International Journal of Current Advanced Research
ISSN: O: 2319-6475, ISSN: P: 2319-6505, Impact Factor: SJIF: 5.995
Available Online at www.journalijcar.org
Volume 6; Issue 12; December 2017; Page No. 8523-8526
DOI: http://dx.doi.org/10.24327/ijcar.2017.8526.1378

Research Article

OPTIMAL AREA PREDICTION FOR WASTE WATER DISCHARGE BY EMPLOYING SUPERVISED LEARNING

Brintha Malar C[1] and Akilandeswari S[2]

[1]Department of Physics, Manomaniam Sundaranar University, Tirunelveli, India
[2]Department of Physics, Annamalai University, Chidambaram, India

A R T I C L E I N F O

Article History:
Received 6th September, 2017
Received in revised form 25th October, 2017
Accepted 4th November, 2017
Published online 28th December, 2017

Key words:
Waste water discharge, pollution control, supervised learning.

A B S T R A C T

Water is the lifeblood of all living beings and hence, this natural resource has to be conserved with all possible efforts. However, due to the increasing trend of urbanization and industrialization, several wastes are produced and are mixed up with water. This causes severe health and sanitation hazard, which should be dealt on time. To regulate this issue, the pollution control board has fixed some standard quality metrics, with which the quality of water can be measured. The pollution control board presets the preferable discharge areas of waste water depending on the quality of the waste water. With this knowledge, this work presents an automatic optimal area prediction system for waste water disposal by employing Extreme Learning Machine (ELM). The ELM is trained with the standards set by the pollution control board and the three classes considered are inland surface water, irrigation and marine coastal areas. The performance of ELM is compared with other classifiers and ELM proves its efficiency in terms of accuracy, sensitivity and specificity rates.

INTRODUCTION

Water is the quintessential component and is intertwined with human lives. Life without water is absolutely impossible, however the source of water is restricted. For instance, the primary source of water is rain and other reliable sources of water are rivers, lakes, bores and wells. Due to the skyrocketing population rise and rapid urbanization, most of the water resources are being deteriorated. Additionally, most of the water resources are polluted by some means. It is stated that the water pollution contributes the maximum in regard to worldwide death rates. A report claims that people of India depart their life everyday, on account of water pollution [1].

Nature has gifted the earth with abundant natural resources. However, due to the urbanization and industrialization, the water resources are abruptly polluted by industrial wastes released by textile, paper industry, health care units, chemical industries and so on. As these wastes are mixed up with the good water, the quality of the water gets degraded and becomes unsuitable for consumption. In order to put the seal onto the deteriorating quality of water, both the central and the state government formed a set of regulations by establishing a separate board in the year 1974. The goal of these government bodies is to protect the water from pollution.

Corresponding author: **Brintha Malar C**
Department of Physics, Manomaniam Sundaranar University, Tirunelveli, India

The pollution control bodies have regulated certain rules, which indicate the standard policies for water discharge. By following these standards, the extent of water pollution can be controlled. This can be attained by aerobic or anaerobic processes. Anaerobic process is proven to be perfectly suitable for industries that generate more wastes. The discharged waste water is processed in the laboratory with the help of anaerobic reactors. Some of the popular anaerobic reactors are Anaerobic Contact (AC), Fluidized Bed Reactor (FBR) and Up-flow Anaerobic Filter (UAF). The standard metrics to test the nature of the waster water are Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD), phosphorous, nitrogen and so on.

Due to the advancement of technologies, several computing techniques are proposed in the literature to predict the quality of water. Hence, this article intends to present a predictive algorithm which can categorise the waste water based on their quality and suggests about the preferable discharging area. The entire work of the research is classified into pre-processing, attribute selection, feature extraction and classification. The pre-processing phase processes the input data, so as to make the input suitable for the forthcoming phase. As the input data contains numerous attributes, it is necessary to select the preferable attributes and is achieved in the attribute selection phase. The potential features are then extracted from the input data. Finally, the classification phase can determine the quality of waste water and preferable discharging areas. The highlights of this work are as follows.

- The pre-processing activity attempts to fill the columns with null value and normalizes the values available.
- The attribute selection process is necessary to choose the important attributes for extracting the features.
- The statistical features are then extracted from the data, which are sufficient to distinguish between the categories.
- Finally, Extreme Learning Machine (ELM) classifier is employed to determine the preferable discharging areas.

The remainder of this article is organized as follows. Section 2 presents the related review of literature with respect to waste water treatment and discharge. The proposed technique is presented in section 3 and the performance of the proposed approach is analysed in section 4. The conclusions of this work are drawn in section 5.

Related Work

This section attempts to review the related literature with respect to waste water treatment and the corresponding discharge areas.

In [2], the performance of a waste water treatment plant is predicted by means of Artificial Neural Networks (ANN). The parameters being used by this work are Biochemical Oxygen Demand (BOD) and Suspended Solids (SS). The relationship between the data is studied by the ANN, which makes it possible to predict the performance of waste water treatment plant. A three layer ANN is proposed for predicting the Chemical Oxygen Demand Removal Efficiency (CODRE) in the cotton textile waste water. The training process of this work is attained by means of Back Propagation (BP) training in association with Principal Component Analysis (PCA) [3]. The performance of ANN is studied and compared.

In [4], the waste water treatment process of paper mill is studied by means of neural networks. The relationship between the entities of the data is studied by means of multilayer back propagation neural networks. This system has shown better prediction rates. A hypersaline oily waste water is processed by ANN by means of Feed Forward Neural Network (FFNN). The neural network is trained by back propagation algorithm and the parameters being used are COD, Total Organic Carbon (TOC) and concentrations of oil and grease [5].

In [6], a ANN based dye containing waste water treatment is proposed. This neural network works in association with Wiener-Laguerre model. The parameters being selected by this work are COD, Mixed Liquor Volatile Suspended Solids (MLVSS) and reaction time. The parameters are modified by Levenberg-Marquardt (LM) algorithm.

Marine oily wastewater being polluted by polycyclic aromatic hydrocarbon, which is called naphthalene is treated by ANN in [7]. The parameters of this system include fluence rate, salinity, temperature, initial concentration and reaction time. The ANN is configured by feed-forward based LM algorithm. In [8], machine learning algorithms such as Support Vector Regression (SVR) and Regression Trees (RT) are employed to predict the quality of water by the indicators. It is shown that SVR performs better than RT.

Motivated by these works, the proposed approach aims to present a supervised learning approach to predict the water quality and to make a decision about the discharge areas. ELM is employed as a classifier, which is given knowledge about the general standards released by the pollution control board. From the knowledge gained during the training process, the ELM can make decision about the discharge areas.

Proposed Approach

The proposed supervised learning based approach aims to provide the basic knowledge of the general standards for the water discharge which contains about 31 parameters. Initially, the input data is pre-processed and normalized. The statistical features are extracted from the pre-processed data. Finally, during the process of classification, the classifier makes a decision about the preferable discharge areas. The overall flow of the work is depicted in figure 1. The detailed explanation of all the phases in the proposed approach is presented as follows.

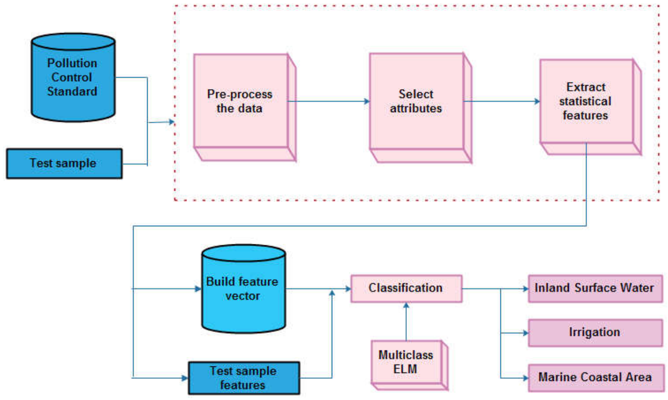

Fig 1 Overall flow of the proposed approach

The entire classification work is classified into two phases, which are training and testing. In the training phase, the classifier is trained with the standard range of values for several parameters. This way of knowledge gaining process helps the classifier to distinguish between the values. In the testing phase, the classifier extracts the features of the testing sample and compares it against the train feature vector. Finally, the classifier suggests the better discharging area, as per the regulation of the pollution control board.

Data Pre-processing

The input database of this work contains nearly thirty one attributes and the range of these attributes is not standard. This data pre-processes the input data by performing the autofill operation. For instance, certain columns in the dataset may not contain values and the pre-processing step attempts to add zero in that place. For the better execution of any algorithm, it is good to avoid empty fields. Some of the important attributes being considered by this work are Total Suspended Solids (TSS), pH, BOD, temperature, COD, phenolic compounds, Fluoride (F), Sulphide (S), pesticide, detergents and so on.

Attribute Selection

The pre-processed data is then passed to the next phase called attribute selection. This phase makes the test sample data to comply with the training data. For instance, the test sample data may not contain all the attributes being present in the train dataset. Hence, this operation considers only the attributes of the test sample and during the process of comparison, all the attributes that are not the part of the train dataset are modified as zero. By this way, the attribute selection process work and is performed only during the testing process.

Statistical Feature Extraction

The classifier is in need of some characteristic features to classify between the available classes. Hence, the classifier is trained with some features, which are sufficient for effective differentiation. The feature vector is formed by means of the extracted features for each record and the feature vectors are saved for future reference. As far as this work is concerned, all the values are numerical and the range of values is limited. Taking this into account and to reduce the time and computational complexity, this work extracts two different features such as mean and standard deviation. These features are computed by the following formulae.

$$\mathrm{M} = \frac{1}{n}\sum_{i=1}^{n} D_i \quad (1)$$

$$SD = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}|D_i - \mathrm{M}|^2} \quad (2)$$

In the above equation, D is the input data and n is the total number of entities being present in the input data. The mean and standard deviation of the train data are computed and the feature vector is formed as follows.

$$fv(TD) = \{|FV_i(f(M, SD)|\}; i = 1,2, \dots, n \quad (3)$$

In equation 3, TD is the training data and the feature vector is computed for all the data items in the train dataset, which comprises mean and standard deviation. The classifier is trained with the so formed feature vector, which makes it viable for suggesting the discharge areas of the waste water, by taking the pollutants into account.

ELM Classification

Initially, the ELM is trained with the feature vector, which is computed in the last phase. The ELM learns about the nature of the data and by applying the knowledge obtained in the training phase, ELM can effectively distinguish between the classes. The main reasons for the choice of ELM are its better and promising performance [9].

Consider A training samples, which is denoted as (m_i, j_i), where $m_i = [m_1, m_2, \dots, m_k]^T \in ID^k$. $j_i = [j_1, j_2, \dots, j_l]^T \in D^l$ indicates the class labels and l is the total number of classes. To achieve classification a Single hidden Layer Feed-Forward Neural Network (SLFN) is constructed by means of an activation function $atv(x)$ and NU neurons, is represented as follows.

$$\sum_{i=1}^{NU} \alpha_i qs(wt_i . m_i + pk_i) = rs_i; j = 1,2, \dots, n \quad (4)$$

In the above equation, wt_i is the weight and is represented by $wt_i = [wt_{i1}, wt_{i2}, \dots, wt_{in}]^T$. The weight vector interconnects the i^{th} hidden neuron with the input neurons, where $i = \left[i_1, i_2, \dots, i_p\right]^T$. The bias of the i^{th} hidden neuron is represented by $bias_i$. The weights and the bias of the ELM can be fixed randomly. The SLFN is denoted as

$$\sum_{i=1}^{NU} \alpha_i atv(wt_i . m_i + bias_i) = n_i; i = 1,2, \dots, n \quad (5)$$

Consider H_{DL} as the ELM's hidden layer output matrix, which means that the i^{th} column of H_{DL} represents the i^{th} hidden neurons output vector by taking the inputs $u_{j1}, u_{j2}, \dots, u_{jn}$ into account.

$$HDL = \begin{bmatrix} atv(wt_1 . m_1 + bias_i) & \dots & atv(wt_{NU} . m_1 + bias_G) \\ \vdots & \vdots & \vdots \\ atv(wt_1 . m_n + bias_i) & \dots & atv(wt_{NU} . m_n + bias_G) \end{bmatrix} \quad (6)$$

$$\alpha = \begin{bmatrix} \alpha_1^T \\ \vdots \\ \alpha_G^T \end{bmatrix} \quad (7)$$

$$J = \begin{bmatrix} j_1^T \\ \vdots \\ j_n^T \end{bmatrix} \quad (8)$$

These equations are represented in matrix format as follows.

$$HDL\alpha = J \quad (9)$$

The output weights are calculated by norm least-square solution as presented below.

$$\alpha = HDL^{\dagger} J \quad (10)$$

In the above equation, $HDL^{\dagger}$ is the HL's Moore-Penrose generalized inverse. During the training process, the ELM requires the total number of classes tc, activation function $atv(x)$ and total number of neurons NU.

By this way, the ELM is trained and with the acquired knowledge, ELM is capable of distinguishing between the optimal discharge areas by considering the quality of the water. The total number of classes involved in this work is three and they are inland surface water, irrigation and marine coastal areas. On passing the test sample, the best suitable class for the test sample is determined by the ELM. The performance of this work is analysed in the forthcoming section.

RESULTS AND DISCUSSION

This work trains the classifier with the standard data by the pollution control board, which is downloaded from http://www.environmentallawsofindia.com/tolerance-limits-for-trade-effluents.html. This standard contains about thirty one attributes. The quality of the water can be predicted by means of this standard and based on the quality, the water is suggested to get discharged in specific areas. Based on this standard, the ELM is trained and when a test data sample is passed as input, the ELM pre-processes the data, selects the attribute, extracts the feature and compares the test feature vector with the train feature vector. By this way, the ELM determines the best suitable discharging area by taking the quality of water into account.

The performance of the proposed approach is tested in terms of standard performance measures such as accuracy, sensitivity and specificity and is compared against other classifiers such as k Nearest Neighbour (k-NN) and Support Vector Machine (SVM). From the experimental results, it is evident that ELM performs better than SVM. The formulae for computing the accuracy, sensitivity and specificity measures are as follows.

$$Ac = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \times 100 \quad (11)$$

$$Sen = \frac{T_p}{T_p + F_n} \times 100 \quad (12)$$

$$Spc = \frac{T_n}{F_p + T_n} \times 100 \quad (13)$$

Where T_p, T_n, F_p, F_n are True Positive, True Negative, False Positive and False Negative rates respectively. The following graphs exhibit the performance of different classifiers such as k-NN, SVM and ELM.

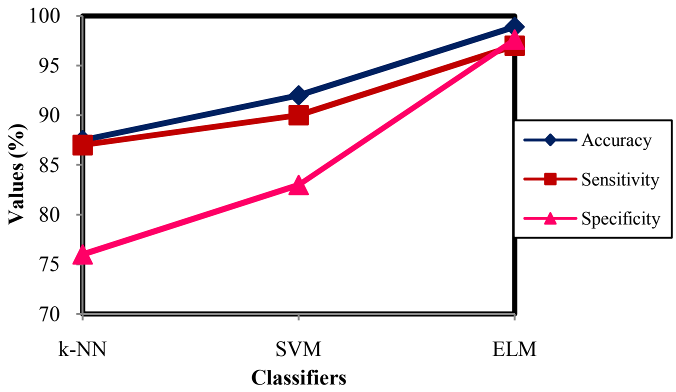

Fig 2 Performance analysis w.r.t classifiers

On analysing the results, it is obvious that the ELM outperforms the other two classifiers in terms of all the three performance measures. The maximum accuracy rate is 98.9 percent and is shown by ELM and the least accuracy rate is shown by k-NN with 87.5 percent. The sensitivity and specificity rates of any classification algorithm must be greater, as it indicates low false positive and false negative rates. Even in sensitivity and specificity analysis, ELM shows the maximum performance with 97 and 97.6 percent respectively. On the other hand, k-NN classifier shows 87 and 76 percent as sensitivity and specificity. ELM is the best performing classifier and k-NN is the least performing classifier. Hence, the ELM takes the standards into account and thereby suggesting the best area in which the waste water can be discharged.

CONCLUSION

This work presents a supervised learning based system to determine the optimal area for waste water discharge. The quality of the water is checked by means of the standard parameters formed by the pollution control board. By taking the quality of water into account, the classifier ELM suggests the best areas for waste water discharge. The areas or the classification classes being considered by this work are inland surface water, irrigation and marine coastal areas. Based on the quality of the water, ELM chooses the best possible area for waste water discharge. The experimental results of the proposed approach are found to be satisfactory in terms of accuracy, sensitivity and specificity.

References

1. Michael Greenstone, Rema Hanna, "Environmental Regulations, Air and Water Pollution, and Infant Mortality in India", The American Economic Review, Vol. 104, No.10, pp.3038-3072, 2014.
2. Maged M Hamed, Mona G Khalafallah, Ezzat A Hassanien, "Prediction of wastewater treatment plant performance using artificial neural networks", *Environmental Modelling & Software*, Vol.19, No.10, pp: 919-928, 2004.
3. Kaan Yetilmezsoy, Zehra Sapci-Zengin, "Stochastic modeling applications for the prediction of COD removal efficiency of UASB reactors treating diluted real cotton textile wastewater", Stochastic Environmental Research and Risk Assessment, Vol.23, No.1, pp.13-26, 2009.
4. G.M.Zeng, X.S.Qin, L.He, G.H.Huang, H.L.Liu, Y.P.Lin, "A neural network predictive control system for paper mill wastewater treatment", *Engineering Applications of Artificial Intelligence,* Vol.16, No.2, pp.121-129, 2003.
5. Ali Reza Pendashteh, A. Fakhru'l-Razi, Naz Chaibakhsh, Luqman Chuah Abdullah,Sayed Siavash Madaeni, Zurina Zainal Abidin, "Modeling of membrane bioreactor treating hypersaline oily wastewater by artificial neural network", *Journal of Hazardous Materials,* Vol.192, pp.568-575, 2011.
6. Yasaman Sanayei, Naz Chaibakhsh, Ali Chaibakhsh, Ali Reza Pendashteh, Norli Ismail,Tjoon Tow Teng, "Long-Term Prediction of Biological Wastewater Treatment Process Behavior via Wiener-Laguerre Network Model", *International Journal of Chemical Engineering,* Vol. 2014, pp. 1-7, 2014.
7. Liang Jing, Bing Chen, Baiyu Zhang, "Modeling of UV-Induced Photodegradation of Naphthalene in Marine Oily Wastewater by Artificial Neural Networks", Water, Air, & Soil Pollution, vol.225, 2014.
8. Francesco Granata 1, Stefano Papirio, Giovanni Esposito, Rudy Gargano, Giovanni de Marinis, "Machine Learning Algorithms for the Forecasting of Wastewater Quality Indicators", Water, Vol.9, No. 105, pp. 1-12, 2017.
9. Guang-Bin Huang, Hongming Zhou, Xiaojian Ding, and Rui Zhang, Extreme Learning Machine for Regression and Multiclass Classification, IEEE Transactions on systems, Man and Cybernetics - Part B, Vol.42, No.2, pp.513-529, 2012.

How to cite this article:

Brintha Malar C and Akilandeswari S (2017) 'Optimal Area Prediction for Waste Water Discharge by Employing Supervised Learning', *International Journal of Current Advanced Research,* 06(12), pp. 8523-8526.
DOI: http://dx.doi.org/10.24327/ijcar.2017.8526.1378

****