

## **SVM Model to Predict the Water Quality Based on Physicochemical Parameters**

**Manisha Koranga**

Department of Information Technology,  
D. S. B. Campus, Nainital, Uttarakhand, India.  
*Corresponding author:* koranga.manisha@gmail.com

**Pushpa Pant**

Department of Geography,  
M. B. Government P.G. College, Haldwani, Uttarakhand, India.  
E-mail: pushpapant01@gmail.com

**Durgesh Pant**

School of Computer Sciences & Information Technology,  
Uttarakhand Open University, Haldwani, Uttarakhand, India.  
E-mail: dpant@uou.ac.in

**Ashutosh Kumar Bhatt**

Department of Computer Science,  
Birla Institute of Applied Sciences Bhimtal, Uttarakhand, India.  
E-mail: ashutoshbhatt123@gmail.com

**R. P. Pant**

Department of Allied Sciences,  
Graphic Era Hill University, Bhimtal Campus, Bhimtal, Uttarakhand, India.  
E-mail: pant\_rp@rediffmail.com

**Mangey Ram**

Department of Mathematics; Computer Science and Engineering,  
Graphic Era Deemed to be University, Dehradun, Uttarakhand, India.  
E-mail: mangeyram@geu.ac.in

**Tarun Kumar**

Department of Chemistry,  
M. B. Government P. G. College, Haldwani, Uttarakhand, India.  
E-mail: tarunksj@gmail.com

(Received October 12, 2020; Accepted January 5, 2021)

### **Abstract**

Analysis of water quality is a very important and challenging task in the management of water bodies and requires immediate attention as it adversely affects the health of living beings. Three parameters namely, pH, Total Dissolved Solids (TDS), and Turbidity were used for data analysis. In this study for mapping of training samples from input space to higher dimensional feature space, LibSVM, (a library of SVM) was used with the use of two kernel function types Radial Basis Function and Polynomial function. For performing the experiment, the three parameter combinations (C, d,  $\gamma$ ) were evaluated based upon the kernel by taking various range values to obtain the best type of kernel functions through a 10-fold cross-validation process. After performing all experiments, a comparative analysis was done to evaluate the best parameter combination (C, d,  $\gamma$ ) and the values of performance measures. The result shows that the

optimum model developed using LibSVM with the use of Polynomial Kernel function which gives an accuracy of 99.434% in predicting water quality.

**Keywords-** Support vector machine, Water quality, LibSVM, Kernel functions, Nainital lake.

## 1. Introduction

Water is mainly responsible for the survival of all biotic species of the whole ecosystem which maintain human health and welfare. The main sources of water in India are Rivers, Lakes, Oceans, etc. Water quality mainly depends upon physiochemical and biological characteristics like conductivity, TDS, color and taste of water, hydrogen ion concentration (pH value), Turbidity, Temperature, inorganic heavy metals (magnesium, aluminum, titanium, iron, copper, tin, silver, gold, platinum, gallium, thallium, and hafnium), Biological oxygen demand (BOD), Dissolved Oxygen (DO), etc., (Sharma, 2014; Pareek et al., 2018; Tadesse et al., 2018). But the water quality of the water bodies contaminated regularly. Water contamination depends upon several factors such as industrial activities, geographical condition of area, artificial products (gasoline, oil, road salts and chemicals), and agricultural activities. It is further classified in to following five types; inorganic, organic, microorganisms, radio nuclides, and disinfectants (Nollet, 2000). In which, inorganic chemicals are mainly responsible for the contamination of water quality as compared to other factors. The striking reason behind conducting this research is growing water pollution in the hilly areas of Uttarakhand. Nainital is one of the largest, historical tourist places of Uttarakhand. It is famous for this beautiful lake and its true natural beauty. But it has been facing water problems from the last few years, due to over-dependence on lake water for the drinking water supply.

As the place is presently the middle of tourism, consequently that is critical to assess the ingesting water of the Nainital Lake to guarantee secure ingesting water for citizens and tourists. But it was found that water quality of Nainital Lake was deteriorating rapidly as there are many anthropogenic reasons like increasing population, tourism movement in the city which leads to pollution, intense boating activity, construction activity, and disposal of solid waste and throwing of edible material into the lake besides these factors, other factors are eutrophication, untreated sewage, lack of drainage maintenance, landslides, etc. (Sharma, 2014). Nainital Lake is considered to be the lifeline for the resident here and its surrounding areas. Therefore, it is very essential to maintain and monitor the water quality of the Lake.

Several research studies have been going in this field by using different techniques, because of the non-linearity nature of the Environmental System (Sarkar and Pandey, 2015). Nowadays various machine learning techniques have been employed to find the solution of the problem. To assess the water quality authors have conducted an acute toxicity test on Zebrafish (*Danio rerio*) in a toxic environment by the use of four metal ions. The behavior response of Zebrafish was visualized to assess the water quality using techniques Biomonitoring and Multiclass SVM. The experimental result represents that the response of Zebrafish to one out of four metal ions comes out to be strong and study also revealed that the response of Zebrafish using method of Biomonitoring results in highly cost effective (Liao et al., 2011). To develop an efficient model authors have performed an experiment to determine the water quality of the Langat River Basin using Support Vector Machine. In this study, six parameters that are found in the catchment area of dual reservoir were used for the experiment. The experiment performed in two ways: Scenario 1 and Scenario 2. The result represents that scenario 1 performed better outcome as compared to

Scenario 2 on the basis of MSE and Correlation Coefficient (Yahya et al., 2019). To conduct a case study on Wen-Rui River, China authors have used of 11 hydro chemical variables to determine the concentration of dissolved oxygen using Support Vector Machine in hypoxic river systems. They used four different calibration models, such as General Regression Neural Network (GRNN), Multiple Linear Regression (MLR), Backpropagation Neural Network (BPNN), and Support Vector Machine (SVM). The prediction accuracy of models was evaluated and compared using Mean Square, Determination coefficient ( $R^2$ ) and Nash-Sutcliffe (NS) to assess the performance of the models. The experimental result represents that SVM model outperformed on models and can efficiently predict water quality (Ji et al., 2017). To predict water quality using Least squares support vector machine method authors have conducted experiments on time-series data. The experimental result represents that LS-SVM performed better than Multilayer Back propagation and Radial-Basis Neural Network to better meet the requirements of water quality prediction and perfectly works well on small sample cases (Tan et al., 2012).

To analyze the performance for predicting water quality components of Tireh River situated in the Southwest part of Iran authors used three Artificial Intelligence techniques consists of Artificial Neural Network (ANN), Support Vector Machine (SVM) and Group method of data handling (GMDH). For developing ANN and SVM model different transfer functions and kernel functions were evaluated. The result showed that tansig and RBF as transfer and kernel function outperformed among others. The comparison of GMDH model result with ANN and SVM model has been done and it was found that GMDH model has acceptable performance but SVM come out to be the best model with lowest DDR value and high accuracy (Haghiabi et al., 2018).

The study conducted by authors to explores data-based methods for forecasting short-term water quality. The prediction of water quality indicators was done with Support Vector Machine and fed to particle swarm algorithm based optimal wavelet neural network to forecast the complete status index of water quality. To examine the effectiveness of proposed approach Gubeikou monitoring section of Miyun reservoir were used. The result showed that proposed model performed better than other data driven models including traditional BP neural network model, wavelet neural network model and Gradient Boosting Decision Tree model in terms of stability and consumes less time (Zhang et al., 2017).

To predict the monthly Stream flow authors have used two soft computing techniques including Artificial Neural Network (ANN) and Support Vector Machines (SVM). The result of both models was compared on the basis of determination coefficient ( $R^2$ ), root mean square error (RMSE) and mean absolute error (MAE) to evaluate the performance. The result showed that SVM model is superior as compared to ANN model (Adnan et al., 2017).

The result of this research reveals that SVM is a powerful technique to resolve several kinds of environmental issues. Through literature review, it was found that only a few chemical and statistical analysis related research work was reported for Nainital lake previously. So, in this paper, the author proposed an efficient model for the prediction of water quality based upon LIBSVM. This kind of prediction model by using a machine learning technique was reported first time for Nainital Lake. (As far the best knowledge of the authors).

## 1.1 Problem Statement and Objective

Increasing tourism movement, improper treatment of sewage waste, and unplanned construction in the catchment area in Nainital city have increased the level of pollution. Due to the pollution, the quality of Nainital lake water is deteriorating day by day as it is the only source of drinking water for the people here. Therefore, it is very important to maintain the water quality of the lake. Several studies in the past conducted on Nainital lake reveals that the water quality of the lake is unsuitable for drinking (Sharma, 2014; Dalakoti et al., 2018). In this paper the efficiency of Support Vector Machine models is evaluate in predicting water quality of Nainital lake.

The objective of this research study:

- a) To proposed a powerful and efficient machine learning model for the prediction of water quality for Nainital Lake.
- b) To overcome the limitations of traditional methods and develop computational models that efficiently and accurately work.

## 2. Material and Methods

### 2.1 Study Area

Nainital Lake, as shown in Figure 1 situated amidst the township of Nainital in the Uttarakhand, India. It is a major tourist spot and an important source of fresh water. It is a natural kidney-shaped having uniform temperature and density from top to bottom, situated at 290 24' N latitude and 790 28' E longitudes. The surface area covers by lake is 48 hectares. The maximum depth of the lake is 27.3m and mean depths of the lake is 16.2m, respectively.



**Figure 1.** Nainital Lake (Source: <https://www.postoast.com/nainital-lake>).

## 2.2 Selection of Appropriate Inputs

Experimental data used in this study was obtained from Uttarakhand Jal Sansthan, Nainital, and Sensor-based device installed in Nainital lake for determining water quality by Uttarakhand Science Education and Research Centre (USERC), Dehradun. Time series data of 1.5 years from 2018 to 2019 were utilized in this study. Three water quality parameters namely, pH, Total Suspended Solids, and Turbidity were chosen for SVM Modelling. Statistical chart corresponding to selected parameters shown in Table 1.

**Table 1.** Statistical chart corresponding to selected parameters for Nainital Lake.

Basic Statistics	pH (pH)	Total Dissolved Solids (mg/L)	Turbidity (NTU)
Max	8.9	460.84	2.8
Min	8.1	430	1.1
Mean	8.362	440.249	1.34
Standard Deviation	0.152	6.048	0.215
Coefficient of Variation	1.8177	1.373768	16.044776

**Table 2.** Drinking water quality parameters, its units, Indian Standards and Recommending water-related agencies.

Water Quality	Unit in Measured	Desirable Limit	Associated Agencies
pH	pH	6.5-8.5	BIS/ICMR
Turbidity	NTU	5	BIS
Total Dissolved Solids (TDS)	mg/L	500	WHO

Above Table 2, shows the selected water quality parameters, their units, Desirable Limit and the Associated agencies.

## 2.3 Support Vector Machine

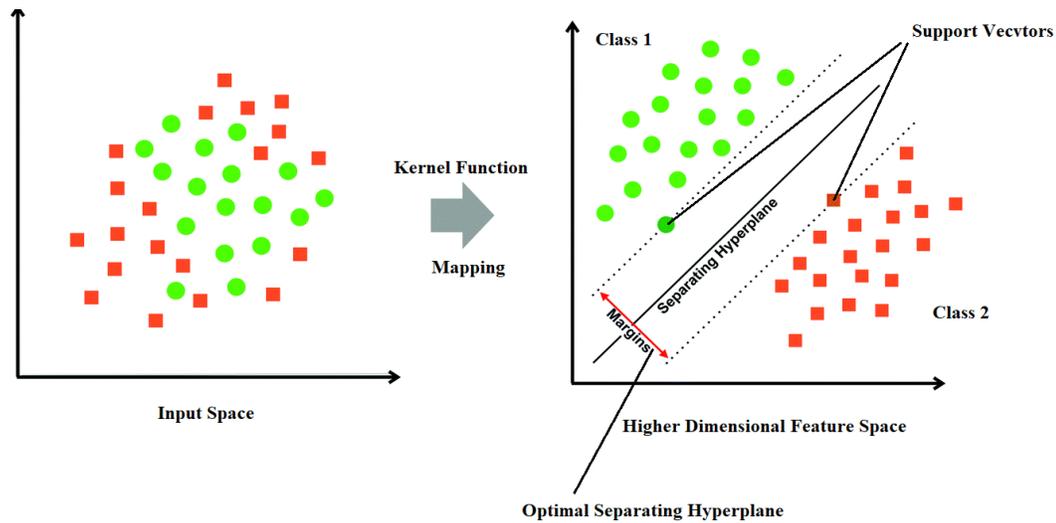
Support vector machine is one of the most popular Supervised machine learning algorithms which is used to solve the problems related to Classification, Regression, Recognition and time-series (Kara et al., 2011). But it is generally used in Classification Problem.

### 2.3.1 Basic Principle

It is based on the statistical theory of learning proposed by Vapnik and Chervonekis in the late 1960's (Vapnik and Chervonekis, 1971). SVM becomes quite popular due to some features such as Good generalization ability, high prediction accuracy, faster training speed. SVM capable of working perfectly with small sample data and it has high ability to classify both linear as well as non-linear data. In two-classification problem, if there exist some original training samples which are represented as  $x_i = \{x_1, x_2, x_3 \dots x_n\}$  in the input space where  $x_i \in R^d$  and their corresponding labels are represented as  $y_i \in \{-1, +1\}$  (Kara et al., 2011).

Then the main objective of SVM is to construct a hyperplane between training samples which divides the training samples by maximal margin so as to mapped them from input space to higher dimensional feature space. So, that the +1 labeled vector lies on one side of the hyperplane and -1 labeled vectors lie on the other side of the hyperplane.

The training samples that lie closest to the hyperplane are called support vectors (Tong and Koller, 2001). The maximum margin hyperplane also known as Optimal Separating Hyperplane (OSH). The mapping of training samples from input space to higher dimensional space is performed with the help of kernel function  $k(x_i, x_j)$  as shown in Figure 2.



**Figure2.** Transformation of input space to corresponding high dimensional feature space using kernel function.

### 2.3.2 Kernel Function

There are four kernel functions types present in SVM such as Linear, Polynomial, Radial Basis and Sigmoid. But in this paper, we have used only two main kernel functions types which are Polynomial and Radial basis kernel functions. The equation of these Kernel functions are as follows:

a) Polynomial Function:

$$k(x_i, x_j) = (x_i \cdot x_j + 1)^d \quad (1)$$

where,  $d$  is the degree of polynomial function.

b) Radial Basis Function:

$$k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (2)$$

where,  $\gamma > 0$ ,  $\|x_i - x_j\|$  denotes the Euclidean Distance from center  $x_i$  to the input vector  $x_j$  (Liao, 2012).

### 3. Experiment

In SVM, the selection of kernel function plays a very critical role to achieve a higher prediction accuracy. In this research paper, LibSVM has been used for classification with the use of two kernel function types Polynomial and Radial Basis functions. Several levels of degree used in polynomial function ( $d$ ) along with regularization parameter ( $C$ ) and gamma constant ( $\gamma$ ) were evaluated in parameter setting experiments (Tarmizi et al., 2014; Larios and Villarica, 2019).

In order to train the model, test it with a certain part of the data the research explores Cross-validation technique were used. In Cross Validation technique the input data splits into k subsets and iterates over all the subsets considering k-1 subsets as training data and keeping 1 subset as testing data. We split the data into k=10 subsets and ran cross validation. The result obtained using both the kernel types were compared to analyze, the best SVM predictive model. All experiments were executed on WEKA software using LibSVM a library developed for SVM implementation (Chang and Lin, 2011).

Table 3 shows the summary of parameter values which was used in the experiment with respect to kernel function types and their parameters.

**Table 3.** SVM parameter level evaluated in parameter setting experiment.

SVM parameter level evaluated in parameter setting experiment			
Classifier	Parameters		
LibSVM	Degree (d)	Gamma in Kernel functions ( $\gamma$ )	Regularization (C)
Radial Basis	-	0.1, 0.2, 0.3.....2.0	1,10,100,1000
Polynomial	1, 2, 3	0.1, 0.2, 0.3.....2.0	1,10,100,1000

In this research, to know how efficient our predictive models are? The following classification measures such as Accuracy, Precision, Recall, F1\_measure, ROC was used (Goutte and Gaussier, 2005; Sokolova et al., 2006). The classification measures are as follows:

a) Accuracy

It is a fraction of correct number of predictions made by the model over all the observed values. Accuracy measured by Equation (3), where TP refers to true positive, TN refers to true negative, FP refers to false positive and FN refers to false negative.

$$Accuracy = \frac{TN+TP}{(TP+FP+TN+FN)} \quad (3)$$

b) Precision

It is the fraction of correct classified instances of the positive class out of the total classified instances of that class. Precision measured by Equation (4), where TP refers to true positive, TN refers to true negative, FP refers to false positive and FN refers to false negative.

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

c) Recall

It is the fraction of correct classified instances of the positive class over total instances that actually classified as positive. Recall measured by Equation (5), where TP refers to true positive, TN refers to true negative, FP refers to false positive and FN refers to false negative.

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

d) F1\_measure

It is the Harmonic mean between Precision and Recall. Its range lies between 0 and 1. F1\_measure measured by Equation (6), where TP refers to true positive, TN refers to true negative, FP refers to false positive and FN refers to false negative.

$$F1\_measure = 2 * \frac{precision*recall}{precision+recall} \tag{6}$$

e) Area Under the Curve-Receiver Operating Characteristic Curve

It is graph illustrating the performance of classification models at various thresholds settings. ROC represents the probability curve and AUC represents the degree of separability. It is one of the most important evaluation metrics for checking the performance of classification models.

**4. Result and Discussion**

In order to determine efficient parameter combination a total of 320 experiments were evaluated in parameter setting experiment. A total of 80 parameter combination was evaluated for radial basis function and a total of 240 parameter combinations were evaluated corresponding to three levels of degree for Polynomial function.

The predicted accuracy of Polynomial SVM model ranges between 78.6792 % and 99.434%. On the contrary, the predicted accuracy of the Radial Basis SVM Model ranges between 78.6792% and 98.6792%. The best parameter combination (C, d,  $\gamma$ ) of SVM models evaluated for both Kernel types is represented in Table 4.

**Table 4.** Best Parameter Combination (C, d,  $\gamma$ ) of SVM model to achieve high Accuracy tested for both LibSVM kernel type.

Best Parameter Combination of SVM Models					
S. No	Kernel Type	C	d	$\gamma$	Accuracy (%)
1	Radial Basis	1000	-	0.1	98.679
2	Polynomial	1	2	0.1,0.2,0.4,0.8,0.9,1.0,1.1,1.3,1.4,1.5,1.6,1.8,1.9	99.434
3	Polynomial	1	3	0.1,0.2,0.4,1.3,1.5	99.434
4	Polynomial	10	3	0.1	99.434

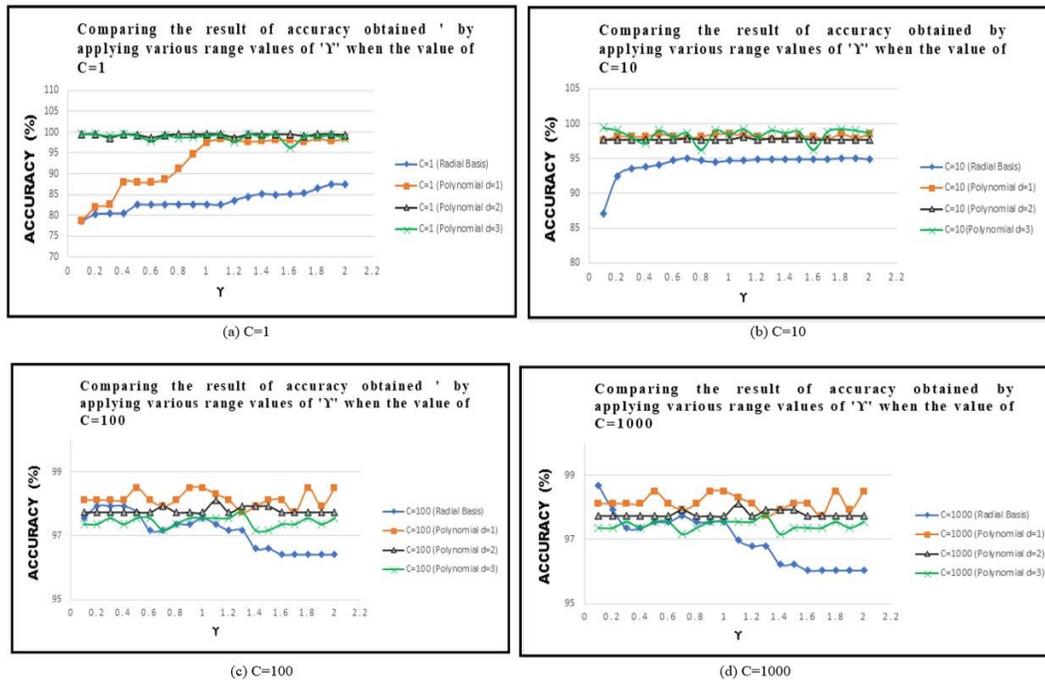
It is clear that Polynomial SVM model attained the maximum accuracy i.e., 99.434% at degree(d) =2 and degree(d)=3 when the value of C=1 and C=10 at various values of  $\gamma$ . The result of Accuracy measured in case of Radial Basis SVM model is less as compared to Polynomial SVM model. The maximum accuracy of Radial Basis SVM Model was reported 98.679%, when C=1000 and  $\gamma$ = 0.1.

Figure 3, illustrates the graphs of the result of predicted accuracy after parameter setting experiment.

The graphs in Figure 4, illustrates the result of Precision measure after parameter setting experiment. The performance of Precision measure in case of Radial basis SVM model ranges between 0.787 and 0.987. On the contrary, the performance of Precision measure in case of the Polynomial SVM Model ranges between 0.787 and 0.994. It was observed that the Polynomial SVM model attained maximum value when C=1 and C=10.

The graphs in Figure 5, illustrates the result of Recall measure after parameter setting experiment. The performance of Recall measure in case of Radial basis SVM model ranges between 0.787 and 0.987. On the contrary, the performance of Recall measure in the case of the Polynomial SVM Model ranges between 0.787 and 0.994. It was observed that the Polynomial SVM model attained maximum value when  $C=1$  and  $C=10$ .

The graphs in Figure 6, illustrates the result of F1-Measure after parameter setting experiment. The performance of F1\_measure measure in case of Radial basis SVM model ranges between 0.881 and 0.987. On the contrary, the performance of F1\_measure measure in the case of the Polynomial SVM Model ranges between 0.881 and 0.994. It was observed that the Polynomial SVM model attained maximum value i.e., 0.994 achieved when  $C=1$  and  $C=10$ .



**Figure 3.** Results of predicted accuracy after parameter setting experiment.

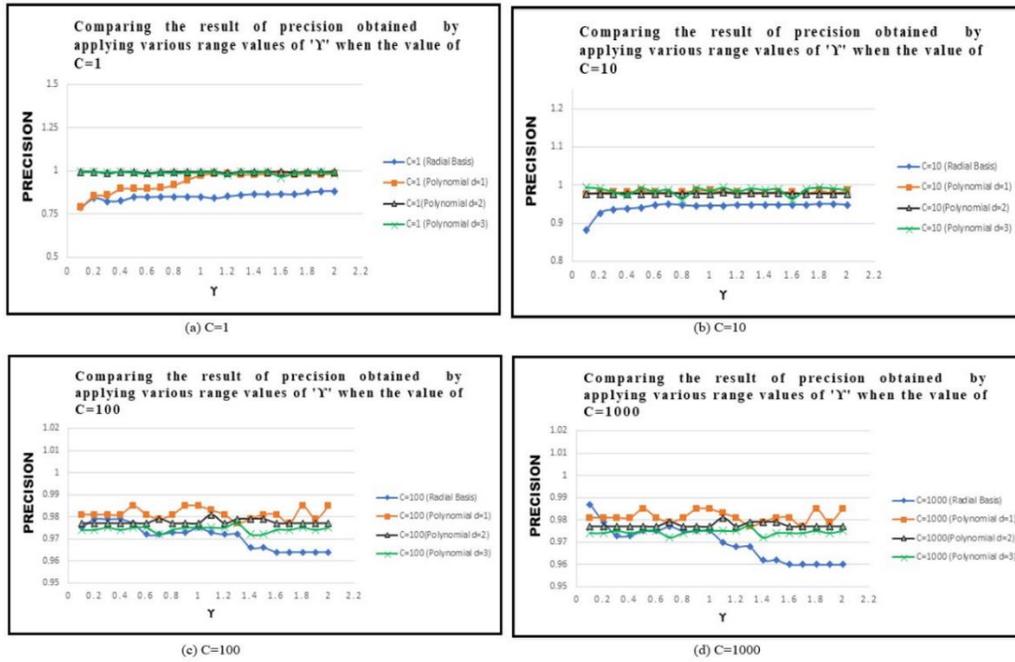


Figure 4. Results of precision measure after parameter setting experiment.

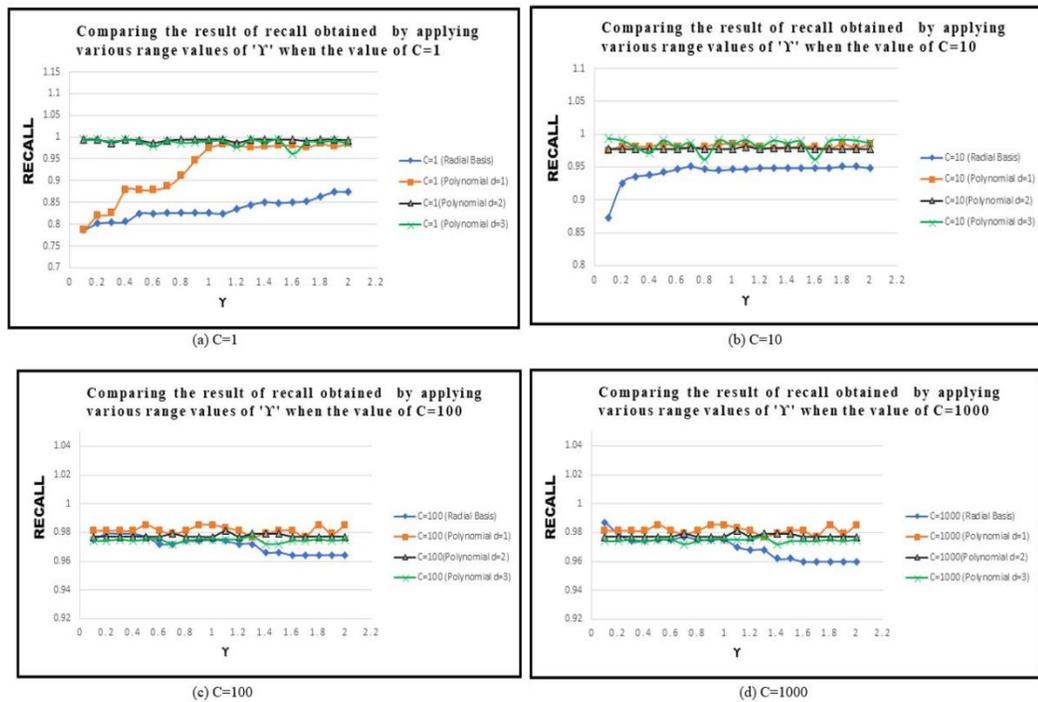
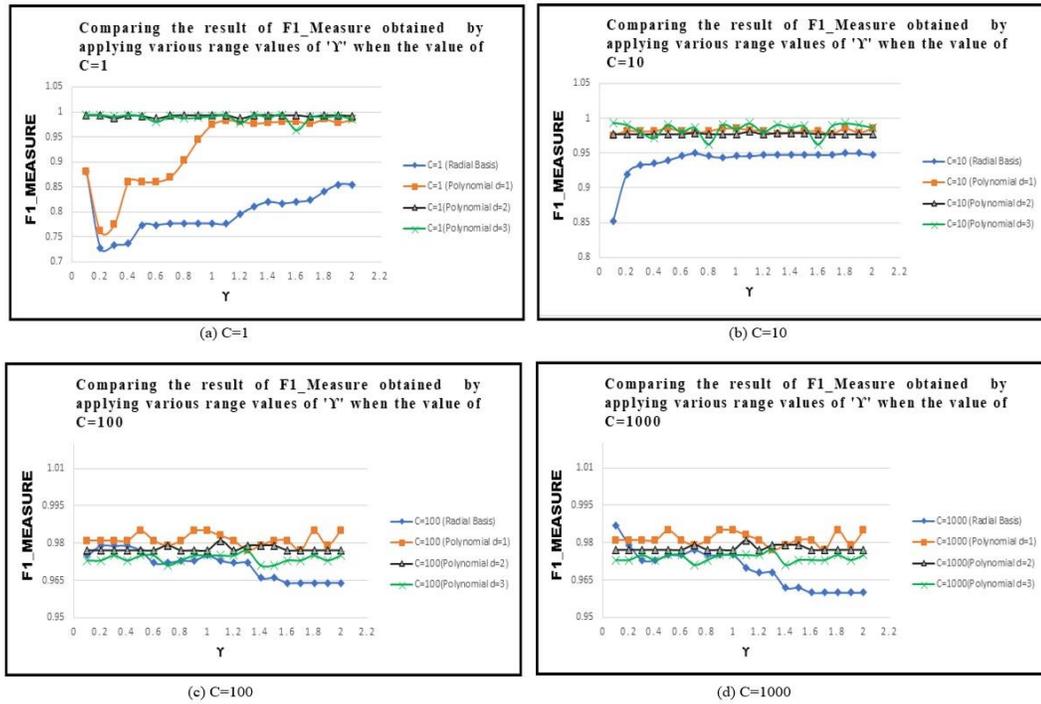


Figure 5. Results of recall measure after parameter setting experiment.



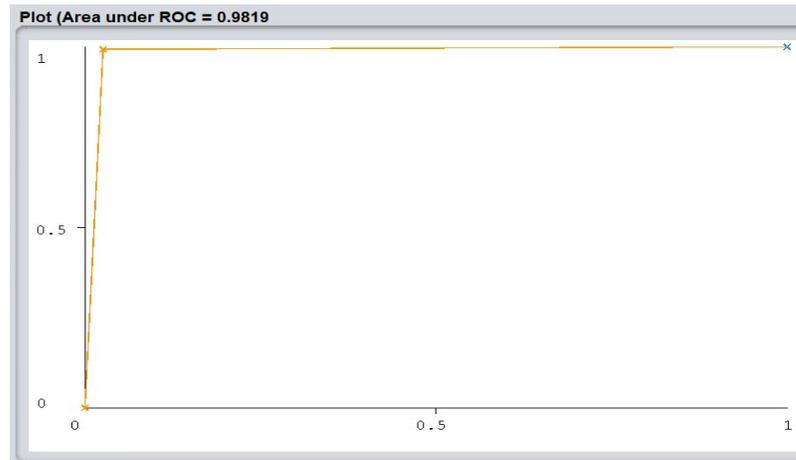
**Figure 6.** Results of F1\_measure metric after parameter setting experiment.

Table 5 illustrates the comparison table created based on the performance of LibSVM kernel functions. It was observed that LibSVM with the use of polynomial kernel function outperforms with an accuracy of 99.434%.

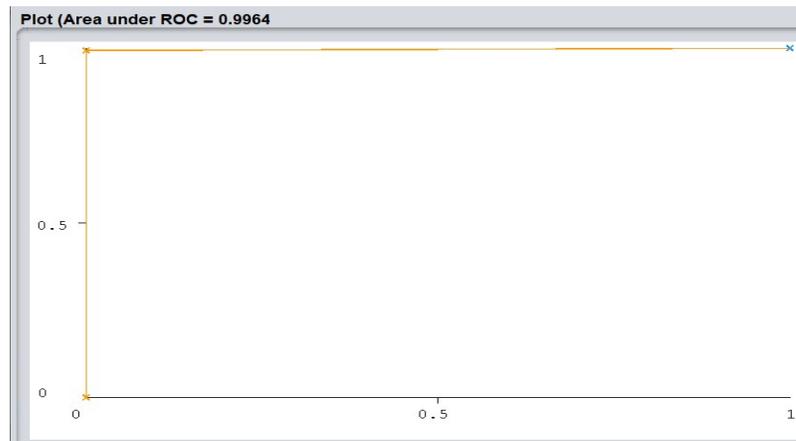
**Table 5.** Comparison table on the basis of performance measures of LibSVM kernel.

Classifiers		Performance Measures				
LibSVM	Kernel Function	Precision	Recall	F1_Measure	ROC	Accuracy (%)
	Radial Basis	0.987	0.987	0.987	0.982	98.679
	Polynomial	0.994	0.994	0.994	0.996	99.434

The Figure 7, illustrates the plot of Receiver Operating Characteristic Curve as mention in Table 5. It was clearly observed in the above plotted graphs that the Area under ROC evaluated in Polynomial Kernel is about 0.9964 is very close to 1.0 as compared to Radial Basis Kernel i.e. 0.9819.



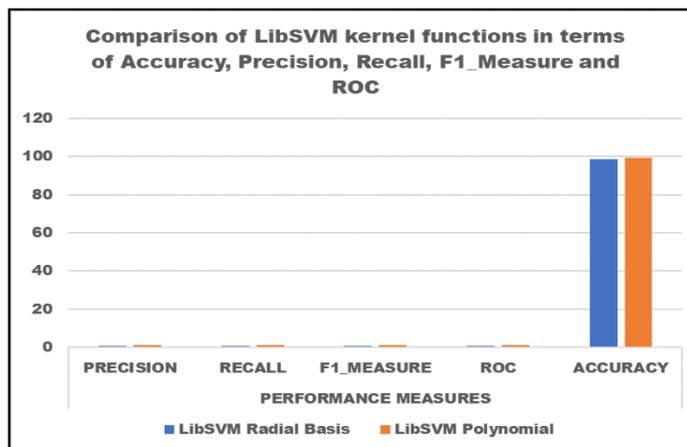
(a) ROC curve of Radial Basis Kernel



(b) ROC curve of Polynomial Kernel

**Figure 7.** ROC Curves of SVM models.

Figure 8 shows the comparison between LibSVM models based on performance metrics which represents that LibSVM with the use of Polynomial Kernel Function with the degree(d)=2 and degree(d)=3 attained maximum accuracy of 99.434% which is greater than Radial Basis function which achieved accuracy i.e., 98.679%. The experimental result concludes that the optimum model developed using LibSVM with the use of a Polynomial Kernel function gives an accuracy of 99.434% in predicting water quality.



**Figure 8.** Comparison of LibSVM models.

## 5. Conclusion

Predicting the quality of water before its use for drinking is the necessity of today's time. Based on the experiment carried out on this study and after the literature review, many conclusions can be drawn. Heavy tourism activity, unplanned construction, and improper solid waste disposal have somehow contributed to the deterioration of water quality. Nainital Lake is the only source of drinking water for the people here, so its planning and management is important. In this study for mapping of training samples from input space to higher dimensional space, LibSVM has been used with the use of two kernel function type and their results were compared. The experiment results demonstrate that LibSVM with polynomial kernel function achieved the best classification accuracy of 99.434%. In conclusion, SVM has proven its efficiency and effectiveness in predicting water quality of Nainital lake.

## Conflict of Interest

The authors confirm that there is no conflict of Interest to publish the paper in the journal.

## Acknowledgment

The authors thanks to the Uttarakhand Science Education and Research Centre (USERC), Dehradun, Uttarakhand, India and Jal Sansthan Nainital, Uttarakhand, India for their support.

## References

- Adnan, R.M., Yuan, X., Kisi, O., & Yuan, Y. (2017). Stream flow forecasting using artificial neural network and support vector machine models. *American Scientific Research Journal for Engineering, Technology, and Sciences*, 29(1), 286-294.
- Chang, C.C., & Lin, C.J. (2011). LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(27) 1-27. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>, 2.
- Dalakoti, H., Mishra, S., Chaudhary, M., & Singal, S.K. (2018). Appraisal of water quality in the lakes of Nainital district through numerical indices and multivariate statistics, India. *International Journal of River Basin Management*, 16(2), 219-229.

- Goutte, C., & Gaussier, E. (2005, March). A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In *Proceedings of the European Conference on Information Retrieval* (Vol. 27, pp. 345-359). Springer, Berlin, Heidelberg.
- Haghiabi, A.H., Nasrolahi, A.H., & Parsaie, A. (2018). Water quality prediction using machine learning methods. *Water Quality Research Journal*, 53(1), 3-13.
- Ji, X., Shang, X., Dahlgren, R.A., & Zhang, M. (2017). Prediction of dissolved oxygen concentration in hypoxic river systems using support vector machine: a case study of Wen-Rui Tang River, China. *Environmental Science and Pollution Research*, 24, 16062-16076. Doi: [org/10.1007/s11356-017-9243-7](https://doi.org/10.1007/s11356-017-9243-7).
- Kara, Y., Boyacioglu, M.A., & Baykan, Ö.K. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: the sample of the Istanbul Stock Exchange. *Expert Systems with Applications*, 38(5), 5311-5319.
- Lerios, J.L., & Villarica, M.V. (2019). Pattern extraction of water quality prediction using machine learning algorithms of water reservoir. *International Journal of Mechanical Engineering and Robotics Research*, 8(6), 992-997.
- Liao, Y., Xu, J.Y., & Wang, Z.W. (2011). A method of water quality assessment based on biomonitoring and multiclass support vector machine. *Procedia Environmental Sciences*, 10, Part A, 451-457.
- Liao, Y., Xu, J.Y., & Wang, Z.W. (2012). Application of biomonitoring and support vector machine in water quality assessment. *Journal of Zhejiang University Science B*, 13(4), 327-334.
- Nainital Lake image source: <https://www.postoast.com/nainital-lake>.
- Nollet, L.M.L. (2000). *Handbook of water analysis*. Marcel Dekker, New York.
- Pareek, R.K., Khan, A.S., Srivastva, P., & Roy, S. (2018). Assessment of physicochemical properties of Ghaggar river. *Journal Biological Chemical Chronicles*, 4(1), 19-25.
- Sarkar, A., & Pandey, P. (2015). River water quality modelling using artificial neural network technique. *Aquatic Procedia*, 4, 1070-1077.
- Sharma, M. (2014). Water quality assessment of the central Himalayan lake, Nainital. *Advances in Environmental Chemistry*, 2014, Article ID 473074, <https://doi.org/10.1155/2014/473074>.
- Sokolova, M., Japkowicz, N., Szpakowicz, S. (2006). Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. In: Sattar A., Kang B. (eds) *AI 2006: Advances in Artificial Intelligence*. AI 2006. Lecture Notes in Computer Science, vol 4304. Springer, Berlin, Heidelberg. DOI: [org/10.1007/11941439\\_114](https://doi.org/10.1007/11941439_114).
- Tadesse, M., Tsegaye, D., & Girma, G. (2018). Assessment of the level of some physico-chemical parameters and heavy metals of Rebu river in Oromia region, Ethiopia. *MOJ Biology and Medicine*, 3(4), 99-118.
- Tan, G., Yan, J., Gao, C., & Yang, S. (2012). Prediction of water quality time series data based on least squares support vector machine. *Procedia Engineering*, 31, 1194-1199.
- Tarmizi, A., Ahmed, A.N., & El-Shafie, A. (2014). Dissolved oxygen prediction using support vector machine in Terengganu River. *Middle-East Journal of Scientific Research*, 21(11), 2182-2188.
- Tong, S., & Koller, D. (2001). Support vector machine active learning with applications to text classification. *Journal of Machine Learning Research*, 2, 45-66.
- Vapnik, V., & Chervonenkis, A. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability & its Applications*, 16(2), 264-280.

Yahya, A.S.A., Ahmed, A.N., Othman, F.B., Ibrahim, R.K., Afan, H.A., El-Shafie, A., Fai, C.M., Hossain, M.S., Ehteram, M., & Elshafie, A. (2019). Water quality prediction model-based support vector machine model for ungauged river catchment under dual scenarios. *Water*, 11(6), 1231.

Zhang, L., Zou, Z., & Shan, W. (2017). Development of a method for comprehensive water quality forecasting and its application in Miyun reservoir of Beijing, China. *Journal of Environmental Sciences*, 56, 240-246. DOI: [org/10.1016/j.jes.2016.07.017](https://doi.org/10.1016/j.jes.2016.07.017).



Original content of this work is copyright ©International Journal of Mathematical, Engineering and Management Sciences. Uses under the Creative Commons Attribution 4.0 International (CC BY 4.0) license at <https://creativecommons.org/licenses/by/4.0/>