (TNTAs) and their composite photocatalysts for environmental and energy applications: Fundamentals, recent advances and applications, Coord. Chem. Rev. 499 (2024) 215495. https://doi.org/10.1016/j.ccr.2023.215495.

[11] M.A. Hossen, A. Abd Aziz, R.R. Ikreedeegh, A. Din Muhammad, N. Yaacof, K.H. Leong, L. Wu, Optimization of anodizing parameters for the morphological properties of TiO_2 nanotubes based on response surface methodology, Next Mater. 2 (2024) 100061. https://doi.org/10.1016/j.nxmate.2023.100061.

[12] W. Li, M. Sohail, U. Anwar, T.A. Taha, A.G. Al-Sehemi, S. Muhammad, A.A. Al-Ghamdi, M.A. Amin, A. Palamanit, S. Ullah, A. Hayat, Z. Ajmal, Recent progress in g-C_3N_4–Based materials for remarkable photocatalytic sustainable energy, Int. J. Hydrogen Energy 47 (2022) 21067–21118. https://doi.org/10.1016/j.ijhydene.2022.04.247.

[13] K.M. Alam, N. Chaulagain, E. Shahini, M. Masud Rana, J. Garcia, N. Kumar, A.E. Kobryn, S. Gusarov, T. Tang, K. Shankar, Low bandgap carbon nitride nanoparticles incorporated in titania nanotube arrays by in situ electrophoretic anodization for photocatalytic CO_2 reduction, Chem. Eng. J. 456 (2023) 141067. https://doi.org/10.1016/j.cej.2022.141067.

[14] M.A. Hossen, R.R. Ikreedeegh, A.A. Aziz, A.Y. Zerga, M. Tahir, Carbon-based nanomaterials (CNMs) modified TiO_2 nanotubes (TNTs) photo-driven catalysts for sustainable energy and environmental applications: A comprehensive review, J. Environ. Chem. Eng. 12 (2024) 114088. https://doi.org/10.1016/j.jece.2024.114088.

[15] Y. Liu, Q. Ge, T. Wang, K. Li, Y. Deng, W. You, L. Xie, L. Zhang, Investigating the impact of pretreatment strategies on photocatalyst for accurate CO_2RR productivity quantification: A machine learning approach, Chem. Eng. J. 473 (2023) 145255. https://doi.org/10.1016/j.cej.2023.145255.

[16] C.M. Kim, Z.H. Jaffari, A. Abbas, M.F. Chowdhury, K.H. Cho, Machine learning analysis to interpret the effect of the photocatalytic reaction rate constant (k) of semiconductor-based photocatalysts on dye removal, J. Hazard. Mater. 465 (2024) 132995. https://doi.org/10.1016/j.jhazmat.2023.132995.

[17] H.N. Nguyen, Q.T. Tran, C.T. Ngo, D.D. Nguyen, V.Q. Tran, Solar energy prediction through machine learning models: A comparative analysis of regressor algorithms, PLoS One 20 (2025) 1–23. https://doi.org/10.1371/journal.pone.0315955.

[18] Y. Zhang, X. Xu, Machine Learning Band Gaps of Doped-TiO_2 Photocatalysts from Structural and Morphological Parameters, ACS Omega 5 (2020) 15344–15352. https://doi.org/10.1021/acsomega.0c01438.

[19] A. Mikolajczyk, E. Wyrzykowska, P. Mazierski, T. Grzyb, Z. Wei, E. Kowalska, P.N.A. Caicedo, A. Zaleska-Medynska, T. Puzyn, J. Nadolna, Visible-light photocatalytic activity of rare-earth-metal-doped TiO_2: Experimental analysis and machine learning for virtual design, Appl. Catal. B Environ. 346 (2024) 123744. https://doi.org/10.1016/j.apcatb.2024.123744.

[20] X. Li, P.M. Maffettone, Y. Che, T. Liu, L. Chen, A.I. Cooper, Combining machine learning and high-throughput experimentation to discover photocatalytically active organic molecules, Chem. Sci. 12 (2021) 10742–10754. https://doi.org/10.1039/d1sc02150h.

[21] Q. Tao, T. Lu, Y. Sheng, L. Li, W. Lu, M. Li, Machine learning aided design of perovskite oxide materials for photocatalytic water splitting, J. Energy Chem. 60 (2021) 351–359. https://doi.org/10.1016/j.jechem.2021.01.035.

[22] V.Y. Yurova, K.O. Potapenko, T.A. Aliev, E.A. Kozlova, E. V. Skorb, Optimization of g-C_3N_4 synthesis parameters based on machine learning to predict the efficiency of photocatalytic hydrogen production, Int. J. Hydrogen Energy 81 (2024) 193–203. https://doi.org/10.1016/j.ijhydene.2024.07.245.

[23] A. Jose, E. Devijver, N. Jakse, R. Poloni, Informative Training Data for Efficient Property Prediction in Metal-Organic Frameworks by Active Learning, J. Am. Chem. Soc. 146 (2024) 6134–6144. https://doi.org/10.1021/jacs.3c13687.

[24] Q. Liu, K. Pan, Y. Lu, W. Wei, S. Wang, W. Du, Z. Ding, Y. Zhou, Data-driven for accelerated design strategy of photocatalytic degradation activity prediction of doped TiO_2 photocatalyst, J. Water Process Eng. 49 (2022) 103126. https://doi.org/10.1016/j.jwpe.2022.103126.

[25] Y. Ahmed, K.R. Dutta, S.N.C. Nepu, M. Prima, H. AlMohamadi, P. Akhtar, Optimizing photocatalytic dye degradation: A machine learning

and metaheuristic approach for predicting methylene blue in contaminated water, Results Eng. 25 (2025) 103538. https://doi.org/10.1016/j.rineng.2024.103538.

[26] I. Salahshoori, M. Namayandeh Jorabchi, A. Baghban, H.A. Khonakdar, Integrative analysis of multi machine learning models for tetracycline photocatalytic degradation with MOFs in wastewater treatment, Chemosphere 350 (2024) 141010. https://doi.org/10.1016/j.chemosphere.2023.141010.

[27] M.F. Javed, M.Z. Shahab, U. Asif, T. Najeh, F. Aslam, M. Ali, I. Khan, Evaluation of machine learning models for predicting TiO_2 photocatalytic degradation of air contaminants, Sci. Rep. 14 (2024) 1–25. https://doi.org/10.1038/s41598-024-64486-7.

[28] J. Zhou, Z. Wu, C. Jin, J.X.J. Zhang, Machine learning assisted dual-functional nanophotonic sensor for organic pollutant detection and degradation in water, Npj Clean Water 7 (2024). https://doi.org/10.1038/s41545-023-00292-4.

[29] R. Mohammadzadeh kakhki, M. Mohammadpoor, Machine learning-driven approaches for synthesizing carbon dots and their applications in photoelectrochemical sensors, Inorg. Chem. Commun. 159 (2024) 111859. https://doi.org/10.1016/j.inoche.2023.111859.

[30] Y. Haghshenas, W.P. Wong, D. Gunawan, A. Khataee, R. Keyikoğlu, A. Razmjou, P.V. Kumar, C.Y. Toe, H. Masood, R. Amal, V. Sethu, W.Y. Teoh, Predicting the rates of photocatalytic hydrogen evolution over cocatalyst-deposited TiO_2 using machine learning with active photon flux as a unifying feature, EES Catal. 2 (2023) 612–623. https://doi.org/10.1039/d3ey00246b.

[31] D. Saadetnejad, B. Oral, E. Can, R. Yıldırım, Machine learning analysis of gas phase photocatalytic CO_2 reduction for hydrogen production, Int. J. Hydrogen Energy 47 (2022) 19655–19668. https://doi.org/10.1016/j.ijhydene.2022.02.030.

[32] S. Özsoysal, B. Oral, R. Yıldırım, Analysis of photocatalytic CO_2 reduction over MOFs using machine learning, J. Mater. Chem. A 12 (2024) 5748–5759. https://doi.org/10.1039/d3ta07001h.

[33] C. Zhang, J. Liu, X. Huang, D. Chen, S. Xu, Multistage Polymerization Design for g-C_3N_4 Nanosheets with Enhanced Photocatalytic Activity by Modifying the Polymerization Process of Melamine, ACS Omega 4 (2019) 17148–17159. https://doi.org/10.1021/acsomega.9b01510.

[34] R.R. Ikreedeegh, M.A. Hossen, M. Tahir, Noble-Metal-Free Modified TiO_2 Nanotube Arrays (TNTAs) for Efficient Photocatalytic Reduction of CO_2 to CO Under Visible Light, ChemistrySelect 9 (2024) 1–11. https://doi.org/10.1002/slct.202403536.

[35] Y. Shan, Q. Wu, H. Yuan, W. Liu, Develop machine learning-based model

and automated process for predicting liquid heat capacity of organics at different temperatures, Fluid Phase Equilib. 584 (2024) 114132. https://doi.org/10.1016/j.fluid.2024.114132.

[36] A. Hosseinzadeh, J.L. Zhou, J. Zyaie, N. AlZainati, I. Ibrar, A. Altaee, Machine learning-based modeling and analysis of PFOS removal from contaminated water by nanofiltration process, Sep. Purif. Technol. 289 (2022) 120775. https://doi.org/10.1016/j.seppur.2022.120775.

[37] A. Casas, D. Rodríguez-Llorente, G. Rodríguez-Llorente, J. García, M. Larriba, Machine learning screening tools for the prediction of extraction yields of pharmaceutical compounds from wastewaters, J. Water Process Eng. 62 (2024) 105379. https://doi.org/10.1016/j.jwpe.2024.105379.

[38] L. Yang, A. Shami, On hyperparameter optimization of machine learning algorithms: Theory and practice, Neurocomputing 415 (2020) 295–316. https://doi.org/10.1016/j.neucom.2020.07.061.

[39] S. Hanifi, A. Cammarono, H. Zare-Behtash, Advanced hyperparameter optimization of deep learning models for wind power prediction, Renew. Energy 221 (2024) 119700. https://doi.org/10.1016/j.renene.2023.119700.

[40] J. Li, X. Liu, H. Wang, Y. Sun, F. Dong, Prediction and interpretation of photocatalytic NO removal on g-C_3N_4-based catalysts using machine learning, Chinese Chem. Lett. 35 (2024) 108596. https://doi.org/10.1016/j.cclet.2023.108596.

[41] Z. Wu, J. Luo, D. Rincon, P.D. Christofides, Machine learning-based predictive control using noisy data: evaluating performance and robustness via a large-scale process simulator, Chem. Eng. Res. Des. 168 (2021) 275-287. https://doi.org/10.1016/j.cherd.2021.02.011.

[42] C. Adán, J. Marugán, E. Sánchez, C. Pablos, R. Van Grieken, Understanding the effect of morphology on the photocatalytic activity of TiO_2 nanotube array electrodes, Electrochim. Acta 191 (2016) 521–529. https://doi.org/10.1016/j.electacta.2016.01.088.

[43] P. Bamola, S. Rawat, C. Dwivedi, M. Sharma, B. Singh, H. Sharma, Effect of nanotube diameter on the photocatalytic activity of bimetallic AgAu nanoparticles grafted 1D-TiO_2 nanotubes, J. Mater. Sci. Mater. Electron. 32 (2021) 1427–1444. https://doi.org/10.1007/s10854-020-04914-2.

[44] M.A. Hossen, F. Khatun, R.R. Ikreedeegh, A.D. Muhammad, A. Abd Aziz, K.H. Leong, L.C. Sim, W. Lihua, M.U. Monir, Enhanced Photocatalytic CO_2 Reduction to CH_4 Using Novel Ternary Photocatalyst RGO/Au-TNTAs, Energies 16 (2023). https://doi.org/10.3390/en16145404.

[45] R.R. Ikreedeegh, M. Tahir, Photocatalytic CO_2 reduction to CO and CH_4 using g-C3N4/RGO on titania nanotube arrays (TNTAs), J. Mater. Sci. 56

(2021) 18989–19014. https://doi.org/10.1007/s10853-021-06516-7.

[46] I. Arora, H. Chawla, A. Chandra, S. Sagadevan, S. Garg, Advances in the strategies for enhancing the photocatalytic activity of TiO_2: Conversion from UV-light active to visible-light active photocatalyst, Inorg. Chem. Commun. 143 (2022) 109700. https://doi.org/10.1016/j.inoche.2022.109700.

[47] T. Ishaq, Z. Ehsan, A. Qayyum, Y. Abbas, A. Irfan, S.A. Al-Hussain, M.A. Irshad, M.E.A. Zaki, Recent Strategies to Improve the Photocatalytic Efficiency of TiO_2 for Enhanced Water Splitting to Produce Hydrogen, Catalysts 14 (2024). https://doi.org/10.3390/catal14100674.

[48] L.C. Sim, K.S. Koh, K.H. Leong, Y.H. Chin, A.A. Aziz, P. Saravanan, In situ growth of g-C_3N_4 on TiO_2 nanotube arrays: Construction of heterostructures for improved photocatalysis properties, J. Environ. Chem. Eng. 8 (2020). https://doi.org/10.1016/j.jece.2019.103611.

[49] F. Zhang, J. Liu, H. Yue, G. Cheng, X. Xue, Construction of g-C_3N_4 nanoparticles modified TiO_2 nanotube arrays with Z-scheme heterojunction for enhanced photoelectrochemical properties, J. Mater. Sci. 58 (2023) 2676–2688. https://doi.org/10.1007/s10853-022-07730-7.

[50] S. Xin, X. Ma, J. Lu, G. Zhang, S. Huo, M. Gao, P. Xu, W. Liu, W. Fu, Enhanced visible light photoelectrocatalytic degradation of o-chloronitrobenzene through surface plasmonic Au nanoparticles and g-C_3N_4 co-modified TiO_2 nanotube arrays photoanode, Appl. Catal. B Environ. 323 (2023) 122174. https://doi.org/10.1016/j.apcatb.2022.122174.

[51] M.A. Hossen, M. Hasan, Y. Ahmed, A. Abd, N. Yaacof, K. Hon, Experimental and AI-driven enhancements in gas-phase photocatalytic CO_2 conversion over synthesized highly ordered anodic TiO_2 nanotubes, Energy Convers. Manag. 327 (2025) 119544. https://doi.org/10.1016/j.enconman.2025.119544.

[52] Y. Ahmed, A.A. Siddiqua Maya, P. Akhtar, M.S. Alam, H. AlMohamadi, M.N. Islam, O.A. Alharbi, S.M. Rahman, A novel interpretable machine learning and metaheuristic-based protocol to predict and optimize ciprofloxacin antibiotic adsorption with nano-adsorbent, J. Environ. Manage. 370 (2024) 122614. https://doi.org/10.1016/j.jenvman.2024.122614.

[53] A.S. Albahri, A.M. Duhaim, M.A. Fadhel, A. Alnoor, N.S. Baqer, L. Alzubaidi, O.S. Albahri, A.H. Alamoodi, J. Bai, A. Salhi, J. Santamaría, C. Ouyang, A. Gupta, Y. Gu, M. Deveci, A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion, Inf. Fusion 96 (2023) 156–191. https://doi.org/10.1016/j.inffus.2023.03.008.

[54] J.T. Hancock, T.M. Khoshgoftaar, CatBoost for big data: an

interdisciplinary review, J. Big Data 7 (2020). https://doi.org/10.1186/s40537-020-00369-8.

[55] M.F. Javed, M. Fawad, R. Lodhi, T. Najeh, Y. Gamil, Forecasting the strength of preplaced aggregate concrete using interpretable machine learning approaches, Sci. Rep. 14 (2024) 1–28. https://doi.org/10.1038/s41598-024-57896-0.

[56] S. Subba, S. Chatterjee, Machine learning-driven determination of key absorber layer parameters in perovskite solar cells, Mater. Today Commun. 42 (2025) 111113. https://doi.org/10.1016/j.mtcomm.2024.111113.

[57] S. Delavari, N.A.S. Amin, M. Ghaedi, Photocatalytic conversion and kinetic study of CO and CH_4 over nitrogen-doped titania nanotube arrays, J. Clean. Prod. 111 (2016) 143–154. https://doi.org/10.1016/j.jclepro.2015.07.077.

[58] H. Li, Y. Gao, Z. Xiong, C. Liao, K. Shih, Enhanced selective photocatalytic reduction of CO_2 to CH_4 over plasmonic Au modified g-C_3N_4 photocatalyst under UV–vis light irradiation, Appl. Surf. Sci. 439 (2018) 552–559. https://doi.org/10.1016/j.apsusc.2018.01.071.

[59] R.R. Ikreedeegh, M. Tahir, Facile fabrication of well-designed 2D/2D porous g-C_3N_4–GO nanocomposite for photocatalytic methane reforming (DRM) with CO_2 towards enhanced syngas production under visible light, Fuel 305 (2021) 121558. https://doi.org/10.1016/j.fuel.2021.121558.

[60] J. Fu, K. Jiang, X. Qiu, J. Yu, M. Liu, Product selectivity of photocatalytic CO_2 reduction reactions, Mater. Today 32 (2020) 222–243. https://doi.org/10.1016/j.mattod.2019.06.009.

[61] D.P. Kumar, K.H. Do, A.P. Rangappa, J. Lee, J. Wang, R. Boppella, M. Gopannagari, K.A.J. Reddy, D.A. Reddy, T.K. Kim, Highly stable and durable $ZnIn_2S_4$ nanosheets wrapped oxygen deficient blue TiO_2(B) catalyst for selective CO_2 photoreduction into CO and CH_4, J. Colloid Interface Sci. 651 (2023) 264–272. https://doi.org/10.1016/j.jcis.2023.07.197.

[62] Y. Zhang, R. Huang, Y. Fang, J. Wang, Z. Yuan, X. Chen, W. Zhu, Y. Cai, X. Shi, Modulation of Fe–MOF via second-transition metal ion doping (Ti, Mn, Zn, Cu) for efficient visible-light driven CO_2 reduction to CH_4, Sep. Purif. Technol. 336 (2024) 126164. https://doi.org/10.1016/j.seppur.2023.126164.