EASR

Engineering and Applied Science Research

https://www.tci-thaijo.org/index.php/easr/index

Published by the Faculty of Engineering, Khon Kaen University, Thailand

Artificial neural network for modelling the removal of pollutants: A review

Siti Fatimah*[1)] and Wiharto[2)]

[1)]Department of Chemical Engineering, Universitas Muhammadiyah Surakarta, Surakarta 57162, Indonesia
[2)]Department of Informatics, Universitas Sebelas Maret, Surakarta 57162, Indonesia

Received 21 September 2019
Revised 18 December 2019
Accepted 15 January 2020

Abstract

Modeling of pollutant degradation using artificial neural networks (ANN) has been done well. The techniques used to model degradation vary. This literature review was done to examine the development of the use of ANN modeling from year to year. It will provide an overview of predictive studies from a degradation treatment condition that will produce optimal conditions. These conditions will be supported by experimental data so that costs and time can be reduced at laboratory scale. Some relevant techniques include separation methods, coagulation, advanced oxidation processes, and chemical oxidation. The algorithmic approaches used are ANN-LM, ANN-BP, ANN-BP (SCG), and ANN-BFGS. Modelling using ANN has very high potential for further development. The perfomance indicator of an ANN method is a strong coefficient of determination (R^2), with good RMSE, MAPE, and MSE values.

Keywords: Review, Pollutants, Removal, Degradation, Artificial neural network

Abbreviations

ANN	: Artificial Neural Network	SSE	: Sum Square Error
MLP	: Multi-layer Perceptron	PCA-ANN	: Principal Component Analysis Artificial Neural Network
ANN-LM	: Artificial Neural Network-Levenberg-Marquardt	BFGS	: Broyden–Fletcher–Goldfarb–Shanno
ANN-BP	: Artificial Neural Network Backpropagation	SSE	: Sum Of Squared Errors
MGD	: Momentum Gradient Descent	PSO	: Particle Swam Optimization
GD	: Gradient Descent	ANFIS	: Adaptive Neuro-Fuzzy Inference System
SCG	: Scaled Conjugate Gradient	ANN-GA	: Artificial Neural Network-Genetic Algorithms
SOM	: Self Organizing Map	RBFNN	: Radial Basis Function Neural Network
MSE	: Mean Square Error	MAPE	: Mean Absolute Percentage Error

1. Introduction

Water is a basic human need for life. Human activities produce wastewater that contains of pollutants that can be harmful to life if not handled properly. Common pollutants include synthetic dyes, heavy metals, and persistent organic chemical pollutants [1]. Synthetic dyes are often used to improve the quality of some products. Based on research by Katheresan [2], the types of human activities that use dyes include textiles 54%, dying 21%, paper and pulp 10%, tannery and paint 8%, and dye manufacture 7%. This shows that the use of dyes is very great. In the textile industry, the biggest use of dyes is to impart colors to cloth (55%), bleaching (25%), wet finishing (18%), and scouring (2%).

Heavy metals and organic pollutants are species that are very dangerous to humans. Cr (VI), Cr (III), Pb (II), Fe (II), Ni (II), Cd (II), Mn (II), Co (II) and As (III) are examples heavy metals that can endanger humans. Other harmful pollutants include phosphates, naphthalene, aniline, toluene and triamterene. Heavy metals are toxic and these other pollutants can be carcinogenic. Physical, chemical and biological methods have been examined by numerous researchers to overcome the problems of dyestuff waste, heavy metals and other pollutants [3].

The treatment of the wastewater considers a variety of variables that differ based on the type waste and treatment methods used. To reduce the cost of the experimental process, modelling and optimization are done so that the waste processing can be performed optimally. Optimal processing will have a significant impact in terms of reduced costs and working times. The main key to modelling and optimization is to produce a removal with the lowest costs and operating time. Therefore, prediction of the maximum removal percentage is necessary. Modelling using highly

*Corresponding author.
Email address: sf120@ums.ac.id
doi: 10.14456/easr.2020.36

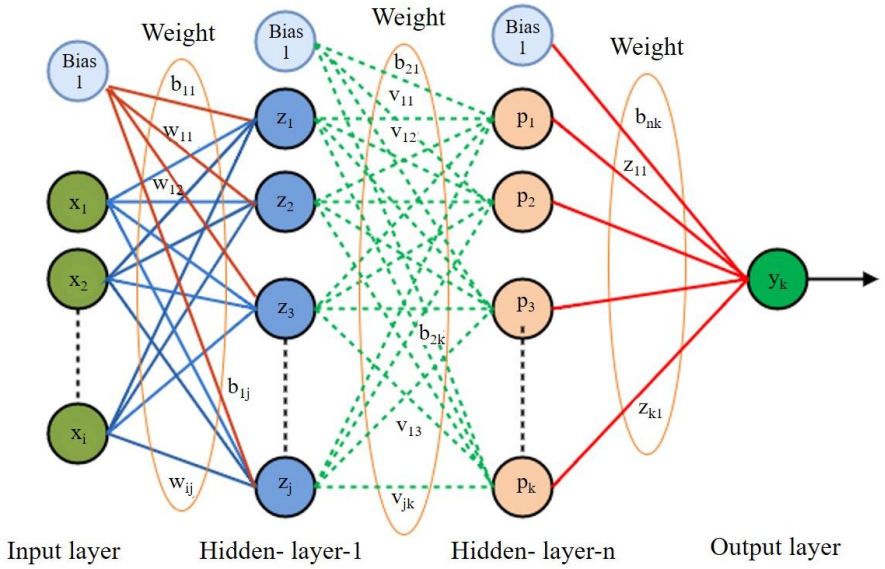

Figure 1 The architecture of a multi-layer neural network

flexible computational intelligence is employed to predict optimum removal. Artificial neural networks (ANNs) have good performance. One object of this article is to describe the fundaments of ANNs and their use in predicting removal of dyes, heavy metals and other pollutants from liquid wastes. This review will also discuss the use of multilayer perceptron neural networks, fuzzy neural networks, and radial basis functions in modelling the degradation of wastewater.

2. Artificial Neural Network

An artificial neural network is a model that can mimic human thought and brain function, making conclusions from pieces of information. It uses the same structure as the human brain. The ability of an ANN to learn makes them useful for solving problems involving classification, prediction and forecasting in various fields. Artificial neural networks can be grouped by type of architecture, training method and learning algorithm [4].

There are three types of ANN architecture, single-layer, multiple-layer and competitive layer. These three architectures are multi- and competitive-layered. ANNs can be grouped according to their training methods as supervised, unsupervised and hybrid. Multi-layer architectures use supervised training methods, while competitive-layer architecture employs unsupervised training. Most well-known supervised learning algorithms are backpropagation and learning vector quantization, while the unsupervised methods include self-organizing maps (SOM), and radial basis function (RBF) algorithms [4-5].

2.1 The architecture of multi-layer perceptron backpropagation

A perceptron multi-layer architecture can be trained using a number of algorithms, including backpropagation, Levenberg-Marquard, and quasi-Newton algorithms. Multi-layer architecture with a backpropagation learning algorithm is an ANN model with good ability in predictive applications. The learning method used in this model is supervised. The backpropagation algorithm uses an error output to change the value of the weights in the reverse direction. To determine the error, we need to do a forward propagation first. A multi-layer architecture is shown in Figure 1. The following is a multi-layer perceptron backpropagation algorithm with one hidden layer [4]:

Step 0:

Initialize the weights for each layer w_{ij} and v_{jk} (single hidden layer)

Step 1:

Repeat steps 2-8, if epoch <maximum and MSE> target error

Do for each training pair

Feedforward:

Step 2: Each input unit $X_i, i = 1,2,3, \dots . , n$ receives an input signal x_i and broadcasts this signal to all units in the layer above (the hidden units).

Step 3. In each hidden unit $Z_j, j = 1,2,3, \dots , p$, sum the weighted input signals. Then, this sum is multiplied by the weights, w_{ij} as:

$$z\_in_j = b1_j + \sum_{i=1}^{n} x_i \, w_{ij} \tag{1}$$

Apply an activation function to compute the output from the hidden layer as:

$$z_j = f(z\_in_j) \tag{2}$$

and send this signal to all units in the layer above (output units). Step 3 is done for each of the hidden layers used.

Step 4: At each output unit $Y_k, k = 1,2,3, \dots , m$, sum the weighted input signals, and multiply this result by the weight, V_{jk}.

$$y\_in_k = b2_k + \sum_{j=1}^{p} z_j \, v_{jk} \tag{3}$$

Apply this activation function to compute an output signal.

$$y_k = f(y\_in_k) \tag{4}$$

$$f(y_{\_in_k}) = \frac{1}{1 + e^{-y\_in_k}} \tag{5}$$

Backpropagation:

Step 5. Each ouput unit $Y_k, k = 1,2,3, \dots, m$, receives a target pattern corresponding to the input training pattern and computes its error information term $t_k, k = 1,2,3, \dots, m$, which given by:

$$\delta 2_k = (t_k - y_k) f'(y_{in_k}) \quad (6)$$

$$\varphi 2_{jk} = \delta_k z_j \quad (7)$$

$$\beta 2_k = \delta_k \quad (8)$$

Then, the weight correction term is determined that will be used later to update v_{jk}.

$$\Delta v_{jk} = \alpha \varphi 2_{jk} \quad (9)$$

A bias correction term is calculated (and used to update $b2_k$ later).

$$\Delta b2_k = \alpha \beta 2_k \quad (10)$$

Step 5 is done for all hidden layers.

Step 6: Each hidden unit, $Z_j, j = 1,2,3, \dots, p$, sums its delta inputs from units of the layer above:

$$\delta\_in_j = \sum_{k=1}^{m} \delta 2_k v_{jk} \quad (11)$$

Then, this term is multiplied by the derivative of its activation function to calculate an error information term:

$$\delta 1_j = \delta\_in_j f'(z_{in_j}) \quad (12)$$

$$\varphi 1_{ij} = \delta 1_j x_j \quad (13)$$

$$\beta 1_j = \delta 1_j \quad (14)$$

A weight correction term is calculated and later used to update w_{ij}.

$$\Delta w_{ij} = \alpha \varphi 1_{ij} \quad (15)$$

A bias correction term is calculated and later used to update $b1_{j.}$

$$\Delta b1_j = \alpha \beta 1_j \quad (16)$$

Step 7: Each output unit $Y_k, k = 1,2,3, \dots, m, j = 1,2,3, \dots, p$ updates its bias and weight.

$$v_{jk}(new) = v_{jk}(old) + \Delta v_{jk} \quad (17)$$

$$b2_k(new) = b2_k(old) + \Delta b2_k \quad (18)$$

Each hidden unit, $Z_j, j = 1,2,3, \dots, p$, i=1,2,3,…,n updates its bias and weight.

$$w_{ij}(new) = w_{ij}(old) + \Delta w_{ij} \quad (19)$$

$$b1_j(new) = b1_j(old) + \Delta b1_j \quad (20)$$

Step 8: Calculate the MSE or SSE (perfomance indicator), where t_k is an output target.

$$MSE = \frac{\sum_{k=1}^{n} (t_k - y_k)^2}{n} \quad (21)$$

$$SSE = \sum_{k=1}^{n} (t_k - y_k)^2 \quad (22)$$

Back Step 1

The backpropagation algorithm is performed after the training stage. The testing process is different from the training process. In a MLP, the testing algorithm is only feedforward, using weight and bias results from training. The backpropagation algorithm can use a number of improved algorithms, such as the momentum gradient descent (MGD) [6], scaled conjugate gradient (SCG) [7], and Bayesian regularization (BR) [8]. ANN performance for predictions is measured using the coefficient of determination, R^2 expressed in Equation (23) [9], RMSE in Equation (24), and MAPE in Equation (25).

$$R = \frac{\sum_{k=1}^{n} (t_k - \mu_T)(y_k - \mu_o)}{\sqrt{\sum_{k=1}^{n} ((t_k - \mu_T)^2} \sqrt{(y_k - \mu_o)^2}} \quad (23)$$

$$RMSE = \sqrt{\frac{\sum_{k=1}^{n} (t_k - y_k)^2}{n}} \quad (24)$$

$$MAPE = \frac{\sum_{k=1}^{n} \left| \frac{\sum_{k=1}^{n} t_k - y_k}{y_k} \right|}{n} \quad (25)$$

where μ_T is the average of the target, while μ_o is the average of the predicted output.

2.2 The architecture of a multi-layer quasi-Newton (BFGS) algorithm

Quasi-Newton multi-layer architecture is a feedforward ANN with a learning algorithm that uses a numerical optimization technique, namely the quasi-Newton approach [10]. The quasi-Newton algorithm belongs to the conjugate gradient category, which more rapidly converges to an optimum value. The basic concept of weight change in ANN is:

$$W_{t+1} = W_t - H_t^{-1}.g_t \quad (26)$$

where:
W is weight and bias vector
g_t is the gradient vector
H is a Hessian matrix

The quasi-Newton method provides for rapid training, but it requires considerable resources in each epoch. The Hessian matrix contains second derivatives that have been replaced by a function for the gradient. This training algorithm is known as the BFGS algorithm.

2.3 The architecture of multi-layer Levenberg-Marquardt

The Levenberg-Marquardt (LM) multi-layer is a feedforward algorithm that is also based on a Hessian matrix to determine the weights in the ANN. The Hessian matrix is a derivative of a performance function for each weight and bias component [11]. The performance function is derived using Equation (22). The change process employs a gradient

function. If the performance function uses the sum of squared errors (SSE), then, the Hessian matrix can be estimated using Equation (27):

$$H = J^T J + \eta I \quad (27)$$

where:
η: the Marquardt parameter
I: Identity matrix
J: a Jacobian matrix that consists of the first derivative of network error on each component of weight and bias.

2.4 Multi-layer radial basis function architecture (RBF)

The RBF multi-layer architecture is included in feedforward ANN, but it uses hybrid training methods, namely a combination of supervised and unsupervised learning algorithms [12]. In the first layer, the ANN employs unsupervised trained, usually using clustering algorithms, while the hidden layer goes to the output layer in training using a supervised method, such as a perceptron.

2.5 Optimization of ANN

ANN sometimes provides good performance when tested using training data, but poor performance using testing data. These problems can be overcome by combining ANN with a number of algorithms including particle swarm optimization (PSO) and genetic algorithms (GA). PSO is an algorithm that models the behavior of animals in a swarm when they are influenced by the behavior of individuals and their groups. A population in the form of particles is treated as a point in a given dimension of space. Then, there are two factors that update the status of particles in the search space, their position and velocity [13]. The best position and the best group for each iteration are stored as a solution.

Genetic algorithms are search algorithms that are based on the mechanism of natural selection and genetics. Genetic algorithms begin with the process of population generation and evaluation of fitness. Then, population selection is done to get parent candidates. In the next stage, crossover, mutation, and reevaluation of individuals in the population occurs. This reshapes the new population that survives. The process is repeated up to the maximum number of generations desired [14]. ANN optimization with PSO or GA can be done in two ways. First, PSO or GA is used to determine an initial weight for an ANN. After that, training is done with algorithms such as LM. Second, ANN is trained using algorithms such as Levenberg-Marquardt (LM), and then optimized using PSO or GA.

3. Literature review

Review of studies of applications from artificial neural networks was objectively done. The methods used for modelling pollutant degradation processes, as well as for mitigation of azo dyes and heavy metal vary greatly. The uses of artificial neural networks in this review are shown in Table 1.

4. Results and analysis

Pollutants, dyes, and heavy metals are very dangerous species if they are released into the environment without appropriate treatment. Azo dyes are one type of synthetic chemical that are widely used in industry. These dyes have a complex structure because of the presence of azo (-N=N-) bonds that are toxic and slowly decompose in the environment. The azo bonds found in dyes actually have two azo bonds called diazo bonds. Some have three azo bonds and are called poliazo. Most azo compounds have an aromatic structure and with sulphate substituent groups or nitrate salts. These substituent groups cause azo compounds to have strong intermolecular strength, a physical appearance that is sharp and bright in color, and difficult to degrade in a simple way. The molecular bond strength of azo compounds with fabric or paper substituents is very great because the presence of Van der Waals or hydrophobic bonds makes these compounds are very durable. Industries that use dyes include the textiles, paper, paint, and cosmetics industries.

Human activities can produce pollution that is carried away in waste. The processes of coloring, dyeing, or painting will create waste streams carrying these species and discharge them into the aquatic environment. Techniques are needed to degrade the dyestuff structures so that when they enter the environment, they becomes species that do not endanger the aquatic biota or humans. Methods continue to be developed to maximally degrade pollutants and azo dyes, as well as sequester heavy metals. These methods are depicted in Figure 2. Based on the results of a review of the reports in Table 1, it can be concluded that techniques for promoting degradation of organic pollutants and azo dyes, as well as removing heavy metals, vary widely. The degradation processes can be modelled using artificial neural networks (ANN).

ANN applications predict optimal conditions for pollutant degradation and/or mitigation with relatively good performance. The use of ANN in predicting optimal conditions uses many training algorithms, but numerous studies used ANN with backpropagation training algorithms. The backpropagation training algorithm has good prediction capability. The weakness of the backpropagation algorithm is in terms of its training speed. A number of subsequent studies used ANN with the Levenberg-Marquardt (ANN-LM) training algorithm. LM algebra has training capabilities that are faster than other algorithms. Additionally, LM is quick to converge and also provides relatively better performance compared to a number of other algorithms. Model development, can also employ Fuzzy approaches such as adaptive fuzzy inference systems (ANFIS) as was done by Porhemmat et al. [15]. Fuzzy based degradation models tend to have poor performance in terms of optimization, so this type of research is relatively sparse [16].

Development of prediction system models to obtain optimal conditions was done from 2006-2018. Research using this algorithm ranks second among the use of training algorithms in ANN, as shown in Figure 3. Development was done using ANN and its optimization. The distribution of research carried out in modelling from 2006-2018 is depicted in Figure 4. In this figure, it can be seen that the model predictors of optimum degradation using ANN and optimization were most numerous in 2017. In 2018, the trend relaxed slightly. From 2006-2018 modeling research was very active. Development was done emphasizing several areas, including various ANN methods, objects modeled, variables of the objects used and degradation methods modeled. This, of course, will greatly affect the parameters predicting degradation.

Table 1 Summary of investigation on modeling of pollutant removal

Author (s)	Dyes	Method of Degradation	Parameters of Method	Method of ANN modelling	Optimum Number of Neurons	Performance Degradation
Dyes						
Elemen [17], 2012	Reactive Red 141	Organoclay using hexadecyltrimethylammonium bromide	Contact time, initial dye concentration, adsorbent dosage, temperature	ANN-BP (SCG)	4-5-1	R^2 = 0.978 MSE = 0.027364
Daneshvar [18], 2006	Basic Yellow 28	Electrocoagulation	Current density, initial pH, time electrolysis, initial dye concentration, the distance between the electrodes, retention time, solution conductivity.	ANN-BP (SCG)	7-10-1	R^2 = 0.93
Khataee [19], 2009	Basic Red 46	Photocatalytic using TiO_2 nanoparticle	Initial dye concentration, UV light intensity, time and initial pH	ANN-BP	6-10-1	R^2 = 0.95
Aleboyeh [20], 2008	Acid orange 7	Photochemical which consist of combination UV and H_2O_2.	Initial dye concentration, the concentration of H_2O_2, pH, time of UV irradiation	ANN-BP (SCG)	4-8-1	R^2=0.996
Ghaedi [21], 2017	Methylene blue (MB) and brilliant green (BG)	Graphite oxide nanoparticle	pH, initial dye concentration, contact time and adsorbent dosage	ANN-LM ANN-BP (GD)	3-8-1	MB: R^2= 0.9816, RMSE=0.0012 BG: R^2 = 0.981 RMSE = 0.001
Dil [22], 2016	Crystal Violet (CV)	Zinc oxide nanorods	Initial dye concentration, pH, adsorbent dosage, sonication time	ANN-BP	4-4-1	MSE=0.0011 R^2=0.9815
Ali [23], 2018	Amido black	Iron composite nanomaterial	Dye amount, agitation time, pH, nanomaterial dosage, temperature	ANN-BP	4-11-1	R^2= 0.94
Fatimah [24], 2017	Acid orange 7	Ozonation.	Initial dye concentration, temperature, pH, ozone contact	ANN-BFGS	4-30-25-1	MAPE=0.00776
Dutta [25], 2010	Reactive black	TiO_2	pH, amount of TiO_2, initial dye concentration, time	ANN-LM	3-11-1	R^2= 0.993
Salari [26], 2009	Basic yellow	Peroxide coagulation	Electrolysis time, initial pH, applied current, initial dye concentration	ANN-BP	4-16-1	R^2= 0.97
Porhemmat [15], 2017	Methylene blue and sunset yellow	Nanocomposites	The concentration of dye, nanocomposite dosage	ANFIS	4-10-1	R^2= 0.999 for MB R^2= 0.999 for SY
Khataee [27], 2012	Acid blue 92	Phytoremediation	Initial dye, pH, temperature, amount of plant and reusability of L. minor (day)	ANN-BP (SCG)	5-14-1	R^2= 0.954
Zarei [28], 2010	Basic blue 3, Malachite green, basic red 46, basic yellow 2	Carbon nanotube-PTFE	Electrolysis time, initial pH, applied current, initial dye concentration	ANN-BP (SCG)	4-14-1	R^2= 0.989
Ardekani [29], 2017	Metylne blue	zinc hydroxide nanoparticles	pH, adsorbent dosage, sonication time	ANN-LM	4-11-1	R^2= 0.98 MSE = 0.0529
Aber [30], 2007	Acid orange 7	Powder activated carbon.	Initial dye concentration, initial pH, temperature	ANN-BP (SCG)	3-2-1	R^2= 0.998
Kasiri [31], 2008	Acid Red 14	Photo-Fenton	Dye concentration, catalyst concentration, pH, the concentration of H_2O_2	ANN-BP	5-10-1	R^2= 0.976
Ghaedi [32], 2015	Malachite Green	Adsorption onto copper nanowires	Solution pH, adsorbent dosage, contact time, and initial MG concentration	ANN-LM	5-11-1	R^2= 0.989 MSE = 0.00023
Ghaedi [33], 2014	Phenol red	Gold and titanium dioxide nanoparticles	pH, dye concentration, adsorbent dosage, contact time	ANN-LM	4-15-19-1	R^2= 0.9962 MSE = 0.00003
Hajati [34], 2015	Acid yellow 41 and sunset yellow	SnO_2 nanoparticle	Contact time, adsorbent dosage, pH, AY41 concentration and SY concentration	PCA-ANN-LM	4-20-1	R^2= 0.95 (AY41) R^2= 0.98 (SY)
Malekbala [35], 2015	Acid red 57	mesoporous carbon-coated monoliths	Initial pH, initial dye concentration, and contact time	ANN-BP	4-8-11	R^2= 0.99 MSE = 0.003
Soleymani [36], 2011	Direct Red 16 (DR16)	UV/$K_2S_2O_8$ process	Contact time, pH, potassium peroxydisulfate dosage, temperature, dye concentrations	ANN-BFGS	5-9-1	R^2 = 0.9968 MSE = 0.000256
Fatimah [37], 2017	Acid Black 1 (AB1) and Acid Yellow 19 (AY19)	Ozonation	Dye concentration, ozone concentration, pH, and temperature	ANN-LM	4-20-15-1	AB1 : MAPE=0.00448 MSE=0.00756 AY19 : MAPE=0.00441 MSE=0.007

Table 1 (continued) Summary of investigation on modeling of pollutant removal

Author (s)	Dyes	Method of Degradation	Parameters of Method	Method of ANN modelling	Optimum Number of Neurons	Performance Degradation
Organic pollutants						
Hasani [38], 2018	Humic acid	Electrocoagulation-flotation.	Initial humic acid concentration, pH, electrical conductivity, time pulse, number of pulse, voltage.	NN-BP (GDM) And ANN-LM	5-9-1 5-6-1	R^2=0.971, MSE=0.0031 R^2=0.999, MSE=0.00006
Nidheesh [3], 2018	Organic pollutants	Electrochemical Oxidation Process (EAOPs)	Dosage of oxidators, current density, initial pH, time electrolysis, initial dye concentration, distance between the electrodes, retention time, solution conductivity	ANN	4-8-1	R^2 = 0.998
Dlamini [39], 2014	Trichlorophenol	Chromolaena odorata stem	pH, initial dye concentration, adsorbant dosage, time contact,	ANN-LM	5-10-1	R^2= 0.98 MSE = 0.30
Deshmukh [40], 2012	Toluene	Biolfilter	time period by daily measurement of inlet VOC concentration, retention time, pH, temperature and packing moisture content	RBFNN	5-27-1	R^2= 0.9755 MSE = 0.03
Xie [41], 2018	Textile dye sludge	Combination of O_2/N_2 and O_2/CO_2	Temperature, mass concentration rate, residual mass, comprehensive stability index, comprehensive combustion index	ANN-BP (BR)	4-16-1	R^2= 0.9989 RMSE = 0.8506
Metal pollutants						
Hoseinian [42], 2017	Ni(II)	Ion flotation	pH, collector concentration, frother concentration, impeller speed and flotation time, the removal percentage of Ni(II) ions and water during ion flotation	ANN-BP + GA	5-27-1 for Ni(II) 5-12-1 for water	R^2= 0.943 RMSE = 0.0771
Kiran [43], 2017	Cadmium	Biosorption with *Spirulina* sp.	Initial concentration, biosorbant dosage, agitation speed, pH.	ANN	4-9-1	R^2= 0.99
Mohan [44], 2015	Cr(VI)	CuO nanoparticles	Different initial Cr(VI) concentrations, pH, temperature and CuONPs dosage.	ANN-BP-GA	4-10-1	R^2= 0.9947 MSE = 0.021

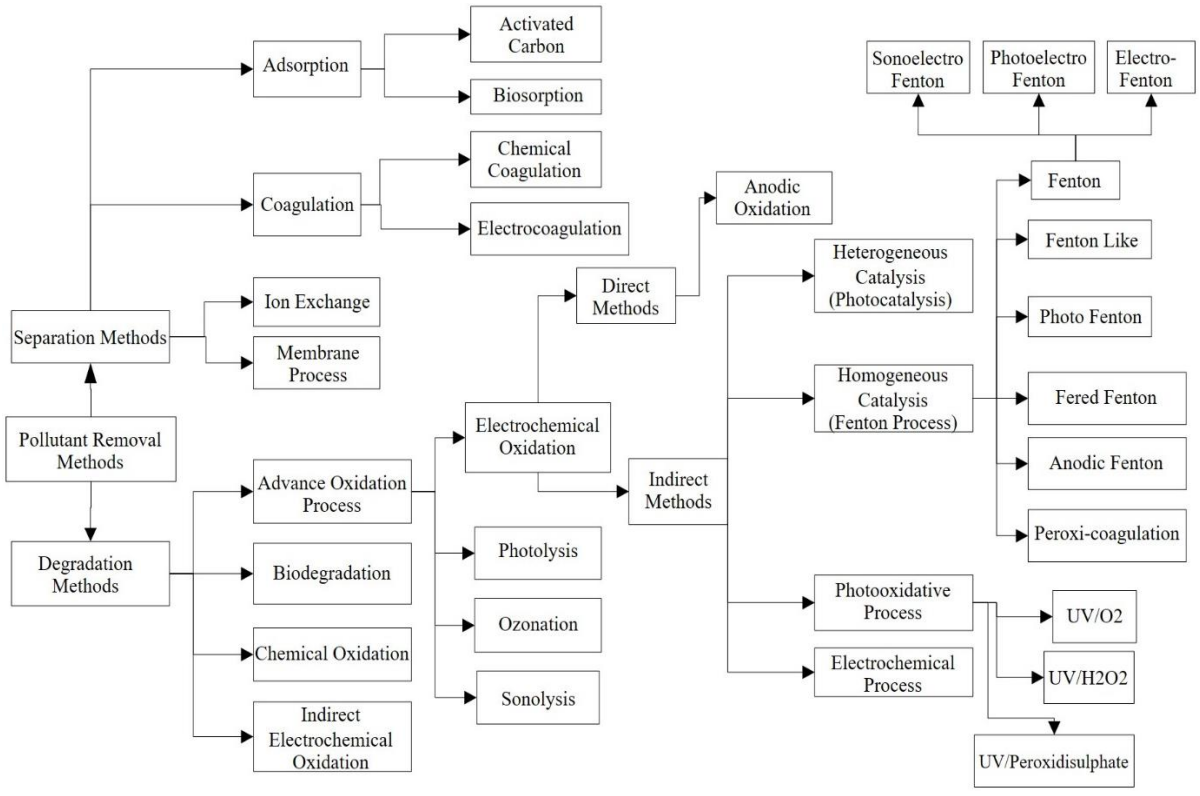

Figure 2 Methods of removing organic pollutants, azo dyes, and heavy metals

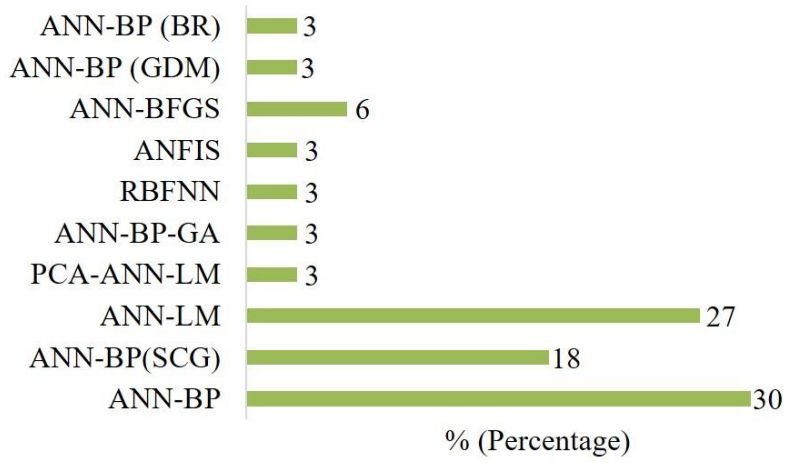

Figure 3 The distribution of ANN training algorithms

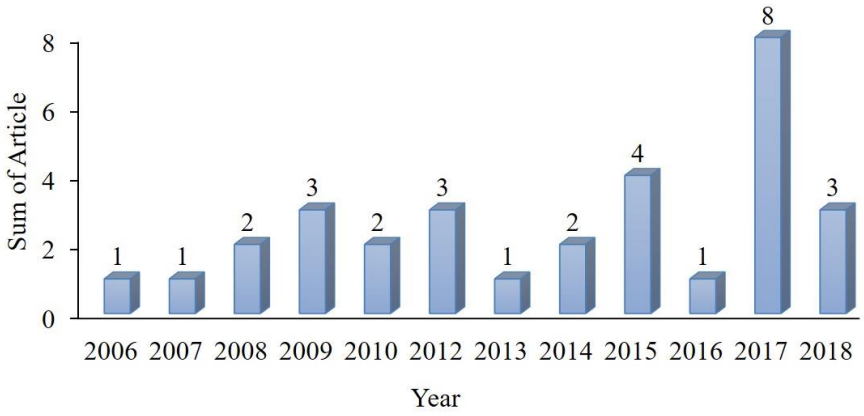

Figure 4 The distribution of ANN and optimization articles from 2006-2018

The ANN methods for modeling pollutant removal will produce variable performance depending on the training algorithm used. Modeling dye degradation using ANN usually produces very high R^2 values, close to 1 for all types of training algorithms used. There are a number of studies that require a complex ANN architecture using two hidden layers, such as Ghaedi [33] and Fatimah [24]. The complexity of the ANN architecture, in addition to the number of hidden layers, is also influenced by the number of neurons in the hidden layer. The influence of the numbers of hidden layers and neurons will affect the time required for training. In the structure with one hidden layer, the fewest neurons with good performance is two [30]. In the case of other types of dye, for example Acid Orange 7, a different ANN architecture can be employed depending on the degradation method used. When using an ozonation degradation method modeled with the ANN-BFGS algorithm, a more complex architecture is required than for ANN-BP (SCG), which is indicated by the number of hidden layers. This condition was demonstrated in research conducted by Fatimah [24], Aber [30] and Aleboyeh [20]. A similar case also occurs for an ozonation method to treat wastewater for AB1 and AY19 dyes. This also required a two hidden layer architecture using ANN-LM. The structure of Acid Orange 7 presents a sodium sulfonate salt group. In the presence of activated carbon, the sulfonate salt group will be easily adsorbed into activated carbon pores. This adsorption process causes more effective interaction between the molecules of Acid Orange 7 and activated carbon. Then Acid Orange 7 splits into intermediate compounds.

Modeling organic pollutants with ANN-BP (GDM) and ANN-LM is simpler compared with RBF-NN and ANN-BP (BR). There are fewer neurons in the hidden layer. Modeling a heavy metal pollutant with ANN without optimization provides performance that is comparable to ANN optimized with GA. The use of GA gives insignificant performance improvement with a requirement of additional computational effort. Another consideration is that GA has a number of random processes, such as selection, crossover and gene mutation. These processes may produce different results when the ANN-GA model is retested. The use of GA in ANN also requires a relatively greater number of neurons in the hidden layer compared to other ANN training algorithms. Modeling metal pollutant mitigation requires more hidden layers. In a study conducted by Hoseinian [42] on Ni (II) removal, there were 27 hidden layers that produced an R^2 value of 0.943. The research of Kiran [43] and Mohan [44] modeled Cd (VI) and Cr (VI) removal. This study produced an R^2 value of 0.99, which is an improvement over Hoseinian's modeling. Removal techniques that involve ion flotation will have more steric considerations due to molecular competition.

From this analysis, using ANN in modeling pollutant removal will provide suggestions for optimizing laboratory scale processes. ANN modeling provides specific descriptions of the parameters of each method used. These parameters can be optimized through estimation by ANN for greater efficiency. The disadvantages of using ANN in modeling pollutant removal include the need for relatively large amounts of training and testing data to produce accurate

results. Another drawback is that the method may identify local rather than global optima, so more testing is needed.

There are several things that must be considered in ANN modelling to predict the degradation of pollutants and azo dyes, as well as heavy metal mitigation. First, the experiments must be designed to capture data measuring the parameters that will be affected. This stage determines the independent and dependent variables. Second is inputing the data and output of each parameter. The third is normalization of input and output data to increase the performance of network training. Fifth is random separation of the generated data into training and testing data sets. A training data set is employed to enable the ANN to learn the input model. The testing of the data sets is used externally with the parameters of the trained ANN. Sixth is construction and training of the ANN model and then optimizing it to produce the most accurate results. Last is selection of the best ANN model based on performance criteria and extraction of results from the trained model.

5. Conclusions

Review of the research on modeling of pollutant degradation processes has been done by numerous researchers. Various modeling studies using ANN show that this approach is growing in popularity and application. The algorithms used also vary. ANN-LM, ANN-BP, ANN-BP (SCG), and ANN-BFGS have been employed. Degradation techniques also vary. ANN modelling can provide good performance and high accuracy. Therefore, further modelling studies need to be done combining ANN with other algorithms for optimization and feature selection.

6. Acknowledgement

This research is the result of collaboration between the Department of Chemical Engineering, Universitas Muhammadiyah Surakarta, Indonesia and the Department of Informatics of Universitas Sebelas Maret, Surakarta, Indonesia.

7. References

[1] Fan M, Hu J, Cao R, Ruan W, Wei X. A review on experimental design for pollutants removal in water treatment with the aid of artificial intelligence. Chemosphere. 2018;200:330-43.

[2] Katheresan V, Kansedo J, Lau SY. Efficiency of various recent wastewater dye removal methods: a review. J Environ Chem Eng. 2018;6(4):4676-97.

[3] Nidheesh PV, Zhou M, Oturan MA. An overview on the removal of synthetic dyes from water by electrochemical advanced oxidation processes. Chemosphere. 2018;197:210-27.

[4] Fausett L. Fundamentals of neural networks architecture, algorithms, and applications. New Jersey: Prentice Hall; 1994.

[5] Fu L. Neural networks in computer intelligence. Singapore: McGraw Hill; 1994.

[6] Phansalkar VV, Sastry PS. Analysis of the back-propagation algorithm with momentum. IEEE Trans Neural Network. 1994;5(3):505-6.

[7] Møller MF. A scaled conjugate gradient algorithm for fast supervised learning. Neural Network. 1993;6(4):525-33.

[8] Kayri M. Predictive abilities of bayesian regularization and Levenberg–Marquardt algorithms in artificial neural networks: a comparative empirical study on social data. Math Comput Appl. 2016;21(20):1-11.

[9] Kardan AA, Sadeghi H, Ghidary SS, Sani MRF. Prediction of student course selection in online higher education institutes using neural network. Comput Educ. 2013;65:1-11.

[10] Mukherjee I, Routroy S. Comparing the performance of neural networks developed by using Levenberg–Marquardt and Quasi-Newton with the gradient descent algorithm for modelling a multiple response grinding process. Expert Syst Appl. 2012;39(3):2397-407.

[11] Ling C, Wang G, He H. A new Levenberg–Marquardt type algorithm for solving nonsmooth constrained equations. Appl Math Comput. 2014;229:107-22.

[12] de Leon-Delgado H, Praga-Alejo RJ, Gonzalez-Gonzalez DS, Cantú-Sifuentes M. Multivariate statistical inference in a radial basis function neural network. Expert Syst Appl. 2018;93:313-21.

[13] Xu G, Yu G. On convergence analysis of particle swarm optimization algorithm. J Comput Appl Math. 2018;333:65-73.

[14] Israeli E, Gilad E. Novel genetic algorithm for loading pattern optimization based on core physics heuristics. Ann Nucl Energ. 2018;118:35-48.

[15] Porhemmat S, Ghaedi M, Rezvani AR, Azqhandi MHA, Bazrafshan AA. Nanocomposites: synthesis, characterization and its application to removal azo dyes using ultrasonic assisted method: modeling and optimization. Ultrason Sonochem. 2017;38:530-43.

[16] Lin CT, Lee G. Neural fuzzy systems. London: Prentice Hall; 1996.

[17] Elemen S, Akçakoca Kumbasar EP, Yapar S. Modeling the adsorption of textile dye on organoclay using an artificial neural network. Dyes Pigments, 2012;95(1):102-11.

[18] Daneshvar N, Khataee AR, Djafarzadeh N. The use of artificial neural networks (ANN) for modeling of decolorization of textile dye solution containing CI basic yellow 28 by electrocoagulation process. J Hazard Mater. 2006;137(3):1788-95.

[19] Khataee AR. Photocatalytic removal of C.I. Basic Red 46 on immobilized TiO2 nanoparticles: artificial neural network modelling. Environ Tech. 2009;30(11):1155-68.

[20] Aleboyeh A, Kasiri MB, Olya ME, Aleboyeh H. Prediction of azo dye decolorization by UV/H2O2 using artificial neural networks. Dyes Pigments. 2008;77(2):288-94.

[21] Ghaedi AM, Vafaei A. Applications of artificial neural networks for adsorption removal of dyes from aqueous solution: a review. Adv Colloid Interface Sci. 2017;245:20-39.

[22] Dil EA, Ghaedi M, Ghaedi A, Asfaram A, Jamshidi M, Purkait MK. Application of artificial neural network and response surface methodology for the removal of crystal violet by zinc oxide nanorods loaded on activate carbon: kinetics and equilibrium study. J Taiwan Inst Chem Eng. 2016;59:210-20.

[23] Ali I, Alharbi OML, Alothman ZA, Badjah AY, Alwarthan A, Basheer AA. Artificial neural network modelling of amido black dye sorption on iron composite nano material: kinetics and thermodynamics studies. J Mol Liq. 2018;250:1-8.

[24] Fatimah S, Wiharto W. The use of artificial neural network for modeling the decolourization of acid

orange 7 solution of industrial by ozonation process. IOP Conf Ser Mater Sci Eng. 2017;172:1-8.
[25] Dutta S, Parsons SA, Bhattacharjee C, Bandhyopadhyay S, Datta S. Development of an artificial neural network model for adsorption and photocatalysis of reactive dye on TiO2 surface. Expert Syst Appl. 2010;37(12):8634-8.
[26] Salari D, Niaei A, Khataee A, Zarei M. Electrochemical treatment of dye solution containing C.I. Basic Yellow 2 by the peroxi-coagulation method and modeling of experimental results by artificial neural networks. J Electroanal Chem. 2009;629(1-2):117-25.
[27] Khataee AR, Movafeghi A, Torbati S, Salehi Lisar S Y, Zarei M. Phytoremediation potential of duckweed (*Lemna minor* L.) in degradation of C.I. acid blue 92: artificial neural network modeling. Ecotoxicol Environ Saf. 2012;80:291-8.
[28] Zarei M, Niaei A, Salari D, Khataee AR. Removal of four dyes from aqueous medium by the peroxi-coagulation method using carbon nanotube–PTFE cathode and neural network modeling. J Electroanal Chem. 2010;639(1-2):167-74.
[29] Ardekani PS, Karimi H, Ghaedi M, Asfaram A, Purkait MK. Ultrasonic assisted removal of methylene blue on ultrasonically synthesized zinc hydroxide nanoparticles on activated carbon prepared from wood of cherry tree: experimental design methodology and artificial neural network. J Mol Liq. 2017;229:114-24.
[30] Aber S, Daneshvar N, Soroureddin SM, Chabok A, Asadpour-Zeynali K. Study of acid orange 7 removal from aqueous solutions by powdered activated carbon and modeling of experimental results by artificial neural network. Desalination. 2007;211(1-3):87-95.
[31] Kasiri MB, Aleboyeh H, Aleboyeh A. Modeling and optimization of heterogeneous photo-fenton process with response surface methodology and artificial neural networks. Environ Sci Technol. 2008;42(21):7970-5.
[32] Ghaedi M, Shojaeipour E, Ghaedi AM, Sahraei R. Isotherm and kinetics study of malachite green adsorption onto copper nanowires loaded on activated carbon: Artificial neural network modeling and genetic algorithm optimization. Spectrochim Acta Mol Biomol Spectros. 2015;142:135-49.
[33] Ghaedi M, Zeinali N, Ghaedi AM, Teimuori M, Tashkhourian J. Artificial neural network-genetic algorithm based optimization for the adsorption of methylene blue and brilliant green from aqueous solution by graphite oxide nanoparticle. Spectrochim Acta Mol Biomol Spectros. 2014;125:264-77.
[34] Hajati S, Ghaedi M, Mahmoudi Z, Sahraei R. SnO2 nanoparticle-loaded activated carbon for simultaneous removal of acid yellow 41 and sunset yellow; derivative spectrophotometric, artificial neural network and optimization approach. Spectrochim Acta Mol Biomol Spectros. 2015;150:1002-12.
[35] Malekbala MR, Hosseini S, Masoudi Soltani S, Malekbala R, Choong TSY, Eghbali Babadi F. Development, application, and evaluation of artificial neural network in investigating the removal efficiency of acid red 57 by synthesized mesoporous carbon-coated monoliths. Desalination Water Treat. 2015;56(8):2246-57.
[36] Soleymani AR, Saien J, Bayat H. Artificial neural networks developed for prediction of dye decolorization efficiency with UV/K2S2O8 process. Chem Eng J. 2011;170(1):29-35.
[37] Fatimah S, Wiharto W. Modeling of decolorisation dyes by ozonation techniques using Levenberg-Marquardt neural network. Mal J Fund Appl Sci. 2017;13(4):737-41.
[38] Hasani G, Daraei H, Shahmoradi B, Gharibi F, Maleki A, Yetilmezsoy K, et al. A novel ANN approach for modeling of alternating pulse current electrocoagulation-flotation (APC-ECF) process: Humic acid removal from aqueous media. Process Saf Environ Protect. 2018;117:111-24.
[39] Dlamini D, Mishra A, Mamba B. Artificial neural network simulations and experimental results: removal of trichlorophenol from water using *Chromolaena odorata* stem. Water SA. 2014;40(2):369-78.
[40] Deshmukh SC, Senthilnath J, Dixit RM, Malik SN, Pandey RA, Vaidya AN, et al. Comparison of radial basis function neural network and response surface methodology for predicting performance of biofilter treating toluene. J Software Eng Appl. 2012;5(8):595-603.
[41] Xie C, Liu J, Zhang X, Xie W, Sun J, Chang K, et al. Co-combustion thermal conversion characteristics of textile dyeing sludge and pomelo peel using TGA and artificial neural networks. Appl Energ. 2018;212: 786-95.
[42] Hoseinian FS, Rezai B, Kowsari E. The nickel ion removal prediction model from aqueous solutions using a hybrid neural genetic algorithm. J Environ Manag. 2017;204:311-7.
[43] Siva Kiran RR, Madhu GM, Satyanarayana SV, Kalpana P, Subba Rangaiah G. Applications of Box–Behnken experimental design coupled with artificial neural networks for biosorption of low concentrations of cadmium using *Spirulina (Arthrospira) spp*. Resour Effic Technol. 2017;3(1):113-23.
[44] Mohan S, Singh Y, Verma DK, Hasan SH. Synthesis of CuO nanoparticles through green route using Citrus limon juice and its application as nanosorbent for Cr(VI) remediation: process optimization with RSM and ANN-GA based model. Process Saf Environ Protect. 2015;96:156-66.