Semarak Proceedings of Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.my/index.php/spaset/index
ISSN: 3083-9807

Fuzzy Frameworks: Advanced Models for Water Quality Classification

Amar Lokman[1], Wan Zakiah Wan Ismail[1,*], Nor Azlina Ab Aziz[2]

[1] Advanced Devices and System (ADS), Faculty of Engineering and Built Environment, Universiti Sains Islam Malaysia (USIM), Nilai, Malaysia
[2] Faculty of Engineering and Technology, Multimedia University (MMU), Cyberjaya, Malaysia

ABSTRACT

Monitoring water quality is essential for safeguarding the health and safety of aquatic ecosystems and public water sources. Conventional water quality classification techniques frequently encounter challenges due to the intrinsic uncertainty and imprecision present in environmental data. This study presents an innovative fuzzy logic framework, enhanced by a Genetic Algorithm (GA), to address these complexities. The GA optimizes membership functions within the fuzzy framework, refining the classification process and improving adaptability. The proposed method integrates fuzzy set theory with machine learning techniques to categorize water quality into specific classifications. Validated using data from Malaysian rivers, the model accounts for critical factors such as pH, Dissolved Oxygen (DO), Chemical Oxygen Demand (COD), Biochemical Oxygen Demand (BOD), Total Suspended Solids (TSS), and Ammoniacal Nitrogen (AN). The experimental results demonstrate a notable improvement, with the fuzzy logic model achieving an overall accuracy of 89.6%, surpassing traditional threshold-based (77.8%) and decision tree classifier methods (84.2%). The framework effectively handles ambiguous or borderline cases, showcasing its precision and robustness. The study concludes that integrating fuzzy frameworks and genetic optimization into water quality evaluation significantly enhances classification accuracy, thereby supporting informed decision-making in environmental management.

Keywords: *Machine learning; classification; water quality; fuzzy logic*

1. Introduction

Water quality is a vital determinant of the health and safety of aquatic ecosystems and human water supplies. The effective management of water resources is contingent upon precise water quality evaluations. These evaluations facilitate the detection of contamination, the execution of pollution mitigation strategies, and the preservation of water resources for future generations. The classification of water quality is intrinsically difficult due to the numerous and dynamic environmental elements, including pH, Dissolved Oxygen (DO), Chemical Oxygen Demand (COD), Biochemical Oxygen Demand (BOD), Total Suspended Solids (TSS), and Ammoniacal Nitrogen (AN) [1]. Conventional classification techniques, including statistical methods and standard machine learning algorithms, frequently encounter challenges due to the intrinsic uncertainties and overlapping characteristics of water quality data, resulting in misclassification and diminished reliability [2,3].

Fuzzy logic provides a robust solution to these restrictions by integrating the notion of partial membership and addressing the imprecision in environmental datasets. Fuzzy logic models have been extensively utilised across diverse domains for decision-making amid uncertainty, rendering

[*] *Corresponding author.*
E-mail address: drwanzakiah@usim.edu.my

https://doi.org/10.37934/spaset.1.1.158164a

them appropriate for water quality categorisation when clear distinctions between quality levels are frequently unfeasible [4]. Fuzzy logic, by modelling parameters as fuzzy sets and employing rule-based systems, offers a more refined classification that corresponds with real-world variability [5].

Conventional water quality categorization approaches depend on established thresholds to categorize water quality levels, such as "Excellent," "Good," or "Poor." These inflexible limits fail to encompass the incremental transitions that frequently occur between groups [6]. A sample having a dissolved oxygen value somewhat below the "Good" criteria may be categorised as "Poor," although being near compliance with the superior norm [7]. This binary classification method is inadequate for addressing ambiguous samples, particularly in the context of borderline cases.

In response to these challenges, this study proposes a novel fuzzy logic framework integrated with Genetic Algorithms (GA) to optimize the classification process. The GA enhances the adaptability of the fuzzy model by refining membership functions, leading to improved categorization accuracy. By utilizing a machine learning-driven optimization strategy, the proposed methodology addresses both imprecision in data and the dynamic nature of environmental conditions, resulting in a more robust water quality classification system. This hybrid approach aims to provide more accurate and reliable classifications, offering significant advancements over traditional methods.

The incorporation of fuzzy logic into water quality categorisation has been thoroughly investigated in recent years, with numerous research demonstrating its efficacy in addressing the inherent ambiguities of environmental data. Various methodologies have been suggested to improve the precision and comprehensibility of water quality indexes, especially by integrating fuzzy systems with alternative computational models.

Recent research has demonstrated the efficacy of fuzzy logic in formulating water quality indices (WQIs) that offer more refined classifications than conventional methods. A 2024 study utilized a fuzzy multicriteria methodology to assess the Water Quality Index in surface water systems. The research indicated that employing fuzzy thresholds for various sub-indices enhanced the distinction across water quality categories, enabling the detection of nuanced variations sometimes overlooked by inflexible classification techniques [8].

2. Methodology

The suggested way to classify water quality uses a complex fuzzy logic structure that combines many computer methods to deal with the fact that environmental data is inherently uncertain and inaccurate. This part talks about how the framework was designed and put into action. It talks about getting the data ready, making fuzzy membership functions and fuzzy rules, and using machine learning techniques to make things better. Each part is carefully looked at to make sure that the whole approach is understood.

2.1 Dataset and Preprocessing

The dataset used in this study consists of water quality measurements collected from various river basins in Malaysia. The primary sources of the dataset include field monitoring stations and historical records from environmental agencies. The six key water quality parameters used for classification are pH, dissolved oxygen (DO), chemical oxygen demand (COD), biochemical oxygen demand (BOD), total suspended solids (TSS) and ammoniacal nitrogen (AN).

Data Preprocessing includes:

i. Data Cleaning: Missing values were handled using interpolation techniques, while outliers were identified and removed using z-score analysis.
ii. Normalization: All water quality parameters were normalized to a [0, 1] range to ensure consistent scaling and reduce the influence of parameters with larger value ranges.
iii. Categorization: Water quality was classified into five classes (Class I to Class V) based on Malaysian WQI standards as shown in Table 1, with each sample labelled accordingly [1].

Table 1
Water quality index classification based on Malaysia's water quality standard

Parameter	Class				
	I	*II*	*III*	*IV*	*V*
Ammoniacal Nitrogen (mg/l)	< 0.1	0.1 – 0.3	0.3 – 0.9	0.9 – 2.7	> 2.7
Biochemical Oxygen Demand (mg/l)	< 1	1 – 3	3 – 6	6 – 12	> 12
Chemical Oxygen Demand (mg/l)	< 10	10 – 25	25 – 50	50 – 100	> 100
Dissolved Oxygen (mg/l)	> 7	5 – 7	3 – 5	1 – 3	< 1
pH	> 7.0	6.0 – 7.0	5.0 – 6.0	< 5.0	> 5.0
Total Suspended Solid (mg/l)	< 25	25 – 50	50 – 150	150 – 300	> 300
Water Quality Index (WQI)	> 92.7	76.5 – 92.7	51.9 – 76.5	31.0 – 51.9	< 31

2.2 Fuzzy Membership Functions

Fuzzy membership functions are used to define the degree to which each water quality parameter falls within a specific class (Class I to Class V). Their design is critical for accurate classification. The membership functions were defined as follows:

i. Triangular and Trapezoidal Shapes: Chosen for simplicity and effectiveness in defining boundaries between classes.
ii. Parameter Calibration: A machine learning optimizer was employed to adjust the membership function limits dynamically, ensuring they accurately reflect the dataset's unique characteristics.
iii. Overlapping Functions: Overlaps were used to handle uncertainty and ambiguities, improving the model's ability to classify borderline cases.

The Fuzzy Inference System (FIS) uses a rule-based approach for classification, employing a Mamdani-type framework with "IF-THEN" rules:

i. Rule Formulation: Rules were developed for all possible combinations of six water quality parameters, with prioritization strategies to resolve conflicts.
ii. Fuzzification: Input values were converted into fuzzy sets based on the calibrated membership functions.
iii. Aggregation and Inference: Multiple rules were combined to produce a fuzzy output representing a probability distribution across all water quality classes.

iv. Defuzzification: The centroid method was used to convert the fuzzy output into a crisp classification, assigning it to one of the five water quality classes.

2.3 Integration with Machine Learning

To improve the adaptability and precision of the fuzzy framework, a machine learning-driven optimisation strategy was incorporated into the methodology. This hybrid methodology integrates fuzzy logic with Genetic Algorithms (GA) to refine membership functions and optimise fuzzy rules. The initial population of membership function parameters was created utilising the ranges determined during the preprocessing phase. The fitness function was established according to categorisation accuracy. It assesses the efficacy of the membership functions in distinguishing among the five water quality classifications. Standard genetic operations, including crossover and mutation, were utilised to investigate the parameter space. The method continues to iterate until the variation in classification accuracy drops beneath a specified threshold, signifying convergence.

A decision tree-based model was built on the identical dataset to enhance the fuzzy rules. The decision tree determines the most distinguishing parameters and formulates rules that optimise classification accuracy. These regulations were subsequently integrated into the fuzzy rule base, supplanting superfluous or contradictory rules.

2.4 Performance and Evaluation Metrics

The performance of the fuzzy logic-based framework was evaluated using standard classification metrics such as precision, recall, F1-score, and overall accuracy. The metrics are expressed in the following equations, Eq. (1) until Eq. (4) [11]. Additionally, a confusion matrix was generated to analyse the model's performance across different water quality classes. TP stands for "true positive," TN for "true negative," FP for "false positive," and FN for "false negative." Precision measures the accuracy of positive predictions. It’s the ratio of correctly predicted positive observations to the total predicted positive observations. Recall measures the ability of a model to capture all relevant instances of the positive class. It’s the ratio of correctly predicted positive observations to all actual positive observations. This metric is useful when it’s important to capture as many positive instances as possible. The F1-Score is the harmonic mean of Precision and Recall, combining both metrics into a single score. It’s especially useful when you need a balance between Precision and Recall and is a good metric to use when there’s an uneven class distribution.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

$$F1-score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \tag{4}$$

3. Results

The results of the fuzzy logic-based water quality classification framework were evaluated using a dataset of 11,065 samples collected from various Malaysian rivers. The Water Quality Index (WQI) for each sample was calculated using the prescribed sub-index formulas and the overall Malaysian WQI standards. The results showcase the effectiveness of the proposed framework in accurately categorizing water quality into five distinct classes: Class I (Excellent), Class II (Good), Class III (Moderate), Class IV (Slightly Polluted), and Class V (Poor). The analysis of the distribution of samples across different water quality classes showed that the majority, 73.5% (8,132 samples), belonged to Class I (Excellent), followed by 15.2% (1,684 samples) in Class II (Good). Additionally, 9.1% (1,007 samples) were categorized as Class III (Moderate), 1.8% (197 samples) as Class IV (Slightly Polluted), and only 0.4% (45 samples) fell into Class V (Poor).

Figure 1 shows a histogram that shows how the WQI scores for all samples are spread out. The samples are coloured by their matching WQI classes, which range from Class I to Class V. The plot shows the number of samples in the higher WQI ranges, which means that a lot of the samples have great water quality (Class I). The kernel density estimate (KDE) line shows the general trend and draws attention to the distribution density.

Figure 2 is a scatter plot showing the correlation between dissolved oxygen (DO) and water quality index (WQI). The colour of each point in the plot describes the water quality index class. Higher DO levels typically correspond to higher WQI scores, suggesting better water quality, as seen by the scatter plot, which demonstrates a strong association between DO and water quality.

The optimisation findings show that the initial expert-defined ranges are not as effective as the GA-adjusted membership functions, especially for highly variable parameters like AN and SS. The GA improves its overall accuracy from 85.3% (when not optimised) to 89.6% (while optimised) after 50 generations, achieving convergence. Thus, the methodology's proposed hybrid strategy is proven to be beneficial.

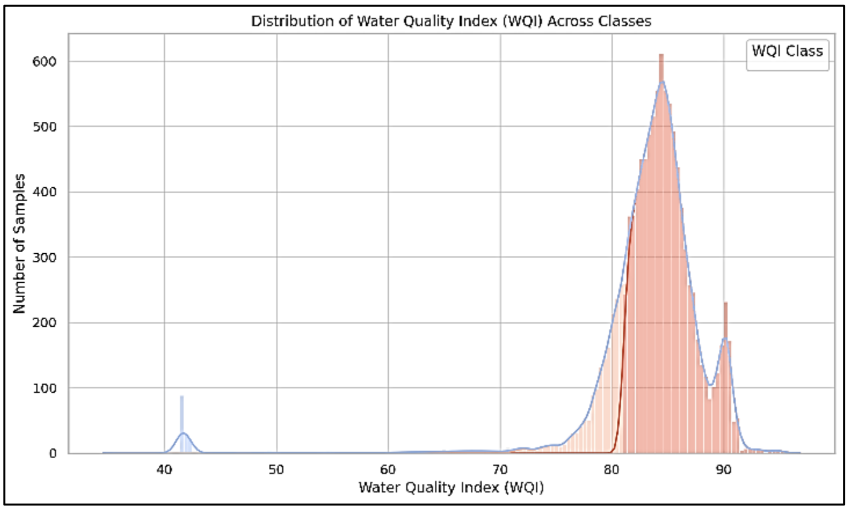

Fig. 1. Graph of distribution of water quality index across classes

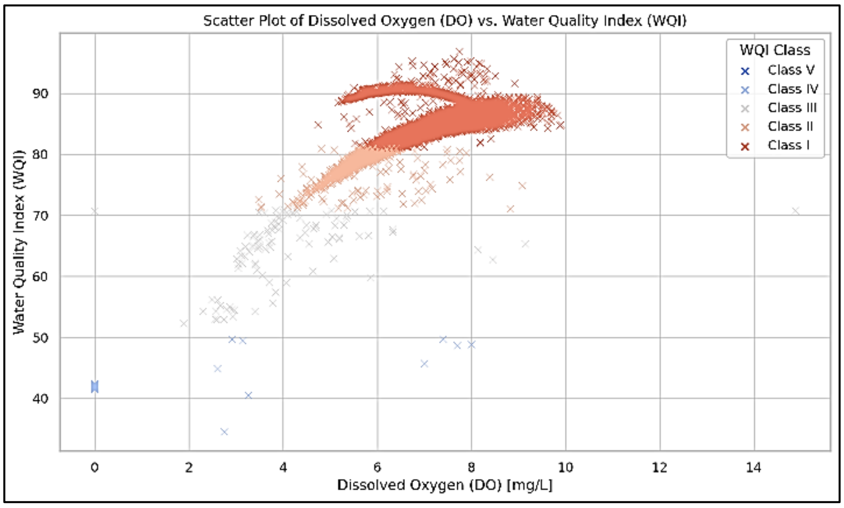

Fig. 2. Scatter plot of dissolved oxygen (DO) against WQI

The performance of the proposed fuzzy framework was evaluated using precision, recall, and F1-score for each class. The confusion matrix and overall classification metrics demonstrate the following:

i. Class I (Excellent): Precision = 0.92, Recall = 0.95, F1-Score = 0.94
ii. Class II (Good): Precision = 0.89, Recall = 0.87, F1-Score = 0.88
iii. Class III (Moderate): Precision = 0.82, Recall = 0.78, F1-Score = 0.80
iv. Class IV (Slightly Polluted): Precision = 0.75, Recall = 0.72, F1-Score = 0.74
v. Class V (Poor): Precision = 0.68, Recall = 0.65, F1-Score = 0.67

These values confirm that the framework handles the varying quality classes well, with the highest accuracy observed for Class I and II. The decrease in performance for Class V is attributed to the small number of samples in this category, which result in a less robust learning process.

The proposed fuzzy logic framework was compared against traditional threshold-based models and decision tree classifiers as shown in Table II. The fuzzy logic framework achieved superior performance, demonstrating the ability to handle imprecise and uncertain data more effectively than traditional approaches.

Table 2
Performance metric for each model

Model	Overall Accuracy	Precision	Recall
Threshold-Based	77.8%	0.76	0.77
Decision Tree Classifier	84.2%	0.83	0.82
Proposed Fuzzy Model	89.6%	0.88	0.89

4. Conclusions

We have presented a fuzzy logic-based framework for water quality classification that addresses the limitations of traditional methods. By leveraging fuzzy set theory and integrating machine learning-based optimization, the proposed model offers improved flexibility and accuracy in classifying water quality. Experimental results demonstrate that the fuzzy framework performs well even in the presence of ambiguity and overlapping class boundaries.

Future work will focus on expanding the model to include additional water quality parameters and testing it on diverse datasets from different geographical regions. Additionally, the integration of real-time data from IoT-based water quality monitoring systems could further enhance the practical applicability of the proposed framework in dynamic environments. Overall, this research contributes to the growing body of knowledge on environmental management by providing a robust, adaptable, and interpretable solution for water quality classification.

Acknowledgement
We would like to acknowledge RedTone company for providing data from the MRANTI lake and Department of Environment (DOE), Malaysia for providing data of rivers in Malaysia. We also would like to acknowledge, Faculty of Engineering and Built Environment, Universiti Sains Islam Malaysia and Ministry of Higher Education, Malaysia under FRGS grant (FRGS/1/2024/WAS02/USIM/02/1) for the funding and support.

References
[1] "National Water Quality Standards and Water Quality Index – Department of Environment," n.d.
[2] Ahmad, Gulzar, Sehrish Aqeel, Zobia Zafar, and Kanza Fatima. "Assessment of Water Quality Using Multi-Layered Mamdani Fuzzy Inference Expert System." In *Harnessing High-Performance Computing and AI for Environmental Sustainability*, pp. 334-352. IGI Global, 2024. https://doi.org/10.4018/979-8-3693-1794-5.ch015
[3] Oladipo, Johnson O., Akinola S. Akinwumiju, O. S. Aboyeji, and Adedeji A. Adelodun. "Comparison between fuzzy logic and water quality index methods: A case of water quality assessment in Ikare community, Southwestern Nigeria." *Environmental Challenges* 3 (2021): 100038. https://doi.org/10.1016/j.envc.2021.100038
[4] Wagh, Sarita Jibhau, Pradip M. Paithane, Mohan D. Sangale, and G. M. Pondhe. "Computation of Groundwater Quality of Baramati with the help of Fuzzy Water Quality Index (FWQI)." *International Research Journal of Multidisciplinary Technovation* 5, no. 6 (2023): 43-57. https://doi.org/10.54392/IRJMT2364.
[5] Icaga, Yilmaz. "Fuzzy evaluation of water quality classification." *Ecological Indicators* 7, no. 3 (2007): 710-718. https://doi.org/10.1016/J.ECOLIND.2006.08.002
[6] Barzegar, Yas, Irina Gorelova, Francesco Bellini, and Fabrizio D'ascenzo. "Drinking water quality assessment using a fuzzy inference system method: a case study of Rome (Italy)." *International journal of environmental research and public health* 20, no. 15 (2023): 6522. https://doi.org/10.3390/IJERPH20156522
[7] Bokingkito Jr, Paul B., and Lomesindo T. Caparida. "Using Fuzzy logic for real-time water quality assessment monitoring system." In *Proceedings of the 2018 2nd International Conference on Automation, Control and Robots*, pp. 21-25. 2018. https://doi.org/10.1145/3293688.3293695
[8] Ribas, José Roberto, Juliana Crenitte Ribas Severo, and Letícia Correa Gonçalves. "A Fuzzy Multicriteria Approach to Estimate the Water Quality Index of Hydroelectric Reservoirs." *Water Resources Management* 38, no. 6 (2024): 2155-2170. https://doi.org/10.1007/S11269-024-03749-4
[9] Dewanti, Novita Ayu, and Agus Maman Abadi. "Fuzzy logic application as a tool for classifying water quality status in Gajahwong River, Yogyakarta, Indonesia." In *IOP Conference Series: Materials Science and Engineering*, vol. 546, no. 3, p. 032005. IOP Publishing, 2019. https://doi.org/10.1088/1757-899X/546/3/032005
[10] Dhaoui, Oussama, Belgacem Agoubi, Isabel Margarida Antunes, Lotfi Tlig, and Adel Kharroubi. "Groundwater quality for irrigation in an arid region—application of fuzzy logic techniques." *Environmental Science and Pollution Research* 30, no. 11 (2023): 29773-29789. https://doi.org/10.1007/s11356-022-24334-5
[11] Khan, Md Saikat Islam, Nazrul Islam, Jia Uddin, Sifatul Islam, and Mostofa Kamal Nasir. "Water quality prediction and classification based on principal component regression and gradient boosting classifier approach." *Journal of King Saud University-Computer and Information Sciences* 34, no. 8 (2022): 4773-4781. https://doi.org/10.1016/J.JKSUCI.2021.06.003