frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 19 September 2022
DOI 10.3389/fenrg.2022.980360

Check for updates

OPEN ACCESS

EDITED BY
Sina Ardabili,
University of Mohaghegh Ardabili, Iran

REVIEWED BY
Reza Sedghi,
University of Tehran, Iran
Esmail Khalife,
Cihan University-Erbil, Iraq

*CORRESPONDENCE
SK Safdar Hossain,
snooruddin@kfu.edu.sa
Bamidele Victor Ayodele,
Bamidele.ayodele@utp.edu.my

SPECIALTY SECTION
This article was submitted to Advanced Clean Fuel Technologies,
a section of the journal
Frontiers in Energy Research

RECEIVED 28 June 2022
ACCEPTED 29 August 2022
PUBLISHED 19 September 2022

CITATION
Safdar Hossain SK, Sadiq Ali S, Cheng CK and Ayodele BV (2022), Performance analysis and modeling of bio-hydrogen recovery from agro-industrial wastewater.
Front. Energy Res. 10:980360.
doi: 10.3389/fenrg.2022.980360

Performance analysis and modeling of bio-hydrogen recovery from agro-industrial wastewater

SK Safdar Hossain[1]*, Syed Sadiq Ali[1], Chin Kui Cheng[2] and Bamidele Victor Ayodele[3,4]*

[1]Department of Chemical Engineering, College of Engineering, King Faisal University, Al-Ahsa, Saudi Arabia, [2]Centre for Catalysis and Separation (CeCaS), Department of Chemical Engineering, College of Engineering, Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates, [3]Department of Chemical Engineering, Universiti Teknologi Petronas, Perak, Malaysia, [4]Centre of Contaminant Control and Utilization (CenCoU), Institute of Contaminant Management for Oil and Gas, Universiti Teknologi Petronas, Perak, Malaysia

Significant volumes of wastewater are routinely generated during agro-industry processing, amounting to millions of tonnes annually. In line with the circular economy concept, there could be a possibility of simultaneously treating the wastewater and recovering bio-energy resources such as bio-hydrogen. This study aimed to model the effect of different process parameters that could influence wastewater treatment and bio-energy recovery from agro-industrial wastewaters. Three agro-industrial wastewaters from dairy, chicken processing, and palm oil mills were investigated. Eight data-driven machine learning algorithms namely linear support vector machine (LSVM), quadratic support vector machine (QSVM), cubic support vector machine (CSVM), fine Gaussian support vector machine (FGSVM), binary neural network (BNN), rotation quadratic Gaussian process regression (RQGPR), exponential quadratic Gaussian process regression (EQGPR) and exponential Gaussian process regression (EGPR) were employed for the modeling process. The datasets obtained from the three agro-industrial processes were employed to train and test the models. The LSVM, QSVM, and CSVM did not show an impressive performance as indicated by the coefficient of determination (R2) < 0.7 for the prediction of hydrogen produced from wastewaters using the three agro-industrial processes. The LSVM, QSVM, and CSVM models were also characterized by high prediction errors. Superior performance was displayed by FGSVM, BNN, RQGPR, EQGPR, and EQGPR models as indicated by the high $R^2 > 0.9$, an indication of better predictability with minimized prediction errors as indicated by the low root mean square error (RMSE), mean square error (MSE), and mean absolute error (MAE).

KEYWORDS

agro-industrial wastewater, support vector machine, Gaussian process regression, binary neural network, bio-hydrogen

Introduction

The agro-industrial often required a huge amount of water for the processing of its agricultural feedstocks to value-added products (Freitas et al., 2021; Martinez-Burgos et al., 2021). This invariably results in a substantial amount of wastewater usually obtain from the process (Libutti et al., 2018). The wastewater generated from agro-industrial processing is increasing at an alarming rate throughout the world (Zaharia et al., 2021). As shown in Figure 1, agro-industrial processing of animals, oil palm, cassava, milk, cheese whey, and vinasse generated a billion liters of wastewater globally as reported Martinez-Burgos et al. (Martinez-Burgos et al., 2021) Wastewater from agricultural and industrial processes often contains high levels of nutrients like phosphorus and nitrogen, which encourage the growth of microorganisms and aquatic plants as well as microalgae (Robles et al., 2020). As a result of eutrophication, the water bodies that receive these effluents become unsuitable for various purposes because they destabilized the ecosystems. To forestall the environmental and health effects of the enormous amount of wastewater from agro-industries, the circular economy concept that utilizes innovative integrated processes of energy recovery and the treatment of wastewater could be developed (Dutta, Arya, and Kumar, 2021).

Several studies have delved into the application of the circular economy concept to harness the opportunities from agro-industrial wastewater. Omran and Baek (Omran and Baek, 2022), reported that agro-industrial biowaste can be valorized to produce green nanomaterials suitable for use in the treatment of wastewater. The potential of producing bio-hydrogen from various agro-industrial wastewater has been reported by Marone *et al.* (2017) and Kumar *et al.* (2022). A combination of dark fermentation and microbial electrolysis displayed a promising alternative for maximizing the conversion of agro-industrial wastewaters and byproducts into bio-hydrogen, as demonstrated by the findings. Marone *et al.* (2017) investigated the possibility of producing bio-hydrogen from microbial electrolysis cells utilizing palm oil mill effluent. The study revealed that factors such as the incubation temperature, initial pH, and influent dilution rate significantly influence the bio-hydrogen production from the palm oil mill effluent. The use of fermentation liquid of waste-activated sludge for biohydrogen production in a microbial electrolysis cell has been reported by Khongkliang *et al.* (2019). The study demonstrated that bio-hydrogen may be recovered from activated sludge by integrating microbial electrolysis cells with active sludge disposal. The recovery of biohydrogen from the conversion of acidogenic effluents in a microbial electrolysis cell has been reported by Lenin Lenin Babu et al., 2013. The study revealed that applied potential conditions in a microbial electrolysis cell are a huge potential for simultaneously producing hydrogen and wastewater treatment.

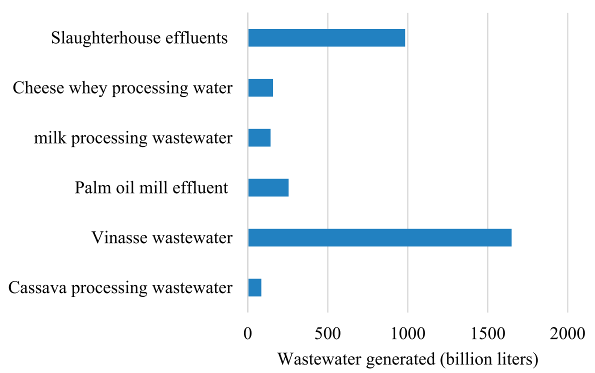

FIGURE 1
Agro-industrial wastewater generated from various processes (Martinez-Burgos et al., 2021).

Although, several experimental studies have established the potential of bioenergy recovery from agro-industrial wastewaters, nevertheless how the various parameters influenced and relate to the various bioenergy recovered from the wastewater is still understudied. A huge amount of data is often generated from the experimental runs capturing the process parameters and the output. A data-driven modeling approach can be adopted to explore the relationship that exists between these input parameters and the targeted output (Sharabiani et al., 2022). As shown in Table 1, various machine learning algorithms such as support vector machine (SVM), Gaussian process regression (GPR) and artificial neural networks (ANN), boost regression, and random forest regression, have been widely employed for modeling different processes involving wastewater treatment. SVM has been reported to be robust in modeling microbial lipid fermentation from cellulosic ethanol wastewater as reported by Zhang, Chao, and Zhang, (2020). As indicated by R^2 of 0.9959 obtained for the data training, the findings show that the SVM model has a great potential to optimize fermentation conditions and could be a useful tool in the future. The modeling of microalgae-based wastewater treatment using SVM was investigated by Hossain *et al.* (2022). A global optimal treatment condition was achieved as indicated by the high removal efficiency of nitrogen and phosphate from microalgae-based wastewater. Hosseinzadeh *et al.* (2022) reported the modeling of biohydrogen recovery from wastewater using SVM. The SVM displayed a significant ability to predict hydrogen production from the wastewater with an R^2 of 0.885. GPR has been employed to model full-scale wastewater treatment and carbon-based material adsorption of organic pollutants from wastewater (Hvala and Kocijan, 2020; Hosseinzadeh et al., 2022). GPR and ANN were effective in modeling the prediction of antibiotics removal from industrial wastewater (Hamza et al., 2022). The GPR was reported to offer a good prediction of the treatment of the wastewater effluent from full-scale wastewater (Hvala and Kocijan, 2020). Bagheri et al. (Bagheri et al., 2015) and Dewasme (Dewasme, 2020) reported the use of ANN for modeling

TABLE 1 Summary of related studies on the application of various machined learning models of wastewater processes.

S/N	Process	Machine learning type	References
1	Microbial lipid fermentation from cellulosic ethanol wastewater	SVM	Zhang et al. (2020)
2	Microalgae-based wastewater treatment	SVM	Hossain *et al.* (2022)
3	Biohydrogen production from wastewater	SVM	Hosseinzadeh *et al.* (2022)
4	Full-scale wastewater treatment plant	GPR	Hvala and Kocijan, (2020)
5	Modeling carbon-based material adsorption of organic pollutants from wastewaters	GPR and ANN	Hamza *et al.* (2022)
6	Prediction of sludge in wastewater treatment plant	ANN	Bagheri et al. (2015)
7	Modeling key-component estimation from brewery wastewater treatment plant	ANN	Dewasme (2020)
8	Prediction of effluent quality parameters	Ada Boost Regression, Gradient Boost Regression (GBR) and Random Forest Regression (RFR	Sharafati, Asadollah and Hosseinzadeh (2020)
9	Modeling sludge bulking of wastewater treatment process	Data-knowledge-driven diagnosis	Han et al. (2021)
10	Effluent prediction of wastewater treatment plant	Random vector functional link integrated with manta ray	Elmaadawy *et al.* (2021)

the prediction of sludge in the wastewater treatment plant and key-component estimation from brewery wastewater treatment plant. The training and validation of the ANN models demonstrated a nearly perfect agreement between the experimental and ANN predicted values. Other machine learning algorithms such as Ada Boost Regression, Gradient Boost Regression, and Random Forest Regression have also been employed for modeling the prediction of effluent quality parameters, and sludge bulking of the wastewater treatment process (Sharafati, Asadollah, and Hosseinzadeh, 2020; Elmaadawy et al., 2021; Han, Dong, and Qiao, 2021). To the best of the authors' knowledge the use of SVM (incorporated with various kernel functions), GPR (incorporated with various kernel functions), and Bi-layer neural network (BNN) for the modeling the effect of various parameters on bio-hydrogen recovery from agro-industrial wastewater has not been reported in the literature. Data is fed into the kernel, and it performs the necessary transformations. This study therefore employed SVM and GPR incorporated with various kernel functions as well as BNN for modeling bio-hydrogen recovery from three agro-industrial wastewater namely dairy wastewater, chicken processing wastewater, and palm oil mill effluent.

Experimental details of biohydrogen production and model development

Experimental on biohydrogen production from wastewaters

The biohydrogen under consideration was produced from dairy wastewater, chicken processing wastewater, and palm oil mill effluent. A detailed description of the processes involved in bio-hydrogen production from dairy wastewater, chicken processing wastewater, and palm oil mill effluent has been reported by Gadhe et al. (Gadhe, Sonawane, and Varma, 2013), Thirugnanasambandham et al. (Thirugnanasambandham, Sivakumar and Prakasmaran, 2015), and Kadier et al. (Kadier et al., 2021). The relationship between maximal biohydrogen production from a given concentration of substrate, pH, COD/Nitrogen ratio, and COD/Phosphorus ratio was investigated (Gadhe, Sonawane and Varma, 2013). For the chicken processing wastewater, the effect of current density, hydraulic retention time, and electrode surface area on the biohydrogen production from the chicken processing wastewater in an electrochemical reactor was investigated (Thirugnanasambandham, Sivakumar and Prakasmaran, 2015). Also, the effect of process variables such as temperature, initial pH of the palm oil mill effluent, and the influent COD concentration on bio-hydrogen production in microbial electrolysis cell was investigated (Kadier et al., 2021). A total of 64 datasets comprised of the various process variables and targeted output was employed to train and validate the machine learning algorithms.

Model development

The stages involved in the model development are represented in Figure 2. The stages include the data acquisition from the experimental runs, data preprocessing, model configuration, model training, model validation, and model deployment for the prediction of the hydrogen produced from wastewater. After the data acquisition from the experimental runs, it ensured that the data are preprocessed for any missing values or outliers. The model configuration entailed the setting of the various models that

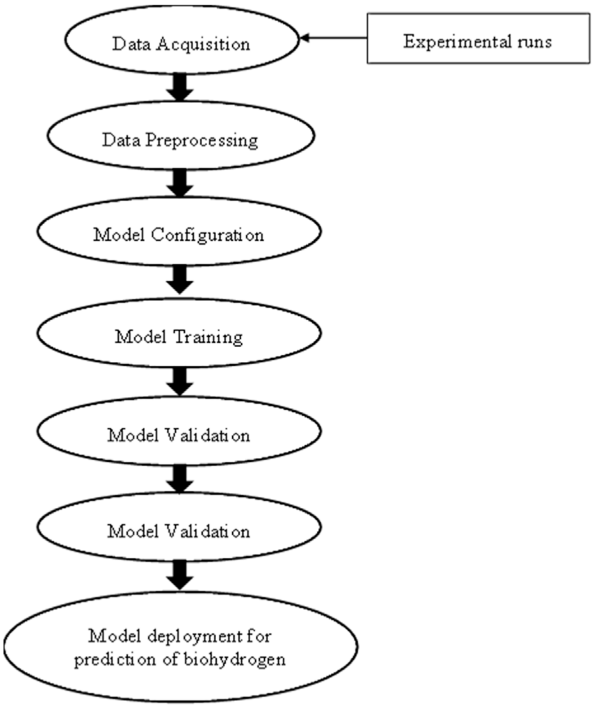

FIGURE 2
Schematic representation of the steps involved in the modeling process.

would be employed for the modeling the prediction of the hydrogen. Thereafter, the models are trained with a portion of the data to ensure that the relationship between the predictors and the targeted variable is well learned. While the remaining portion of the data is employed to validate the trained model. The performance of the model is tested before deployment for predicting hydrogen production.

Eight machine learning algorithms namely LSVM, QSVM, CSVM, FGSVM, BNN, RQGPR, EQGPR, and EQGPR were configured for modeling the non-linear relationship between the various input parameters to the wastewater treatment processes and the biohydrogen produced from the wastewater. The effect of kernel functions such as linear, quadratic, cubic, and fine Gaussian on the performance of the SVM was investigated (Leong et al., 2021). While the effect of kernel functions such as rotational quadratic, squared exponential, and exponential on the performance of the GPR was also investigated. Altogether, a total of eight different models were considered (Zeng, Ho and Yu, 2020).

The main objective of the SVM is to use various forms of kernel functions to project nonlinearly separable samples onto a higher-dimensional environment. Kernel functions are frequently referred to as "generalized dot products" since they compute the dot product of two vectors X and y in a (very high-dimensional) feature space (Zanaty and Afifi, 2020). Kernel functions are important in SVM for bridging the gap between linearity and nonlinearity. In the higher dimensional space, the linear model $f(X, \psi)$ for SVM is as follows:

$$f(X, \psi) = \sum_{i=1}^{n} \psi_i g_i(X) + b \tag{1}$$

$g_i(x)$denotes a set of linear transformations, the bias term is denoted by b.

The polynomial kernel function which includes, quadratic, and cubic compares input samples not just on their individual properties, but also on their combinations. The polynomial kernel represented in Eq. 2 produces enlarged features using n original features and d polynomial degrees (Koschwitz et al., 2018).

$$k\left(X_i, X_j\right) = \left(X_i . X_j + 1\right)^d \tag{2}$$

SVM regression analysis may be utilized to circumvent the challenges of utilizing linear functions in the high-dimensional feature space, and the optimization issue is turned into dual convex quadratic algorithms (D Koschwitz et al., 2018). Errors larger than or equal to the threshold are penalized by applying the loss function to the regression. As a result, the sparse representation of the decision rule provides considerable advantages in terms of algorithmic and representational efficiency.

Just like the SVM, the GPR is a robust machine learning algorithm that can be applied to modeling bioenergy recovery from agro-industry wastewater (Gao et al., 2018). The fact that GPR is non-parametric means that it may be used to handle a broad range of supervised learning problems, even though only a limited amount of information is provided. Any subset of the GPR's random variables can be said to be jointly Gaussian as represented in Eq. 3 (Bang, Yoon and Jeon, 2020).

$$p(x) = \frac{1}{\left((2\pi)^d \Delta \left|\sum\right|\right)^{1/2}} e^{\left(-\frac{1}{2}(x-\mu)^T \left(\sum^{-1}(x-\mu)\right)\right)}, \quad x = \left[x_i \ldots x_j\right]^T \in \mathbb{R}^d \tag{3}$$

In Eq 3, d depicts the number of random variables, μ represents the vector of mean values, Σ is the covariance matrix of the random variables, x is a set of random variables between i and j. Given observed training data, GPR uses this data to compute the parameters of a posterior Gaussian distribution for targets over the test points x. A Gaussian distribution may be thought of as being predicted at each test point.

The BNN consists of the hidden and the output layer. Input signals into the BNN are combined linearly, and the activation function is used to transform the output (Zhu, Duong and Liu, 2020). The BNN configurations are made up of layers of neurons that feed each other's output till the ultimate output is reached. Training the network means learning the relationship between the inputs and the targets that the network is presented with (Martinez et al., 2020). At each iteration (epoch), the difference between the target data and the network output was computed, and the network

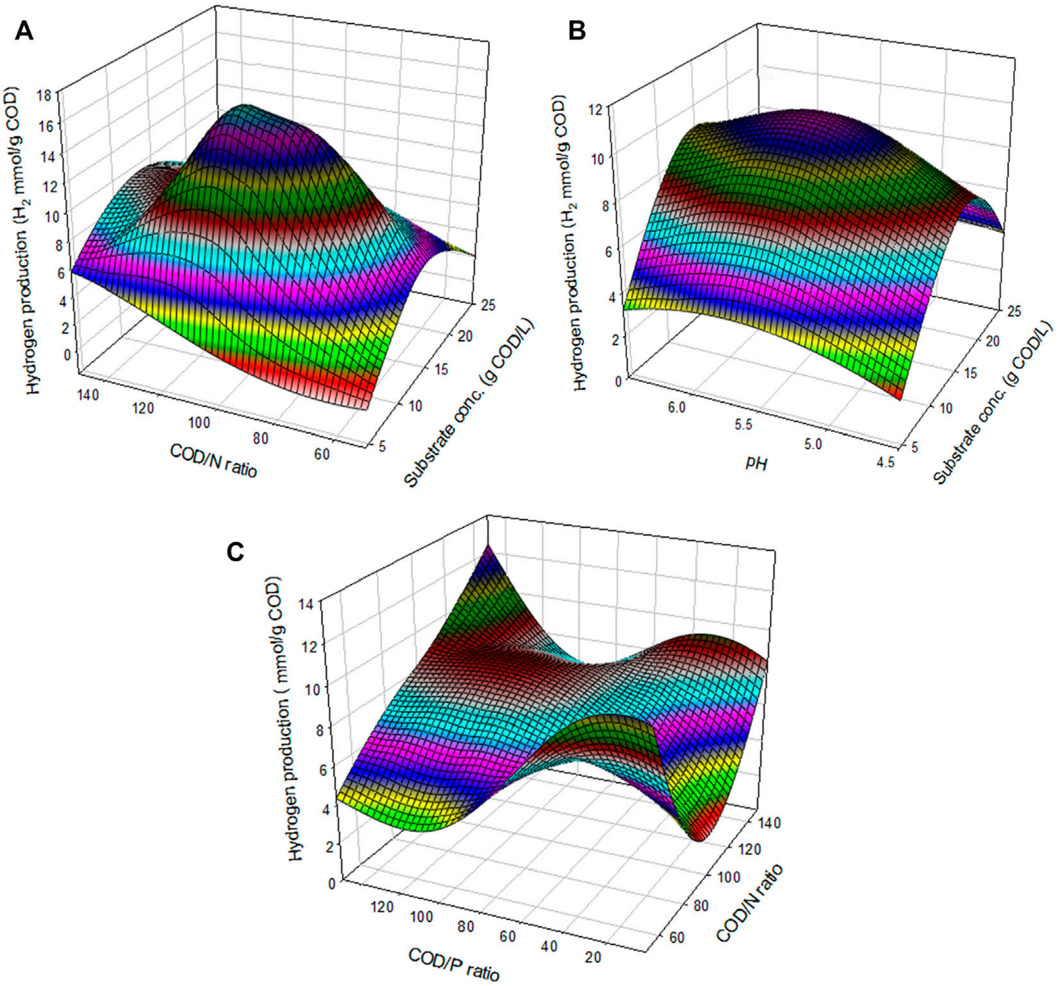

FIGURE 3
Non-linear relationship between **(A)** COD/N ratio and substrate concentration **(B)** pH and substrate concentration and **(C)** COD/P ratio and COD/N ratio on hydrogen produced from dairy wastewater.

weights were updated until a low mean standard error (MSE) was achieved. The MSE of the targeted output on the training set is computed as weights are provided to the training set at each epoch. Every epoch, the MSE of the validation set is computed and training is stopped when the MSE of the validation set rises.

The configuration of the SVM, GPR, and BNN was performed using a regression learner application in the Mathlab environment. K-fold cross-validation was to prevent data overfitting. For this study, 2-fold cross-validation was employed. Each data sample is divided into a certain number of groups by a single parameter called k in this technique. In applied machine learning, cross-validation is used to measure the model's ability to learn from new data. A small sample may be utilized to get an idea of how well the model will perform when it is applied to data that was not included in the training process. The performance of each of the models was evaluated using mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE) and coefficient of determination (R^2) defined in Eqs 4–7 (Ayodele et al., 2020).

$$MSE = \sum_{i=1}^{n}\left(\frac{\left(z_{pi} - z_{ai}\right)^2}{n}\right) \tag{4}$$

$$RMSE = \sum_{i=1}^{n}\left(\frac{\left(z_{pi} - z_{ai}\right)^2}{n}\right)^{1/2} \tag{5}$$

$$MAE = \frac{\sum_{i=1}^{n}\left|z_{pi} - z_{ai}\right|}{n} \tag{6}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(z_{pi} - z_{ai}\right)^2}{\sum_{i=1}^{n}\left(z_{pi} - \bar{z}_{ai}\right)^2} \tag{7}$$

where z_{pi}, z_{ai} are the predicted and actual outputs for each data set i, respectively, n is the number of observed datasets, $\bar{z}_{ai}$ is the mean actual output.

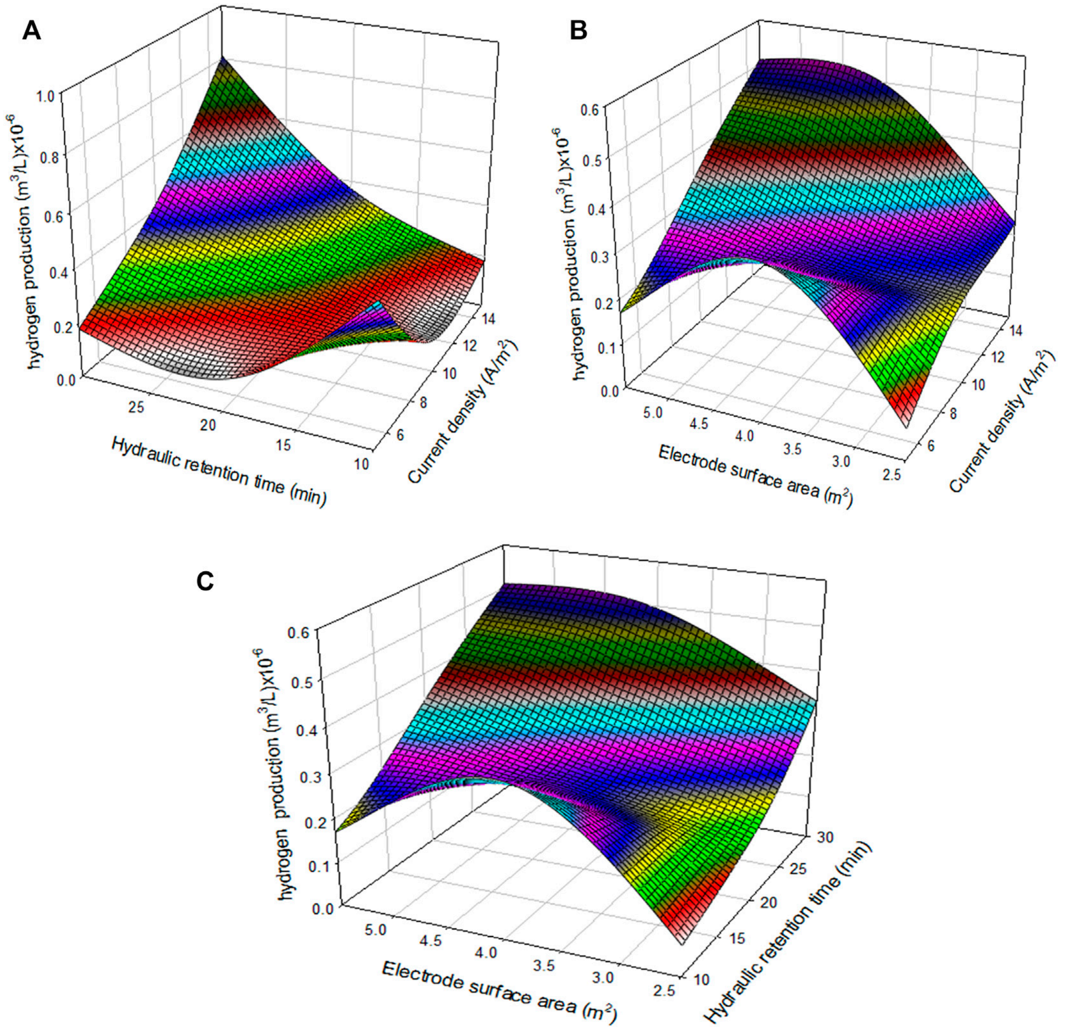

FIGURE 4
Non-linear relationship between **(A)** hydraulic retention time and current density **(B)** electrode surface area and current density and **(C)** electrode surface area and hydraulic retention time on hydrogen produced from chicken processing wastewater.

Results and discussion

Parametric analysis of input and target variables

Three different wastewaters from a dairy, chicken processing, and palm oil mill were investigated for the possibilities of biohydrogen production. The hydrogen from the dairy wastewater was produced from the batch fermentation process considering the effect of COD/nitrogen ratio, COD/phosphorus ratio, and substrate concentration. The relationship between the various input variables and the hydrogen produced from the dairy wastewater is represented in Figure 3. In Figure 3A, a non-linear relationship exists between the COD/N ratio, substrate concentration, and hydrogen production. An increase in the COD/N ratio resulted in a corresponding increase in hydrogen production which is consistent with the work of Liu et al. (2022) for hydrogen production from herbal wastewater. The presence of nitrogen in the wastewater medium helps to facilitate the breaking down of the organic matters in wastewater to release biohydrogen (Goswami et al., 2021). It can be seen that hydrogen production from dairy wastewater is promoted using substrate concentrations ranging from 5 to 15 g COD/L (Gadhe, Sonawane and Varma, 2013). A decline in hydrogen production has been observed at a substrate concentration >15 g COD/L. In Figure 3B, an increase in the pH of the fermentation medium produces an increase in hydrogen production. Higher hydrogen production is favoured at 5.6. Similarly in Figure 3C, an undulating effect of COD/P ratio on hydrogen production is observed. A higher concentration of phosphorus in the

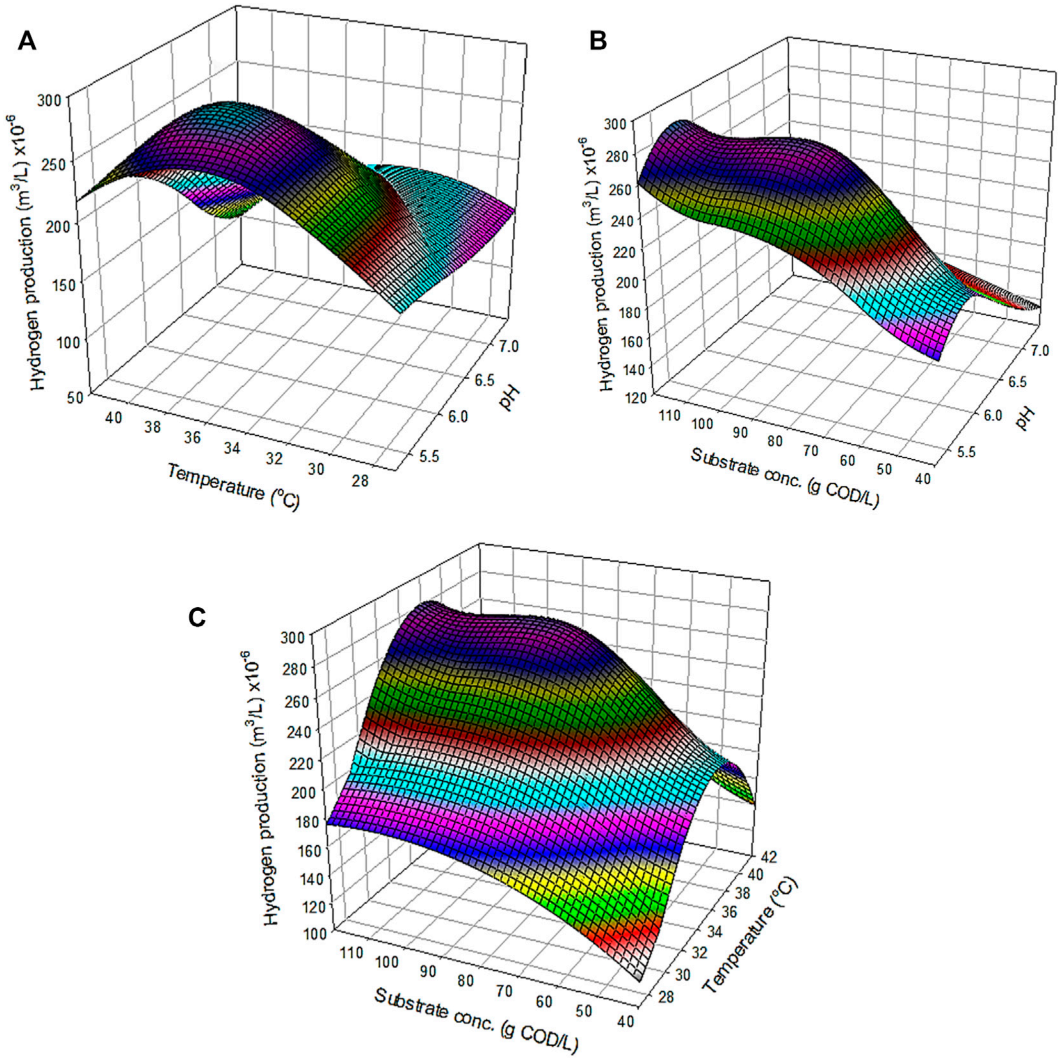

FIGURE 5
Non-linear relationship between **(A)** Temperature and pH **(B)** Substrate concentration and pH and **(C)** Substrate concentration and Temperature time on hydrogen produced from palm oil mill effluent wastewater.

fermentation facilitated microbial decomposition of the substrates to release hydrogen.

Figure 4 displays the relationship between the various input variables like hydraulic retention time, current density, electrode surface area, and hydrogen produced in an electrochemical reactor. The relationship depicted in Figure 4A revealed that the hydrogen produced from the electrochemical reactor is favoured at high current density and low retention time (Sharma and Li, 2010; Kirkaldy et al., 2018). Whereas, using an electrode surface of 3.8 m^2 produces maximum hydrogen (Figure 4B). A decline in hydrogen production is observed at an electrode surface area >3.8 m^2. In Figure 4C, increasing the hydraulic retention time promotes an increasing hydrogen production as a result of the interaction with the electrode surface area.

The relationship between the input variables on the hydrogen produced from palm oil mill effluent using microbial fermentation is represented in Figure 5. The increase in batch reactor temperature from 28 to 36 °C favours an increase in the hydrogen production from the palm oil mill effluent as shown in Figure 5A (Norfadilah et al., 2016). For the interaction between the two variables (temperature and pH), hydrogen production is favoured at pH of 5.5. Using a higher amount of substrate concentration also promotes a high volume of hydrogen production as shown in Figures 5A, B (Cisneros-Pérez et al., 2015). It can be seen that the highest hydrogen production of 280×10^{-6} m^3/L is obtained with the interaction between the substrate concentration and pH (Figure 4B) as well as substrate concentration and temperature (Figure 5C).

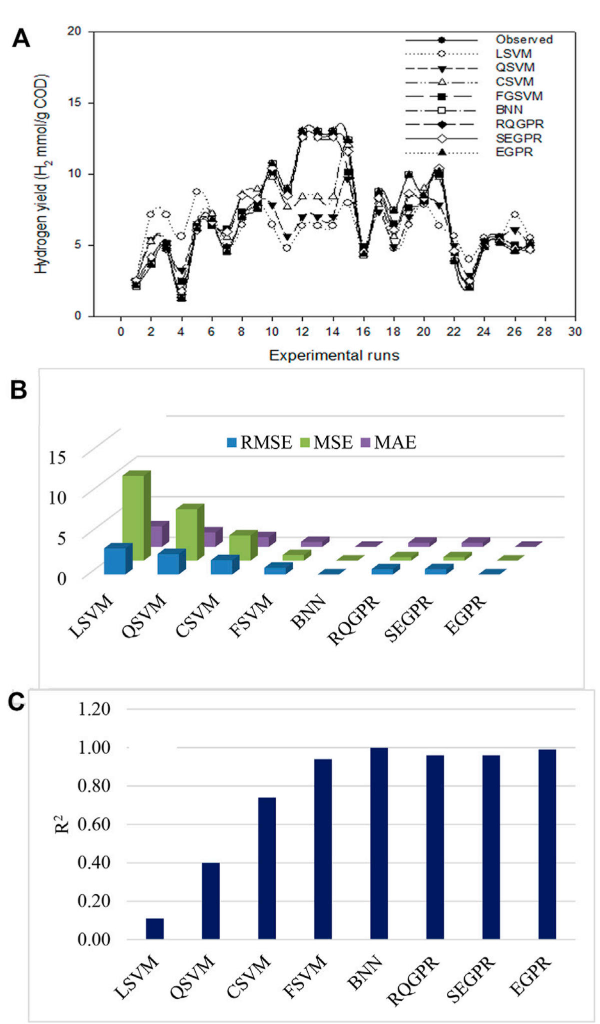

FIGURE 6
(A) Dispersion plot of actual and predicted hydrogen produced from the dairy wastewater **(B)** error analysis of the various models and **(C)** performance of each of the models in terms of R^2.

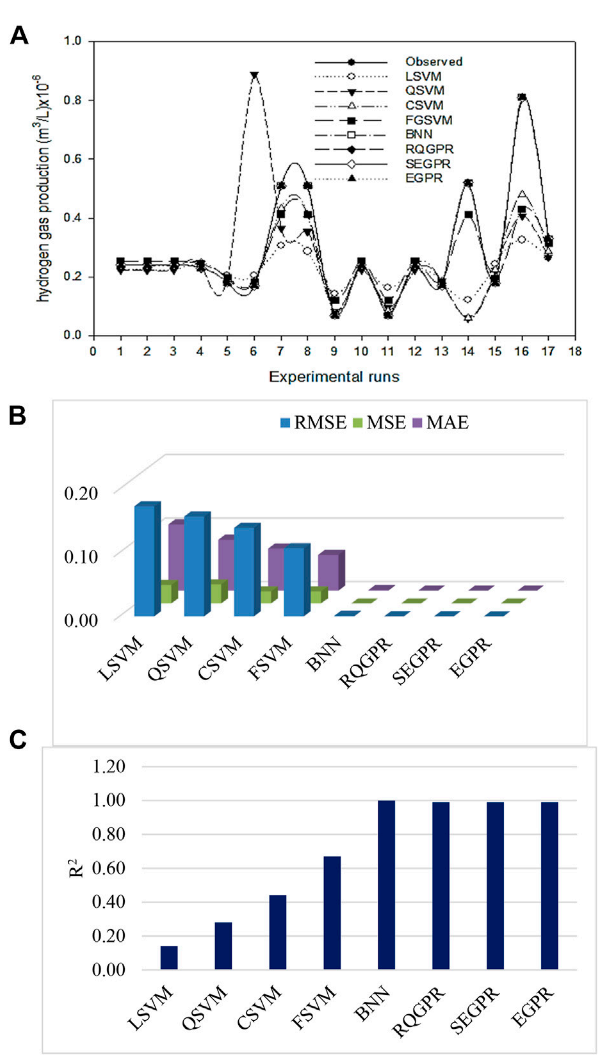

FIGURE 7
(A): Dispersion plot of actual and predicted hydrogen produced from the chicken processing wastewater **(B)** error analysis of the various models and **(C)** performance of each of the models in terms of R^2.

Performance analysis of the models

The production of hydrogen from the dairy wastewater, chicken processing, and palm oil mill effluent was modeled using eight machine learning algorithms namely, LSVM, QSVM, CSVM, FGSVM, BNN, RQGPR, SEGPR, and EGPR. The performance of the eight models in modeling hydrogen production from dairy wastewater is depicted in Figure 6. Figure 6A represents the performance of the models as a function of comparison between the actual and the predicted hydrogen production. As shown in Figure 6A, the SVM did not show impressive performance in modeling the prediction of the hydrogen production from the dairy wastewater. There is a huge deviation between the actual and the predicted values of hydrogen production even with the incorporation of the linear, quadratic and cubic kernel functions. However, it is interesting to note that the performance of the SVM increases with an increase in the degree of polynomial from linear to fine Gaussian. As shown in Figure 6B, higher RMSE, MSE, and MAE were obtained for the LSVM, QSVM, and CSVM compared to the QSVM. Also, lower R^2 values of 0.11, 0.40, and 0.74 were obtained for the LSVM, QSVM, and CSVM, respectively compared to 0.94 obtained for the FGSVM. The performance of the FGSVM compared to the LSVM, QSVM, and CSVM could be attributed to its unique advantage. The fine Gaussian kernels are universal kernels, which implies that when used in conjunction with adequate regularisation, they ensure the creation of an optimum predictor that minimizes both the estimate and approximation

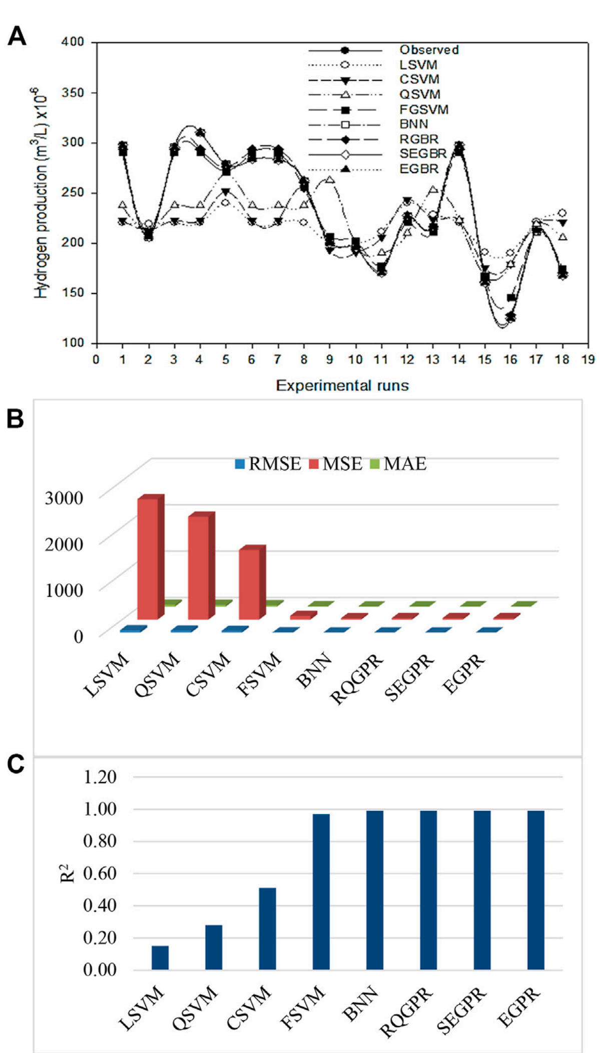

FIGURE 8
(A): Dispersion plot of actual and predicted hydrogen produced from palm oil mill effluent **(B)** error analysis of the various models and **(C)** performance of each of the models in terms of R^2.

errors of a predicted value (Bang, Yoon and Jeon, 2020). The FGSVM however displayed a lesser performance when compared to other models like the BNN, RQGPR, SEGPR, and EGPR. As shown in the dispersion plots, the predicted hydrogen production from the BNN, RQGPR, SEGPR, and EGPR models is in close agreement with the actual values. This can also be confirmed by the low values of the RMSE, MSE, and MAE as well as the high R^2 in Figures 6B, C. The prediction of the hydrogen production from the dairy wastewater resulted in R^2 of 0.999, 0.960, 0.960, and 0.990, respectively.

Figure 7 depicts the performance of the eight models in terms of the dispersion plot which compares the actual and the predicted hydrogen production, the error analysis, and the R^2 analysis. As shown in Figure 7A, the predicted hydrogen produced from the chicken processing wastewater by the LSVM, QSVM, CSVM, and FGSVM is a variant of the actual values. This is evident in the high values of the RMSE, MSE and MAE obtained for the prediction of the hydrogen as depicted in Figure 7B. The R^2 values of 0.140, 0.280, 0.440, and 0.670 obtained for LSVM, QSVM, CSVM, and FGSVM, respectively imply that only the short range of the dataset can be generalized by the models. A better performance was obtained using the BNN, RQGPR, SEGPR, and EGPR, as indicated by the proximity of the predicted and the actual hydrogen production from the chicken processing wastewater as indicated in Figure 7A. Very low RMSE, MSE, and MAE were obtained for the BNN, RQGPR, SEGPR, and EGPR models compared to the SVM-based models. The R^2 values of 0.999, 0.990, 0.990, and 0.990 obtained for the BNN, RQGPR, SEGPR, and EGPR models, respectively are indications of better generalization of the models.

Figure 8 represents the performance of the eight models as a function of the dispersion plots, the error analysis, and the R^2. As established in the previous sections, the LSVM, QSVM, and CSVM did not show impressive performance in modeling the hydrogen production from the palm oil mill effluent as indicated in Figure 8A. The predicted hydrogen production obtained by LSVM, QSVM, and CSVM models largely deviate from the actual values obtained from the experimental runs. A large error analysis was obtained for the prediction of hydrogen production as indicated in Figure 8B. The R^2 values of 0.15, 0.28, and 0.51 obtained for LSVM, QSVM, and CSVM, respectively are an indication of the low generalization ability of the models. However, the incorporation of the fine Gaussian kernel functions into the SVM showed a significant improvement as indicated by R^2 of 0.97. This can be attributed to the robustness of the fine Gaussian kernel functions in the generalization of non-linear functions. Better performance in modeling hydrogen production is obtained using the BNN, RQGPR, SEGPR, and EGPR as indicated by Figure 8A The predicted and the actual hydrogen production from the wastewater are in close agreement. The models predicted the hydrogen production with minimum errors as depicted in Figure 8B. An R^2 of 0.999 obtained for each of the BNN, RQGPR, SEGPR, and EGPR models depicted in Figure 8C indicated that a large proportion of the datasets can be generalized with minimum error.

Comparison of the best models with literature and practical implications of the study

The comparison between the four best models in this study namely BNN, RQGPR, SEGPR, and EGPR, and those reported in

TABLE 2 Comparison of the best models with literature.

Model types	Process	Performance matrix	References
BNN	Dairy wastewater	R^2 = 0.999, RMSE = 9.93 × 10^{-7}	This study
	chicken processing wastewater	R^2 = 0.999, RMSE = 8.73 × 10^{-4}	
	palm oil mill effluent	R^2 = 0.990, RMSE = 5.38	
RQGPR	Dairy wastewater	R^2 = 0.960, RMSE = 0.65	This study
	chicken processing wastewater	R^2 = 0.990, RMSE = 1.72 × 10^{-7}	
	palm oil mill effluent	R^2 = 0.990, RMSE = 5.58	
SEGPR	Dairy wastewater	R^2 = 0.960, RMSE = 0.66	This study
	chicken processing wastewater	R^2 = 0.990, RMSE = 1.72 × 10^{-5}	
	palm oil mill effluent	R^2 = 0.990, RMSE = 5.58	
EGPR	Dairy wastewater	R^2 = 0.990, RMSE = 5.72 × 10^{-4}	This study
	chicken processing wastewater	R^2 = 0.990, RMSE = 1.53 × 10^{-5}	
	palm oil mill effluent	R^2 = 0.990, RMSE = 5.58	
Random forest	Industrial wastewater	R^2 = 0.902, RMSE = 0.126	Hosseinzadeh *et al.* (2022)
ANFIS	Industrial wastewater	R^2 = 0.930, RMSE = 0.089	Taheri *et al.* (2021)
BPNN	Distillery wastewater	R^2 = 0.929, RMSE = N. R*	Sridevi, Sivaraman and Mullai, (2014)
MLPNN	Confectionery wastewater	R^2 = 0.996, APE = 0.0004	Yogeswari, Dharmalingam and Mullai, (2019)
ANN	Fermentative medium	R^2 = 0.900, RMSE = N. R	Sewsynker, Kana and Lateef, (2015)
SVM	Industrial wastewater	R^2 = 0.998, RMSE = 0.983	Raji *et al.* (2022)

*N.R, not reported.

the literature for similar processes are summarized in Table 2. The four models are robust in modeling the prediction of biohydrogen from dairy wastewater, chicken processing water, and palm oil mill effluent. This is evidenced by the high R^2 values (>0.9) and low RMSE values. An indication that the predicted biohydrogen produced from the various processes is consistent with the values obtained from the experimental runs. It implies that the models' algorithms efficiently learn the non-linear relationship between the various input variables and the biohydrogen produced from the wastewaters. The performances of the BNN, RQGPR, SEGPR, and EGPR are comparable with other machine learning algorithms such as random forest, Adaptive neuro-fuzzy inference system (ANFIS) (Hosseinzadeh et al., 2022), Backpropagation neural network (BPNN) (Sridevi, Sivaraman and Mullai, 2014), multilayer perceptron neural network (MLPNN) (Yogeswari, Dharmalingam and Mullai, 2019) and SVM (Raji et al., 2022). The modeling of biohydrogen production from industrial wastewaters, distillery wastewater, confectionery wastewater, and fermentative medium results in an accurate prediction with high R^2 and low RMSE. Generally, studies have shown that machine learning algorithms are highly efficient in modeling processes with a non-linear relationship between the input and the targeted variables. With the help of the machine learning algorithms, biohydrogen production from the various wastewaters can be optimized in real-time thereby improving the process efficiency as well as enhance energy and material utilization. The historical data from the processes can be employed to continuously improve the process performance and optimize desired products.

Conclusion

The potential of producing bio-hydrogen from agro-industrial wastewater has been established in this study. Dairy, poultry processing, and palm oil mill wastewaters all have promising potential for bio-hydrogen generation. Hydrogen was produced from a variety of wastewater sources, and the datasets acquired from the experimental investigations were used to model the relationship between the input factors and the desired result. Eight machine learning models were used in the study, all of which demonstrated promising results when tasked with learning the non-linear connection between the input and the goal variables. The LSVM, QSVM, and CSVM models performed poorly in terms of generalizing the datasets and making predictions about hydrogen production as shown by the low R^2 values. Predictions of hydrogen production was improved using the SVM with fine Gaussian kernels. The BNN, RQGPR, SEGPR, and EGPR models however outperformed the SVM-based models. Each of the BNN, RQGPR, SEGPR, and EGPR models performed exceptionally well in predicting hydrogen production from the dairy, chicken processing, and palm oil mill, with an $R^2 > 0.9$. Indicated by low RMSE, MSE, and MAE values, the models can generalize well for the task of predicting hydrogen recovered from agro-industrial effluent with as little error in their predictions as possible. In the event of a scaleup, the included BNN, RQGPR, SEGPR, and EGPR algorithms may aid in increasing the efficiency of the process. The impacts of input and output variables on process safety, material utilization, and energy efficiency may be monitored if their interdependencies are understood.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material further inquiries can be directed to the corresponding authors.

Author contributions

SS: Conceptualization, Writing—Review and Editing, Supervision, Project administration, Funding acquisition. SS: Writing—Review and Editing CC: Writing—Review and Editing. BA: Conceptualization, Methodology, Formal analysis, Investigation, Writing—Original Draft, Visualization.

Funding

Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia (Project No. Grant 736).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ayodele, B. V., Alsaffar, M. A., Mustapa, S. I., and Vo, D. N. (2020). Back-propagation neural networks modeling of photocatalytic degradation of organic pollutants using TiO_2 -based photocatalysts. *J. Chem. Technol. Biotechnol.*, jctb.6407–11. (January). doi:10.1002/jctb.6407

Babu, M. L., Subhash, G. V., Sarma, P. N., and Mohan, S. V. (2013). Bio-electrolytic conversion of acidogenic effluents to biohydrogen: An integration strategy for higher substrate conversion and product recovery. *Bioresour. Technol.* 133, 322–331. doi:10.1016/j.biortech.2013.01.029

Bagheri, M., Mirbagheri, S. A., Bagheri, Z., and Kamarkhani, A. M. (2015). Modeling and optimization of activated sludge bulking for a real wastewater treatment plant using hybrid artificial neural networks-genetic algorithm approach. *Process Saf. Environ. Prot.* 95, 12–25. doi:10.1016/j.psep.2015.02.008

Bang, H.-T., Yoon, S., and Jeon, H. (2020). Application of machine learning methods to predict a thermal conductivity model for compacted bentonite. *Ann. Nucl. Energy* 142, 107395. doi:10.1016/j.anucene.2020.107395

Cisneros-Pérez, C., Carrillo-Reyes, J., Celis, L. B., Alatriste-Mondragon, F., Etchebehere, C., and Razo-Flores, E. (2015). Inoculum pretreatment promotes differences in hydrogen production performance in EGSB reactors. *Int. J. Hydrogen Energy* 40 (19), 6329–6339. doi:10.1016/j.ijhydene.2015.03.048

Dewasme, L. (2020). Brewery wastewater treatment plant key-component estimation using moving-window recurrent neural networks. *IFAC-PapersOnLine* 53 (2), 16808–16813. doi:10.1016/j.ifacol.2020.12.1173

Dutta, D., Arya, S., and Kumar, S. (2021). Industrial wastewater treatment: Current trends, bottlenecks, and best practices. *Chemosphere* 285, 131245. doi:10.1016/j.chemosphere.2021.131245

Elmaadawy, K., Abd Elaziz, M., Elsheikh, A. H., Moawad, A., Liu, B., Lu, S., et al. (2021). Utilization of random vector functional link integrated with manta ray foraging optimization for effluent prediction of wastewater treatment plant. *J. Environ. Manag.* 298, 113520. doi:10.1016/j.jenvman.2021.113520

Freitas, L. C., Barbosa, J. R., da Costa, A. L. C., Bezerra, F. W. F., Pinto, R. H. H., and Carvalho Junior, R. N. d. (2021). From waste to sustainable industry: How can agro-industrial wastes help in the development of new products? ', *Resour. Conservation Recycl.* 169, 105466. doi:10.1016/j.resconrec.2021.105466

Gadhe, A., Sonawane, S. S., and Varma, M. N. (2013). Optimization of conditions for hydrogen production from complex dairy wastewater by anaerobic sludge using desirability function approach. *Int. J. Hydrogen Energy* 38 (16), 6607–6617. doi:10.1016/j.ijhydene.2013.03.078

Gao, W., Karbasi, M., Hasanipanah, M., Zhang, X, and Guo, J. (2018). Developing GPR model for forecasting the rock fragmentation in surface mines. *Eng. Comput.* 34 (2), 339–345. doi:10.1007/s00366-017-0544-8

Goswami, R. K., Mehariya, S., Verma, P., Lavecchia, R., and Zuorro, A. (2021). Microalgae-based biorefineries for sustainable resource recovery from wastewater. *J. Water Process Eng.* 40, 101747. doi:10.1016/j.jwpe.2020.101747

Hamza, M. A., Althobaiti, M. M., Al-Wesabi, F. N., Alabdan, R., Mahgoub, H., Hilal, A. M., et al. (2022). Gaussian process regression and machine learning methods for carbon-based material adsorption. *Adsorpt. Sci. Technol.* 2022. doi:10.1155/2022/3901608

Han, H. G., Dong, L. X., and Qiao, J. F. (2021). Data-knowledge-driven diagnosis method for sludge bulking of wastewater treatment process. *J. Process Control* 98, 106–115. doi:10.1016/j.jprocont.2021.01.001

Hossain, S. M. Z., Sultana, N., Mohammed, M. E., Razzak, S. A., and Hossain, M. M. (2022). Hybrid support vector regression and crow search algorithm for modeling and multiobjective optimization of microalgae-based wastewater treatment. *J. Environ. Manag.* 301, 113783. doi:10.1016/j.jenvman.2021.113783

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

Hvala, N., and Kocijan, J. (2020). Design of a hybrid mechanistic/Gaussian process model to predict full-scale wastewater treatment plant effluent. *Comput. Chem. Eng.* 140, 106934. doi:10.1016/j.compchemeng.2020.106934

Kadier, A., Wang, J., Chandrasekhar, K., Abdeshahian, P., Islam, M. A., Ghanbari, F., et al. (2021). Performance optimization of microbial electrolysis cell (MEC) for palm oil mill effluent (POME) wastewater treatment and sustainable Bio-H2 production using response surface methodology (RSM). *International journal of hydrogen energy*, 1–16. doi:10.1016/j.ijhydene.2021.09.259

Khongkliang, P., Jehlee, A., Kongjan, P., Reungsang, A., and O-Thong, S. (2019). High efficient biohydrogen production from palm oil mill effluent by two-stage dark fermentation and microbial electrolysis under thermophilic condition. *Int. J. Hydrogen Energy* 44 (60), 31841–31852. doi:10.1016/j.ijhydene.2019.10.022

Kirkaldy, N., Chisholm, G., Chen, J.-J., and Cronin, L. (2018). A practical, organic-mediated, hybrid electrolyser that decouples hydrogen production at high current densities. *Chem. Sci.* 9 (6), 1621–1626. doi:10.1039/c7sc05388f

Koschwitz, D., Frisch, J., and van Treeck, C. (2018). Data-driven heating and cooling load predictions for non-residential buildings based on support vector machine regression and narx recurrent neural network: A comparative study on district scale. *Energy* 165, 134–142. doi:10.1016/j.energy.2018.09.068

Kumar, A., Verma, L. M., Sharma, S., and Singh, N. (2022). Overview on agricultural potentials of biogas slurry (BGS): Applications, challenges, and solutions. *Biomass Conversion and Biorefinery* 4, 1–41. doi:10.1007/s13399-021-02215-0

Leong, W. C., Bahadori, A., Zhang, J., and Ahmad, Z. (2021). Prediction of water quality index (WQI) using support vector machine (SVM) and least square-support vector machine (LS-SVM). *Int. J. River Basin Manag.* 19 (2), 149–156. doi:10.1080/15715124.2019.1628030

Libutti, A., Gatta, G., Gagliardi, A., Vergine, P., Pollice, A., Beneduce, L., et al. (2018). Agro-industrial wastewater reuse for irrigation of a vegetable crop succession under Mediterranean conditions. *Agric. Water Manag.* 196, 1–14. doi:10.1016/j.agwat.2017.10.015

Liu, R., Lin, Y., Xu, G., and Li, Y. (2022). Hydrogen production from herbal wastewater via anaerobic fermentation with diatomite-immobilized sludge. *Asia. Pac. J. Chem. Eng.* 17 (1), e2642. doi:10.1002/apj.2642

Marone, A., Ayala-Campos, O. R., Trably, E., Carmona-Martinez, A. A., Moscoviz, R., Latrille, E., et al. (2017). Coupling dark fermentation and microbial electrolysis to enhance bio-hydrogen production from agro-industrial wastewaters and by-products in a bio-refinery framework. *Int. J. Hydrogen Energy* 42 (3), 1609–1621. doi:10.1016/j.ijhydene.2016.09.166

Martinez, B., Yang, J., Bulat, A., and Tzimiropoulos, G. (2020). Training binary neural networks with real-to-binary convolutions. *ICLR 2020*. Available at.1–11.

Martinez-Burgos, W. J., Bittencourt Sydney, E., Bianchi Pedroni Medeiros, A., Magalhaes, A. I., de Carvalho, J. C., Karp, S. G., et al. (2021). Agro-industrial wastewater in a circular economy: Characteristics, impacts and applications for bioenergy and biochemicals. *Bioresour. Technol.* 341, 125795. doi:10.1016/j.biortech.2021.125795

Norfadilah, N., Raheem, A., Harun, R., and Ahmadun, F. (2016). Bio-hydrogen production from palm oil mill effluent (pome): A preliminary study. *Int. J. Hydrogen Energy* 41 (28), 11960–11964. doi:10.1016/j.ijhydene.2016.04.096

Omran, B. A., and Baek, K.-H. (2022). Valorization of agro-industrial biowaste to green nanomaterials for wastewater treatment: Approaching green chemistry and circular economy principles. *J. Environ. Manag.* 311, 114806. doi:10.1016/j.jenvman.2022.114806

Raji, M., Tahroudi, M. N., Ye, F., and Dutta, J. (2022). Prediction of heterogeneous Fenton process in treatment of melanoidin-containing wastewater using data-based models. *J. Environ. Manag.* 307, 114518. doi:10.1016/j.jenvman.2022.114518

Robles, Á., Aguado, D., Barat, R., Borras, L., Bouzas, A., Gimenez, J. B., et al. (2020). New frontiers from removal to recycling of nitrogen and phosphorus from wastewater in the Circular Economy. *Bioresour. Technol.* 300, 122673. doi:10.1016/j.biortech.2019.122673

Sewsynker, Y., Kana, E. B. G., and Lateef, A. (2015). Modelling of biohydrogen generation in microbial electrolysis cells (MECs) using a committee of artificial neural networks (ANNs)', *Biotechnology & Biotechnological Equipment. Biotechnol. Biotechnol. Equip.* 29 (6), 1208–1215. doi:10.1080/13102818.2015.1062732

Sharabiani, V. R., Kaveh, M., Taghinezhad, E., Abbaszadeh, R., Khalife, E., Szymanek, M., et al. (2022). Application of artificial neural networks, support vector, adaptive neuro-fuzzy inference systems for the moisture ratio of parboiled hulls. *Appl. Sci.* 12 (4), 1771. doi:10.3390/app12041771

Sharafati, A., Asadollah, S. B. H. S., and Hosseinzadeh, M. (2020). The potential of new ensemble machine learning models for effluent quality parameters prediction and related uncertainty. *Process Saf. Environ. Prot.* 140, 68–78. doi:10.1016/j.psep.2020.04.045

Sharma, Y., and Li, B. (2010). Optimizing energy harvest in wastewater treatment by combining anaerobic hydrogen producing biofermentor (HPB) and microbial fuel cell (MFC). *Int. J. Hydrogen Energy* 35 (8), 3789–3797. doi:10.1016/j.ijhydene.2010.01.042

Sridevi, K., Sivaraman, E., and Mullai, P. (2014). Back propagation neural network modelling of biodegradation and fermentative biohydrogen production using distillery wastewater in a hybrid upflow anaerobic sludge blanket reactor. *Bioresour. Technol.* 165, 233–240. doi:10.1016/j.biortech.2014.03.074

Taheri, E., Amin, M. M., Fatehizadeh, A., Rezakazemi, M., and Aminabhavi, T. M. (2021). Artificial intelligence modeling to predict transmembrane pressure in anaerobic membrane bioreactor-sequencing batch reactor during biohydrogen production. *J. Environ. Manag.* 292, 112759. doi:10.1016/j.jenvman.2021.112759

Thirugnanasambandham, K., Sivakumar, V., and Prakasmaran, J. (2015). Optimization of process parameters in electrocoagulation treating chicken industry wastewater to recover hydrogen gas with pollutant reduction. *Renewable Energy* 80, 101–108. doi:10.1016/j.renene.2015.01.030

Yogeswari, M. K., Dharmalingam, K., and Mullai, P. (2019). Implementation of artificial neural network model for continuous hydrogen production using confectionery wastewater. *J. Environ. Manag.* 252, 109684. doi:10.1016/j.jenvman.2019.109684

Zaharia, C., Leon, F., Curteanu, S., and Iacob-Tudose, E. T. (2021). Textile wastewater treatment in a spinning disc reactor: Improved performances—experimental, modeling and SVM optimization. *Processes* 9 (11), 2003. doi:10.3390/pr9112003

Zanaty, E. A., and Afifi, A. (2020). Generalized Hermite kernel function for support vector machine classifications. *Int. J. Comput. Appl.* 42 (8), 765–773. doi:10.1080/1206212X.2018.1489571

Zeng, A., Ho, H., and Yu, Y. (2020). Prediction of building electricity usage using Gaussian Process Regression. *J. Build. Eng.* 28, 101054. doi:10.1016/j.jobe.2019.101054

Zhang, L., Chao, B., and Zhang, X. (2020). Modeling and optimization of microbial lipid fermentation from cellulosic ethanol wastewater by Rhodotorula glutinis based on the support vector machine. *Bioresour. Technol.* 301, 122781. doi:10.1016/j.biortech.2020.122781

Zhu, S., Duong, L. H. K., and Liu, W. (2020). "XOR-Net: An efficient computation pipeline for binary neural network inference on edge devices," in 2020 IEEE 26th international conference on parallel and distributed systems (ICPADS), 124–131. doi:10.1109/ICPADS51040.2020.00026