https://doi.org/10.24867/GRID-2022-p89 Review paper

MACHINE LEARNING AS A SUPPORT TOOL IN WASTEWATER TREATMENT SYSTEMS – A SHORT REVIEW

Sanja Radović [1], Sabolc Pap [1,2], Maja Turk Sekulić [1]
[1] University of Novi Sad, Faculty of Technical Sciences, Department of Environmental Engineering and Occupational Safety and Health, Trg Dositeja Obradovica 6, 21 000 Novi Sad, Serbia
[2] Environmental Research Institute, UHI North Highland, University of the Highlands and Islands, Thurso, Scotland, KW14 7JD, UK

Abstract: *Machine learning (ML) is a subset of artificial intelligence (AI). It is based on teaching computers how to learn from data and how to improve with experience. This valuable technique has been increasingly supporting different spheres of life. This includes ML application in enhancement and optimisation of many ecological and environmental engineering solutions, such as wastewater treatment systems (WWTS). Complexity of processes triggers challenges in ensuring good effluent quality by adequate response to dynamic process conditions. That is why techniques such as ML which, after being trained, have strong prediction ability, have been applied in WWTS. ML facilitates understanding of correlation between input features and output targets through a data-driven approach. Different ML models have been used for this purpose. Some of the commonly used were artificial neural network (ANN) or deep neural network (DNN) model, support vector machine (SVM) and its variation support vector regression (SVR) model, random forest (RF) model and many others. More often authors apply a few different models in order to obtain the one that most appropriately works for specific problem. In wastewater management those problems are various, and could include modelling of WWT processes, prediction of certain technology performance, optimisation of technology working parameters, optimisation of the production of the materials there are being used in WWT technology etc. For instance, there are several articles which describes ML power in optimisation of material synthesis (e.g., biochar production). Application of ML led to reduction in number of runs which were necessary for obtaining the best results by applied production procedure, which saved time and was also cost-beneficial. Indeed, ML incorporation in solving or avoiding potential problems within WWTS is a promising approach which has gained more attention in recent years due to the exponential technology development and progress in artificial intelligence application.*

Key words: artificial intelligence, machine learning, wastewater treatment technology, prediction and process optimisation

1. INTRODUCTION

In order to more efficiently evaluate complex data and improve traditional types of data processing within the environmental management field, different advanced computer-based techniques could be applied (Corominas et al., 2018). One of the subsets of artificial intelligence which is recognized and more often applied for the mentioned purpose is machine learning (Sundui et al., 2021). Among versatile environmental science and engineering scientific fields, machine learning analytical tool seems to be mostly applied in the field of water management (Zhong et al., 2021). ML algorithms have proven their efficiency in enhancing many segments of wastewater treatment, such as: energy cost modelling (Torregrossa et al., 2018), monitoring programme, anomalie detection, evaluation of wastewater infrastructure, optimization of treatment technologies, detection of process hot spots by life cycle assessment (Zhong et al., 2021), optimization of material synthesis and prediction of material or process behaviour (Li et al., 2019; Paula et al., 2022).

To address complex problems, ML approach is using so called "training data" to build models/algorithms which will be able to make predictions or decisions, without being explicitly programmed for that (Zhong et al., 2021). Among many available ML algorithms, artificial neural network (ANN) or deep neural network (DNN) model, support vector machine (SVM) and its variation support vector regression (SVR) model as well as random forest (RF), were among most frequently used.

This paper provides a short overview of good practice examples in the mentioned field, summarising and reviewing the most frequently used ML models applied to different wastewater management problems.

Furthermore, knowledge gaps and some recommendations for future studies were included and emphasised.

2. MACHINE LEARNING ALGORITHMS AND THEIR APPLICABILITY DOMAIN IN WWT TECHNOLOGY

ML algorithms are based on input data called "independent variables", which are connected with correlated outputs called "dependent variables". In order to "train" the ML model and enable it to improve with experience, large datasets are usually used (Zhong et al., 2021). After being "trained" algorithms can make certain predictions. In general, when more "training data" is available the more efficient predictions will be generated by the proposed ML model (Zhong et al., 2021). Fig. 1 shows typical ML model workflow, which depicts the data transformation process from raw data to generation of the best possible solution (Sundui et al., 2021). ML algorithm is consisted of three key components: the structure of the algorithm (e.g., RF, DNN); the goal which is aimed to achieve (e.g., prediction accuracy); and the training method to achieve the goal (e.g., stochastic gradient descent) (Montavon, Orr & Müller, 2012). In this work, the focus is put on the mostly applied ML models in WWT systems: ANN and DNN, RF, SVM and SVR.

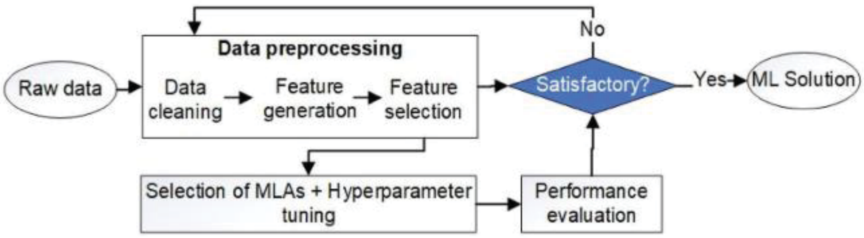

Figure 1: Machine learning model workflow (Sundui et al., 2021)

2.1 ANN models

ANN model consists of set of input, hidden and output processing units (i.e., artificial neurons) which are generating the network of predefined structure. The Type of ANN model is defined by architecture and activation functions (Moreno-Pérez et al., 2018). Single neuron performance is limited to simple calculations. However, their interconnections within hierarchical organizations make them suitable for carrying out complex tasks (El-Din & Smith, 2002). It is very important to test and validate ANN (with new data sets) after training it with previously available data. The performance of the obtained ANN is based on its ability to generalize from the training to the validation data set (El-Din & Smith, 2002). Neural networks with more than two hidden layers are known as deep neural networks (DNN) (Sundui et al., 2021).

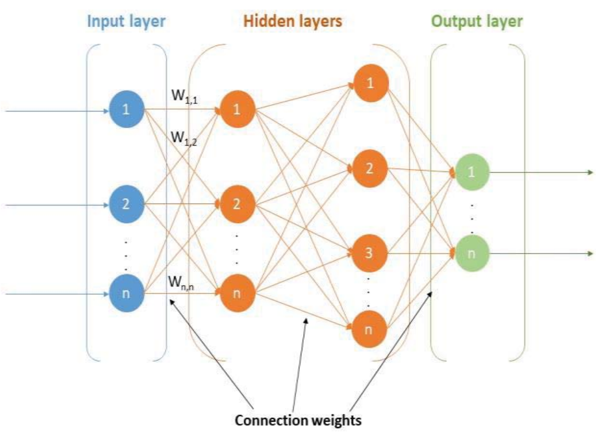

Figure 2: Schematic interpretation of ANN model

2.2 RF model

The Random Forest building process implies generation of new training data sets by applying the bootstrap method on the original data set. Each training data set forms a regression tree which afterwards generates separate prediction. Finally, the general prediction is calculated as the mean value of all separated predictions (Salem et al., 2022). All trees in RF have the same distribution. The generalization error of a RF depends on the strength of the individual trees in the forest and their correlations (Jin et al., 2020). Sometimes another model (e.g. ANN) could be used for the validation of RF model (Wang et al., 2021), however, RF has its advantages over other methods (e.g., importance of every input variable) as the representativeness of training data highly affects model performance (Salem et al., 2022).

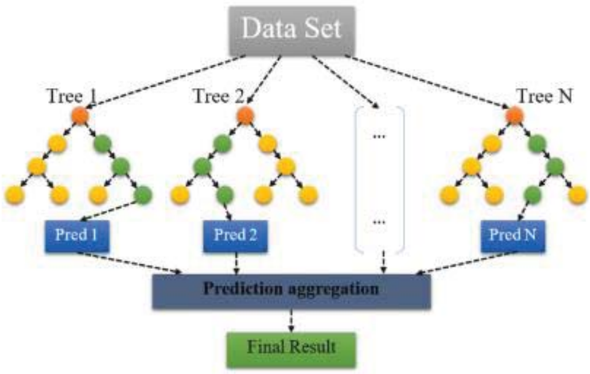

Figure 3: Random forest architecture (Bagherzadeh et al., 2021)

2.3 SVM models

Support vector machine is another powerful algorithm that can be used for addressing a variety of environmental problems (Sundui et al., 2021). While standard, linear SVM presents the fundamental formulation of SVM and is used for classification and regression and, least squares SVM can deal with more different complex problems (Yang, Guergachi & Khan, 2006). General operation of SVM considers observing the end points of data sets, and generating a decision boundary near extreme values, consequently separating classes into two spaces. Basically, the SVM algorithm represents a border between two classes, and the hyperplane that separates the classes is also known as a support vector (Subramaniam & Kaur, 2019). An Important feature of SVM is that it can simultaneously minimize estimation residuals and model dimensions (Foroughi et al., 2020).

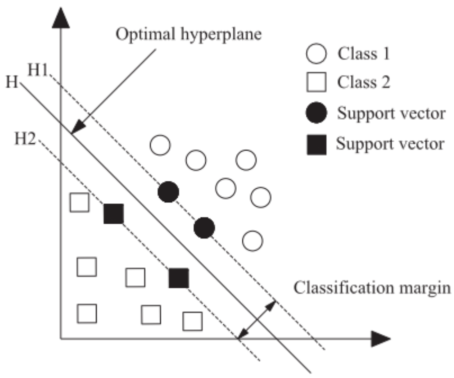

Figure 4: Interpretation of SVM model (Li et al., 2019)

3. EXAMPLES OF GOOD PRACTICE

WWTPs are complex systems which daily receive large quantities of wastewater of dynamic quality. In spite of all variations, WWTPs have to produce an effluent which adheres to the certain standards (Corominas et al., 2018). In order to achieve that, monitoring and control of WWTP performance must be maintained on a high level. In general, as stated by Corominas et al. (2018) there are several driving forces that triggered an intensified development of computer-based approaches for transformation of data into knowledge, namely: control implementation (in order to ensure stability of the process and optimize resource utilization, including energy and chemical consumption) and transformation of data graveyards into data mines (great number of data generated within WWTPs calls for efficient and advanced software processing and transformation into valuable information). Utilization of ML models for mentioned and similar purposes within wastewater treatment management, has especially rapidly increased during the last two decades, which can be seen from the number of published articles within the mentioned field of research. Several examples of good practices are summarized in Table 1.

Table 1: Examples of different ML models utilization within WWT systems

Algorithm	Application	Reference
ANN and DNN	Improving effluent quality control in WWTP (as validation of RF model used for the same purpose) Generation of energy cost model in WWTPs Prediction of breakthrough curves in adsorption study Modelling and optimization of the extraction process Support of modelling arsenic removal by adsorption process Wastewater inflow prediction Optimization of coagulant dosage by modelling jar-test experiments Prediction of ciprofloxacin adsorption Estimation of phosphorus reduction Generation of support sensors Development of software sensors Optimization of naproxen adsorption by biochar	(Wang et al., 2021) (Torregrossa et al., 2018) (Moreno-Pérez et al., 2018) (Genuino et al., 2017) (Rodríguez-Romero et al., 2020) (El-Din and Smith, 2002) (Haghiri, Daghighi & Moharramzadeh, 2018) (Salawu, Han & Adeleye, 2022) (Kumar & Deswal, 2020) (Dürrenmatt & Gujer, 2012) (Bhattacharya et al., 2021)
RF	Improving effluent quality control in WWTP Generation of energy cost model in WWTPs Prediction of phosphorus content in hydrochar Modelling and evaluation of the performance of a full-scale subsurface constructed wetland plant (prediction of pollutants removal) Wastewater inflow prediction Estimation of phosphorus reduction Monitoring of odor in WWTPs Ozone-membrane process optimization Development of software sensors	(Wang et al., 2021) (Torregrossa et al., 2018) (Djandja et al., 2022) (Salem et al., 2022) (Zhou et al., 2019a) (Kumar & Deswal, 2020) (Cangialosi, Bruno & De Santis, 2021) (Mousavi et al., 2022) (Dürrenmatt & Gujer, 2012)
SVM and SVR	Prediction of an adsorption performance Performance prediction of biological WWTP Nitrogen removal process modelling Optimization and modelling of tetracycline removal from wastewater Wastewater inflow prediction Optimization of flocculation conditions	(Li et al., 2019) (Manu & Thalla, 2017) (Yang, Guergachi & Khan, 2006) (Foroughi et al., 2020) (Szelag et al., 2017) (Li, Hu & Wang, 2021)

Herein, ML utilization within WWT sector will be overviewed through 3 basic groups of application, i.e., prediction, monitoring and optimization (process optimization and optimization of resource usage).

3.1 Prediction models

Back in 2002 ANN were used for the prediction of inflow wastewater (El-Din & Smith, 2002), and after several years another ML algorithms were applied for the same purpose, including SVM and RF in 2017 (Szelag et al., 2017) and RF in 2019 (Zhou et al., 2019a). In the first case, the ANN model was built in order to predict the quantity of wastewater influent in case of storms as the maximum capacity of WWTP can be achieved only in cases when there is a warning on incoming increased flows. The model was built on the rainfall data collected from 8 gauges from the Edmonton city drainage area. The modelling was based on a feed-forward neural network with a back-propagation training algorithm. This model has an advantage of insensitivity to the selection of the value of the learning and momentum factors - however, it faces challenges in terms of determination of the appropriate number of training cycles. The Obtained model was proven to be valuable for the prediction of quantity of WWTP influent and can be integrated into a real-time control system, where complete pollution minimization from WWT systems could be set as an ultimate goal (El-Din & Smith, 2002). Beside feed forward back propagation neural network, other ANN training algorithms could be used, such as feed forward back propagation neural network with distributed time delay, cascade forward neural network and elman neural network, which were tested and compared in terms of their viability to predict asymmetric breakthrough curves obtained from the multi-component heavy metal ions adsorption on a biochar (Moreno-Pérez et al., 2018). Within their research, Szelag et al. (2017) additionally highlighted the importance and advantage of the application of different modelling methods instead of restricting the research on only one. SVM, RF, k-nearest neighbour and of Kernel regression methods were used and tested for the prediction of sewage inflow into the sewage treatment plant. Among tested methods, SVM was superior in 9 out of 12 cases, having the smallest prediction errors. On the contrary, RF provided the best fit and the smallest error in case of two input and one explanatory variable (Szelag et al., 2017). Single technology performance could also be predicted by mentioned models. For instance, ML algorithms were efficient in prediction of an adsorption performance (Li et al., 2019), performance of full-scale constructed wetlands (Salem et al., 2022) or uncertain performances of the biological wastewater treatment process (Sundui et al., 2021).

3.2 Monitoring models

ML can be used for regular monitoring within WWTP or they can be used for detecting different anomalies, unexpected working conditions or contamination events. For instance, within anomaly detection, new observations were compared with learned data distribution in order to identify statistically improbable deviations (Zhong et al., 2021). Timely detected anomalies can help to avoid irregularities or unreliable operations (Zhong et al., 2021). Zhou et al. (2019b) used deep learning models in order to accurately locate pipe bursts in water distribution networks, which was important as that information could help efficient and time-relevant repair of the pipes and restoration of water supply (Zhou et al., 2019b). On the other hand, Cangialosi, Bruno & De Santis, (2021) tested ML algorithms, namely RF and ANN for fenceline monitoring of odor classes and concentrations at the WWTP. By utilization of instrumental odor monitoring systems and application of ML models, the most important sources of odor could be identified and their concentrations could be compared to permissible limits stated in the regulative (Cangialosi, Bruno & De Santis, 2021).

3.3 Optimization models

3.3.1 Process optimization

Different process optimizations could be achieved by application of ML algorithms. Optimization of different technologies performances and removal of different pollutants could be done considering the fact that the main goal is to achieve the highest technology performances by minimal time and resource consumption. Bhattacharya et al. (2021) conducted a study in which the ANN model was compared with a response surface methodology (RSM) in order to optimize naproxen removal by activated rice straw biochar. The input variables (time, adsorbent dosage and solution pH) were the same for both models ANN and RSM and the output was the percentage of naproxen removal. Both models showed high performances, however, ANN had a higher degree of correlation and could be declared as a better fitting model for the described purpose. Besides providing an optimal solution, the significance of parameters interactions was provided. The results proved that the application of computer-based technologies such as RSM and ANN was justified in the field of adsorption process optimization (Bhattacharya et al., 2021). There are other studies which confirmed a superior performance of ML models (such as SVM) in comparison to RSM (Foroughi et al., 2020). Although both approaches gives invaluable results in the field of wastewater management optimization, a slight advantage of ML approaches probably originates from the fact that ML present more modern multivariate mathematics able to solve and models complex, non-linear problems, while RSM's quantitative analysis are based on the more traditional mathematical models (e.g., linear, polynomial) and validation analysis were done by several statistical tools (Foroughi et al., 2020).

Another commonly used wastewater treatment technology is coagulation-flocculation. However, its performance depends on different factors such as wastewater and flocculants properties etc. (Li, Hu &

Wang, 2021). In the study of Li, Hu & Wang (2021), the optimization of hydraulic conditions was conducted through 16 experiments consisting of different combinations of hydraulic conditions and two ML methods were applied for its evaluation. Support vector regression was superior in comparison to Gaussian process regression for the optimization of flocculation conditions in deinking WWT. Finally, determination of optimal hydraulic conditions (high and low-stirring speed and temperature) led to the improvement of the flocculation process (Li, Hu & Wang, 2021).

3.3.2 Optimization of resource usage

In today's rapidly developing world, resource depletion represents one of the most significant problem society is facing. Different resources can be used in WWT systems, from a wide variety of chemicals to energy consumption. Hence, utilizing computer-based techniques which can lead to optimization and savings are highly valued. NN and RF algorithms were recently used for the generation of energy cost models in WWTPs (Torregrossa et al., 2018). Until now, energy cost models have usually been generated using exponential, logarithmic or linear functions that were useful, but can sometimes struggle to accurately operate with complex data sets. In order to make progress Torregrossa et al. (2018) investigated utilization of machine learning approach for this purpose. The most significant variables for modelling were determined using databases from 317 WWTP, while energy price was for the first time used as one of the model parameters. The Study included two-steps methodology, which consisted of the regression model performance identification followed by the determination of parameters importance. In comparison to traditional approaches, the ML model provided better performance and the robustness of the model was guaranteed by the independent tests. On the other hand, the ML methodology required more time and input data than traditional methodologies. Hence, ML is not always the most viable methodology, but in the case of large data sets and unsatisfactory results of traditional models the, ML approach could be a cost-beneficial option as a single tool or as a part of integrations with the traditional approaches.

In the context of resource optimization usage, optimization of chemical utilization in some processes can be a part of th3 general process optimization (section 3.3.1). For instance, when the coagulant dosage is optimized by modelling the jar-test experiments, the excessive usage of the coagulant was avoided. The Mentioned optimization avoids additional-costs or excessive sediment in filtrate, while maintaining the good quality of the coagulant, and performance. The driving force for ANN application was to accurately and time-efficiently overcome limitations of jar-test experiments caused by changing characteristics of influent (Haghiri, Daghighi & Moharramzadeh, 2018). Djandja et al. (2022) highlighted that traditional optimization can be a time-consuming process in which a lot of resources can be wasted for the conduction of an excessive number of trial experiments. Hence, in their study, Djandja et al. (2022) used a RF model to generate a valuable predictor of phosphorus content in hydrochar, which could guide the production of sewage sludge based hydrochar with the desired content, without a conduction of large number of experiments, as the model was based on data collected from available literature (Djandja et al., 2022).

4. KNOWLEDGE GAPS AND FUTURE PERSPECTIVES

Generally, ML approach is a beneficial tool for processing large amounts of complex data, which might be insufficiently understood and interpreted by traditional statistical approaches. ML is a time and cost-beneficial technique, however, it is in the early stages of application in environmental science and engineering field. Lack of knowledge about its proper employment might lead to incorrect applications of ML algorithms to certain data sets (Zhong et al., 2021). Hence, those and similar problems that might occur if ML is used inadequately should be considered before the actual application of the ML. More articles which include several algorithms utilizations and viability comparison for the same purpose should be included, if possible. That way, the most appropriate model with the highest performance could be chosen. Furthermore, comparison with traditional models might give an additional justification of ML application in further studies. Currently, there are not many articles which included this aspect. In the field of resource management, literature is sparse with research which include energy cost optimization, which could be characterized as one of the most influential parts of cost-benefit analysis of the wastewater treatment. Additionally, there is an increment in the application of circular-economy principles within adsorption technology, where different waste materials could be used as starting materials for the adsorbent production. This way self-life of a material is prolonged and less raw materials were used. In order to additionally prevent the excessive use of raw materials, different traditional tools have been used, such as response surface methodology. Little is known about the utilization of ML algorithms as advantageous and advanced tools for adsorbent production process optimization. This can be valuable as the advanced ML approach might obtain more precise/trustworthy results in cases where complex and large data sets should be processed or when the traditional approach does not give a satisfactory response. A Similar approach might be investigated for optimizing other material production processes. Hence, those knowledge gaps could potentially be a part of some future research.

5. CONCLUSIONS

ML approach has been increasingly valued in the wastewater treatment sector as it provides a viable, flexible and high performing tool for optimization, prediction, monitoring and other enhancements of wastewater quality management. Its further implementation in environmental engineering and the complex wastewater technology sector might lead to a further decrease in resource depletion, energy and time consumption, as well as to th3 development of real-time control systems and a consequently timely reaction on extreme conditions such as accidental situations which can lead to an occurrence of higher pollutant concentrations in wastewater treatment plant influents etc. The Overall cost and environmental footprint of WWT systems could be optimized by application of ML models, which highly justifies further research in this direction.

6. ACKNOWLEDGEMENTS

The research has been supported by the Ministry of Education, Science and Technological Development, Republic of Serbia, through the project number 451-03-68/2022-14/200156: “Innovative scientific and artistic research from the FTS (activity) domain”.

7. REFERENCES

Bagherzadeh, F., Mehrani, M. J., Basirifard, M. & Roostaei, J. (2021) Comparative study on total nitrogen prediction in wastewater treatment plant and effect of various feature selection methods on machine learning algorithms performance. *Journal of Water Process Engineering*. 41, 102033. Available from: doi: 10.1016/j.jwpe.2021.102033

Bhattacharya, S., Das, P., Bhowal, A. & Saha, A. (2021) Thermal, Chemical and ultrasonic assisted synthesis of carbonized Biochar and its application for reducing Naproxen: Batch and Fixed bed study and subsequent optimization with response surface methodology (RSM) and artificial neural network (ANN). *Surfaces and Interfaces*. 26, 101378. Available from: doi: 10.1016/j.surfin.2021.101378

Cangialosi, F., Bruno, E. & De Santis, G. (2021) Application of machine learning for fenceline monitoring of odor classes and concentrations at a wastewater treatment plant. *Sensors*. 21 (14), 4716. Available from: doi: 10.3390/s21144716

Corominas, L., Garrido-Baserba, M., Villez, K., Olsson, G., Cortés, U. & Poch, M. (2018) Transforming data into knowledge for improved wastewater treatment operation: A critical review of techniques. *Environmental Modelling & Software*. 106, 89–103. Available from: doi: 10.1016/j.envsoft.2017.11.023

Djandja, O. S., Salami, A. A., Wang, Z. C., Duo, J., Yin, L. X. & Duan, P. G. (2022) Random forest-based modeling for insights on phosphorus content in hydrochar produced from hydrothermal carbonization of sewage sludge. *Energy.* 245, 123295. Available from: doi: 10.1016/j.energy.2022.123295

Dürrenmatt, D. J. Ô. & Gujer, W. (2012) Data-driven modeling approaches to support wastewater treatment plant operation *Environmental Modelling & Software*. 30, 47–56. Available from: doi: 10.1016/j.envsoft.2011.11.007

El-Din, A. G. & Smith, D.W. (2002) A neural network model to predict the wastewater inflow incorporating rainfall events. *Water Research*. 36 (5), 1115–1126. Available from: doi: 10.1016/S0043-1354(01)00287-1

Foroughi, M., Rahmani, A. R., Asgari, G., Nematollahi, D., Yetilmezsoy, K. & Samarghandi, M. R. (2020) Optimization and Modeling of Tetracycline Removal from Wastewater by Three-Dimensional Electrochemical System: Application of Response Surface Methodology and Least Squares Support Vector

Machine. *Environmental Modelling & Assessment*. 25, 327–341. Available from: doi: 10.1007/s10666-019-09675-9

Genuino, D. A. D., Bataller, B. G., Capareda, S. C. & De Luna, M. D. G. (2017) Application of artificial neural network in the modeling and optimization of humic acid extraction from municipal solid waste biochar. Journal of Environmental Chemical Engineering. 5 (4), 4101–4107. Available from: doi: 10.1016/j.jece.2017.07.071

Haghiri, S., Daghighi, A. & Moharramzadeh, S. (2018) Optimum coagulant forecasting by modeling jar test experiments using ANNs. *Drinking Water Engineering and Science*. 11, 1–8. Available from: doi: 10.5194/dwes-11-1-2018

Jin, Z., Shang, J., Zhu, Q., Ling, C., Xie, W. & Qiang, B. (2020) RFRSF: Employee Turnover Prediction Based on Random Forests and Survival Analysis. In: *21st International Conference – Web Information Systems Engineering – Proceedings Part II, WISE 2020, 20-24 October 2020, Amsterdam, Netherlands.* New York, Association for Computing Machinery. pp. 503–515. Available from: doi: 10.1007/978-3-030-62008-0_35

Kumar, S. & Deswal, S. (2020) Estimation of phosphorus reduction from wastewater by artificial neural network, random forest and M5P model tree approaches. *Pollution*. 6 (2), 417–428. Available from: doi: 10.22059/POLL.2020.293086.717

Li, M., Hu, K. & Wang, J. (2021) Study on optimal conditions of flocculation in deinking wastewater treatment. *Journal of Engineering and Applied Science*. 68, 1–14. Available from: doi: 10.1186/s44147-021-00044-6

Li, M., Wei, D., Liu, T., Liu, Y., Yan, L., Wei, Q., Du, B. & Xu, W. (2019) EDTA functionalized magnetic biochar for Pb(II) removal: Adsorption performance, mechanism and SVM model prediction. *Separation and Purification Technology*. 227, 115696. Available from: doi: 10.1016/j.seppur.2019.115696

Manu, D. S. & Thalla, A. K. (2017) Artificial intelligence models for predicting the performance of biological wastewater treatment plant in the removal of Kjeldahl Nitrogen from wastewater. *Applied Water Science*. 7, 3783–3791. Available from: doi: 10.1007/s13201-017-0526-4

Montavon, G., Orr, G. B. & Müller, K.-R. (2012) *Neural Networks: Tricks of the Trade*. Berlin, Springer. Available from: doi: 10.1007/978-3-642-35289-8

Moreno-Pérez, J., Bonilla-Petriciolet, A., Mendoza-Castillo, D. I., Reynel-Ávila, H. E., Verde-Gómez, Y. & Trejo-Valencia, R. (2018) Artificial neural network-based surrogate modeling of multi-component dynamic adsorption of heavy metals with a biochar. *Journal of Environmental Chemical Engineering*. 6 (4), 5389–5400. Available from: doi: 10.1016/j.jece.2018.08.038

Mousavi, S. Z., Momeni, S. A., Dehdashti, B., Dehdashti, D., Fatehizadeh, A. & Amin, M. M. (2022) Advanced treatment of industrial estate effluent using ozone-membrane processes based on optimized random forest model. *Journal of Water Process Engineering*. 48, 102897. Available from: doi: 10.1016/j.jwpe.2022.102897

Paula, A. J., Ferreira, O. P., Souza Filho, A. G., Filho, F. N., Andrade, C. E. & Faria, A. F. (2022) Machine Learning and Natural Language Processing Enable a Data-Oriented Experimental Design Approach for Producing Biochar and Hydrochar from Biomass. *Chemistry of Materials*. 34 (3), 979–990. Available from: doi: 10.1021/acs.chemmater.1c02961

Rodríguez-Romero, J. A., Mendoza-Castillo, D. I., Reynel-Ávila, H. E., De Haro-Del Rio, D. A., González-Rodríguez, L. M., Bonilla-Petriciolet, A., Duran-Valle, C. J. & Camacho-Aguilar, K. I. (2020) Preparation of a new adsorbent for the removal of arsenic and its simulation with artificial neural network-based adsorption models. *Journal of Environmental Chemical Engineering*. 8 (4), 103928. Available from: doi: 10.1016/j.jece.2020.103928

Salawu, O. A., Han, Z. & Adeleye, A. S. (2022) Shrimp waste-derived porous carbon adsorbent: Performance, mechanism, and application of machine learning. *Journal of Hazardous Materials*. 437, 129266. Available from: doi: 10.1016/j.jhazmat.2022.129266

Salem, M., EL-Sayed Gabr, M., Mossad, M. & Mahanna, H. (2022) Random Forest modelling and evaluation of the performance of a full-scale subsurface constructed wetland plant in Egypt. *Ain Shams Engineering Journal*. 13 (6), 101778. Available from: doi: 10.1016/j.asej.2022.101778

Subramaniam, P. & Kaur, M. J. (2019) Review of Security in Mobile Edge Computing with Deep Learning. In: *2019 Advances in Science and Engineering Technology International Conferences, ASET 2019, 26 March – 10 April 2019, Dubai, United Arab Emirates*. Piscataway, IEEE. Available from: doi: 10.1109/ICASET.2019.8714349

Sundui, B., Ramirez Calderon, O. A., Abdeldayem, O. M., Lázaro-Gil, J., Rene, E. R. & Sambuu, U. (2021) Applications of machine learning algorithms for biological wastewater treatment: Updates and perspectives. *Clean Technologies and Environmental Policy*. 23, 127–143. Available from: doi: 10.1007/s10098-020-01993-x

Szelag, B., Bartkiewicz, L., Studziński, J. & Barbusiński, K. (2017) Evaluation of the impact of explanatory variables on the accuracy of prediction of daily inflow to the sewage treatment plant by selected models nonlinear. *Archives of Environmental Protection*. 43 (3), 74–81. Available from: doi: 10.1515/aep-2017-0030

Torregrossa, D., Leopold, U., Hernández-Sancho, F. & Hansen, J. (2018) Machine learning for energy cost modelling in wastewater treatment plants. *Journal of Environmental Management*. 223, 1061–1067. Available from: doi: 10.1016/j.jenvman.2018.06.092

Wang, D., Thunéll, S., Lindberg, U., Jiang, L., Trygg, J., Tysklind, M. & Souihi, N. (2021) A machine learning framework to improve effluent quality control in wastewater treatment plants. *Science of The Total Environment*. 784, 147138. Available from: doi: 10.1016/j.scitotenv.2021.147138

Yang, Y. H., Guergachi, A. & Khan, G. (2006) Support Vector Machines for Environmental Informatics: Application to Modelling the Nitrogen Removal Processes in Wastewater Treatment Systems. *Journal of Environmental Informatics*. 7 (1), 14–23. Available from: doi: 10.3808/jei.200600063

Zhong, S., Zhang, K., Bagheri, M., Burken, J. G., Gu, A., Li, B., Ma, X., Marrone, B. L., Ren, Z. J., Schrier, J., Shi, W., Tan, H., Wang, T., Wang, X., Wong, B. M., Xiao, X., Yu, X., Zhu, J. J. & Zhang, H. (2021) Machine Learning: New Ideas and Tools in Environmental Science and Engineering. *Environmental Science & Technology*. 55 (19), 12741–12754. Available from: doi: 10.1021/acs.est.1c01339

Zhou, P., Li, Z., Snowling, S., Baetz, B. W., Na, D. & Boyd, G. (2019a) A random forest model for inflow prediction at wastewater treatment plants. *Stochastic Environmental Research and Risk Assessment*. 33, 1781–1792. Available from: doi: 10.1007/s00477-019-01732-9

Zhou, X., Tang, Z., Xu, W., Meng, F., Chu, X., Xin, K. & Fu, G. (2019b) Deep learning identifies accurate burst locations in water distribution networks. *Water Research*. 166, 115058. Available from: doi: 10.1016/j.watres.2019.115058