Team ID - **PNT2022TMID52625**
Team Lead - **DEVATHARSHINI S**
Team members - **AKASH RAJ N**
PREM B
VIJAY SHANMUGAN S
Project Title - **Efficient Water Quality Analysis and Prediction Using Machine Learning.**

LITERATURE SURVEY

S. NO	Title,authors and year	Methodology	Advantages	Limitations
1	Ali El Bilali, Abdeslam Taleb, "Prediction of irrigation water quality parameters using machine learning models in a semi arid environment",(2020).	● Multiple linear regression. ● Artificial Neural Network. ● Random forest ● Decision tree ● kNN ● Support Vector Machine ● Adaboost	Adaboost prediction accuracy is high for seven parameters compared to other models.	SVM and KNN models showed low performance for predicting contents in water. Different models showed different accuracy.
2	Qinghong Zou, Qingyu Xiong, Hualing Yi, " A water quality prediction method based on the multi-time scale bidirectional LSTM network",(2020).	● (MT-BLSTM)Multi time scale bidirectional LSTM. ● Box-Behnken experimental design. ● Bidirectional RNN.	MT-BLSTM had the highest accuracy compared to LSTM or bidirectional LSTM.	The reusability of the model needs to be improved.
3	Gasim Hayder, Osman Kurniawan, " Implementation of ML models for monitoring and predicting water quality parameters", (2020).	● Artificial Neural Network	As only one method is used, the results were accurate.	These method takes only 6 water quality parameters. Other parameters are not considered.
4	Sudhakar Singha, Srinivas Pasupuleti, Soumya S.	● Random forest. ● Extreme Gradient Boosting(XGBoost) ● Artificial Neural	Relatively higher prediction performance was observed in	A single monsoon dataset is considered in the prediction models

	Singha, Rambabu Singh, Suresh Kumar, "Prediction of groundwater quality using efficient machine learning technique",(2020).	Network.	XGBoost model than RF and ANN models.	is the limitation.
5	Md.Mahedi Hassan, Laboni Akter, Mushfiqur Rahman, Nusrat Jahan, "Efficient Prediction of Water Quality Index(WQI) Using Machine Learning Algorithms",(2021)..	● Random Forest ● Multinomial Logistic Regression(MLR). ● Artificial Neural Network(ANN) ● Support Vector Machine(SVM). ● Bagged Tree Model	Multinomial Logistic Regression has the highest accuracy at 99.83% Random forest has accuracy at 99.63%.	Only 7 parameters were considered to predict water quality.
6	Tianan Deng, Kwok-Wing Chau, Huan-Feng Duan, "Machine learning based marine water quality prediction for coastal hydro-environment management",(2021).	● Artificial Neural Network. ● Support Vector Machine	ANN showed good predicting performances and there is no overfitting problem. SVM performance is better than all ANN models in terms of water quality prediction results.	It takes much longer to train the SVM than the ANN as the SVM takes a quadratic programming with time complexity of O(n3).
7	Smail Dilmi, Mohamed Ladjal, “A novel approach for water quality classification based on the integration of deep learning and feature extraction techniques”,(2020).	● Long short term memory ● Support vector machines ● Linear discriminant analysis ● Independent component analysis	Integration of LSTM with LDA/ICA gives 99.72% accuracy.	This method doesn’t measure chemical parameters that can be measured continuously.
8	S.Sharath, R.Harish, V.Aishwarya and Dr.M.Preetha, “Water quality prediction using	● XG boost ● Random forest classifier	Produces the appropriate results with the mentioned techniques.	Accuracy of the technique is not mentioned.

	machine learning and flask",(2022).			
9	Maryam Imani, Md Mahmudul Hasan, Luiz Fernando Bittencourt, Kent McClymont and Zoran Kapelan, "A novel machine learning application: Water quality resilience prediction Model",(2020).	• Artificial neural network • Resilience	Resilience is a fast growing concept proved to be an effective approach in preparing engineering systems to tackle and cope with emerging challenges.	Model performance in validation dataset decreases abruptly in successive iterations.
10	Huimin Jia, Xiaofeng Zhou, "Water Quality Prediction Method Based on LSTM-BP",(2020).	• Long short term memory • Neural network • Back propagation	The integrated LSTM-BP has better prediction performance of time series and data generalisation ability than the single LSTM model and BP neural network model.	The Back propagation method's convergence speed is slow.
11	Al-Akhir Nayan, Muhammad Golam Kibria, Md. Obaidur Rahman and Joyeta Saha, "River Water Quality Analysis and Prediction Using GBM",(2020).	• Gradient boosting model	The model is capable of producing realistic prediction against the changes made in the dataset and accuracy is very satisfactory.	For quick and reliable on-site usage, IoT should be deployed.
12	Fangling Pu, Chujiang Ding, Zeyi Chao, Yue Yu and Xin Xu, "Water-Quality Classification of Inland Lakes Using Landsat8 Images by Convolutional Neural Networks",(2019).	• Convolutional Neural Network • Transfer learning	CNN has the best classification accuracy.	Improvement in interpretability of CNN models using water quality related variables.

13	Leonardo F. Arias-Rodriguez, Zheng Duan, Rodrigo Sepulveda, Sergio I. Martinez-Martinez and Markus Disse, “Monitoring Water Quality of Valle de Bravo Reservoir, Mexico, Using Entire Lifespan of MERIS Data and Machine Learning Approaches”,(2020).	● Gaussian processes regression ● Support vector machines ● Random forest regression ● Remote sensing	Remote sensing techniques are powerful tools for overcoming limited resources when planning monitoring programs of water quality, even across long time periods.	Greater uncertainties in the results may be present due the variation of 2 days between field and satellite data collection.
14	Paul D. Rosero-Montalvo, Vivian F. López-Batista , Jaime A. Riascos and Diego H. Peluffo-Ordóñez, “Intelligent WSN System for Water Quality Analysis Using Machine Learning Algorithms: A Case Study (Tahuando River from Ecuador)” ,(2020).	● Condensed Nearest Neighbour (CNN), ● Decremental Reduction Optimization Procedures 1 (DROP 1), ● Decremental Reduction Optimization Procedures 2(DROP 2) ● Decremental Reduction Optimization Procedure 3 (DROP3). ● Distance-based: K-Nearest Neighbours (KNN).	A classifier together with a prototype selection is used and it shows a good trade-off between the computational resource usage and the classification performance at detecting the pollution levels.	The battery life is to be more carefully considered by exploring both different methods of extending its duration and alternative sources of energy to supply the nodes.
15	Nur Aqilah Paskhal Rostam , Nurul Hashimah Ahamed Hassain Malim , Rosni Abdullah, Abdul Latif Ahmad , Boon Seng Ooi , And Derek Juinn Chieh Chan, “A Complete Proposed Framework for Coastal Water	● Long Short-term Memory (LSTM) ● ANN(Artificial Neural Networks) ● Support vector machines(SVM) ● Decision Tree(DT) ● Regression method.	A development of the most suitable predictive modelling which can be used to tackle the dynamic issues of algal growth prediction in coastal studies which plays a major role in water quality prediction.	Incorporation of MF that might further improve the performance.

	Quality Monitoring System With Algae Predictive Model",(2021).			
16	Mosleh Hmoud Al-Adhaileh , and Fawaz Waselallah Alsaade,"Modelling and Prediction of Water Quality by Using Artificial Intelligence" ,(2021).	• Adaptive neuro-fuzzy inference system (ANFIS) • Feed-forward neural network (FFNN) • K-nearest neighbours(KNN) • Artificial Neural Networks (ANN)	The ANFIS model showed accuracy of 96.17% for predicting WQI. The FFNN model achieved the highest accuracy 100% for WQC.	The developed models can be enhanced to predict the quality of different types of water around the globe.(As it can classify and predict the quality of only waters in Saudi Arabia).
17	Umair Ahmed , Rafia Mumtaz,, Hirra Anwar, Asad A. Shah , Rabia Irfan and José García-Nieto, "Efficient Water Quality Prediction Using Supervised Machine Learning",(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 Perceptrons (MLP) • Gaussian Naïve Bayes • Logistic Regression • Stochastic gradient descent • K Nearest Neighbour • Decision Tree • Bagging Classifier	The model uses a minimal number of parameters to validate the possibility of its use in real time water quality detection systems.	Integration of the findings of this research in a large-scale IoT-based online monitoring system is needed inorder to classify and predict the real time data.
18	Sang-Soo Baek , Jongcheol Pyo and Jong Ahn Chun,"Prediction of Water Level and Water Quality Using a CNN-LSTM	• Convolutional Neural Network (CNN) • Long Short-Term Memory (LSTM)	The LSTM model in this study well represented the different temporal variations of each pollutant type.	Only three different pollutants were investigated in this study.A further study is recommended to develop models

	Combined Deep Learning Approach",(2020).			so that more pollutants can be simulated.
19	Jianzhuo Yan, Jiaxue Liu, Yongchuan Yu, and Hongxia Xu, "Water Quality Prediction in the Luan River Based on 1-DRCNN and BiGRU Hybrid Neural Network Model",(2021).	● One-dimensional residual convolutional neural networks (1-DRCNN) ● Bidirectional gated recurrent units (BiGRU) ● Long Short-Term Memory (LSTM)	PCA technology was used to enhance the data of water quality parameters to prevent the overfitting of the model.	The prediction effect is not ideal for the dataset with small TP with a value.Therefore, consideration of other factors which affects water quality should be considered.
20	Dao Nguyen Khoi , Nguyen Trong Quan, Do Quang Linh, Pham Thi Thao Nhi and Nguyen Thi Diem Thuy,"Using Machine Learning Models for Predicting the Water Quality Index in the La Buong River, Vietnam ",(2022).	● Adaptive boosting ● Gradient boosting ● Histogram-based gradient boosting ● Light gradient boosting ● Extreme gradient boosting ● Decision tree ● Random forest ● Multilayer perceptron ● Radial basis function ● Deep feed-forward neural network ● Convolutional neural network	It strengthens the fact that ML models, particularly XGBoost, can be utilised for predicting the WQI with a high degree of accuracy, which will further improve water quality management.	