|  |  |  |  |
| --- | --- | --- | --- |
| Author | Title | Best Algorithm | Worst Algorithm |
| (Nguyen et al., 2022) | The Use of Machine Learning Algorithms for Evaluating Water Quality Index: A Survey and Perspective | Deep learning algorithms like CNN, LSTM, and DBN are preferred for predicting water quality parameters | Traditional ML methods. |
| (Rahu et al., 2024) | IoT and machine learning solutions for monitoring agricultural water quality: a robust framework | Support Vector Machine (SVM) | Naïve Bayes |
| (Rahu et al., 2023) | Toward Design of Internet of Things and Machine Learning-Enabled Frameworks for Analysis and Prediction of Water Quality | Random Forest | Support Vector Regression (SVR) |
| (Jayaraman et al., 2024) | Critical Review On Water Quality Analysis Using IoT And Machine | Long Short-Term Memory (LSTM) and Artificial Neural Networks (ANN) | Traditional methods |
| (Goodarzi et al., 2023) | Water Quality Index Estimations Using Machine Learning Algorithms: A Case Study of Yazd-Ardakan Plain, Iran | MARS | GEP |
| (S. Angel Vergina et al., 2020) | A Real Time Water Quality Monitoring Using Machine Learning Algorithm | K-Means Clustering | Not mention |
| (Nair & Vijaya, 2021) | Predictive Models for River Water Quality using Machine Learning and Big Data Techniques - A Survey | The combination of Support Vector Regression (SVR) with Artificial Neural Networks (ANN) and Markov chain achieved the best performance | Not mention |
| (Li et al., 2022) | Predicting Aquaculture Water Quality Using Machine Learning Approaches | Support Vector Machine (SVM) | Back Propagation Neural Network (BPNN) and Least Squares Support Vector Machine (LSSVM) |
| (Nallakaruppan et al., 2024) | Reliable Water Quality Prediction And Parametric Analysis Using Explainable AI Models | Random Forest | Support Vector Machine (SVM) |
| (Xu et al., 2022) | A Machine Learning Predictive Model To Detect Water Quality And Pollution | NuSVC | SVD Imputation |

|  |  |  |
| --- | --- | --- |
| Author | Title | Limitation |
| (Nguyen et al., 2022) | The Use of Machine Learning Algorithms for Evaluating Water Quality Index: A Survey and Perspective | Deep learning algorithms require a large number of samples for training, and they are often considered "black box" models, making it difficult for domain experts to understand the model's outputs |
| (Rahu et al., 2024) | IoT and machine learning solutions for monitoring agricultural water quality: a robust framework | The SVM model, while highly accurate with an accuracy of 96%, is computationally intensive, making it less suitable for real-time applications due to its complexity and resource demands. Naïve Bayes, although simple and efficient, assumes feature independence, which may not hold true in real-world scenarios, affecting its accuracy. The Decision Tree model, despite being interpretable, does not perform as well as other models, indicating a trade-off between interpretability and performance |
| (Rahu et al., 2023) | Toward Design of Internet of Things and Machine Learning-Enabled Frameworks for Analysis and Prediction of Water Quality | The study relies on a reduced set of Water Quality Parameters (WQPs), which may limit the model's ability to capture the full spectrum of water quality variations. Additionally, machine learning models perform better with smaller datasets, which could restrict scalability for extensive real-time monitoring applications |
| (Jayaraman et al., 2024) | Critical Review On Water Quality Analysis Using IoT And Machine | The accuracy of machine learning models is influenced by the quality of training data, complexity of water quality metrics, and monitoring frequency. Traditional methods are time-consuming and prone to errors |
| (Goodarzi et al., 2023) | Water Quality Index Estimations Using Machine Learning Algorithms: A Case Study of Yazd-Ardakan Plain, Iran | The MARS model, while having good estimation accuracy, tends to overfit during the forward phase, which can lead to decreased accuracy when applied to new data. Additionally, the study's models are heavily dependent on the quality and quantity of the data used for training and testing. The models were specifically trained on data from the Yazd-Ardakan Plain, which may limit their applicability to other regions |
| (S. Angel Vergina et al., 2020) | A Real Time Water Quality Monitoring Using Machine Learning Algorithm | K-Means is an unsupervised learning model, which may not always provide the most accurate predictions compared to supervised learning models. Additionally, it might not capture all the complexities and variations in water quality, especially when other contaminants or parameters are involved. The system relies on IoT devices and cloud computing, which require data integrity and security during transmission to ensure reliability |
| (Nair & Vijaya, 2021) | Predictive Models for River Water Quality using Machine Learning and Big Data Techniques - A Survey | Data quality issues affect many factors like battery, device faults, sensor network issues. Lack of integrated real-time monitoring and supervision handling system for river water quality. Smaller data size can lead to overfitting when training models with algorithms like ANN |
| (Li et al., 2022) | Predicting Aquaculture Water Quality Using Machine Learning Approaches | Parameter selection of SVM was a shortcoming, suggesting that parameter optimization methods could be used to obtain better prediction results |
| (Nallakaruppan et al., 2024) | Reliable Water Quality Prediction And Parametric Analysis Using Explainable AI Models | The model's results are based on a specific dataset, which may not be applicable globally. The model needs to be re-trained when applied to new environments. The research lacks the involvement of a subject matter expert. The study identifies solids as a primary influencing factor for potability, but in real-world applications, solids can vary widely in form |
| (Xu et al., 2022) | A Machine Learning Predictive Model To Detect Water Quality And Pollution | The dataset is too small, leading to overfitting and difficulty in discerning training speed differences between models. The labeling process may lack authority, impacting the reliability of results |