International Journal of Scientific Research & Engineering Trends
Volume 6, Issue 6, Nov-Dec-2020, ISSN (Online): 2395-566X

Comparative Study on Environmental Pollution through Software Techniques

Assistant Professor C.Brintha Malar
Department of Physics
Udaya School of Engineering
Nagercoil, India
brinthasivasankar@gmail.com

Associate Professor K.Siva Sankar
Department of IT
Noorul Islam Centre for Higher Education
Nagercoil, India
sivasnakarniu@gmail.com

Abstract- **Water pollution is one of the serious threats to the society, as water is the primary need of every organism thriving on earth. It is necessary to control and detect water pollution by assessing the quality of water. However, the production of wastewater is always there and is inevitable. Hence, it is equally important to treat the wastewater in a better way, such that the environment is not affected. The pollution control board has formulated certain standards, which provides the range of values for each pollutant and the feasible discharge locations. Taking these standards as the input for training the system, this work extracts basic statistical features such as mean, standard deviation, entropy and variance for training the classification system. The ensemble classification is incorporated, which includes k-Nearest Neighbour (k-NN), Support Vector Machine (SVM) and Extreme Learning Machine (ELM). The performance of the proposed approach is evaluated in terms of accuracy, sensitivity and specificity. The results of the proposed approach are found to be satisfactory.**

Keywords **– Classification, Waste water discharge, pollution control, supervised learning, water pollution.**

I.INTRODUCTION

Environmental pollution is one of the serious threats to the society and it means that the environment is made dirty with the pollutants or impurities. The term 'environmental pollution' is generic and it may refer to many kinds of pollution. The prominent forms of pollution are water pollution, air pollution and soil pollution. The following subsections present the summary of these forms of pollution and the pollution depicted in Figure 1.

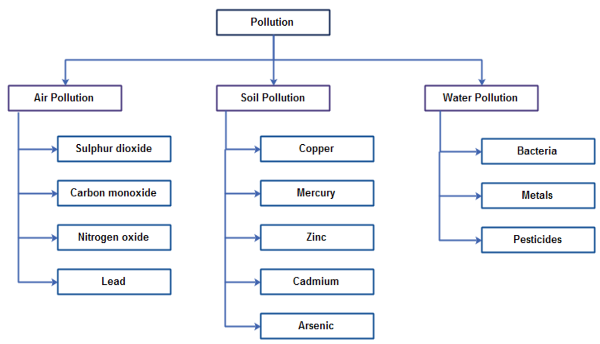

Fig.1. Pollution types with important pollutants.

1. Air pollution

Air pollution is one of the most serious kinds of pollution in the present condition. The major reasons for the cause of air pollution can be the release of smoke from vehicles which constitutes both the vehicles of roadways and railways. Additionally, the smoke is also released by industries and is poisonous. The fresh air is pollutedby means of sulphur dioxide (SO_2), carbon monoxide (CO), nitrogen oxide (NO_X), lead (PB) and so on. Besides this, the people involved in some unfavourable activities that could affect the healthy air. For instance, sometimes people burn the household garbage, tyres and fallen leaves from the trees. All these activities emanate more smoke and seriously affect the atmosphere. Recently, many surveyscite that several diseases of mankind are due to air pollution. The following subsections present some details about the important pollutants of air.

II. LITERATURE REVIEW

The water pollution level is assessed with respect to pyrethroid pesticide and heavy metals over mRibo Nucleic Acid (mRNA) range of Metallothionine and Nile Tilapia in the work of Ghazy H.A. et.al. (2017). The properties of water in different level of heavy metals are tracked and the residues of Iron, Manganese,

Zinc, Copper and Nickel. ,The work proposed by Kar S. et.al. (2016) exploits hyperion data to classify between the river water pollution. The chemical pollutants are found by utilizing hyperspectral images and the heavy metals such as Iron, Arsenic, Cadmium and Lead. Based on the severity of the chemical pollutants, the river water surface is classified based on the SVM classifier. The idea of this work is to utilize hyperspectral images for checking out the quality of water from space.

Rajendran V. et.al. (2018) proposed a quality assessment technique for marine area in the gulf of mannar. This work collects water samples from five different areas before and after the monsoon. The collected samples are tested by means of standard techniques and are found that the microbiological indicators range is greater, such that the water quality is affected.

Li R. et.al. (2016) assess the quality of water in Qu river by means of fuzzy pollution index technique. A fuzzy inference system and the water pollution index are utilized for developing an improved water pollution index. The performance of this work is checked over the river Qu in China. The performance of this work is claimed to be reliable. A simulation model that analyses the water pollution, as a result of cow outwintering is proposed by McGechan M.B. et.al. (2017). The experiments are carried out on two winters at two different sites and the simulation results prove that considerable water pollution is observed due to outwintering of cows with the emission of ammonium and phosphorous. This work concludes that the cow outwintering must be minimized.

Wang Q. and Yang Z. (2016) present the health hazards of China with respect to industrial water pollution. The experimental data is collected from the China Health and Retirement Longitudinal Study (CHARLS) during the years 2011 and 2013. Random effects Logit model is utilized to correlate the health and water pollution, whereas the mediator model is employed to analyse the effects of water pollution.

Wu Z. et.al. (2017) measure the water pollution by involving the energy theory. To validate this work, the samples of Qingyi river is used and the key pollutants such as NH_3-N and COD are computed. In all the tested areas NH_3-N is greater than COD and is concluded that the pollutant discharge ratio vary with respect to location.

Reddy D.H.K. (2017) discusses the water pollution control technologies by considering the source and fate of water pollutants. Additionally, water pollution detection, prediction and monitoring systems are reviewed and discussed in detail. The overview of water pollution and pollutants is presented.

Schweitzer L. and Noblet J. (2018). This work wants to ensure that the water must be utilised for some purpose, if not at all for drinking. The sources of pollution being caused due to the human activities are presented. The measurement and monitoring of the biochemical oxygen demand (BOD) play an important role in the planning and operation of wastewater treatment plants have been reported byBaki, O. T. et al., (2018). The study has been carried out in a wastewater treatment plant in Turkey (Hurma WWTP) to estimate the biochemical oxygen demand in shorter time with a lower cost. Estimation was performed using artificial neural network (ANN) method. The root mean squared error (RMSE) and the mean absolute error (MAE) values were used in evaluating performance criteria for each model. As a result of the general evaluation, the ML-ANN method provided the best estimation results both training and test series with 0.8924 and 0.8442 determination coefficients, respectively.

Mehmet KazimYetik (2017) has reported the relationship between BOD and COD. In this study 4 years experimental water quality parameter values were analyzed for Araç steam in Turkey. The results show the correlation coefficient r is 0.72 between BOD and COD. In addition that regression analyze result gives a line equation between them this line equation is y= 0.834x + 4.55 and in final step MSE was checked. The obtained results indicated that there is a good correlation between BOD and COD in Araç stream,however more survey and data are also required to find exact correlation between BOD and COD for Araç stream.

The prospective of an artificial neural network technique (ANN) was examined by comparing the results of observed and estimated BOD and DO in the Mahanadi River have been studied by Nibedita G. et al., (2013). It was found that for prediction of the BOD and DO in the Mahanadi River lying in Odisha an ANN model appears to be a useful tool. The accuracy performance of training, validation and prediction of seasonal water quality parameters has been tested. To test the validity of ANN model, correlation coefficient (R) statistical model was used.The results are comparable and in some cases better. To test the validity of these models, performance evaluation was done using error statistic.

Vijayarengan P.et al., (2013) have studied the 50 and 100 mg kg level of zinc in the soil wasbeneficial for the growth of tomato plants. The level ofzinc in the soil above 150 mg kg proved to be toxic.The results indicated that the zinc levels 50 to 100 mg kg 1can be applied for increasing the growth and yield oftomato plants.

Jayakumar, K., (2007) have studied aCo treatment at all levels tested (except 50 mg kg^{-1} soil) decreased the various growth parameters, such as root and shoot length, and total leaf area, biochemical contents (pigment, sugar, starch, amino acid, and protein), mineral status of leaves (macro- and micronutrients), and antioxidant enzyme activity (CAT) of R. sativus plants. However, antioxidant enzymes (POX and PPO) increased with an increase in the Co level of soil. Based on the reported results, it can be concluded that the 50 mg Co kg^{-1} soil treatment was beneficial to the growth of R. sativus plants.

Jadia, C. D., et al., (2009) have reported the mobility, bioavaliability and plant response to the presence of soil heavy metals. A greenhouse experiment was conducted to determine the phytotoxic effect of heavy metals such as Ni and Pb on the growth of Georgina wild. The selected metals were dosed at various concentration ranging from 0.5, 1.0, 1.5, 2.0 mg/kg for nickel and 5.0, 10, 15, 20 mg/kg for lead separately in soil. The results revealed that as the concentration of heavy metal in soil increases, the overall length of plant decreases with respect to control.

Manivasagaperumal, R., et al., (2011) have studied the low copper concentration (50 mg kg^{-1}) had stimulatory effect on growth, dry matter yield and mineral nutrient content of green gram. Application beyond these levels (100-250 mg kg^{-1}) adversely affected the growth, dry matter yield and nutrient content.

Nourani, V., et al., (2018) have reported the prediction of BOD, the ensemble models of simple averaging ensemble (SAE), weighted averaging ensemble (WAE) and neural network ensemble (NNE), increased the performance efficiency of Artificial Intelligence (AI) modeling up to 14%, 20% and 24% at verification phase, respectively and less than or equal to 5% for both COD eff and TN eff in calibration phase. Obtained results showed that, NNE model is more robust and reliable ensemble method for predicting the NWWTP performance due to its non-linear averaging kernel.

Amosa, M. K., et al., (2015) have studied the SEM microphotographs showed opened micro pores existing in the PAC structure with a surface area of 886.2 m^2/g. The FTIR spectra revealed the three major peaks exhibited by the surface of the activated carbon at exactly wave numbers of 1737.61, 1365.10 and 1216.91 cm^{-1}. Maximum COD reduction of 84 % (227 ppm residual) was achieved from an initial concentration of 1387 ppm. This study being the first optimization process with the utility of EFB-based PAC in the treatment of the high-strength multicomponent biotreated POME; hence, the results could serve as requisite data for upscaling and/or future investigations in the utility of the precursor as a viable adsorbent.

Nasr, M., et al., (2016) have reported the benefits of unused water. Results showed that the optimum operating condition is: current-density = 12.5 mA/cm^2, electrolysis-time = 30 min, and inter-electrode gap = 0.5 cm. Results of pollutants removal efficiency and electrical energy consumption encourage the applicability of EC method in the treatment of grey water. By entering new inputs, the developed network can be used for prediction and control. The application of ANFIS indicated that current-density is the most influential input on turbidity removal.

Maachou R.et al., (2015) have demonstrated the ANFIS model has been proposed and applied to simulate recirculated sludge in activated sludge process to optimize the elimination yields of parameters COD, BOD and SS. Simulation and prediction have shown that the resulting model can adequately predict the process with satisfactory results. These last were obtained during the learning and validation periods, revealing the advantages of fuzzy reasoning and justifying the predictive power of the model developed to simulate the amount of the recycle sludge in activated sludge process.

The vertical and horizontal migrations of leach ate pollutant into landfill soil have been studied by Kebria, D. Y., et al., (2018). The physical and chemical properties of the landfill soil showed that the leach ate greatly affected the Atterberg limits and water content. The concentration of chloride, phosphor, COD, and heavy metals of leach ate in the soil samples was also measured. According to the measured and modeled results, it can be concluded that HCO_3^- concentrates between boreholes B and C and migrated deeper in the 12 years since the landfill became active, indicated to monitor the landfill because it can introduce pollutants into the soil and ground water and pose dangerous health risks to humans.

Sahooa, M. M., et al., (2015) have demonstrated the water quality values predicted by ANFIS model lies between 21-52, from this range of water quality it can be said that he water of the Brahmani River can be used for drinking water as well as domestic purposes up to a certain extent. The degree of the coefficient of determination (R_2) is 0.970 and 0.792 for the regression plots between actual WQI and predicted WQI via ANFIS model for training and testing data respectively. The mean absolute percentage error for training and testing data are 0.37 and 1.09 showed the validity of ANFIS model.

Khaled, B., et al., (2018) have reported the superiority of the ANFIS-SC models over ANFIS-GP models in term of the various performance criteria, meaning the ANFIS-SC model is more reliable than ANFIS-GP for modelling BOD_5. The best set of input variables for ANFIS-SC models is [COD, TIN]. The study suggests that BOD_5 could be computed from COD using two ANFIS approaches (ANFIS-SC and ANFIS-GP). Adaptive neuro-fuzzy inference system for water quality modeling could be successfully demonstrated by

Ahmed, A. M. et al., (2015). This may not necessarily mean that ANFIS always performs better than other conventional conceptual models; however, in this particular case the proposed ANFIS could be applied for biochemical oxygen demand forecasting with a reasonable performance. The ANFIS-I model with all the variables shows slightly better performance both in the training and testing periods than the ANFIS-II model. From the study, it is resolved that the developed ANFIS-I model is better than ANFIS-II model in prediction of biochemical oxygen demand (BOD).

Solgi, A., et al., (2017) have reported themonthly time series of BOD index was used in Karun River in Mollasani station and also, covariates like Dissolved Oxygen (DO), monthly temperature, and river flow were used from 2002 to 2014. The results indicated that the SVR model with RMSE = 0.0338 mg/l and R_2 = 0.843 has better performance than the ANFIS model with R_2 = 0.828. Also, applying the wavelet transform on input data of the SVR model improved the results to R_2 = 0.937 and RMSE = 0.0210 mg/l. Therefore, combining the SVR with the wavelet transform (WSVR) was a good idea to improve the prediction of the BOD value in Karun River. Finally, the combination was recognized as a suitable method and the BOD was predicted in six months.

Neuro-fuzzy models are based on the extraction of knowledge from data collected upstream and downstream of a treatment plant have been reported by Maachou, R et al (2015). The historical values of the observed yields associated with the energy consumed during the study period enable the prediction of the energy needed for a validation period. The input parameters used in this study includes the removal yields of organic pollutants parameters and energy consumption as a decision parameter with respect to the discharge standards. The predictive power of energy shows the feasibility and robustness of the simulation approach with a filtered data.

Nadiri, A. A. et al., (2018) has demonstrated the Linear combination of the three FL models via the committee FL (CFL) algorithm outperforms individual FL models by improving the prediction result. The trained ANN model provides a better prediction model for the Tabriz wastewater treatment plant than the individual FL models and the CFL models. The SCFL model improved the CFL-WA results by approximately 30% for BOD, 31% for COD and 23% for TSS in the testing step. The results provide evidence that the SCFL model is capable of predicting more accurate effluent water quality parameters of TWWTP than the other alternatives tested in this research. The nature of non-linearity is easily modeled in GP which is difficult in ANN was case out by Vanitha, S., (2016). GP quite accurately models a natural ecosystem even with very less training data set, and hence can be used for modeling of aquatic systems. It is recommended to carry out more detailed research with additional parameters governing the ecosystem to have a deeper understanding of the physics of the ecosystem.

Neuro-fuzzy inference system (ANFIS) modeling to assess the removal efficiency of Kjeldahl Nitrogen of a full-scale aerobic biological wastewater treatment plant was reported by Manu, D. S., et al., (2017). The errors related to the prediction of effluent Kjeldahl Nitrogen concentration by the SVM modeling appeared to be reasonable when compared to that of ANFIS models with Gbell and trapezoidal MF. From the performance evaluation of the developed SVM model, it is observed that the approach is capable to define the interrelationship between various wastewater quality variables and thus SVM can be potentially applied for evaluating the efficiency of aerobic biological processes in WWTP. Neuro-fuzzy inference system (ANFIS) and a neural network model with exogenous inputs (ANN) were used for removal of reactive red dye prediction was studied by Kahkha, M. R. R., et al., (2016). Results indicated that both models are favorable for simulation and prediction of results. By investigating effect of adsorbent (ZnO nanoparticle) dosage on removal efficiency and lowest amount of MAE, RMSE and highest amount of R^2 obtained from ANFIS model found that ANFIS model is more accurate for this parameters. By studying effect of dye concentration on removal efficiency and amount of MAE, RMSE and R2 obtained from ANN model found that ANN more matches with actual data for these parameters.

Bagheri, M., et al., (2015) have studied the low experimental values of input data to train ANNs, the MLPANN-GA as compared with RBFANN-GA is more accurate due to higher R2 and lower RMSE values. The accuracy of all models increased when GA was applied to the MLPANN and RBFANN models. The modeling results using the RBFANN-GA and the MLPANN-GA demonstrated an approximately normal distribution of residuals produced by the RBFANN-GA and the MLPANN-GA models. The normal distribution (Gaussian curve) demonstrates that our results are symmetrical and their axis approach zero for all, train and test data sets.

Mohammadpour, R. et al., (2015) have demonstrated the SVM and two methods of ANNs, namely FFBP and RBF, were employed to investigate the WQI in the free surface constructed wetland. Seventeen points of the wetland were monitored twice a month over a period of 14 months, and an extensive dataset was collected for 11 water quality variables. The high value of the coefficient of correlation (R_2=0.9984) and low error (MAE=0.0052) indicated that the SVM model provides better prediction compared to theRBF network with R_2 = 0.9960 and MAE = 0.0080. Furthermore, the result provided by SVM was comparable with that of the FFBP network (R_2=0.9988 and MAE= 0.0044). This research highlights that the SVM and FFBP can be successfully used as valuable methods for the prediction of water quality in the wetlands.

The application of FFNN model was used to predict the CODeff in Nicosia municipal wastewater treatment plant, for the comparison MLR model was also adopted have been studied by Abba, S. I.et al., (2017). The sensitivity analysis was conducted using various statistical performance tools. According the results the determination coefficient and root mean square error was found to be 0.7034and 0.1018, respectively, which on the basis of comparing the result, the ANNs turned to be high in term of performance and efficiency to the MLR for modelling the wastewater treatment plant. It was also revealed from the analysis that, other water quality indices and experimental model could not produce a better prediction of COD in this wastewater treatment plant.

Csábrági, A et al., (2015) have studied by training and testing of the ANNs, it was found that the GRNN model provided better predictions of DO than the BPNN, and so the use of GRNN is justified not only due to its better performance, but also on account of it quickness, as, in contrast to BPNN, it is a one-pass training algorithm that does not necessitate an iterative training process. A comparison of the ANNs with the conventional MLR shows that the ANNs demonstrated better performance indicators than the MLR when every model was trained and tested by the same data sets and input variables. Conclusions have shown that the two ANNs, and especially the GRNN are practical methods for predicting DO concentrations in a river.

The consortial system (consortium-PG) between G. pulchella and P. monteilii ANK showed a greater potential of decolorization of textile dyes and effluents compared to their individual cultures have been reported by Kabra, A. N., (2013). Within the consortium-PG two different enzyme systems from both the sources worked together, resulting in efficient and faster decolorization of the dyes. Differential fate of metabolism of Scarlet RR by G. pulchella and P. monteilii and their consortium were proposed. The developed consortium reactor performed efficient removal of dye, TOC, COD and BOD from the dye mixture and textile effluents.

Phytoremediation using green plants has been explored and studied extensively regarding the absorption or decolorization of dyestuffs originating from various sources throughout the world Tahir, U. et al., (2015). Hence a lot of research is needed for exploration of dye degrading plant species, their operating remediation mechanisms, applications of various additives and influence of associated microbial activities for improvement of phytoremediation processes. Moreover the applied molecular techniques and creation of transgenic plants will decipher the specific metabolic pathways involved in dye metabolism, which will further boost the applicability of phytoremediation technologies for alleviating the impacts of dye laden wastes on various environmental compartments.

Nezhad, M. F et al., (2018) have demonstrated the nonlinear formula was developed using AHP methods and sub-indices based on the legal limitations for reuse and disposal. In the second part, the applicability of Bayesian networks for modeling EQI was assessed. Comprehensive quality parameters were assessed and selected using the AHP method. Results revealed that Bayesian methods offered considerable potential for the estimation of the effluent quality index. This method can be employed for data with uncertainty. BN model enabled the probabilistic modeling of the EQI and identified the importance and contribution of quality parameters.

III. PROPOSED WASTEWATER DISPOSAL AREA SPECULATION SYSTEM

This section describes the idea of the proposed approach, in addition to the overview of the approach, as depicted in figure 3.16. The entire work cycle is subdivided into two phases, which are training and testing. In the training phase, the classification system is given knowledge about the range of values, which should be there for each and every attribute.

The training phase is the knowledge gaining phase in which the classifier is made to learn about the extremity of each and every attribute or chemical pollutant. Hence, the performance of the classifier depends on the knowledge it has gained from the samples. The knowledge gaining process solely based on the phase of feature extraction. The better the feature extraction, the more accurate is the classification. Yet feature extraction is the most crucial step, as the feature set should not be maximal or minimal.

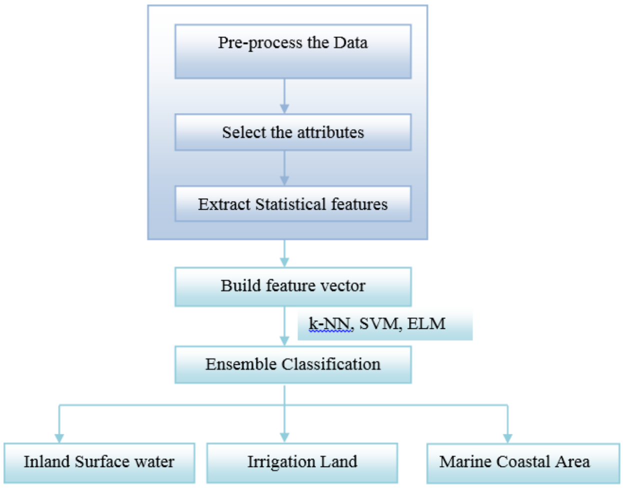

Fig.2. Overall flow of the work

The larger feature set consumes more memory, time and computational complexity. The performance of the system must also be taken into account and not just the accuracy rates. On the other side, smaller feature set may not involve or miss out important features, which may lead to inaccuracy. Hence, building the feature set with significant features is important. Taking this into account, this work builds the feature vector by considering the most important features that would suffice for better classification. In the testing phase, the test sample is passed on to the classification system, which performs the same process as in training phase such as pre-process and extract features from the data.

Finally, the classification system chooses the best wastewater disposal area by taking the range of chemical pollutant into account. The phases involved in this work are presented in the following subsections.

International Journal of Scientific Research & Engineering Trends
Volume 6, Issue 6, Nov-Dec-2020, ISSN (Online): 2395-566X

1. Data Pre-Processing

This work takes the pollution control standards into account for accomplishing the classification task. The data of pollution controls standards comprise about thirty one attributes, however certain fields are left as empty. When the data is processed, the system cannot perform well and may throw warnings. In order to overcome this issue, the proposed approach fills all the empty fields as zero and hence, the void spaces are filled up. By this way, the data pre-processing step is carried out and this may seems to be simple but essential.

2. Attribute Selection

Attribute selection is the second phase that attempts to compare the fields available in the train and test samples. The pollution control standards contain all the possible chemical pollutant, however it is not always possible to measure the range of all the chemical pollutants in the water, as it is waste of time. Additionally, as the system is trained with all the possible range of values of chemical pollutants, the test sample can be processed with any feasible pollutant. This is the merit of this system and it attempts to extract the chemical pollutants that are computed by the test sample from the train sample. The remaining fields are set to zero. This feature of this system brings in the utmost flexibility to the system.

3. Statistical Feature Extraction

Feature extraction is the most important step for any classification system. The efficiency of this phase determines the performance of the classification. The time consumption and accuracy of the system is determined by this step and so it is given more importance. As all the values of the data involve numerical values, this work intends to extract basic statistical features such as mean, standard deviation, entropy and variance from the data. All these features are computed by the following equations.

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

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

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

$$E = -\sum_{i=1}^{n} P(D_i).\log_2 P(D_i) \quad (4)$$

All these basic statistical features are extracted from the data and the feature vector is formed as follows. The feature vector, which is the outcome of the feature extraction phase is built by

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

In the above given equations, the mean, standard deviation, variance and entropy are represented by M, SD, V and E respectively. The input data is represented by D and the total number of entities is denoted by n. The overall algorithm of this work is presented as follows.

Algorithm for Wastewater Disposal Area Speculation System

// 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 ($\boldsymbol{M}$), standard deviation ($\boldsymbol{SD}$), variance (V) and entropy (E)
Features;
Construct $\boldsymbol{fv}(\boldsymbol{TD})$ and store it in the local database;
Feed the knowledge to ensemble 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 $\boldsymbol{M}, \boldsymbol{SD}, \boldsymbol{V}, \boldsymbol{E}$ features;
Construct $\boldsymbol{fv}(\boldsymbol{TD})$;
Apply Ensemble classifier to match the test and train samples;
Collect the classification results of k-NN, SVM and ELM;
Choose the dominant result as the final result;
Analyse the performance;
End;
End;

The feature vector is represented by fv and TD is the training data. All these basic statistical features are extracted from the input data and the feature vector is formed. With the so formed feature vector, the ensemble classifier is trained as presented in the following section.

3. Ensemble Classification

The classifiers are provided with enough knowledge by means of the features being extracted from the data. The knowledge is given in the form of feature vectors and the classifier learns about the nature of data. The classifier then equips itself to attain better classification by analysing the data. Though the working principles of the classifiers vary, their objective is the same. The classification problem can contain two or any number of classes. This work incorporates ensemble classifier, which is a group of classifiers and the classification results are obtained from all the classifiers. Finally, the decision with maximum hits is declared as the final.

International Journal of Scientific Research & Engineering Trends
Volume 6, Issue 6, Nov-Dec-2020, ISSN (Online): 2395-566X

This way of classification improves the accuracy rates of the classification by reducing the false positives and false negative rates as well. The classifiers being utilised for this work are k-NN (as presented in section 3.2.2.5), SVM (as presented in section 3.3.3.4) and ELM (as presented in section 3.4.3.4). The final classification results of the k-NN, SVM and ELM are collected and the maximal occurring result is detected. This work employs three classifiers and hence eight different cases are possible. To exemplify this idea, when two or three classifiers arrive at the same result, then the result is considered as final. This kind of classification is more reliable than employing a singe classifier. Besides this, the choice of classifiers to be a part of ensemble classification is important and so, this work has chosen the best performing and promising classifiers. The false positive and false negative rates are considerably reduced, which in turn improves the sensitivity and specificity rates.

IV. 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 following graphs exhibit the performance of different classifiers such as k-NN, SVM and ELM.

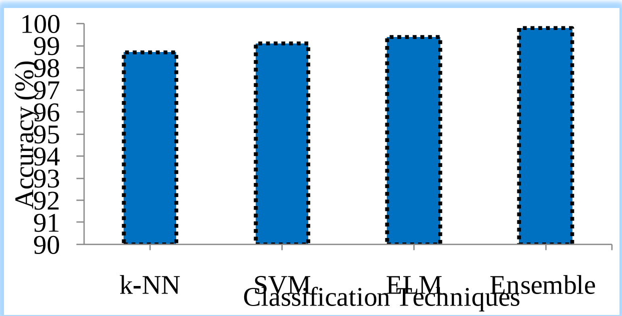

Fig.3. Accuracy rate analysis

The accuracy rate of the ensemble classification is evaluated and compared with the other classification techniques such as k-NN, SVM and ELM. The performance of ensemble classification is better than the comparative techniques. The ensemble classification technique shows the accuracy rate of 99.8 percent and is closely followed by the ELM classifier with the accuracy rate of 99.4 percent. The SVM and k-NN classifier proves an accuracy rate of 99.1 and 98.7 percent respectively. The following figure presents the sensitivity rate of the proposed ensemble classification approach.

The sensitivity of the proposed approach is analysed and is found to be satisfactory with reasonable sensitivity rates. The ensemble classifier shows 99.6 percent as the sensitivity and the ELM classifier shows 98.6 percent. Though the accuracy rates of the ensemble and ELM classifier is closer to each other, the sensitivity rates show some difference. The SVM classifier has shown 97.6 percent as sensitivity rate. The reason for the better sensitivity rates of ensemble classification is the minimal false negative rates. The specificity rate analysis of the proposed approach is depicted in figure 3.19.

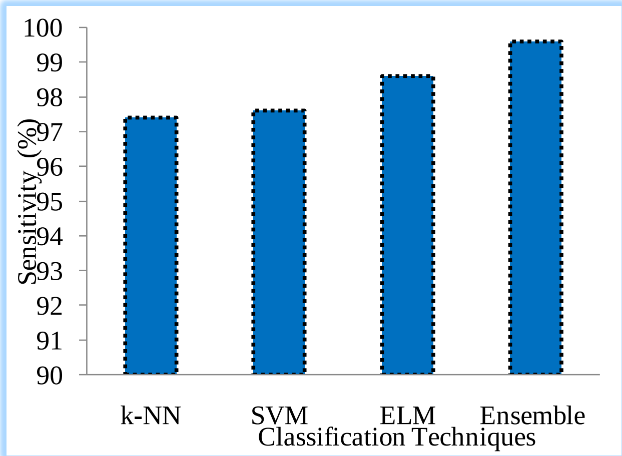

Fig.4. Sensitivity rate analysis

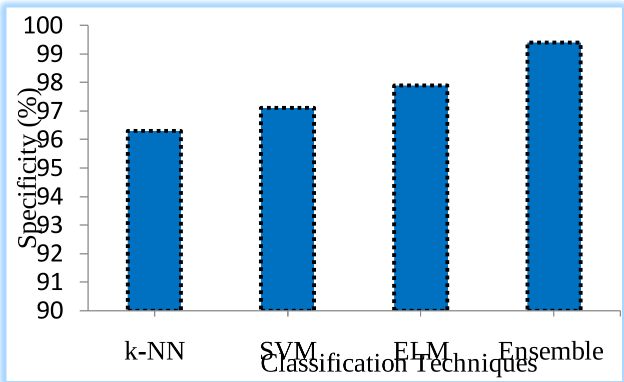

Fig.5. Specificity rate analysis

The specificity rate analysis proves that the ensemble classifier performs better with the greatest specificity rate. The ensemble classifier shows 99.4 percent as the specificity rate and is followed by the ELM classifier with 97.9 percent. The reason for the greater specificity rate shown by the ensemble classifier is that the classification decision is not made by a single classifier but a group of classifiers. As the final classification decision depends on the individual decisions of three classifiers, the possibility of obtaining false positive results is comparatively low, when compared to other classifiers. The ensemble classifier proves better results with respect to accuracy, sensitivity and specificity rates. The time consumption analysis of the proposed approach is depicted in figure 3.20.

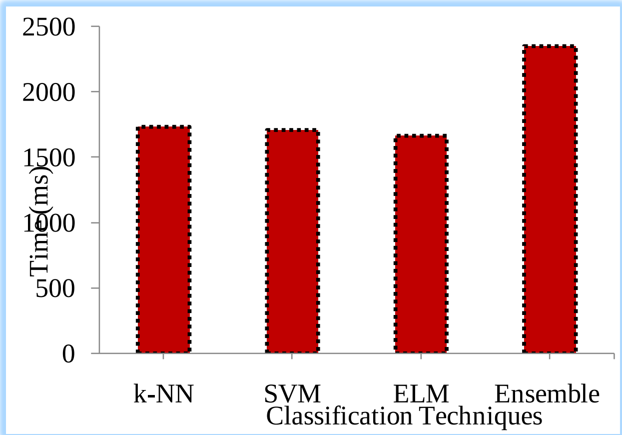

Fig.6. Classification Techniques.

On observing the results of time consumption analysis, it is noted that the time consumption of ensemble classification is greater than the other classification techniques. The main reason for more time consumption of the ensemble classification is the decision dependency of more classifiers. However, the time consumption is tolerable and the better results are attained at the cost of tolerable time consumption. From the experimental analysis, it is evident that the time consumption of ensemble classifier is more than the individual classifiers. The reason for more time consumption is that this work relies on the decision of three individual classifiers, rather than the decision of a single classifier. This increases the accuracy rates at the cost of time consumption. However, the time consumption is tolerable and justifiable. By this way, the wastewater disposal area is selected by taking the intensity of water pollutant into account.

V. CONCLUSION

This phase presents a novel wastewater disposal area speculation system based on ensemble classification. It is important to control and monitor the water pollution and is equally important to plan for the discharge location for the wastewater by taking the intensity of the pollutants. This work achieves the research goal by means of four key phases such as pre-processing, attribute selection, feature extraction and classification. The pre-processing phase aims to fill out the empty fields of the data and attributes are selected from the train data with respect to the test data sample. The statistical features are extracted from the data and the classification is performed by ensemble classifier, which is a combination of k-NN, SVM and ELM.

REFERENCES

[1] 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.
[2] KaanYetilmezsoy, ZehraSapci-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.
[3] 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.
[4] 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.
[5] YasamanSanayei, NazChaibakhsh, 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.
[6] 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.
[7] 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.
[8] Taufiqurrahman ;Ni'am Tamami ; DitoAdhi 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.
[9] XuLuo; 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.

[10] AsmaaHashemSweidan ; 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.

[11] 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.

[12] XuLuo; 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.

[13] 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.

[14] ZhengFeng; Jun Yang ; XuLuo, "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.

[15] C. Brintha Malar, Dr. S. Akilandeswari, “Optimal Area Prediction for Waste Water Discharge by Employing Supervised Learning” International Journal of Current Advanced Research, Vol.6, No.12, pp.8523-8526, 2017.

[16] C. Brintha Malar, Dr. S. Akilandeswari, “Auto-Suggesting Discharge Location System for Wastewater Treatment”, International Journal of Current Research and Modern Education, Volume 3, Issue 1, Page Number 213-217, 2018.