Journal of Engineering Technology Vol. 6: 14-21, 2018
ISSN 2231-8798

Data Based Modeling of a Wastewater Treatment Plant by using Machine Learning Methods

Nor Azuana Ramli[1] & Mohd Fairuz Abdul Hamid[1]
[1]Section of Electrical Engineering
Universiti Kuala Lumpur British Malaysian Institute

Corresponding email: norazuana@bmi.unikl.edu.my

Abstract: It is observed through studies and recent data, energy consumption increases proportionally to the electricity costs. In order to reduce the costs, energy consumption needs to be analyzed and one way to reduce the energy consumption is by implementing the energy efficiency. In this study, the focus is on wastewater treatment plant where energy efficiency is implemented to reduce the costs so that organized industrial water supply and sewerage can be brought to a larger population in the developing countries. If clean water and wastewater treatment can be offered more affordably in less developed regions, the quality of living be generally improved and thus, can save lives. There are few ways to implement energy efficiency such as optimizing the machine and equipment that used lots of energy or predict the energy consumption by using data based modeling which will be done in this study. The main objective of this study is to obtain a suitable method that can predict energy consumption with least error and provides maximum amount of cost saving. By predicting the energy output, the energy consumption for the next month can be determined and in that way, the energy consumption can be reduced. This paper presents data based modeling by using machine learning methods to predict the energy consumption. From the analysis, it is found that the prediction by using artificial neural network is better than other methods as it produces least error with root mean square error is 52084. By using the energy prediction results that were obtained through neural network, we can save as much as RM156499 which is 2.23 percent from current usage.

Keywords: wastewater treatment; energy efficiency; machine learning; energy prediction

1.0 INTRODUCTION

Wastewater treatment plant consumes a large proportion of total energy consumption. In other words, it devours on the already tight supplies of fossil fuel and coal while making it worse by releasing greenhouse gasses. This waste-to-energy conversion reduces greenhouse gas emissions in two ways. Electricity is generated which made it almost able to run off the grid. The greenhouse gas emissions are significantly reduced by preventing methane emissions from landfills. The urge of ending the life of this non-sustainable traditional plant, leads to a total game changer from solely waste disposal to resource recovery (water, nutrients and energy).

Hence, energy efficiency need to be implemented in order to optimize the energy consumption and to achieve a cost saving system. The energy efficiency is defined as the ratio of electricity generated to the electricity needed to operate the wastewater treatment plant. There are lots of successful examples showing the enormous potential of implementation energy efficiency in wastewater treatment plant.[1] One of the examples is the Strass wastewater treatment plant in Austria, which has reached 108% of energy recovery through implementation of energy efficiency.[2]

Generally, in order to achieve higher percentage of energy efficiency, energy management and optimization works across entire wastewater treatment system. In terms of energy management and optimization, there are some equipment and processes that its energy consumption can be optimized. For instance, blowers and aerators in aeration tanks are particularly used lots of energy but at the same time afford opportunities for major savings.

Journal of Engineering Technology Vol. 6: 14-21, 2018
ISSN 2231-8798

Implementation of energy efficiency in this study is focus on the data based modeling instead of energy optimization. Data based modeling refers to methods that do not applied any physical or chemical parameters of the process, but only handling measured process data as the basis for modeling. These methods are also known as computational methods. The aims of data based modeling have generally been used to predict influent flow, the quality of effluent, biogas production by anaerobic digestion and so on while the aim in this study is to predict energy consumption. Different methodologies applied from machine learning are proposed as the methods in this study in order to compare which method provides more accuracy.

Data based modeling is not something new as it was already done by many researchers. The most popular method applied in data based modeling in wastewater treatment plant case study is artificial neural networks. It has been used to predict components of biogas and the substrate in the process of anaerobic digestion, prediction of methane profiles[3], prediction of effluent oil and grease, prediction of wastewater treatment plants performance[11] and many more. There are also other methods used in data based modeling such as adaptive neuro-fuzzy inference system to model anaerobic digestion system of primary sludge of Kayseri municipal wastewater treatment plant [4], principal component analysis and fuzzy c-means clustering to monitor and control of biological treatment plants during extreme events,[12] linear autoregressive integrated moving average (ARIMA) to analyze the characteristics of wastewater process variables and watershed variables,[13] support vector machines to predict wastewater treatment plant performance,[14] wavelet packet decomposition to predict wastewater treatment plant performance[5] and self-organizing map and clustering for wastewater treatment monitoring.[6]

Data based modeling to predict energy consumption has been implemented in most installation in Malaysia in order to get a baseline for the energy analysis. The methodology used by most Registered Electrical Energy Manager or industries in this country is linear regression since it is available in Microsoft Excel and easy to understand. To see the difference between linear regression and proposed methodologies in this study, a graphical and statistical comparison will be made.

2.0 DESCRIPTION OF THE STUDY SITE

The wastewater treatment plant for this study is located at peninsular Malaysia. There are two systems used at this plant which are aerated lagoon (AL) and conventional activated sludge (CAS). The plant operates and maintains to produce an effluent quality that comply to Standard A under the provisions of the Environmental Quality Act 2009. The major indices are those of Biochemical Oxygen Demand (BOD), suspended solids, Chemical Oxygen Demand (COD), oil and grease, Ammoniacal Nitrogen, Nitrate Nitrogen and Total Phosphorus. Table 1 provides the characteristic of sewage and effluent standards for this plant and Figure 1 shows the wastewater treatment scheme.

Table 1: Characteristic of sewage and effluent standards

Items	**Influent Sewage**	**Effluent Standards**
BOD	120 – 180 mg/l	2 – 2.5 mg/l
Suspended Solids (SS)	180 – 220 mg/l	2 – 4 mg/l
Total Nitrogen (T-N)	28 – 33 mg/l	5 – 12 mg/l

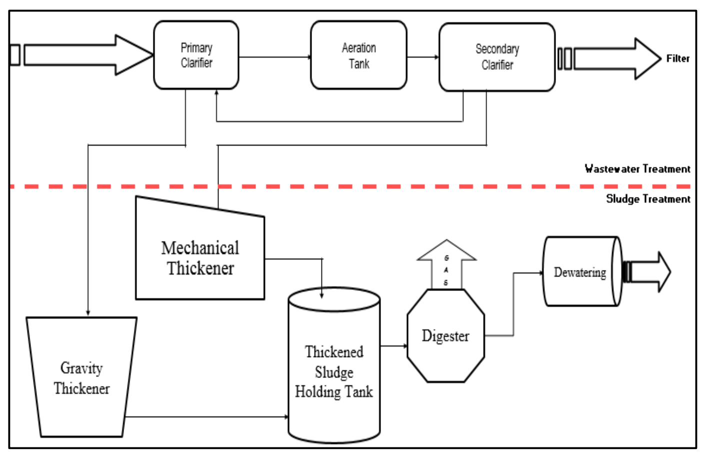

Figure 1: Process flow diagram of the wastewater treatment plant

3.0 METHODOLOGY AND DATA COLLECTION

The indicator used in this study is cubic meter (m^3) of sewage treated per day while the energy units will be measured in kWh/m^3 sewage treated per day. The current energy performance indicator (EnPI) before the analysis conducted was 0.44 kWh/m^3. The target and expected reduction is about 10%. Thus, it is expected that the EnPI after the analysis is 0.40 kWh/m^3. Based on an average flow of 1.3million

m^3/month of flow, the reduction in electrical energy usage is 1,300,000 X (0.44 – 0.40) = 52,000 kWh and estimated saving is 52,000 X RM0.337/kWh = RM 17,524/month. As the plant equipment and processes do not have any sub meter and there is no electrical meter for monitoring, the only way to verify the savings is through the Tenaga National Berhad (TNB) bills. All the electrical bills from TNB are collected from March 2011 until February 2015 and the data are graph as in Figure 2.

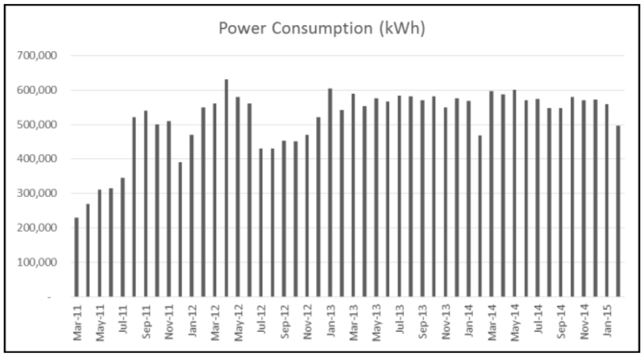

Figure 2: Energy consumption of wastewater treatment plant from March 2011 until February 2015

A. Artificial Neural Network

Artificial neural networks are one of the most popular method used in wastewater treatment plant study. There are two methods in artificial neural networks which are multilayer perceptron (MLP) and radial basis function (RBF). Radial basis function is applied in this analysis since this approach are keeping the mathematics simple and the computations relatively cheap. Furthermore, RBF networks are particular type of linear model which made it comparable with linear regression. It also had been proven by Moody and Darken[7] where they demonstrated that the RBF type networks learn faster than MLP networks.

Radial functions are simply a class of functions. Generally, they could be employed in any sort of model either linear or nonlinear and any sort of network. The RBF network can be nonlinear if the basis functions can move or change size or the number of hidden layer is more than one. In this study, SPSS software is used to run the analysis on energy prediction. Under the analyze menu, there are two options under neural networks which are radial basis function and multilayer perceptron.

The RBF procedure fits a radial basis function neural network, which is feedforward, supervised learning network with an input layer, a hidden layer called the radial basis function layer, and an output layer. The connections between the input and the hidden layers are not weighted. The hidden layer transforms the input vectors into radial basis functions but the input reach the hidden layer nodes unchanged. For an input X^i, the jth hidden node produces a response h_j given by;

$$h_j = e^{\left\{\frac{-\left\|X^i - U_j\right\|}{2\sigma_j^2}\right\}} \tag{1}$$

where $\left\|X^i - U_j\right\|$ is the distance between the point representing the input X^i and the center of the jth hidden node as measured by some norm. The output, y_{ik} of the network at the output node is given by;

$$y_{ik} = \sum_{j=1}^{L} h_j w_{kj} \tag{2}$$

where w_{kj} are weights. The weight updating therefore follows of the following rule:

$$w_{kj}^{new} = w_{kj}^{old} + \frac{\partial E}{\partial w_{kj}}$$

$$\frac{\partial E}{\partial w_{kj}} = -\eta \frac{\partial E}{\partial z_i} \frac{\partial z_i}{\partial y_k} \frac{\partial y_k}{\partial w_{kj}} \tag{3}$$

where η is the learning rate which controls the size of the gradient descent step.

Learning rate can be very sensitive in choosing the right value. Theoretically, too small learning rate would lengthen the estimation time while too large learning rate may cause network oscillation in the weight space. The momentum parameter is added to modify the basic back propagation so that the oscillation in weight changes and a weight decay term that penalizes the overly complex network with large weights can be controlled. In this study, these values are fixed where learning rate and momentum are set to 0.3. The learning rate and momentum are set to 0.3 based on the results of the experiment on the main effects of the learning rate, momentum rate and number of hidden nodes done by Okafor and Adetona[8]. The combination of low learning rate and momentum rate gives much lower mean prediction error than a high learning rate of 0.6 and momentum rate of 0.6.

The RBF procedure trains the network in two stages where; firstly, the procedure determines the radial basis functions using clustering methods. The center and width of each radial basis function are determined here. Secondly, the procedure estimates the synaptic weights given the radial basis functions. The sum-of-squares error function with identity activation function for the output layer is used for prediction. Ordinary Least Squares regression is used to minimize the sum-of-squares error. Due to this two-stage training approach, the RBF network is in general trained much faster than multilayer perceptron. Table 2 shows the network information on RBF that we obtained from SPSS software.

Table 2: Network information on RBF

Input Layer	Factors	Flow
	Number of Units	33
Hidden Layer	Number of Units	9
	Activation Function	Softmax
Output Layer	Dependent Variables	Power
	Number of Units	1
	Rescaling Method for Scale Dependents	Standardized
	Activation Function	Identity
	Error Function	Sum of Squares

B. K-nearest neighbor method

In machine learning, the k-nearest neighbor method (k-NN) is also known as lazy learning because of its training is held up to run time. This classifier is also one of the most straightforward and simplest since classification of the datasets is based on their nearest neighbors' class. The data sets are consequently allocated to the class that's more similar and k must be a positive integer. The value of k is usually small. When k=1, the datasets are basically allocated to the class of its nearest neighbor.

At first, the classifier was studied in 1951 by Fix and Hodges in the US Air Force School of Aviation Medicine. Then, Cover and Hart[9] formalized the idea and invented the main properties of this method. They also described this classifier more properly and found the upper error bound of the method to be twice of Bayes' error probability.

The k-NN classifier performance depends on the choice of a distance that is used**.** There are four different types of distance in the k-NN but in this analysis, Manhattan Distance is used since it is better than Euclidean distance[10]. The Manhattan distance is also known as rectilinear distance, city block distance, Minkowski's L_1 distance or taxi cab metric. The Manhattan name is taken from the place itself where the grid layout of most streets and it is based on the shortest path that anything could take between two intersections in the area to have equality length to the intersections' distance. If an attribute is numeric, thus the local distance function can be written as:

$$dist'(x, y) = \sum_{i=1}^{n} |x_i - y_i| \tag{4}$$

Manhattan distance refers to the global distance that is calculated as the sum of these local distances.

The feature of the training data is proportional to the results of the classification. Simultaneously, the choices of the number of k also playing the important role as for the different number of k will give the different outcome. In computing the high-dimensional data, the k-NN algorithm is a straightforward method among others. However, if the set of test, train, and data dimension are too large, the computational complexity will be massive and it takes a long time to operate. To improve the efficiency of the algorithm, we use optimization which is Linear Nearest Neighbor Searching as it can lower precision. In this study, SPSS software is used to predict the energy consumption in wastewater treatment plant.

C. Support Vector Machines

Support Vector Machine (SVM) classifier is one of the most effective machine learning algorithms for classification problems. SVM is a supervised learning algorithm that belongs to the class of discriminative models. The main idea of SVM is to achieve the maximum generalization ability by constructing the optimal hyperplane which has the largest distance between the different types of sample sets in sample space or feature space. When the sample is linearly separable, SVM solves the maximal-margin solution in the sample space. When the sample is linearly nonseparable, SVM is mapping sample set to high dimensional space through proper kernel function, thus the sample set is linear separable in high dimensional space.

In linear separable condition, (x_i, y_i) is the sample of data set where $x \in R^d, y \in \{-1, +1\}$ and i = 1, 2, …, n. The general form of linear classification function in d-dimensional space is $g(x) = w \cdot x + b$ where *w* is the normal vector to the hyperplane and *b* is the bias. For the hyperplane, the equation is

$$w.x + b = 0 \quad (5)$$

The parameter $b/\ \|w\|$ determines the offset of the hyperplane from the origin along the normal vector w. By using geometry, the distance between these two hyperplanes is $2/\| w \|$. Maximizing the classes margin equal to minimize $\| w \|$ or $\| w \|^2$. If a hyperplane can correctly classify all the samples, it must meet:

$$y_i[(w.x) + b - 1] \geq 0, i = 1,2, \dots, n \quad (6)$$

A hyperplane which meet the above condition and let $\| w \|^2$minimize is called optimal hyperplane.

The solution of the optimal hyperplane can be transformed into the following constrained optimization problem. That is by finding the minimum value of the following functions in the constraints of the equation 6.

$$\varphi(w) = \frac{1}{2}\|w\|^2 = \frac{1}{2}(w.w) \quad (7)$$

After solving the above problems, the optimal classification function is;

$$f(x) = \text{sgn}\{(w^*.x) + b^*\}$$
$$= \text{sgn}\{\textstyle\sum_{i=1}^{n} a_i{}^* y_i (x_i.x) + b^*\} \quad (8)$$

In which sgn() is the sign function. Equations above applied only to the condition that samples are strictly linear separable. A relaxation $\xi_i \geq$ is added in the equation 6 to meet the conditions if samples are linearly nonseparable. Hence, the equation 6 becomes

$$y_i[(w.x) + b - 1] + \xi_i \geq 0, i = 1,2, \dots, n \quad (9)$$

At the same time, the objective function also changes to solve the minimum of the following function.[15]

$$\Phi(w, \xi) = \frac{1}{2}\|w\|^2 + C|\textstyle\sum_{i=1}^{n} \xi_i| \quad (10)$$

That is getting the generalized optimal hyperspace by considering both least wrong sub-sample and maximize classes margin. In the equation, $C > 0$ is a constant which is the degree of punishment of wrong sub-samples. Nevertheless, it is still incomplete to deal with nonlinear complex system only through the method above. The most prominent advantage of SVM is that is capable of solving nonlinear problems and not just linearly separable problems. It achieves linear classification and gets the optimal hyperspace by transforming the nonlinear problem into a linear problem in high dimensional space. The input space is transformed into a high dimensional space by an inner product defined nonlinear transform function.

This nonlinear transformation is achieved by defining the suitable inner product function which is also known as kernel function. Different kernel will have a different SVM and different formation of algorithms. Equation 11 shows Radial Basis Function (RBF) kernel.

$$K(x, x_i) = \exp\{-\frac{|x-x_i|^2}{\sigma^2}\} \quad (11)$$

There is a big difference between the RBF for SVM and the traditional RBF neural network where each center of its basis function corresponds to a support vector and output values are automatically determined by the algorithm. In this analysis, Statistica software is used to obtain the prediction of energy consumption.

D. Linear Regression

Linear regression method is selected as a comparison with our proposed method in this analysis because it is the most common method used by industries in our country. In this analysis, a scatter plot is created by using Microsoft Excel and the data taken are the active power (kWh) and influent flow (m^3). Regression line is created to describe the relationship between independent and dependent variable where in this case is influent flow and the active power respectively. Year 2012 is taken as our baseline and the linear equation that we obtained is used to predict the energy consumption for the next month. By using Microsoft Excel, the linear regression equation that we obtained is:

$$y = 0.2149x + 234600 \quad (12)$$

Journal of Engineering Technology Vol. 6: 14-21, 2018
ISSN 2231-8798

4.0 RESULTS AND DISCUSSION

From equation 12 that we obtained through regression plot, the next month energy consumption can be predicted. The results of energy prediction by using linear regression is plotted on the same graph with energy prediction that we obtained by three methods from machine learning and actual power consumption. The plotted graph in Fig. 3 shows that the prediction by using artificial neural network (ANN) is more close to actual power consumption compared to other methods.

Besides plotted graph, the prediction by using ANN is better than other methods as it produces least error where the root mean square error is 52084 compared to others as showed in Table 3. ANN also provides second maximum cost saving and better than other methods although one of the problems that occur during neural network training is when the values keep changing. This phenomenon is called overfitting where the values keep changing, thus, the error on the training set is driven to a very small value, but when new data is presented to the network the error is large. The network has memorized the training examples, but it has not learned to generalize to new situations. There are a few methods that can be applied in order to prevent overfitting such as regularization, early stopping, use a network that is large enough to provide an adequate fit and others. In our analysis, we employed regularization to reduce the overfitting.

The amount of cost saving from energy consumption is calculated based on the prediction results. The amount of costs from energy consumption that can be saved by using linear regression is RM 132,732 while the amount that we obtained by using RBF is RM 156,499 which is RM 23,767 difference.

Table 3: Root mean square error and the amount of cost saving for each methods used in this study

Method	Root Mean Square Error	Amount of Cost Saving (RM)
Linear regression	67505	132,732
k-nearest neighbor	60715	109,461
Support Vector Machine	62280	158,939
Artificial Neural Network	52084	156,499

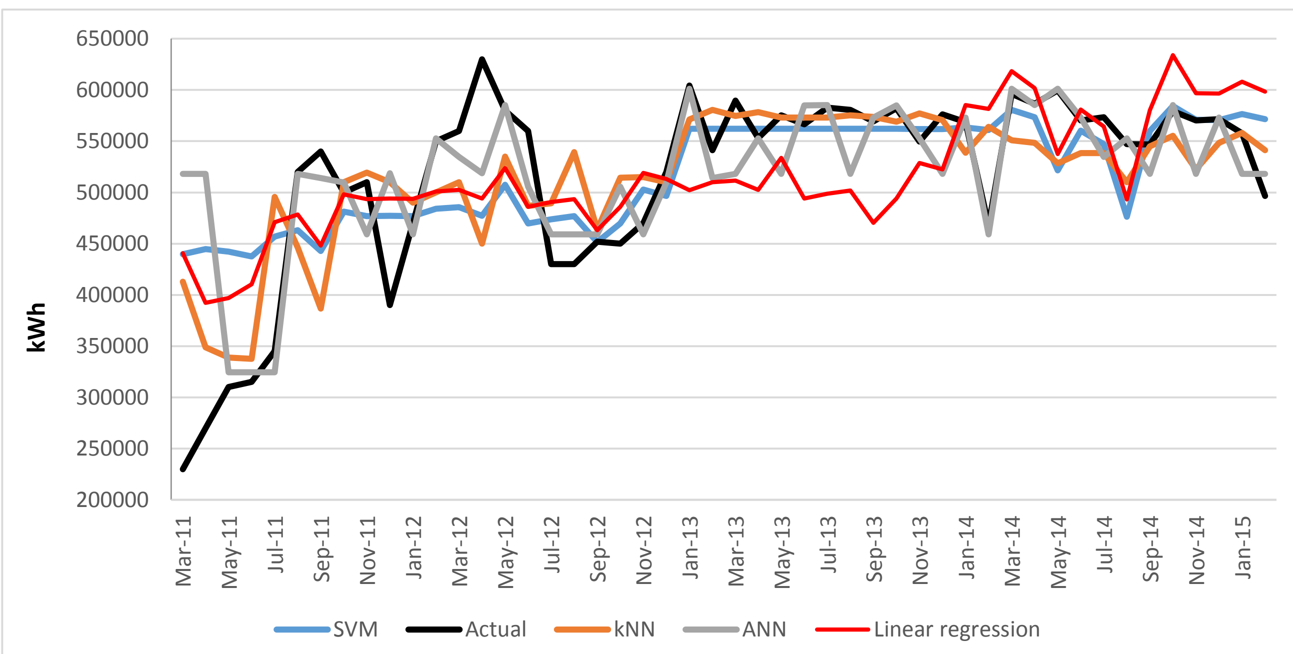

Fig. 3 Comparison of plotted graph of energy consumption between all methodologies applied in this study

5.0 CONCLUSIONS

Electricity cost is the major expenses in operating wastewater treatment plant. In order to reduce the cost, energy consumption need to be optimized. To optimize the energy consumption, a few actions can be taken such as proposal on optimization measures for both aerators and blowers or predict the energy consumption by using data based modeling like in this study. From the results that we obtained by using machine learning methods and linear regression, the maximum amount of production costs that can be saved is by using support vector machine which RM 158,939. However, support vector machine has second highest root mean square error. Method that provide least error and maximum amount of production is artificial neural network based on results in Table 3. The findings from this study is useful for industries as it can be used as a baseline study. Moreover, most of industries still use linear regression method as their mathematical tools in predicting energy production.

To sum up, the objective of this study is achieved without compromise the main purpose of wastewater treatment plant which is to treat sewage in comply with the Environmental Quality Act 2009 and Malaysian Sewerage Industry Guideline. For future study, these methodologies can be applied in prediction effluent oil and grease, components of biogas and others.

REFERENCES

[1] C.Y. Shi, *Mass Flow and Energy Efficiency of Municipal Wastewater Treatment Plants,* London: IWA Publishing, 2011.

[2] B. Wett, “Development and implementation of a robust deammonification process,” *Wst. Sci. & Technol 56(7),* pp. 81-88, 2007.

[3] P. Kianmehr, W. Mansoor, F.A. Kfoury, “Prediction of biogas generation profiles in wastewater treatment plants using neural networks,” *Journal of Clean Energy Technologies 2(3)*, pp. 201 – 205, 2014.

[4] M. Cakmakci, “Adaptive neuro-fuzzy modeling of anaerobic digestion of primary sedimentation sludge,” *Bioprocess and Biosystems Engineering 30(5)*, pp. 349-357, 2007.

[5] D. Hanbay, I. Turkoglu, Y. Demir Y, “Prediction of wastewater treatment plant performance based on wavelet packet decomposition and neural networks,” *Expert Systems with Applications 34(2)*, pp. 1038-1043, 2008.

[6] H.L. Garcia, I.M. Gonzalez, “Self-organizing map and clustering for wastewater treatment monitoring,” *Environmental Applications of Artificial Intelligence 17*, pp. 215-225, 2004.

[7] J. Moody, C.J. Darken, “Fast learning in networks of locally-tunes processing units,” *Neural Computation 1(2)*, pp. 281-294, 1989.

[8] A.C. Okafor, O. Adetona, “Multi-sensor integration using neural networks for predicting quality characteristics of end-milled parts. Part II: Interaction effect of training parameters,” in *Intelligent Materials*, Second International Conference Proceedings, C.A. Rogers, Ed. Pennsylvania: CRC Press, 1949, pp 946-959.

[9] T.M. Cover, P.E. Hart, “Nearest neighbor pattern classification,” *IEEE Transactions on Information Theory 13(1)*, pp. 21-27, 1967.

[10] N.A. Ramli, M.T. Ismail, H.C. Wooi, “Designing Early Warning System: Prediction Accuracy of Currency Crisis by using k-nearest Neighbour Method,” *World Academy of Science, Engineering and Technology 79*, pp. 1023-1028, 2013.

[11] F.S. Mjalli, S. Al-Asheh, H.E. Alfada, “Use of artificial neural network black-box modeling for the prediction of wastewater treatment plants performance,” *Journal of Environmental Management 83(3)*, pp. 329-338, 2007.

[12] C. Rosen, Z. Yuan, “Supervisory control of wastewater treatment plants by combining principal component analysis and fuzzy c-means clustering,” *Water Science and Technology 43(7),* pp. 147-156, 2001.

[13] S.A. Dellana, D. West, “Predictive modeling for wastewater applications: Linear and nonlinear approaches,” *Environmental Modeling and Software 24*, pp. 96-106, 2009.

[14] D. Ribeiro, A. Sanfins, O. Belo, “Wastewater treatment plant performance prediction with support vector machines,” in *ICDM’13 Proceedings of the 13th International Conference*

on Advances in Data Mining: Applications and Theoretical Aspects, pp. 99-111.

[15] S.R. Gunn, *Support vector machines for classification and regression*. Dissertation, University of Southampton, 1998.

[16] A.W. Jayawardena, D.A.K. Fernando, M.C. Zhou MC, "Comparison of Multilayer Perceptron and Radial Basis Function networks as tools for flood forecasting," Destructive Water: Water-Caused Natural Disasters, their Abatement and Control, IAHS Publication 239, pp. 173-181, 1997.