**Title Page:**

**Improving the Accuracy in Assessing the Drinking Water Quality using Convolutional Neural Network in comparison with Naïve Bayes Algorithm**

Poleboina Radha Devi1, Dr. V. Gokula Krishnan2

Poleboina Radha Devi1

Research Scholar,

Department of Computer Sciences Engineering,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105

[radhadevip1651.sse@saveetha.com](mailto:radhadevip1651.sse@saveetha.com)

Dr. V. Gokula Krishnan2

Project Guide, Corresponding Author,

Department of Quantum Intelligence,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105

[gokulakrishnan.sse@saveetha.com](mailto:gokulakrishnan.sse@saveetha.com)

**Keywords**: Convolutional Neural Network, Naïve Bayes Algorithm, Accuracy, Drinking Water, Quality Assessment.

**ABSTRACT**

**Aim:** This study aims to enhance the accuracy of assessing drinking water quality by employing a Convolutional Neural Network (CNN) and comparing its performance with the Naïve Bayes Algorithm. **Materials and Methods:** Drinking water quality assessment is performed by: Convolutional Neural Network in comparison Algorithm and Naïve Bayes Algorithm, Total sample size of 20 is calculated using G power with a pretest power of 0.8 and alpha 0.05. The dataset consists of Drinking water comprised diverse water quality parameters, and the CNN was trained to recognize patterns indicative of water contamination. Additionally, we employed the Naïve Bayes Algorithm as a benchmark for comparison. The dataset included various parameters relevant to water quality assessment. **Result:** The mean accuracy of the Naïve Bayes Algorithm in assessing drinking water quality was notably superior to the Convolutional Neural Network in comparison, with an accuracy of 85% compared to 83%. Statistical analysis revealed a significance level of 0.001, emphasizing a substantial statistical distinction between the two algorithms in terms of accuracy. **Conclusion:** The mean accuracy of the drinking water quality assessment in the Convolutional Neural Network is better than the Naïve Bayes Algorithm.

**Keywords**: Convolutional Neural Network, Naïve Bayes Algorithm, Accuracy, Drinking Water, Quality Assessment.

**INTRODUCTION**

Evaluating the physical, chemical, and biological state of water(Karim, Guha, and Beni 2020) entails a comprehensive analysis that considers its interplay with natural elements, assesses the impact of human activities, and anticipates future applications. Securing access to safe water is crucial for public health, spanning needs from drinking and household tasks to food cultivation and recreation (Craddock et al. 2021). The improvement of water supply, sanitation, and resource management not only fosters economic growth in nations but also plays a pivotal role in alleviating poverty. Regular monitoring and testing facilitate (Suva et al. 2003) the application of appropriate treatment and purification methods, (Kılıçarslan and Argun 2023) ensuring the provision of clean and safe drinking water.



**Fig. 1.** Drinking water quality assessment

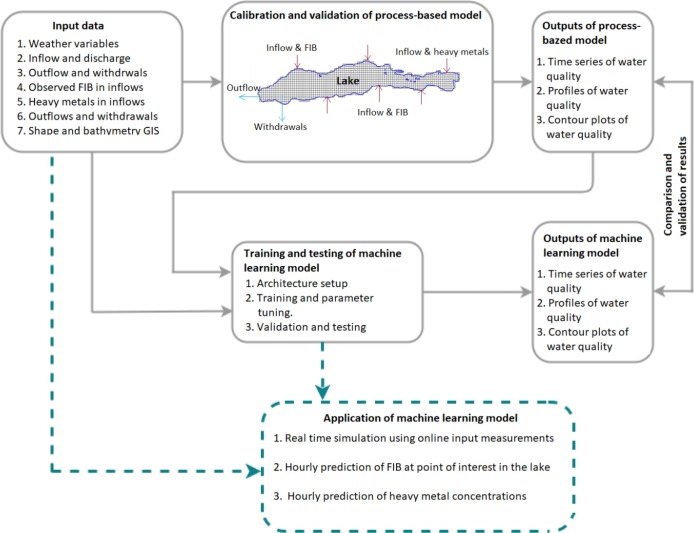
Many articles published in this topic over the past years in at least 2 databases. Some of them are Water Quality Assessment Through Predictive Machine Learning (Kavitha et al. 2020). Enhancing Water Potability Assessment Using Hybrid Fuzzy-Naïve Bayes Algal Morphological Identification in Watersheds for Drinking Water Supply (Wang, Asefa, and Thornburgh 2022) Using Neural Architecture Search for Convolutional Neural Network A review of the application of machine learning in water quality evaluation Performance of Naïve Bayes Tree with ensemble learner techniques for groundwater potential mapping and the author was Tran van Phong.

The condition of water mirrors the state of the environment, influencing how humans perceive their quality of life (Etale and Siegrist 2018). Variables linked to coastal water quality serve as indicators, reflecting human perceptions of their social and cultural surroundings, ultimately impacting psychological and physical aspects of life quality. This study adds value by suggesting a novel validation approach for machine learning algorithms. It presents a perspective on validation that can steer future research in constructing predictive models for categorizing coastal waters, aligning with current regulations on pathogenic microbial load.

**MATERIALS AND METHODS**

Various machine learning techniques were explored, with CNN and Naïve Bayes algorithms considered alongside others. In the realm of neural networks, we formulated a sequential model with layers tailored to our data and fine-tuned the parameters for optimal performance. No. of groups is 2 and, the total sample size is 20. The power G is 0.8 and alpha value is 0.05.

Colorimeters and Photometers play a vital role in analysing water samples, suspended sediment, and bottom material for their inorganic and organic constituents. Meanwhile, test strips offer an economical method for conducting spot checks on water quality. Various technical associations have crafted manuals outlining water sample analysis methods. Recognizing the need for uniformity, efforts are underway to develop practical, standardized methods across these manuals. For instance, the determination of chloride may be covered by essentially the same method in manuals from different associations.



**Fig. 2.** The methods of hydrodynamic and water quality model

In recent decades, the method endeavours to convert a calibrated hydrodynamic and water quality model, characterized as process-based, into a series of LSTM models. These LSTM models are trained using the outputs of the process-based model as their targets. Eight water potability-related fields in a dataset underwent machine learning analysis. A compilation of 16 algorithms, comprising 11 shallow and 5 deep learning models, was employed to predict water potability. The assessment resulted in a peak testing accuracy of 85.03%, comparing the algorithms' performance based on chemical and laboratory measurements.

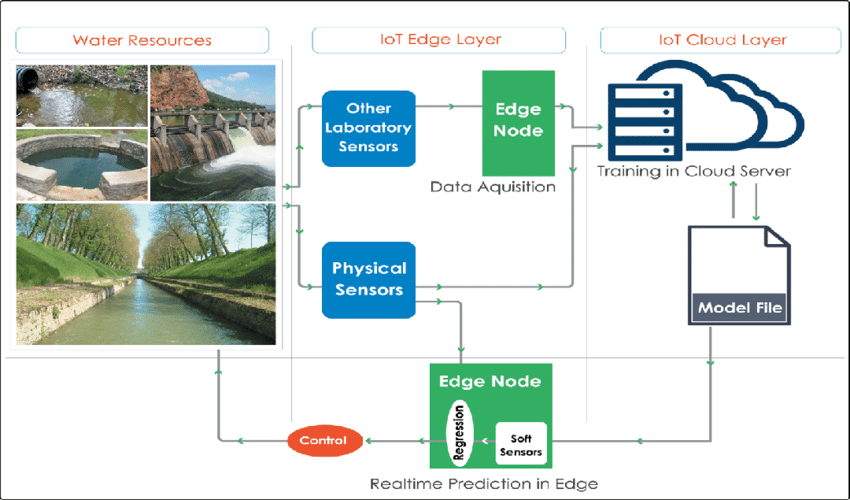
The World Health Organization designates water as a vital element for life, spanning 70% of Earth's surface, comprising 50–60% of human weight, and constituting 90% of our cells. Established by the United Nations in 1992, World Water Day, celebrated annually on March 22, aims to enhance public awareness of the significance of drinking water and the sustainable management of water resources.

Although water is a crucial nutrient for human health, its ready availability often leads to underappreciation, obscuring the importance of proper hydration. Experts highlight water as the second essential source of oxygen, vital for life, given that two-thirds of the human body consists of water. While a person can endure several days without food, survival without water is limited to just a few days. As the primary component of the human body, water plays a pivotal role in all stages of human development.

Our research work was based on a publicly available environmental pollution from human activities has surged, significantly impacting water quality. Discharges from sources like factories or sewage treatment plants directly influence water quality. Additionally, pollution from diffuse sources, such as nutrients and pesticides from agricultural activities, and pollutants released into the air by industries, which subsequently settle on land and sea, also contribute to these concerns.

**CONVOLUTIONAL NEURAL NETWORK**

Convolutional Neural Networks (CNNs) excel in image-related tasks but extend their applicability to diverse data types, including those relevant to drinking water quality assessment. In the domain of machine learning for evaluating drinking water quality, CNNs can be deployed to analyse images capturing water samples, such as microscopic images or photos of water quality indicators’ play a crucial role in identifying contaminants, particles, or distinctive patterns signalling potential water quality issues. By learning hierarchical features from these images, the algorithm adeptly recognizes intricate patterns that may pose challenges for traditional methods. It’s important to note that the effectiveness of CNNs in this context hinges on the availability and quality of labelled training data. Moreover, integrating CNNs with other machine learning techniques or incorporating them into a comprehensive system can amplify the overall accuracy of drinking water quality assessment.



**Fig. 3.** Real time prediction of water

**STEPS:**

1. Acquire a labelled dataset of images representing water samples.
2. Design a CNN architecture suitable for drinking water quality assessment feature extraction.
3. Train the CNN model on the labelled dataset.
4. Fine-tune hyperparameters, such as learning rate and kernel sizes, for optimal performance.
5. Employ the trained CNN to analyse new water sample images and identify contaminants or patterns.
6. Evaluate the CNN model's accuracy and effectiveness in detecting drinking water quality assessment issues.

**NAÏVE BAYES ALGORITHM**

The Naive Bayes algorithm is applicable to assess drinking water quality assessment through classification. In this scenario, it employs Bayes' theorem with an assumption of feature independence, hence the term "naive." Here's how it can be utilized:

1. Data Preparation: Gather and preprocess a dataset with water quality-related features like chemical concentrations, physical properties, and microbial counts.

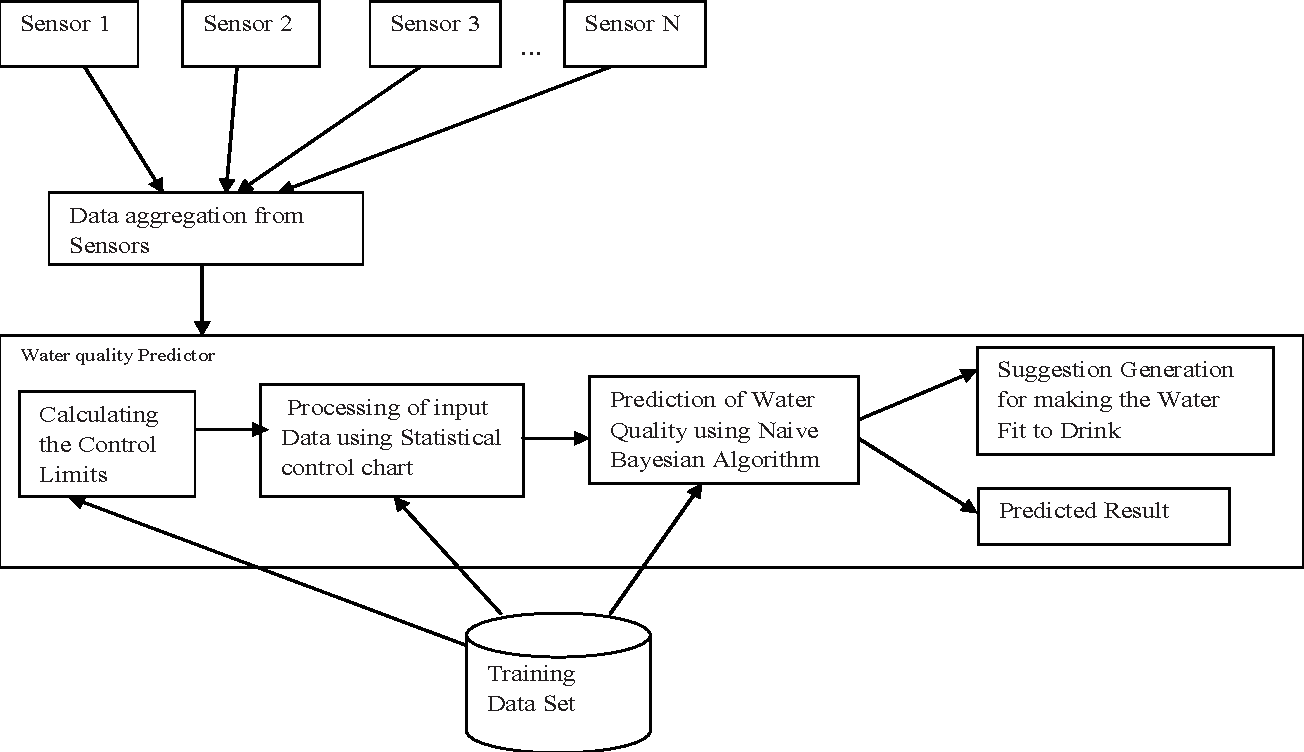
2. Feature Independence Assumption: Despite its simplifying assumption, Naive Bayes effectively manages features independently, making it suitable for specific aspects of water quality assessment.

3. Model Training: Train the Naive Bayes model on a labelled dataset, where labels denote different water quality categories (e.g., safe, contaminated).

4. Prediction: Utilize the trained model to predict the water quality category of new samples based on their feature values. The algorithm calculates the probability of a sample belonging to each category, selecting the one with the highest probability.

5. Evaluation: Assess the model's performance using metrics like accuracy to measure its effectiveness in classifying water samples.

Although Naive Bayes is commonly used for discrete data tasks like text classification and spam filtering, its simplicity and efficiency make it a viable choice for specific aspects of drinking water quality assessment where the assumption of feature independence holds reasonably well. Nevertheless, the algorithm selection should account for the specific characteristics and complexity of the drinking water quality assessment data under consideration.



**Fig. 4.** Drinking water dataset

**STEPS:**

1. Obtain a labelled dataset comprising images depicting water samples.
2. Devise a Naive Bayes algorithm tailored for water quality assessment.
3. Train the Naive Bayes model using the labelled dataset.
4. Fine-tune parameters, considering features' independence, for optimal performance.
5. Utilize the trained Naive Bayes model to analyse new water sample images, discerning contaminants or patterns.
6. Evaluate the model's accuracy and effectiveness in identifying water quality issues, considering metrics.

**STATISTICAL ANALYSIS**

The algorithm was executed on a Windows 11 64-bit laptop with 8 GB RAM, utilizing a robust internet connection. Additionally, a Collab Notebook with Python was employed. Analysis involved utilizing features from both cracked and uncracked images. Model iteration was considered in T-Test calculations, comparing the performance of Convolutional Neural Networks (CNNs) and Naïve Bayes Algorithms based on loss and precision percentage. Statistical significance indicates a difference between the two algorithms. Cities, seasons, and dates are independent factors, while accuracy and loss serve as dependent variables. The analysis considers both independent and dependent variables to enhance accuracy.

**RESULT**

The input of the Support Vector Machine (SVM) is taken in the form of features extracted from the dataset and identifies the asthma prediction rate with an accuracy of 85%. Table 8 shows the value Random Forest Algorithm and Support Vector Machine (SVM) after comparing the accuracy and loss. It can be observed that a Support Vector Machine (SVM) Algorithm has a better accuracy value than a Random Forest with a value of p=0.000. The accuracy and loss of both algorithms are depicted in the bar graph.

The Mean, Standard Deviation, and Standard Error Mean for the Random Forest and SVM Algorithms are displayed in Table 4 for the group statistical analysis. Table 5 displays the findings of the independent sample T-test that was run on two graphs for significance and determining standard error. With a 95% confidence interval, the p-value is 0.001 less than 0.05 and is deemed statistically significant.

**DISCUSSION**

**CONCLUSION**

**DECLARATION**

**Conflicts of Interest**

No conflict of interest in this manuscript.

**Authors' Contribution**

Author PRD was involved in data collection, analysis, and manuscript writing. Author VGK was involved in the conceptualization, guidance, and critical review of the manuscript.

**Acknowledgments**

The authors would like to express their gratitude towards SIMATS School of Engineering, Saveetha Institute of Medical and Technical Sciences (formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

**Funding**

We thank the following organizations for providing financial support that enabled us to complete this study.

1. Saveetha School of Engineering.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.

**REFERENCES**

Craddock, Hillary A., Younes Rjoub, Kristal Jones, Clive Lipchin, and Amy R. Sapkota. 2021. “Perceptions on the Use of Recycled Water for Produce Irrigation and Household Tasks: A Comparison between Israeli and Palestinian Consumers.” *Journal of Environmental Management* 297 (113234): 113234.

Etale, Anita, and Michael Siegrist. 2018. “Perceived Naturalness of Water: The Effect of Biological Agents and Beneficial Human Action.” *Food Quality and Preference* 68 (September): 245–49.

Karim, Kaleh, Sujata Guha, and Ryan Beni. 2020. “Comparative Analysis of Water Quality Disparities in the United States in Relation to Heavy Metals and Biological Contaminants.” *Water* 12 (4): 967.

Kavitha, B., P. Parthiban, M. Goel, K. Ravikumar, A. Das, J. S. Sudarsan, and S. Nithiyanantham. 2020. “Assessment and Recurrence of Kidney Stones through Optimized Machine Learning Tree Classifiers Using Dietary Water Quality Parameters and Patient’s History.” *Advanced Science, Engineering and Medicine* 12 (10): 1219–23.

Kılıçarslan, Müberra Nur, and Mehmet Emin Argun. 2023. “Holistic Effect of Operating Conditions and Purification Methods on the Pollution Characteristics of Washing Water and Biodiesel Quality.” *Journal of Water Process Engineering* 54 (103989): 103989.

Suva, Jeff A. K., Paul E. Mariner, Hans W. Meinardus, and Jon S. Ginn. 2003. “Monitoring Partitioning Tracer Testing and Surfactant Flooding by In‐line Gas Chromatography Techniques.” *Ground Water Monitoring & Remediation* 23 (1): 85–92.

Wang, Hui, Tirusew Asefa, and Jack Thornburgh. 2022. “Integrating Water Quality and Streamflow into Prediction of Chemical Dosage in a Drinking Water Treatment Plant Using Machine Learning Algorithms.” *Water Science & Technology: Water Supply* 22 (3): 2803–15.

**TABLES AND FIGURE**

**Table 1.** SVM Accuracy

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

