APPLICATION OF IoT FOR MONITORING WATER QUALITY SUNGAI PUSU CASE STUDY

NASSERELDEEN KABBASHI [1*], TAHSIN FUAD HASAN [1], MD ZAHANGIR ALAM [1], T SALEH[2], AISHA HASSAN [3], ABDULRAHMAN H. NOUR[4] AND MUHAMMAD ANIQ QAWIEM SAZALY[2]

[1]*Department of Chemical Engineering and Sustainability, International Islamic University Malaysia, Gombak, 53100 Kuala Lumpur, Malaysia.*
[2]*Department of Mechatronics Engineering, International Islamic University Malaysia, Gombak, 53100 Kuala Lumpur, Malaysia*
[3]*Department of Electrical Engineering, International Islamic University Malaysia, Gombak, 53100 Kuala Lumpur, Malaysia*
[4]*Faculty of Chemical and Process Engineering Technology, Universiti Malaysia Pahang, Malaysia*

**Corresponding authors:* nasreldin@iium.edu.my

ABSTRACT: The availability of clean water, a critical natural resource essential for supporting diverse ecosystems, is increasingly threatened by sediment accumulation, which negatively impacts rivers, oceans, and coastal environments. During rainy seasons, excessive sediment, including clay particles known as total suspended solids (TSS), contaminates water, compromising its quality, altering its color, and driving up treatment costs. Additionally, sediment can carry pollutants that reduce water clarity, harm aquatic life, and disrupt ecosystem functions. Conventional water treatment methods—such as coagulation, flocculation, sedimentation, and filtration are commonly employed to address these challenges. However, traditional water quality monitoring relies heavily on laboratory tests that require specialized personnel, chemicals, and expertise, which may not always be adequate for timely and effective intervention. The advent of Internet of Things (IoT) technology offers a promising alternative by enabling the real-time collection of water quality data. Furthermore, integrating soft computing technology into water quality assessment presents a more efficient, rapid, and environmentally sustainable alternative to traditional laboratory-based approaches. This research utilizes an IoT device to monitor the performance of a water treatment system and gather data on its water quality indicators. Integrating IoT devices into water treatment systems enables real-time monitoring, automated control, and predictive maintenance, leading to improved efficiency, cost savings, and proactive water quality management. These technologies enable real-time monitoring and analysis, reducing the need for manual sampling and laboratory testing, which lowers labor and operational costs. Early detection and automated intervention help prevent costly repairs and environmental damage by addressing issues before they escalate. A novel IoT-based monitoring system of integrating soft computing into water quality assessment has been achieved.

KEY WORDS: *IoT, Water quality, TSS, Turbidity, Sungai Pusu*

1. INTRODUCTION

Rivers are vital natural infrastructure, essential for providing water for agricultural, urban, and industrial purposes, and they play a crucial role in the sustainable development of any country. However, rapid economic growth and urbanization have led to continuous river pollution, severely damaging river ecosystems. Large volumes of domestic and industrial wastewater are discharged into rivers, resulting in significant pollution and compromising their ability to serve as vital resources. This has led to a marked deterioration in urban ecology and water quality. The problems of urban river pollution and ecological degradation are becoming increasingly urgent, with few natural rivers remaining by the early 20th century, as noted by [1].

Water pollution is a global issue, with organic and inorganic contaminants from agricultural, industrial, and domestic sources polluting water bodies. This pollution has severe repercussions for human health and agriculture, leading to the bioaccumulation of toxic metals in the food chain. Overall, pollution poses a significant threat to modern society, as many water sources have been compromised by the release of harmful substances. In 2015, water and soil pollution were responsible for 16% of global deaths, with approximately 92% of these occurring in developing countries [2]. River pollution also harms ecosystems and diminishes the benefits rivers provide for agriculture, urban and industrial water supply, and irrigation.

According to data from the Department of Environment (DoE) (2017), the number of rivers in Malaysia decreased from 579 in 2008 to 477 in 2019. Along with this decline, the quality of river water has also deteriorated, making it increasingly difficult to use for various purposes. The widespread effects of industrialization, combined with the accumulation of excess sediment in bodies of water, including rivers and oceans, pose significant risks to ecosystems, reduce the usability of water resources, and increase treatment costs.

To address these challenges, there is an urgent need to explore and develop innovative approaches, such as utilizing machine learning and IoT technologies, to create more robust monitoring systems capable of continuously assessing water quality parameters in real-time.

2. IoT AND WATER MANAGEMENT

The Internet of Things (IoT) enables the interconnection of devices through the internet, offering significant benefits for automating water distribution and monitoring for leaks. The fundamental structure of IoT consists of three layers as shown in Figure 1: the physical layer, where sensors collect data from the environment; the network layer, where this data is transformed into digital streams for processing; and the application layer, which delivers specific services to the user. To prevent network congestion, data must be processed immediately upon collection or stored in the cloud, because cloud structure have not a single point of control [3]. Numerous sensors are available for water monitoring, including fabricated sensors, capacitive sensors, turbidity sensors, and soil moisture sensors.

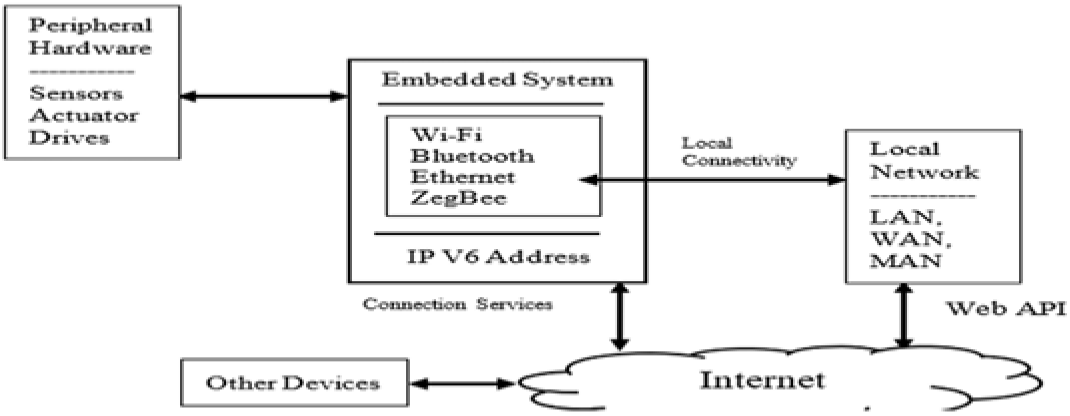

Fig. 1. Basic architecture of IoT [3]

In 2016, [4] proposed an IoT-based solution for monitoring in-pipe water quality. Their model tests water samples and analyzes data posted online, enhancing the accuracy of water measurements and identifying deviations from predefined values. Similarly, K. Gupta et al. [5] introduced a device that can be continuously monitored via a mobile app from any location. This device offers complete intelligent automation, is stable, easy to install, and has a compact design.

V. Ranjan et al. [6] proposed a smart rainwater harvesting concept using IoT. The model includes a segregation system that divides the collected water into two tanks in a 60-40 percent ratio. A rainfall detection sensor is placed atop the system to detect precipitation. The model was tested using two water samples—one containing regular drinking water and the other mixed with a mildly acidic solution, simulating varying water conditions.

Hamid et al. [7] developed a smart water quality monitoring system (SWQMS) designed for assessing and evaluating water quality in swimming pools, analyzing factors such as pH and temperature. This allows for immediate adjustments, ensuring optimal water quality and safety. The system also automates routine checks, reducing the need for manual testing, and can predict maintenance needs, thereby improving efficiency and reducing operational costs. By maintaining consistent water quality, the SWQMS helps prevent health issues and prolongs the life of pool infrastructure.

Overall, smart water management through IoT is a technology designed to collect and analyze data on a city's water supply, pressure, and delivery systems to improve the efficiency of water transportation and usage.

2.1 Research Method

The initial phase of this research project involves conducting a comprehensive literature review to establish a well-defined set of objectives. Following this, sensors and IoT modules will be carefully selected based on insights from the literature. Concurrently, various machine learning models will be rigorously analyzed using pre-existing classified datasets.

Once the sensors selection are finalized to monitor the water quality parameters like pH, turbidity, dissolved oxygen, and temperature the IoT framework will be set up and configured using technologies like Node-RED and Arduino components for monitoring of applications' performances. A preliminary trial will then be conducted to collect data from various river profile points at different time intervals. Simultaneously, an exhaustive search for the most optimal machine learning model will be carried out, with the pre-existing classified dataset serving as a benchmark for testing and selection.

The third critical phase focuses on data refinement, ensuring that any extraneous or irrelevant information is thoroughly eliminated. This meticulous data cleaning process is essential to remove potential inaccuracies and discrepancies, ensuring the reliability of the final analysis. The ultimate goal is to enable precise feature extraction, laying a solid foundation for robust and accurate water quality assessment.

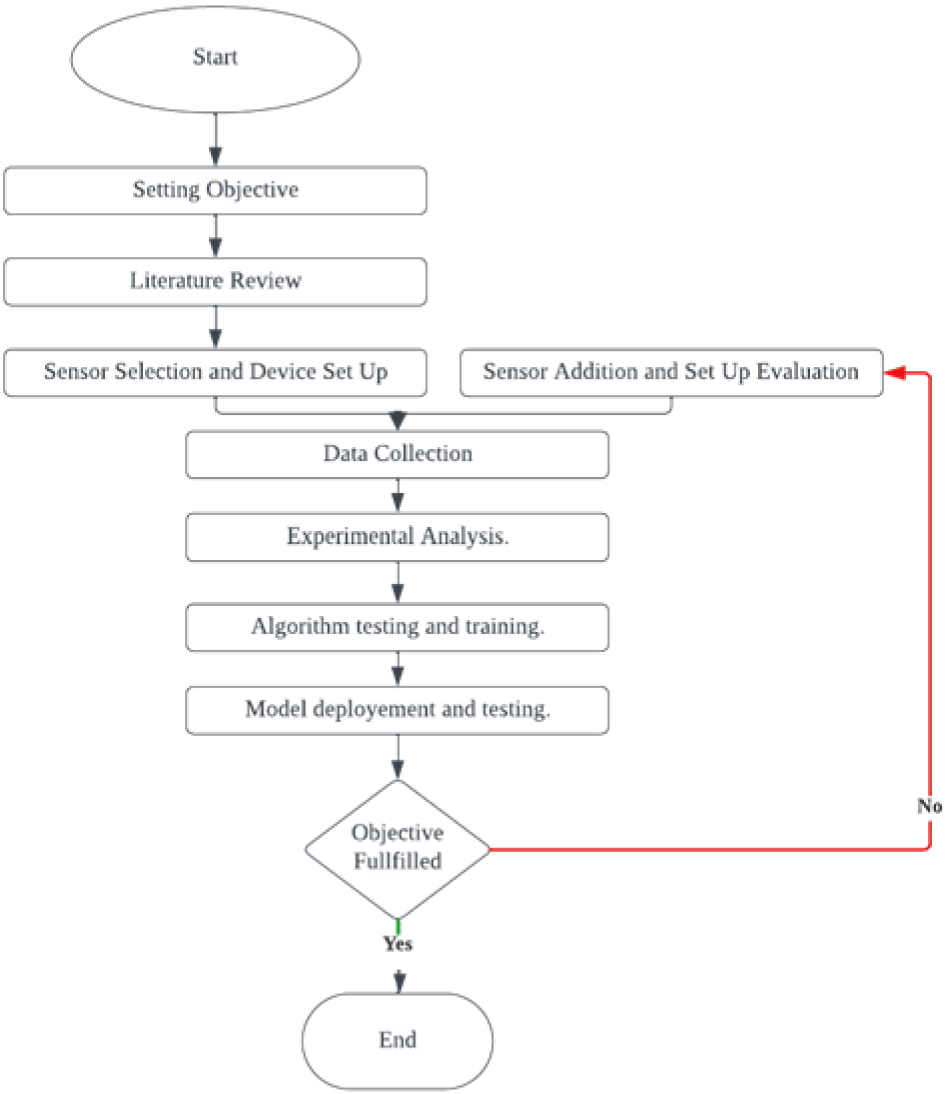

Fig. 2. Research Flowchart

2.2 System Architecture

The foundation of this system was established through the careful selection of a microcontroller, with the Arduino UNO ATmega328 emerging as the cornerstone. This choice was primarily motivated by its cost-effectiveness and, importantly, Arduino's inherent versatility in seamlessly integrating a variety of sensor systems—a crucial factor in optimizing our data collection capabilities. Sensors connect to the Arduino UNO via analog or digital pins, where the Arduino processes the data, converts it into meaningful readings, and then transmits it through serial communication or wireless modules for real-time monitoring and analysis.

Our initial phase of data collection and experimentation relied on the integration of four essential water parameter sensors, each playing a critical role in capturing specific aspects of water quality as highlighted in Table 1:

Temperature Sensor: This sensor provided vital insights into water temperature fluctuations, a key factor affecting aquatic life and various biochemical processes.

pH Sensor: Essential for determining the acidity or alkalinity of the water, the pH sensor delivered invaluable data for assessing the water's suitability for a wide range of applications.

Turbidity Sensor: This sensor addressed water clarity and cloudiness, serving as a crucial indicator of particulate matter and sediment levels—factors that are pivotal for maintaining ecosystem health.

TDS (Total Dissolved Solids) Sensor: Instrumental in measuring the concentration of dissolved substances in the water, the TDS sensor helped gauge overall purity and the water's suitability for different uses.

In the second phase of data collection and testing, we expanded our monitoring capabilities by including an additional sensor:

Dissolved Oxygen Sensor: The Dissolved Oxygen sensor was introduced in the second phase to provide a more comprehensive assessment of water quality, as oxygen levels are a key indicator of aquatic health. The expanded monitoring capabilities allowed the system to detect variations in oxygen content, enhancing the ability to identify potential issues such as pollution or eutrophication, thus improving the overall effectiveness and reliability of the water quality monitoring system.

Table 1: Sensor details.

Sensor	Brand/Product ID	Range of Measurement
Temperature	DS18b20	-55 – 125 C
pH	Gravity: Analog pH Sensor/Meter Kit V2 by DFRobots	0-14
Turbidity	Gravity-Analog Turbidity Sensor For Arduino	0-4.5 V
TDS	SEN0244 by Dfrobot	0 to 1000 ppm
Dissolved Oxygen	Gravity: Analog Dissolved Oxygen Sensor by DFRobots	0-20 mg/L

3. RESULTS AND DISCUSSION

The dataset included over 877 entries and upon initial data cleaning and processing which included removal of duplicates, null values and faulty data set moved the final dataset to 526 entries from the 3 dataset classes. The "faulty data" removed during the data cleaning process typically included sensor readings that were inconsistent, out-of-range, or clearly erroneous due to sensor malfunctions, environmental interference, or transmission errors. This data could also include incomplete records or anomalies caused by temporary system glitches. Removing this faulty data was crucial to ensure the accuracy and reliability of the final analysis.

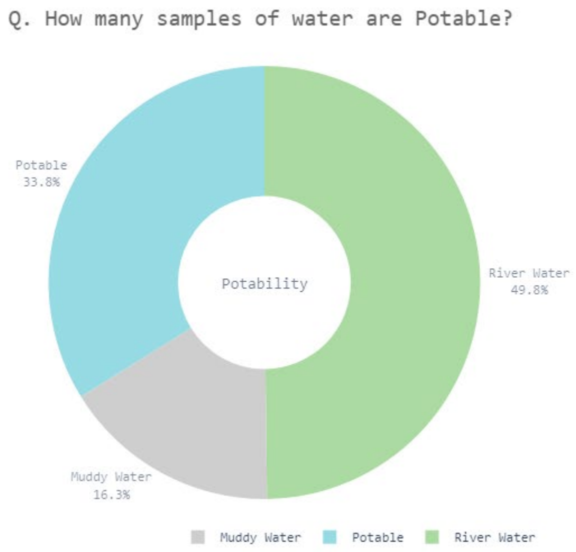

Fig 3: The distribution of the dataset

Figure 3 illustrates the distribution of the dataset according to the potability class, revealing a certain degree of class imbalance. Specifically, river water accounts for over 45% of the dataset, potable water represents 33.8% (indicating a mild imbalance), and muddy water comprises 16.3% (indicating a moderate imbalance). These class imbalances present challenges, particularly due to the overrepresentation of the river water class, which could potentially bias predictive outcomes.

The correlation matrix as in Figure 4 provides valuable insights into the impact of each parameter on water quality. Notably, the temperature sensor readings show no significant correlation with other features or with potability itself. This lack of correlation can be attributed to the multifactorial nature of temperature, which can vary widely based on factors like the collection point and time of day, regardless of water potability. As a result, the temperature feature was excluded from the model training and testing phases due to its limited predictive value.

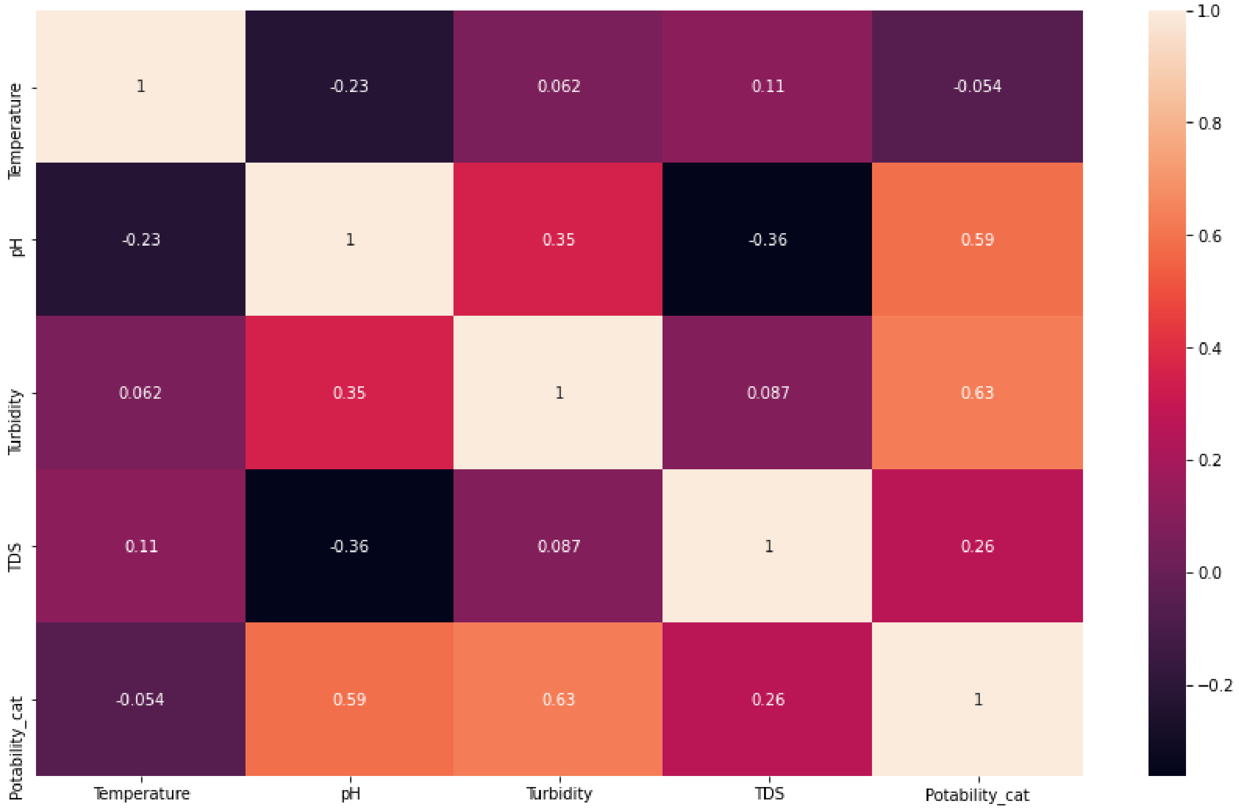

Fig 4: Correlation Matrix for dataset

In contrast, turbidity and pH exhibit a strong overall correlation with potability, approximately 50%, highlighting their importance in classification tasks. In the context of feature selection or model tuning, the correlations between pH and turbidity were leveraged to improve the predictive accuracy of the machine learning models. By analyzing the relationship between these two parameters, the model could identify patterns that are indicative of water quality issues. For instance, a strong correlation might suggest that changes in pH often accompany variations in turbidity, which could be a signal of pollution or other water quality problems. Understanding these correlations allowed for more informed feature selection, where the model prioritized parameters with stronger predictive power, ultimately leading to more accurate and reliable water quality assessments. Additionally, during model tuning, these correlations helped refine the model's sensitivity to variations in pH and turbidity, ensuring that it could better detect and respond to changes in water quality.

On the other hand, total dissolved solids (TDS) show a lower correlation with potability at 23%. Additionally, minimal correlation was observed between each feature and the other features in the dataset. These findings emphasize the distinct contributions of each parameter in assessing water quality and underscore the critical role of feature selection in model development.

The datasets achieved an accuracy exceeding 85% across all models, largely due to the dataset's size and the existing clusters among the classes. Gradient Boosting excelled with an impressive accuracy of 99.3%. The model was optimized through careful selection of features and fine-tuning of hyperparameters to achieve high accuracy. Techniques such as crossvalidation, grid search, and regularization were employed to ensure that the model was not overfitting and could generalize well to new data. However, achieving 99.3% accuracy likely involved trade-offs, such as increased computational complexity or the need for a larger dataset to prevent overfitting. Additionally, there might have been a balance between maximizing accuracy and maintaining real-time processing efficiency, particularly in a resource-constrained environment. Similarly, other ensemble models, such as Random Forests, along with non-linear models like Decision Trees and K-Nearest Neighbors (KNN), demonstrated exceptional performance in classifying water potability across multiple classes. In contrast, linear models such as Logistic Regression and Support Vector Machines (SVMs) faced challenges with the non-linear relationships in the data, which likely impacted their classification accuracy in this multi-class scenario, these models assume linear separability, which is not suitable for the complex relationships in the dataset.

ACKNOWLEDGEMENT

The author wishes to acknowledge the support received from the KOE and the CHES department, as well as express gratitude to all the technicians and colleagues involved. Special thanks are extended to RMC IIUM for enabling this grant under Project ID: IUMP-SRCG22-005-0005 and Project Title: "Investigation of Water Quality Monitoring Using IoT and Machine Learning Techniques: Sungai Pusu River IIUM Gombak Case Study."

REFERENCES

[1] Wang, J., Liu, X. D., & Lu, J. (2012). Urban River Pollution Control and Remediation. Procedia Environmental Sciences, 13(2011), 1856–1862. https://doi.org/10.1016/j.proenv.2012.01.179.

[2] Landrigan, P. J., Fuller, R., Acosta, N. J. R., Adeyi, O., Arnold, R., Basu, N. (Nil), Baldé, A. B., Bertollini, R., Bose-O'Reilly, S., Boufford, J. I., Breysse, P. N., Chiles, T., Mahidol, C., Coll-Seck, A. M., Cropper, M. L., Fobil, J., Fuster, V., Greenstone, M., Haines, A., Zhong, M. (2018). The Lancet Commission on pollution and health. The Lancet Commissions, 391(10119), 462–512. https://doi.org/10.1016/S0140- 6736(17)32345-0.

[3] Yasin, H. M., Zeebaree, S. R., Sadeeq, M. A., Ameen, S. Y., Ibrahim, I. M., Zebari, R. R., & Sallow, A. B. (2021). IoT and ICT based smart water management, monitoring and controlling system: A review. Asian Journal of Research in Computer Science, 8(2), 42-56.

[4] Geetha S, Gouthami S. (2016) Internet of things enabled real time water quality monitoring system. Smart Water.; 2:1-19.

[5] Gupta K, Kulkarni M, Magdum M, Baldawa Y, Patil S. (2018) Smart water management in housing societies using IoT. Proceedings of the International Conference on Inventive Communication Computational Technologies,1609-1613.

[6] Ranjan V, Reddy MV, Irshad M, Joshi N. (2020) The Internet of Things (IOT) based smart rain water harvesting system. IEEE, 6th International Conference on Signal Processing and Communication, ICSC 2020;302-305.

[7] Hamid, S.A., Rahim, A.M., Fadhlullah, S.Y., Abdullah, S.B., Muhammad, Z., & Leh, N.A. (2020). IoT based Water Quality Monitoring System and Evaluation. 2020 10th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), 102-106.