# Forecasting the Environmental Parameters of Water Resources Using Machine Learning Methods

Farshid Faraj[1] and Haojing Shen[2]

[1]School of Chemical Engineering, College of Engineering, University of Tehran, Tehran, Iran
[2]Department of Civil Engineering, College of Engineering, University of Texas at Arlington, Arlington, Texas, USA
Email: `farshid_faraji@ut.ac.ir`

**Abstract** Careful monitoring of the quality and quantity of water resources plays a significant role in modern environmental management. Thus, to achieve this aim, the quantitative and qualitative parameters of water resources should be measured and controlled with a desirable accuracy. However, it is not always possible to measure these parameters with easy, inexpensive, precise, and quick experimental methods. Therefore, today to solve this type of problems, new methods including smart methods are used which have a great potential in many computational areas. Considering the variety of the studies and absence of a comprehensive review paper, this research should be conducted. The aim of this paper is to comprehensively review application of the smart methods of artificial neural network and support vector machine in the area of water resources and to develop a comprehensive study source for the researchers interested in this field. The results of these studies all show that these advanced smart methods are more efficient, accurate, economical, and faster than other computational methods to predict quantitative and qualitative parameters of water resources.

**Keywords:** Qualitative and quantitative parameters of water resources, artificial neural network and support vector machine intelligent methods, modern environmental management.

## 1 Introduction

Today, determination of the qualitative and quantitative characteristics of water resources is one of the most important scientific concerns around the world. Thus, monitoring these parameters is important for presenting essential information in the area of water resources and environmental management (Akram et al., 2013; Nahvi et al., 2018). Among the important qualitative and quantitative parameters of waters are salinity, hardness, acidity, alkalinity, turbidity, color, odor, taste, temperature, total dissolved solids (TDS), total suspended solids, dissolved oxygen (DO), chemical oxygen demand (COD), biological oxygen demand (BOD), bacteriological characteristics, electrical conductivity, flow rate, flow velocity, hydraulic gradient, pressure, hydraulic conductivity, and water table slope (Shen et al., 2017). Precise measurement of some of these parameters such as BOD is not possible due to low accuracy and precision of the laboratory equipment, being time-consuming, and staggering costs of analyses. Accordingly, today novel computational methods including artificial intelligence are used for estimating, classifying, and optimizing this type of parameters (Zadeh, 1997; Palani et al., 2008; Daghighi and Nahvi, 2014; Maier and Dandy, 2001).

## 2 Overview of Intelligent Methods

In recent years, intelligent methods have been under use in different areas of engineering and sciences to solve computationally complex and bulky problems. Artificial neural networks (ANNs), genetic algorithm (GA), fuzzy systems (FS), swarm intelligence (SI), evolutionary algorithms (EA), etc. are among intelligent methods (Daghighi, 2017; Kecman, 2001). Many natural living and non-living systems are more efficient than human-made systems. Self-guidance and reproduction are rules in biological and natural systems, some of which can hardly be obtained even in human-made systems. This has prompted researchers to care about natural systems and the intelligence governing their behavior as a model for designing engineering

systems (Nahvi et al., 2018). Such attempts have resulted in development of a set of optimization methods adapted from the nature or based on artificial intelligence. They are categorized into two major subsets: methods based on living nature and methods adapted from non-living nature. Most methods that are based on living nature are inspired by the evolutionary process of creatures in the nature. In this process, the characteristics of animals are improved so that they can better adapt to the environment and overcome the competition with other animals to obtain the limited natural resources.

The optimization methods based on intelligent methods are dependent on the group of direct methods and are very similar to stochastic search methods. Stochastic search methods only enjoy general search strategy, while methods based on artificial intelligence run general and local search strategy simultaneously (Daghighi et al., 2017). Indeed, intelligent methods are intelligent dynamic systems based on experimental data, which transfer the knowledge or law beyond data to the network structure by processing experimental data. For this reason, these methods are called intelligent, as they encompass general rules based on computation on numerical data. In these methods, observational data are trained to the model and after the training, the model performs the prediction and simulation with a suitable accuracy. Some of these intelligent methods and their application in prediction of quantitative and qualitative parameters of water resources are further investigated in this paper.

## 3   Application of Artificial Neural Network

Artificial neural network is a set of links among units or nodes, which try to present a function like the set of human brain neurons (Cilek, 2002). The network's processing potential emanates from the type of relationship between units and weights of the links, which is either obtained through learning process or already predetermined. A neuron is the smallest unit of information processing (Specht, 1991). Once several neurons that act in parallel assemble together, a layer is formed. Every layer has a special task and with combination of layers, artificial neural networks are developed. Considering the objective for which the artificial neural network is designed, various layers are considered (Laberge et al., 2000). Artificial neural networks are an information processing system which are developed with simulation and simplification of neural networks of the human body. The output of other neurons is multiplied by different weights and reaches a new neuron as input. These inputs leave the neuron after aggregation and passing through a function called transfer function. Therefore, each neuron has a large number of inputs and only one output. Connection of a large number of these neurons to each other can form an artificial neural network (Hagan et al., 1996).

An artificial neural network consists of the constituent layers and weights. The network behavior is dependent on the interrelationships between the members (Demuth and Beale, 1993). Generally, there are three types of neural layers in an artificial neural network (Demuth and Beale, 1993):

The input layer: receiving raw information that has been fed into the network.

Hidden layer: the function of the hidden layer is determined by the inputs and the weight of the relationship between them and the hidden layers.

The output layer: the function of the output layer is contingent upon the activity of the hidden layer and the weight of the relationship between the hidden and output unit.

Both layers of a network are interrelated through weights and indeed the links. In ANNs, there are several types of link or weighted bond:

Feed-forward: the majority of links are of this type, in which the signals move in only one direction. There is no loop from the input to the output. The output of each layer has no effect on that layer.

Recurrent: the data are fed back into the nodes of the low layers from the nodes of the higher layer.

Lateral: the outputs of the nodes of each layer are used as the input of the nodes of the same layer.

Any ANN goes through training, validation, and running stages. Indeed, ANNs can be used for solving problems for which precise mathematical relations do not exist between its inputs and outputs. These networks consist of simple elements which act in parallel. Inspired by biological networks, each of these elements is called neuron (Haghiri et al., 2018). Training ANNs is nothing but adjusting the relational weights of these neurons in response to receiving different examples, so that the network output converges to a desirable output (Sadeghiamirshahidi and Ardejani, 2013).

ANNs can process the input signals according to their design and covert them into their desired output signals. Typically, once an artificial neural network was designed and implemented, the parameters

of the weight and bias per sets of input signal should be adjusted in a way that the output signals of the network develop the desired output. Such as process is called ANN training (in the first training stage, the bias and weight values are chosen randomly, because as long as these parameters have no value, the ANN will not be usable). During ANN training (i.e. gradually concurrent with the increase in the times of adjustment of parameter values to achieve a more desirable output), the value of the parameters approaches its real and final value. Generally, there are two methods for training ANN: Supervised and unsupervised methods (Rajabi et al., 2010).

Next, we investigate the papers published in the recent decade on application of this method for environmental issues of water resources:

Maier et al examined use of ANN to monitor the treated water quality. In this paper, temperature, acidity, turbidity, total hardness, and calcium level before Zarinehroud water treatment procedure were considered as the input variables, while the post-treatment total dissolved solids and electrical conductivity were regarded as the output neuron of the model. To better interpret the model efficiency, the results were compared with other classic and applied models, proving the competence of ANN to predict the parameters of treated water (Maier et al., 2004).

Kuo et al employed ANN with post-propagation training algorithm for predicting the reservoir eutrification of Te-Chi reservoir, which is a key factor in management of water resources in Taiwan. The reason of choosing this method has sprung from a number of complex factors including limited flow data, meteorological uncertainty, dramatic changes in the reservoir water level, overdevelopment of the basin with agricultural activities and constant increase in nutrient loads into the reservoir. In this study, four models have been developed to predict four indices of dissolved oxygen, total phosphorus (TP), chlorophyll-a (Chl-a), and Secchi Disk Depth (SD) separately. The results of this study have had a high validity for the models, as the correlation coefficient for the TP in the training and test stages has been 0.88 and 0.83, respectively. For the SD, DO, and Chl-a, the values have been 0.85 and 0.81; 0.75 and 0.72; and 0.92 and 0.86, respectively, which are acceptable correlations (Kuo et al., 2007).

Zhao et al used a post-propagation ANN model with Levenberg-Marquadt training algorithm for predicting the required chemical oxygen and dissolved oxygen of the next day using 10 water quality variables on a daily basis including water temperature, turbidity, acidity, alkalinity, chloride, ammonium nitride, nitrogen nitrite, hardness, COD, and DO. To evaluate the results, the values predicted for COD and DO were compared with the real measured values. The results revealed good efficiency of the developed model (Zhao et al, 2007).

Dogan et al used feed-forward ANN to model BOD in Melen River in Turkey. In that research, COD, nitrogen ammonia, Chl-a, nitrogen nitrite, temperature, nitrogen nitrate, DO, and water flow rate were used for the model, and eight ANN models were developed using each parameter individually. To find the most effective input parameter, sensitivity analysis was used. The correlation coefficient was determined for each of the parameters, with the most effective parameter being COD with the highest correlation coefficient of 0.7458. The model results suggest its potential in predicting the parameter of interest (Dogan et al., 2009). Singh et al modeled Gomti River water quality in India by calculating the level of DO and BOD using feed-forward ANN. They chose water acidity, total alkalinity, total hardness, total suspended solids, COD, ammonia nitrogen, nitrogen nitrate, chloride, phosphate, potassium, and sodium as the independent variables. Selection of the influential factors may significantly affect the model output, such that in this study, the coefficient of determination for BOD for the training, validation, and test stages was obtained as 0.85, 0.85, and 0.77, respectively, while for DO, the values were 0.70, 0.74, and 0.76, respectively. The relatively better performance of the BOD model compared to DO model suggests that the factors affecting selection (input variables) have a relatively greater effect on BOD than on DO. The results show that desirable networks are able to capture the long-term trend observed (10-year) for qualitative variables of water. They can also be used in other areas to better understand the river's contamination trend (Singh et al., 2009).

He et al used the data related to land use, fertilizer, and hydro-meteorological data taken from 59 river basins across Japan to simulate the monthly total nitrogen concentration (TNC) in streams using feed-forward ANN model. Sixteen input variables were used for model development, for which 16 ANN models were also developed using the independent variables separately. In general, good correspondence was achieved between the observed and modeled values for all of the predicted data (training, test, and

validation). This suggests that the results obtained from the model are very good and implementable for predictions of total nitrogen concentration (He et al., 2011).

Jiang et al evaluated the risk of water quality through Mont-Carlo simulation and post-propagation ANN (as an alternative method to finite element method). They significantly reduced the number of water quality models that should have been calculated with losing accuracy. Small generalization error and good correlation (0.9954 for training data and 0.9946 for experimental data) between the general and calculated results suggest that the model can be used for computing the iteration of water qualitative models with a good accuracy (Jiang et al., 2013; Nahvi et al., 2018).

In this study, total dissolved solids were considered as the model output as well as water salinity criterion. The linear regression analyses presented acceptable results for the estimated values against real values for training, validation, and test stages, suggesting the good efficiency of the method proposed in the research.

## 4  Application of Support Vector Machine

For many engineering sometimes were face problems, for which we inevitably deal with a series of sporadic experimental data. To model and use these data, they are typically compared and separated in terms of the number of classes that lie in them. Support vector machine is an effective method for data modeling. This method provides an integrated working framework for most models by enhancing the dimensions of the problem and using Kernel mapping, providing the possibility of comparison. In summary, in the support vector machine method, the input vectors are mapped onto a multi-dimensional space.

By having a known dataset $T = \{x_i, y_i\}_i^m = 1$ where $x_i \in R^n$, representing the input vectors $yi \in \{+1, -1\}$, denoting two-class state, while m shows the number of samples, finding polyhedron f(x)=0 for separating the input data is possible if the two separable classes are linear.

(1) $f(x) = w, x + b = \sum_{k=1}^m (w_k, x_k + b) = 0$

In the above equation, w is weighted vector and b is the bias. w and b are used for determining the position of the separator polyhedron. This separator should satisfy the following constraints.

(2) $y_i, f(x) = y_i, (w_i, x_i + b) \geq 1, i + 1, 2, 3, ..., m$

By introducing multiple Lagrangian $\alpha_i$, the problem of finding the optimal separator polyhedron is formulated as follows:

(3) Maximize $L(\alpha) = \sum_{i=1}^m (\alpha_i) = -\frac{1}{2} \sum_{i,j=1}^m (\alpha_i \alpha_j y_i y_j (x_i, x_j))$

Subject to $\sum_{i,j=1}^m (\alpha_i y_i = 0, \alpha_i \geq 0, i = 1, 2, ..., m)$

The linear decision function is developed by solving the above optimization problem, which is defined as follows:

(4) $f(x) = sign(\sum_{i,j=1}^m (\alpha_i y_i (x_i, x_j) + b)$

Support vector machine method is used in nonlinear classification through focal function. Using the nonlinear mapping function $\varphi$ where there is linear classification, the main data x is mapped onto the feature space with the above dimensions.

(5) $k(x_i, x_j) = exp(-|x_i - x_j|/2\sigma 2)$

$K(x_i, x_j)$ is called the focal function and $K(x_i, x_j) = \varphi(x_i)\varphi(x_j)$. In the support vector machine method, the radical basis function (RBF). The support vector machine created by the RBF has a great potential in nonlinear classification. For this reason, in most studies, RBF function is used in the support vector machine algorithm. Here, C and $\alpha$ are the parameters specified by the designer. Good selection of the parameters plays a significant role in the performance of support vector machine. The optimal value of these parameters can be obtained using optimization algorithms.

Note that considering the breadth of content about support vector machine method, there are some contents which have not been mentioned for the sake of brevity. Thus, to better understand the support vector machine method, it is recommended to study the references (Shimawaki et al., 2006; Ng and Gong, 2002; Shin et al., 2005).

The papers of the recent decades on application of this method in environmental issues of water resources are discussed further. Bashi et al investigated the method to point out an unknown contamination source (positioning of the source and extent of leakage) in groundwater reservoir systems in the southern part of Tehran Refinery using probabilistic support vector machines (PSVM) and probabilistic neural

networks (PNN). In this study, based on the water quality data available and the storage volume of the reservoirs, total petroleum hydrocarbon has attracted attention as a detector of water quality. The results suggest that in this paper, the accuracy of PSVM is more than that of PNN, with the former having lower error in practice (Ayvaz, 2010). By integrating biological monitoring methods and computer vision technology, Liao et al studied the toxic effects of copper ion in water considering the behavior of a special type of organisms. In this study, support vector machine method has been used for comprehensive evaluation of information with regards to the behavioral differences of organisms in different toxic aqueous environments, for which three different classes of toxicity degree have been developed. This high accuracy of this estimation suggests its suitable efficiency in indirect evaluation of heavy metals in aqueous environments (Liao et al., 2011).

Tan et al predicted the timeseries data of water quality based on Least Squares Support Vector Machine (LS-SVM) method. In this paper, to achieve the intended objective, comparisons were made between LS-SVM, post-propagation neural network, and radial basic network. The minimum mean squared error and relative mean error of the first method were far lower compared to the two other methods. Accordingly, it can be concluded that the LS-SVM method had a greater accuracy and potential for qualitative studies of water with regards to timeseries in this study (Tan et al., 2012).

Zheng et al investigated application of SVM in identification of mutagenicity of contaminating compounds in groundwaters of China. In this study, instead of considering the total contamination profile, the first 10 compounds were considered based on mass distribution. It was found that consideration the 10 more important compounds suggests greater predictability of efficiency compared to the total contamination profile (Zheng et al., 2013).

**Table 1.** ANN vs. vector machine summary table

| ANN Method | | | |
|---|---|---|---|
| Author | Year | Aim | Output |
| Maier | 2004 | monitoring the treated water quality | total dissolved solids and electrical conductivity |
| Kuo | 2007 | predicting the reservoir eutrification | dissolved oxygen, total phosphorus (TP), chlorophyll-a (Chl-a), and Secchi Disk Depth (SD) separately |
| Zhao | 2007 | predicting the required chemical oxygen and dissolved oxygen | water temperature, turbidity, acidity, alkalinity, chloride, ammonium nitride, nitrogen nitrite, hardness, COD, and DO |
| Dogan | 2009 | | COD, nitrogen ammonia, Chl-a, nitrogen nitrite, temperature, nitrogen nitrate, DO, and water flow rate |
| Singah | 2009 | Modeling water quality | DO and BOD |
| He | 2011 | simulating the monthly total nitrogen concentration (TNC) | total nitrogen concentration |
| Jiang | 2018 | evaluating the risk of water quality | total dissolved solids |
| Vector Machine | | | |
| Author | Year | Aim | Output |
| Bashi | 2010 | Monitoring contamination source in groundwater reservoir | total petroleum hydrocarbon |
| Liao | 2011 | studing the toxic effects of copper ion in water | heavy metals level |
| Tan | 2012 | predicting the timeseries data of water quality | |
| Zheng | 2013 | identification of mutagenicity of contaminating compounds in groundwaters | |
| Cao | 2016 | evaluating absorption of copper ion from water on membrane absorbent | |

Cao et al evaluated absorption of copper ion from water on membrane absorbent modified with the intelligent method of LS-SVM. The results revealed that the developed model can present a good estimation using experimental data. The statistical tests performed also showed that all of the experimental data lie within an acceptable range with suitable confidence intervals. The correlation coefficients for all of the stages of training, test, and validation were obtained to be above 0.99, confirming the high power of the developed model (Cao et al., 2016).

## 5  Conclusion

Water resources are among the principal pillars of life on the Earth. Today, due to the growth of population and overuse, these resources have been experiencing severe crises such as contamination and reduction of storage volume. Thus, for proper management of these resources under the current conditions, measurement of the qualitative and quantitative indices of waters seems to be essential. Nevertheless, it is not economical with experimental methods and sometimes novel methods can be used instead of traditional methods. In this paper, to achieve this aim, the efficiency of two types of intelligent methods was investigated and confirmed. The results of this paper suggest desirable efficiency of these methods in the area of studying water resources. In this regard, methods such as artificial neural network and support vector machine have a better potential in estimation and prediction of unknown parameters using a certain number of parameters, while methods such as genetic algorithm are better for optimizing the parameters used in estimator models.

## References

1. Laberge, C., Cluis, D. and Mercier, G., 2000. Metal bioleaching prediction in continuous processing of municipal sewage with Thiobacillus ferrooxidans using neural networks. Water Research, 34(4), pp.1145-1156.
2. Akram, M., Azari, A., Nahvi, A., Bakhtiari, Z. and Safaee, H.D., 2013. Subsurface drainage in Khuzestan, Iran: environmentally revisited criteria. Irrigation and Drainage, 62(3), pp.306-314.
3. Shimawaki, T., Miura, J., Sakiyama, T. and Shirai, Y., 2006, May. Ball route estimation in broadcast soccer video. In Proc. ECCV-2006 Workshop on Computer Vision Based Analysis in Sport Environments (pp. 26-37).
4. Ayvaz, M.T., 2010. A linked simulationâĂŞoptimization model for solving the unknown groundwater pollution source identification problems. Journal of Contaminant Hydrology, 117(1-4), pp.46-59.
5. Zadeh, L.A., 1997. What is soft computing?. Soft computing, 1, pp.1-2.
6. Palani, S., Liong, S.Y. and Tkalich, P., 2008. An ANN application for water quality forecasting. Marine Pollution Bulletin, 56(9), pp.1586-1597.
7. Daghighi, A. 2017. Harmful Algae Bloom Prediction Model for Western Lake Erie Using Stepwise Multiple Regression and Genetic Programming, Electronic Thesis, available at: https://etd.ohiolink.edu/ (last access: 24 January 2018).
8. Daghighi, A., Nahvi, A. and Kim, U. 2017. Optimal Cultivation Pattern to Increase Revenue and Reduce Water Use: Application of Linear Programming to Arjan Plain in Fars Province. Agriculture, 7(9), 73.
9. Daghighi, A., Nahvi, A.: Effect of different additives on fatigue behaviour of asphalt mixtures. In: Ekolu, S.O., Dundu, M., Gao, X. (eds.) Construction Materials and Structures: Proceedings of First International Conference on Construction Materials and Structures, pp. 601–607. IOS Press, Amsterdam (2014)
10. Demuth H, Beale M.1993. Neural network toolbox for use with MATLAB
11. Dogan E, Sengorur B, Koklu R.2009. Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique. Journal of Environmental Management.90:1229-1235
12. Sadeghiamirshahidi, M. and Ardejani, F.D., 2013. Application of artificial neural networks to predict pyrite oxidation in a coal washing refuse pile. Fuel, 104, pp.163-169.
13. Rajabi, M., Bohloli, B. and Ahangar, E.G., 2010. Intelligent approaches for prediction of compressional, shear and Stoneley wave velocities from conventional well log data: A case study from the Sarvak carbonate reservoir in the Abadan Plain (Southwestern Iran). Computers and Geosciences, 36(5), pp.647-664.
14. Hagan MT, Demuth HB, Beale MH .1996. Neural network design. Pws Pub. Boston
15. Haghiri, S., Daghighi, A., and Moharramzadeh, S. 2018. Optimum coagulant forecasting by modeling jar test experiments using ANNs. Drinking Water Engineering and Science, 11(1), 1-8.
16. He, B., Oki, T., Sun, F., Komori, D., Kanae, S., Wang, Y., Kim, H. and Yamazaki, D., 2011. Estimating monthly total nitrogen concentration in streams by using artificial neural network. Journal of Environmental Management, 92(1), pp.172-177.

17. Jiang, Y., Nan, Z. and Yang, S., 2013. Risk assessment of water quality using Monte Carlo simulation and artificial neural network method. Journal of environmental management, 122, pp.130-136.

18. Cilek, E.C., 2002. Application of neural networks to predict locked cycle flotation test results. Minerals Engineering, 15(12), pp.1095-1104.

19. Kecman, V., 2001. Learning and soft computing: support vector machines, neural networks, and fuzzy logic models. MIT press.

20. Kuo J-T, Hsieh M-H, Lung W-S, She N.2007. Using artificial neural network for reservoir prediction. Ecological Modelling.200:171-177.

21. Liao Y, Xu J, Wang W.2011. A method of water quality based on biomonitoring and multiclass support vector machine. Procedia Environmental Sciences.10:451-457.

22. Maier HR, Dandy GC.2001. Neural network based modelling of environmental variables: a systematic approach. Mathematical and Computer Modelling.33:669-682

23. Maier HR, Morgan N, Chow CW.2004. Use of artificial neural networks for predicting optimal alum doses and treated water quality parameters. Environmental Modelling and Software.19:485-494

24. Nahvi, A., Daghighi, A. and Nazif, S. 2018. The environmental impact assessment of drainage systems: a case study of the Karun river sugarcane development project. Archives of Agronomy and Soil Science, 64(2), 185-195.

25. Nahvi, A., Sadati, S.S., Cetin, K., Ceylan, H., Sassani, A. and Kim, S., 2018. Towards resilient infrastructure systems for winter weather events: Integrated stochastic economic evaluation of electrically conductive heated airfield pavements. Sustainable Cities and Society, 41, pp.195-204.

26. Ng J, Gong S.2002. Composite support vector machines for detection of faces across views. Image and Vision Computing.20:359-368

27. Cao, H., Xin, Y. and Yuan, Q., 2016. Prediction of biochar yield via least squares support vector machine intelligent approach. Bioresource technology, 202, pp.158-164.

28. Shen, W., Ceylan, H., Gopalakrishnan, K., Kim, S., and Nahvi, A. 2017. Sustainability assessment of alternative snow-removal methods for airport apron paved surfaces. No. DOT/FAA/TC-17/34 https://trid.trb.org/view/1472632.

29. Shin, K.S., Lee, T.S. and Kim, H.J., 2005. An application of support vector machines in bankruptcy prediction model. Expert Systems with Applications, 28(1), pp.127-135.

30. Singh, K.P., Basant, A., Malik, A. and Jain, G., 2009. Artificial neural network modeling of the river water quality—a case study. Ecological Modelling, 220(6), pp.888-895.

31. Specht, D.F., 1991. A general regression neural network. IEEE transactions on neural networks, 2(6), pp.568-576.

32. Tan, G., Yan, J., Gao, C. and Yang, S., 2012. Prediction of water quality based on least squares support vector machine. Procedia Engineering, 31, pp.1194-1199.

33. Zhao, Y., Nan, J., Cui, F.Y. and Guo, L., 2007. Water quality forecast neural network at Yuqiao reservoir. Journal of Zhejiang University-Science A, 8(9), pp.1482-1487.

34. Zheng, W., Tian, D., Wang, X., Tian, W., Zhang, H., Jiang, S., He, G., Zheng, Y. and Qu, W., 2013. Support vector machine: classifying mutagenicity of complex mixtures based on pollution profiles. Toxicology, 313(2-3), pp.151-159.