International Journal of Current Research and Modern Education (IJCRME)
Impact Factor: 6.925, ISSN (Online): 2455 - 5428
(www.rdmodernresearch.com) Volume 3, Issue 1, 2018

AUTO-SUGGESTING DISCHARGE LOCATION SYSTEM FOR WASTEWATER TREATMENT

Dr. S. Akilandeswari* & C. Brintha Malar**

* Assistant Professor, Department of Physics, Annamalai University, Chidambaram, Tamilnadu

** Research Scholar, Department of Physics, Manomaniam Sundaranar University, Tirunelveli, Tamilnadu

Cite This Article: Dr. S. Akilandeswari & C. Brintha Malar, "Auto-Suggesting Discharge Location System for Wastewater Treatment", International Journal of Current Research and Modern Education, Volume 3, Issue 1, Page Number 249-254, 2018.

Abstract:

Water is the most basic and primary need of all living organism. It is an absolute need to preserve the available water. Yet, due to the technological advancement, most of the wastages from the industry are released to water sources, which lead to severe pollution. In order to combat with this issue, the government bodies formed a pollution control board. The pollution control board sets some limits for the pollutants or the chemical parameter of the wastewater. This work attempts to propose an auto-suggesting discharge location system for wastewater treatment. The discharge locations being considered by this work are inland surface, irrigation land and marine coastal area. The better discharge locations are suggested by the classifier Support Vector Machine (SVM), which is trained by the statistical features. The performance of the proposed approach is tested against accuracy, sensitivity and specificity rates by varying the kernel functions.

Key Words: Water Pollution, Wastewater Treatment & Classification System

1. Introduction:

The term 'environment' may seem to be a single term but it spans the entire coverage of the earth. This makes sense that it encloses all the entities being present in the earth, which can either be living or non-living organisms. The essence of this term is the relationship between the living and non-living organisms in the super colossal earth. The earth grants the living organisms with all their basic requirements of sustainment. The major components of environment are air, water and soil. These components must be managed so carefully, so as to maintain the balanced ecosystem. In case of any disturbance to the ecosystem, the quality of life of all the living organisms is seriously affected.

Hence, it is quintessential to safeguard the environment for the better lifestyle of organisms. Understanding the necessity and importance, the government has enforced several laws and policies for maintaining a good balance on the environment. Yet, pollution is one of the inevitable threats to the environment. Pollution is defined as the unwanted deterioration of environment by means of harmful stuffs, which are the end products of human activities. The environment can tolerate the mass of harmful stuffs to some extent and when the saturation point is reached, the environment cannot handle the aberration. This is a serious threat to the environment and the organisms that live upon it as well.

The central and the state governments have composed a pollution control board that is responsible for managing the level of polluting activities. The term 'environmental pollution' may refer to different forms of pollutions. However, the most occurring and influenced pollutions are soil, water and air pollution, as most of the industrial wastages mixed up with these natural resources. Realizing the significance of the issue, this work aims to present a small contribution for predicting the feasible place for waste water discharge by taking the waste water discharge standards formed by pollution control board into account.

The research goal of this work is attained by four important phases such as data pre-processing, attribute selection, feature extraction and classification. The pollution control board systematizes the wastewater discharge issue by considering the degree of the chemical components and the preferable range of components for each location is fixed. This work takes the waste water discharge standard as input and decides the best discharge location of the waste water based on the range of pollutants. This is made possible by training the prediction system with the wastewater discharge standards and then the system can predict the best place for wastewater discharge. The training process is done by means of simple statistical features such as mean, standard deviation and variance. As the standards possess limited number of chemical components, simpler statistical features can make the system works accurately. The contributions of this work are as follows.

- ✓ The results of the proposed decision making system are accurate, as the system is trained effectively.
- ✓ As the basic statistical features are utilized for training the decision making system, the computational, time and memory complexities are overthrown.
- ✓ This work employs Support Vector Machine (SVM) as classifier for making decision about the suitable discharge place.

- ✓ The proposed decision making system can be utilized along with wastewater treatment system and the real inputs can be fed into the system for deciding the best suitable discharge place of the wastewater.
- ✓ The performance of the proposed approach is tested in terms of accuracy, sensitivity, specificity and time consumption.

The remainder of this paper is organized as follows. Section 2 presents the review of literature with respect to water pollution. The proposed decision making system is elaborated in section 3. The performance of the proposed work is analysed and discussed in section 4 and the conclusions of the work are presented in section 5.

2. Review of Literature:

This section reviews the related literature with respect to water pollution detection and control.

In [1], the quality of the water is tested by four different indicators such as acidity, total dissolved solids, water transparency and temperature. The thresholds for all these indicators are set as follows, pH is set as 6.5 to 8.5, TDS is set to 1000 RPM, transparency is lesser than 5 and the temperature is set lesser than plus or minus three Celsius than the air temperature. All the sensors are processed by precision test and the average percentage error of the parameters of the sensor is 1.46% pH, 1.09% TDS, 2% turbidity sensor and 0.83% temperature. The quality of water is classified as good, less good and bad.

The source of water pollution is detected by means of Wireless Sensor Networks (WSN) in [2]. The static and dynamic water pollution sources are detected by WSN and the hypothesis testing is performed. The pollution source detection techniques are tested for accuracy. In [3], the quality of water is computed by means of automatic classification approach based on fish liver histopathology. The idea of this work is to process the microscopic images of fish liver for detecting the water pollution.

The proposed approach relies on three important phases such as image pre-processing, feature extraction and classification. The level of water quality is identified by coloured histogram and gabor wavelet. The features of the microscopy fish liver images are extracted by means of Principal Component Analysis (PCA) and the water quality is classified by the classifier SVM. The classification process is attained by training the system with 125 images and 45 images are used for testing. The accuracy of proposed water quality classification is 93.3%.

The environmental monitoring of water pollution based on residual antibiotics is presented in [4]. The aim of this work is to control the antibiotic traces in the water and these specific antibiotic traces are detected by Ultra Violet spectrophotometer and low frequency impedence. The visual absorbance level of the solution is computed by spectrophotometer and the range of capacitance is tracked by impedence analyser. The accuracy rate of this work varies with respect to the kind of antibiotic and its concentration in water. In [5], the current progress of water pollution and pollution source localization with WSN is discussed. Additionally, the issues and challenges associated with the water pollution monitoring are discussed.

The air and water pollution emitted by the Baghdad power plant are evaluated in [6]. The proposed work is tested over twenty one and six sites inside and outside the power plant respectively. The waste water samples are tested and the concentration of metals such as Pb, Ni, Zn and is detected from the waste water. A source of water pollution is localized in three dimensional space by means of sensor networks in [7]. A spatial-temporal Unscented Kalman Filter (UKF) is proposed to localize the water pollution source in terms of time and space. The performances of spatio-temporal and temporal UKF are compared and the performance of spatio-temporal UKF is proven with better performance.

The work presented in [8] claims that the process of data pre-processing has the great influence over the accuracy rate of the classification system. Hence, the paper discusses about the different data pre-processing techniques and the results of different clustering, regression and classification results are presented. A real time water pollution monitoring system is proposed for the lake Toba in [9]. The quality of the water is estimated by means of Dissolved Oxygen Level, pH value, temperature, air humidity and air temperature. The sensors are distributed throughout the river and the water quality is measured finally.

In [10], a water pollution accident source is proposed to be detected by means of optimal search theory. The term “accident source” means that the sudden source of pollution in the river. Whenever the water is polluted, the research data of target distribution and detection function constructs the optimal search model for finding the accident source of water pollution. In [11], a hazard management system for detecting water pollution and radiation dispersion is proposed. This system predicts and makes decision about the water quality.

The pollution in Luton Hoo lake is detected by means of several physical and chemical parameters in [12]. The physical parameters such as temperature, conductivity and turbidity are considered and the chemical parameters like dissolved oxygen, pH and ammonium are considered. The water quality index is calculated for different locations in a lake and the water quality is analysed.

Most of the existing works focus on determining the quality of water by checking the degree of pollution. Motivated by the above presented works, this work aims to find the most suitable location for discharging the waste water based on the degree of impurity. The following section presents the proposed approach in detail.

3. Proposed Automated Water Discharge Location Prediction System:

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 3.1. The detailed explanation of all the phases in the proposed approach is presented as follows.

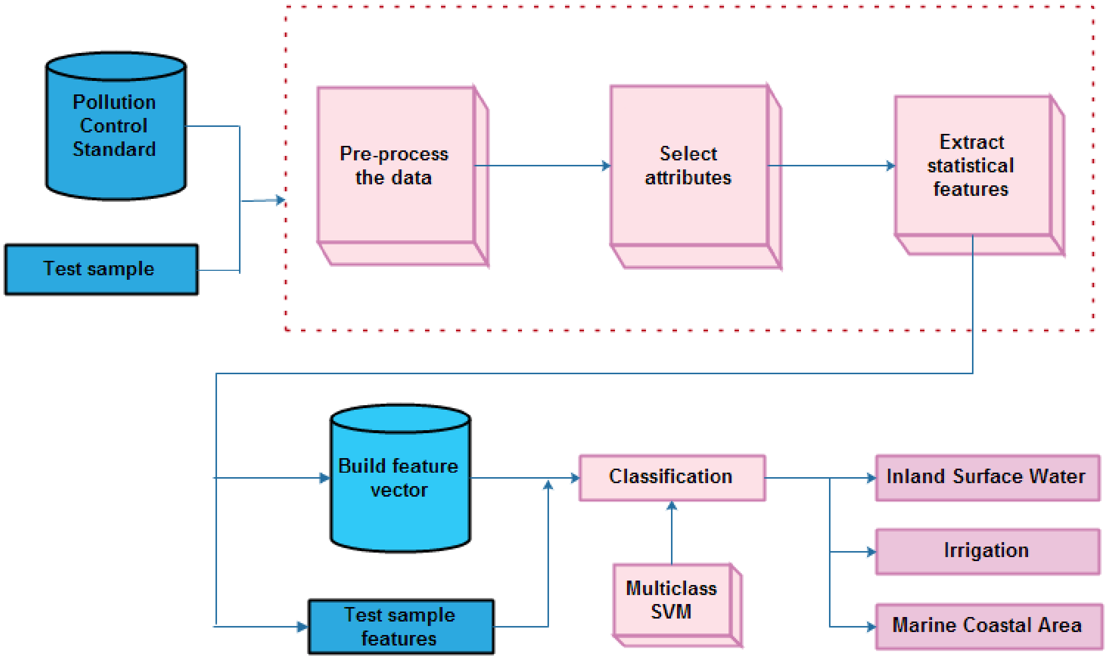


Figure 3.1: Overall flow of the diagram

3.1 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.

3.2 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.

3.3 Statistical Feature Extraction:

As soon as the attributes are selected, the statistical features are extracted from the data. Feature extraction is the heart of the classification system and the efficiency of the classification system depends on the potential of the features being utilized. Hence, the features should be more precise and crispy, such that the classification task can be performed effectively. When the training data is fed into the discharge location classification system, the system extracts the statistical features such as mean, standard deviation and variance of the training data and forms the feature vector. The so formed feature vector is saved in the database for performing future classification tasks. A short feature vector may be inefficient, as the details of the features are insufficient for performing accurate classification. On the other hand, a large feature set involves time, memory and computational complexity.

Taking this into account, the proposed approach attempts to build an efficient and sharp feature vector that can bring in better accuracy with lesser time complexity. The mean, standard deviation and variance are computed as follows.

$$\mathrm{M} = \frac{1}{n}\sum_{i=1}^{n} D_i \tag{1}$$

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

$$V = \frac{\sum D^2}{n} - M^2 \quad (3)$$

All the three features work in a blended fashion and arrives at better feature set and the feature vector is formed by

$$fv(TD) = \{|FV_i(f(M, SD, V)|\}; i = 1,2, \dots, n \quad (4)$$

In the above equations, D_i is the input data, TD is the train data. The overall algorithm of this work is presented as follows.

Algorithm for Auto-suggesting Discharge Location System for Wastewater

// Training
Input: General standards from pollution control board
Output : Knowledge gaining
Begin
Pre-process the general standards by autofill operation;
For all records
do
Extract mean (M), standard deviation (SD) and Variance (V) features;
Construct $fv(TD)$ and store it in the local database;
Feed the knowledge to SVM classifier;
End;
End;
// Testing
Input: Measurement of pollutants in the sample wastewater
Output : Optimal discharge area suggestion
Begin
Pre-process the pollutant list by autofill operation;
For the test sample
do
Extract M, SD, V features;
Construct $fv(TD)$;
Apply SVM classifier to match the test and train samples;
Utilise $Poly\ SVM$, $Sigmoid\ SVM$ and $RBF\ SVM$ for matching;
Analyse the performance;
End;
End;

The computed feature vectors are stored in database for future classification problem. When the training process gets completed, the system is ready for performing the testing phase. In this stage, the known values of the chemical parameters are passed in to the automated discharge location classification system. The proposed work suggests the best possible discharge location after analysing the parameters of the wastewater. The following section presents the SVM classification process.

3.4 SVM Classification:

This work utilizes multiclass SVM for selecting the best possible discharge location of the wastewater after analysing the quality of water. As this work considers three different discharge locations such as irrigation land, inland surface and marine coastal area, multiclass SVM is employed. SVM can be incorporated in two different ways for dealing with a multiclass classification issue. The first way employs multiple binary SVM classifiers and each classifier is trained to handle a single class out of a set of classes. In the second way, multiple classifiers are processed over every pair of classes and the data is assigned to the class with maximal votes. This technique exploits $n(n-1)/2$ classifiers and finally max-vote strategy is followed [13]. This work performs multiclass classification by tackling all the classes at the same instant of time by solving a single objective function.

The classification problem with n different classes is denoted by a single optimization problem and is written as

$$\min_{w,b,\omega} \frac{1}{2} \sum_{y=1}^{n} w_y^p\, w_y + C \sum_{i=1}^{l} \sum_{y \neq s_i} \omega_{i,y} \quad (4)$$

Where $$w_{s_i}^p \rho(x_i) + b_{s_i} \geq w_y^p \rho(x_i) + b_y + 2 - \omega_{i,y};\ \omega_{i,y} \geq 0 \quad (5)$$

Where $i = 1,2, \dots l$ are training samples and $y \in \{1,2, \dots n\}$. The final decision is obtained by the below given equation.

$$dec_{fn} = \max_{y=1,2,..n} (w_y^p \rho(x_i) + b_y) \quad (6)$$

This way of classification conserves more time and is efficient. Besides this, the requirement of support vectors is lesser, when compared to the usage of multiple binary SVMs. Thus, the multiclass SVM can serve its purpose, irrespective of the class count. The classification process is carried out by this way and the performance of the system is analysed as follows.

4. Performance Analysis:

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 SVM is trained and when a test data sample is passed as input, the SVM 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 SVM determines the best suitable discharging area by taking the quality of water into account.

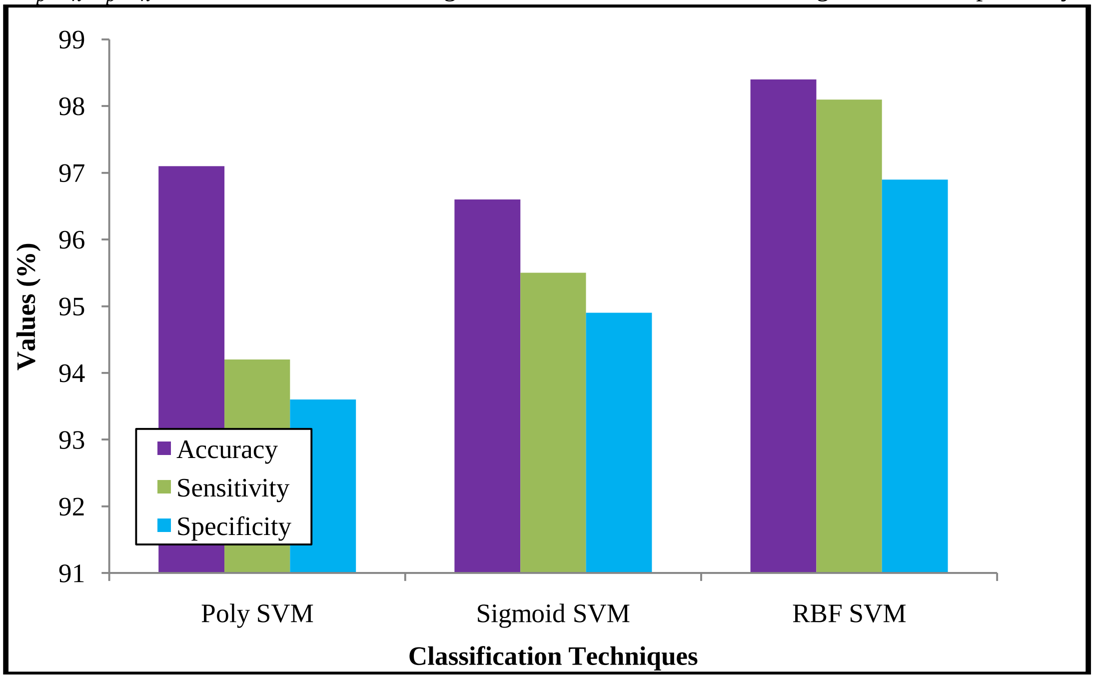
The performance of the proposed approach is tested in terms of standard performance measures such as accuracy, sensitivity and specificity and is compared against different kernels of SVM such as poly, RBF and sigmoid. From the experimental results, it is evident that RBF SVM performs better than the other two. The formulae for computing the accuracy, sensitivity and specificity measures are as follows.

$$Ac = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \times 100 \qquad (7)$$

$$Sen = \frac{T_p}{T_p + F_n} \times 100 \qquad (8)$$

$$Spc = \frac{T_n}{F_p + T_n} \times 100 \qquad (9)$$

Where T_p, T_n, F_p, F_n are True Positive, True Negative, False Positive and False Negative rates respectively.



On analysing the performance of different kernels of SVM, it is found that the RBF kernel of SVM serves well than the other two. The greatest accuracy, sensitivity and specificity values of SVM are proven by RBF SVM. Hence, this work concludes that the performance of RBF SVM is satisfactory in terms of better accuracy, sensitivity and specificity rates.

The time consumption of the proposed approach is compared by varying the kernel functions of SVM. Though the time consumption of all the three kernel functions of SVM are more or less similar to each other, RBF SVM consumes 1742 ms, which is the least and the remaining classification techniques consume 1832 and 1807 ms respectively.

5. Conclusion:

This article presents an automated discharge location suggestion system for wastewater treatment. The quality of water is affected when more harmful substances mix up with the water and the water is not fit for consumption. Additionally, the environment is also polluted by the wastewater. The proposed automated discharge location suggestion system is based on four phases such as data pre-processing, attribute selection, statistical feature extraction and classification. The statistical features such as mean, standard deviation and variance are extracted from the data and it forms the base of the classification task. Multiclass SVM with different kernel functions are utilized for achieving the classification task and this work concludes that RBF SVM performs well than poly SVM and sigmoid SVM. In future, better features are planned to be extracted and the optimal features alone are to be processed by feature selection algorithm.

6. References:

1. Taufiqurrahman; Ni'am Tamami; Dito Adhi Putra; Tri Harsono, "Smart sensor device for detection of water quality as anticipation of disaster environment pollution", International Electronics Symposium, Denpasar, Indonesia, 29-30 Sep, 2016.
2. Xu Luo; Jun Yang; Li Chai, "Water pollution source detection in wireless sensor networks", IEEE International Conference on Information and Automation, Lijiang, China, 8-10 Aug, 2015.
3. Asmaa Hashem Sweidan ; Nashwa El-Bendary ; Aboul Ella Hassanien ; Osman Mohammed Hegazy ; Abd El-karim Mohamed, "Machine Learning based Approach for Water pollution detection via fish liver microscopic images analysis", International Conference on Computer Engineering & Systems, Cairo, Egypt, 22-23 Dec, 2014.
4. M. Soprani ; O. Korostynska ; A. Mason et.al, "Low-Frequency Capacitive Sensing for Environmental Monitoring of Water Pollution with Residual Antibiotics", International Conference on Developments in eSystems Engineering, Liverpool, UK, 31 Aug-2 Sep, 2016.
5. Xu Luo ; Jun Yang, "Problems and challenges in water pollution monitoring and water pollution source localization using sensor networks", Chinese Automation Congress, Jinan, China, 20-22 Oct, 2017.
6. Salah Farhan A. Sharif ; Ghada Mahdi Al-Saadi, "Evaluation of air and water pollution caused by South Baghdad Power Plant South Baghdad Power Plant", International Conference on Environmental Impacts of the Oil and Gas Industries: Kurdistan Region of Iraq as a Case Study (EIOGI), Koya-Erbil, Iraq, 17-19 April, 2017.
7. Zheng Feng ; Jun Yang ; Xu Luo, "A water pollution source localization method in three-dimensional space using sensor networks", IEEE Conference on Industrial Electronics and Applications, Siem Reap, Cambodia, 18-20 June, 2017.
8. Eva C. Serrano Balderas; Laure Berti-Equille; Maria Aurora Armienta Hernandez; Corinne Grac, "Principled Data Preprocessing: Application to Biological Aquatic Indicators of Water Pollution", International Workshop on Database and Expert Systems Applications, Lyon, France, 28-31 Aug, 2017.
9. Romi Fadillah Rahmat ; Athmanathan ; Mohammad Fadly Syahputra ; Maya Silvi Lydia, "Real time monitoring system for water pollution in Lake Toba", International Conference on Informatics and Computing, Mataram, Indonesia, 28-29 Oct, 2016.
10. Ying Liu; Dan Liu; Fan Zhang; Qingxin Zhu, "Application of Optimal Search Theory in the Detection of Water Pollution Accident Source", International Multi-symposiums on Computer and Computational Sciences, Zhejiang, China, 20-24 June, 2006.
11. Anca Daniela Ionita ; Mariana Mocanu, "Modeling framework for hazard management applied to water pollution and radiation dispersion", International Conference on Intelligent Computer Communication and Processing, Cluj-Napoca, Romania, 8-10 Sep, 2016.
12. Tochukwu Kene Anyachebelu; Marc Conrad; David Rawson; Tahmina Ajmal, "Application of water quality index for pollution detection at Luton Hoo Lake", IEEE Sensors, Busan, South Korea, 1-4 Nov, 2015.
13. Hsu, C-W. and Lin, C-J., A comparison of methods for multiclass support vector machines,‖ IEEE Trans. Neural Networks, vol. 13, pp. 415-425, 2002.