International Journal of Computer Science & Communication (ISSN: 0973-7391)
Volume 17 • Issue 2 pp. 299-303 March 2026-Sept 2026 www.csjournals.com

Comparative analysis of Spatio Temporal Water Quality using Machine Learning Models

Mamta[1], Dr. Sahil Saharan[2], Dr. Anupam Bhatia[3]

[1]MCA Student, [3]Asst. Prof., [2]Associate Professor
Deptt. of Comp. Sc. and Appl, Ch. Ranbir Singh University, Jind, Haryana, India

Abstract: It is noted that by rapid increment in the population the demand of the water is also increased. This study represents the comparative analysis of water quality using the four models LR,DT,RF and SVM. The model performance is measured using different techniques like confusion matrix ,accuracy, precision and recall. These results indicates that the ML achieves 90%+ accuracy with balanced precision and recall. These finding reveal that ML performs best and effective in accuracy.

*LR = Linear Regression , *DT= Decision Tree , *RF=Random Forest , * SVM = Support Vector Machine,*ML=ML,*ST-GCN=Spatio-Temporal Graph Convolutional Network.

1. Introduction

Water is the lifeblood of our earth. Approximately 71 percent of the earth surface is made up of water. Every small living things need water .So, its our responsibility to keep water safe. However, the highly nonlinear and complex spatio-temporal dependencies of river water pose significant challenges. saptio is the word which is the combination of space and Time. It is Difficult to predict the water quality in many areas .So , There is need of classify model of ML for prediction. As a long time , nature clean river itself. but now pollution increases at very high volume .Rivers like Ganga , yamuna are polluted at very high volume like increasing the phosphates and nitrites in water that make the water unsafe for many of the peoples. Also occurs when harmful substances such as sewage, industrial waste, chemicals, or microorganisms contaminate water bodies like rivers, lakes, and ground water. This causes many types of diseases in human being like cholera, typhoid fever and many more . That is very harmful for human being. In past few years ST-GCN model Plays an important role in prediction of water quality. Researcher tells us that ML and deep learning models are best for the prediction. We collected data from the government sites and after training and testing apply models on it...

1.1 Motivation

Scientists and researcher realizes that water changes the quality depending on season, rain and the location. To protect from the environment pollution scientists uses tool Water Quality Index(WQI) , that helps to know about the health of water like oxygen in water, how acidic it is and pH of it and tell it in single number that it is ”Excellent” ,”Poor” or ”Bad” that all. After using models and algorithms accuracy must be between 80 to 90 percent than we can say that the model is ”good”. WQI is high time consuming so we use ML models to check the quality of water. By using these models, we create a faster, cheaper, and more efficient way to guard our water resources and respond to pollution before it becomes a disaster. To keep people safe, we used to send someone to a river, take a bottle of water,

and bring it to a lab. By the time the lab results came back a week later, the pollution had already moved downstream. Plus, testing for 30 different chemicals is incredibly expensive and slow. As it is difficult to take testing tube so we are trying to make 3 4 keys like(Temperature, pH and solidity) so that we can easily check the water at a time which will be less expensive with the help of artificial intelligence or ML .

We uses ML models for our model that make the prediction model more effective and helpful. We can predict polluted water by its ph value,Turbidity and solidity in it. These high-precision tools serve as a foundation for intelligent, real-time monitoring in sustainable urban water management systems.

2. Literature Review

Recent years have seen a shift toward ML techniques for predictive water quality modeling. Xu et al. (2024) demonstrated that various ML models could accurately predict the Water Quality Index using only a small subset of key parameters. This significantly reduces laboratory testing costs and improves efficiency. To address spatial and temporal dependencies in water systems, deep learning architectures have been introduced. Huan et al. (2023) proposed a Spatio -Temporal Graph Convolutional Network (STGCN) for river monitoring stations. This model effectively captured relationships between different monitoring locations, leading to improved prediction accuracy

S. No./Ref.	Year	Authors	Proposed Model	Best Findings
1 [1]	2020	MOHAMED ABU HASHIM	WQI with PCA and CA	Identified 8–10 significant parameters (e.g., NH_4^+-N) explaining more than 70% variance in water quality.
2 [2]	2023	Huan et al.	STGCN	Captured spatial and temporal dependencies among monitoring stations, improving prediction accuracy.
3 [3]	2025	Huang et al.	Hybrid DL and ML framework	Achieved high-precision water quality forecasting by combining strengths of multiple models.
4 [4]	2018	Lin & Hwang	Orbital Resonance Analysis	Improved satellite observation efficiency, supporting large-scale environmental monitoring.
5 [5]	2016	Romanescu et al.	Spatio-temporal chemical analysis	Distinguished between anthropogenic and natural influences on river water chemistry over time.
6 [6]	2024	Xu et al.	Machine Learning models	Predicted WQI using fewer key parameters, reducing monitoring costs.
7 [7]	2018	Zelenakova et al.	Statistical trend analysis	Identified long-term seasonal trends in pollutant concentrations across river systems.

.

IJCSC 0973-7391
International Journal of Computer Science & Communication (ISSN: 0973-7391)
Volume 17 • Issue 2 pp. 299-303 March 2026-Sept 2026 www.csjournals.com

3. Methodolgy

3.1 Data description

We collected data from the government sites that have multiple rows and columns like humidity, turbidity, pH value, hardness, Turbidity etc.

This dataset contain different physiochemical properties of water that helps in analysis of water quality.

3.2 Handle missing values

This dataset contains various missing values that may effect the performance and accuracy of model. So we handle missing value by median imputation because it is robust to outliers. This is necessary before model training.

3.3 Feature Scaling

Feature Scaling using Standard Scaler from scikit learn library for normalizing the input features .These are use for algorithms that are sensitive like LR,SVM.

The scaling process increase the model performance.

3.4 Model Development

Models were used for comparative analysis of water quality :

I) LR Model: It is Supervised ML model that is used to predict categorical outcomes. This was applied to classify water quality condition based on physiochemical parameters. This predict whether a sample belong to particular class such as safe or unsafe class. This was selected because of its Simplicity , Fast training speed and ability to provide better results.

II) DT Classifier: This is used for classification task. This was used by splitting the dataset into smaller subsets using decision tree based feature values. This model divides the until the stopping criteria is satisfied. It is commonly based on Gini impurity .It handle non-linear relationships effectively and does not require heavy mathematical assumptions.

III) RF Classifier: It works by combining the multiple Decision tree to improve prediction accuracy and reduce over fitting. Instead on single tree it create multiple trees and produce the final prediction based on majority.

IV) SVM: Its main goal is to

Create a boundary line that separates the different group of data as clear as possible. It tries to keep maximum distance between two classes so that prediction become more accurate.

International Journal of Computer Science & Communication (ISSN: 0973-7391)
Volume 17 • Issue 2 pp. 299-303 March 2026-Sept 2026 www.csjournals.com

3.5 Training and Testing: Dataset is divided into Two mode. Training set have 70% and 30% is contained by validation and Testing. From the 30 % , 15% is for validation and 15% for test set.

3.6 Model Evaluation: Model evaluation is technique where we check the model's accuracy and performance using various terms or techniques like precision,recall,f1-score and other visualization terms on data.

Flow Chart:

4. Results and Discussions

We did a comparative analysis of four models(LR,DT , RF and SVM) in this work. To check the model accuracy, precision, recall and f1-score were used.

Model	Accuracy	Precision	Recall	F1-score
LR	0.88	0.84	0.90	0.95
DT	0.95	1.00	0.96	0.98
RF	0.92	0.89	0.95	0.95
SVM	0.80	0.83	0.91	0.95

The **DT** is one of the best model, achieving the highest overall Accuracy (0.95), perfect Precision (1.00), and a leading F1-score (0.98).

 The **RF** also demonstrates strong, balanced performance with an Accuracy of 0.92.

 In contrast, the **SVM** logs the lowest baseline metrics, trailing with an Accuracy of 0.80 and Precision of

0.83.

Conclusion and Future Work:

In this model we compared four models for comparative analysis. The accuracy increased form baseline model to transfer learning model. Results are evaluated from accuracy,precision,f1-score and recall. They were gave very good results but not that much that must be given.

So the limitation is that it can give better accuracy by applying different models and algorithms. That can classify results much better than that . We used only simple models not higher classify models in this work. And also not clear how decision are created.

In Future work , dataset must be change and more parameters can also be added . So higher classification models will be used that will classify the model more accurately .

References:-

[1] N. T. Giao, P. K. Anh, and H. T. H. Nhien, "Spatiotemporal analysis of surface water quality in Dong Thap province, Vietnam using water quality index and statistical approaches," *Water*, vol. 13, no. 3, p. 336, 2021.

[2] J. Huan, W. Liao, Y. Zheng, X. Xu, H. Zhang, and B. Shi, "A deep learning model with spatio-temporal graph convolutional networks for river water quality prediction," *Water Supply*, vol. 23, no. 7, pp. 2940–2953, 2023.

[3] T. Huang, K. Y. Chau, S. Zhan, and S.-H. Pan, "Hybrid deep learning and ML framework for high-precision water quality prediction in urban systems," *Environment, Development and Sustainability*, Jun. 2025.

[4] S.-F. Lin and C. Hwang, "Orbital resonances of Taiwan's Formosat-2 remote sensing satellite," *Acta Astronautica*, vol. 147, pp. 71–85, 2018.

[5] G. Romanescu *et al.*, "Spatio-temporal analysis of the water quality of the Ozana river," *Revista de Chimie (Bucharest)*, vol. 67, no. 1, pp. 42–46, 2016.

[6] J. Xu *et al.*, "Assessing and predicting water quality index with 6key water parameters by ML models in coastal cities, China," *Heliyon*, vol. 10, no. 14, p. e33695, 2024.

[7] M. Zelenakova *et al.*, "Spatio-temporal variations in water quality parameter trends in river waters," *Revista de Chimie (Bucharest)*, vol. 69, no. 10, pp. 2940–2947, 2018.