A comparison of artificial intelligence models for predicting phosphate removal efficiency from wastewater using the electrocoagulation process

Majid Gholami Shirkoohi , Rajeshwar D. Tyagi ,
Peter A. Vanrolleghem , Patrick Drogui

PII: S2772-5081(22)00034-5
DOI: https://doi.org/10.1016/j.dche.2022.100043
Reference: DCHE 100043

To appear in: *Digital Chemical Engineering*

Received date: 22 April 2022
Revised date: 30 June 2022
Accepted date: 4 July 2022

Please cite this article as: Majid Gholami Shirkoohi , Rajeshwar D. Tyagi , Peter A. Vanrolleghem , Patrick Drogui , A comparison of artificial intelligence models for predicting phosphate removal efficiency from wastewater using the electrocoagulation process, *Digital Chemical Engineering* (2022), doi: https://doi.org/10.1016/j.dche.2022.100043

A comparison of artificial intelligence models for predicting phosphate removal efficiency from wastewater using the electrocoagulation process

Majid Gholami Shirkoohi[a,d], Rajeshwar D. Tyagi[b], Peter A. Vanrolleghem[c,d], Patrick Drogui[a,d,*] patrick.drogui@ete.inrs.ca

[a]Institut National de la Recherche Scientifique (INRS), Centre-Eau Terre Environnement, Université du Québec, 490, Rue de la Couronne, Québec, QC, G1K 9A9, Canada

[b]BOSK Bioproducts, 399 Rue Jacquard, suite 100, Québec, QC, G1N 4J6, Canada

[c]modelEAU, Département de génie civil et de génie des eaux, Université Laval, 1065 av. de la Médecine, Québec, QC, G1V 0A6, Canada

[d]CentrEau, Centre de recherche sur l'eau, Université Laval, Québec, QC, Canada

[*]Corresponding author

Abstract

In this study, artificial intelligence (AI) models including adaptive neuro-fuzzy inference systems (ANFIS), artificial neural networks (ANN), and support vector regression (SVR) were applied to predict the removal efficiency of phosphate from wastewaters using the electrocoagulation process. The five input variables used in this study were current intensity, initial phosphate concentration, initial pH, treatment time, and electrode type. The optimal hyperparameters of the ANN and SVR models were found by integrating evolutionary algorithms such as genetic algorithms (GA) and particle swarm optimization (PSO) to these models. To increase the reliability and robustness of the developed AI models, a search for optimal hyperparameters was conducted based on repeated random sub-sampling validation instead of a single split approach. The results demonstrated that the effectiveness of the data-driven model depends on how the data is distributed to the training, validation, and test sets. However, hybrid ANN models

outperformed other models and PSO-ANN models showed exceptional generalization performance for the different sub-datasets. The average MSE, R^2, and MAPE values of the 10 test subsets for PSO-ANN were determined as 7.201, 0.981, and 2.022, respectively. The EC process was interpreted for phosphate removal efficiency using the trained PSO-ANN model. The two input factors with the greatest influence on the effectiveness of phosphate removal, according to the results, are the electrode type and initial phosphate concentration. Additionally, it was found that lowering the pH and initial phosphate concentration and increasing the current intensity and treatment time enhance the removal efficiency.

Keywords

Data-driven model, Electrochemical process, Hyperparameters, Metaheuristic algorithm, Modelling, Phosphorus removal

Nomenclature

AI	Artificial intelligence
ANFIS	Adaptive neuro-fuzzy inference systems
ANN	Artificial neural networks
EC	Electrocoagulation
FCM	Fuzzy c-mean clustering
GA	Genetic algorithms
MAPE	Mean absolute percentage error
MSE	Mean square error
PSO	Particle swarm optimization
R^2	Correlation coefficient
RBF	Radial Basis Function
SVR	Support vector regression

1. Introduction

Human activities have significantly increased the soluble forms of phosphorus compounds in aquatic environments over the last 50 years, owing to the widespread use of detergents and chemical fertilizers, animal manure, wastewater effluents, and plant residues. Environmental concerns such as eutrophication can be caused by soluble forms of phosphorus in water, compromising the quality and sustainability of water bodies. This phenomenon can deplete oxygen levels in water due to algae breakdown, which can harm fish and other aquatic life, resulting in decreased biodiversity (Ano *et al.*, 2019; Li *et al.*, 2022; Tran *et al.*, 2012). Chemical precipitation (Lavanya *et al.*, 2021), adsorption (Gizaw *et al.*, 2021), ion exchange (Bektaş *et al.*, 2021), and biological processes (Zhang *et al.*, 2022) have all been developed to lower phosphate levels in wastewater before it is released into the environment. An alternative method for removing phosphates is electrochemical treatment, such as electrocoagulation (EC) (Kobya *et al.*, 2021). The main advantages of the (EC) process are the ease of use of the equipment, the ease of automation, and the process efficiency in the treatment of a wide range of pollutants. Furthermore, because this process does not necessitate the use of chemicals, treatment costs are reduced. Metal cations are released in situ by electrodissolution of an Al or Fe anode immersed in the effluent, as opposed to chemical precipitation (Jing *et al.*, 2021).

Modelling and optimization of the electrochemical process are seen as a key part of the study in order to examine the efficacy of the process. The concentration of pollutants, the applied current density and electrical potential, the types of electrodes, the electrolyte type and concentration, and chemical interactions between contaminants are all important aspects in electrochemical processes for water and wastewater treatment (Drogui *et al.*, 2007). Phenomenological and empirical modelling approaches are generally used for water treatment processes. Electrochemical processes for water treatment are highly complicated nonlinear systems due to the complex relationships between input parameters and outputs. This is due to the fact that several mechanisms usually happen at the same time in an electrochemical system. For instance, In the electrocoagulation process, detailed mechanisms of charge transport, electrochemical

kinetics, thermodynamics, adsorption isotherms and kinetic models, flocculation, flotation, settling, and complexation should be known (Hakizimana *et al.*, 2017). Also, in electrooxidation, the concentration of every compound in an electrochemical cell depends on time and space, that is, their distance from the electrode surface. Describing the profile of compounds under such conditions involves a number of partial differential equations, which are often difficult to solve and involve many model parameters. The complexity of these models depends on the number of species included in the model. In a multivariable model, all the significant species in an electrochemical cell are included. This however requires further knowledge on reaction pathways to account for subsequent formations and transformations (Cañizares *et al.*, 2004a; Cañizares *et al.*, 2004b). Empirical (regression) modelling is one of the alternative modelling approaches for phenomenological modelling. In most cases, a quadratic linear regression model will be selected that is often not adequate to describe the nonlinearities of the systems (Nandi *et al.*, 2004). Therefore, it is not always the best option to use phenomenological or conventional empirical models to model, simulate, and optimize the processes. Given the diversity of aspects, modelling, simulation, and optimization of these processes are challenging. Artificial intelligence methods such as artificial neural networks (ANN), adaptive neuro-fuzzy inference systems (ANFIS), support vector regression (SVR), and evolutionary algorithms such as genetic algorithms (GA) and particle swarm optimization (PSO) have emerged as appealing approaches for modelling and optimizing these nonlinear processes. These data-driven models are based on empirical data and linkages between process input and output variables rather than process knowledge.

Some recent studies have shown the application of AI models in electrochemical processes for water and wastewater treatment with reasonable accuracy (Farzin *et al.*, 2020; Gholami Shirkoohi *et al.*, 2021; Taheri *et al.*, 2013). Farzin *et al.* (2020) applied different approaches to data mining, including ANN, SVM, M5 model tree, and ANFIS for electrochemical removal of Ciprofloxacin (CIP) as a model pollutant. The interpolation method was used as an augmentation technique to increase the number of data samples in the dataset. Based on multi-criteria decision-making, it was found that M5 and the SVM model (tuned by the firefly optimization algorithm) had the best

performance and could be used for different tasks, such as determining the optimal removal of the drug, and investigating the impact of different parameters on drug removal process. Curteanu *et al.* (2014) applied two machine learning techniques (artificial neural networks and support vector machines) for the prediction of the performance of an electrooxidation method to decrease the organic compounds and remove micro-organisms from activated sludge effluent. It was reported that overall, the SVM outperformed the ANN models when comparing correlation coefficients. ANFIS, along with ANN, has also been studied for the treatment of greywater using electrocoagulation by Nasr et al. (Nasr *et al.*, 2016). Comparison between the mentioned AI techniques in other fields of science has also been studied (Azad *et al.*, 2019; Zakeri *et al.*, 2022).

AI models include built-in hyperparameters that should be fine-tuned so that the model can solve the machine learning problem to its full potential. The learning process is controlled by these hyperparameters, which directly impact the model performance. In the case of ANN models, these hyperparameters include training algorithms, number of epochs, maximum validation failure, number of hidden layers and hidden neurons, and transfer functions (Viana *et al.*, 2018). For SVR models, the penalty factor, margin of tolerance for errors, and the type of kernel function and the kernel parameters should be optimally selected (Rui *et al.*, 2019). Membership functions and the number of clusters affect the performance of ANFIS models (Abdulshahed *et al.*, 2015). Grid search and random search are often used to search for the optimal values within the space of hyperparameters of the AI models. In the case of a large space to investigate, grid search would be too computationally intensive and slow, and with random search, there is a chance not to be able to find the optimal hyperparameters (Menapace *et al.*, 2021). Recently, metaheuristic algorithms such as genetic algorithm (Gu *et al.*, 2011), particle swarm optimization (Huang *et al.*, 2021), firefly algorithm (FA) (Zhang *et al.*, 2019), ant colony optimization algorithm (ACO) (Jiang *et al.*, 2020), and bat algorithm (BA) (Hafezi *et al.*, 2015) have been used to efficiently tune and optimize the AI models' parameters. For instance, Viana *et al.* (2018) used the PSO algorithm to optimize neural network model hyperparameters, including the hidden neuron number, the transfer function, and the learning rate.

With the significant investment of time and money in experimental work, only a limited number of samples is available in datasets for data-driven models. As the use of data-driven models in the field of electrochemical processes for water and wastewater treatment expands, model reliability and robustness become increasingly important. Apart from hyperparameters, the allocation of the data to training and testing sets influences the performance of the AI model generated with relatively small sample sizes. The use of single split training and test sets (hold-out cross-validation) methodologies is a standard research technique, but it is problematic given the limited sample sizes of experimental datasets accessible from electrochemical processes (Singh *et al.*, 2021). There have been some studies focusing on the tuning hyperparameters of the AI models based on single split cross-validation using trial-and error approach for electrochemical processes in wastewater treatment (Ahmed Basha *et al.*, 2010; da Silva Ribeiro *et al.*, 2019; Piuleac *et al.*, 2010; Sangal *et al.*, 2015; Valente *et al.*, 2014). One of the most thorough studies on the effect of various network architectures and parameters on the modelling performance was performed by Hasani *et al.* (2018) for the modelling of alternating pulse current electrocoagulation-flotation (APC-ECF) for humic acid (HA) removal. Their study focused on the effect of various network architectures and parameters (e.g., two different ANN architectures as MLP and generalized feedforward, number of hidden neurons, transfer functions, and learning parameters) on the modelling performance. However, the reliability of AI models used in electrochemical processes for water and wastewater treatment in the context of the mentioned issue has not been considered thoroughly in the literature.

In this paper, different AI models including ANFIS, ANN, and SVR are developed to predict the removal efficiency of phosphate from wastewaters using the electrocoagulation process. To optimize the hyperparameters of the SVR and ANN models, GA and PSO have been integrated as the proposed approach. To increase the robustness of the AI models with the optimal hyperparameters with respect to the division of the data between training and testing sets, repeated random sub-sampling validation has been utilized for the hybrid models. This would help to find hyperparameters that are optimal values for the different variations of data distributions to reduce the related uncertainty and improve the reliability and robustness of the

developed AI models. In order to illustrate the predicting performance of the proposed models, results were compared based on statistical indices. The best AI model trained with optimal hyperparameters found was used to interpret the phosphate removal efficiency by electrocoagulation process.

2. Development of the AI models

2.1. Electrocoagulation process

Electrocoagulation (EC), developed from chemical coagulation, produces coagulant agents (Fe^{2+}/Fe^{3+} or Al^{3+}) in-situ to effectively remove pollutants by deposition on the cathode or by floatation caused by the generation of hydrogen gas at the cathode (Asselin *et al.*, 2008a). The schematic of the EC process is shown is Fig. 1. The following equations describe the main reactions occurring in an EC cell:

$$\text{At the anode: } M_{(s)} \rightarrow M_{(aq)}^{n+} + ne^{-}$$

$$\text{At the cathode: } 2H_2O + 2e^{-} \rightarrow 2OH^{-} + H_2$$

$$\text{In the bulk solution: } M_{(aq)}^{n+} + nOH^{-} \rightarrow M(OH)_{n(s)}$$

where M(s) is the metal, $M_{(aq)}^{n+}$ refers to the metallic ion (iron or aluminum ion), $M(OH)_{n(s)}$ represents the metallic hydroxide, and ne^{-} is the number of electrons transferred in the reaction at the electrode. It is worth mentioning that Eq. **Error! Reference source not found.** describes a simple case of metallic hydroxide formation. In fact, depending on the pH and the type of metal involved, the formation of different metallic complex species is possible (Dia *et al.*, 2017).

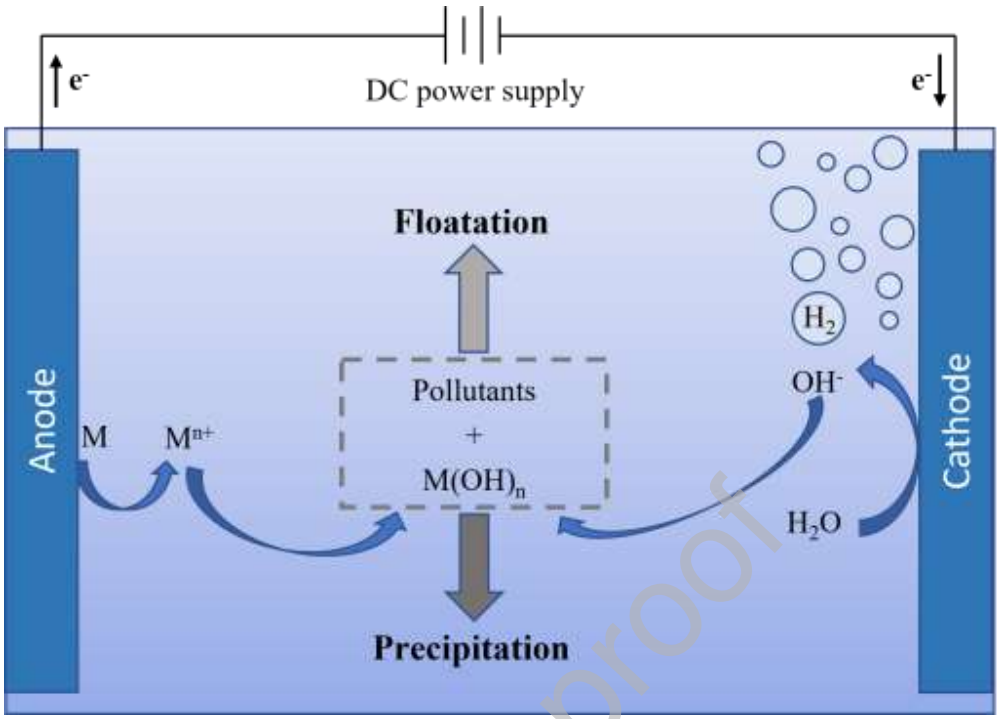


Fig. 1. Schematic of the electrocoagulation process

EC has several advantages over chemical coagulation, such as easy automation, low salinity of the effluent after treatment, low footprint, and reduced production of solid residuals. On the other hand, some disadvantages may include the necessity for regular replacement of sacrificial anodes, electrode passivation, and operating cost where electricity is not abundant (An *et al.*, 2017; Drogui *et al.*, 2007). The EC process has been widely studied for environmental applications to treat drinking water, urban wastewater, textile wastewater, restaurant wastewater, refractory oily wastewater, and heavy metal containing wastewaters (Al-Shannag *et al.*, 2015; Asselin *et al.*, 2008b; Daghrir *et al.*, 2012; Elazzouzi *et al.*, 2017; Kobya *et al.*, 2014; Mólgora *et al.*, 2013).

Several factors affect the efficiency of the electrocoagulation process (Moussa *et al.*, 2017), including:

- Current density: The quantity of metal ions discharged from the electrodes is controlled by current density, which is the current per area of electrode.

- Electrode material: Since it affects the processes that might occur, choosing the right electrode material is essential. Due to their availability and reliability, aluminum and iron electrodes are most frequently utilized.
- Initial pH: When it comes to electrocoagulation, pH is a crucial factor since it has an impact on the solution's conductivity, zeta potential, and electrode dissolution.
- The concentration of ions: The destabilizing characteristics of metal ions are affected differently by the presence of various anions. Additionally, a significant component that impacts EC efficiency and power consumption is the solution's conductivity; the higher the conductivity, the lower EC's power consumption.
- Electrode arrangement: Monopolar or bipolar electrode configurations in series or parallel connections are employed when a large electrode surface area is required.

2.1. Data acquisition

To develop the AI models, a total number of 62 experimental data for the removal of phosphate from synthetic wastewaters using an electrocoagulation process was gathered from Ano *et al.* (2019). In their study, factorial design (FD) and central composite design (CCD) were used as response surface methodology (RSM) to investigate the effect of current intensity, initial phosphate concentration, initial pH, treatment time, and electrode type. Table 1 shows the description and statistical parameters of the dataset used in this study. The experimental matrix is obtained by combining the coded values [-2, -1, 0, 1, 2] of all factors. The last two rows of the Table 1 show the coded and actual experimental domain for all the independent variables in which the 62 experimental runs were conducted. The FD consisting of 32 experiments is designed at the coded experimental domain of [-1, 1] for all the independent variables. The CCD, including the other 30 experiments, is obtained at the coded experimental domain of [-2, 0, 2]. The CCD consists of 16 experiments at the axial points ([-2, 2]) and 14 experiments at the center points.

Table 1. Description of the dataset from Ano *et al.* (2019) used in this study

Statistical parameters	Independent variables/Inputs					Dependent variable/Output
	Current Intensity (A)	Initial phosphate concentration (mg/L)	pH	Treatment time (min)	Electrode type	Removal efficiency (%)
Number of samples	62	62	62	62	62	62
Range	0.25-1.25	15-75	2-10	10-90	Al/Fe	29.2-100
Average	0.75	45	6	50	-	74.0
Standard deviation	0.22	13.3	1.77	17.7	-	20.6
Coded experimental domain	[-2, -1, 0, 1, 2]	[-2, -1, 0, 1, 2]	[-2, -1, 0, 1, 2]	[-2, -1, 0, 1, 2]	-	-
Actual experimental domain	[0.25, 0.5, 0.75, 1, 1.25]	[15, 30, 45, 60, 75]	[2, 4, 6, 8, 10]	[10, 30, 50, 70,90]	-	-

2.2. Adaptive neuro fuzzy inference system

ANFIS, introduced by Jang (Jang, 1993) is a hybrid technique of artificial intelligence in which a Sugeno-type Fuzzy Inference System (FIS) and an artificial neural network (ANN) are combined. Fuzzy Logic produces fuzzy rules that map the inputs to an output based on a given input-output data set. Using the ANFIS hybrid approach, an initial fuzzy model based on fuzzy logic from the input-output data of the system is derived. Then, the neural network learning process is utilized to fine-tune the rules of the initial fuzzy model to generate the optimal ANFIS model of the system. Therefore, ANFIS benefits from the advantages of Fuzzy Logic and neural networks in a single integrated system (Buragohain, 2009).

The FIS structure can be generated by different strategies, including grid partitioning, Fuzzy C-mean clustering (FCM) and the subtractive clustering method (SCM). FCM integrated with ANFIS helps obtaining a relatively small number of rules which prevents the model from being too complex and reduces the risk of overfitting. The FCM clustering method partitions the input data into different clusters and is used to identify the fuzzy membership functions and fuzzy rule base for the ANFIS model (Melin *et al.*,

2014). In this study, FCM has been used for the ANFIS model and the number of clusters will be manually selected for the best generalization performance. The architecture of the ANFIS models with two clusters (which is identified in section 3.1) used for phosphate removal efficiency is given in Table 2.

Table 2. The ANFIS model characteristics

Parameters	Value/description
Fuzzy structure	Takagi-Sugeno
Initial FIS generated	Fuzzy C-Means clustering
Input membership function type	Gaussian ('gaussmf')
Output membership function type	Linear
No. of clusters	2
Optimization method	Hybrid (least-squares and backpropagation gradient descent method)
Number of linear parameters	10
Number of nonlinear parameters	16
Total number of parameters	26
No. of fuzzy rules	2
No. of maximum epoch	100
Number of inputs	5
Number of outputs	1

2.3. Support vector regression

Support vector regression (SVR) is an extension of support vector machines (SVM), first presented by Vapnik (1963), used for prediction and regression problems. Due to its promising generalization performance, SVR has been widely applied to regression prediction problems (Saradhi *et al.*, 2007). In SVR, the main goal is to obtain a predictor function $f(x)$ that describes the relationship between input and output data with an error value less than ε for all the training data. At the same time, the function $f(x)$ is required to be as flat as possible, meaning that the errors are not significant as long as they are less than ε, but any deviation larger than this amount is not tolerated. The SVR models were developed based on the standardized data of five input variables and one output

variable. In this work, dummy variables were generated for the categorical input of electrode type.

2.3.1. SVR parameters optimization

The performance of the SVR model highly depends on the accurate selection of its hyperparameters. These include the box constraint (C), the acceptable error epsilon (ε), the type of kernel function, and the kernel parameter. The box constraint C is a trade-off between model complexity and generalization ability. The magnitude of the penalty for samples with losses greater than ε is determined by the C-value. With too small C-values, the model will be simpler (less complex), but the training error will increase. On the other hand, when C is too large, the empirical risk (the second term in the regression risk R) will be the dominant term for the minimization objective, which results in the overfitting issue (Rui *et al.*, 2019). The ε hyperparameter influences the number of support vectors and, hence, the performance of the SVR by determining the size of the ε-insensitive zone. The kernel function and its relevant parameter maps nonlinear input data into the higher dimensional feature space to help SVR handle nonlinear problems (Alade *et al.*, 2019).

Therefore, the three hyperparameters of C, ε, and kernel parameter (σ for RBF and q for polynomial kernel function) should be selected carefully in view of the effectiveness of the SVR model. In this study, PSO and GA are applied to find the optimal values of these hyperparameters. It should be noted that in the case of a linear kernel function, only C and ε will be optimized by the optimization algorithms.

2.4. Artificial neural networks

An ANN imitates the essential characteristics of the human brain (which itself is a highly nonlinear, complex, and parallel computer), such as self-adaptability, self-organization, and error tolerance (Haykin, 1998). Considering how the different neurons are positioned and connected to each other as well as the composition of layers, various ANN architectures can be generated. The multilayer perceptrons (MLP) feedforward network is one of the most common ANN architectures. These networks are applied to

diverse problems, including function approximation, pattern classification, system identification, process control, process optimization, and so on (Gholami Shirkoohi *et al.*, 2021). The weights of the connections between the neurons are adjusted in the training process. Multiple optimization algorithms can be used for the training process, such as gradient descent, Levenberg–Marquardt, and Bayesian Regularization backpropagation. MLP-ANN models with five input neurons and one output neuron have been developed to predict the phosphate removal efficiency of the EC process with the standardized data.

2.4.1. ANN parameters optimization

The hyperparameters of an ANN model, which define its topology and learning options, influence the accuracy and effectiveness of the trained model. The numbers of hidden layers and neurons in each hidden layer, training algorithm, transfer functions, the regularization parameter, the learning algorithm, and the maximum validation failure are considered as ANN hyperparameters (Shirkoohi *et al.*, 2021; Sinha *et al.*, 2021; Valencia *et al.*, 2021; Viana *et al.*, 2018). In this study, the selection of the number of hidden neurons, the training algorithm, the type of transfer function in the hidden layer, and the number of maximum validation failures are considered for tuning with PSO and GA. As training algorithms Gradient descent with momentum and adaptive learning rate backpropagation (traingdx), Levenberg-Marquardt backpropagation (trainlm), and Bayesian Regularization backpropagation (trainbr) have been selected for the optimization process. Also, transfer functions including Log-sigmoid (logsig), hyperbolic tangent sigmoid (tansig), and positive linear (poslin) were chosen as options.

2.5. Hybrid ANN and SVR models

Metaheuristic optimization algorithms, including GA and PSO will be used in this study to search for optimal hyperparameters of the developed SVR and ANN models. Genetic Algorithms, first proposed by Holland (Holland, 1992) are evolutionary search and optimization algorithms based on natural selection. GAs, thanks with good global searching ability, flexibility, no need for gradient information of the objective (fitness)

functions, and ease of operation, have been a powerful technique for optimization problems (Curteanu *et al.*, 2007; Ding *et al.*, 2011). PSO, first introduced by Kennedy *et al.* (1995), is based on the social behaviour simulation of a flock of birds, called 'swarm', searching for food. PSO is a stochastic population-based optimization approach in which particles, a swarm of potential solutions, fly in the problem space to find better regions and finally the optimal solution, while cooperating and competing with other ones (Chen *et al.*, 2010). GA and PSO have been used to find optimal hyperparameters of AI models in the other disciplines of science (Amar *et al.*, 2020; Gao *et al.*, 2020; Haznedar *et al.*, 2018; Mandal *et al.*, 2015; Wang *et al.*, 2015).

Fig. 2 shows the flowchart of the integrated GA and PSO to find the optimal hyperparameters of the models. The repeated random sub-sampling validation or Monte Carlo cross-validation (Picard *et al.*, 1984) is used in this study as applied before in literature (Altaf *et al.*, 2016; Cao *et al.*, 2020; Severeyn *et al.*, 2019). In the simplest and most common cross-validation technique, single-split (hold-out cross-validation), the dataset is divided into training, validation, and test sets where the model will be trained and tuned on the training and validation sets and tested on the test set. In the repeated random sub-sampling validation, which is a variation of k-fold cross-validation in case that k is the number of the times the model will be trained, not the number of folds. In this cross-validation technique, on every iteration, samples will be randomly selected as training, validation, and test sets. To have an equal distribution of data for the different AI models, at first, 10 sub-datasets are generated randomly, each containing 62 data points. From these 10 sub-datasets, 42 data points will be selected for training, 10 data points for validation, and 10 data points for testing. The search for the optimal hyperparameters will be conducted using all 10 sub-datasets. This helps to find hyperparameters that result in the best performance available for the 10 sub-datasets and to overcome the uncertainty related to the use of single split training, validation and test sets method with limited data points available.

To evaluate the performance of the AI models, the mean square error (MSE), the correlation coefficient (R^2), and the mean absolute percentage error (MAPE) were used

as comparison criteria. To train and validate the hybrid models, MSE was used as the error function. These functions were calculated as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

where y_i is the experimental value, $\hat{y}_i$ is the predicted value, and $\bar{y}$ is the average value of the experimental data.

MATLAB (version R2019a) has been used in our study to develop and apply different AI techniques. Hardware used was Intel(R) Core(TM) i7 with 16 GB RAM and equipped and an NVIDIA T500. Overall computation times for ANFIS, PSO-SVR, GA-SVR, PSO-ANN, and GA-ANN were recorded as 110 s, 1127 s, 1148 s, 4851 s, and 4897 s, respectively. Since the ANFIS approach does not include an optimization algorithm, it has much shorter computation times than the hybrid models.

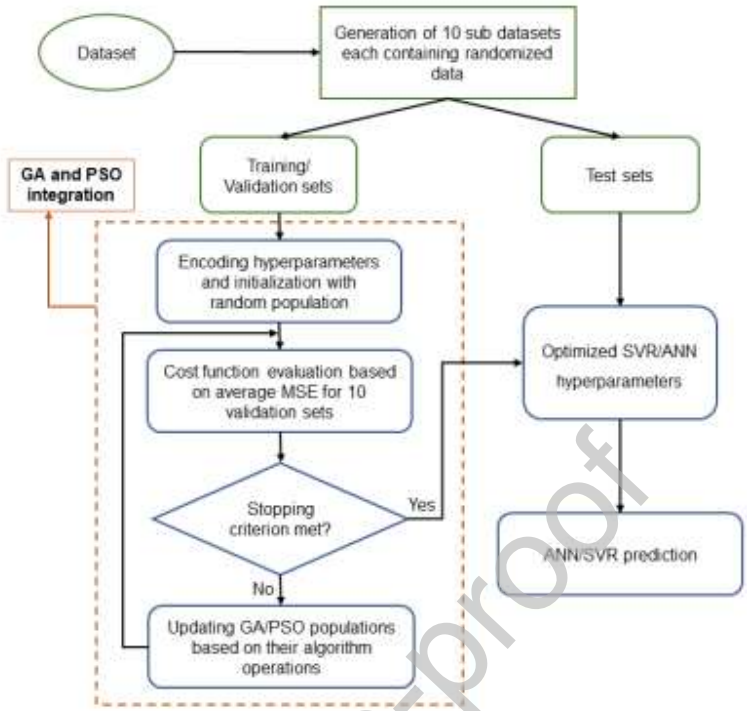


Fig. 2. Flowchart of the proposed hybrid models

3. Results

3.1. ANFIS model

The number of clusters in FCM-FIS generation affects the performance of the ANFIS model. Table 3 shows the impact of increasing the number of clusters on the performance of trained ANFIS models in training, validation, and test. As can be seen, setting a higher number of clusters for an ANFIS model results in having better performance on the training set by making the model more complex. However, this causes the model to lose generalization capability and to risk overfitting. This is evident from the performance of the models on the validation and test sets when higher number of clusters are adopted. Although both models with two and three clusters perform well on the validation and test stage, the model with two clusters would have a lower total number of parameters than a model with three clusters (38 parameters compared to 57)

and hence leads to a simpler model. Therefore, ANFIS models with two clusters were selected.

Table 3. Error analysis as a function of the number of clusters in ANFIS

No. of clusters	MSE			R^2			MAPE		
	Train	Validation	Test	Train	Validation	Test	Train	Validation	Test
2	**31.642**	**68.733**	**60.763**	**0.924**	**0.794**	**0.835**	**6.550**	**9.322**	**9.275**
3	17.647	58.422	96.994	0.958	0.832	0.757	5.066	8.176	10.823
4	16.695	103.391	120.240	0.961	0.709	0.676	4.506	11.294	12.706
5	8.306	151.575	179.181	0.981	0.557	0.504	2.690	12.961	14.650
6	1.558	932.835	1030.522	0.996	-0.986	-2.050	0.757	19.982	22.588

3.2. PSO-SVR and GA-SVR

As mentioned before, the selection of kernel function and its inherent parameter influence the performance of the SVR model. Therefore, it was necessary to select the kernel function of the SVR model before fine-tuning the hyperparameters. For this purpose, the three kernel functions (linear, polynomial, and RBF) were tested on the data using the 5-fold cross-validation method. This method applies the training process on 4 folds of observations and (each time) leaves one fold of observations out to calculate the generalization error of the models. Table 4 shows the results for the different kernel functions used. The generalization error calculated is the out-of-sample MSE. It should be mentioned that the MSE obtained in Table 3 is based on the default values of the hyperparameters and the optimal hyperparameters of the SVR models with the three different kernel functions on the validation sets. As can be seen, the polynomial function leads to the best results on the 5-fold cross-validation method with the default hyperparameters. The kernel parameter, C, and ε should be optimally selected by the hyperparameter optimization algorithm. The three kernel functions were also tested for hyperparameter optimization by PSO-SVR and GA-SVR on the 10 data subsets. It can be seen that optimal SVR models with polynomial kernel function obtain the lowest MSE on the validation sets of PSO and GA algorithms with a population size of 50.

Table 4. Effect of the kernel function on the SVR performance

Kernel functions	MSE	
	Default hyperparameters	PSO and GA optimal hyperparameters
Linear function	119.96	72.89
Polynomial function	105.48	65.43
RBF	361.61	66.09

3.3. GA-ANN and PSO-ANN

GA and PSO were used to find the optimal hyperparameters of the ANN models. A population size of 50 has been considered for the algorithms. Although population size is problem-dependent and no universal value can be proposed, there are studies suggesting that population sizes between 20 and 50 would be appropriate for solving optimization problems (Lobo *et al.*, 2007; Poli *et al.*, 2007; Wang *et al.*, 2018; Zhang *et al.*, 2005). Other control parameters of the GA and PSO are the same as suggested in the literature as described in section 2.5. The number of maximum iterations was set to 200 and 100 for GA and PSO, respectively, to have an equal number of function calls for both algorithms (with respect to convergence), allowing for an unbiased comparison. Indeed, an equal number of iterations for each test, and not an equal number of function calls, may result in a better performance of the optimization algorithm that is attributed to a larger number of function calls (Piotrowski *et al.*, 2020).

Table 5 represents the optimal hyperparameters of the ANN models found by GA and PSO. The prediction accuracy of the hybrid models is given as the average MSE based on the 10 validation sets. It can be seen that both GA and PSO perform reasonably on the validation subsets. The best performances, in terms of average MSE over 10 validation sets, are calculated as 7.686 and 7.830 for GA and PSO, respectively. However, their performance will be evaluated in the next section for the test subsets as well. The Levenberg-Marquardt training algorithm and log-sigmoid transfer function were found to be optimal for both GA and PSO techniques. The Levenberg-Marquardt method has shown its good performance including its ability to converge 10–100 times faster than the conventional gradient descent backpropagation algorithm (Da Silva *et al.*, 2017; Kamosi *et al.*, 2010).

Table 5. Optimal hyperparameters and prediction accuracy results for the hybrid GA and PSO models

Technique	Hyperparameters				MSE
	No. of hidden neurons	Training algorithm	Transfer function	No. of maximum validation failures	
GA-ANN	7	trainlm	logsig	5	7.686
PSO-ANN	9	trainlm	logsig	7	7.830

3.4. Performance comparison of the proposed models

The objective of the proposed AI models including ANFIS, PSO-SVR, GA-SVR, PSO-ANN, GA-ANN was to predict the removal efficiency of phosphate from wastewaters using the electrocoagulation process. The performance of each data-driven model was evaluated by its correspondence with experimental data on training, validation, and test sets. Table 6 represents the statistical analyses of the proposed models with the optimal parameters. Results are provided for training, validation, and test sets as the average of the 10 sub-datasets.

The results in Table 6 show that hybrid ANN models perform better than ANFIS and hybrid SVR models in all training, validation, and test stages. Both PSO-SVR and GA-SVR models showed similar performance results on the datasets. This comes from the fact that both evolutionary algorithms found the same SVR parameters as the optimal parameters. The optimal parameters (C, q, ε) of PSO-SVR and GA-SVR were identified as (1, 3, 3.171). While the ANFIS models performed better in terms of generalization for test sets, it was outperformed by PSO-SVR and GA-SVR in the training and validation steps. The best performance both in training and generalization was obtained by the PSO-ANN models, while the GA-ANN models also showed their effectiveness. The average MSE, R^2, and MAPE values of the 10 sub-datasets for PSO-ANN are determined as 7.201, 0.981, and 2.022 for the test sets, respectively.

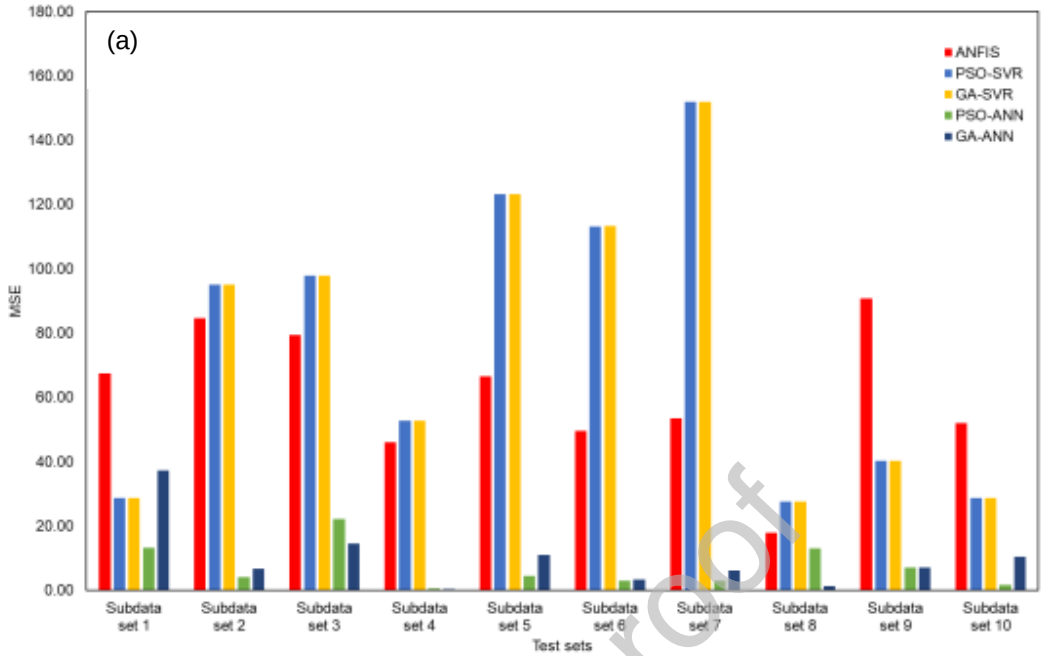
Fig. 3 shows the performance comparison of the AI models on the test sets of each of the 10 data subsets. As can be observed, the effectiveness of the data-driven model depends on how the data is distributed over the training, validation, and test sets. For instance, the hybrid SVR models have low MSE, MAPE, and high R^2 for sub-dataset 1.

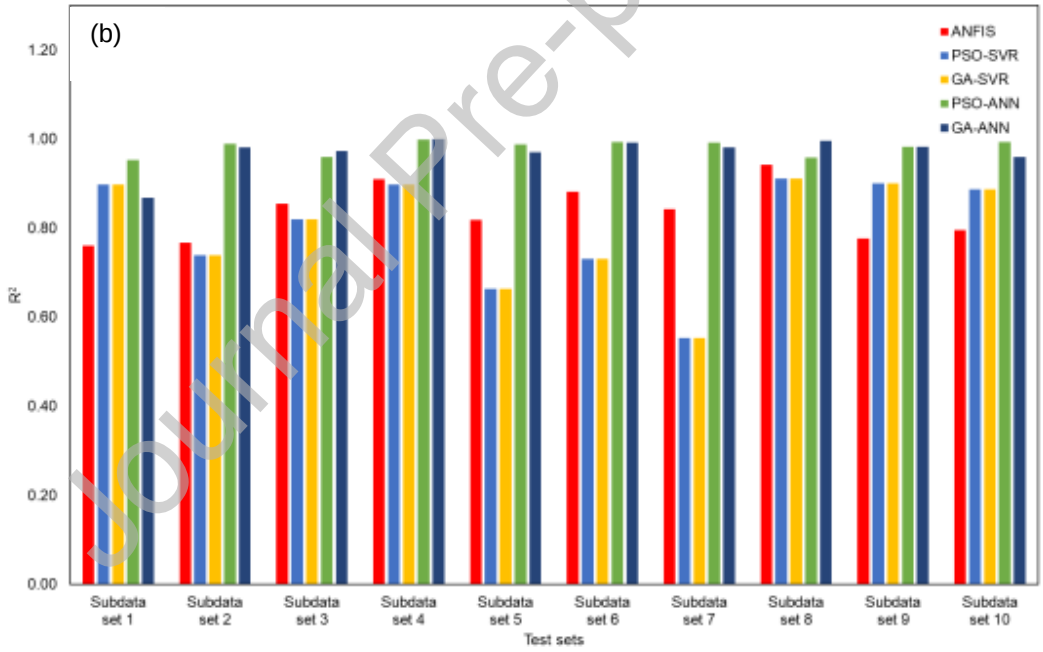
Their performance is comparable to the hybrid ANN models. However, the hybrid ANN models outperform ANFIS and hybrid SVR models for the test sets of other data subsets. Fig. 4 presents the box plots of the performance indices of the AI models for the test sets of 10 data subsets. It shows that there is a greater variability for the hybrid SVR models than for the hybrid ANN models. It can be concluded that generally, hybrid SVR models are more sensitive than hybrid ANN models to the distribution of the data points among the training, validation, and test sets. Nevertheless, hybrid ANN models show less dispersed performance for the test sets of the different data subsets.

According to the mentioned analyses, it can be concluded that the proposed hybrid ANN models have great performance in the prediction of the removal efficiency of phosphate using the electrocoagulation process. The results confirmed that PSO-ANN models have exceptional generalization performance for the different data subsets.

Table 6. Performance evaluation of the developed AI models

Model	MSE			R^2			MAPE		
	Train	Validation	Test	Train	Validation	Test	Train	Validation	Test
ANFIS	31.642	68.732	60.763	0.924	0.793	0.835	6.550	9.322	9.275
PSO-SVR	9.374	65.433	75.893	0.978	0.808	0.800	4.377	8.998	10.779
GA-SVR	9.374	65.433	75.894	0.978	0.808	0.800	4.377	8.998	10.779
PSO-ANN	7.259	7.830	7.201	0.983	0.978	0.981	1.958	2.286	2.022
GA-ANN	8.765	7.686	9.759	0.979	0.978	0.970	2.520	2.747	2.774





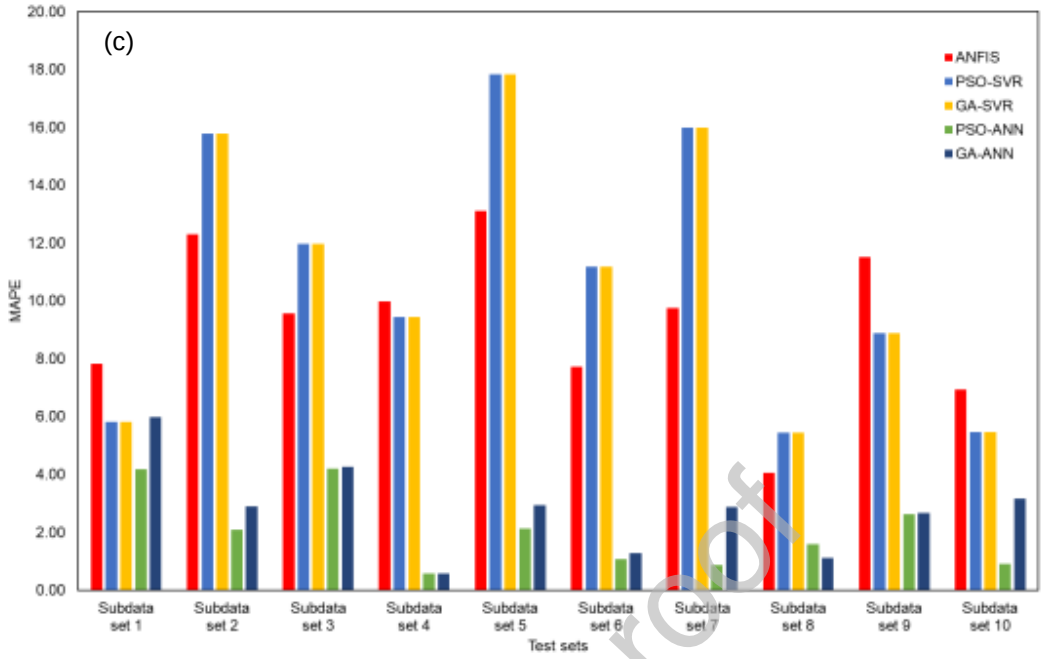


Fig. 3 Performance evaluation of the developed AI models on the test sets of the 10 subsets; (a) MSE, (b) R^2, (c) MAPE

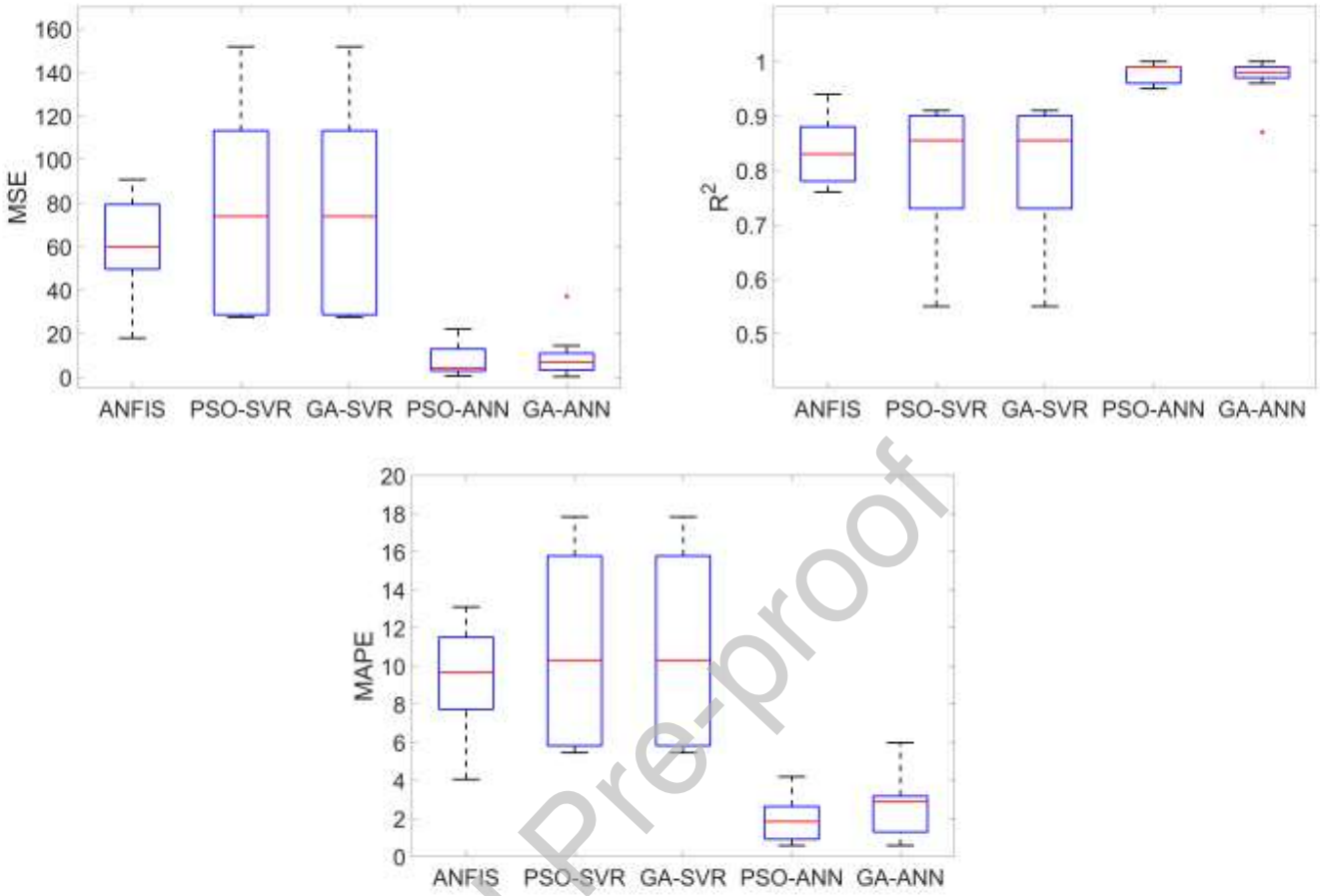


Fig. 4 Boxplot comparison of the performance of the different AI models for the test sets using three performance criteria

3.5. Interpretation of the phosphate removal process with the optimal model

In this section, the optimal PSO-ANN model found will be used to interpret the electrocoagulation process for phosphate removal from wastewater. In this matter, the ANN model will be trained with the optimal hyperparameters found in section 3.3. The performance of the trained model is shown in Fig. 5. As can be seen, there is a good agreement between actual and predicted values for the phosphate removal efficiency (R^2=0.974). Also, the residual error analysis and the histogram shows that the errors are approximately symmetrically distributed around zero with a mean value of -0.19109 and a standard deviation of 3.1788.

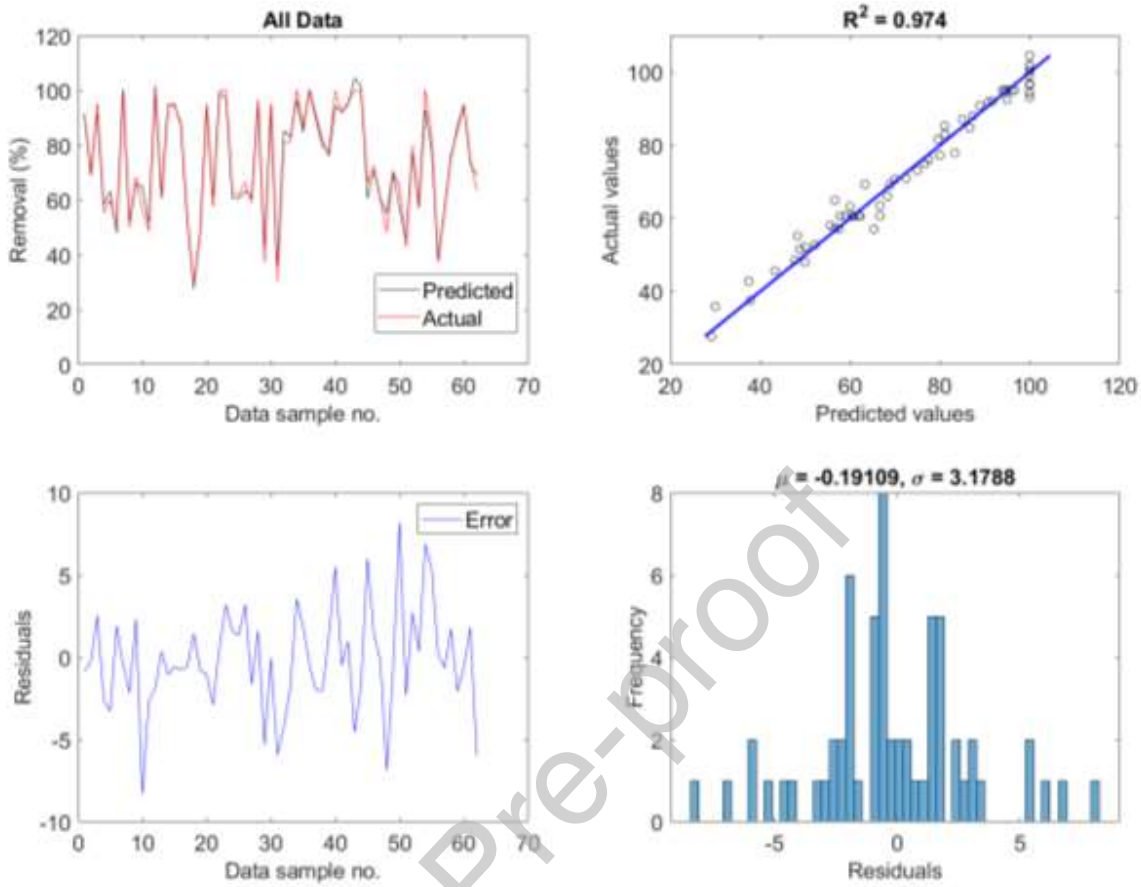


Fig. 5 Performance of the trained ANN model using optimal hyperparameters found by PSO

The weights acquired during ANN training represent synaptic strengths between axons and dendrites in a real neuron in the brain. These weights, like in real life, determine the proportion of the incoming signal that is delivered to the neuron's body (Khataee *et al.*, 2010). Despite the black-box nature of ANNs, a sensitivity analysis on the ANNs may be performed to evaluate the effect of several independent factors on the output. The neural connection weight matrix may be used to determine the relative relevance of each input independent variable on the intended output. First, Garson (Garson, 1991) and then Goh (Goh, 1995) presented a method for partitioning the connection weights in order to establish the relative importance of the various inputs. This method basically involves partitioning the hidden-output connection weights of each hidden neuron into components associated with each input neuron (Zhang *et al.*, 2018). In essence, this strategy entails partitioning each hidden neuron's hidden-output connection weights into

components related to each input neuron. The relative importance of each input variable on the output variable is calculated as follows:

$$I_j = \frac{\sum_{m=1}^{m=N_h}\left(\left(\left|W_{jm}^{jh}\right| / \sum_{k=1}^{N_i}\left|W_{km}^{ih}\right|\right) \times \left|W_{mn}^{ho}\right|\right)}{\sum_{k=1}^{k=N_i}\left[\sum_{m=1}^{m=N_h}\left(\left|W_{km}^{ih}\right| / \sum_{k=1}^{N_i}\left|W_{km}^{ih}\right|\right) \times \left|W_{mn}^{ho}\right|\right]}$$

where I_j is the relative importance of the jth input variable on the output variable, N_i and N_h are the numbers of input and hidden neurons, respectively; the Ws are connection weights, the superscripts i, h and o refer to input, hidden, and output layers, respectively. Also, the subscripts k, m and n refer to input, hidden and output neurons, respectively.

As can be seen in Fig. 6, the relative importance of each individual variable was found as: electrode type > initial phosphate concentration > treatment time > current intensity > pH. The electrode type and initial phosphate concentration are the two most influential input variables on the phosphate removal efficiency. This is in accordance by the results of the Ano *et al.* (2019) obtained by the RSM approach. Surface plots of the ANN model for phosphate removal efficiency as a function of initial phosphate concentration and treatment time at the center point of the current intensity and pH are illustrated in Fig. 7. Generally, it can be said that treatment time has a positive effect on the removal efficiency for both electrodes meaning that with an increase in the treatment time, the removal efficiency improves. However, the initial phosphate concentration has a negative impact on the phosphate removal efficiency.

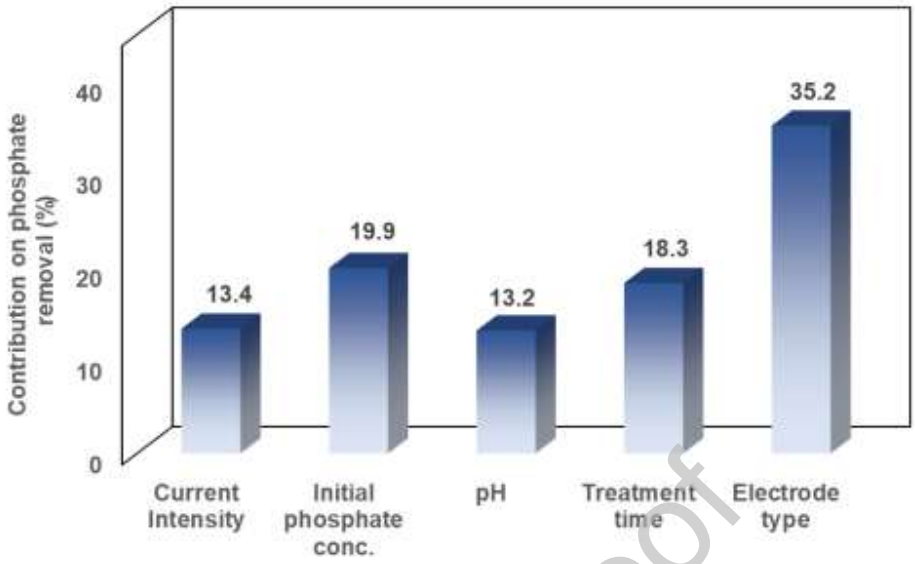
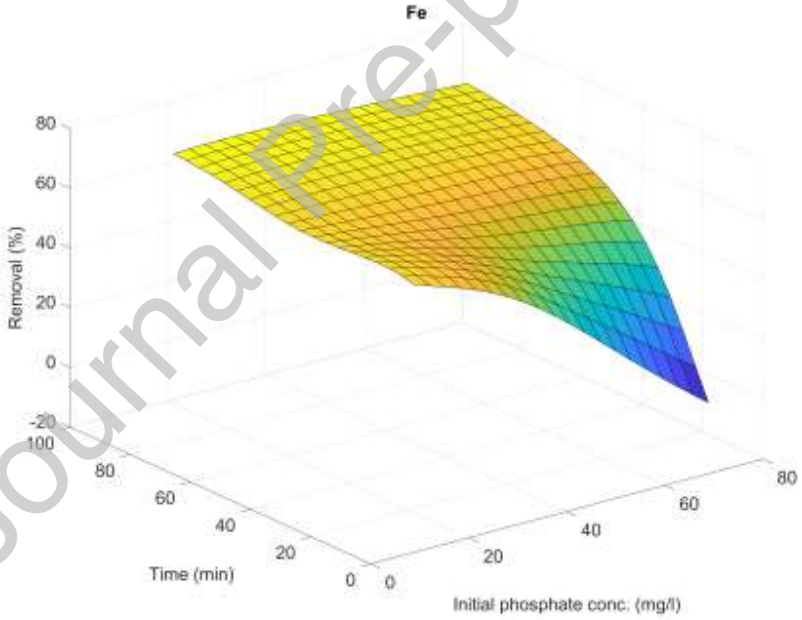


Fig. 6. Relative importance of the input variables on the phosphate removal efficiency



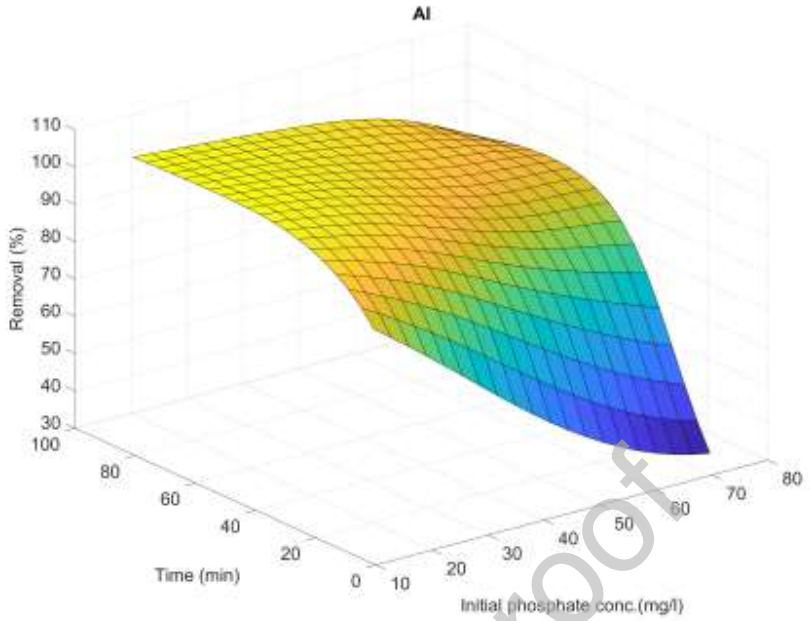


Fig. 7. Surface plots of the ANN model for phosphate removal efficiency as function of initial phosphate concentration and treatment time at a current intensity of 0.75 A and pH of 6.

The main effects of each independent variable on the phosphate removal efficiency are represented in Fig. 8. For each variable, other independent variables were set at their mean values. In all the four plots, it is evident that the Al electrode has better removal efficiency than the Fe electrode. Because of their demonstrated durability and availability for electrocoagulation, Al and Fe electrodes are the most commonly utilized. However, due to its lower positive charge, Fe (II) is a poor coagulant when compared to Fe (III). A lower positive charge suggests a poorer capacity of the ion to compress the electrical double layer/destabilize colloids. Most studies have shown that Al electrodes improve the effectiveness of eliminating pollutants better than Fe electrodes (Moussa *et al.*, 2017). The removal efficiency is improved by increasing the current intensity and electrolysis time. In fact, Faraday's law states that the amount of metal cations produced in the medium, which is dependent on the electrolysis time and current intensity, determines the removal efficiency of phosphate. The higher current intensity and longer electrolysis time help the generation of more metal cations. On the other hand, increasing initial phosphate concentration results in a decrease in removal efficiency for both electrodes. This is because the amount of metal cations formed at the given values

for other variables will likewise be fixed, which may not be enough to coagulate the high phosphate concentrations. Finally, it is shown that the removal efficiency is lower in higher pH (basic condition) than in lower pH (acidic condition) for both electrodes. Due to the predominance of the Al^{3+} and Fe^{3+} cations at acidic pH, a significant number of metal cations would precipitate with the phosphate ions as $AlPO_4$ and $FePO_4$, respectively. The different precipitates $AlPO_4$, $FePO_4$, $Al(OH)_3$, and $Fe(OH)_3$ become more soluble as the pH rises which would account for the decrease in efficiency (Attour *et al.*, 2014). Also, It is worth mentioning that several authors have suggested that in the case of Al electrodes, the maximum performance of the EC system can be reached at around neutral pH, as predicted in this study (Katal *et al.*, 2011; Terrazas *et al.*, 2010). The optimal ANN model found by the PSO helped to interpret the electrocoagulation process for the removal of phosphate from wastewater.

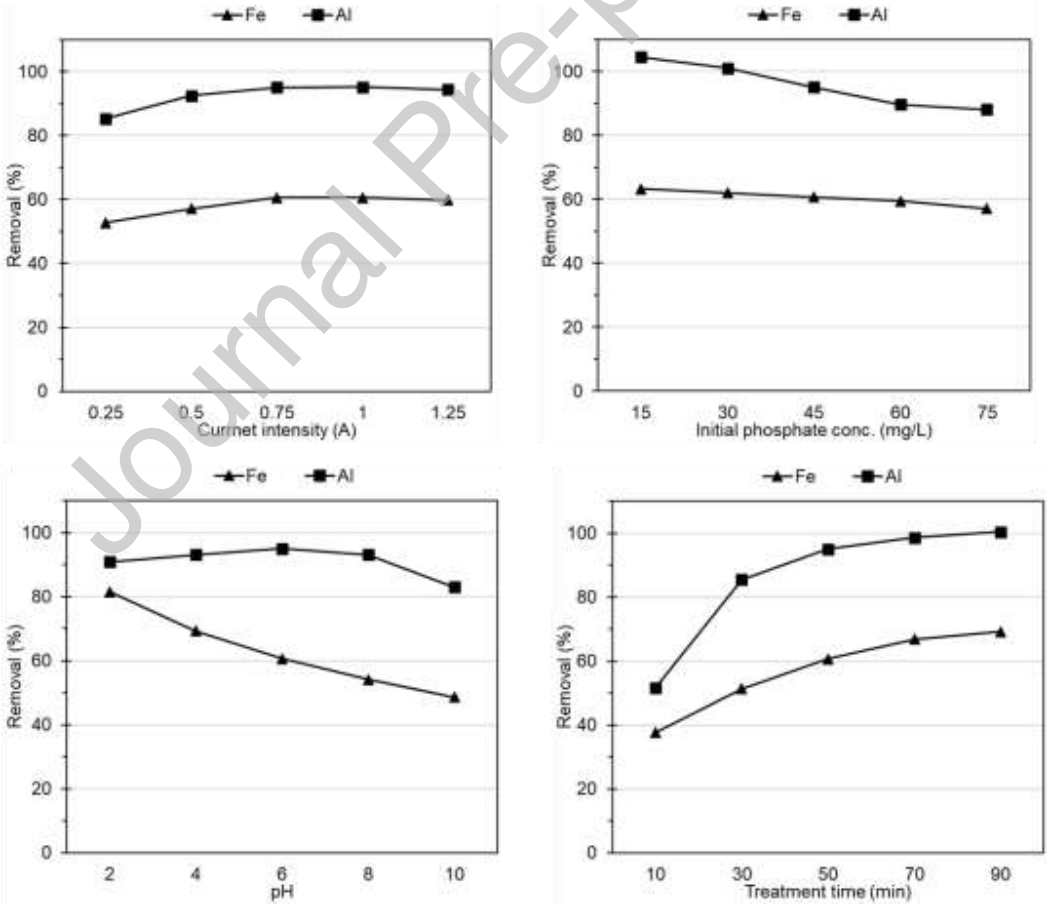


Fig. 8. Main effect plot of each independent variable on caffeine removal efficiency

4. Conclusion

In this study, GA and PSO algorithms were used to optimize the hyperparameters of the SVR and ANN models to forecast the removal efficiency of phosphate from wastewaters using the electrocoagulation process. The current intensity, initial phosphate concentration, initial pH, treatment time, and electrode type were considered as models' inputs. The performance criteria (MSE, R^2, MAPE) comparison of models showed that the effectiveness of the data-driven models depends on how the data is distributed to the training, validation, and test sets. The ANFIS and hybrid SVR models were more sensitive than hybrid ANN models to the distribution of data points. The hybrid ANN models showed greater accuracy than the ANFIS and hybrid SVR models that they were compared to using different performance criteria and indicated less dispersed performance for the test sets of the different sub-datasets. Remarkably, PSO-ANN models illustrated exceptional generalization performance for the 10 data subsets examined. The trained PSO-ANN model was used to interpret the EC process for phosphate removal efficiency. Results showed that the electrode type and initial phosphate concentration are the two most influential input variables on the phosphate removal efficiency. Also, it was found that the removal efficiency is improved by increasing the current intensity and treatment time and decreasing the initial phosphate concentration and the pH. Further research in the application of hybrid evolutionary algorithms and AI models may be carried out in electrochemical processes for water and wastewater treatment with respect to the reliability and robustness of the models.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Sincere thanks are extended to the National Sciences and Engineering Research Council of Canada and the CREATE TEDGIEER program for their financial contribution to this study. Peter A. Vanrolleghem holds the Canada Research Chair on Water Quality Modelling.

References

Abdulshahed AM, Longstaff AP, Fletcher S (2015) The application of ANFIS prediction models for thermal error compensation on CNC machine tools. *Applied Soft Computing* 27:158-168.

Ahmed Basha C, Soloman PA, Velan M, Miranda LR, Balasubramanian N, Siva R (2010) Electrochemical degradation of specialty chemical industry effluent. *Journal of Hazardous materials* 176(1):154-164.

Al-Shannag M, Al-Qodah Z, Bani-Melhem K, Qtaishat MR, Alkasrawi M (2015) Heavy metal ions removal from metal plating wastewater using electrocoagulation: Kinetic study and process performance. *Chemical Engineering Journal* 260:749-756.

Alade I, Saleh T, Abd Rahman MA (2019) Modeling and prediction of the specific heat capacity of Al2O3/water nanofluids using hybrid genetic algorithm/support vector regression model. *Nano-Structures & Nano-Objects* 17:103-111.

Altaf MAB, Zhang C, Radakovic L, Yoo J (2016) Design of energy-efficient on-chip EEG classification and recording processors for wearable environments. *2016 IEEE International Symposium on Circuits and Systems (ISCAS).* (Montreal, Canada, 22-25 May), p 1126-1129.

Amar MN, Zeraibi N, Jahanbani Ghahfarokhi A (2020) Applying hybrid support vector regression and genetic algorithm to water alternating CO2 gas EOR. *Greenhouse Gases: Science and Technology* 10(3):613-630.

An C, Huang G, Yao Y, Zhao S (2017) Emerging usage of electrocoagulation technology for oil removal from wastewater: A review. *Science of the Total Environment* 579:537-556.

Ano J, Assémian AS, Yobouet YA, Adouby K, Drogui P (2019) Electrochemical removal of phosphate from synthetic effluent: A comparative study between iron and aluminum by using experimental design methodology. *Process Safety and Environmental Protection* 129:184-195.

Asselin M, Drogui P, Benmoussa H, Blais JF (2008a) Effectiveness of electrocoagulation process in removing organic compounds from slaughterhouse wastewater using monopolar and bipolar electrolytic cells. *Chemosphere* 72(11):1727-1733.

Asselin M, Drogui P, Brar SK, Benmoussa H, Blais JF (2008b) Organics removal in oily bilgewater by electrocoagulation process. *Journal of Hazardous materials* 151(2-3):446-455.

Attour A, Touati M, Tlili M, Ben Amor M, Lapicque F, Leclerc JP (2014) Influence of operating parameters on phosphate removal from water by electrocoagulation using aluminum electrodes. *Separation and Purification Technology* 123:124-129.

Azad A, Manoochehri M, Kashi H, Farzin S, Karami H, Nourani V, Shiri J (2019) Comparative evaluation of intelligent algorithms to improve adaptive neuro-fuzzy inference system performance in precipitation modelling. *Journal of Hydrology* 571:214-224.

Bektaş TE, Uğurluoğlu BK, Tan B (2021) Phosphate removal by Ion exchange in batch mode. *Water Practice and Technology* 16(4):1343-1354.

Buragohain M (2009) *Adaptive Network based Fuzzy Inference System (ANFIS) as a Tool for System Identification with Special Emphasis on Training Data Minimization*. DOCTOR OF PHILOSOPHY (Indian Institute of Technology Guwahati, India).

Cañizares P, Garcia-Gomez J, Lobato J, Rodrigo MA (2004a) Modeling of wastewater electro-oxidation processes Part I. General description and application to inactive electrodes. *Industrial and Engineering Chemistry Research* 43(9):1915-1922.

Cañizares P, Garcia-Gomez J, Lobato J, Rodrigo MA (2004b) Modeling of wastewater electro-oxidation processes Part II. Application to active electrodes. *Industrial and Engineering Chemistry Research* 43(9):1923-1931.

Cao W, Liu X, Ni J (2020) Parameter optimization of support vector regression using Henry Gas Solubility optimization algorithm. *IEEE Access* 8:88633-88642.

Chen W, Yu X, Wang H, Wen B-g (2010) PSO-GA on Endmember extraction for hyperspectral imagery. *2010 International Conference on Computer Application and System Modeling (ICCASM 2010).* (Taiyuan, China, 22-24 Oct.), p V7-459-V457-464.

Curteanu S, Cazacu M (2007) Neural Networks and Genetic Algorithms used for modeling and optimization of the siloxane-siloxane copolymers synthesis. *Journal of Macromolecular Science Part A-Pure and Applied Chemistry* 45(1):23-36.

Curteanu S, Godini K, Piuleac CG, Azarian G, Rahmani AR, Butnariu C (2014) Electro-Oxidation Method Applied for Activated Sludge Treatment: Experiment and Simulation Based on Supervised Machine Learning Methods. *Industrial & Engineering Chemistry Research* 53(12):4902-4912.

Da Silva IN, Hernane Spatti D, Andrade Flauzino R, Liboni LHB, Dos Reis Alves SF (2017) Artificial Neural Networks : a Practical Course.

da Silva Ribeiro T, Grossi CD, Merma AG, dos Santos BF, Torem ML (2019) Removal of boron from mining wastewaters by electrocoagulation method: Modelling experimental data using artificial neural networks. *Minerals Engineering* 131:8-13.

Daghrir R, Drogui P, François Blais J, Mercier G (2012) Hybrid process combining electrocoagulation and electro-oxidation processes for the treatment of restaurant wastewaters. *Journal of Environmental Engineering* 138(11):1146-1156.

Dia O, Drogui P, Buelna G, Dube R, Ihsen BS (2017) Electrocoagulation of bio-filtrated landfill leachate: Fractionation of organic matter and influence of anode materials. *Chemosphere* 168:1136-1141.

Ding S, Su C, Yu J (2011) An optimizing BP neural network algorithm based on genetic algorithm. *Artificial Intelligence Review* 36(2):153-162.

Drogui P, Blais J-F, Mercier G (2007) Review of electrochemical technologies for environmental applications. *Recent Patents on Engineering* 1:257-272.

Elazzouzi M, Haboubi K, Elyoubi MS (2017) Electrocoagulation flocculation as a low-cost process for pollutants removal from urban wastewater. *Chemical Engineering Research and Design* 117:614-626.

Farzin S, Nabizadeh Chianeh F, Valikhan Anaraki M, Mahmoudian F (2020) Introducing a framework for modeling of drug electrochemical removal from wastewater based on data mining algorithms, scatter interpolation method, and multi criteria decision analysis (DID). *Journal of Cleaner Production* 266:122075.

Gao W, Han J (2020) Prediction of Destroyed Floor Depth Based on Principal Component Analysis (PCA)-Genetic Algorithm (GA)-Support Vector Regression (SVR). *Geotechnical and Geological Engineering* 38(4):3481-3491.

Garson GD (1991) Interpreting Neural-Network Connection Weights. *AI Expert* 6:46-51.

Gholami Shirkoohi M, D. Tyagi R, Vanrolleghem PA, Drogui P (2021) Artificial neural networks and genetic algorithms: An efficient modelling and optimization methodology for active chlorine

production using the electrolysis process. *The Canadian Journal of Chemical Engineering* 99(S1):S389-S403.

Gizaw A, Zewge F, Kumar A, Mekonnen A, Tesfaye M (2021) A comprehensive review on nitrate and phosphate removal and recovery from aqueous solutions by adsorption. *Journal of Water Supply: Research and Technology-Aqua* 70(7):921-947.

Goh ATC (1995) Back-propagation neural networks for modeling complex systems. *Artificial Intelligence in Engineering* 9(3):143-151.

Gu J, Zhu M, Jiang L (2011) Housing price forecasting based on genetic algorithm and support vector machine. *Expert Systems with Applications* 38(4):3383-3386.

Hafezi R, Shahrabi J, Hadavandi E (2015) A bat-neural network multi-agent system (BNNMAS) for stock price prediction: Case study of DAX stock price. *Applied Soft Computing* 29:196-210.

Hakizimana JN, Gourich B, Chafi M, Stiriba Y, Vial C, Drogui P, Naja J (2017) Electrocoagulation process in water treatment: A review of electrocoagulation modeling approaches. *Desalination* 404:1-21.

Hasani G, Daraei H, Shahmoradi B, Gharibi F, Maleki A, Yetilmezsoy K, McKay G (2018) A novel ANN approach for modeling of alternating pulse current electrocoagulation-flotation (APC-ECF) process: Humic acid removal from aqueous media. *Process Safety and Environmental Protection* 117:111-124.

Haykin S (1998) *Neural Networks. A Comprehensive Foundation.* Prentice Hall PTR, USA

Haznedar B, Kalinli A (2018) Training ANFIS structure using simulated annealing algorithm for dynamic systems identification. *Neurocomputing* 302:66-74.

Holland JH (1992) *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence.* MIT Press, United States

Huang W, Liu H, Zhang Y, Mi R, Tong C, Xiao W, Shuai B (2021) Railway dangerous goods transportation system risk identification: Comparisons among SVM, PSO-SVM, GA-SVM and GS-SVM. *Applied Soft Computing* 109:107541.

Jang JR (1993) ANFIS: adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics* 23(3):665-685.

Jiang BT, Zhou J, Huang XB, Wang PF (2020) Prediction of critical heat flux using Gaussian process regression and ant colony optimization. *Annals of Nuclear Energy* 149:107765.

Jing G, Ren S, Pooley S, Sun W, Kowalczuk PB, Gao Z (2021) Electrocoagulation for industrial wastewater treatment: an updated review. *Environmental Science: Water Research & Technology* 7(7):1177-1196.

Kamosi M, Hashemi AB, Meybodi MR (2010) A new Particle Swarm Optimization Algorithm for dynamic environments. *Swarm, Evolutionary, and Memetic Computing.* (Berlin, Heidelberg, 2010) Springer, p 129-138.

Katal R, Pahlavanzadeh H (2011) Influence of different combinations of aluminum and iron electrode on electrocoagulation efficiency: Application to the treatment of paper mill wastewater. *Desalination* 265(1):199-205.

Kennedy J, Eberhart R (1995) Particle Swarm optimization. *ICNN'95 - International Conference on Neural Networks.* (Perth, WA, Australia, 27 Nov.-1 Dec.), p 1942-1948 vol.1944.

Khataee AR, Kasiri MB (2010) Artificial neural networks modeling of contaminated water treatment processes by homogeneous and heterogeneous nanocatalysis. *Journal of Molecular Catalysis A: Chemical* 331(1-2):86-100.

Kobya M, Gengec E, Sensoy MT, Demirbas E (2014) Treatment of textile dyeing wastewater by electrocoagulation using Fe and Al electrodes: optimisation of operating parameters using central composite design. *Coloration Technology* 130(3):226-235.

Kobya M, Omwene PI, Sarabi SM, Yildirim S, Ukundimana Z (2021) Phosphorous removal from anaerobically digested municipal sludge centrate by an electrocoagulation reactor using metal (Al, Fe and Al-Fe) scrap anodes. *Process Safety and Environmental Protection* 152:188-200.

Lavanya A, Sri Krishnaperumal Thanga R (2021) Effective removal of phosphorous from dairy wastewater by struvite precipitation: process optimization using response surface methodology and chemical equilibrium modeling. *Separation Science and Technology* 56(2):395-410.

Li Y, Lu D, Liu X, Li Z, Zhu H, Cui J, Zhang H, Mao X (2022) Coupling of cathodic aluminum dissolution and anodic oxidation process for simultaneous removal of phosphate and ammonia in wastewaters. *Chemical Engineering Journal* 427:130944.

Lobo FG, Lima CF, Michalewicz Z (2007) *Parameter Setting in Evolutionary Algorithms.* Springer Publishing Company, Heidelberg, Germany

Mandal S, Mahapatra SS, Adhikari S, Patel RK (2015) Modeling of Arsenic (III) Removal by Evolutionary Genetic Programming and Least Square Support Vector Machine Models. *Environmental Processes* 2(1):145-172.

Melin P, Castillo O (2014) A review on type-2 fuzzy logic applications in clustering, classification and pattern recognition. *Applied Soft Computing* 21:568-577.

Menapace A, Zanfei A, Righetti M (2021) Tuning ANN hyperparameters for forecasting drinking water demand. *Applied Sciences* 11(9):4290.

Mólgora CC, Domínguez AM, Avila EM, Drogui P, Buelna G (2013) Removal of arsenic from drinking water: A comparative study between electrocoagulation-microfiltration and chemical coagulation-microfiltration processes. *Separation and Purification Technology* 118:645-651.

Moussa DT, El-Naas MH, Nasser M, Al-Marri MJ (2017) A comprehensive review of electrocoagulation for water treatment: Potentials and challenges. *Journal of Environmental Management* 186:24-41.

Nandi S, Badhe Y, Lonari J, Sridevi U, Rao BS, Tambe SS, Kulkarni BD (2004) Hybrid process modeling and optimization strategies integrating neural networks/support vector regression and genetic algorithms: study of benzene isopropylation on Hbeta catalyst. *Chemical Engineering Journal* 97(2):115-129.

Nasr M, Ateia M, Hassan K (2016) Artificial intelligence for greywater treatment using electrocoagulation process. *Separation Science and Technology* 51(1):96-105.

Picard RR, Cook RD (1984) Cross-validation of regression models. *Journal of the American Statistical Association* 79(387):575-583.

Piotrowski AP, Napiorkowski JJ, Piotrowska AE (2020) Population size in Particle Swarm Optimization. *Swarm and Evolutionary Computation* 58:100718.

Piuleac CG, Rodrigo MA, Cañizares P, Curteanu S, Sáez C (2010) Ten steps modeling of electrolysis processes by using neural networks. *Environmental Modelling & Software* 25(1):74-81.

Poli R, Kennedy J, Blackwell T (2007) Particle swarm optimization. *Swarm Intelligence* 1(1):33-57.

Rui J, Zhang H, Zhang D, Han F, Guo Q (2019) Total organic carbon content prediction based on support-vector-regression machine with particle swarm optimization. *Journal of Petroleum Science and Engineering* 180:699-706.

Sangal V, Kaur P, Kushwaha J (2015) Modeling and evaluation of electro-oxidation of dye wastewater using artificial neural networks. *RSC Advances* 5:34663-34771.

Saradhi VV, Karnick H (2007) Classifier complexity reduction by support vector pruning in kernel matrix learning. *Computational and Ambient Intelligence.* (Berlin, Heidelberg, Springer, p 268-275.

Severeyn E, Velásquez J, Herrera H, Wong S (2019) Random sub-sampling cross validation for empirical correlation between heart rate variability, biochemical and anthropometrics parameters. *Information and Communication Technologies of Ecuador (TIC.EC).* (Cuenca, Ecuador, November 27-29) Springer International Publishing, p 347-357.

Shirkoohi MG, Doghri M, Duchesne S (2021) Short-term water demand predictions coupling an artificial neural network model and a genetic algorithm. *Water Supply* 21(5):2374-2386.

Singh V, Pencina M, Einstein AJ, Liang JX, Berman DS, Slomka P (2021) Impact of train/test sample regimen on performance estimate stability of machine learning in cardiovascular imaging. *Scientific Reports* 11(1):14490.

Sinha T, Sikka K, Lall R (2021) Artificial neural networks and bayesian techniques for flip-chip package thermo-mechanical analysis. *IEEE 71st Electronic Components and Technology Conference (ECTC).*1 June-4 July), p 1442-1449.

Taheri M, Alavi Moghaddam MR, Arami M (2013) Techno-economical optimization of Reactive Blue 19 removal by combined electrocoagulation/coagulation process through MOPSO using RSM and ANFIS models. *Journal of Environmental Management* 128:798-806.

Terrazas E, Vázquez A, Briones R, Lázaro I, Rodríguez I (2010) EC treatment for reuse of tissue paper wastewater: Aspects that affect energy consumption. *Journal of Hazardous materials* 181(1):809-816.

Tran N, Drogui P, Blais J-F, Mercier G (2012) Phosphorus removal from spiked municipal wastewater using either electrochemical coagulation or chemical coagulation as tertiary treatment. *Separation and Purification Technology* 95:16-25.

Valencia F, Arcos H, Quilumba F (2021) Prediction of stress in power transformer winding conductors using artificial neural networks: Hyperparameter analysis. *Energies* 14(14):4242.

Valente GFS, Mendonça RCS, Pereira JAM, Felix LB (2014) Artificial neural network prediction of chemical oxygen demand in dairy industry effluent treated by electrocoagulation. *Separation and Purification Technology* 132:627-633.

Vapnik V (1963) Pattern recognition using generalized portrait method. *Automation and Remote Control* 24:774-780.

Viana DF, Salazar-Banda GR, Leite MS (2018) Electrochemical degradation of Reactive Black 5 with surface response and artificial neural networks optimization models. *Separation Science and Technology* 53(16):2647-2661.

Wang D, Tan D, Liu L (2018) Particle swarm optimization algorithm: an overview. *Soft Computing* 22(2):387-408.

Wang J-S, Ning C-X (2015) ANFIS Based Time Series Prediction Method of Bank Cash Flow Optimized by Adaptive Population Activity PSO Algorithm. *Information* 6(3):300-313.

Zakeri MS, Mousavi S-F, Farzin S, Sanikhani H (2022) Modeling of Reference Crop Evapotranspiration in Wet and Dry Climates Using Data-Mining Methods and Empirical Equations. *Journal of Soft Computing in Civil Engineering* 6(1):1-28.

Zhang C, Guisasola A, Baeza JA (2022) A review on the integration of mainstream P-recovery strategies with enhanced biological phosphorus removal. *Water Research* 212:118102.

Zhang J, Teng Y-F, Chen W (2019) Support vector regression with modified firefly algorithm for stock price forecasting. *Applied Intelligence* 49(5):1658-1674.

Zhang L, Ding W, Qiu J, Jin H, Ma H, Li Z, Cang D (2018) Modeling and optimization study on sulfamethoxazole degradation by electrochemically activated persulfate process. *Journal of Cleaner Production* 197:297-305.

Zhang LP, Yu HJ, Hu SX (2005) Optimal choice of parameters for particle swarm optimization. *Journal of Zhejiang University: Science* 6 A(6):528-534.