LITERATURE SURVEY

TOPIC: EFFICIENT WATER QUALITY ANALYSIS AND PREDICTION USING MACHINE LEARNING

TEAM MEMBERS:

- SUBIKSHA DEVI M S
- DEEPTHY SHIVANI S
- DIVYA K
- SOUNDARYA M

S.NO	Paper Title & Author	Paper Source	Year	Methodology Used	Advantages	Disadvantages
1	WaterNet: A Network for Monitoring and Assessing Water Quality for Drinking and Irrigation Purposes OLASUPO O. AJAYI , ANTOINE B. BAGULA , HLONIPHANI C. MALULEKE, ZAHEED GAFFOOR , NEBO JOVANOVIC, AND KEVIN C. PIETERSEN	IEEE	2022	Random Forest (RF) Logistic Regression (LR) Support Vector Machine (SVM) low power long range protocol for data transmission	Coverage in remote areas Easy to deploy Cheaper	Susceptible to destruction Requires erecting long antennas
2	Quality Risk Analysis for Sustainable Smart Water Supply Using Data Perception Di Wu , Hao Wang , Hadi Mohammed, Razak Seidu	IEEE	2020	Adaptive Frequency Analysis	Integrates indicators, geography and time domains Cost less data analysis	This method is difficult to use for the data sets which do not have significant frequency effects.

3	Proposition of New Ensemble Data-Intelligence Models for Surface Water Quality Prediction ALI OMRAN AL-SULTTANI, MUSTAFA AL-MUKHTAR, ALI B. ROOMI AITAZAZ AHSAN FAROOQUE , KHALED MOHAMED KHEDHER , AND ZAHER MUNDHER YASEEN	IEEE	2021	Quantile regression forest , Random Forest, radial support vector machine, Stochastic Gradient Boosting, Gradient Boosting Machines Genetic Algorithm Principal Components Analysis	The model exhibited less approximation of the input parameters that are extremely for the catchments with less environmental or ecological information.	Tuning the internal parameters of the SVM model with other advanced non-linear function
4	Hybrid Predictive Model for Water Quality Monitoring Based on Sentinel-2A L1C Data GEHAD HASSAN, MOHAMED E. GOHER , MASOUD E. SHAHEEN SHEREEN A. TAIE	IEEE	2021	Binary Whale Optimization Algorithm Artificial Neural Network (BWOA-ANN)	The proposed model was found to outperform the ANN with an R2 value higher by 40% and 52% for the optical and non-optical parameters	Harder to train
5	Real-Time Monitoring and Prediction of Water Quality Parameters and Algae Concentrations using Microbial Potentiometric Sensor Signals and Machine Learning Tools Daniel Saboe, Hamidreza	Elsevier	2020	Long short-term memory (LSTM)	technology can be used to supervise the reliability and performance of other monitoring tools	water quality parameters stem from the fact that a certain amount of data needs to be generated to train the ML/AI tools.

	Ghasemi, Ming Ming Gao, Mirjana Samardzic, Kiril D. Hristovski, Dragan Boscovic, Scott R. Burge, Russell G. Burge, David A. Hoffman					
6	River Water Quality Analysis and Prediction using GBM Al-Akir Nayan,Muhammad Golam Kibria,Md. Obaidur Rahman and Joyeta Saha	IEEE	2020	Gradient boosting model	The model is capable of producing realistic prediction even when the changes made in the dataset.	For quick and reliable on-site usage, IOT should be used.
7	Water Quality Prediction Method Based on LSTM-BP Huimin Jia, Xiaofeng Zhou	IEEE	2020	Neural network,Back propagation, Long short term memory	Integrated LSTM-BP has better prediction performance of time-series than the single LSTM and BP.	Back propagation method's convergence speed is slow.
8	Efficient Water Quality Prediction Using Supervised Machine Learning Umair Ahmed , Rafia Mumtaz, Hirra Anwar, Asad A. Shah, Rabia Irfan and José García-Nieto	MDPI	2019	Multiple Linear Regression, Polynomial Regression, Random Forest,Gradient Boosting Algorithm, Support Vector Machines, Ridge Regression, Lasso Regression, Elastic Net Regression, Neural net/Multi-layer perceptron,	Gradient boosting and polynomial regression performed better in predicting WQI, whereas MLP performed better in predicting WQC.	Integrating in a large-scale IoT-based online monitoring system by using only the sensors of the required parameters, it will predict the water quality immediately based on the real-time data fed

				Gaussian Naive Bayes, Logistic Regression, Stochastic Gradient Descent, K Nearest Neighbor, Decision Tree, Bagging Classifier.		from the IoT system.
9	Prediction of Water Level and Water Quality Using a CNN-LSTM Combined Deep Learning Approach Sang-Soo Baek, Jongcheol Pyo, Jong Ahn Chun	MDPI	2020	Convolutional Neural Network (CNN), Long Short-Term Memory(LSTM)	It combined two deep learning models CNN and LSTM. CNN was used to stimulate the water level while LSTM was used to stimulate the concentration of the pollutants it also represented the different temporal variations of each pollutant type.	Three pollutants were only used in this study. Further study to develop deep learning models so that more pollutants included like chlorophyll, algae, dissolved oxygen, and fecal bacteria can be simulated
10	A novel approach for water quality classification based on the integration of deep learning and feature extraction techniques Smail Dilmi ,	IEEE	2021	Long Short term memory, Recurrent neural network, Support Vector machine, Principle Component Analysis,Linear Discriminant	The clustering feature using LDA and ICA was better than PCA.The Random-Holdout technique is a effective method for	The chemical parameters cannot be measured continuously.

	Mohamed Ladjal			Analysis, Independent Discriminant Analysis	estimating the performance of time series prediction models. Integration of LSTM with LDA, and LSTM with ICA gives 99.72% accuracy .	
11	Predictive Models for River Water Quality using Machine Learning and Big Data Techniques - A Survey	IEEE	2021	Time series Analysis for WQP , Autoregressive (AR) , Autoregressive Moving Average (ARMA), Linear Regression, Logistic Regression, Artificial Neural Network, Recurrent Neural Networks (RNNs) , Long Short-Term Memory Networks (LSTMs) ,	High level hybrid approach by integrating remote sensing, IOT, Artificial intelligence and GIS can be developed to find the dynamic flow of water, time series on both single and multiple input, dynamic standard effect on various levels of water and to develop a best model for evaluation and monitoring	major problem is smaller data size which will overfit the model while training data with algorithms like ANN, SVM, decision tree and affects the result badly
12	Predicting and Analyzing Water Quality using Machine Learning: A Comprehensive Model	IEEE	2016	Artificial Intelligence (AI) techniques like Bayesian Networks (BN), Artificial Neural Networks (ANN), Neuro-Fuzzy	A suitable hybrid of multiple models for forecasting and prediction gives better results than using a single	More work needs to be done in terms of effectiveness , reliability, accuracy as well as usability of

				Inference, Support Vector Regression (SVR), Decision Support System (DSS) and Auto-Regressive Moving Average (ARMA)	one. Different methodologies have been proposed and applied for analysis and monitoring of water quality as well as time series analysis. The methodologies range from statistical techniques, visual modeling, analysis algorithms and prediction algorithms and decision making.	the current water quality management methodologies.Besides further improvements in prediction accuracy, there needs to be a more user-centric approach
13	Water Quality Prediction for Smart Aquaculture Using Hybrid Deep Learning Models	IEEE	2022	LSTM DL NEURAL NETWORK, GRU DL NEURAL NETWORK, HYBRID CNN LSTM/GRU DL NEURAL NETWORK,	combining RNN and CNN, making a new hybrid model for WQP having the advantages of both models.the CNN-LSTM hybrid model provides significant improvement in prediction accuracy as well as computation time compared to the baseline DL models.	RNN has a vanishing gradient problem while training with large data and affects the learning from large datasets
14	Efficient Prediction of Water Quality	atlantis-press	2021	Neural Network,	The highly useful	The model has trained,

	Index (WQI) Using Machine Learning Algorithms			Random Forest, Multinomial Logistic Regression, Support Vector Machine, and Bagged Tree Model	generated models are utilized to identify the mode of the water contamination process, assisting decision-makers in making the appropriate decisions at the appropriate time	validated, and has been continually validated using monthly data.so n data Normalizatio n and then Min-Max Normalizatio n, which is a crucial procedure in data analysis, to improve data quality for all of the datasets
15	Designing Efficient and Sustainable Predictions of Water Quality Indexes at the Regional Scale Using Machine Learning Algorithms	MDPI	2022	Decision Trees (DT), K-Nearest Neighbors (KNN), Discriminants Analysis (DA), Support Vector Machine (SVM), and Ensemble Trees (ET)	The SVM model is a simple and effective empirical model to simulate water quality, and the method presented in this work is sufficiently general to be applied to a wide range of arid areas.	The techniques that are Used here are reliable tools,but they can be costly and time-consuming.
16	Water Quality Index Prediction for Improvement of Treatment Processes on Drinking Water Treatment Plant Goran Volf, Ivana	MDPI	2022	Rule-based regression models for numeric prediction. Four models were built to predict WQI with a time step of	Obtained prediction models can help with the optimization and management of treatment processes	correlation coefficients for the proposed models decrease as the number of prediction days

	Sušanj Čule, Elvis Žic and Sonja Zorko			one, five, ten, and fifteen days in advance	especially during the summer months when the quality of raw water in the Butoniga reservoir is the worst	increases
17	Simple Prediction of an Ecosystem-SpecificWater Quality Index and the Water Quality Classification of a Highly Polluted River through Supervised Machine Learning Alberto Fernández del Castillo, Carlos Yebra-Montes, Marycarmen Verduzco Garibay, José de Anda, Alejandro Garcia-Gonzalez 4, and Misael Sebastián Gradilla-Hernández	MDPI	2022	Multiple linear regression and generalized additive models for regression tasks, and logistic regression and linear discriminant analysis for classification tasks	Proposed model involved less water quality parameters to ease the time and costs associated with water quality monitoring and, consequently, to extend the number of sampling points that are regularly monitored in this large basin. The models applied in this work are easy to reproduce for water quality evaluation due to their relatively simple structure and practical programming.	It is important to maintain the monitoring of the 20 current sampling points with the 17 original parameters as the water quality trends could change over time, especially if corrective actions are applied to reduce contaminatio n
18	Designing Efficient and Sustainable	MDPI	2022	Multiple classification	The method presented in	Comparative ly it gives

	Predictions of Water Quality Indexes at the Regional Scale Using Machine Learning Algorithms Abdessamed Derdour, Antonio Jodar-Abellan, Miguel Ángel Pardo, Sherif S. M. Ghoneim and Enas E. Hussein			techniques such as Decision Trees (DT), K-Nearest Neighbors (KNN), Discriminants Analysis (DA), Support Vector Machine (SVM), and Ensemble Trees (ET) we're used and compared. Support Vector Machine (SVM) algorithms classify groundwater quality with high accuracy (95.4%) with standardized data and accuracy of 88.88% for raw data.	this work is sufficiently general to be applied to a wide range of arid areas.	lower accuracy of 88.88% for raw data.
19	Prediction of the groundwater quality index through machine learning in Western Middle Cheliff plain in North Algeria Yamina Elmeddahi, Ragab Ragab	ResearchGate	2022	Neuro-Sensitivity Analysis (NSA) and principal component analysis (PCA) techniques for features extraction and selectionselection. Proposed 3 machine learning models with different parameter combinations: multilayer perceptron	Results suggest that NSA-MLPNN model could be a robust and cost-effective model for classifying groundwater quality levels. Another benefit of this study is the possibility of applying this model in different locations where the	Due to the small number of studies on groundwater modeling in Algeria, it is recommended that additional research be conducted on this topic by applying new hybrid models and more advanced methods that could

				neural network (MLPNN), support vector regression (SVR) and decision tree regression (DTR). Multiple linear regressions are used as the base model for comparisons	costs of estimating multiple water quality variables are high and can be generally restrictive.	improve WQI modeling.
20	Robust machine learning algorithms for predicting coastal water quality index Md Galal Uddin, Stephen Nash, Mir Talas Mahammad Diganta, Azizur Rahman, Agnieszka I. Olbert	Elsevier	2022	we compared eight commonly used algorithms, including Random Forest (RF), Decision Tree (DT), KNearest Neighbors (KNN), Extreme Gradient Boosting (XGB), Extra Tree (ExT), Support Vector Machine (SVM), Linear Regression (LR), and Gaussian Naïve Bayes (GNB).	Tree-based (DT and ExT) and ensemble-based (XGB and RF) algorithms could be effective and robust for predicting the Coastal WQI.	This study's inadequacy to assess the water quality in terms of temporal resolution constitutes one of its limitations. Further studies should be carried out in order to validate the other algorithms