Benefits from Using an Artificial Neural Network as a Prediction Model for Bio-hydrogen Production

WALID M. ALALAYAH, YAHIA ALHAMED, ABDULRAHIM AL-ZAHRAI, GABER EDRIS, HAMAD A. AL-TURAIF

King Abdulaziz University KAU, College of Engineering, Chemical and Materials Engineering Department, P.O. Box 80204, Jeddah 21589 Saudi Arabia

The performance of the glucose-based production of H2 in a batch reactor was predicted by an artificial neural network (ANN). The potential of utilizing an ANN modeling approach to simulate and predict the hydrogen production of Clostridium saccharoperbutylicum N1-4 (ATCC 13564) was investigated. Sixty experimental data records have been utilized to develop the ANN model. In this paper, a unique architecture has been introduced to mimic the inter-relationship between three input parameters: initial substrate concentration, initial medium pH and temperature (10 g/l, 6.0 ± 0.2, 37°C, respectively). A comparative analysis with a traditional Box-Wilson Design (BWD) statistical model proved that the ANN model output significantly outperformed the BWD model at similar experimental conditions. The results showed that the ANN model provides a higher level of accuracy for the H2 prediction and fewer errors and that it overcomes the limitation of the BWD approach with respect to the number of records, which merely considers a limited length of stochastic patterns for H2 prediction.

Keywords: Hydrogen production, anaerobic fermentation, bioprocess modeling, artificial neural network model.

Hydrogen is one alternative fuel available to meet our energy requirements, and its consumption as a fuel is completely devoid of carbon dioxide emissions, which distinguishes it from conventional fuels. Hydrogen can be used as a clean transportation fuel and to produce electricity via fuel cells. Hydrogen gas can be produced by a chemical process, but the use of anaerobic microorganisms to produce hydrogen from biomass has been declared an innovative and promising biotechnology [1]. Biomass-based hydrogen production technologies are still under development; hence, there are limited plant and operation data that are available at this point. It is, therefore, difficult for policy makers to identify promising technologies and the advantages and disadvantages of each technology [2]. Bio-hydrogen is a sustainable energy resource due to its non-polluting nature and high energy density. The principal energy options in the transport sector are ‘green’ electric energy for electric vehicles, bio-fuels and bio-hydrogen [3]. Fermentative hydrogen production can contribute to both waste reduction and energy production by using organic wastes as the substrate [4-6]. Many factors, such as temperature, initial pH and substrate concentrations, can influence the fermentative hydrogen production, as these factors can affect the activity of essential enzymes such as hydrogenases, which changes the activity of the hydrogen-producing bacteria [7,8]. The H2 yield from microbial fermentation is dependent upon the microbial communities present, type of substrate and operational and environmental factors (e.g., the organic loading rate, initial pH, or temperature) [9]. A traditional model had been developed to predict the hydrogen yield utilizing Box–Wilson design (BWD) approach by [10]. In fact, BWD approach could provide acceptable level of accuracy for predicting the hydrogen yield; however, BWD approach has some limitations. It could be applied for limited records of the available data set due to its mathematical procedure. As a result, it is still required to investigate other methods that could provide robust model that mimic the pattern and generalized the hydrogen yield with respect to the input pattern. Many systematic approaches have been introduced to facilitate the investigations regarding the influence of these parameters on the production yield. Significant progress in the field of nonlinear pattern recognition and system control theory has made advances in a branch of nonlinear system theoretic modeling called artificial neural network (ANN). The quantification and prediction of the hydrogen yield (HY) for any experimental approach is considered a highly stochastic process and with dynamics that could be linear or non-linear mathematical procedure experiences a linear/non-linear mathematical procedure [11,12]. ANN models have been used successfully to model complex nonlinear input-output time series relationships in a wide variety of fields [13]. ANN has been used extensively by scientists in the fields of science and engineering; it is popular as an effective and efficient way to model the hydrogen production process. Applying ANN prediction will help in investigating the effects of factors that affect on the production. The main objective of this work is to construct a suitable model to predict the hydrogen production by observing variables, such as initial glucose concentration, initial medium pH and reaction temperature. Few references in the field of bio-hydrogen production for energy that feature an application of the ANN technique with an acceptable accuracy, most notably using Clostridium saccharoperbutylacetonicum N1-4 (ATCC 13564; CSN1-4) using dark fermentation. This study was developed as a mathematical model for the previous work, which used CSN1-4 [1]. Sixty experimental runs were used for the ANN and compared with a Box-Wilson Design, which could not predict the output for more than 15 runs.

* email: wulalayah@kau.edu.sa
Experimental part
Materials and methods
Microbial strain and Culture media

CSN1-4 culture stock was obtained from a culture collection maintained at the Chemical Engineering Department, UKM and reported previously by [1, 14]. A solution of 15% PG medium per litre of distilled water was used as a growth medium for the inoculum. This medium was incubated in boiling water for one hour and then filtered through cotton cloth. The filtrate was sterilized in an autoclave at 121°C for 15 minutes. TYA medium was used for the preculture as well as main culture, and the composition of this medium per litre of distilled water was 40 g glucose, 2 g yeast extract, 6 g Bacto-Tryptone, 3 g ammonium acetate; 10 mg FeSO4·7H2O, 0.5 g KH2PO4, and 0.3 g MgSO4·7H2O per litre of distilled water as reported by [14; 15].

Box-Wilson design (BWD)

BWD is a technique to investigate the impact of the experimental variables on the response (output) that uses central composite design (CCD) to create a response surface that is commonly chosen for response optimization and depends on the values of the variables, as mentioned [19]. The experimental data were used to build a second-order polynomial mathematical model by a regression method. This mathematical model was taken as the objective function and was optimized using statistical software 7.0. The BWD technique can effectively be used with three variables to relate the inputs of temperature, substrate concentration, and initial glucose concentration to the output activity Y based on a selected transfer function. The growth of microorganisms can be modeled by the Monod equation.

\[
\mu = \frac{\mu_{\text{max}} \cdot S}{K_s + S}
\]

where \(\mu\) is the specific growth rate, \(\mu_{\text{max}}\) is the maximum specific growth rate, \(K_s\) is the saturation constant and \(S\) is the limiting substrate concentration. The values of \(\mu_{\text{max}}\) and \(K_s\) are usually estimated following the Monod model by regression analysis of the Line weaver-Burk linearized equation as reported by [15,19].

Hydrogen yield prediction using ANN

It is very important to predict \(H_2\) production using a comprehensive model for the design, monitoring and management of bio-hydrogen producing bioreactors. There are several works regarding the control of complex bioprocess and biosystems in environmental and industrial applications [11]. This work examined the development of a predictive ANN model for \(H_2\) production and compared it with BWD. It has been proven for ANN that a network of two layers that utilizes a sigmoid and a linear transfer function in its first and second layers, respectively, can be trained to model any non-linear relation as reported by [23]. The network model of the multi-layer perceptron architecture is based on units, which compute a non-linear function of the scalar product of the input vector and the weight vector. An alternative architecture of ANN is one in which the distance between the input vector and a certain weight vector. An alternative architecture of ANN is one in which the distance between the input vector and a certain weight vector.

\[
y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_1^2 + a_5x_1x_2 + a_6x_2x_3
\]

where the terms \((a_1, a_2)\) in the above model are coefficients of the regression analysis; \((T, pH, S)\) are temperature, initial medium \(pH\) and initial glucose concentration; and \(\varepsilon\) is the error, which is normally distributed with mean = 0 according to the observed response. Many reports employed the BWD model to evaluate the effects of variables on the response [19,20].

Artificial Neural Networks (ANN)

An artificial neural network (ANN) consists of densely interconnected processing units that use parallel computation algorithms. ANN is also referred to as connectionism, parallel distributed processing, neuron-computing, natural intelligent systems and machine learning algorithms. The basic advantage of ANNs is that they can learn from representative examples without providing special programming modules to simulate special patterns in the data set [21, 22]. The ANN can be trained to perform a particular function by tuning the values of the weights (connections) between these elements [23]. The ANN training procedure is performed so that a particular input leads to a certain target output, as shown in figure 1 [24]. In other words, an ANN learns from examples (e.g., of known input/output sequences) and exhibits some capability for generalization beyond the training data [25,26]. Multi-layer perceptron function neural network (MLP-NN) gives an approximation of any input/output relationship as a linear combination of the radial basis functions, which are a special class of functions with the characteristic feature that their response decreases (or increases) monotonically with distance from a central point [27]. Each neuron in the hidden layer provides a value for the degree of membership for the input pattern with respect to the basis vector of the respective hidden unit. The output layer is composed of linear neurons. The numbers of neurons in the input and the output layers of any network are equal to the number of the inputs and outputs of the system, respectively [27]. Backpropagation (BP) algorithm is a generalized learning rule that is based on the gradient descent algorithm. It is commonly used with multi-layer networks that use non-linear transfer functions. The total weighted input at any neuron \(X_i\) and its output activity \(Y_i\) based on a selected transfer function is computed as described in the results section [28]. In this study, it was of interest to develop a forecasting model that can predict the hydrogen yield as a function of three different factors. The ANN model, with its nonlinear and stochastic modeling capabilities, was developed using MATLAB R2010 7.10.0.499 software and 60 records; this work also studied the comparison to BWD for the same records.
layer were used to examine the best performance. The choice of the number of hidden layers and the number of neurons in each layer is based on two performance indices [21, 29]. The first is the root mean square value of the prediction error, and the second is the value of the maximum error. To optimize the neuron number, the number of neurons was varied from 5 to 30 neurons in increments of five, and each number was run ten times. Then, the mean-squared and iteration numbers were separately evaluated for the neuron numbers. With increasing neuron numbers, the MSE decreased for the training set [11]. To accelerate the training procedure and to achieve the minimum mean square estimation error (MSE), the inflow data were normalized, and the steps to calculate the MSE are presented in the results section. The ANN-based architecture is employed in this study to provide the hydrogen yield as a response to different variables as presented in figure 2.

Results and discussions

Results of the modeling abilities of the BWD statistical technique

BWD is a statistical technique used to investigate the impact of experimental variables on the response (output) that uses central composite design (CCD) to create a response surface, which is commonly chosen for response optimization. BWD is based on the Newton statistical method and depends on the values of the variables [30,31]. The BWD technique can effectively be used with three variables to relate the response output and variable’s inputs. However, it has a drawback that it cannot exceed 15 experiments, as reported in the literature. Eq. (2) was used to fit the hydrogen yield experimental data to construct the model. It was used previously by [19,20], who reported that the BWD model was used to relate the response and three variables inputs. The values of these coefficients and the statistically insignificant terms for the model that represent the suitable form of the mathematical model relating the hydrogen yield (y) to the three variables in terms of levels is listed in (table 1). A nonlinear least-squares regression program based on the Gauss Newton method (GNM) was used to fit eq. (2), and this fitting provides the
predicted hydrogen yield \( y \), the residual error, the coefficients \( (a_n) \) of this equation and the fitted response presented as eq. (3).

\[
y = -649.6 - 26.47T + 58.66pH + 4.99S - 0.32T^2 - 3.25pH^2 - 0.042S^2 + 0.05
\] (3)

This model was used to verify form 2 by using ten experimental runs for calibration and fitted another five experimental runs for validation. Figure 3 presented the statistical prediction of hydrogen with a correlation coefficient \( R^2 = 0.895 \) and a minimum MSE of \( 0.0851 \pm 0.001 \), which was obtained from equation 4. This is in contrast with the results from 15 experiments that were performed with ANN, which obtained the best correlation coefficient \( R^2 = 0.984 \) and MSE of \( 0.0521 \pm 0.001 \). Figure 4 recorded the calibration and validation regression error of the hydrogen yield experimentally, and the results showed the error was between (±15). 

**Results of the modeling abilities of ANN**

Few reports have been studied on the field of hydrogen production using artificial neural networks ANN, especially with CSN1-4. In this work, the ANN modeling technique has the ability to predict 60 experimental runs and more by using multilayer perceptron (MLP-NN) as presented in figure 2. The proposed ANN model was examined by using 60 records of hydrogen yield experiments associated with temperature, \( pH \) and initial glucose concentration. It is important to evaluate the performance of the prediction model considering a wide range of the stochastic pattern of the hydrogen yield. Therefore, the proposed ANN model architecture in figure 2 is re-arranged to consider a total of 60 records of hydrogen yield experiments, from which 50 records were fixed as the training session and the last 10 records for the testing session. Figure 5 illustrates that the proposed ANN model could provide a hydrogen yield prediction with an error of less than 5%, except in two cases exp # 2 and exp # 38 that are during the training stage. On the other hand, during the testing session, the ANN model achieved a prediction error below 20% as shown in figure 6. This is due to the highly stochastic pattern experienced in the data records (exp # 51 to exp # 60).

The mean-squared error was significant for iteration numbers that were separately evaluated from the neuron numbers.
2) as presented in figure 2. They were employed in the training and testing stages to obtain the best values compared to the output and to perform the BP algorithm learning rules. Therefore, as the number of neurons increased, the size of the training set decreased, which was in agreement with [32-34], this study presented different methods for investigating model performance. These were mean absolute relative error (MARE), mean absolute error (MAE) and MSE. To accelerate the training procedure they evaluated models by the following eqs.(4-6):

\[
\text{MARE} = \frac{1}{N} \sum \frac{|y - y'|}{y} \times 100
\]

\[
\text{MAE} = \frac{1}{N} \sum \frac{|y - y'|}{N} \times 100
\]

\[
\text{MSE} = \frac{\sum (y - y')^2}{N} \times 100
\]

where, \(y\) is the experimental Hydrogen production, \(y'\) predicted Hydrogen production by model and \(N\) is the total number of data as reported by [33,35]. This work used 15 experiments to compare with the BWD technique, which merely considers the limited length of 15 records for the stochastic pattern for hydrogen yield.

Also figure 2 shows the numerical data of the hydrogen yield from ten experiments associated with temperature, \(pH\) and initial glucose concentration. It was used to train the ANN model in the eqs. (4-6) to achieve the MARE, MAE and (MSE) target successfully, and the other five were used as testing stage. The training curve for the proposed ANN architecture presented in figure 2 is demonstrated in figure 7, which shows convergence to the target MSE of 0.0001 after 127 iterations. The results based on the ANN model found a high degree of accuracy and efficiency in achieving prediction errors lower than those in the central composite design, which agrees with [36], who reported that the root mean square error and the standard error of prediction for the neural network model were much smaller than those for the response surface methodology model. This indicates that the neural network model had a much higher modeling ability than the response surface methodology model [32,34]. Typically, many such input/target pairs are used to train a network. Backpropagation (BP) uses input vectors and corresponding target vectors to train an ANN. The neural networks with a sigmoid and linear output layer are capable of approximating any function with a finite number of discontinuities [37]. The standard BP algorithm is a gradient descent algorithm in which the network weights are changed along the negative of the gradient of the performance function [11]. There are a number of variations of the basic backpropagation algorithm, which are based on other optimization techniques, such as conjugate gradient and Newton methods. Figure 8 shows the performance of the proposed ANN using the same dataset presented for the BWD, which includes 15 experiments. It can be concluded that ANN outperformed the BWD model and provided a higher and a more consistent level of accuracy for the hydrogen yield at the same conditions with a correlation coefficient \(R^2 = 0.984\), while the standard error in BWD was less than in ANN. This result follows the work performed by [36], which investigated the effect of temperature, initial \(pH\) and glucose concentration on fermentative hydrogen production by mixed cultures in a batch test and found the
neural network model to be a much better model. Figure 9 indicates the calibration and validation regression error was found to be more streamlined and more accurate with ANN to within (±6).

For further analysis, the prediction error distribution as a statistical index for the model evaluation was used as in the following simple equation:

$$PE = \left( \frac{H_{\text{actual}} - H_{\text{predicted}}}{H_{\text{actual}}} \right) \times 100$$  \hspace{1cm} (7)

Where PE is the prediction error and H is the hydrogen yield. Figures (10, 11) show the error distribution for the model output during training (exp# 1 to 10) and during the testing session (exp # 11 to 15), respectively. It can be observed from figure 10 that the ANN model could provide significantly higher accuracy, with errors below 6%. On the other hand, higher levels of errors have been observed during the testing session, as presented in figure 11, due to the significant changes in the input pattern for the model. However, the ANN model still provides acceptable levels of error, lower than 10%, except in one case (exp # 13). This result showed that the neural network could be successfully used to describe the effects of the temperature, initial pH and glucose concentration on the hydrogen yield, and it is in agreement with the report by [34, 38, 39]. For more verification, table 2 shows a comparison of the prediction of hydrogen yield using the ANN model and the BWD model using equation 4. The ANN model was able to reduce the prediction error in the hydrogen yield to be less than ± 6%; in contrast, the BWD model was not able to achieve a similar level of accuracy. The performance of the ANN model in columns 5, 6 of table 2 shows only one case similarity (exp # 4) and provides relatively lower accuracy for (exp # 8), whereas ANN outperformed the BWD model for the predicted hydrogen yield in 13 experiments out of 15. As a result, it is much more advantageous to use the ANN model to predict the hydrogen yield with utmost accuracy using a number of variables and experimental patterns instead of the BWD model. The predicted maximum hydrogen yield was 81.8% at the optimum operating condition of 10 g/L initial glucose concentration, 37°C reaction temperature, and 6.0±0.2 initial medium pH. The observed experimental yield was approximately 4.30% lower than the model prediction based on the experimental conditions.

Further Developments in ANN for Predicting Hydrogen Yield

It is common in ANN development to train several different networks with different architectures and to select the best one based on the performance of the networks with testing/validation sets. A major disadvantage of such an approach is that it assumes that the performance of the networks for all other possible testing sets will usually be similar, which is statistically incorrect. Moreover, observing the performance of the fifteen developed ANN models tested with the four testing sets makes it obvious that no single network has the optimal prediction for all the testing data sets. Therefore, better accuracy compared to the best reported by any single network can be achieved if an optimized algorithm can be developed to use all of these networks. Another interesting observation is that the effect of the transfer function is as important as the number of layers and neurons in each layer. This can be observed when comparing the performance of two networks with similar number of hidden layers and neurons but with different transfer functions. Further discussion on the effect of the optimal...
Table 2
EXPERIMENTAL DESIGN FOR CONSTRUCTING STATISTICAL AND ANN MODELS

<table>
<thead>
<tr>
<th>Error (%)</th>
<th>Error (%)</th>
<th>H2 Y (%)</th>
<th>H2 Y (%)</th>
<th>Statistical (BWD)</th>
<th>H2 Yield (%)</th>
<th>Exp. No</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLP-NN</td>
<td>BWD</td>
<td>ANN</td>
<td>MLP-NN</td>
<td>BWD</td>
<td>ANN</td>
<td>MLP-NN</td>
</tr>
<tr>
<td>-5.71</td>
<td>-5.89</td>
<td>29.60</td>
<td>29.65</td>
<td>28.00</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>5.20</td>
<td>-7.96</td>
<td>56.81</td>
<td>58.3</td>
<td>54.00</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>-3.05</td>
<td>-10.00</td>
<td>44.53</td>
<td>47.53</td>
<td>43.21</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4.18</td>
<td>-4.18</td>
<td>69.34</td>
<td>69.34</td>
<td>66.56</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>-5.06</td>
<td>17.03</td>
<td>52.32</td>
<td>41.32</td>
<td>49.80</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>5.49</td>
<td>8.47</td>
<td>63.40</td>
<td>61.40</td>
<td>67.08</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>-5.01</td>
<td>10.23</td>
<td>68.90</td>
<td>58.90</td>
<td>65.61</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>5.29</td>
<td>-3.23</td>
<td>81.86</td>
<td>80.26</td>
<td>77.75</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>-3.56</td>
<td>11.18</td>
<td>42.52</td>
<td>36.47</td>
<td>41.06</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>-3.11</td>
<td>8.90</td>
<td>68.67</td>
<td>60.67</td>
<td>66.60</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>-4.74</td>
<td>12.65</td>
<td>48.20</td>
<td>40.20</td>
<td>46.02</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>2.38</td>
<td>6.06</td>
<td>72.76</td>
<td>66.76</td>
<td>71.07</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>-9.34</td>
<td>10.49</td>
<td>49.30</td>
<td>40.36</td>
<td>45.09</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>2.36</td>
<td>-7.87</td>
<td>72.75</td>
<td>76.66</td>
<td>71.07</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>-2.36</td>
<td>-7.87</td>
<td>77.75</td>
<td>76.66</td>
<td>71.07</td>
<td>15</td>
<td></td>
</tr>
</tbody>
</table>

combination of different transfer function for specific applications is beyond the scope of this study.

Conclusions
In this research, ANN successfully predicted hydrogen yield using CSN 1-4 with three variables: reaction temperature, initial medium pH and initial glucose concentration. The proposed ANN-based model reliably predicts hydrogen yield and could be used as a predictive controller for management and operation of large-scale hydrogen-fermenting systems. The neural network with its non-linear architecture could provide a significant level of accuracy in predicting hydrogen yield under different stochastic patterns of temperature, initial pH and glucose concentration. Sixty experimental data records have been used to develop the ANN model. The results showed that the proposed ANN widens the range of the hydrogen yield prediction with consideration of the different levels of stochastic pattern of the input up to 60 records of hydrogen yield experiments, out of which 50 records were fixed as the training session and the last 10 records for the testing session. The results also showed that the proposed ANN model achieved a consistent level of accuracy for (HY), while in the training and testing stages for (HY) prediction, the accuracy was within a maximum error of (± 6%).

Moreover, using 15 data records yielded the same error. Furthermore, a comparison analysis with a traditional Box-Wilson Design (BWD) statistical approach has been introduced and shows that the ANN model output significantly outperformed the BWD. Consequently, the ANN overcomes the limitation of the BWD approach, which merely considers a limited length of stochastic patterns for hydrogen yield (15 records).

Acknowledgement: The authors would like to thank King Abdulaziz University for providing instrumental help and laboratory facilities during the research work.

References
2. MILTNER, A., WUKOVITS, W., PRÖLL, T., FRIEDL, A., J of Cleaner Production. 18, 2010, p 551
3. URBANIEC, K., FRIEDL, A., HUISINGH, D., CLAESSEN, P., J of Cleaner Production. 18, 2010, p 51
7. SINGHAL, A., GOMES, J., PRAVEEN, V.V., RAMACHANDRAN, K.B., Biotecnol Prog. 14, 1998, p 645
9. MOHAMMADI, P., IBRAHAM, S., ANNUR, M., LAW, S., J of Cleaner Production, 19, 2011, p 1654
12. PEREZ, T., CAVALCANTI, IOP Science, 2013, p 1
15. FRANK, M.N., KRAJEWSKI, W.F., CUKENDAL, R.R., J. Hydrology, 137, 1992, p 1
18. TANASE, C., CARAMIHAI, M., MUNTEAN, O., Rev. Chim. (Bucharest), 64 no. , 2013, p 182
20. SONMEZ, I., CEBECI, Y., Fuel, 85, 2006, p 289
27. MAIER, H.R., DANDY, G.C., Environmental Modeling Software, 15, 2000, p 101
30. ABHANG, L.B., HAMEEDULLAH, M., J of Engineering Research, ISSN 17266009, 9, 2012, p 210
38. WOO, S. H., JO J. H., LEE, M. W., LEE D. S., J of Nano electronics and Optoelectronics. 6, 2011, p 1

Manuscript received: 6.01.2014