International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0527

e ISSN: 2584-2854
Volume: 03
Issue: 12 December 2025
Page No: 3332 - 3339

Water Quality Assessment of the River Cauvery: Prediction of Treatment Recommendations Using LR-GB Stacking Ensemble Model

Karishma L[1], Shyamala K[2]
[1]*Research Scholar, Dr. Ambedkar Government Arts College, Vyasarpadi, Chennai, India.*
[2]*Associate Professor, Dr. Ambedkar Government Arts College, Vyasarpadi, Chennai, India.*
Emails: *karishmal.research@gmail.com[1], shyamalakannan2000@gmail.com[2]*

Abstract

Machine learning techniques are increasingly being applied in environmental protection, particularly for analyzing the quality of air, water and soil. One of the most important problems in current environment is water pollution and it's our responsibility to protect one of our primordial elements of the earth. Many researchers are currently concentrating in preventing pollution by developing advanced monitoring systems, sustainable treatment methods and data-driven predictive models for early detection and control. The water quality parameter values were used to determine the contamination level and it helps to decide the treatment recommendations for the Cauvery river dataset. In this work, the LR-GB Stacking ensemble model was proposed and efficiently used to predict the future treatment recommendations for the Cauvery river stations. This model achieved higher prediction accuracy, precision and recall when compared with the other machine learning models.

Keywords: *Water Pollution, Cauvery river, machine Learning, stacking ensemble model, water parameter limit, treatment suggestion.*

1. Introduction

Water quality is essential for environmental sustainability, public health and economic development, yet urban expansion and industrial agriculture significantly threaten fresh water availability by degrading and depleting surface and groundwater resources [1]. Water pollution is a significant cause, resulting in numerous deaths annually due to waterborne diseases [2]. Assessing water quality is a complex process that involves identifying contaminants in water sources and determining whether it is safe for human consumption. In recent years, water pollution has become a serious problem affecting water quality. There are several factors that influence the quality of water in urban environments [3]. Certain exceeded level of water pollution leads to environmental consequences like eutrophication, oxygen depletion, toxic accumulation and human health impacts like waterborne diseases, chemical poisoning, bioaccumulation risks and it also leads to socio-economic Impacts. Therefore, to design a model that predicts and suggest water treatment suggestion was a most needed for our society to control water pollution, as well as to know which river station of Cauvery river was in danger zone and needs immediate water treatment [4]. Rising issues in water pollution and ecosystem management requires advanced analysis methods. With the growth of monitoring technologies and large-scale observations, Machine learning has become essential for effective analysis and prediction. Unlike conventional methods typically used in water studies machine learning based data-driven approaches are capable of handling complex and nonlinear problems effectively. It offers promising solutions for pollution control, water quality enhancement and the sustainable management of watershed ecosystems. Machine learning has also shown significant potential in advancing waste water treatment such as improving treatment learning [5]. This study proposed a stacked LR-GB base model. The Stacking ensemble model was reinforced using Logistic

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0527
e ISSN: 2584-2854
Volume: 03
Issue: 12 December 2025
Page No: 3332 - 3339

Regression and Gradient Boosting algorithms. It is efficiently used to predict the future treatment recommendations for the Cauvery river stations. This LR-GB stacking ensemble model was compared with six traditional machine learning models, including Random Forest (RF), AdaBoost (ADB), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naive Bayes (NB) and Neural Network (NN) to predict the treatment suggestion [6]. These mentioned models were trained and tested using a dataset of 26 parameters [7]. The performance of the LR-GB model was evaluated based on different Evaluation Metrics Shown in Table 1.

Table 1 Water Parameter Limits

WQ Parameter Name	WQ Parameter Limits
Ammonia	5.4
Hardness	298
Calcium	75
Magnesium	110
Sulphate	200
Sodium	182
TSS	300
TDS	1000
TFS	300
Phosphate	1.5
Boron	0.1
Potassium	29
BOD	3
Fluoride	1.5
DO	7.5
Nitrate	0.503
Total Coliform	100
Fecal Coliform	60
pH	8.7
Conductivity	400
Turbidity	5
Phenolphthalein Alkalinity	26
Total Alkalinity	200
Chloride	215
COD	24
TKN (Total Kjeldahl Nitrogen)	39

1.1. Treatment Suggestion Framework

The dataset utilized in this research was collected from the Cauvery River, covering 27 monitoring stations across Tamilnadu and Karnataka for the period of January to December 2023. Each water quality parameter in the dataset has a defined threshold limit as prescribed by the National and International Water Quality Standards (BIS, WHO). When these limits are exceeded, it is an indication of water quality deterioration due to different forms of pollution. To systematically analyze the nature of pollution and provide appropriate treatment recommendations, the parameters have been grouped into physical, chemical and biological categories. The categorization not only provides clarity in identifying the pollution source, but also enables the suggestion of relevant treatment suggestion.

1.2. Physical Parameters and Treatment

The physical parameters include Turbidity, Total Fixed Solids (TFS), Conductivity, Total Dissovled Solids (TDS), Total Suspended Solids (TSS), Hardness, Calcium and Magnesium. Exceedences in these parameters indicate pollution from suspended solids, inorganic salts and mineral particles. Such pollution generally originates from soil erosion, urban runoff or discharge of untreated effluents containing sediments and mineral residues. The Physical treatment processes typically involve screening, coagulation and sedimentation aimed at removing coarse solids and particulate matter. These steps serve as the preliminary stage in water treatment, creating a foundation for further chemical and biological treatment if required.

1.3. Chemical Parameters and Treatment

The chemical parameter consists of pH, Dissolved Oxygen (DO), Biological Oxygen Demand (BOD), Nitrate, Phenolphthalein Alkalinity, Total Alkalinity, Chloride, Chemical Oxygen Demand (COD), Total Kjeldahl Nitrogen (TKN), Ammonia, Sulphate, Sodium, Phosphate, Boron, Potassium and Fluoride. Exceedance in these values reflects pollution due to industrial discharges, nutrient enrichment, oxygen depletion, salinity and chemical residues. For example, elevated COD and BOD values indicate high organic and chemical load, while excess fluoride, nitrate and phosphate concentrations reflect

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0527
e ISSN: 2584-2854
Volume: 03
Issue: 12 December 2025
Page No: 3332 - 3339

industrial and agricultural inputs. To address such forms of pollution, chemical treatment methods are suggested. These include screening, coagulation, sedimentation and filtration. The objective of these processes is to neutralize acidity/alkalinity imbalances, reduce dissolved inorganic and organic contaminants and restore the chemical balance of water. These treatments are essential for reducing issues like eutrophication, chemical toxicity and oxygen imbalance.

1.4. Biological Parameters and Treatment

The Biological parameters include Total Coliform and Fecal Coliform, which are recognized as indicators of microbial contamination. Exceedance of these values, points to the presence of bacteria, viruses and pathogens. These are typically originating from domestic sewage, agricultural runoff or untreated wastewater. This form of contamination directly threatens public health and requires stringent treatment. In such cases, biological treatment methods are recommended. These methods integrate all processes of chemical treatment with an additional disinfection stage, such as chlorination, zonation and ultraviolet (UV) treatment to ensure the effective inactivation of pathogenic microorganisms. Disinfection is an essential step to achieve microbiologically safe water, particularly in river systems that serve as drinking water sources.

1.5. Decision for Treatment Suggestion

By classifying the parameters into these three distinct groups, the treatment suggestion framework provides a structured methodology for water quality management. This framework enables a direct linkage between parameter exceedance and the corresponding treatment process. For example, if

Table 2 Categorization of Water Parameters Based on Treatment Type

S. No	Treatment Type	Parameter	Reason For Classification	Suggested Treatment
1	Physical	Turbidity	Indicates suspended particles, silt, clay, debris	Sedimentation
2	Physical	TFS	Indicates inorganic fixed solids	Filtration
3	Physical	Conductivity	Indicates dissolved salts and ions	Filtration
4	Physical	TDS	Indicates dissolved solids and salts	Filtration
5	Physical	TSS	Indicates suspended solids	Sedimentation
6	Physical	Hardness	Indicates dissolved salts of Ca & Mg	Softening
7	Physical	Calcium	Indicates hardness due to calcium	Softening
8	Physical	Magnesium	Indicates hardness due to magnesium	Softening
9	Chemical	pH	Indicates acidity/alkalinity	Neutralization
10	Chemical	DO	Indicates oxygen balance	Aeration
11	Chemical	BOD	Indicates organic load	Filtration
12	Chemical	Nitrate	Indicates nutrient pollution	Ion Exchange
13	Chemical	Phenolphthalein Alkalinity	Indicates carbonate alkalinity	Neutralization
14	Chemical	Total Alkalinity	Indicates buffering capacity	Neutralization
15	Chemical	Chloride	Indicates salinity/industrial input	Filtration
16	Chemical	COD	Indicates chemical pollution	Coagulation

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0527
e ISSN: 2584-2854
Volume: 03
Issue: 12 December 2025
Page No: 3332 - 3339

17	Chemical	TKN	Indicates organic nitrogen content	Filtration
18	Chemical	Ammonia	Indicates sewage/organic pollution	Chlorination
19	Chemical	Sulphate	Indicates mineral salt pollution	Precipitation
20	Chemical	Sodium	Indicates salinity	Ion Exchange
21	Chemical	Phosphate	Indicates nutrient pollution	Chemical Precipitation
22	Chemical	Boron	Indicates industrial/irrigation input	Ion Exchange
23	Chemical	Potassium	Indicates agricultural input	Ion Exchange
24	Chemical	Fluoride	Indicates excess minerals	Adsorption
25	Biological	Total Coliform	Indicates microbial contamination	Disinfection
26	Biological	Fecal Coliform	Indicates fecal pollution	Disinfection

elevated turbidity and hardness are observed, physical treatment will be prioritized, whereas high COD and nitrate concentrations will necessitate chemical treatment. Similarly, detection of coliforms mandates biological treatment with disinfection. This structured approach ensures that the suggested treatment measures are evidence-based, targeted and efficient by improving the applicability of machine learning. Table 2. Presents the classification of water quality parameters based on their treatment type, reason for classification and suggested treatment method. Each parameter is categorized as physical, chemical or biological depending on its effect on water quality. The suggested treatment is determined according to the parameter exceedances of permissible limit to ensure appropriate water quality management Shown in Figure 1.

2. Materials and Methods

2.1. About Cauvery River

In this study, the dataset utilized has been derived from the Cauvery River, one of the most prominent and lifeline rivers of South India, flowing through the states of Karnataka and Tamilnadu. The Cauvery, often referred to as the “Ganga of the South” holds immense ecological, cultural and economic significance, serving as a major source of irrigation, drinking water, hydroelectric power and supporting a rich aquatic biodiversity. The river is also regarded as sacred and plays a vital role in sustaining agriculture and livelihoods across its basin.

Figure 1 Cauvery River Map

2.2. About Dataset

For this study, the Cauvery River water quality dataset for the year 2023 was obtained from the Tamilnadu Pollution Control Board (TNPCB) [8] which systematically monitors the river at various locations to assess pollution levels and environmental sustainability. The dataset comprises observations from 27 monitoring stations, capturing measurements across 26 critical water quality parameters including Dissolved Oxygen (DO), pH, Conductivity (COND), Biochemical Oxygen Demand (BOD), Nitrate, Fecal Coliform (FC), Total Coliform (TC), Turbidity, Phenolphthalein Alkalinity (PA), Total Alkalinity (TA), Chloride, Chemical Oxygen Demand (COD), Total Kjeldahl Nitrogen (TKN), Ammonia, Hardness, Calcium, Magnesium, Sulphate, Sodium, Total Dissolved Solids (TDS), Total Fixed Solids (TFS), Total Suspended Solids (TSS),

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0527
e ISSN: 2584-2854
Volume: 03
Issue: 12 December 2025
Page No: 3332 - 3339

Orthophosphate, Boron, Potassium and Fluoride. Altogether, the dataset consists of 325 records with 26 feature variables, providing a comprehensive representation of the physio-chemical and biological status of the Cauvery river. The diverse set of parameters and wide geographical coverage across multiple stations make this dataset highly valuable for applying machine learning techniques to analyze pollution trends, identify key influencing factors and the prediction of future water quality scenarios.

2.3. Ensemble Learning Methodology

In this study, ensemble learning was proposed to understand the dataset accurately, to enhance the prediction accuracy and reliability of the treatment suggestion model for river water quality analysis. Ensemble methods are based on the concept of combining the predictive capabilities of multiple base learners to obtain a more robust and generalized model. Among the various ensemble approaches, stacking was selected because it allows integration of diverse machine learning algorithms, enabling the model to capture both linear and nonlinear relationships among the input water quality parameters [9][10][11]. The stacking ensemble was implemented using the Orange Data mining tool, which provides a visual workflow-based environment for model design, training and evaluation. For the stacking process, Logistic Regression was chosen for its simplicity, strong performance, interpretability on linearly separable patterns, while Gradient Boosting was employed for its ability to capture complex non-linear relationships. Combining these two models in a stacking ensemble enhances overall performance and results in higher generalization capability. The meta-learner was responsible for learning in the optimal way to combine the base model outputs, thereby refining the overall prediction. Orange's stacking widget was configured to automatically partition the dataset for training and testing, ensuring unbiased evaluation. Cross-validation techniques were applied to validate the ensemble's stability and prevent overfitting.

3. Results and Discussions

The Ensemble stacking model achieved approximately 3% higher accuracy than the standalone models, indicating that the combined approach was able to capture additional patterns within the dataset than the single models. The improvement confirms that the ensemble learning effectively reduces model variance.

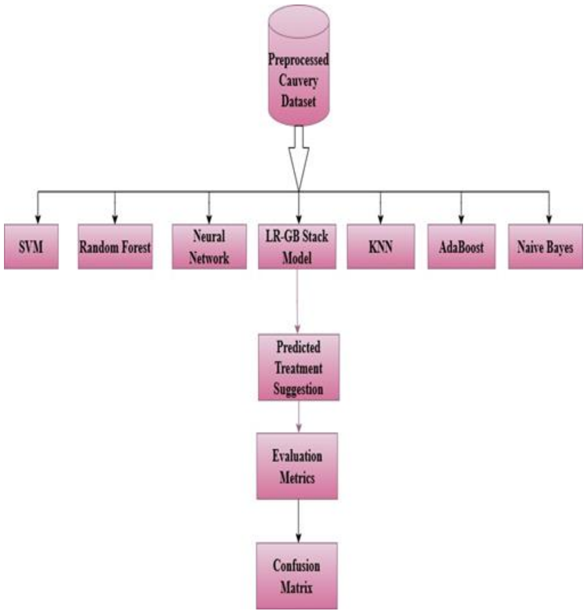

Figure 2 **Workflow of this Study**

This methodology, thus establishes stacking as a suitable approach for water quality-based treatment suggestion combining the interpretability of Logistic Regression with the adaptive learning capacity of Gradient Boosting for improved environmental decision making. Figure 2. Shows the architecture of the workflow of this study. The base learners used in this research included seven supervised machine learning models such as Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Neural Network (NN), AdaBoost Naïve Bayes and LR-GB stacking model. Each model was trained on the preprocessed Cauvery River dataset to predict the treatment type physical, chemical or biological based on the observed water quality parameters. The enhanced accuracy of the stacking model reflects its ability to provide more reliable treatment suggestions based on parameter exceedance levels. The meta-learning layer successfully synthesized the outputs of the base models, thereby improving classification precision and reducing prediction uncertainty. Such

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0527
e ISSN: 2584-2854
Volume: 03
Issue: 12 December 2025
Page No: 3332 - 3339

improvement is especially valuable in environmental studies, where accurate identification of the required treatment method directly supports effective pollution management.

Table 3 Evaluation Metrics

Model	AUC	Accuracy	F1	Precision	Recall
RF	99.7	96.3	96.2	96.4	96.3
NN	99.3	94.6	94.7	95	94.6
AdaBoost	94.5	94.6	94.4	94.5	94.6
SVM	99.5	93.9	93.5	93.9	93.9
kNN	98.4	91.2	89.1	92	91.2
NB	98.9	87.5	88.3	91.7	87.5
LR-GB	99.2	98	98	98.1	98

Table 3. represents the performance comparison of various machine learning models based on their evaluation metrics demonstrates that the LR-GB Stacking model outperforms all other algorithms in predicting water quality. It achieved the highest accuracy of 98%, an AUC of 99.2 and a Matthews Correlation Coefficient (MCC) of 96.4, indicating excellent model reliability and generalization. The Random Forest model also performed strongly, with an accuracy of 96.3% and F1-score of 96.2%, showing its robustness in handling complex, nonlinear data patterns. The Neural Network and Adaboost models followed closely with comparable accuracies of 94.6%, demonstrating stable predictive ability across multiple metrics. Although the Support Vector Machine (SVM) achieved a high AUC of 99.5, its overall accuracy and F1-score (93.9% and 93.5) were slightly lower than ensemble models, reflecting moderate sensitivity to class variations. The K-Nearest Neighbor (KNN) and Naïve Bayes models exhibited relatively lower accuracies of 91.2% and 87.5% respectively, suggesting limitations in handling diverse feature distributions and complex relationships among parameters. Overall, the comparative evaluation clearly indicates that the ensemble-based models particularly the Stacking Model (LR-GB) provide superior performance and greater consistency across all key metrics, making them highly suitable for reliable water quality prediction and classification tasks. These findings confirm that ensemble learning is a powerful approach for water quality analysis, offering improved accuracy and robustness for treatment recommendation systems in river monitoring applications.

4. Error Analysis

The error analysis of the Stacking Model (LR-GB) shows that the model achieved an overall high classification accuracy, with the majority of samples correctly predicted across all classes. As seen from the confusion matrix, Classes 0 and 1 were predicted with 100% accuracy, indicating the model's strong capability in identifying these water quality levels without any misclassification. Class 2 also demonstrated excellent performance with 98.4% correct predictions, though a small proportion (1.6%) of samples were incorrectly classified as Class 1. Similarity, Class 3 achieved an accuracy of 88.2% with minor misclassifications of 11.8% into Class 2, suggesting that the boundary between these two pollution levels is slightly overlapping. For Class 4, 96% of the samples were correctly identified, while 4% were misclassified into Class 2. These minimal misclassification rates can be attributed to the close numerical ranges of certain water quality parameters such as BOD, nitrate and turbidity in borderline cases. Overall, the model exhibits high reliability and generalization capability with misclassifications concentrated mainly between adjacent classes, which is acceptable in environmental data where parameter variations are gradual rather than discrete Shown in Figure 3 - 9.

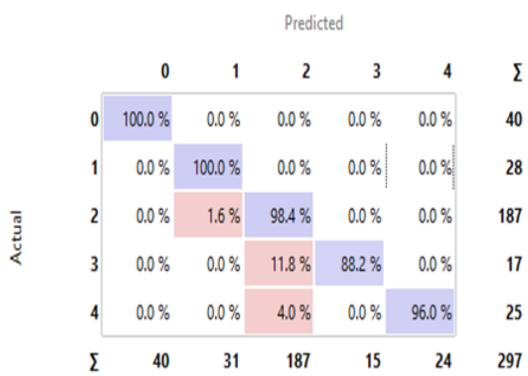

Figure 3 Confusion Matrix of Stacking LR-GB

e ISSN: 2584-2854
Volume: 03
Issue: 12 December 2025
Page No: 3332 - 3339

Model

Actual \ Predicted	0	1	2	3	4	Σ
0	100.0 %	0.0 %	0.0 %	0.0 %	0.0 %	40
1	0.0 %	92.9 %	7.1 %	0.0 %	0.0 %	28
2	0.0 %	1.1 %	98.4 %	0.0 %	0.5 %	187
3	0.0 %	0.0 %	29.4 %	64.7 %	5.9 %	17
4	0.0 %	0.0 %	0.0 %	0.0 %	100.0 %	25
Σ	40	28	191	11	27	297

FIGURE 4 Confusion Matrix of Random Forest

Actual \ Predicted	0	1	2	3	4	Σ
0	97.5 %	0.0 %	0.0 %	0.0 %	2.5 %	40
1	0.0 %	82.1 %	10.7 %	0.0 %	7.1 %	28
2	0.0 %	1.6 %	96.3 %	0.0 %	2.1 %	187
3	0.0 %	0.0 %	5.9 %	94.1 %	0.0 %	17
4	0.0 %	8.0 %	0.0 %	0.0 %	92.0 %	25
Σ	39	28	184	16	30	297

Figure 5 Confusion matrix of Neural Network

Actual \ Predicted	0	1	2	3	4	Σ
0	100.0 %	0.0 %	0.0 %	0.0 %	0.0 %	40
1	0.0 %	85.7 %	14.3 %	0.0 %	0.0 %	28
2	0.0 %	1.6 %	97.3 %	0.5 %	0.5 %	187
3	0.0 %	5.9 %	29.4 %	58.8 %	5.9 %	17
4	0.0 %	0.0 %	0.0 %	0.0 %	100.0 %	25
Σ	40	28	191	11	27	297

Figure 6 Confusion matrix of Adaboost

Actual \ Predicted	0	1	2	3	4	Σ
0	100.0 %	0.0 %	0.0 %	0.0 %	0.0 %	40
1	0.0 %	92.9 %	3.6 %	0.0 %	3.6 %	28
2	0.0 %	2.7 %	95.7 %	0.0 %	1.6 %	187
3	0.0 %	5.9 %	82.4 %	5.9 %	5.9 %	17
4	0.0 %	0.0 %	0.0 %	0.0 %	100.0 %	25
Σ	40	32	194	1	30	297

Figure 7 Confusion matrix of KNN

Actual \ Predicted	0	1	2	3	4	Σ
0	100.0 %	0.0 %	0.0 %	0.0 %	0.0 %	40
1	0.0 %	92.9 %	0.0 %	0.0 %	7.1 %	28
2	0.0 %	7.5 %	81.3 %	3.7 %	7.5 %	187
3	0.0 %	0.0 %	0.0 %	100.0 %	0.0 %	17
4	0.0 %	0.0 %	0.0 %	0.0 %	100.0 %	25
Σ	40	40	152	24	41	297

Figure 8 Confusion matrix of Naïve Bayes

Actual \ Predicted	0	1	2	3	4	Σ
0	100.0 %	0.0 %	0.0 %	0.0 %	0.0 %	40
1	0.0 %	57.1 %	39.3 %	0.0 %	3.6 %	28
2	0.0 %	0.5 %	99.5 %	0.0 %	0.0 %	187
3	0.0 %	0.0 %	17.6 %	82.4 %	0.0 %	17
4	0.0 %	4.0 %	4.0 %	0.0 %	92.0 %	25
Σ	40	18	201	14	24	297

Figure 9 Confusion matrix of SVM

5. Limitations

The analysis was conducted using Cauvery River water quality data from a single year 2023, which may not capture seasonal or long-term variations of the pollution patterns. Additionally, the dataset had limited representation for some pollution categories, which could affect the model's generalization. The exclusion of certain biological and environmental factors also restricts the comprehensiveness of this study. Future studies incorporating multi-year datasets and additional parameters could provide a more robust and reliable evaluation of water quality trends. Furthermore, external environmental factors such as rainfall, temperature and land-use changes were not explicitly integrated into the model. Incorporating these dynamic variables in future research could enhance the robustness and adaptability of the predictive framework for real-time water quality monitoring and treatment planning.

6. Future Scope

Future research can extend this work by incorporating multi-year datasets to capture seasonal and climatic variations that influence river water quality. The inclusion of additional physical, chemical and biological parameters would enhance the model's capability to represent complex interactions within aquatic systems. Integrating real-time data collection through IOT based sensors and remote sensing technologies could further improve prediction accuracy and support timely decision-making for water treatment and pollution control. Moreover, exploring deep learning and hybrid ensemble techniques may provide more adaptive and scalable models for large-scale water quality management across different river basins. This study addresses a critical research gap in water quality monitoring, specifically within the Cauvery river basin, where substantial contamination poses significant risks to both human health and aquatic ecosystems.

References

[1]. Choudhary R, Kumar A, C P, Naik MM, Choudhury M, Khan NA. Predicting water quality index using stacked ensemble regression and SHAP based explainable artificial intelligence. Sci Rep. 2025 Aug 24;15(1):31139. doi: 10.1038/s41598-025-09463-4. PMID: 40850991; PMCID: PMC12375761.

[2]. Vellingiri J et al 2024 Environ. Res. Commun. 6 075025 DOI 10.1088/2515-7620/ad6061 AO-SVM: a machine learning model for predicting water quality in the Cauvery river.

[3]. Muldoon, Conor, et al. "Mining Explainable Predictive Features for Water Quality Management." arXiv preprint arXiv:2212.04419 (2022).

[4]. Azrour, M., Mabrouki, J., Fattah, G. et al. Machine learning algorithms for efficient water quality prediction. Model. Earth Syst. Environ. 8, 2793–2801 (2022). https://doi.org/10.1007/s40808-021-01266-6

[5]. Nagalapalli Satish, JagadeeshAnmala et al. A stacking ANN ensemble model of ML models for stream water quality prediction of Godavari River. Ecological Informatics,Volume 80,2024,102500,ISSN 15749541,https://doi.org/10.1016/j.ecoinf.2024.102500.

[6]. Mengyuan Zhu, Jiawei Wang et al. A review of the application of machine learning in water quality evaluation, Eco-Environment & Health, Volume 1, Issue 2, 2022, Pages 107-116, ISSN 2772-9850, https://doi.org/10.1016/j.eehl.2022.06.001

[7]. Alqahtani A, Shah MI, Aldrees A, Javed MF. Comparative Assessment of Individual and Ensemble Machine Learning Models for Efficient Analysis of River Water Quality. Sustainability. 2022; 14(3):1183. https://doi.org/10.3390/su14031183.

[8]. https://tnpcb.gov.in/

[9]. W. Jiang, Z. Chen, Y. Xiang, D. Shao, L. Ma and J. Zhang, "SSEM: A Novel Self-Adaptive Stacking Ensemble Model for Classification," in IEEE Access, vol. 7, pp. 120337-120349, 2019, doi: 10.1109/ACCESS.2019.2933262.

[10]. Lu, M., Hou, Q., Qin, S., Zhou, L., Hua, D., Wang, X., & Cheng, L. (2023). A Stacking Ensemble Model of Various Machine Learning Models for Daily Runoff Forecasting. Water, 15(7), 1265. https://doi.org/10.3390/w15071265

[11]. Gupta, Aditya, Vibha Jain, and Amritpal Singh. "Stacking ensemble-based intelligent machine learning model for predicting post-COVID-19 complications." New Generation Computing 40.4 (2022): 987-1007.