REAL-TIME DRINKING WATER QUALITY ASSESSMENT USING AN OPTOELECTRONIC SYSTEM

Salauat Kengesbayev
PhD Student, Department of Broadcasting Systems, Tashkent University of Information Technologies, E-mail: salawatkenesbaev@gmail.com

Abstract: *This article proposes the concept of an optoelectronic liquid monitoring system based on the Attenuated Total Reflection (ATR) technology. The system integrates an ATR sensor module, an Internet of Things (IoT) device (Raspberry Pi 4B), and cloud-based Machine Learning (ML) algorithms, enabling real-time assessment of liquid quality. The proposed approach is designed to identify and predict categorical states of liquids by analyzing their optical parameters, with potential applications in industry, environmental monitoring, and food safety.*

Keywords: *ATR, Optoelectronic sensor, IoT, Raspberry Pi 4B, Machine learning, Liquid quality, Optical parameters, Real-time monitoring.*

INTRODUCTION

Real-time monitoring of liquid quality has become a critical task in environmental protection, food safety, pharmaceutical manufacturing, and water supply systems [1]. Traditional laboratory-based testing methods are often time-consuming, costly, and incapable of providing real-time data. Consequently, in recent years, there has been a significant increase in the demand for monitoring systems based on IoT technologies and intelligent sensor networks [2], [3].

IoT modules, particularly microcontroller and microcomputer platforms such as Raspberry Pi 4B and ESP32, enable the collection, processing, and transmission of data from various sensors in real time to a central server or cloud platform via a network [4], [5]. These modules are highly adaptable, supporting multiple communication protocols (Wi-Fi, Ethernet, LoRa, ZigBee), thereby facilitating large-scale deployment of the system.

ML algorithms hosted on the cloud infrastructure analyze the collected data and perform three main functions:

1. Anomaly Detection – identifies deviations of system parameters from the normal range, thereby enabling early warning of potential malfunctions.

2. Predictive Analytics – studies time series and historical data (database) to forecast future changes, for example, predicting when the liquid quality will become unsuitable after a certain period.

3. Automated Response – upon detecting hazardous parameters through liquid state prediction, triggers the alarm system in real time, stops the flow via a relay, or sends warning notifications [6].

In recent years, the direct numerical method has been widely applied in solving complex differential equations. This approach enables the analysis of bifurcation diagrams

and Lyapunov exponents to study the behavior of systems under uncertainty, as well as to predict physicochemical parameters in liquid monitoring in real time [7].

The integration of IoT and ML significantly enhances the flexibility of the system (rapid adaptation to real-time conditions), accuracy (error reduction through ML), and efficiency (resource savings via automated control). Additionally, cloud platforms enable remote monitoring and centralized data analysis, which reduces maintenance costs and increases system reliability.

Adequate IoT integration and the application of machine learning algorithms.

The proposed system combines the advantages of this technology with IoT and ML to provide accurate, fast, and predictive control.

Existing liquid monitoring systems face issues such as the lack of real-time prediction capability, limited long-term stability, latency between the sensor and the cloud, and a restricted set of measurable parameters.

The aim of this study is to develop a stable, real-time, and predictive liquid monitoring system through the integration of an ATR-based optoelectronic sensor, IoT, and ML.

MAIN PART

The proposed IoT-based liquid quality monitoring system is organized according to a three-layer architecture consisting of the measurement and processing layer, intermediate layer, and application layer. Each layer plays a crucial role in ensuring accurate and real-time monitoring and control of water quality. These layers enable the efficient collection, transmission, and analysis of data by leveraging modern technologies and communication frameworks.

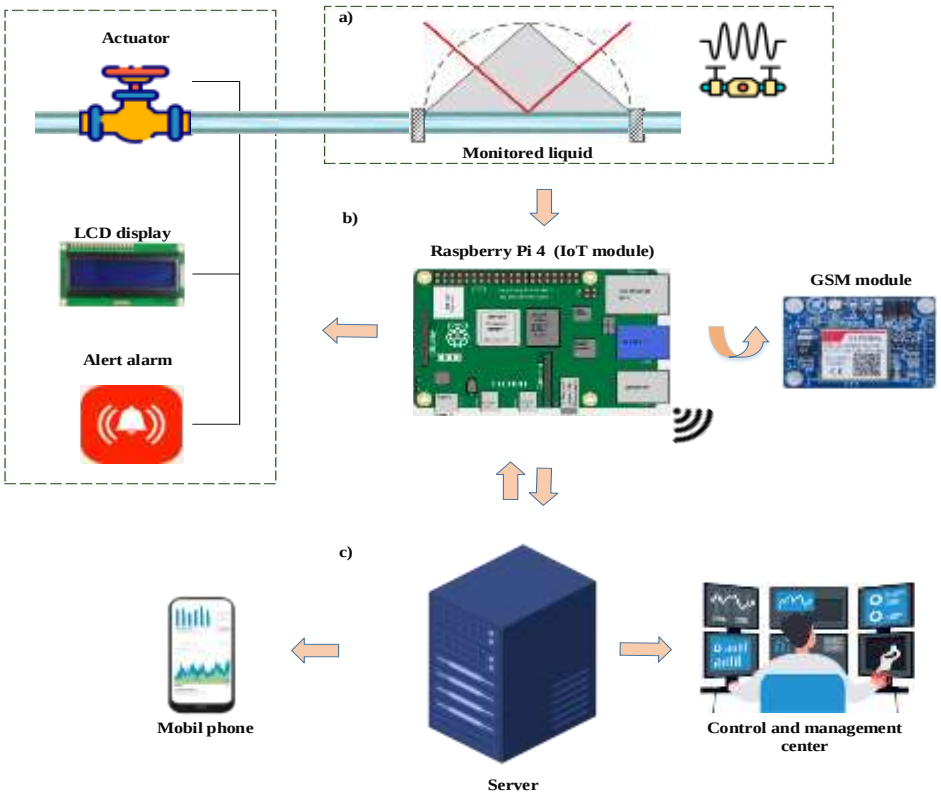

Fig. 1. Proposed layered architecture: (a) measurement and actuation layer,(b) communication and server layer,(c) control and management layer.

The proposed system employs machine learning (ML) techniques to significantly enhance the accuracy and efficiency of optoelectronic control systems based on the ATR effect. In spectral data, the relationship between absorbance and concentration is typically described by the Beer–Lambert law. However, in multi-component liquid mixtures, interactions between signals often introduce nonlinear dependencies, making conventional analytical models insufficient. Therefore, the application of ML algorithms becomes essential for accurately identifying these complex relationships.

The Multioutput Regressor is a wrapper model that extends single-output regression algorithms to handle multiple target variables simultaneously. As illustrated in Figure 3, a separate regression model is trained for each output variable. This approach is particularly advantageous for predicting the quality indicators of liquids based on their spectral data, as it enables simultaneous estimation of several chemical and physical parameters from the same input dataset[8].

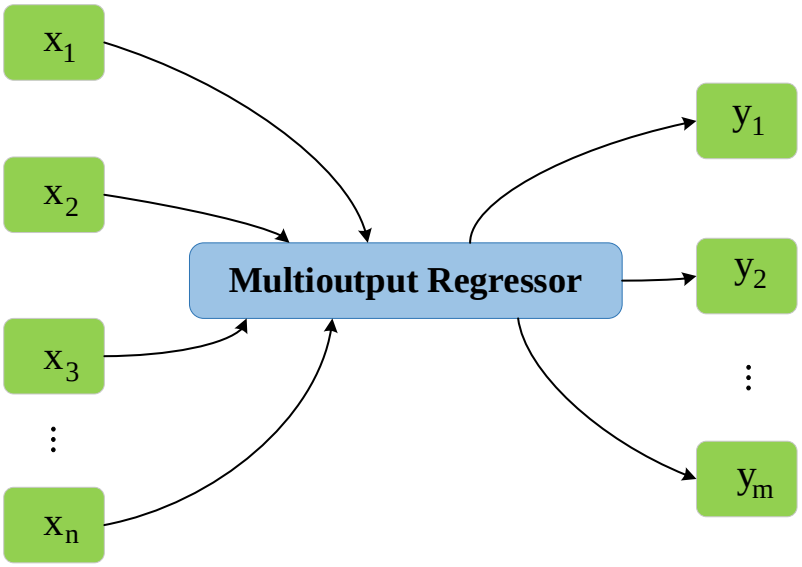

Fig. 2. Input and output data flow in the multi-output regression model.

In our research, XGBoost is employed as the core model within the Multioutput Regressor framework. This approach enables the simultaneous prediction of multiple physicochemical parameters of liquids using their spectral data (absorbance values). For each output variable, an individual XGBoost regressor is trained, and the results are then aggregated to produce the final output.

An additional advantage of XGBoost lies in its capability to analyze feature importance, which reveals which wavelengths (spectral ranges) are most influential in determining specific target parameters. Consequently, the model not only performs accurate predictions but also provides valuable insights into the key spectral features affecting the liquid's characteristics.

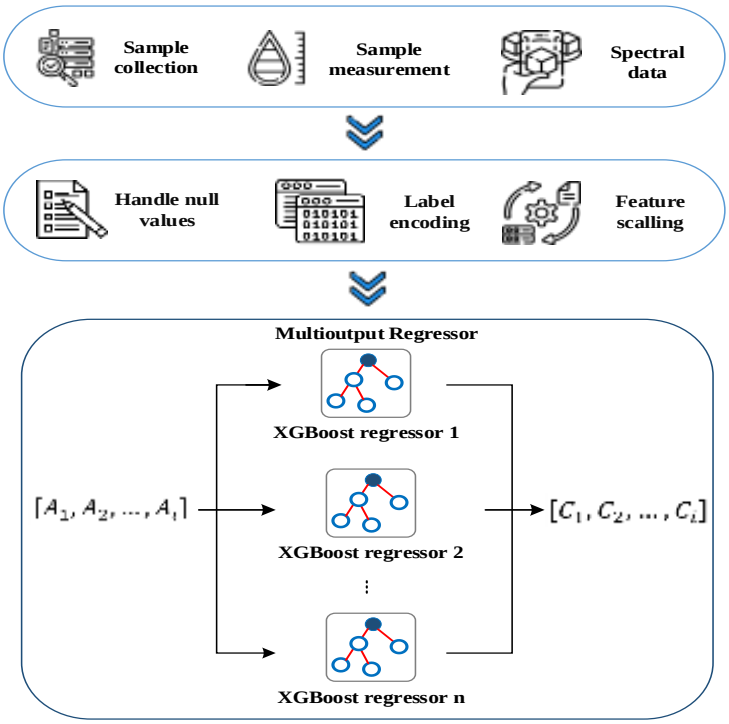

Fig. 3. Training stages of the "Intelligent Measurement" model.

The model training process is carried out in three main stages. In the first stage, samples are collected, and during measurement, their absorbance values and concentration levels are determined. In the second stage, the acquired data are cleaned, missing values are filled in, and the dataset is scaled and normalized to prepare it for model training. In the third stage, XGBoost regressors are trained within a Multioutput Regressor framework, producing separate prediction results for each parameter.

As a result, an "Intelligent Measurement" model is developed, capable of simultaneously determining multiple physical and chemical parameters of liquids in real time.

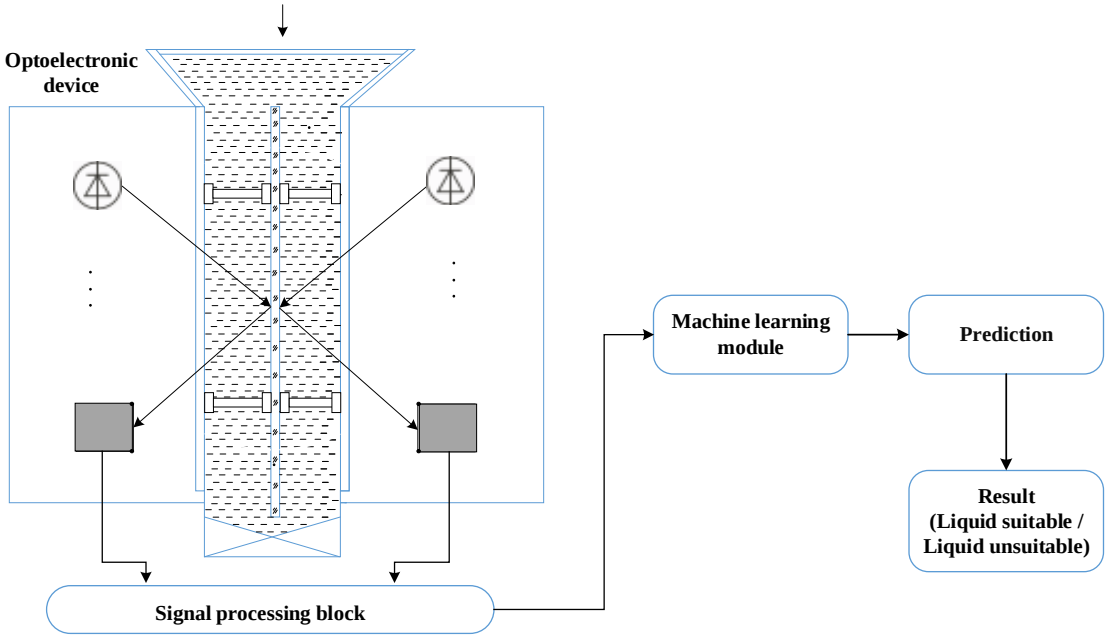

Fig. 4.Functional diagram of the "Intelligent Measurement" model for liquid quality assessment.

When new samples are introduced into the model, the results obtained from each regressor are aggregated to produce the final prediction. In the final stage, the predicted physicochemical parameters are visualized and compared with the threshold values, after which a conclusion is made regarding whether the liquid is "suitable" or "unsuitable."

RESULTS AND DISCUSSION

During the experimental phase, the primary focus was placed on key indicators commonly found in drinking water, which are considered essential from both ecological and sanitary perspectives. In particular, special attention was given to determining the parameters of Dissolved Organic Carbon (DOC), ammonium ions (NH_4^+), sulfates (SO_4^{2-}), orthophosphates (PO_4^{3-}), nitrates (NO_3^-), and nitrites (NO_2^-). These components are regarded as the main indicators defining the organic pollution level, nutrient content, and overall environmental quality of water.

Based on the dataset formed during the research-comprising 487 samples with 277 spectral features-various machine learning algorithms such as Linear Regression (LR), Random Forest (RF), Support Vector Regression (SVR), and XGBoost were tested. These algorithms were applied to predict the physicochemical parameters of liquids (using drinking water as an example) from BTIQ spectroscopic absorbance data. Their performance was evaluated using the Root Mean Square Error (RMSE) and Coefficient of Determination (R^2) metrics.

The obtained results (Figure 4) demonstrated that:

- The XGBoost algorithm achieved the highest accuracy and lowest error rate overall, making it the most optimal approach among the tested models.
- The SVR algorithm showed relatively high R^2 values for certain parameters but exhibited lower stability compared to XGBoost.
- The Random Forest (RF) algorithm produced moderately stable results, though with higher error levels for some parameters.
- The Linear Regression (LR) model served as a baseline comparison method and showed the lowest accuracy among all evaluated algorithms.

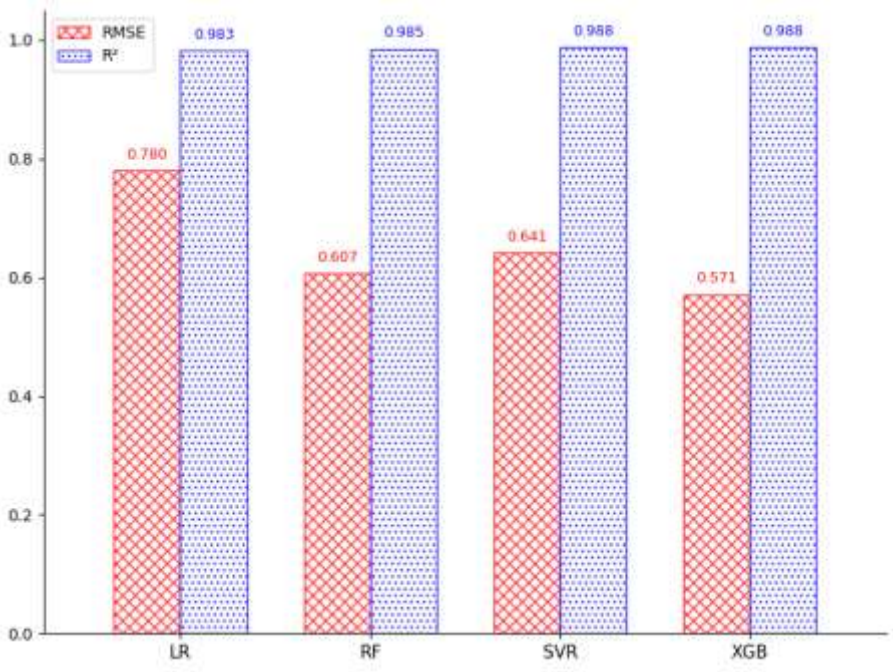

Fig. 5. Comparison of machine learning algorithms based on average RMSE and R^2 metrics.

Thus, in this study, the XGBoost algorithm is recommended as the primary choice for multi-parameter physicochemical prediction. The machine learning module was trained on the dataset using regression models and subsequently tested on a separate test set.

REFERENCES:

[1]G. M. E. Silva, D. F. Campos, J. A. T. Brasil, M. Tremblay, E. M. Mendiondo, and F. Ghiglieno, "Advances in technological research for online and in situ water quality monitoring—A review," Sustainability, vol. 14, no. 9, p. 5059, Apr. 2022, doi: 10.3390/su14095059.

[2] J. Kumar, R. Gupta, S. Sharma, T. Chakrabarti, P. Chakrabarti and M. Margala, "IoT-Enabled Advanced Water Quality Monitoring System for Pond Management and Environmental Conservation," in IEEE Access, vol. 12, pp. 58156-58167, 2024, doi: 10.1109/ACCESS.2024.3391807.

[3] M. Govindasamy, K. Jayanthi and S. Rajagopan, "IoT Product on Smart Water Quality Monitoring System (Iot Wq-Kit) for Puducherry Union Territory," 2023 Second International Conference on Advances in Computational Intelligence and Communication (ICACIC), Puducherry, India, 2023, pp. 1-5, doi: 10.1109/ICACIC59454.2023.10435019.

[4] N. S. Kumar and P. Nirmalkumar, "A Novel Architecture of Smart Healthcare System on Integration of Cloud Computing and IoT," 2019 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2019, pp. 0940-0944, doi: 10.1109/ICCSP.2019.8698048.

[5] D. Abhiram, K. D. Reddy, P. K. Reddy and M. Belwal, "IoT-based Environmental Monitoring System," 2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), Bengaluru, India, 2025, pp. 764-769, doi: 10.1109/IDCIOT64235.2025.10915186.

[6] U. G. Sharanya, K. M. Birabbi, B. H. Sahana, D. M. Kumar, N. Sharmila and S. Mallikarjunaswamy, "Design and Implementation of IoT-based Water Quality and Leakage Monitoring System for Urban Water Systems Using Machine Learning Algorithms," 2024 Second International Conference on Networks, Multimedia and Information Technology (NMITCON), Bengaluru, India, 2024, pp. 1-5, doi: 10.1109/NMITCON62075.2024.10698922.

[7] S. Tashmetov, J. C. Chedjou, K. Nosirov, A. Berdiyev and E. Norov, "Computational Approaches for Solving Complex Differential Equations: Implementation of the Direct Numerical Method with Modeling and Analysis of Bifurcation Diagrams and Lyapunov Exponent," 2024 International Conference on Information Science and Communications Technologies (ICISCT), Seoul, Korea, Republic of, 2024, pp. 151-155, doi: 10.1109/ICISCT64202.2024.10957210.

[8] Borchani, H.; Varando, G.; Bielza, C.; Larranaga, P.; Discovery, K. A survey on multi-output regression. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 2015, 5, 216–233.