Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Artificial Neural Network–Driven Optimization of Photocatalytic Glycerol Reforming for Sustainable Hydrogen Production

Wei Lin[1], Yixian Li[1], Chien Yong Goh[2,*], Sin Yuan Lai[1,*]

[1] School of Energy and Chemical Engineering, Xiamen University Malaysia, Jalan Sunsuria, Bandar Sunsuria, Selangor Darul Ehsan, 43900 Sepang, Selangor, Malaysia
[2] Department of Mathematics, Xiamen University Malaysia, Jalan Sunsuria, Bandar Sunsuria, 43900 Sepang, Selangor, Malaysia

ARTICLE INFO

Article history:
Received 27 March 2026
Received in revised form 21 May 2026
Accepted 2 June 2026
Available online 22 June 2026

ABSTRACT

The increasing production of biodiesel in Malaysia has led to a surplus of glycerol by-products, posing environmental and resource management challenges. Photocatalytic reforming of glycerol using TiO_2-based photocatalysts presents a promising pathway for sustainable hydrogen generation. However, process optimization is complicated by nonlinear interactions among light intensity, catalyst dosage, glycerol concentration, and reaction duration. This study integrates Artificial Neural Network (ANN), Genetic Algorithm (GA) optimization, and Garson sensitivity analysis to systematically model and optimize hydrogen production under ultraviolet (UV), visible, and UV-visible irradiation. Among the ANN models evaluated, Bayesian Regularization (BR) achieved the highest predictive accuracy under UV light (R^2=0.9629), Levenberg-Marquardt (LM) performed best under visible light (R^2 = 0.9194), and Scaled Conjugate Gradient (SCG) excelled under UV-visible light (R^2 = 0.9245), all with low mean square errors. GA optimization revealed light-dependent optimal conditions, yielding a maximum hydrogen reforming rate of 3.33×10^5 μmol·g^{-1}·h^{-1} under UV irradiation, while visible and UV–visible conditions produced lower but optimized rates of 3.80×10^3 and 3.40×10^3 μmol·g^{-1}·h^{-1}, respectively. Correlation and Garson analyses identified the light power and glycerol concentration dominate under UV irradiation, whereas catalyst loading becomes the main factor under visible and UV-visible light. The results show that specific optimization under different light is essential for maximizing hydrogen production efficiency. The ANN-GA-Garson framework is a data-driven strategy for supporting the waste-to-wealth concept, whereby transforming the biodiesel-derived waste to sustainable hydrogen production.

Keywords:

Process optimization; artificial neural networks; glycerol photoreforming; H_2 production; genetic algorithm

1. Introduction

The global biodiesel industry is a big player in renewable energy, making more than 50 billion gallons of biodiesel every year. But as it grows quickly, crude glycerol, a byproduct that creates around 0.1 to 0.2 kilograms for every liter of biodiesel, becomes a huge problem [1]. Each year, roughly 4–5 million metric tons of this waste glycerol pile up, which is about three times higher than

* *Corresponding author.*
E-mail address: sinyuan.lai@xmu.edu.my; gcyong@xmu.edu.my

https://doi.org/10.37934/araset.60.5.481520

what the market actually needs. Traditionally, crude glycerol has been seen as industrial waste due to impurities like methanol, soaps, and fatty acids, requiring expensive and energy-intensive processes such as distillation and pH neutralization to purify it for pharmaceutical use [2]. Consequently, about 60% of crude glycerol worldwide is either burned, placed in landfills, or discharged into waterways, contributing to environmental problems such as soil acidification, decreased oxygen levels in aquatic ecosystems, and annual CO_2 emissions of approximately 12 million metric tons [3].

Glycerol photoreforming offers a viable and sustainable solution to this issue. Unlike traditional water electrolysis that requires ultrapure water and external voltages greater than 1.23 V, photoreforming utilizes sunlight to concurrently transform crude glycerol into valuable hydrogen fuel [4]. The photoreforming reaction can be expressed as Eq. (1):

$$C_3H_8O_3 + 3H_2O\ (l) \rightarrow 7H_2 + 3CO_2 \quad (1)$$

Photoreforming uses glycerol's α-hydrogen atoms as electron donors, which makes it 30–50% more thermodynamically efficient than typical water splitting. This removes the oxygen evolution reaction (OER), which uses a lot of energy [5]. Recent improvements in photocatalyst materials, such as platinum-coated titanium dioxide nanotubes and Z-scheme MoS_2/g-C_3N_4 heterojunctions, have led to hydrogen evolution rates of up to 23.4 $mmol \cdot g^{-1} \cdot h^{-1}$ under normal solar light. [6]. Photoreforming efficiently converts unrefined glycerol (with 60–80% purity), skipping the costly step of purification. Interestingly, typical impurities in crude glycerol, such as 15–20% methanol, can even boost hydrogen production by about 18%, making the process even more beneficial. Besides turning waste glycerol into useful energy, photoreforming offers significant environmental advantages. For every metric ton of crude glycerol processed, around 120 kg of hydrogen can be produced, providing about 4,000 kWh of energy. At the same time, it can cut CO_2 emissions by roughly 2.8 metric tons annually per biodiesel production facility, enhancing sustainability and reducing environmental impact [7].

Photoreforming uses different photocatalysts to make hydrogen from glycerol when light is shone on it. Some common photocatalysts are titanium dioxide (TiO_2), cadmium sulfide (CdS), zinc oxide (ZnO), graphitic carbon nitride (g-C_3N_4), and bismuth-based compounds as $BiVO_4$ and Bi_2WO_6. Choosing the right materials is very important for getting the best results from photoreforming because they all have different light absorption, charge separation, and stability under reaction circumstances. TiO_2 is one of the most common photocatalysts since it can oxidize things well, is quite stable, and doesn't cost much. The band gap is just right (approximately 3.2 eV for anatase), and it can easily make electron-hole pairs when exposed to ultraviolet (UV) radiation. Also, TiO_2 is safe to use and doesn't change chemically when it's in water, which makes it a great candidate for long-term hydrogen production. But because it has a big band gap, it can only absorb UV light, which is only a small fraction of the solar spectrum. To get over this problem, other methods have been tried, like adding metal or nonmetal components, linking with plasmonic nanoparticles, and making heterojunctions with materials like g-C_3N_4 or MoS_2 to improve their activity in visible light.

The biodiesel industry generates large quantities of glycerol, accounting for 10-20% of its production. This surplus has led to innovative sustainable applications, such as glycerol photoreforming, which utilizes sunlight to produce hydrogen more efficiently and eco-friendly compared to traditional water splitting. This method leverages glycerol as a sacrificial agent to enhance hydrogen production while avoiding high-energy processes. TiO_2-based photocatalysts are favored for their non-toxicity, affordability, and durability, making them ideal for sustainable hydrogen production from biodiesel by-products. Photoreforming glycerol is an approach for converting biodiesel by-products into hydrogen, which is beneficial to the environment. This method reduces waste and generates clean energy. However, employing traditional experimental techniques

to achieve optimal photocatalytic performance is time-consuming, costly, and labor-intensive. This frequently entails multiple trial-and-error experiments necessitating meticulous modifications of factors like light intensity, pH, and catalyst loading. While these trials are crucial for system optimization, their repetitive nature and substantial resource requirements impose considerable constraints on practical use. To tackle these issues, machine learning methods, particularly Artificial Neural Networks (ANN), can be utilized to process and evaluate extensive information. Through comprehensive literature review and information acquisition, ANN facilitates the fitting and assessment of experimental data, minimizing errors and enhancing data interpretation. This method can substantially reduce chemical waste and markedly decrease response times, resulting in more efficient resource utilization and generating additional options for investigating connected study domains.

Levenberg-Marquardt (LM) has fast convergence and precision, Bayesian Regularization (BR) is slower but has built-in regularization, and Scaled Conjugate Gradient (SCG) is ideal for large-scale data convergence. These models will be evaluated using high-accuracy criteria (R2 near 1 and low MSE). Pearson, Spearman, and Kendall's Tau correlation analyses will be used to examine how light irradiation reaction time (min), light source power (watt), catalyst loading (g/L), and pollutant concentration (vol%) affect hydrogen production rate efficiency. The analysis could fully explain photoreforming hydrogen yield parameters, ensuring ANN model flexibility and correctness. This research advances waste-to-energy conversion and environmental remediation catalysis, supporting SDG No. 7 (Affordable and Clean Energy) and SDG No. 12 (Responsible Consumption and Production) by promoting process efficiency, waste management, and sustainable energy solutions.

2. Methodology

The data collection involved literature research, concentrating on critical factors like catalyst loading, light irradiation duration, light source power, and pollutant concentration. The identification of pertinent parameters and datasets occurs prior to establishing the quantity of hidden layers in the ANN, which may vary from 1 to 50. The suitable type of ANN is thereafter chosen, with the configuration of transfer functions and the quantity of neurons in hidden layers established. Training techniques include Levenberg-Marquardt, Bayesian Regularization, or Scaled Conjugate Gradient are utilized on 70% of the data in the training phase. Subsequent to training, the ANN is subjected to testing with the residual data to assess its performance. The assessment entails examining the predictive accuracy of the ANN. Should the predictions prove satisfactory, the process will complete. Should this not suffice, modifications are implemented, including the augmentation of hidden layers, until optimal performance is achieved. This methodical methodology guarantees systematic investigation and optimization of the network to improve its forecasting precision.

2.1 Data Acquisition and Integration

The Multilayer Perceptron (MLP) architecture, utilized for forecasting H_2 production rates, comprises three primary layers: the input layer, hidden layers, and the output layer. In the input layer, neurons correspond to critical variables like light source power, irradiation duration, catalyst loading, and pollutant concentration. The data is subsequently processed through one or more concealed layers. Complex data interactions are represented using weighted connections and biases within these levels. The specific relationship is captured by the Eq. (2).

$$Y_i = \sum_{k=1}^{n} \theta(W_{ik}X_k + b_i) \quad (2)$$

Where W_{ik}, X_k, Y_i and i represent the weight, input, output and number of neurons, respectively. θ represent the activation function while b_i represent the bias.

This process progresses to the output layer, where a single neuron outputs the predicted rate of H_2 reforming. The architecture, as illustrated in Figure 1, demonstrates how the MLP efficiently processes input data to yield accurate predictions of hydrogen production.

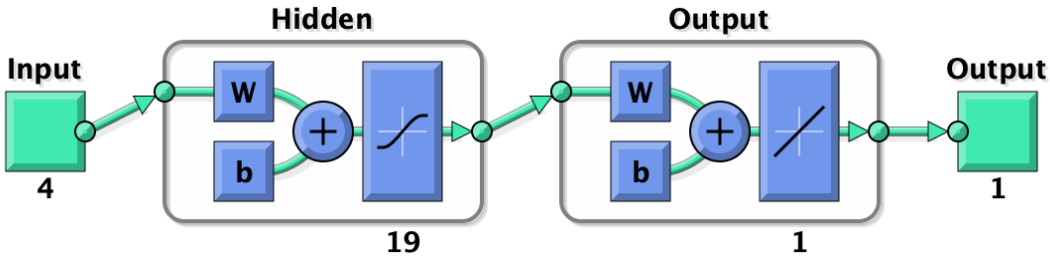

Fig. 1. Conceptual diagram of the ANN model

Figure 2 shows how the artificial neural network (ANN) used in this study is set up. There are 4 nodes in the input layer, selected 19 nodes in the hidden layer (varied based on performance), and 1 node in the output layer. There are weights on each connection between layers, and biases are added to the hidden and output layers.

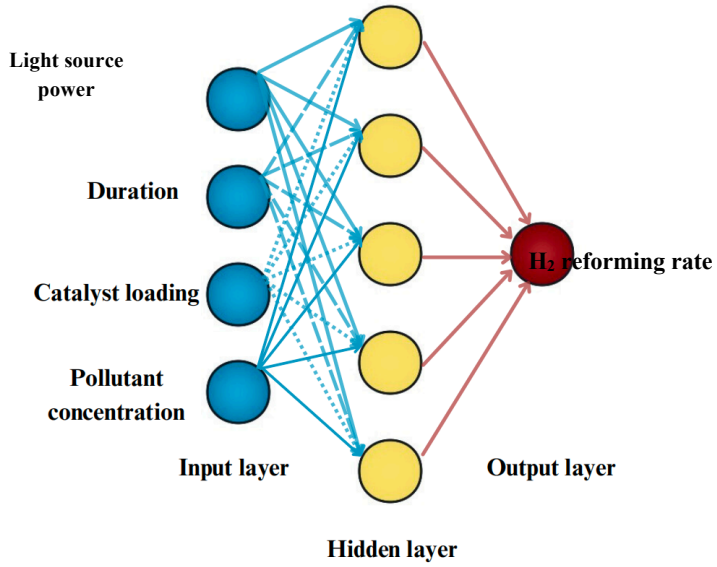

Fig. 2. Diagram of ANN for H2 production estimation process

2.2 Training Algorithms

In the training of artificial neural networks, selecting the appropriate algorithm is essential for attaining optimal performance. The Levenberg–Marquardt (LM) algorithm performs effectively for networks and datasets of intermediate size, providing rapid and dependable convergence. The Scaled Conjugate Gradient (SCG) algorithm is efficient for larger or more intricate neural networks, offering computational efficacy and adaptability. The Bayesian Regularization (BR) approach is typically used for smaller or noisy datasets, as it improves the model's capacity to generalize effectively and reliably to new data.

2.2.1 Levenberg–Marquardt (LM) Algorithm

The Levenberg–Marquardt (LM) method is esteemed as one of the most efficient training algorithms for feed-forward neural networks, particularly for medium-sized networks and datasets. LM is fundamentally a hybrid methodology that integrates the gradient descent technique with Gauss–Newton optimization. It attains swift convergence by adaptively alternating between both strategies based on the intricacy of the error surface.

Initially, the language model resembles gradient descent, facilitating stable convergence, but it transitions towards Gauss–Newton optimization as the network parameters near the optimum. A significant feature is that LM necessitates the computation of the Jacobian matrix, which comprises the partial derivatives of the network errors concerning weights and biases. While this may be computationally demanding for extensive datasets or networks, it leads to markedly quicker convergence for modestly sized issues.

2.2.2 Scaled Conjugate Gradient (SCG) algorithm

The Scaled Conjugate Gradient (SCG) algorithm is an efficient training technique tailored for large-scale challenges and intricate neural network architectures. In contrast to fundamental gradient-descent techniques that rely on user-specified learning rates, SCG autonomously modifies its learning step size throughout the training process, enhancing stability and accelerating convergence.

SCG does not require the explicit computation of the Hessian matrix (second-order derivatives), which enables it to process large-scale or high-dimensional datasets quickly and efficiently. Because SCG avoids calculating and storing extensive derivative matrices, its memory and computational demands are lower than those of other second-order methods, such as LM. For this reason, SCG is frequently applied to large or complex datasets and neural networks.This approach not only accelerates computation but also ensures consistent convergence.

2.2.3 Bayesian Regularization (BR) algorithm

Bayesian Regularization (BR) is a complex way to train a neural network that adds regularization to the training process. The BR algorithm uses Bayesian inference methods to make the model better at generalizing, which helps stop overfitting by punishing networks that are too complex.

Bayesian Regression really finds the best parameters (weights and biases) by looking at a probability distribution for network parameters and finding the best balance between model complexity and fitting accuracy. BR is different from traditional training methods because it doesn't need a separate validation dataset. This is because model selection is built into the training process itself. BR is notably useful for working with small or noisy datasets, which makes neural networks better at handling new data.

2.3 Validation of ANN Models

In the validation of artificial neural network (ANN) models, it is essential to carefully select error metrics to accurately assess the predictions. These measurements quantify the discrepancy between anticipated outcomes and actual observed values. For most measures, the goal is to minimize values; lower values indicate improved accuracy and greater model performance. This study utilized Mean Squared Error (MSE) to evaluate the models.

Conversely, the Coefficient of Determination, referred to as R^2, exhibits distinct behavior. In the context of R^2, elevated values are preferable. An R^2 score approaching 1 signifies that the model's predictions align closely with the actual data, illustrating the model's capacity to replicate and encapsulate the variability of the observed values. The disparity in measurements, with some favoring lower values and others higher, highlights the necessity of a nuanced approach to understanding model validation outcomes.

2.3.1 Mean Squared Error (MSE)

Mean Squared Error (MSE) is a key metric for evaluating the performance of an artificial neural network (ANN) model. It measures the average squared difference between observed responses from experiments, noted as $\mathrm{Y_i}$, and predictions made by the ANN, labeled as $\ddot{\mathrm{Y}}_\mathrm{i}$. The MSE is calculated using Eq. (3).

$$(\mathrm{MSE}) = \frac{1}{\mathrm{n}}\sum_{\mathrm{i=1}}^{\mathrm{n}}\left(\mathrm{Y_i} - \ddot{\mathrm{Y}}_\mathrm{i}\right)^2 \quad (3)$$

n represents the total number of observations. This metric is particularly useful because it emphasizes larger errors by squaring the differences, which places more emphasis on significant deviations from the actual values.

A reduced MSE value indicates that the model's predictions are more closely aligned with the actual data, signifying enhanced accuracy. The Mean Squared Error (MSE) is particularly beneficial in contexts where substantial errors are paramount, as it aids in identifying models that may exhibit overall competence yet still experience sporadic considerable flaws. By concentrating on lowering MSE during the training process, the ANN model can be optimized to diminish prediction errors, resulting in a more dependable predictive model.

2.3.2 Coefficient of determination (R^2)

The Coefficient of Determination, commonly known as quantifies the degree to which a model's predictions approximate actual data points. It is calculated using Eq. (4).

$$R^2 = \left(\frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}\right)^2 \quad (4)$$

In this equation, the letter x stands for observed values, y stands for predicted values, and n stands for the number of observations in this equation. A perfect fit is shown by a R^2 value of 1, which means that the model's predictions match the data exactly, with no differences. On the other hand, a R^2 score that is closer to 0 means that the model does not adequately show the trends in the data. R^2 is a useful tool for evaluating how well a model captures and represents the variability present in the data under analysis. A higher R^2 number usually suggests that the model performs a better job of explaining what is seen in the real world, therefore it's also helpful for comparing how well different models work.

2.4 Correlation Analyses of Inputs and Outputs

In the analysis of influential variables and their outputs, three primary methods of correlation analysis, Pearson, Spearman, and Kendall, are used to examine the relationships between these

variables, each providing a distinct perspective on how closely they are related and helping to identify the key factors that significantly impact the outcomes.

2.4.1 Pearson correlation analysis

Pearson correlation analyzes the linear relationship between variables, highlighting how one variable changes in response to another under the assumption of a linear correlation. This method is particularly sensitive to outliers and is most effective with data that is normally distributed. The Pearson equation is shown in Eq. (5).

Pearson method: $\boldsymbol{r} = \frac{\sum(\boldsymbol{x_i}-\bar{\boldsymbol{x}})(\boldsymbol{y_i}-\bar{\boldsymbol{y}})}{\sqrt{\sum(\boldsymbol{x_i}-\bar{\boldsymbol{x}})^2\sum(\boldsymbol{y_i}-\bar{\boldsymbol{y}})^2}}$ (5)

The above equation measures the linear correlation between two variables, where x_i and y_i are data points, and $\bar{x}$ and $\bar{y}$ are the means.

Range: -1 to 1

Interpretation:

- r = 1: Perfect positive linear correlation.
- r = 0: No linear correlation.
- r = -1: Perfect negative linear correlation.
- Values closer to 1 or -1 indicate stronger relationships, while values near 0 suggest weak or no correlation.

2.4.2 Spearman correlation analysis

Spearman correlation, a non-parametric approach, evaluates how accurately the relationship between two variables can be described by a monotonic function. This involves ranking the data before determining the Pearson correlation of these ranks, which enhances its robustness against non-normal data distributions and outliers compared to the Pearson method.

The Spearman correlation is shown in Eq. (6).

Spearman method: $\boldsymbol{\rho} = \mathbf{1} - \frac{\mathbf{6}\sum \boldsymbol{d_i^2}}{\boldsymbol{n(n^2-1)}}$ (6)

Here, d_i is the difference between the ranks of corresponding variables, and n is the number of data points.

Range: -1 to 1

Interpretation:

- ρ = 1: Perfect positive monotonic relationship.
- ρ = 0: No correlation in ranks.
- ρ = -1: Perfect negative monotonic relationship

2.4.3 Kendall correlation analysis

Kendall correlation, also referred to as Kendall's tau coefficient, gauges the strength of dependence between two variables. It offers a more conservative approach than Spearman and excels with small sample sizes or datasets that feature many tied ranks. The Kendall correlation is shown in Eq. (7)

Kendall method: $\tau = \frac{(C-D)}{\frac{1}{2}n(n-1)}$ (7)

where C is the number of concordant pairs, D is the number of discordant pairs, and n is the total number of pairs.

- τ = 1: Perfect agreement between rankings.
- τ = 0: No association.
- τ = -1: Perfect disagreement between rankings.

In correlation analysis using the Pearson, Spearman, and Kendall methods, positive values signify a direct correlation, while negative values point to inverse relationships, with values close to 1 or -1 indicating stronger correlations.

2.5 Sensitivity analysis using modified Garson algorithm

The influence of each input parameter on the ANN output was evaluated using the modified Garson algorithm. In this approach, the weights connecting the input and hidden layers (W_{ij}) are combined with those linking the hidden and output layers (V_{jK}) to distribute the contribution of each input. The relative importance is calculated according to Eq. (8) [8]:

Garson algorithm: $\varphi_{ik} = \frac{\sum_{k=1}^{l}|W_{ij}V_{jk}|/\sum_{r=1}^{l}|W_{rj}|}{\sum_{i=1}^{n}\sum_{j=1}^{l}|W_{ij}V_{jk}|/\sum_{r=1}^{l}|W_{rj}|}$ (8)

In this formula, W_{ij} is the weight between the $i-th$ input and $j-th$ hidden neuron, $V_{jk}$is the weight between the $j-th$ hidden and $k-th$ output neuron, n is the number of inputs, and l is the number of hidden neurons. After normalization, φ_{ik} gives the relative contribution of each input variable, making it possible to rank the parameters according to their impact on the network output.

3. Results

3.1 UV Light Source

3.1.1 Models evaluation

This section discusses the performance of three intelligent estimators: Levenberg-Marquardt (LM), Bayesian Regularization (BR), and Scaled Conjugate Gradient (SCG) algorithms, as applied to the prediction of hydrogen production rate based on four key input variables: light source power (W), catalyst loading (g/L), glycerol concentration (v/v%), and duration (h), with the output being the hydrogen production rate $\mu mol g^{-1}h^{-1}$. A critical aspect of this study is the size of the dataset available for training (70%), validation (15%), and testing (15%), which comprises only 75 data points for experiments conducted under UV light conditions. The models' efficacy was evaluated based on their Coefficient of Determination (R^2) and Mean Squared Error (MSE) values across training, validation, testing, and the entire dataset.

The availability of only 75 data points for the entire dataset under UV light conditions significantly influences the training dynamics and achievable accuracy of any neural network model. This limited size stems from the inherent practical challenges in simultaneously obtaining experimental data points for all four input variables (light source power, catalyst loading, glycerol concentration, and duration) along with their corresponding hydrogen production rates. Such multi-factorial

experiments are often resource-intensive and time-consuming, making the acquisition of a large, comprehensive dataset inherently difficult. Consequently, the input space is sparsely covered, increasing the risk of overfitting during training for data-hungry neural networks and posing a significant challenge for their generalization ability.

Once the hidden layer size is set to 15, the ANN model is as illustrated in Figure 3(a). This process involves training the model using the provided data, after which the R^2 values for training, validation, and testing sets, as well as the overall data, will be calculated.

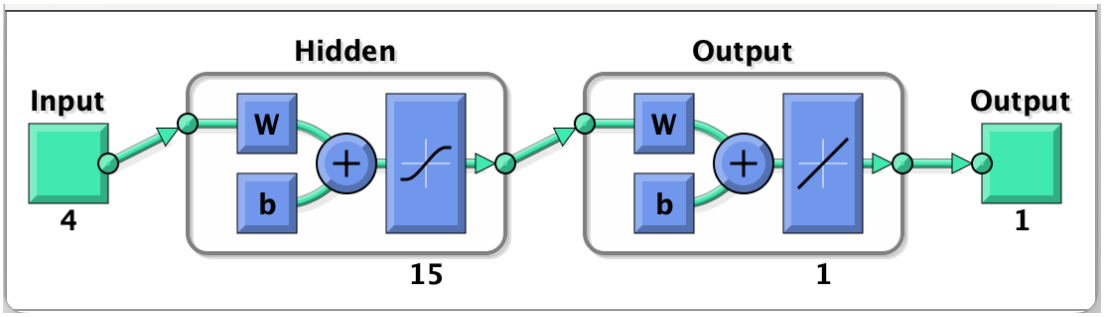

Fig. 3 (a). Conceptual diagram of the ANN model (15 hidden neurons)

The Levenberg Marquardt model got a training R^2 of 0.9679, a validation R^2 of 0.9421, and a testing R^2 of 0.9474 when it was lit by ultraviolet light. The model had an overall R^2 of 0.9575 and a mean squared error (MSE) of 2.45×10^7. These results show that the LM model was able to pick up on the nonlinear patterns in the training data.

As shown in Figure 4(b) and with the R^2 values for training, validation, and testing all exceeding 0.9, it can be concluded that the LM model is highly reliable. This success can be attributed to its use of a second-order optimization method. Second-order optimization methods, unlike first-order methods that only rely on gradients (first derivatives), also take into account the curvature of the objective function by using second derivatives. The LM algorithm uses the Jacobian matrix, which contains the first-order derivatives, to approximate the Hessian matrix, which represents the second-order partial derivatives. The LM algorithm effectively traverses the error surface by employing this approximation, which accounts for the objective function's curvature (Hessian) and direction (gradient) [9].

The apparently large MSE of 2.45×10^7 obtained for the UV LM model reflects the much higher scale of the target variable, since hydrogen production rates in this work range from 10 000 to 100 000 µmol g^{-1} h^{-1} and the MSE squares every prediction error. By contrast, there is one study that shows an MSE of 1.13×10^{-3} when modelling SO_2 solubility in deep eutectic solvents, where the targets range from 0 to 1 mole fraction [10]. Because the output values in this work are about five orders of magnitude larger, squaring the residuals amplifies the gap by another five orders, so an R^2 above 0.95 can reasonably correspond to an MSE in the 10^7 range here, whereas an R^2 of 0.98 yields an MSE near 10^{-3} in the cited study. This comparison confirms that the predictive accuracy of the current model is appropriate for the scale of its outputs.

Because of this, the model can converge more quickly and accurately, which makes it perfect for handling complex, nonlinear relationships in data, like estimating the rates at which hydrogen is produced. The Levenberg-Marquardt (LM) algorithm's lack of built-in regularization, however, is a known drawback that can result in overfitting, particularly when dealing with noisy data or an excessive number of parameters. In spite of this, the model's reliability is confirmed by the high R^2 values obtained from testing, validation, and training, which show how well it captures nonlinear patterns. This pattern fits with [11] when LM-trained networks are used for hydrogen production

rate tasks, they tend to do well on training subsets but not so well on generalization, especially when the dataset sizes are small. The regression fit for the LM model on the training data is visually represented in Figure 3(b).

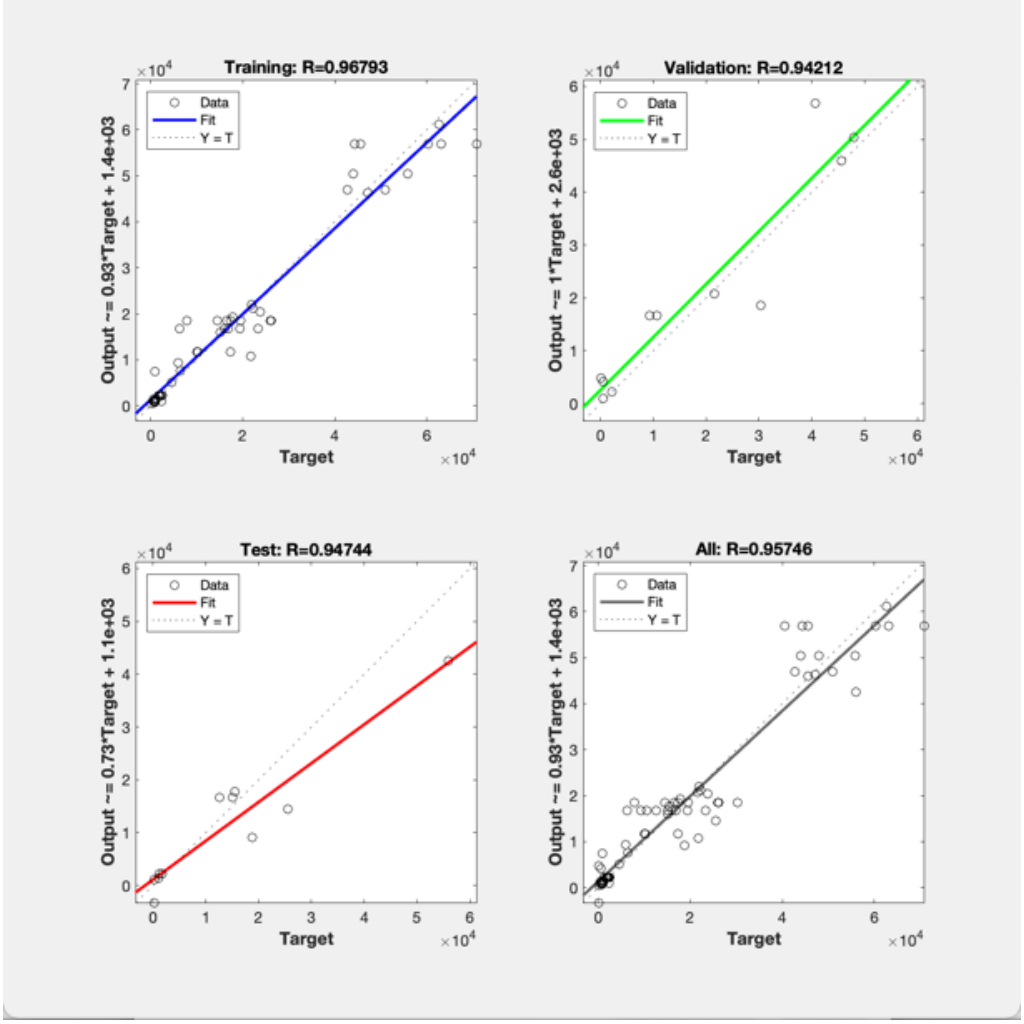

Fig. 3 (b). Regression diagram based on LM model

In contrast to LM, the BR algorithm with 7 hidden neurons (Figure 4c) exhibited remarkable consistency across the datasets. It achieved the highest R^2 on the training set (0.9660) and maintained a robust performance on the testing set with an R^2 of 0.9623. Notably, no separate validation set was used for BR, as its inherent regularization mechanism handles model complexity adaptively. This strong and consistent performance on both seen and unseen data underscores BR's superior generalization ability, especially critical given the limited 75 data points. BR operates within a probabilistic framework that automatically determines optimal regularization parameters, effectively penalizing unnecessary model complexity and shrinking less important weights towards zero. This principled approach prevents overfitting by balancing data fit with model simplicity, leading to a more robust and generalized learned relationship [12]. The excellent training regression achieved by the BR model is illustrated in Figure 3(c) with seven hidden neurons, the BR model achieved training $R^2 = 0.9660$, testing $R^2 = 0.9623$, all-data $R^2 = 0.9629$, and MSE = 2.39×10^7 after the training process as shown in Figure 3(d).

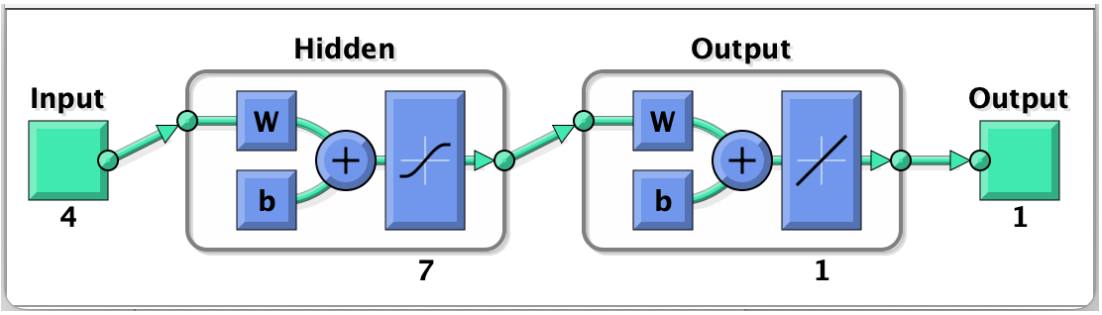

Fig. 3 (c). Conceptual diagram of the ANN model (7 hidden neurons)

In machine learning, regularization is a technique used to prevent overfitting, a phenomenon where a model learns the training data too well, including its noise and outliers, leading to poor performance on unseen data. Essentially, regularization adds a penalty term to the model's loss function. This function is what the algorithm tries to minimize during training. This penalty discourages the model from becoming overly complex or relying too heavily on any single feature or weight. By balancing the fit to the data with model complexity, regularization aims to improve the model's generalization ability, its capacity to make accurate predictions on new, unobserved data. Common regularization techniques often involve penalizing large weight values, effectively "shrinking" them [10].

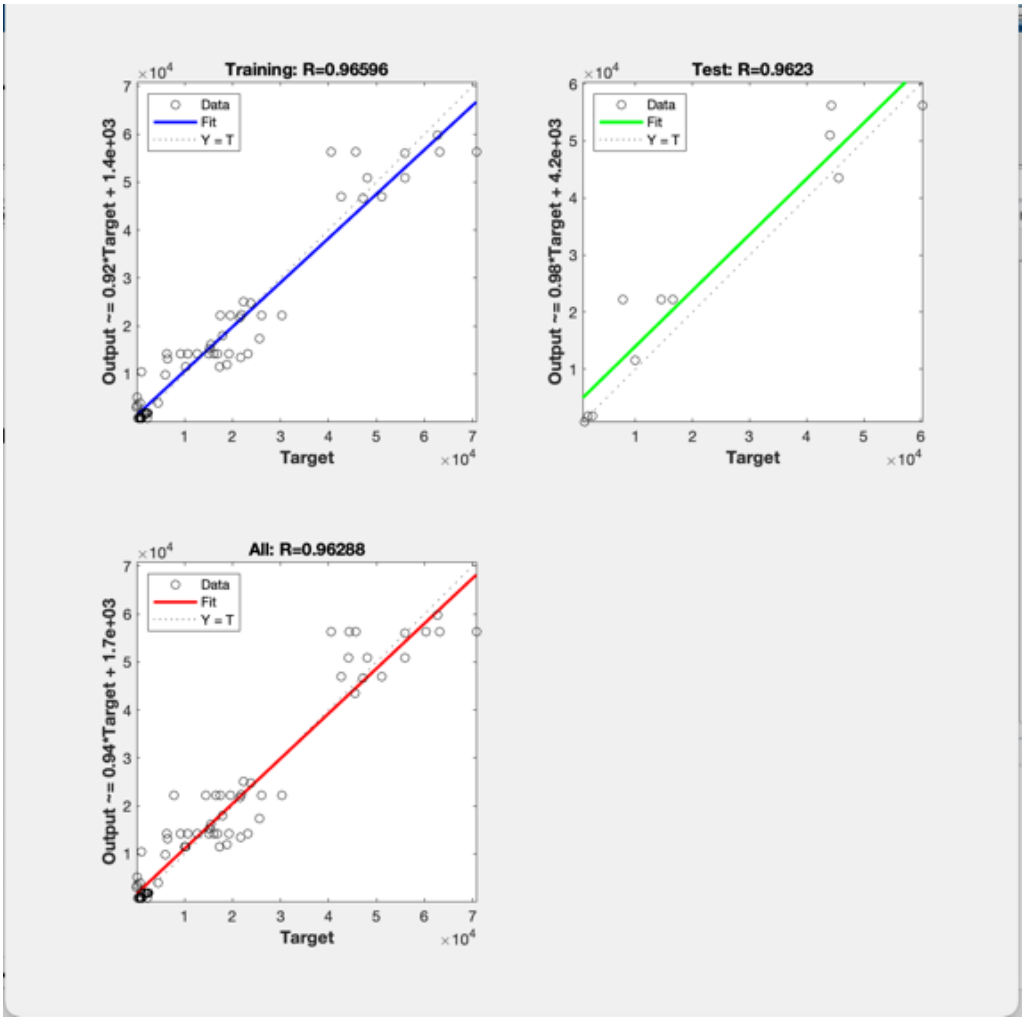

Fig. 3 (d). Regression diagram based on BR model

The regression fit for the SCG model on the training data (15 hidden neurons) is visually represented in Figure 3(e) (15 hidden neurons). The SCG algorithm showed good performance across the datasets by training process as illustrated in Figure 3(e). It achieved an R^2 of 0.9274 on the training set, followed by 0.9259 on the validation set, and then an improvement to 0.9516 on the testing set as illustrated in Figure 3(f). While its training R^2 was lower than LM and BR, because the SCG algorithm uses less memory than the LM algorithm, it is especially well-suited to processing larger datasets. SCG is more scalable when working with high-dimensional data or larger datasets because it only computes the Hessian vector product, as opposed to LM, which explicitly approximates the entire Hessian matrix. When memory resources are scarce or the dataset size is sufficiently large to render LM's full Hessian approximation computationally costly, SCG's efficiency allows it to handle these scenarios. Its lower training R^2 compared to LM, however, indicates that this trade-off comes at the expense of somewhat less accuracy in capturing the intricate nonlinear relationships in the training data.

Although SCG works well with larger datasets, it can achieve higher training R^2 in smaller, easier-to-manage datasets thanks to the more intricate adjustments in LM. However, like LM, SCG does not incorporate an explicit, built in regularization term like BR. Its iterative optimization can still be sensitive to the sparse distribution of data points [12]. Nevertheless, its performance indicates a reasonable balance was achieved for the limited dataset, navigating the multiple dimensional input space fairly well despite the absence of explicit regularization.

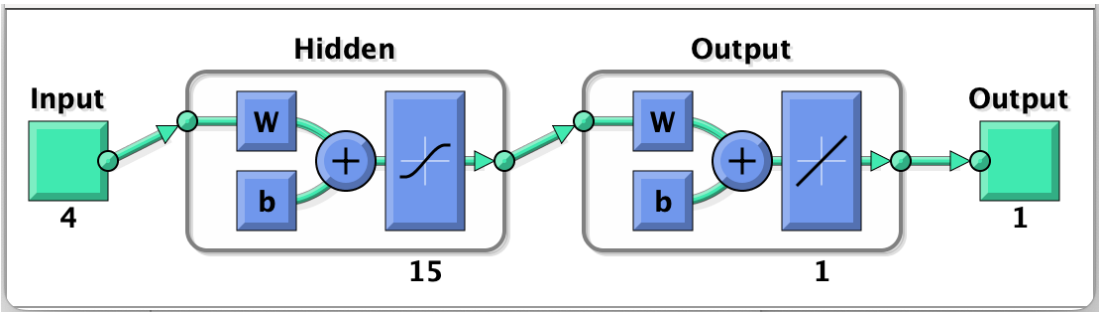

Fig. 3 (e). Conceptual diagram of the ANN model (15 hidden neurons)

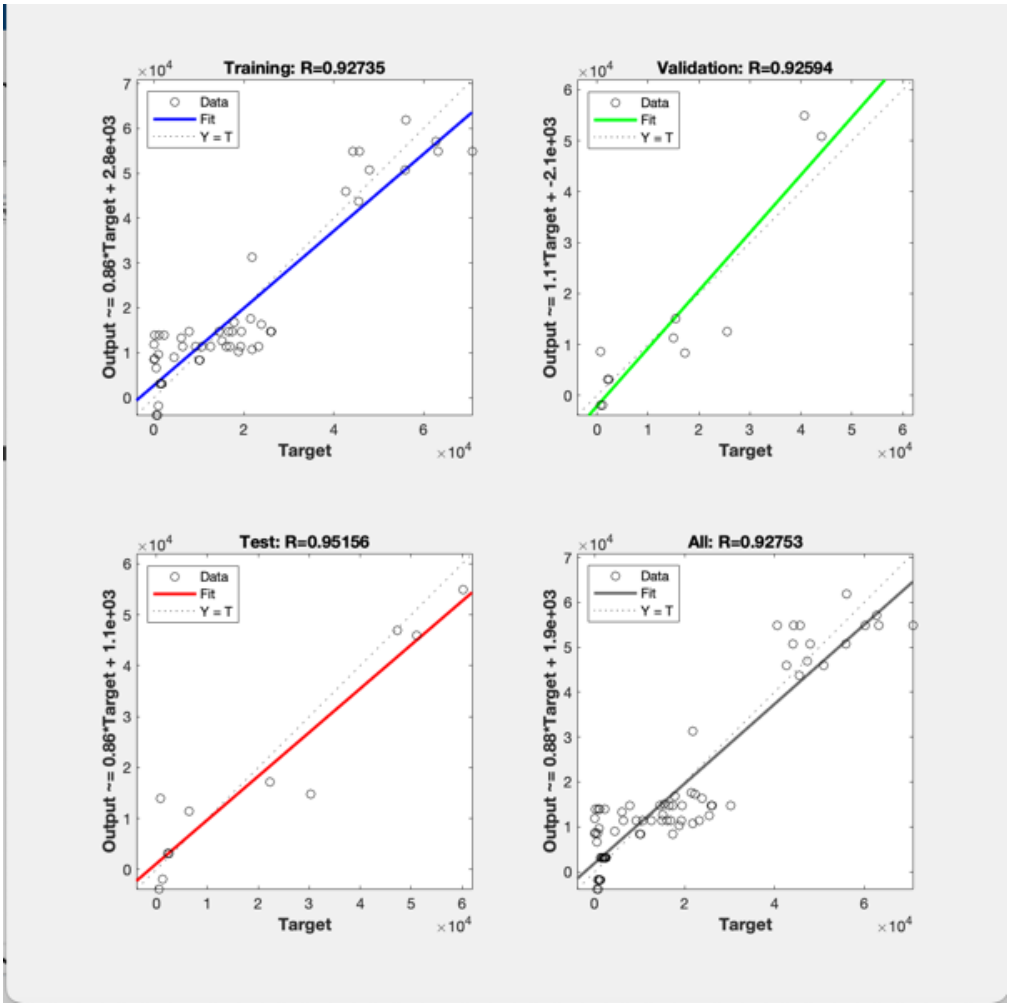

Fig. 3 (f). Regression diagram based on SCG model

Given the constraint of only 75 data points available from UV light experiments, representing a sparse sampling of the input space, the Bayesian Regularization (BR) model is identified as the most effective intelligent estimator. Its intrinsic regularization mechanism allows it to achieve a high R^2 (0.9629) and low MSE (2.39×10^7) and robust generalization by effectively preventing overfitting, a common pitfall with limited and sparsely distributed datasets. This makes BR particularly well suited for training and modeling complex relationships under UV light conditions, all specific data regarding the performance of each intelligent estimator are summarized in Table 1.

Table 1 melintang

3.1.2 Correlation analysis

To further understand the relationship between the model's predictions and the actual experimental data, as well as the inherent dependencies captured by each intelligent estimator, a comprehensive correlation analysis was performed. This analysis utilized three widely recognized correlation coefficients: Pearson, Spearman, and Kendall Tau. The trends of these coefficients for each input variable under the LM, BR, and SCG models are presented in Figures 4(g), 4(h), and 4(i), respectively. The relevance of the figures presented in subsequent sections will be directly linked to the correlation trends described in Table 2. These figures will offer a more intuitive view of the

dependencies between input variables and model predictions, following the classification provided in the table.

Table 2
Interpretation of the physical meaning of the Pearson, Spearman, and Kendall coefficient (η)

Range of Coefficients	Correlation type
$-1 \leq \eta \leq -0.8$	Very strong indirect relationship
$-0.79 \leq \eta \leq -0.6$	Strong indirect relationship
$-0.59 \leq \eta \leq -0.4$	Moderate indirect relationship
$-0.39 \leq \eta \leq -0.2$	Weak indirect relationship
$-0.19 \leq \eta \leq -0$	Very weak indirect relationship
$\eta = 0$	No relationship
$0 \leq \eta \leq 0.19$	Very weak direct relationship
$0.2 \leq \eta \leq 0.39$	Weak direct relationship
$0.4 \leq \eta \leq 0.59$	Moderate direct relationship
$0.6 \leq \eta \leq 0.79$	Strong direct relationship
$0.8 \leq \eta \leq 1$	Powerful direct relationship

The correlation patterns for the LM model, depicted in Figure 4(g), indicate a diverse level of agreement among the three correlation types for our various input variables. LM demonstrated a lack of intrinsic regularization, leading to overfitting. The model excelled on the training subset, although its effectiveness markedly declined on the validation and testing sets. This overfitting often shows up right here: for certain inputs, Pearson's coefficient might lag noticeably behind Spearman's or Kendall's. This divergence means that while LM might have grasped the general upward or downward flow (monotonic trends), its ability to accurately model the precise linear connections across the full data range, particularly with our limited dataset, was compromised [13].

Within the Levenberg Marquardt model, based on the correlation analysis as shown in relevancy Figure 3(g), the light source power shows a powerful direct relationship with hydrogen production according to the Pearson coefficient (0.8 to 1), while the Spearman coefficient (0.2 to 0.39) indicates a weak direct relationship, and the Kendall coefficient (0 to 0.19) suggests a very weak direct relationship. For catalyst loading, both Pearson and Kendall coefficients (−0.59 to −0.4) reflect a moderate indirect relationship, whereas the Spearman coefficient (−0.79 to −0.6) reveals a strong indirect relationship. Glycerol concentration demonstrates a weak direct relationship under the Pearson coefficient (0.2 to 0.39), a strong direct relationship under the Spearman coefficient (0.6 to 0.79), and a moderate direct relationship under the Kendall coefficient (0.4 to 0.59). Duration exhibits a moderate indirect relationship according to both the Pearson and Kendall coefficients (−0.59 to −0.4), while the Spearman coefficient (−0.79 to −0.6) indicates a strong indirect relationship. These classifications are based on the ranges defined in Table 4 and help clarify the strength and direction of each input variable's influence on hydrogen production rate.

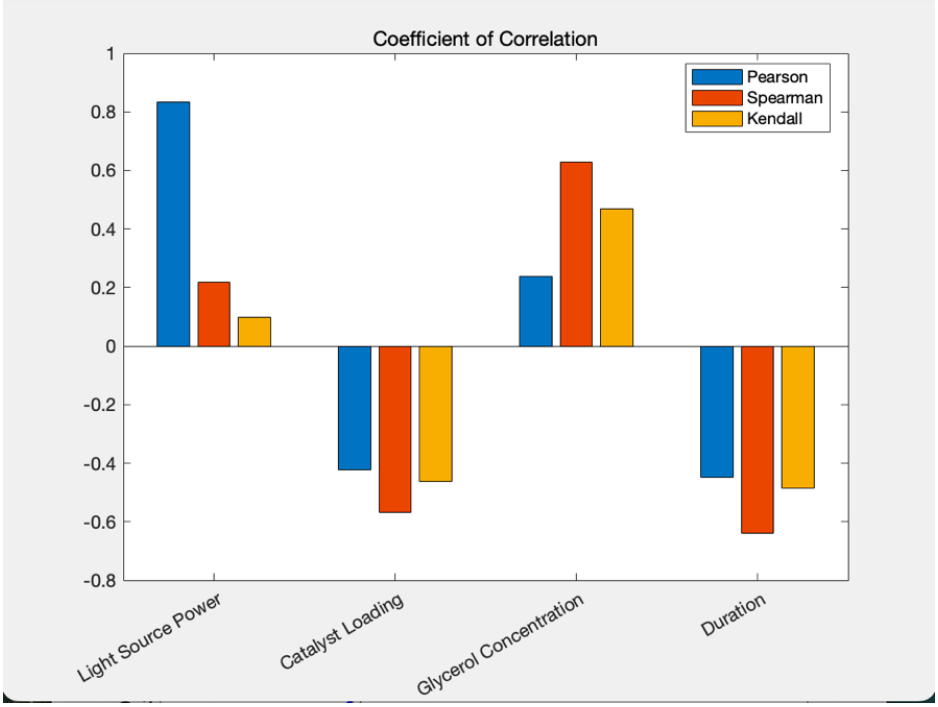

Fig. 3 (g). The relevancy between hydrogen production rate in photoreforming glycerol and light source power, catalyst loading, glycerol concentration and duration under UV light source and LM model training

As demonstrated in Figure 3(h), For the BR model, the correlation analysis reveals distinct patterns across input variables and correlation methods. Light source power exhibits a strong direct relationship under the Pearson coefficient (0.6 to 0.79), while both Spearman and Kendall coefficients (0 to 0.19) indicate a very weak direct relationship. Catalyst loading shows a moderate indirect relationship based on Pearson and Kendall values (−0.59 to −0.4), whereas Spearman (−0.79 to −0.6) suggests a strong indirect correlation. Glycerol concentration displays a weak direct relationship according to the Pearson coefficient (0.2 to 0.39), a strong direct relationship under Spearman (0.6 to 0.79), and a moderate direct relationship based on Kendall (0.4 to 0.59). As for duration, both Pearson and Kendall coefficients (−0.59 to −0.4) fall into the range of moderate indirect relationship, while Spearman (−0.79 to −0.6) again reflects a strong indirect association. These trends, interpreted in line with the ranges defined in Table 4. The seamless agreement among Pearson, Spearman, and Kendall confirms that the Bayesian regularization mechanism prevents overfitting and forces the network to extract relationships that are both robustly linear and reliably monotonic, resulting in superior generalization and a clear depiction of stable input–output dependencies [14].

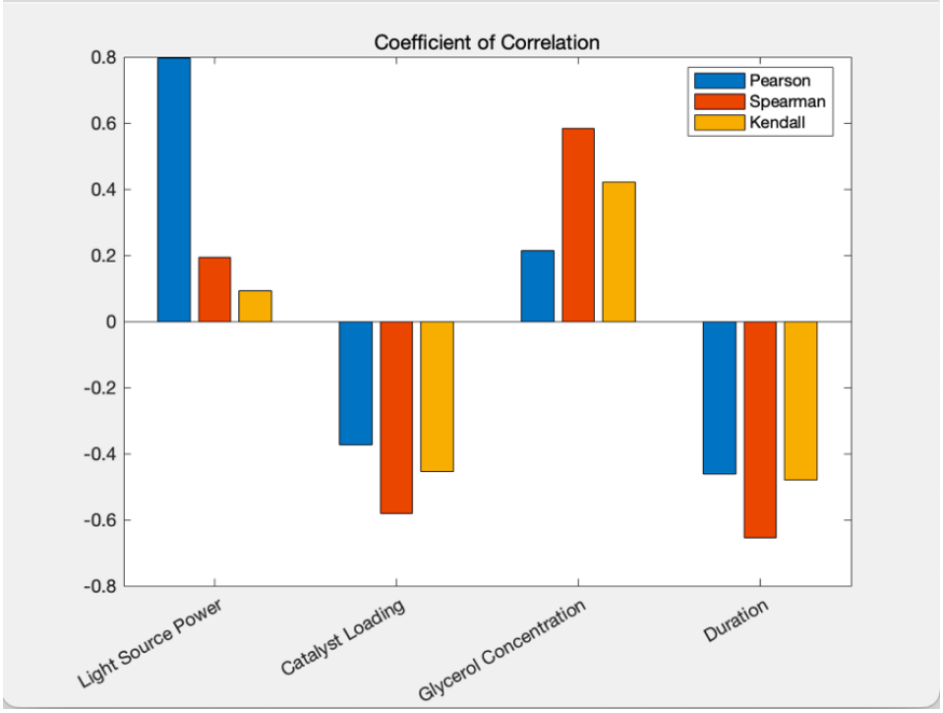

Fig. 3 (h). The relevancy between hydrogen production rate in Photoreforming glycerol and light source power, catalyst loading, glycerol concentration and duration under UV light source and BR model training

The Scaled Conjugate Gradient model, depicted in Figure 3(i), the Pearson coefficient for light source power falls within the range of 0.8 to 1, indicating a powerful direct relationship. However, both Spearman and Kendall coefficients lie between 0 and 0.19, reflecting only a very weak direct relationship. Catalyst loading shows stronger negative associations, with Spearman suggesting a strong indirect relationship (−0.79 to −0.6), while Pearson and Kendall indicate moderate indirect relationships (−0.59 to −0.4). Glycerol concentration correlates positively, with Spearman showing a strong direct relationship (0.6 to 0.79), Kendall indicating a moderate one (0.4 to 0.59), and Pearson suggesting a weaker direct link (0.2 to 0.39). For duration, the correlation is consistently negative: Spearman implies a strong indirect relationship (−0.79 to −0.6), whereas Pearson and Kendall point to moderate indirect relationships (−0.59 to −0.4). These findings, classified according to the scale in Table 4, highlight how different statistical methods can reflect varying dimensions of input-output dependency in the SCG model. Inputs with positive rank correlations can still be increased to lift performance, yet the less uniform alignment across the three measures indicates that SCG offers weaker assurance for precise linear optimization than its Bayesian regularized counterpart [12].

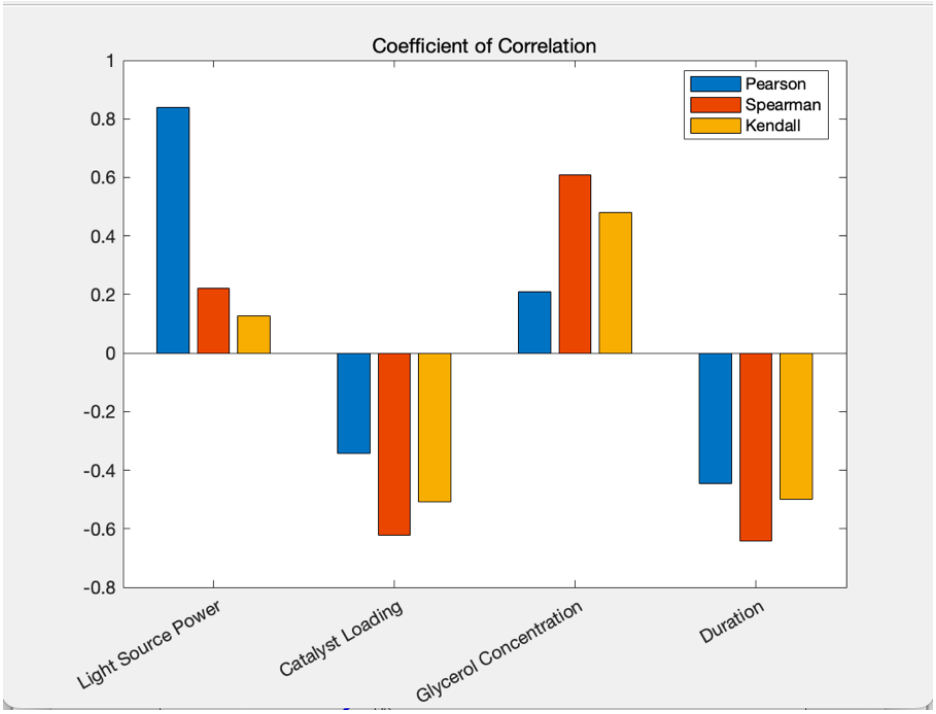

Fig. 3 (i). The relevancy between hydrogen production rate in photoreforming glycerol and light source power, catalyst loading, glycerol concentration and duration under UV light source and SCG model training

Based on the comprehensive correlation analysis using Pearson, Spearman, and Kendall coefficients, the Bayesian Regularization (BR) model has been identified as the most effective among the three evaluated models for predicting hydrogen production under ultraviolet (UV) light irradiation. This makes the correlation results within the BR model particularly reliable. Notably, light source power in the BR model shows a Pearson correlation coefficient which falls within the range of 0.8 to 1 and is classified as a powerful direct relationship according to the criteria defined in Table 2.

Although glycerol concentration demonstrates a consistent increasing trend across all three correlation methods, showing a linear relationship in Pearson and an ordinal increase in both Spearman and Kendall, the overall influence of light source power remains the most significant.

Under ultraviolet conditions, the high energy of the irradiation enhances photocatalytic activity, making light source power the dominant influencing factor. Although glycerol concentration shows a positive correlation with hydrogen production, an excessive amount can reduce light penetration due to increased opacity, which limits the efficiency of the reaction. In contrast, ultraviolet light, with its high photon energy, continues to drive the generation of electron hole pairs effectively. Consequently, despite the increasing trend observed for glycerol concentration, all three models including Levenberg Marquardt, Bayesian Regularization, and Scaled Conjugate Gradient consistently show that under ultraviolet irradiation, light source power remains the most influential variable affecting hydrogen production.

3.2 Visible Light Source

3.2.1 Models evaluation

Figure 4(a) shows the Levenberg-Marquardt (LM) model with 19 hidden layer size, which consistently demonstrating robust performance across all evaluated datasets. It achieved a coefficient of determination (R^2) of 0.9831 on the training set, an exceptionally high R^2 of 0.9904 on

the validation set, and maintained a strong R^2 of 0.9668 on the independent testing set. For the aggregated "All data" set, the R^2 was recorded as 0.9207 as shown in Figure 4(b).

The nearly flawless R^2 reported in the validation set is significant, demonstrating that the model was extensively refined during development and adeptly identified the underlying patterns in the data. The robust performance on the test set further illustrates its remarkable generalization abilities to novel data. The exceptionally high R^2 on the validation set, along with the strong R^2 on the test set and the minimal overall mean squared error (MSE) of 2.49×10^3 across all datasets, substantiates that the LM algorithm effectively utilized its intrinsic robustness and rapid convergence attributes [2] . This allowed the model to get a highly optimized and generalizable fit, emphasizing its distinctive ability to delineate the intricate nonlinear correlations inherent in photocatalytic hydrogen production data. The minimal MSE further substantiates that, on average, its forecasts are nearest to the actual values.

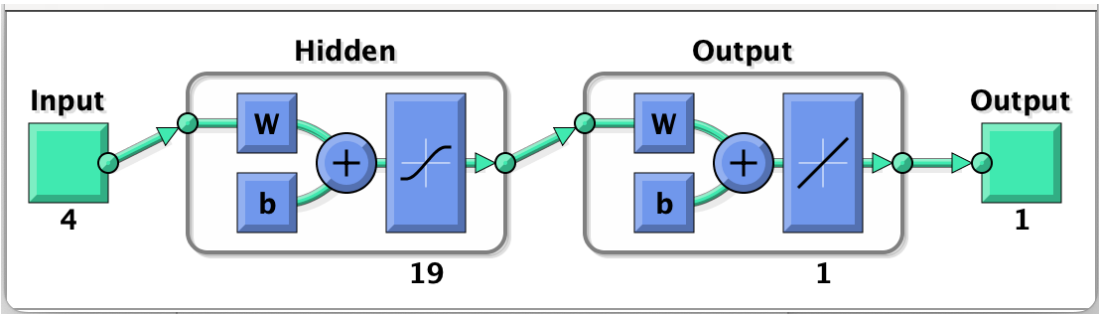

Fig. 4 (a). Conceptual diagram of the ANN model (19 hidden neurons)

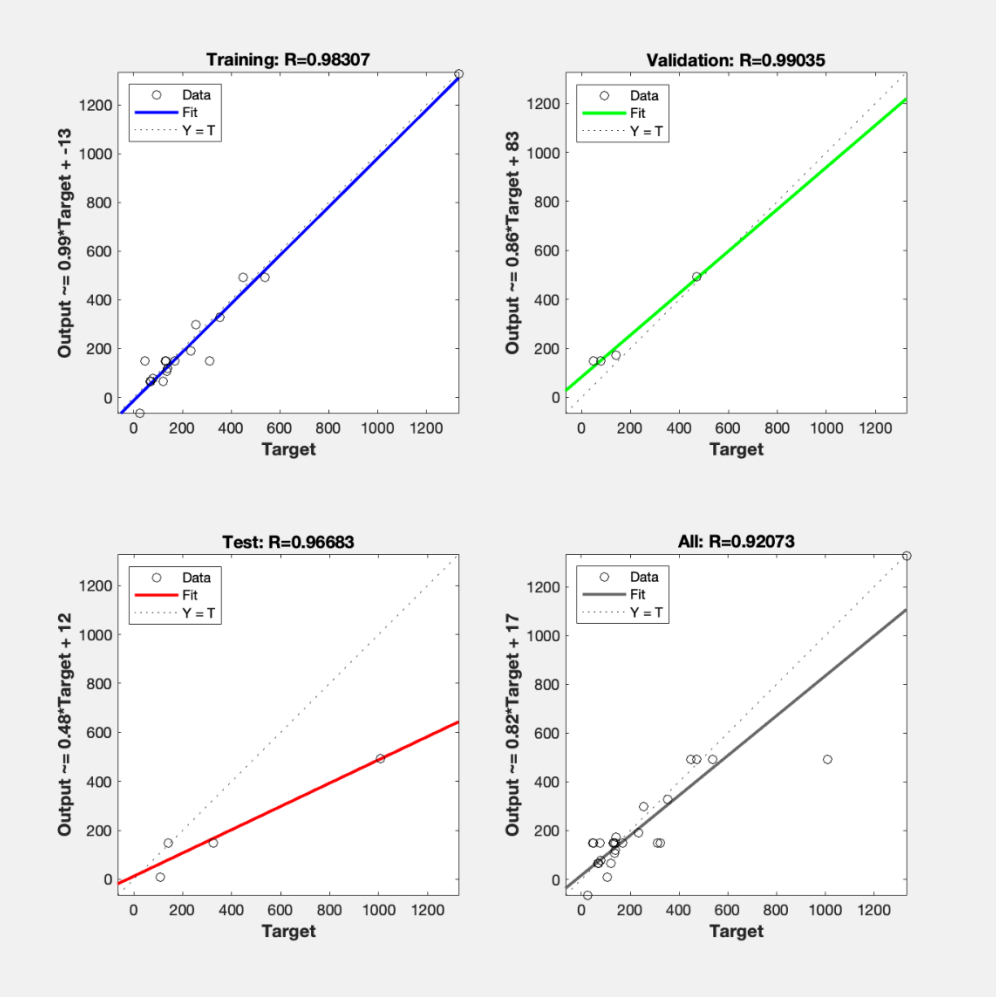

Fig. 4 (b). Regression diagram based on LM model

Figure 4(c) displays a model with 9 hidden neurons with the Bayesian Regularization (BR) model demonstrated strong predictive performance, with an R^2 of 0.9257 on the training set and an exceptionally high 0.9897 on the independent testing set. The R^2 value for the "All data" set was 0.9277 as illustrated in Figure 4(d). It is worth noting that a separate validation R^2 was not reported, which is consistent with the BR model's inherent regularization properties that often reduce the necessity for extensive cross-validation. A particularly notable observation is that the R^2 on the testing set surpassed that of the training set, indicating BR's superior generalization ability and its intrinsic resistance to overfitting [15]. This suggests that the model effectively captured the fundamental patterns within the data while possibly filtering out noise that was more prominent in the training set. In terms of prediction error, the Mean Squared Error (MSE) for the "All data" set was 1.46×10^4. There is one study that reports an MSE of only 0.641 for photocatalytic dye-degradation efficiency [16]. Because the targets in the present work are orders of magnitude larger than those in that study, the resulting MSE of 1.46×10^4 is proportionally higher, even though both models achieve R^2 values above 0.9. This comparison highlights the need to interpret MSE in the context of the target scale or to employ scale-independent error metrics when comparing different studies. Although this is slightly higher than the MSE of the LM model, it remains substantially lower than that of the SCG model, reflecting a commendable level of predictive accuracy.

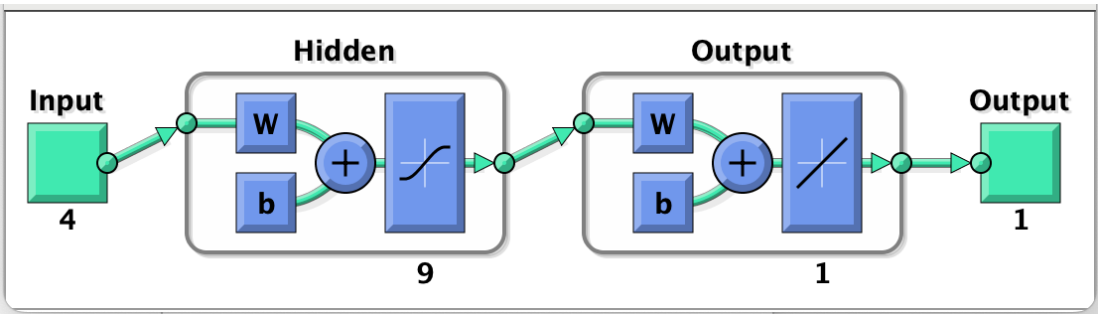

Fig. 4 (c). Conceptual diagram of the ANN model (9 hidden neurons)

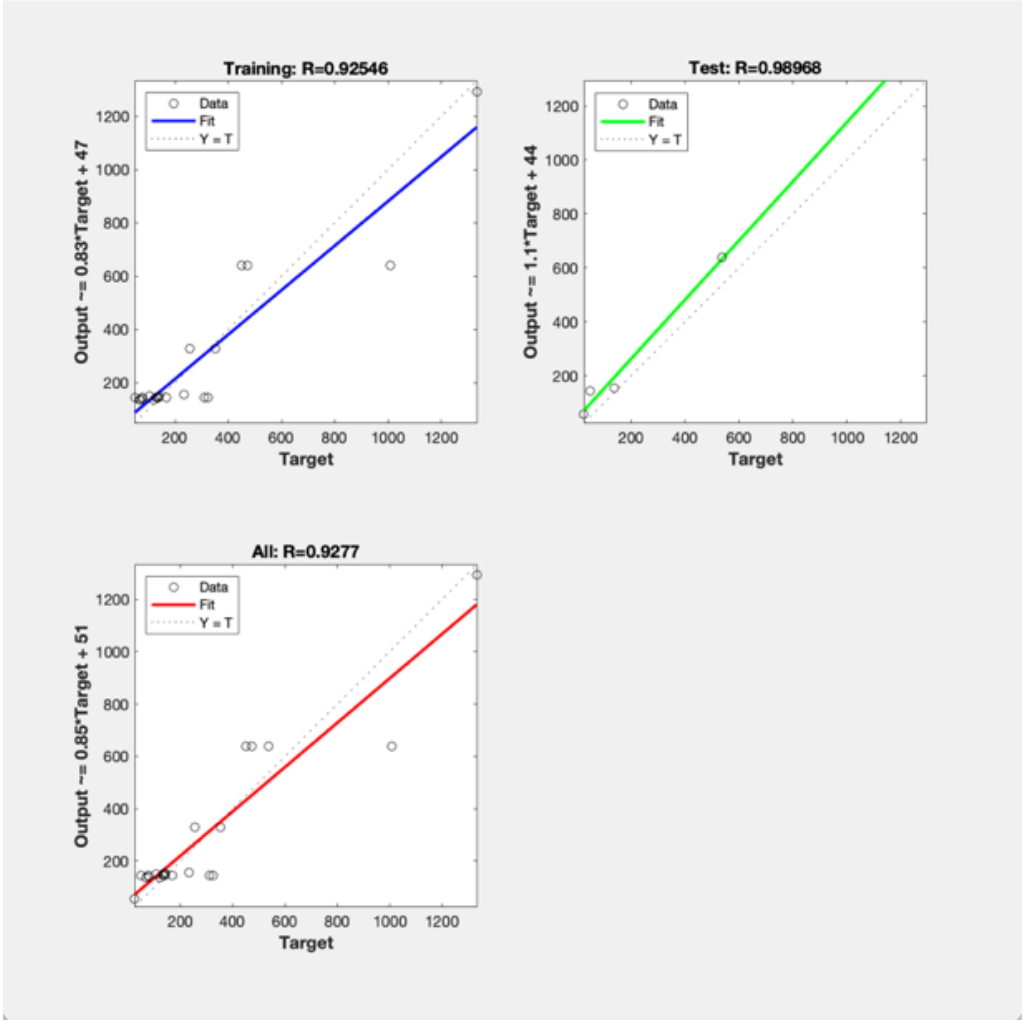

Fig. 4 (d). Regression diagram based on BR model

The Scaled Conjugate Gradient (SCG) model performed quite well on the training process (Figure 4e) with an R^2 of 0.9779 and did even better on the validation set with an R^2 of 0.9964. However, its performance dropped significantly on the independent testing set, where the R^2 fell to 0.9747. When looking at the combined "All data" set, the R^2 was 0.9189. This noticeable decline from validation to testing suggests that the SCG model struggles with generalization. All of these values can be found in Figure 4(f). In other words, it seemed to learn the training and validation data quite well but wasn't able to maintain that level of accuracy when faced with new data.

The Mean Squared Error (MSE) for the full dataset was 3.50×10^3, which is much higher than the MSEs from the LM and BR models. This further confirms that SCG's predictions were less accurate

overall. While the SCG algorithm is known for its fast and stable convergence, in this case, that stability may have caused the model to settle on a solution that worked well for the training data but didn't translate well to unseen data. It might also have been more sensitive to noise or unique patterns in the testing set that weren't present during training.

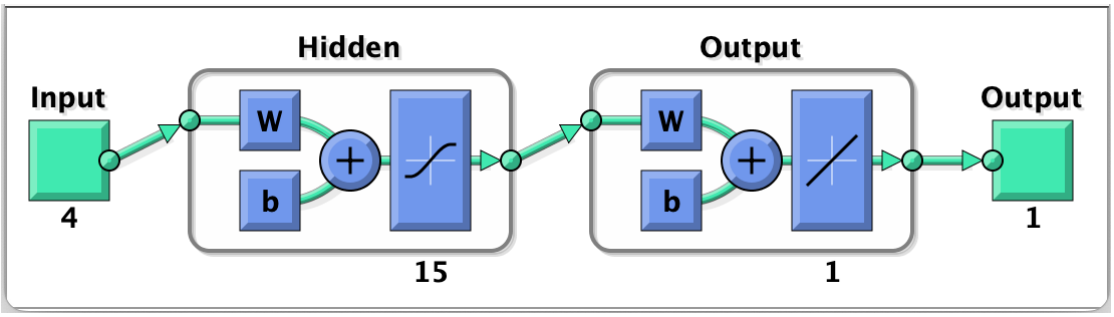

Fig. 4 (e). Conceptual diagram of the ANN model (15 hidden neurons)

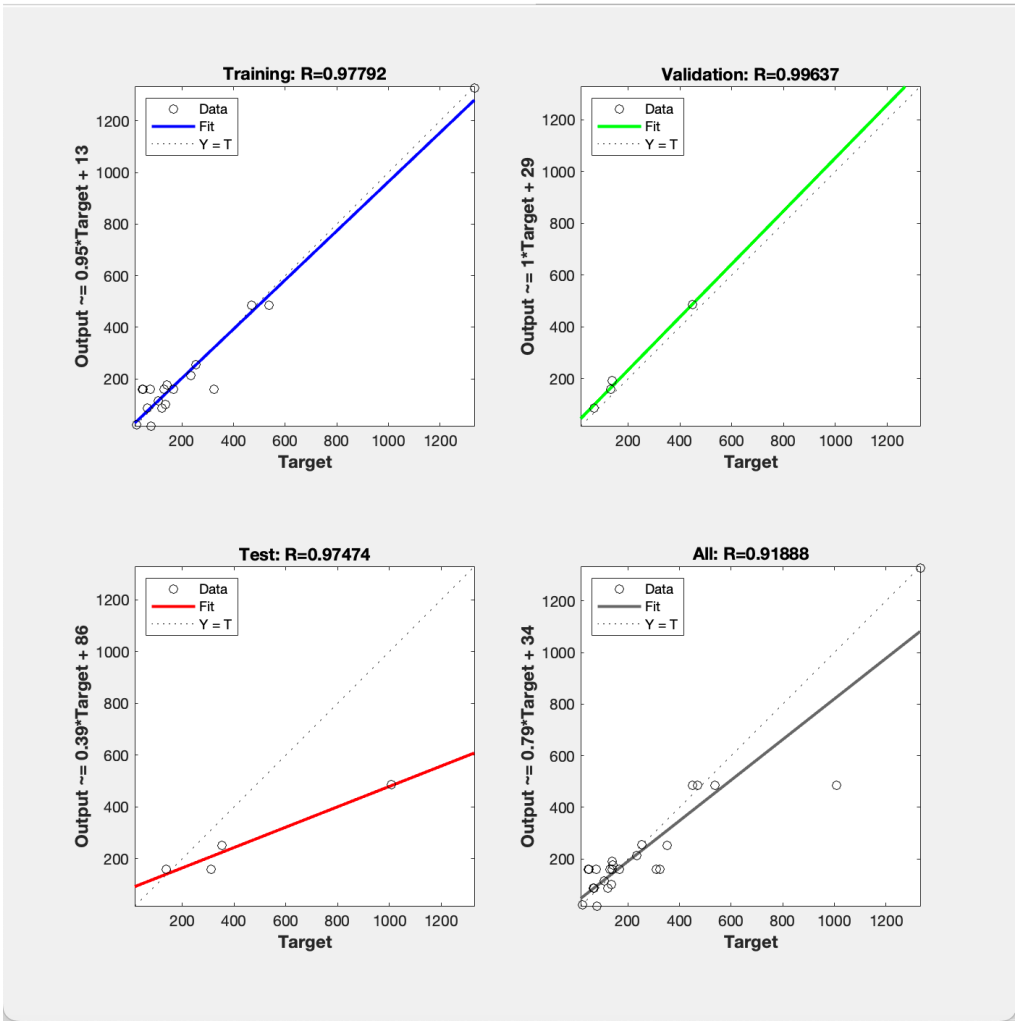

Fig. 4 (f). Regression diagram based on SCG model

Synthesizing the R^2 and MSE evaluations, and prioritizing the practical utility of the model for precise estimation, the Levenberg-Marquardt (LM) algorithm is identified as the overall most suitable model for accurately estimating photocatalytic hydrogen production from glycerol using TiO_2-based photocatalysts under visible light conditions. A model with a very high R^2 but a relatively higher MSE (like BR) might explain a large portion of the data's variance, yet its individual predictions could still

deviate considerably from the true values. Conversely, a model with a slightly lower R^2 but a substantially lower MSE (like LM) might explain marginally less overall variance but produces predictions that are, on average, much closer to the true values. Despite Bayesian Regularization (BR) demonstrating an exceptionally high R^2 on the testing set, indicative of excellent generalization and variance explanation, LM's Mean Squared Error on "All data" (2.49×10^3) was an order of magnitude lower than BR's (1.46×10^4). For engineering and scientific applications, minimizing the absolute magnitude of prediction errors, as reflected by MSE, is often paramount for practical implementation and process control. LM's consistently high R^2 across validation and testing, coupled with its significantly lower MSE, collectively suggest a more accurate and robust model overall. While BR remains a very strong contender and a close second, particularly in its generalization ability (high testing R^2), LM's superior performance in minimizing prediction error renders it the preferred choice for precise quantitative estimation. The SCG model was clearly outperformed by both LM and BR, exhibiting significant generalization issues and substantially higher errors. Table 3 shows all the specific information about how well each intelligent estimator works.

Table 3 melintang

3.2.2 Correlation analysis

When the model is trained using the Levenberg-Marquardt (LM) algorithm as shown is Figure 4(g), Light source power demonstrates a very weak direct relationship in the Pearson coefficient (0.1 to 0.19), while both Spearman and Kendall coefficients (0.2 to 0.39) indicate weak direct relationships. Catalyst loading presents consistently strong negative correlations, with the Pearson coefficient (below −0.8) classified as a very strong indirect relationship, the Spearman coefficient (−0.79 to −0.6) as a strong indirect relationship, and the Kendall coefficient (−0.59 to −0.4) as a moderate indirect relationship. Glycerol concentration shows a weak direct relationship in the Pearson coefficient (0.2 to 0.39), whereas Spearman and Kendall coefficients (−0.19 to 0) reflect very weak indirect relationships. For duration, the Pearson coefficient (around 0.3) suggests a weak direct relationship, while Spearman and Kendall coefficients (−0.19 to 0) again indicate very weak indirect relationships. These interpretations, following the classification ranges defined in Table 4, highlight the varying sensitivity of the LM model to each input parameter in predicting hydrogen production rate.

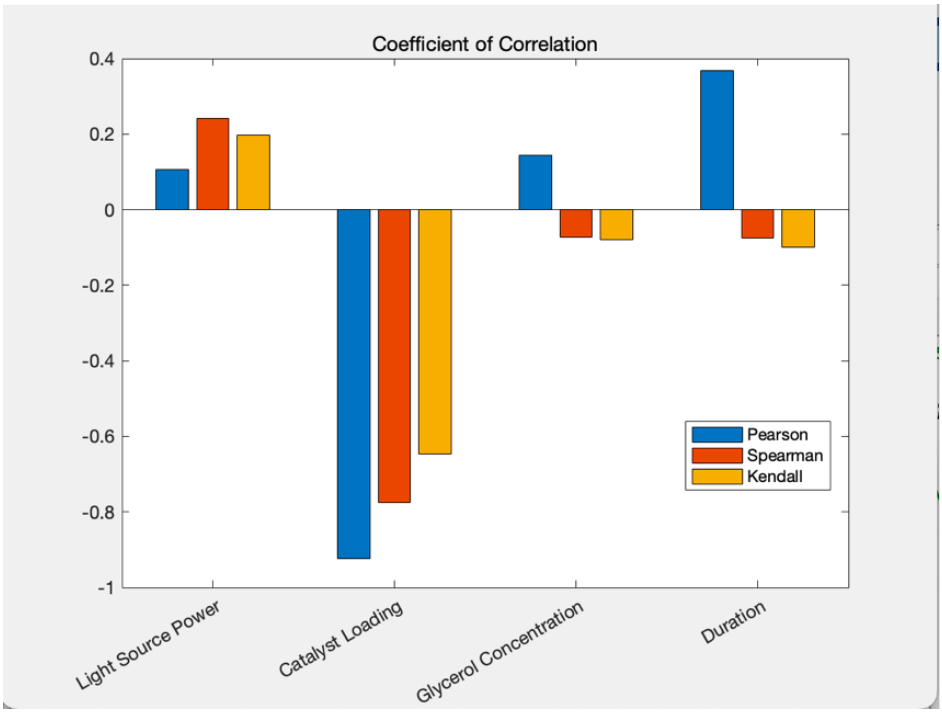

Fig. 4 (g). The relevancy between hydrogen production rate in photoreforming glycerol and light source power, catalyst loading, glycerol concentration and duration under Visible light source and LM model training

From Figure 4(h), it's clear that when the model is trained using the Bayesian Regularization (BR) model, the correlation of hydrogen production rate with the four input parameters shows some notable distinctions from the Levenberg-Marquardt (LM) model. For Light source power shows a very weak direct relationship according to the Pearson coefficient (0 to 0.19), while the Spearman and Kendall coefficients (−0.19 to 0) suggest very weak indirect relationships. Catalyst loading displays the strongest negative correlation, with the Pearson coefficient falling below −0.8, indicating a very strong indirect relationship, while the Spearman and Kendall coefficients (−0.79 to −0.6 and −0.59 to −0.4, respectively) reflect strong and moderate indirect relationships. Glycerol concentration demonstrates a weak direct relationship in the Pearson coefficient (0.2 to 0.39), but the Spearman and Kendall coefficients (−0.19 to 0) suggest very weak indirect correlations. Duration, on the other hand, exhibits the most notable positive influence, with the Pearson coefficient (0.4 to 0.59) indicating a moderate direct relationship, while the Spearman and Kendall coefficients (0 to 0.19) reflect very weak direct relationships.

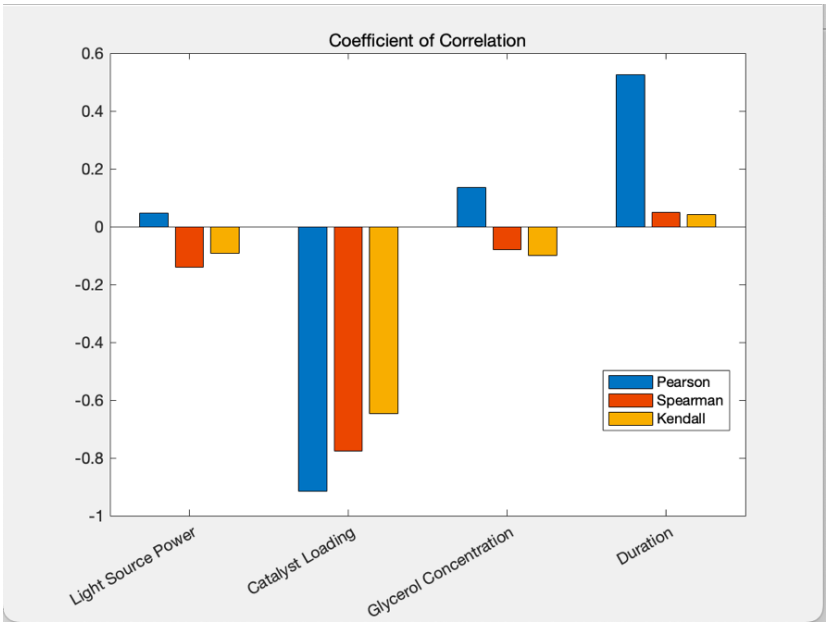

Fig. 4 (h). The relevancy between hydrogen production rate in photoreforming glycerol and light source power, catalyst loading, glycerol concentration and duration under Visible light source and BR model training

As shown in Figure 4(i), the Scaled Conjugate Gradient (SCG) model reveals a range of correlation strengths between the input variables and hydrogen production. Light source power is positively related to the output, with the Pearson coefficient indicating a very weak direct relationship (0 to 0.19), while the Spearman and Kendall coefficients point to weak direct relationships (0.2 to 0.39). Catalyst loading emerges as the most strongly negative factor, with Pearson showing a very strong indirect relationship (below −0.8), and Spearman and Kendall reflecting strong and moderate indirect relationships (−0.79 to −0.6 and −0.59 to −0.4, respectively). Glycerol concentration displays a weak direct relationship through the Pearson coefficient (0.2 to 0.39), though the Spearman and Kendall values (−0.19 to 0) suggest only very weak indirect effects. As for duration, the Pearson coefficient falls within the range of a moderate direct relationship (0.4 to 0.59), whereas the other two coefficients indicate very weak indirect relationships. These results suggest that while catalyst loading significantly hinders hydrogen production, the influence of other variables such as duration and light source power is more limited but generally positive.

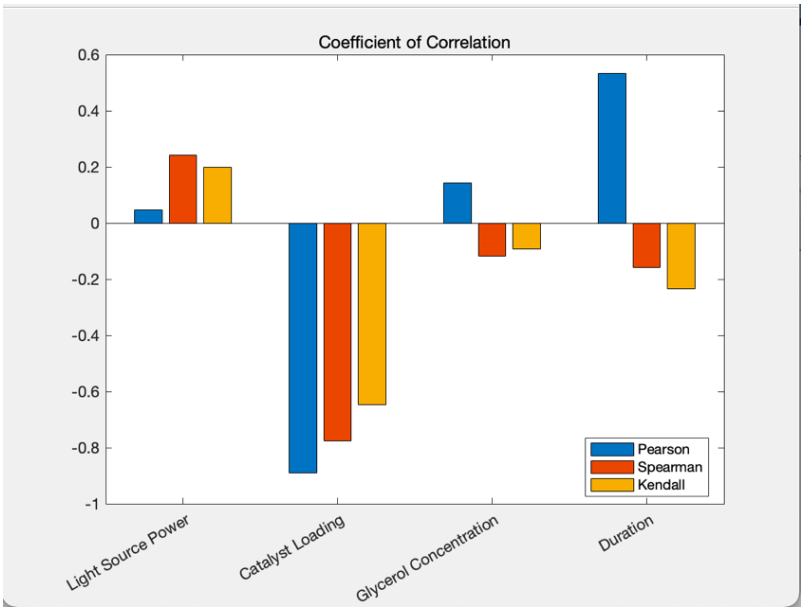

Fig. 4 (i). The relevancy between hydrogen production rate in photoreforming glycerol and light source power, catalyst loading, glycerol concentration and duration under Visible light source and SCG model training

Correlation analysis of key influencing factors revealed several important insights. Spearman correlation indicated a strong positive monotonic relationship between light source power and hydrogen yield, suggesting that increasing light intensity generally enhances hydrogen production, even if the relationship is not strictly linear. Pearson correlation showed that both irradiation time and glycerol concentration maintain a stable and positive linear relationship with hydrogen output within the observed range. When comparing these trends across the three models, namely LM, BR, and SCG, catalyst loading consistently emerges as the most significant and negatively correlated factor, indicating its critical role in limiting light penetration and reducing photocatalytic efficiency when used excessively.

The LM model demonstrated the best overall performance, capturing both linear and ordinal relationships with high accuracy. While light source power and glycerol concentration also show notable correlations in this model, especially under Kendall's ordinal metric, their impact is less pronounced compared to catalyst loading. This distinction becomes particularly clear under visible light conditions. Unlike ultraviolet light, visible light has lower photon energy and limited penetration capability. Catalyst loading shows a consistently strong negative correlation across all methods, with Pearson below −0.8 (very strong), Spearman between −0.79 and −0.6 (strong), and Kendall between −0.59 and −0.4 (moderate indirect relationship). As glycerol concentration increases, it further reduces light transmission through scattering and absorption, making catalyst loading the dominant factor influencing hydrogen production under visible light irradiation.

3.3 UV-Visible Light Source

3.3.1 Models evaluation

The selection of 27 hidden neurons suggests that the LM model benefits from a moderately complex network architecture (Figure 5a), which likely aids in capturing both primary linear trends and subtle nonlinear fluctuations.

As shown in the Figure 5(b), the LM model achieved R^2 values of 0.9257 for the training set, 0.8494 for the validation set, and 0.9308 for the testing set. The overall R^2 reached 0.9193, accompanied by a Mean Squared Error (MSE) of 9.44×10^6. These results reflect a stable and coherent performance across the datasets, with the model demonstrating particularly strong predictive accuracy on the unseen testing data. The slightly lower R^2 on the validation set could indicate limited flexibility in adapting to intermediate variations within the data, possibly a result of the LM algorithm's tendency to prioritize global fit over localized behavior. Nevertheless, the elevated R^2 on the testing set implies that the model was successful in identifying robust, generalizable relationships rather than overfitting to noise. Compared to the SCG model, the LM model's higher MSE points to the presence of localized prediction errors or sensitivity to specific data points, which may affect its precision in real-world applications.

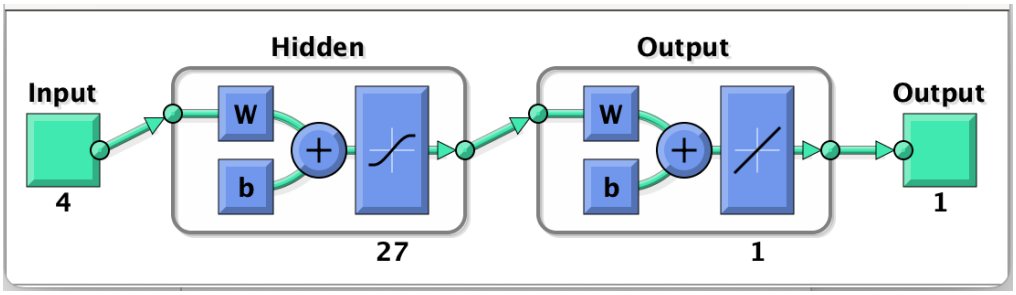

Fig. 5 (a). Conceptual diagram of the ANN model (27 hidden neurons)

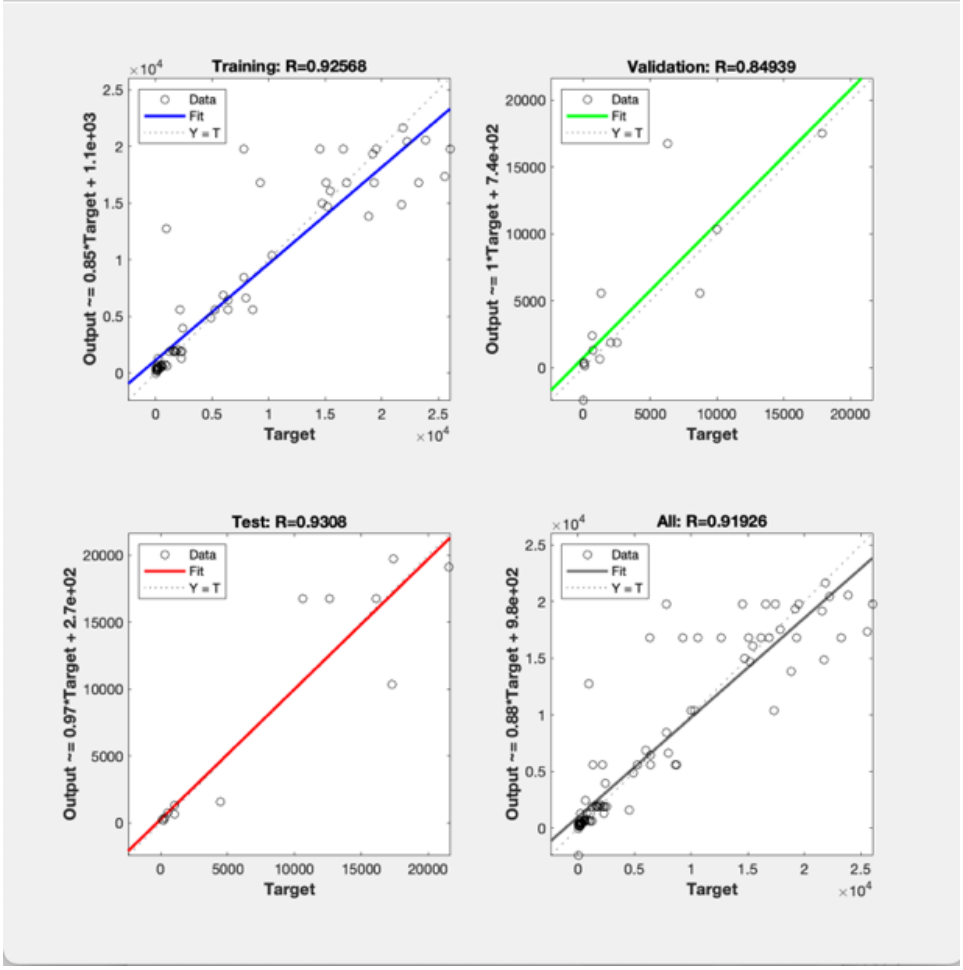

Fig. 5 (b). Regression diagram based on LM model

Figure 5(c) is the conceptual diagram of the BR ANN model (19 hidden neurons). The Figure 5(d) shows that the BR model achieved R^2 values of 0.9070 on the training set and 0.9416 on the testing set, with an overall R^2 of 0.9130. The associated Mean Squared Error (MSE) was 1.16×10^7, which is slightly higher than that of the LM model and significantly higher than that of SCG. Despite this higher error metric, the BR model produced the highest R^2 value on the testing set among all three models as trained by the process, underscoring its remarkable generalization capability. This strength is closely linked to the intrinsic characteristics of Bayesian regularization, which automatically adjusts model complexity by penalizing overfitting during the training process. The final architecture, comprising 19 hidden neurons, reflects an effective balance between model simplicity and representational capacity.

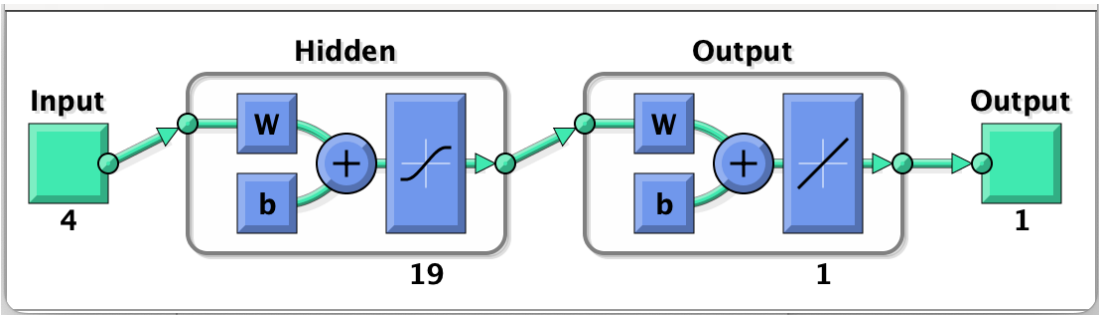

Fig. 5 (c). Conceptual diagram of the ANN model (19 hidden neurons)

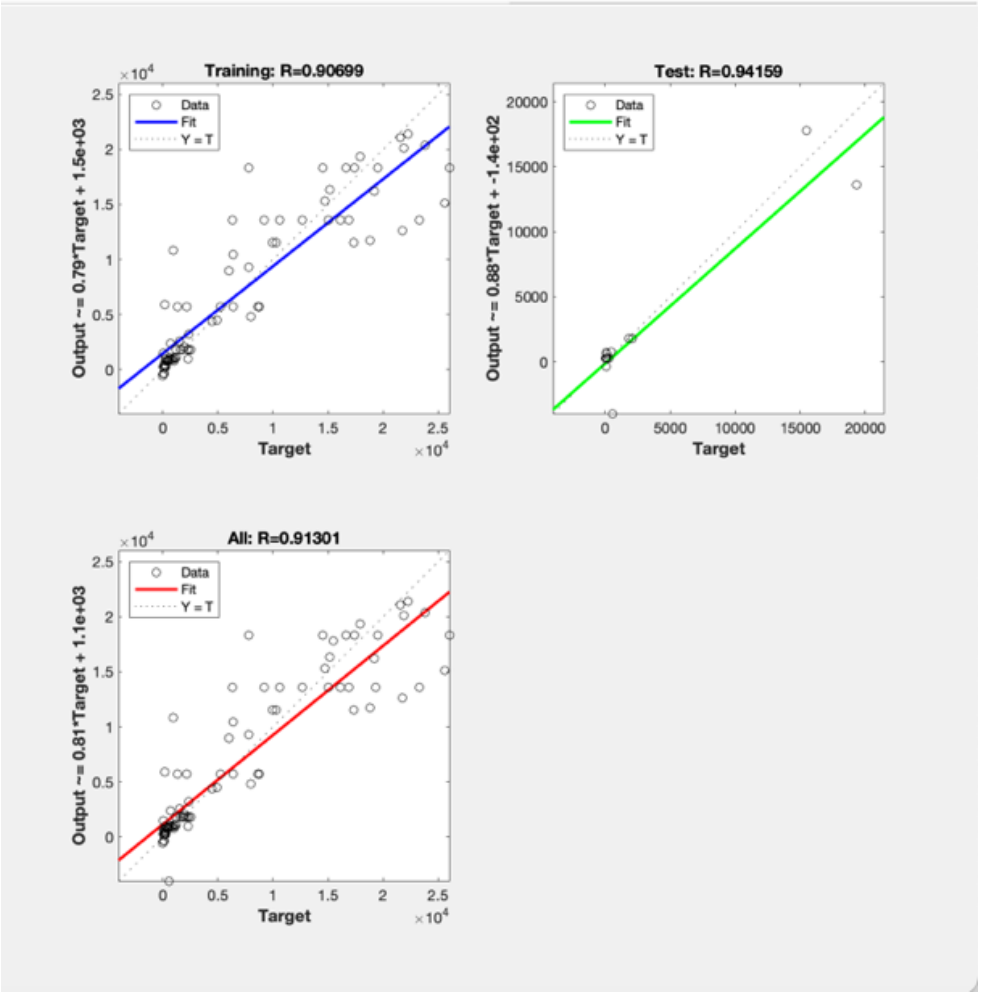

Fig. 5 (d). Regression diagram based on BR model

Figure 5(e) is the conceptual diagram of the SCG ANN model (25 hidden neurons). The SCG model presents a particularly interesting case. As shown in the Figure 5(f), the R^2 values were 0.9330 for the training set, 0.8938 for the validation set, and decreased notably to 0.8102 for the testing set. On the surface, this decline suggests a limitation in the model's ability to generalize. However, the overall Mean Squared Error (MSE) across all data was only 1.39×10^4, the lowest among the three models, indicating that the model maintained minimal average deviation from actual values. This apparent contradiction between a relatively low R^2 on the testing set and a very low MSE suggests that while the model's predictions were numerically close to the target values, it may have failed to replicate the exact trends or variance in specific regions of the testing data. The architecture, which employed 25 hidden neurons (Figure 5e), enabled the model to take advantage of SCG's fast convergence properties and reach a solution that minimizes the overall error effectively. However, this rapid optimization may have led to convergence toward a local minimum that fits the training data well but lacks flexibility when exposed to new patterns. This behaviour, where numerical accuracy is achieved at the expense of trend fidelity.

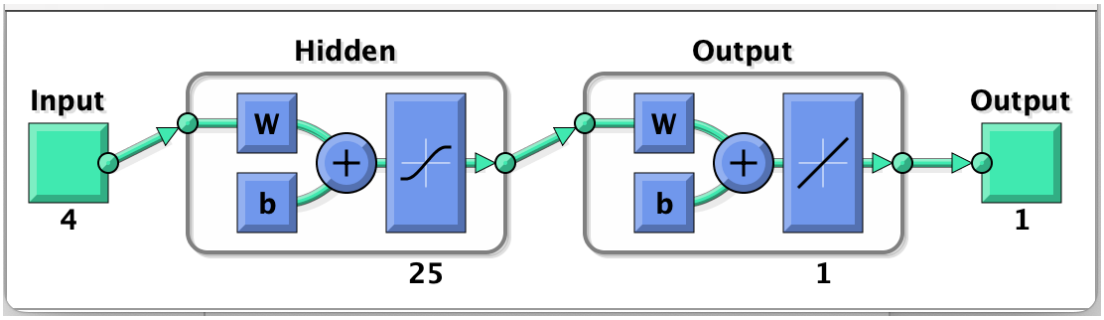

Fig. 5 (e). Conceptual diagram of the ANN model (25 hidden neurons)

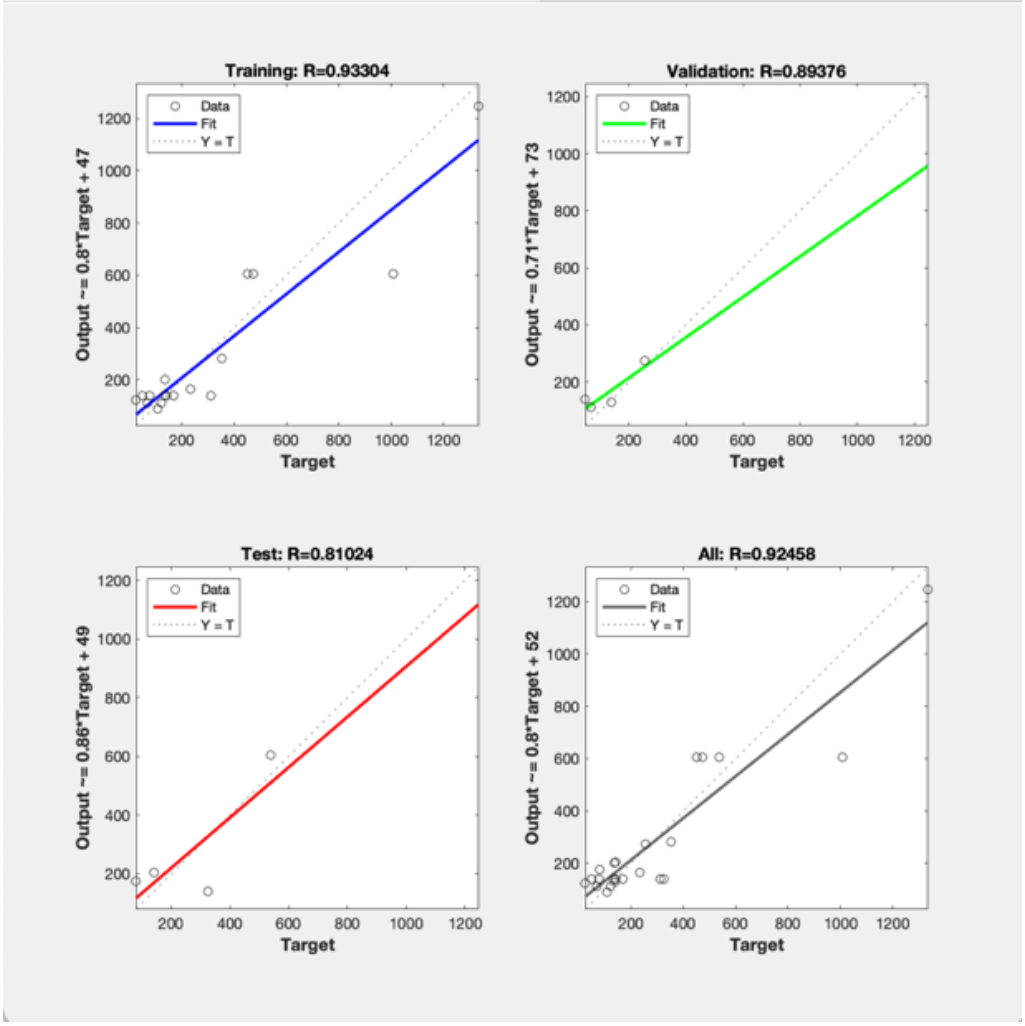

Fig. 5 (f). Regression diagram based on SCG model

The SCG algorithm did the best overall out of the three models tested in UV-visible light. It had the highest R^2 of 0.9245 and the lowest mean squared error (MSE) of 1.39×10^4 on the full dataset, which means it was more accurate and consistent than both LM and BR. The testing R^2 (0.8102) was lower than that of the other two models, but the MSE was so low that it suggests that SCG made very accurate numerical predictions across all data. The model also learnt well, with a training R^2 of 0.9330 and a validation R^2 of 0.8938. The 25 hidden neurons in its architecture made it easy for the model to converge while still being able to generalize. There is one study that modelled photocatalytic palm-oil-mill-effluent degradation, where the response spans 0–100 % and the best Gaussian-process model reported RMSE = 1.0394 with R^2 = 0.9962 [17]. Because the present outputs are several orders of magnitude larger than those percentage values, the corresponding squared errors naturally rise by similar orders, making an MSE on the order of 10^4 entirely reasonable even while R^2 remains above 0.92. These results according to the Table 4 show that SCG is the best and most reliable way to model the nonlinear dynamics of photocatalytic hydrogen production in the UV-visible range.

table 4 melintang

3.3.2 Correlation analysis

Under the LM model, hydrogen production negatively correlates significantly with increased catalyst loading and moderately with higher light source power, possibly due to issues like catalyst aggregation, excessive scattering of photons, or saturation phenomena at higher intensities, as illustrated explicitly in Figure 5(g). Light source power shows a very weak indirect relationship, with Pearson, Spearman, and Kendall coefficients all within the range of (−0.19 to 0). Catalyst loading demonstrates the strongest negative influence, with Pearson below (−0.8), indicating a very strong indirect relationship, Spearman between (−0.79 to −0.6) classified as strong, and Kendall within (−0.59 to −0.4) reflecting a moderate indirect relationship. Conversely, glycerol concentration strongly promotes hydrogen yield, reflecting ample substrate availability that facilitates continuous reaction cycles, supported by prior studies demonstrating glycerol's importance as a sustainable reactant [18]. Glycerol concentration displays a consistent positive effect, with Pearson in the range of (0.2 to 0.39) indicating a weak direct relationship, Spearman within (0.6 to 0.79) as a strong direct relationship, and Kendall in the range of (0.4 to 0.59) showing a moderate direct relationship. Duration exhibits weaker negative correlations, with Pearson between (−0.39 to −0.2) suggesting a weak indirect relationship, and both Spearman and Kendall falling within (−0.19 to 0), indicating very weak indirect relationships.

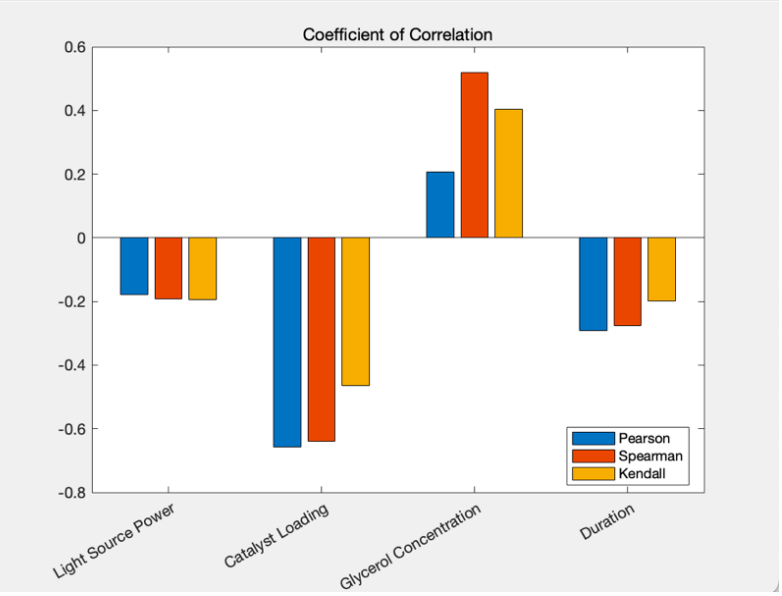

Fig. 5 (g). The relevancy between hydrogen production rate in photoreforming glycerol and light source power, catalyst loading, glycerol concentration and duration under UV-Visible light source and LM model training

Under UV-visible light conditions, the Bayesian Regularization (BR) model exhibits clear patterns (Figure 5h) in the correlation between input variables and hydrogen production. Light source power has minimal impact, as indicated by Pearson, Spearman, and Kendall coefficients all falling within the range of (−0.19 to 0), which corresponds to a very weak indirect relationship. In contrast, catalyst loading shows a consistently strong negative association across all three methods. The Pearson coefficient is below (−0.8), indicating a very strong indirect relationship, while Spearman and Kendall coefficients lie within (−0.79 to −0.6) and (−0.59 to −0.4), representing strong and moderate indirect relationships, respectively.

Glycerol concentration emerges as a positive contributor, with the Pearson coefficient in the range of (0.2 to 0.39) suggesting a weak direct relationship, the Spearman coefficient between (0.6 to 0.79) indicating a strong direct relationship, and the Kendall coefficient within (0.4 to 0.59) reflecting a moderate direct relationship. Duration shows a negative correlation, with Pearson and Spearman coefficients in the range of (−0.39 to −0.2), implying weak indirect relationships, and Kendall slightly weaker, within (−0.19 to 0), indicating a very weak indirect relationship. These findings reinforce that under UV-visible light conditions, catalyst loading continues to be the most influential limiting factor, while glycerol concentration has a more consistent and positive effect on hydrogen production rate.

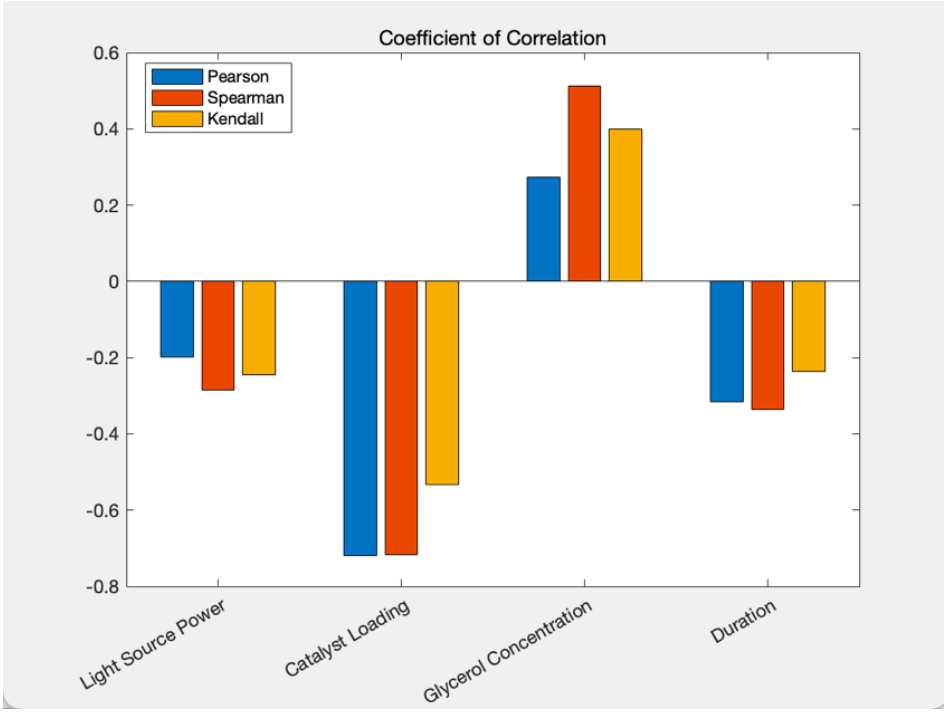

Fig. 5 (h). The relevancy between hydrogen production rate in photoreforming glycerol and light source power, catalyst loading, glycerol concentration and duration under UV-Visible light source and BR model training

In the SCG model (Figure 5i), Light source power shows minimal correlation with hydrogen production, as the Pearson, Spearman, and Kendall coefficients all fall within the range of (−0.19 to 0), indicating a very weak indirect relationship. Catalyst loading once again shows the strongest negative impact, with the Pearson coefficient falling below (−0.8), denoting a very strong indirect relationship. The Spearman coefficient lies within (−0.79 to −0.6), suggesting a strong indirect relationship, while the Kendall coefficient falls between (−0.59 to −0.4), indicating a moderate indirect relationship. Such selective parameter sensitivity aligns with earlier reports showing the paramount importance of optimized catalyst dispersion and adequate light irradiation durations [12]. Therefore, drastically minimizing catalyst loading and considerably extending reaction durations represent the core optimization strategies.

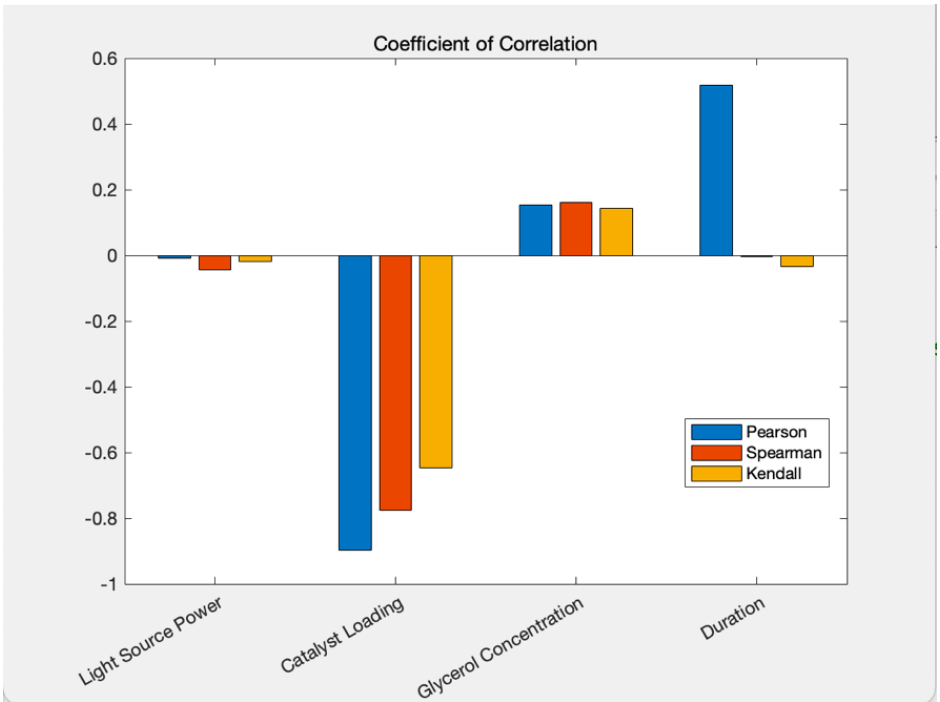

Fig. 5 (i). The relevancy between hydrogen production rate in photoreforming glycerol and light source power, catalyst loading, glycerol concentration and duration under UV-Visible light source and SCG model training

Under optimal conditions in the Scaled Conjugate Gradient (SCG) model, catalyst loading is identified as the most critical factor, as excessive amounts can hinder photocatalytic efficiency through particle agglomeration or site blockage. Duration plays a key positive role, where longer irradiation enhances hydrogen yield. In contrast, glycerol concentration and light source power show weak to very weak correlations but should still be maintained at moderate levels to ensure efficiency and avoid waste. Therefore, the optimal SCG strategy involves minimizing catalyst loading, extending reaction duration, and moderately regulating glycerol and light intensity.

The SCG model demonstrated the best overall predictive accuracy. In this model, catalyst loading again appears as the most significant limiting factor. The Pearson correlation coefficient falls below (−0.8), indicating a very strong indirect relationship. The Spearman coefficient lies within the range of (−0.79 to −0.6), suggesting a strong indirect relationship, while the Kendall coefficient falls between (−0.59 to −0.4), reflecting a moderate indirect relationship. These consistently strong negative correlations indicate that catalyst loading is the most important factor affecting hydrogen production under UV-Visible light conditions.

3.4 Effects of Inputs (Parameters) toward Hydrogen Production Rate
3.4.1 Light source power (W)

As shown in Figure 6, increasing light-source power generates a sharp rise in hydrogen-production rate at low intensities: moving from 80 W to 100 W elevates the median output by nearly twenty-fold. Beyond roughly 420 W the curve plateaus, with most observations clustering between 4×10^4 and 7×10^4 µmol g^{-1} h^{-1}, indicating progressive saturation of photo-active sites. Around 500 W the rate can even drop to about 1.4×10^4 µmol g^{-1} h^{-1} likely due to increased scattering, local heating, or mass-transfer constraints that offset the benefits of higher photon flux. These patterns align with TiO_2 photo physics: ultraviolet photons (≈ 3.1 eV at 400 nm) exceed the band gap, so additional power enhances electron–hole generation only until recombination and surface saturation impose kinetic

limits. Hence lamp power is confirmed as the most influential variable for hydrogen production under UV irradiation.

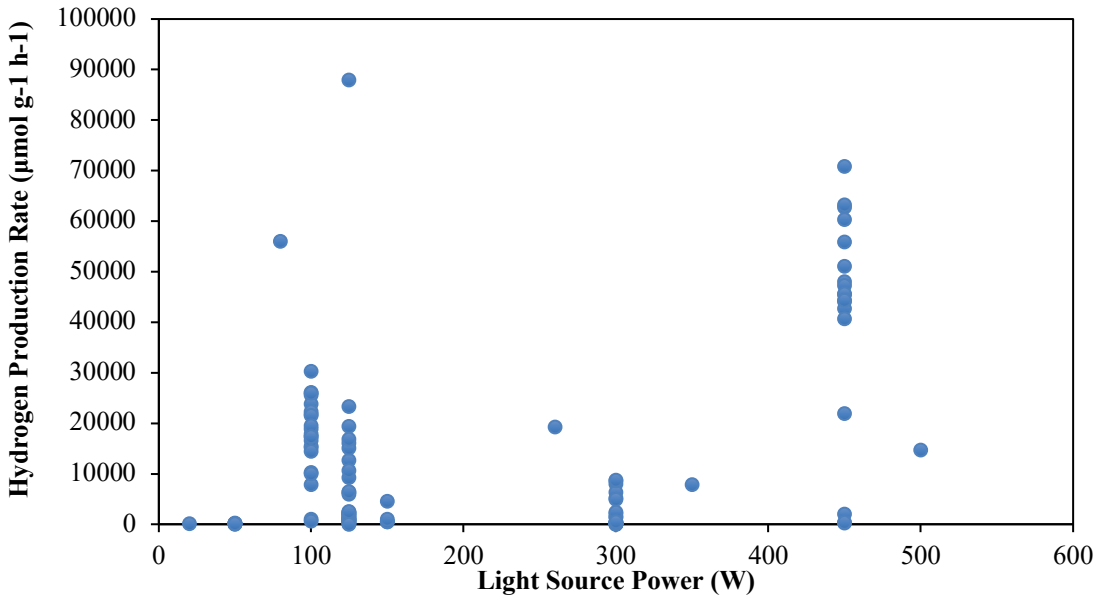

Fig. 6. Hydrogen production rate (µmol g^{-1} h^{-1}) versus light source power (W)

3.4.2 Catalyst loading (g/L)

Figure 7 shows that catalyst loading has a marked but decreasing effect on hydrogen-production rate. At very low loadings below about 0.2 g L^{-1} the suspension remains clear enough for photons to penetrate deeply, and rates can exceed 7 × 10^4 µmol g^{-1} h^{-1} Increasing the concentration to around 0.3 to 0.6 g L^{-1} initially maintains high activity because the additional surface area still outweighs the modest rise in turbidity. Once the loading approaches 1 g L^{-1} the data collapse toward only a few thousand µmol g^{-1} h^{-1} or even zero because extra solids scatter and absorb incident light, shorten photon paths, and favour particle agglomeration so much of the catalyst surface is no longer illuminated. The effect becomes more severe at 1.5 to 2.0 g L^{-1} where almost all points fall below one thousand µmol g^{-1} h^{-1}. Sections 4.2.2 and 4.3.2 confirm that under visible light and under combined UV–Visible light catalyst loading is the most influential negative factor, with very strong indirect correlations reported by every learning model. In these regimes the lower photon energy and shallower penetration depth of visible radiation strengthen the shielding effect: photons are absorbed in the upper slurry layer, electron–hole pairs recombine before they can reach reactive sites, and external mass transfer to deeper layers is restricted. Therefore, selecting a catalyst concentration that is below the optical saturation threshold is essential for maximizing hydrogen production in both visible and UV–visible light systems.

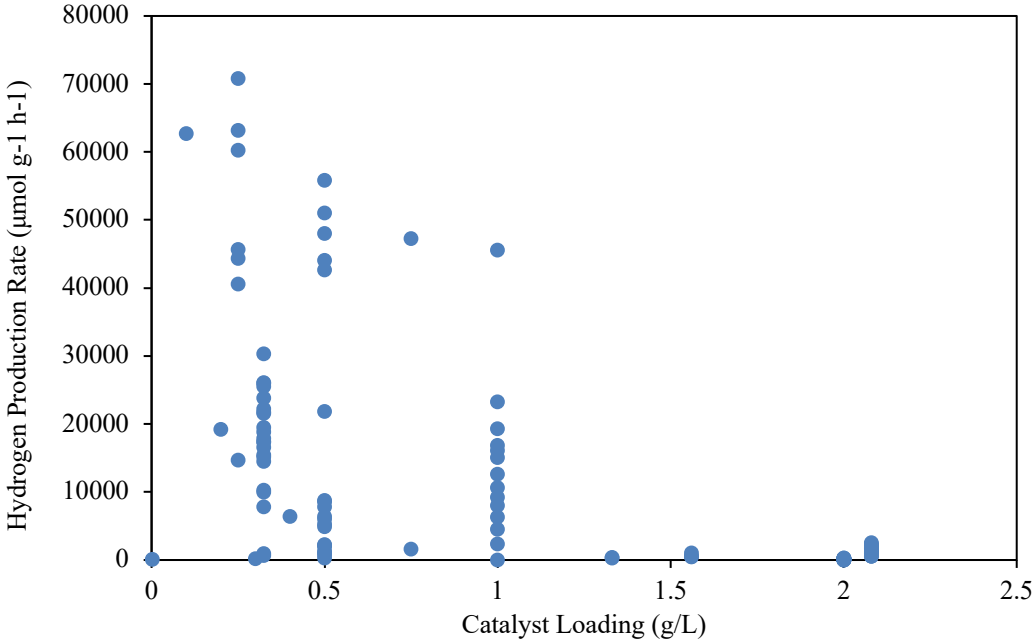

Fig. 7. Hydrogen production rate (µmol g^{-1} h^{-1}) versus catalyst loading (g/L)

3.4.3 Glycerol concentration (v/v%)

As shown in Figure 8, hydrogen-production rate rises sharply as glycerol concentration increases from trace levels to about 15–20 v/v%, where the highest values near 9×10^4 µmol g^{-1} h^{-1} are recorded. In this range glycerol acts as an effective hole scavenger: adsorbed molecules donate electrons rapidly after photo-oxidation, extend charge-carrier lifetimes, and generate strongly reducing conduction-band electrons that promote water reduction to H_2. In addition, glycerol oxidation is exergonic, so its intermediates withdraw photogenerated holes more readily than water, suppressing recombination and amplifying quantum efficiency. Beyond roughly 25 v/v%. The trend reverses and the rate declines progressively.

As glycerol concentration rises, the solution becomes more viscous, which lowers mass-transfer coefficients and restricts the diffusion of dissolved oxygen, protons, and reaction intermediates to and from the catalyst surface. A thicker glycerol layer also competes with water for adsorption sites and thereby limits access to surface hydroxyl groups required for hydrogen evolution [19]. In addition, concentrated glycerol absorbs a portion of the incident light and elevates the medium's refractive index, diminishing photon penetration and increasing internal scattering so that fewer photons arrive at the catalyst's reactive sites. At high concentrations, condensation of glyceryl radicals can produce oligomeric by-products that deposit on the catalyst and shield reactive facets [20]. Taken together, these factors account for the bell-shaped response curve and imply an optimum glycerol concentration that maximizes hole scavenging while minimizing optical and mass-transport losses.

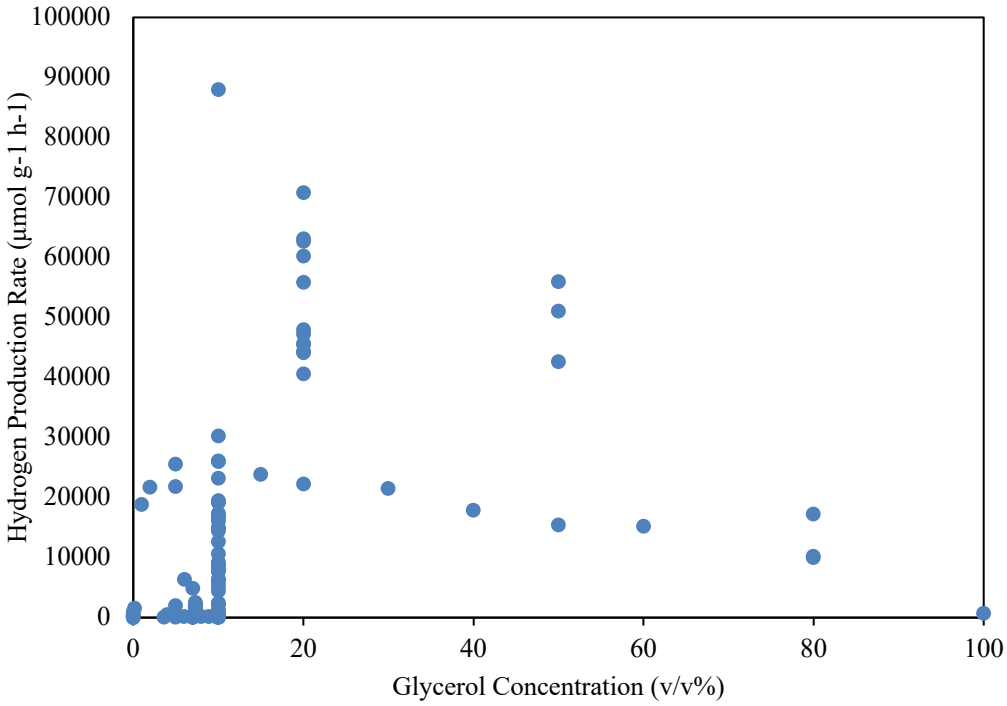

Fig. 8. Hydrogen production rate (µmol g^{-1} h^{-1}) versus glycerol concentration (v/v%)

3.4.4 Duration of light irradiation (h)

Figure 9 reveals that hydrogen-production rate is highest during the first few hours of irradiation then declines markedly as the experiment progresses. At short durations below about two hours the catalyst surface is fresh, glycerol is abundant, and little inhibitory by-product has formed, so photogenerated holes are efficiently scavenged and the specific rate can exceed 7×10^4 µmol g^{-1} h^{-1}. As irradiation extends toward five to ten hours several factors begin to suppress activity. Glycerol and dissolved oxygen are partly consumed, organic radicals and carbonaceous fragments adsorb onto TiO_2, and evolving hydrogen bubbles accumulate at the surface [21]. These changes lower reactant availability at active sites and increase charge-carrier recombination. Prolonged ultraviolet exposure also creates Ti^{3+} centers and oxygen vacancies that trap electrons and further diminish quantum efficiency [22]. After roughly twenty hours donor depletion, intermediate build-up, and surface fouling dominate, driving the rate below 1×10^3 µmol g^{-1} h^{-1}. By fifty hours the rate approaches zero, indicating near-complete passivation of reactive facets and exhaustion of readily oxidizable glycerol. The observed trend therefore reflects an early kinetic regime governed by intrinsic photocatalytic efficiency followed by a period limited by mass transfer, surface poisoning, and catalyst deactivation, highlighting the necessity of catalyst regeneration or sacrificial-agent replenishment for sustained hydrogen production.

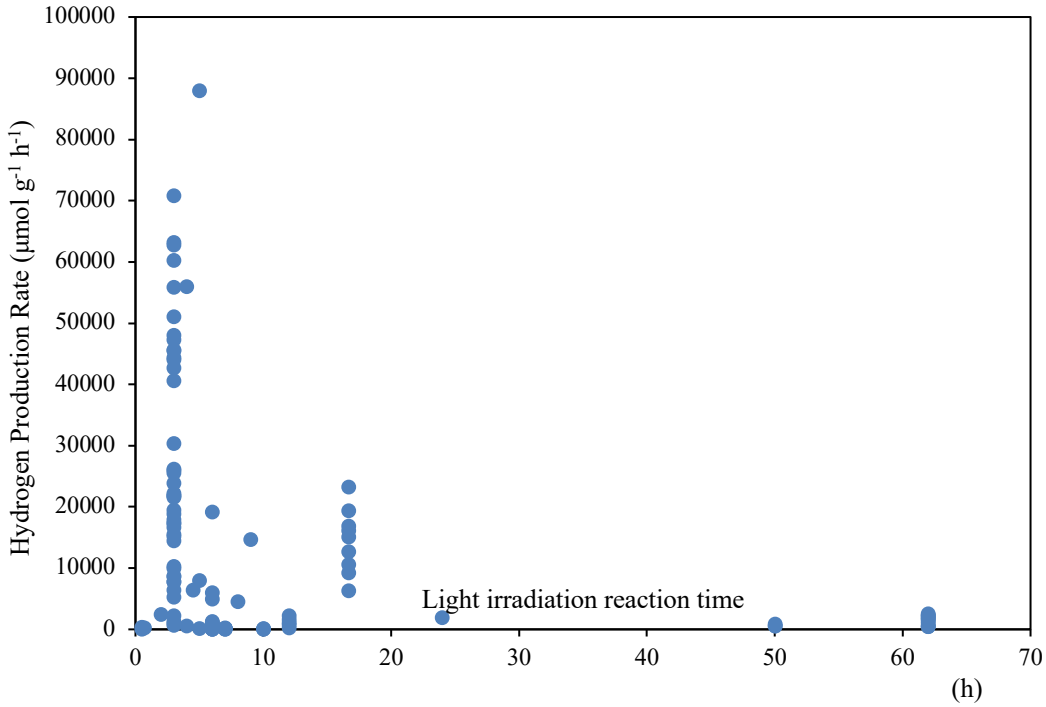

Fig. 9. Hydrogen production rate (µmol g^{-1} h^{-1}) versus light irradiation reaction time (h)

3.4.5 Optimal condition and sensitivity analysis

According to the Genetic Algorithm optimization results shown in Table 5, under UV light the optimal conditions are a catalyst amount of 3.87 g/L, glycerol concentration of 63.35 vol%, reaction time of 39.57 h, light power consumption of 236.00 W, and an H_2 reforming rate of 333282.27 μmol·$g^{-1}h^{-1}$. Under visible light, the optimal parameters are a catalyst amount of 0.56 g/L, glycerol concentration of 0.24 vol%, reaction time of 39.32 h, light power consumption of 450.00 W, and an H_2 reforming rate of 3803.93 μmol·$g^{-1}h^{-1}$. Under UV–visible light, the optimal conditions include a catalyst amount of 0.90 g/L, glycerol concentration of 1.63 vol%, reaction time of 57.14 h, light power consumption of 55.00 W, and an H_2 reforming rate of 3400.04 μmol·$g^{-1}h^{-1}$.

The sensitivity analysis results of the Garson algorithm further reveal the contribution of various factors under different light sources to the hydrogen production rate. As shown in Figure 10 (a), under ultraviolet light sources, the importance of light power consumption and glycerol concentration to the hydrogen production rate is 0.3936 and 0.3452 respectively, while the reaction time is relatively small, only 0.1081. This shows that under the condition of ultraviolet light source, the light power consumption not only directly provides the energy required for the reaction, but also interacts with the concentration of glycerol, which together improves the reaction rate. Unlike ultraviolet light, under visible light sources, according to Figure 10 (b), the importance of light power consumption and glycerol concentration is 0.2187 and 0.2364 respectively, while the catalyst loading is 0.2785, which became the most important factor affecting the hydrogen production rate. This shows that under visible light conditions, the number of surface-active sites of the catalyst becomes more important, which may be because visible light cannot effectively activate the catalyst, therefore, a larger catalyst loading is required to improve the reaction efficiency. Under UV-visible light conditions, the catalyst concentration is still the greatest, reaching 0.3016 which shown in Figure 10 (c). These differences show that under different light sources, the influence of light power consumption and catalyst loading on the hydrogen production rate is different, and the specific impact trend is closely related to the type of light source.

Based on the analyses of Genetic algorithm optimization results and Garson algorithm, an optimization strategy for different light source conditions has been derived: under ultraviolet light sources, increasing light power consumption can significantly improve the hydrogen production rate; under visible light sources, increasing catalyst loading is the key to improving hydrogen production efficiency; under UV-visible light conditions, glycerol concentration has the greatest impact on the reaction rate. Through these optimization strategies, effective operating conditions can be provided for the photocatalytic hydrogen production process.

Table 5 melintang

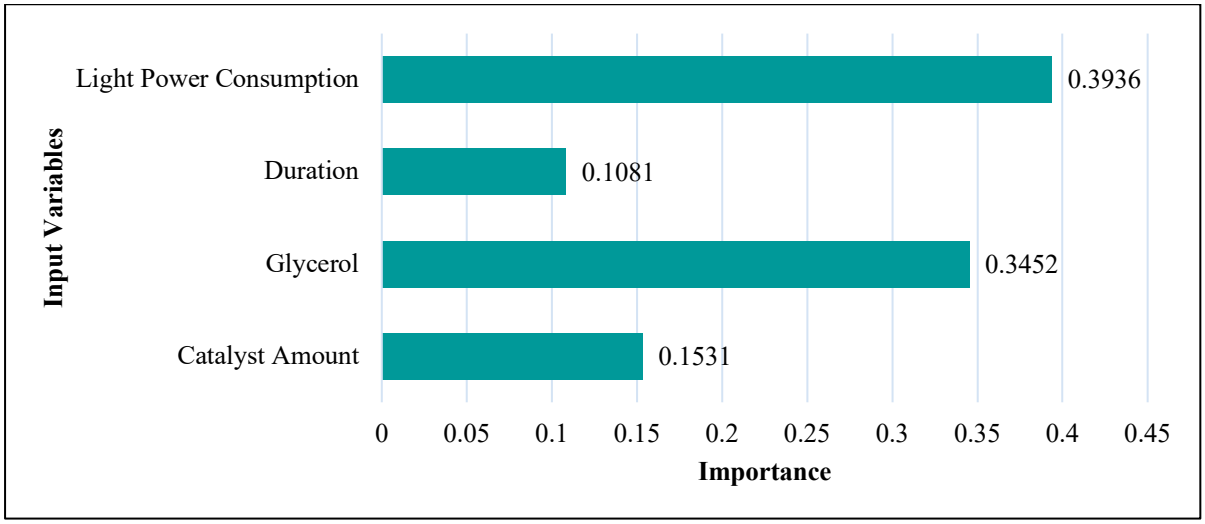

Fig. 10 (a). Influence of input variables on hydrogen production rate under UV range

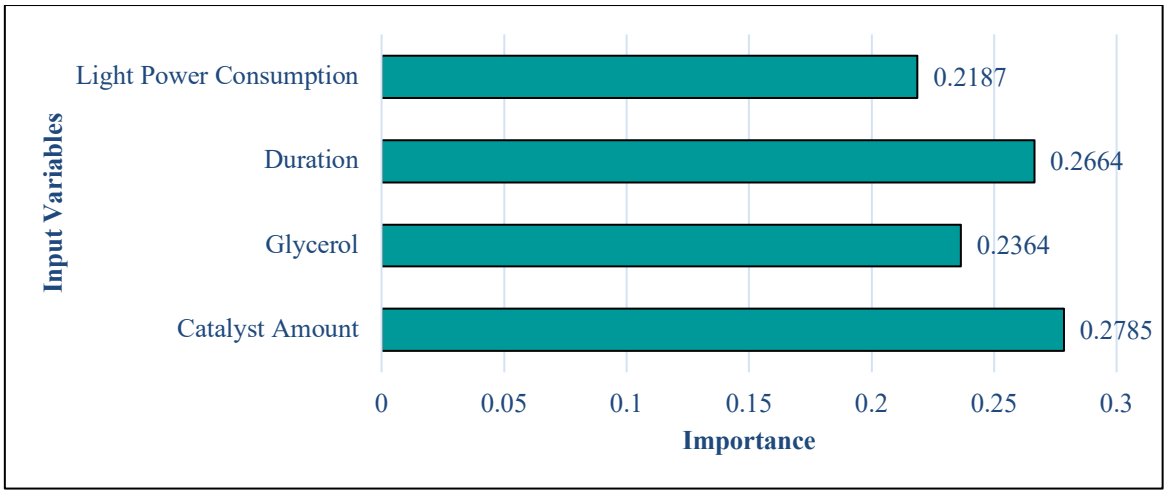

Fig. 10 (b). Influence of input variables on hydrogen production rate under visible range

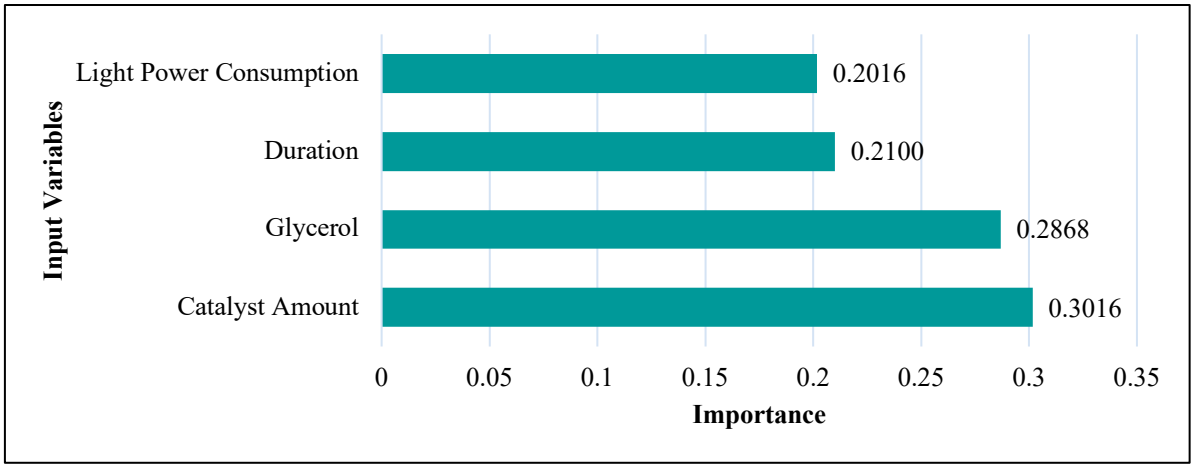

Fig. 10 (c). Influence of input variables on hydrogen production rate under UV-visible range

4. Conclusions

This study explored the application of machine learning techniques for estimating photocatalytic hydrogen production rate from glycerol using TiO_2-based photocatalysts. Specifically, three artificial neural network algorithms were investigated, namely Bayesian Regularization (BR), Linear Model (LM), and Scaled Conjugate Gradient (SCG). These models were developed and evaluated under three illumination conditions: ultraviolet (UV), visible (Vis), and combined ultraviolet visible (UV-Vis) light. Under UV light source conditions, the BR model achieved the best performance with the lowest mean squared error (MSE) of 2.39×10^7 and the highest coefficient of determination (R^2) of 0.9629, indicating strong adaptability. Under Visible light source conditions, the LM model performed best with an MSE of 2.96×10^3 and an R^2 of 0.9194. Under UV Vis conditions, the SCG model delivered the highest accuracy with an MSE of 1.39×10^4 and an R^2 of 0.9245. These results achieved a high prediction accuracy with R^2 values close to 1 and relatively low error metric. The larger MSE values arise from the much higher hydrogen-production scale in this study. Squaring residuals at such magnitudes amplifies the error, so normalizing or scaling the outputs would yield MSE figures that are directly comparable to those in the literature without changing the actual prediction quality. The correlation analyses were conducted to explore the relationships between key process parameters and hydrogen production rate. Under UV light conditions, high photon energy enhances photocatalytic activity, making light source power the dominant influencing factor. All three models, LM, BR, and SCG, support this finding. In the BR model, light source power shows a powerful direct relationship in Pearson (0.8 to 1), a weak direct relationship in Spearman (0.2 to 0.39), and a very weak direct relationship in Kendall (0 to 0.19). Glycerol concentration also shows a positive trend, though excessive levels may reduce light penetration. Under visible light conditions, catalyst loading is the most critical factor due to lower photon energy and limited penetration. All models report consistently negative correlations: a very strong indirect relationship in Pearson (below −0.8), a strong indirect relationship in Spearman (−0.79 to −0.6), and a moderate indirect relationship in Kendall (−0.59 to −0.4). While light source power and glycerol concentration remain positively correlated, their influence is weaker. In the LM model, which performed best under visible light, showing the glycerol concentration increases light scattering, further reinforcing the importance of reducing catalyst loading. Under ultraviolet visible light conditions, the SCG model showed the highest predictive accuracy. The catalyst loading again appears as the most influential factor, with a very strong indirect relationship in Pearson (below −0.8), a strong indirect relationship in Spearman (−0.79 to −0.6), and a moderate indirect relationship in Kendall (−0.59 to −0.4). Duration shows a

moderate direct relationship in Pearson (0.4 to 0.59), while Spearman and Kendall indicate very weak indirect relationships (−0.19 to 0). The glycerol concentration and light source power also show weak to very weak correlations across all methods. These results confirmed the catalyst loading as the most important factor limiting hydrogen production rate under ultraviolet visible light conditions.

Acknowledgements
The work is supported by Xiamen University Malaysia Research Fund (XMUMRF/2025-C15/IENG/0077).

References

[1] Kosamia, Niravkumar Mahendrasinh, Mahdieh Samavi, Bijaya Kumar Uprety, and Sudip Kumar Rakshit. "Valorization of biodiesel byproduct crude glycerol for the production of bioenergy and biochemicals." *Catalysts* 10, no. 6 (2020): 609. https://doi.org/10.3390/catal10060609

[2] Bansod, Yash, Benjamin Crabbe, Luke Forster, Kamran Ghasemzadeh, and Carmine D'agostino. "Evaluating the environmental impact of crude glycerol purification derived from biodiesel production: A comparative life cycle assessment study." *Journal of Cleaner Production* 437 (2024): 140485. https://doi.org/10.1016/j.jclepro.2023.140485

[3] Kolonnage, Thamasha Navodya, and Irene Mei Leng Chew. "Environmental and cost analysis for polyhydroxyalkanoate production from glycerol byproduct: A case study from integrated soap and biodiesel plants." *Cleaner Environmental Systems* 9 (2023): 100123. https://doi.org/10.1016/j.cesys.2023.100123

[4] Escamilla, Juan Carlos, Jesús Hidalgo-Carrillo, Juan Martín-Gómez, Rafael C. Estévez-Toledano, Vicente Montes, Daniel Cosano, Francisco J. Urbano, and Alberto Marinas. "Hydrogen production through glycerol photoreforming on TiO2/mesoporous carbon: influence of the synthetic method." *Materials* 13, no. 17 (2020): 3800. https://doi.org/10.3390/MA13173800

[5] Abhishek, B., A. Jayarama, S. Nagarkar Sanjog, P. Duttagupta Siddhartha, and S. Prabhu Sriganesh. "Challenges in photocatalytic hydrogen evolution: Importance of photocatalysts and photocatalytic reactors." *International Journal of Hydrogen Energy* 81 (2024): 1442-1466. https://doi.org/10.1016/j.ijhydene.2024.07.262

[6] Jia, Jia, Qiqi Zhang, Keke Li, Yating Zhang, Enzhou Liu, and Xin Li. "Recent advances on g–C3N4–based Z-scheme photocatalysts: Structural design and photocatalytic applications." *International Journal of Hydrogen Energy* 48, no. 1 (2023): 196-231. https://doi.org/10.1016/j.ijhydene.2022.09.272

[7] Almeida, Eugênia Leandro, José Eduardo Olivo, and Cid Marcos Gonçalves Andrade. "Production of biofuels from glycerol from the biodiesel production process—a brief review." *Fermentation* 9, no. 10 (2023): 869. https://doi.org/10.3390/fermentation9100869

[8] Garson, G. David. "A comparison of neural network and expert systems algorithms with common multivariate procedures for analysis of social science data." *Social Science Computer Review* 9, no. 3 (1991): 399-434.

[9] Nu, Qadir, and S. M. Smith. "Direct Differentiation Based Hessian Formulation for Training Multilayer Feed forward Neural Networks using the LM Algorithm-Performance Comparison with Conventional Jacobian-Based Learning." *Global J Technol Optim* 9 (2018): 223.. https://doi.org/10.4172/2229-8711.1000223

[10] Zhu, Xiaolei, Marzieh Khosravi, Behzad Vaferi, Menad Nait Amar, Mohammed Abdelfetah Ghriga, and Adil Hussein Mohammed. "Application of machine learning methods for estimating and comparing the sulfur dioxide absorption capacity of a variety of deep eutectic solvents." *Journal of Cleaner Production* 363 (2022): 132465. https://doi.org/10.1016/j.jclepro.2022.132465

[11] Kanthasamy, Ramesh, Imtiaz Ali, Bamidele Victor Ayodele, and Hisham A. Maddah. "Bio-hydrogen production from the photocatalytic conversion of wastewater: Parametric analysis and data-driven modelling using nonlinear autoregressive with exogeneous input and back-propagated multilayer perceptron neural networks." *Fuel* 344 (2023): 128026.https://doi.org/10.1016/j.fuel.2023.128026

[12] Han, Ke, and Yawei Wang. "A review of artificial neural network techniques for environmental issues prediction." *Journal of Thermal Analysis and Calorimetry* 145, no. 4 (2021): 2191-2207. https://doi.org/10.1007/s10973-021-10748-9

[13] Fidalgo, Annita, Sonia Letichevsky, and Brunno F. Santos. "Assessment of TiO2 band gap from structural parameters using artificial neural networks." *Journal of Photochemistry and Photobiology A: Chemistry* 405 (2021): 112870.https://doi.org/10.1016/j.jphotochem.2020.112870

[14] Essam, F., Hashash El, and Shiekh Raga Hassan Ali. "A comparison of the pearson, spearman rank and kendall tau correlation coefficients using quantitative variables." *Asian J. Probab. Stat* 20, no. 3 (2022): 36-48. https://doi.org/10.9734/ajpas/2022/v20i3425

[15] Roy, Rajib Baran, Md Rokonuzzaman, Nowshad Amin, Mahmuda Khatun Mishu, Sanath Alahakoon, Saifur Rahman, Nadarajah Mithulananthan, Kazi Sajedur Rahman, Mohammad Shakeri, and Jagadeesh Pasupuleti. "A comparative performance analysis of ANN algorithms for MPPT energy harvesting in solar PV system." *IEEE Access* 9 (2021): 102137-102152. https://doi.org/10.1109/ACCESS.2021.3096864

[16] Ayodele, Bamidele Victor, May Ali Alsaffar, Siti Indati Mustapa, and Dai-Viet N. Vo. "Backpropagation neural networks modelling of photocatalytic degradation of organic pollutants using TiO2-based photocatalysts." *Journal of Chemical Technology & Biotechnology* 95, no. 10 (2020): 2739-2749. https://doi.org/10.1002/jctb.6407

[17] Tabatabai-Yazdi, Fatemeh-Sadat, Azadeh Ebrahimian Pirbazari, Fatemeh Esmaeili Khalil Saraei, and Neda Gilani. "Construction of graphene based photocatalysts for photocatalytic degradation of organic pollutant and modeling using artificial intelligence techniques." *Physica B: Condensed Matter* 608 (2021): 412869.https://doi.org/10.1016/j.physb.2021.412869

[18] Qureshi, Fazil, Mohammad Yusuf, Amjad Ali Pasha, Huma Warsi Khan, Binash Imteyaz, and Kashif Irshad. "Sustainable and energy efficient hydrogen production via glycerol reforming techniques: A review." *International Journal of Hydrogen Energy* 47, no. 98 (2022): 41397-41420. https://doi.org/10.1016/j.ijhydene.2022.04.010

[19] Li, Junqing, Ying Tu, Kelin He, Chao Chen, Lixing Liang, Chongze Ruan, and Qitao Zhang. "Mechanistic insights into glycerol oxidation to high-value chemicals via metal-based catalysts." *Molecules* 30, no. 6 (2025): 1310. https://doi.org/10.3390/molecules30061310

[20] Kuo, Hsuan-Hung, Truong-Giang Vo, and Yung-Jung Hsu. "From sunlight to valuable molecules: A journey through photocatalytic and photoelectrochemical glycerol oxidation towards valuable chemical products." *Journal of Photochemistry and Photobiology C: Photochemistry Reviews* 58 (2024): 100649. https://doi.org/10.1016/j.jphotochemrev.2023.100649

[21] Nosaka, Yoshio, and Atsuko Nosaka. "Understanding hydroxyl radical (• OH) generation processes in photocatalysis." *ACS Energy Letters* 1, no. 2 (2016): 356-359. https://doi.org/10.1021/acsenergylett.6b00174

[22] Amano, Fumiaki, Masashi Nakata, Akira Yamamoto, and Tsunehiro Tanaka. "Effect of Ti3+ ions and conduction band electrons on photocatalytic and photoelectrochemical activity of rutile titania for water oxidation." *The Journal of Physical Chemistry C* 120, no. 12 (2016): 6467-6474. https://doi.org/10.1021/acs.jpcc.6b01481