Journal of King Saud University - Science 36 (2024) 103409

Contents lists available at ScienceDirect

Journal of King Saud University - Science

journal homepage: www.sciencedirect.com

Full Length Article

Real-time water quality monitoring using AI-enabled sensors: Detection of contaminants and UV disinfection analysis in smart urban water systems

Yeliz Durgun

Department of Biomedical Equipment Technology Program, Turhal Vocational School, Gaziosmanpaşa University, Tokat 60300, Türkiye

ARTICLE INFO

Keywords:
Water quality assessment
Multispectral spectroscopic sensors
Artificial intelligence in environmental monitoring
Machine learning in water safety

ABSTRACT

This study introduces a novel method for assessing water quality, employing a cutting-edge sensor system integrated with artificial intelligence (AI) technologies. Addressing the global challenge of water scarcity and pollution, the research focuses on the innovative use of spectroscopic analysis for real-time water quality monitoring. The study evaluates the effectiveness of this system in distinguishing between clean, contaminated, and UV-disinfected water samples, highlighting its precision in detecting variations in water quality. Central to the research is the deployment of advanced machine learning algorithms, including Random Forest, Support Vector Machines (SVM), and Neural Networks, to process and classify spectral data. These models demonstrate remarkable accuracy in real-time classification, underscoring the synergy between AI and environmental science in addressing critical public health issues. Significantly, the study showcases the potential of UV disinfection in water treatment, as evidenced by the spectral changes observed in disinfected water samples. This aspect of the research emphasizes the role of spectral analysis in verifying the efficacy of water treatment processes. Overall, this study paves the way for more sophisticated and accessible water quality monitoring systems, offering a promising solution to one of the most pressing environmental challenges. The integration of AI and spectral analysis in this research offers a breakthrough in ensuring a safe water supply and effective water resource management. This study utilizes advanced machine learning algorithms, including Random Forest, Support Vector Machines (SVM), and Neural Networks, for water quality assessment. These models process and classify spectral data with high precision, highlighting variations in water quality.

1. Introduction

Water serves as an indispensable resource crucial for the sustenance of future generations and ecosystems. This vital importance is underscored by the fact that while Earth is covered by 2/3 water, only a mere 2.5 % of it is fresh water, a fraction that is rapidly diminishing under the pressures of modern challenges (Ajakwe et al., 2022; Bria et al., 2020). The scarcity of water is becoming a critical global issue, with projections indicating many developing countries may face severe shortages by 2025 (du Plessis and du Plessis (2019); Mancosu et al., 2015; Boretti and Rosa, 2019). Safe drinking water is a cornerstone of public health. It must be free from microbiological and chemical substances that pose health risks (Wu, 2020; World Health Organization, 2019).

The presence of pathogenic microorganisms, often introduced into water sources through human and animal feces, is a major concern. These pathogens are particularly linked to acute gastrointestinal diseases, presenting a significant public health challenge (Egbueri, 2021; Sanderson et al., 2019). Annually, more than 5 million deaths, mostly among children, are attributed to water pollution, highlighting the dire need for effective purification and pathogen removal processes in municipal water supplies (Gürsoy and Atun, 2019).

The challenges of water quality are further exacerbated by modern urbanization and industrialization. Unregulated waste disposal and inefficient sewage systems in rural areas contribute to groundwater pollution, compromising the quality of municipal water supplies. This urban–rural dichotomy creates a complex web of pollution sources that feed into lakes and rivers, eventually contaminating municipal water systems (Menon et al., 2017; Suganthi et al., 2019).

To navigate these multifaceted challenges, innovative approaches leveraging edge computing and machine learning technologies have emerged. These technologies offer promising solutions for developing smart optical sensors for water quality monitoring (Whaiduzzaman et al., 2022; Alshami et al., 2024). These sensors, operating without direct contact with water, enable the rapid detection and monitoring of potable water, significantly contributing to public health and safety (Muncan et al., 2020).

E-mail address: yeliz.durgun@gop.edu.tr.

https://doi.org/10.1016/j.jksus.2024.103409
Received 5 May 2024; Received in revised form 19 July 2024; Accepted 22 August 2024
Available online 27 August 2024

Spectroscopic analysis plays a crucial role in water quality assessment. Techniques such as near-infrared (NIR) spectroscopy have been applied to monitor various stages in water treatment systems, revealing distinct spectral patterns for different water states (Niroumand-Jadidi and Bovolo, 2022). This study leverages spectroscopic measurements to provide a nuanced understanding of water quality and the efficacy of treatment processes.

In the realm of water quality assessment, spectral measurement techniques have played a pivotal role. Near-infrared spectra and other spectral measurements have been applied to monitor various stages in water treatment systems, revealing distinctive spectral patterns for different water states. This kind of spectral analysis provides a nuanced understanding of water quality and treatment efficacy (Niroumand-Jadidi et al., 2022).

Furthermore, the integration of remote sensing and field spectral measurements has been explored for assessing water quality. Studies in this domain have demonstrated significant potential in evaluating water quality, particularly in investigating surface water pollution using an integrated approach that combines remote sensing data with on-site measurements (Ovidiu et al., 2021). Additionally, the development of smart water quality monitoring systems utilizing edge computing has marked a significant advancement. These systems enable automatic water quality measurement through sensor modules that communicate data to a centralized analysis system, making real-time monitoring and management of water resources more efficient and effective (Peterson et al., 2019).

The application of neural networks for multi-component spectral detection in water quality inspection has also shown promising results. This approach provides a reference method for quantitative analysis in environmental assessment, harnessing the power of advanced data analysis techniques to enhance water quality monitoring (Pooja et al., 2020).

Lastly, the use of high-resolution remote sensing data for water quality analysis has been another noteworthy development. This approach has utilized high-resolution imagery and the red-edge portion of the electromagnetic spectrum to monitor various water quality parameters in small water areas, further emphasizing the potential of advanced remote sensing techniques in water quality assessment (Sahu et al., 2022).

In conclusion, the integration of spectral measurements, edge computing, advanced data analysis techniques, and neural networks represents a promising approach for water quality assessment. These advancements provide a solid foundation for the development of more effective and efficient water quality monitoring systems, which are crucial for ensuring safe water access and public health.

Artificial Intelligence (AI) and Machine Learning (ML) technologies offer innovative solutions for water quality assessment. AI is utilized for processing and managing large datasets, while ML algorithms are employed to identify patterns and classify data. In this study, both technologies are integrated to provide real-time evaluations of water quality, demonstrating their potential in environmental monitoring and resource management.

2. Materials and methods

2.1. Artificial intelligence-enabled sensor system design for detection of water pollution

This study introduces an innovative spectroscopic sensor system integrated with artificial intelligence (AI), designed for the detection and analysis of water pollution. The system architecture incorporates various components, each engineered for optimized data collection and processing, as visualized in Fig. 1. The principal component of the system is a circuit board containing the AS7265x sensor, which is serially connected to NODEMCU ESP8266 board via an I2C interface cable. configuration is crucial for capturing detailed spectral data from water samples. The data collected by the spectral sensor is transmitted serially to the NODEMCU ESP8266 and then wirelessly to a personal computer via the ESP8266′s Wi-Fi module. This data to transmission process is a critical step for the subsequent storage and which is fundamental to ensuring accuracy inwater quality analysis.

The components displayed in the visual can be described as follows:

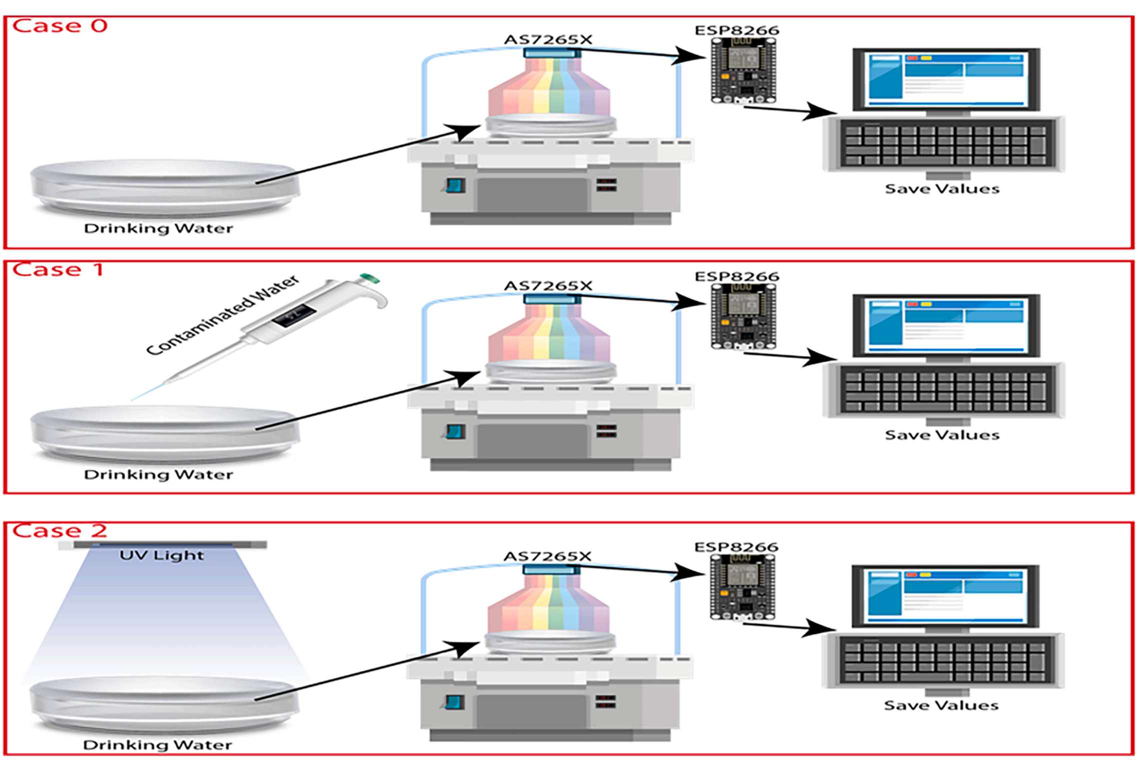

Fig. 1. AI enabled sensor system.

Clean Water Containers (Case 0): The system begins with a container holding clean water samples for spectral measurement. This clean water state provides a reference point for the system's capability to detect pollution.

Contamination Unit (Case 1): In the second phase, contaminants are added to the clean water. This is typically done with a syringe in a precise manner and is used to test the system's ability to detect contamination.

UV Disinfection Unit (Case 2): Finally, the state of water disinfected with UV light is shown. The UV lamp disinfects the water by killing microbial contaminants and demonstrates the sensor system's ability to collect spectral data from the cleaned water.

In each case, the circuit board containing the AS7265x sensor (Durgun, 2024) detects light reflected from the water sample and transmits this data to the NODEMCU ESP8266. This data is then sent to the personal computer via the Wi-Fi module for processing and analysis. In this system, the computer fulfills the tasks of data storage as well as data processing and water quality analysis through AI software. Each of these three scenarios represents specially designed situations to evaluate various aspects of water quality and provides valuable data on how the system design can detect and analyze different states of water.

The sensor system employed in this study consists of an AS7265x sensor serially connected to a NODEMCU ESP8266 board via an I2C interface cable. The AS7265x sensor is pivotal for capturing detailed spectral data from water samples. The collected data is transmitted serially to the NODEMCU ESP8266, and then wirelessly to a personal computer via the ESP8266′s Wi-Fi module. This process is crucial for ensuring accuracy in water quality analysis. Fig. 1 illustrates the system components: Clean Water Containers (Case 0), Contamination Unit (Case 1), and UV Disinfection Unit (Case 2). The AS7265x sensor detects light reflected from the water samples and transmits this data to the NODEMCU ESP8266, which subsequently sends it to a personal computer for storage and analysis. This setup allows for comprehensive data collection and real-time monitoring of water quality.

2.1.1. *Multispectral sensor*

At the core of our design is the SPARKFUN AS7265X module, a state of the art triple spectroscopic of sensor array that integrates three distinct optical sensors: i) AS72651,ii) AS72652, and iii) AS72653. Each sensor is equipped with a set of six on board optical filters, which together provide comprehensive spectral coverage across the ultraviolet (UV), visible (Vis), and near-infrared (NIR) ranges, from 410 nm to 940 nm. This configuration allows for the precise and sensitive capture of spectral information, which is crucial for the detection and quantification of various water contaminants. As illustrated in the accompanying Fig. 2, the 18 channel spectral response of the AS7265X sensors demonstrates the normalized responsivity of each channel across the monitored wavelength spectrum. Each peak represents the responsivity of a particular channel corresponding to a specific wavelength, with peaks labeled from A to U, denoting wavelengths from 410 nm to 940 nm. The varying heights of these peaks reflect the relative sensitivity of each sensor channel at different wavelengths. This range allows for precise detection and quantification of various contaminants. The spectral data collected is then processed and analyzed to determine water quality.

The spectral response graph provides an insight into the sensors' capabilities to discrim- inate between different wavelengths, which is essential for identifying and analyzing a wide range of substances present in water samples. For instance, the peaks in the blue region (410 nm–485 nm) are critical for detecting substances that absorb or scatter light in the UV spectrum, while the red and NIR peaks (610 nm–940 nm) are indicative of the sensors' ability to detect organic and inorganic matter that typically absorbs light in the longer wavelengths.The AS7265X sensors' design enables the concurrent detection of multiple wavelengths, allowing for the simultaneous analysis of various chemical

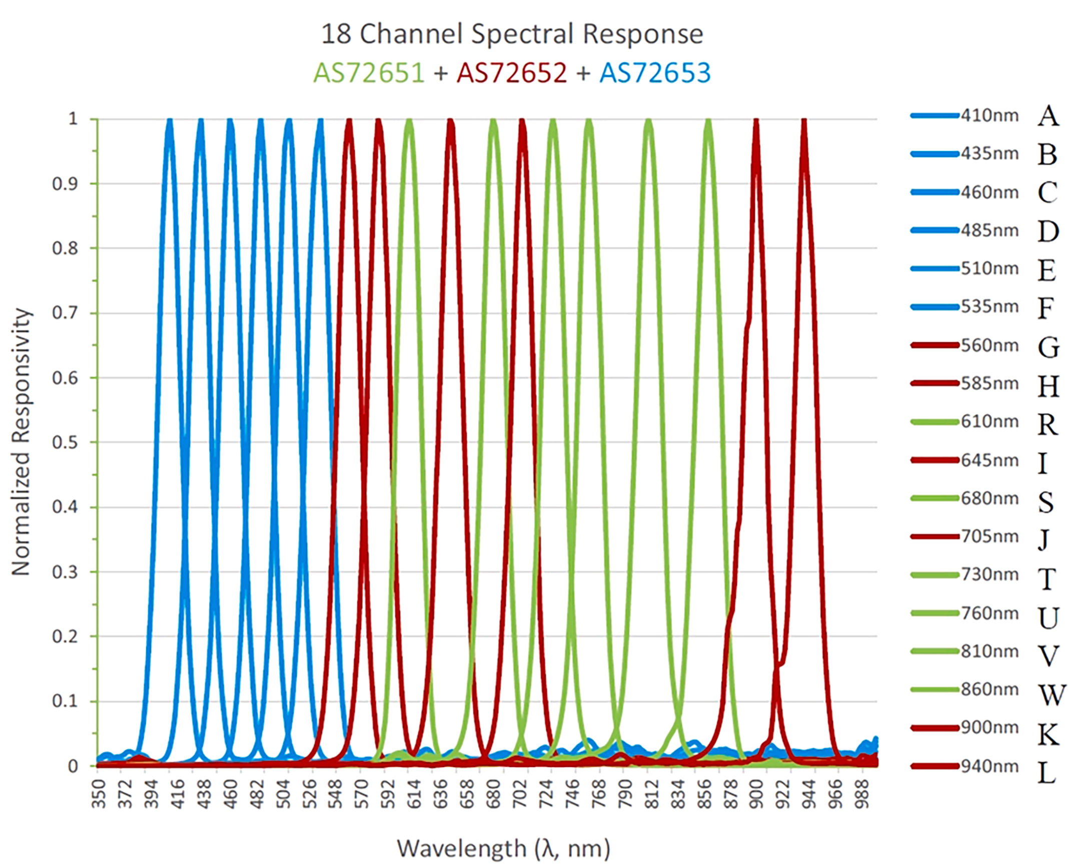

Fig. 2. 18 channel spectral response of AS7265X sensors (Alshami et al., 2024).

and biological constituents in water. This multiplexed approach not only increases the efficiency of the detection process but also significantly enhances the analytical capabilities of the system, making it a robust tool for comprehensive water quality assessment.

2.1.2. NodeMCU ESP8266

NodeMCU, an open-source firmware and development kit supports IoT-based application development, allowing for the flexible programming of the sensor system in various languages using the Arduino IDE. The ESP8266 module enhances the system with Wi-Fi capabilities, integral to data transmission (Dolcel et al., 2021).

2.1.3. Machine learning models

To classify the water samples into categories of clean water, contaminated water, and UV-treated contaminated water, three machine learning models were employed: Random Forest, Support Vector Machines (SVM), and Neural Networks.

The Random Forest model used in this study consists of 100 decision trees. Each tree is trained with a random subset of the data to ensure robustness and reduce overfitting. The maximum depth of each tree is set to 10, and the criterion used for splitting nodes is 'gini impurity'. Hyperparameters were optimized using grid search cross-validation. The key hyperparameters tuned include the number of trees, maximum depth, and the criterion for node splitting. The spectral data collected was split into training (70 %) and testing (30 %) datasets. The model was trained using the training dataset, and its performance was evaluated using the testing dataset. The accuracy, precision, recall, and F1-score metrics were used to assess the model's performance.

The SVM model employed a radial basis function (RBF) kernel, which is effective in handling non-linear data. The regularization parameter (C) was set to 1.0, and the kernel coefficient (gamma) was set to 'scale'. Hyperparameters were tuned using a grid search with cross-validation to find the optimal values for C and gamma. Similar to the Random Forest model, the spectral data was split into training and testing datasets. The SVM model was trained on the training data and evaluated on the testing data using accuracy, precision, recall, and F1-score metrics.

The neural network model used in this study is a feedforward neural network with three hidden layers. The architecture includes an input layer with 18 nodes, corresponding to the 18 spectral channels, three hidden layers with 64, 32, and 16 nodes, respectively, and an output layer with 3 nodes representing the classification categories. The ReLU activation function was used for the hidden layers, and the softmax activation function was used for the output layer. The model was trained using the Adam optimizer with a learning rate of 0.001. The training process included 100 epochs, and early stopping was implemented to prevent overfitting. Hyperparameters such as the number of hidden layers, nodes per layer, and learning rate were tuned to optimize performance. The data was divided into training and testing sets as described for the other models. The neural network was trained on the training dataset and evaluated on the testing dataset using the same performance metrics: accuracy, precision, recall, and F1-score.

The combined flow diagram for the machine learning models is shown in Fig. 3. This diagram illustrates the data flow from collection to deployment for all three models (Random Forest, SVM, and Neural Networks).

2.1.4. Data collection

The spectral data was collected from a total of 1285 different water specimens, including clean, contaminated, and UV-treated samples. These specimens were prepared as follows: Commercially available drinking water was used as clean water samples. Contaminated water samples were prepared by adding *E. coli* bacteria to the clean water to achieve a concentration of 10^6 CFU/ml. The contamination process was carefully controlled and monitored. UV-treated water samples were subjected to UV disinfection using a 14-watt low-pressure mercury lamp emitting UV light at 254 nm for a specific duration to ensure effective disinfection. The spectral data was collected using the AS7265x sensor module with readings taken every 150 ms. The data acquisition process was automated using the Arduino IDE software, and the data was transmitted wirelessly to a personal computer for storage and analysis.

2.1.5. Pre-processing

The collected spectral data underwent several pre-processing steps to ensure its suitability for machine learning analysis. Outlier removal was performed to eliminate any data points that deviated significantly from the expected range. Normalization was applied to ensure that all features had the same scale, which is crucial for the performance of machine learning models. The dataset was then split into training (70 %) and testing (30 %) sets. The training set was used to train the machine learning models, while the testing set was used to evaluate their performance.

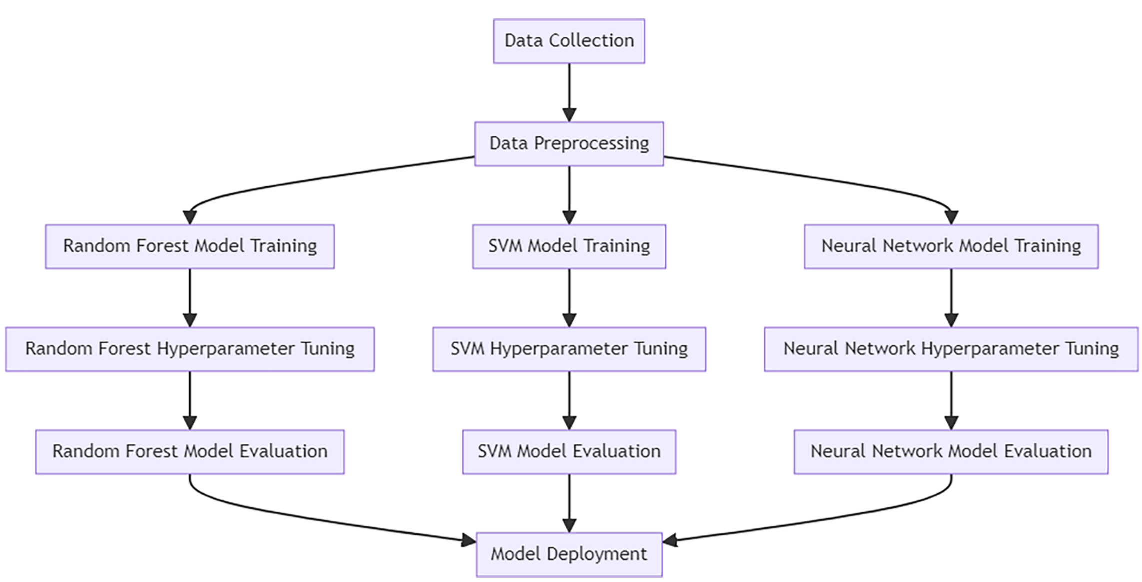

Fig. 3. Combined flow diagram for machine learning models.

2.1.6. UV lamp

The UV lamp is placed in the middle of the reactor in a quartz glass sheath. disinfection; It was carried out with a 14-watt low-pressure mercury lamp (Lighttech) emitting UV light at a wavelength of 254 nm. The UV intensity of the lamp is 40 μW/cm^2. Closed vessel type UV reactors are preferred in drinking water applications. In this type of reactors, contamination by airborne materials is at a minimum level, the user is not exposed to UV rays and has ease of equipment. These reactors, which can have a flow rate of 600 gallons per minute, can provide the necessary UV dose for bacteria and virus inactivation.

2.2. Sample preparation

Commercial drinking water samples were used for this experiment. Drinking water sam- ples containing microorganisms were also taken for this experimental analysis. The samples were taken into the sample tube and measured. In the first stage, a clean drinking water sample weighing about 100 ml was taken in a silver bowl, placed on top of the waterfilled bowl with the distance between the sensor system and the water 1 in. to obtain better observations. Then the relevant spectral readings for the clean drinking water sample are observed. Then, it was prepared by adding the dirty water sample into the clean drinking water sample. And spectral measurements were made. Then, the contaminated water sample was placed in the environment where the closed vessel type UV reactor was located. And the microorganisms in the dirty water were disinfected with UV light. The disinfected water sample was again spectral measurements.

2.3. Spectral data collection

In this segment, we delve into the methodology employed to amass spectral readings from both drinkable and contaminated water specimens. The initial step involves the preparation of the samples. Once the experimental apparatus is fully set up and connected, the spectral readings are captured using the sensor. This is achieved by running the data collection software available in the Arduino IDE library. Data transmission is facilitated by the I2C bus, which sends the information to the ESP8266 microcontroller board. Subsequently, this board forwards the spectral readings to a desktop computer using a wireless connectivity module. In total, spectral information was collected from 1285 different water specimens. The multispectral sensor apparatus operates with a time interval of 150 ms between sam- plings, translating to a frequency of 6 readings every second for all the spectral data of drinkable and tainted water samples.

2.4. Bacterial growth conditions

Escherichia coli ATCC (American Type Culture Collection) 25,922 bacteria was used. The bacteria were grown aerobically at 120 rpm in a 37 ℃ environment inside a 100 ml nutrient broth for one night. Afterwards, the bacteria grown were centrifuged 3939×xg for 10 min at a temperature of 20 ℃ and left to wait in 100 ml phosphate buffered saline solution. Finally, the initial population of 10^3 CFU/ml was diluted for use in the experiment. Afterwards, 24 samples were prepared by diluting as 10^6 CFU/ml. The same dilution coefficients were used for preparing the control groups for each experiment which were not exposed to light (Durgun and Karaman,2019).

3. Result

3.1. Spectral analysis of water samples: implications for quality and safety

This study presents a spectral analysis of different water samples, comparing the characteristics of clean water (Case 0), *E. coli* contaminated water (Case 1), and UV-disinfected water (Case 2). The findings highlight significant indicators of water quality and safety, reflected through changes in optical properties of water.

In the clean water scenario (Case 0), the spectrum typically displayed low intensity values across all wavelengths. This is consistent with the transparency and minimal contamination of the water, showing no significant absorption or reflection characteristics. In contrast, the *E. coli* contaminated water sample (Case 1) displayed a notable increase in intensity values at certain wavelengths. This increase can be interpreted as *E. coli* absorbing or reflecting more light at these wavelengths, indicating a clear alteration in the optical properties of the water due to contamination. Particularly noteworthy are the spectral characteristics observed in the UV-disinfected water sample (Case 2). Post UV disinfection, the spectral profile reverted to a pattern similar to that of clean water (Case 0). This finding suggests that UV light effectively eradicated *E. coli* bacteria, restoring the water's optical properties to a state akin to clean water. This result serves as a strong testament to the efficacy of the UV disinfection process (Fig. 4).

The implications of these findings suggest that spectral analysis can be a potent tool in assessing the quality and safety of water. The use of such analyses for evaluating contamination in water samples and the efficacy of disinfection processes can make significant contributions to the management and preservation of water resources.

3.2. Detailed spectral analysis: distinguishing water quality at key wavelengths

This study delves into the spectral characteristics of water samples at specific wavelengths, notably at 610 nm and 860 nm, to discern the optical variations corresponding to different states of water quality: clean water (Case 0), *E. coli*-contaminated water (Case 1), and UV-disinfected water (Case 2). The findings are illustrated in Fig. 5, which presents the normalized intensity values at these wavelengths for each case.

Analysis at 610 nm Wavelength: The graph depicting normalized intensity at 610 nm demonstrates a significant difference between Case 0 and Case 2, underscoring the effectiveness of UV disinfection in reverting the optical properties of contaminated water to a state resembling that of clean water. Additionally, an increase in intensity is observed for Case 1, indicative of the alteration in optical properties due to *E. coli* contamination. This wavelength proves to be particularly insightful for observing the changes brought about by contamination and subsequent disinfection processes.

Analysis at 860 nm Wavelength: At 860 nm, the disparity in intensity between Case 0 and Case 1 is less pronounced yet discernible, suggesting that *E. coli* contamination exerts a noticeable effect at this wavelength as well. The results at 860 nm further reinforce the findings at 610 nm, providing a comprehensive understanding of the impact of bacterial contamination and UV treatment on water's optical characteristics. These observations demonstrate the utility of spectral analysis in evaluating water quality and safety. By focusing on specific wavelengths where significant differences are observed, it is possible to effectively identify and quantify changes in water properties due to contamination and treatment processes. This approach offers a powerful tool for the assessment and management of water resources, ensuring their safety and suitability for various uses.

3.3. Comparative analysis of machine learning models in real-time spectral water quality assessment

This study embarked on a comparative analysis of three prominent machine learning models—Random Forest, Support Vector Machines (SVM), and Neural Networks—for the classification of water samples into three distinct categories: clean water, contaminated water, and UV-treated contaminated water. The classification was based on spectral measurements, with the primary objective of identifying the most effective model for real-time water quality assessment through spectral

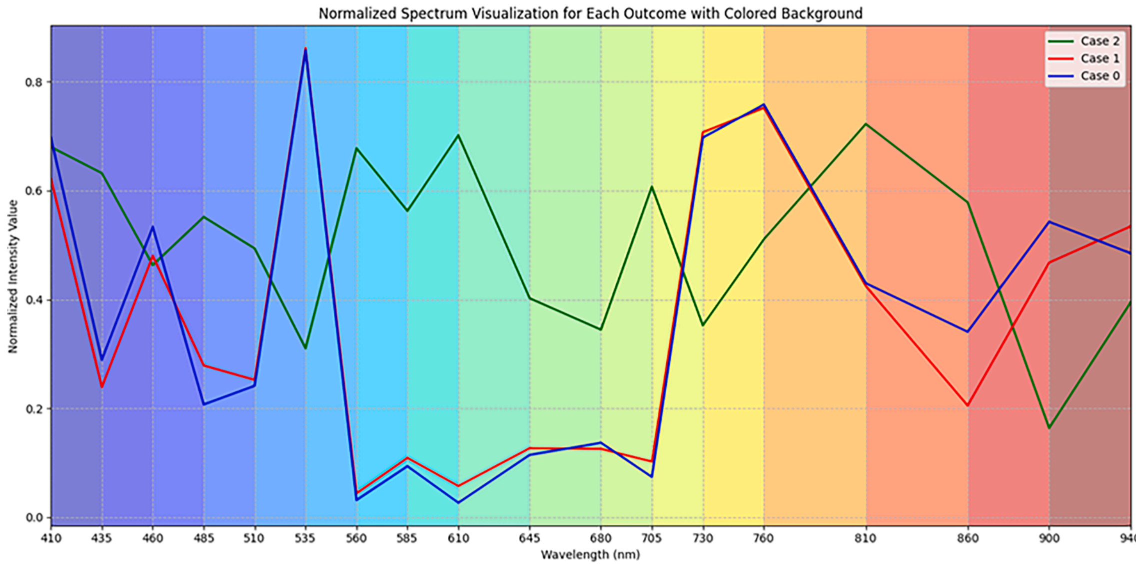

Fig. 4. Spectral analysis of water quality.

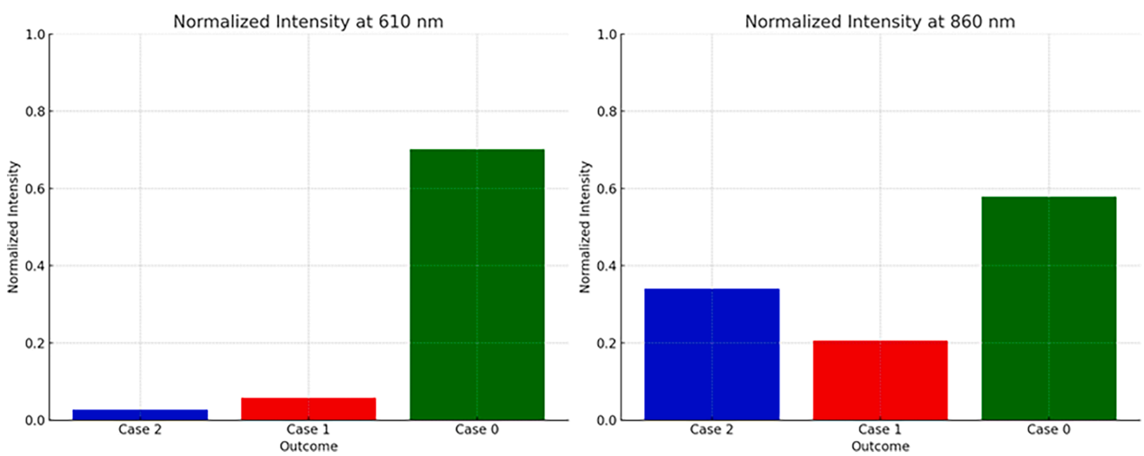

Fig. 5. Normalized intensity at 610 nm and 860 nm for different water samples.

data analysis.

Random Forest Model: This model exhibited a remarkable accuracy of 100 %, showcasing its exceptional capability in handling the spectral data for water quality classification. Its robustness is underscored by perfect scores in precision, recall, and F1-score metrics. These results indicate the Random Forest model's reliability and potential as an essential tool for accurate water quality assessment.

Support Vector Machines (SVM) and Neural Networks Models: Both these models also achieved an impressive accuracy of 100 %, closely paralleling the performance of the Random Forest model. Despite minor variances in individual classifications, these models were highly effective in interpreting spectral data, evidenced by their near-perfect precision and recall rates.

The outcomes of this comparative analysis underscore the potential of utilizing advanced machine learning techniques for real-time water quality monitoring. Each model tested demonstrated high accuracy and reliability, with the Random Forest model slightly leading due to its flawless classification capability. These findings are pivotal for the development of efficient, real-time water quality monitoring systems, which can be integrated into edge computing devices for on-site assessments. Future research endeavors will aim to further optimize these models for enhanced resource efficiency and processing speed, ensuring their suitability for deployment in practical water quality monitoring scenarios.

3.4. *Microbial analysis of water samples: E. coli as a contaminant indicator*

In this study, Escherichia coli ATCC 25922 strain served as the biological contaminant to assess water quality across different cases: Case 0 (clean water), Case 1 (contaminated water), and Case 2 (UV-treated contaminated water). The bacteria were cultured aerobically at 120 rpm in a 37 °C environment in a 100 ml nutrient broth overnight. Post cultivation, the bacteria were centrifuged at 3939×g for 10 min at 20 °C and resuspended in 100 ml phosphate buffered saline solution. The initial bacterial population was set to 10^3 CFU/ml and diluted to 10^6 CFU/ml for the experiment. Control groups for each case were prepared using the same dilution coefficients and were not exposed to light.

After 24 h, the microbial growth in Petri dishes was visually documented, providing a graphical representation of the bacterial proliferation for each case. The graphical outcomes, displayed in Fig. 6, showcase the clarity of clean water with no bacterial growth (Case 0), the high density of bacterial colonies in contaminated water (Case 1), and the substantial reduction of bacterial presence in UV-treated water (Case 2). These results visually corroborate the spectral analysis findings, where Case 1′s significant increase in intensity at certain

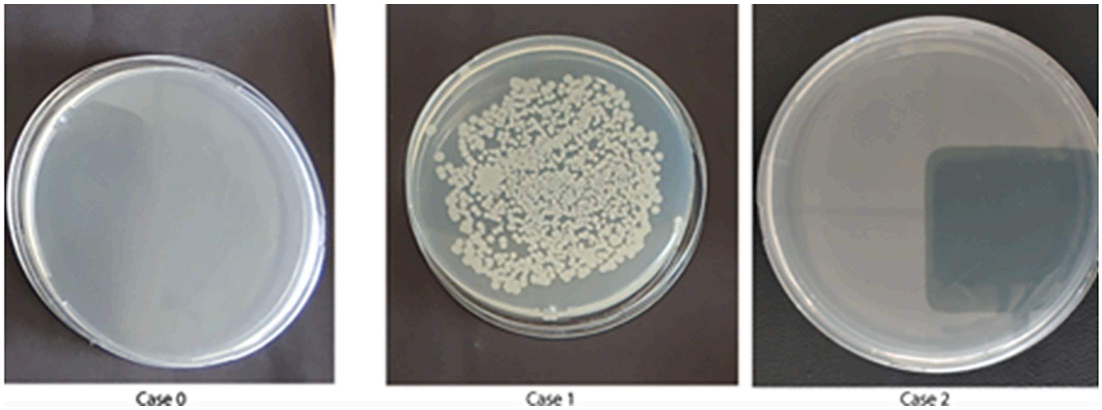

Fig. 6. Microbial cultures of water samples illustrating *E. coli* growth in different cases after 24 hours.

wavelengths indicates the presence of contamination. Similarly, the reduced bacterial growth in Case 2 aligns with the spectral data suggesting effective UV disinfection. The visual evidence from the Petri dishes provides a compelling complementary tool to spectral analysis for water quality assessment.

The study's microbial component enhances the understanding of water contamination and the efficacy of UV treatment, serving as a crucial step towards developing comprehensive water quality monitoring systems. In addition to spectral data, other metrics such as total dissolved solids (TDS), pH levels, and microbial contamination were also used to evaluate water quality. These metrics provide a comprehensive view of water quality changes. For instance, TDS and pH levels are critical for assessing the chemical composition and contamination levels of water. The integration of these metrics enhances the robustness of the water quality assessment.

4. Discussion

The study's innovative integration of spectral analysis and machine learning for water quality assessment is commendable and aligns with recent research trends. For instance, Peterson et al. (2019) demonstrated the effectiveness of machine learning regression techniques in accurately modeling relationships between spectral reflectance and water-quality parameters, such as chlorophyll and turbidity, using remote sensing data (Muncan et al., 2020). Similarly, Niroumand-Jadidi and Bovolo (2022) successfully applied extreme gradient boosting (XGBoost), a machine learning model, for water quality parameter retrieval in optically-complex waters, highlighting the adaptability and precision of such techniques (Niroumand-Jadidi et al., 2022).

The use of spectral data and machine learning in water quality assessment is a growing field of research. The work of Peterson et al. (2019) and Niroumand-Jadidi and Bovolo (2022) complements the findings of this study, demonstrating the robustness and effectiveness of these techniques in diverse water environments. These studies collectively emphasize the potential of machine learning and spectral analysis in enhancing water quality monitoring and management.

While the integration of spectral analysis and machine learning in water quality assessment is a significant advancement, challenges remain in applying these techniques to varied water types and environmental conditions. Egbueri's (2021) study, for example, explored the application of machine learning techniques in assessing water quality across different water types, revealing the need for broader application and validation of these methods (Ovidiu et al., 2021).

The integration of AI and ML technologies has demonstrated effectiveness in real-time water quality assessment within this study. AI facilitates the preprocessing and management of spectral data, while ML models such as Random Forest, SVM, and Neural Networks classify the water samples with high accuracy. This combination enhances the reliability and precision of water quality monitoring, making it a valuable tool for environmental management.

The results underscore the importance of utilizing advanced technologies in ensuring safe water supplies. Future research should focus on validating these methods in different environmental settings and water types. Additionally, long-term studies are needed to evaluate the reliability and consistency of these techniques in real-world applications. Investigating the integration of these methods into existing water monitoring frameworks could also provide valuable insights for enhancing water quality management.

In conclusion, this study's application of spectral analysis and machine learning in water quality assessment is a promising advancement. However, further research is required to expand its applicability and to fully understand its potential in varied environmental conditions and water types.

5. Conclusion

This research marks a significant stride in water quality assessment, showcasing the efficacy of an AI-enhanced multispectral sensor system in detecting and analyzing water pollution. The study confirms the system's ability to differentiate between clean, contaminated, and UV-disinfected water with high precision, illustrating the promise of spectral analysis in public health and environmental monitoring. The successful application of machine learning models in real-time water quality assessment further emphasizes the potential of integrating technology with environmental science.

However, the study also notes the need for broader validation across various water types and environmental conditions. Future research should extend these methods' applicability and explore their integration into existing water monitoring frameworks, ensuring a wider impact on global water quality management. This continued research will be crucial in refining these technologies to achieve more reliable, comprehensive, and practical solutions for water quality monitoring and management across diverse environments.

Funding

The author received no funds, grants, or other financial support for this research project.

CRediT authorship contribution statement

Yeliz Durgun: Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The author would like to thank the experts who took their valuable time to provide feedback for this study.

References

Ajakwe, S.O., Ihekoronye, V.U., Ajakwe, I.U., Jun, T., Kim, D.-S., Lee, J.M., 2022. Connected Intelligence for Smart Water Quality Monitoring System in IIoT. In: 2022 13th International Conference on Information and Communication Technology Convergence (ICTC), pp. 2386–2391. https://doi.org/10.1109/ICTC55196.2022.9952785.

Alshami, A., Ali, E., Elsayed, M., Eltoukhy, A.E., Zayed, T., 2024. IoT innovations in sustainable water and wastewater management and water quality Monitoring: A comprehensive review of advancements, implications, and future directions. IEEE Access. https://doi.org/10.1109/ACCESS.2024.3392573.

Boretti, A., Rosa, L., 2019. Reassessing the projections of the world water development report. NPJ Clean Water 2 (1), 15. https://doi.org/10.1038/s41545-019-0039-9.

Bria, A., Cerro, G., Ferdinandi, M., Marrocco, C., Molinara, M., 2020. An IoT-ready solution for automated recognition of water contaminants. Pattern Recogn. Lett. 135, 188–195. https://doi.org/10.1016/j.patrec.2020.04.019.

Dolcel, H., Durgun, M., Gökrem, L., 2021. IoT tabanli ve makine öğrenmesine dayali seçici sulama sistemi. Avrupa Bilim Ve Teknoloji Dergisi 28, 395–401. https://doi.org/10.31590/ejosat.1002401.

du Plessis, A., du Plessis, A., 2019. Primary water quality challenges, contaminants and the world's dirtiest places. Water as an Inescapable Risk: Current Global Water Availability, Quality and Risks with a Specific Focus on South Africa, pp. 79–114. https://doi.org/10.1007/978-3-030-03186-2_5.

Durgun, Y., 2024. Classification of starch adulteration in milk using spectroscopic data and machine learning. Int. J. Eng. Res. Dev. 16 (1), 221–226. https://doi.org/10.29137/umagd.1379171.

Durgun, Y., Karaman, İ., 2019. Lighting unit providing disinfection for microorganisms via variable frequency and duty cycle. Fresenius Environ. Bull. 28 (11), 8094–8103.

Egbueri, J.C., 2021. Prediction modeling of potentially toxic elements' hydrogeopollution using an integrated Q-mode HCs and ANNs machine learning approach in SE Nigeria. Environ. Sci. Pollut. Res. 28 (30), 40938–40956. https://doi.org/10.1007/s11356-021-13678-z.

Gürsoy, Ö., Atun, R., 2019. Investigating surface water pollution by integrated remotely sensed and field spectral measurement data: A case study. Polish J. Environ. Stud. 28 (4) https://doi.org/10.15244/pjoes/90598.

Mancosu, N., Snyder, R.L., Kyriakakis, G., Spano, D., 2015. Water scarcity and future challenges for food production. Water 7 (3), 975–992. https://doi.org/10.3390/w7030975.

Menon, G.S., Ramesh, M.V., Divya, P., 2017. A low cost wireless sensor network for water quality monitoring in natural water bodies. In: 2017 IEEE Global Humanitarian Technology Conference (GHTC), pp. 1–8. https://doi.org/10.1109/GHTC.2017.8239341.

Muncan, J., Matovic, V., Nikolic, S., Askovic, J., Tsenkova, R., 2020. Aquaphotomics approach for monitoring different steps of purification process in water treatment systems. Talanta 206, 120253. https://doi.org/10.1016/j.talanta.2019.120253.

Niroumand-Jadidi, M., Bovolo, F., 2022. Extreme gradient boosting machine learning for total suspended matter (TSM) retrieval from Sentinel-2 imagery, in: Remote Sensing of the Ocean, Sea Ice, Coastal Waters, and Large Water Regions 2022, Vol. 12263, pp. 30–37. https://doi.org/10.1117/12.2638465.

Ovidiu, C.O., Camen, D., Poşta, G., 2021. General considerations regarding the use of UV-Vis-NIR multispectral sensors to monitor plant health in field conditions.

Peterson, K.T., Sagan, V., Sidike, P., Hasenmueller, E.A., Sloan, J.J., Knouft, J.H., 2019. Machine learning-based ensemble prediction of water-quality variables using feature-level and decision-level fusion with proximal remote sensing. Photogrammetr. Eng. Remote Sens. 85 (4), 269–280. https://doi.org/10.14358/PERS.85.4.269.

Pooja, D., Kumar, P., Singh, P., Patil, S., 2020. Sensors in water pollutants monitoring: role of material. Springer. https://doi.org/10.1007/978-981-15-0671-0.

Sahu, S., Roy, R., Anand, R., 2022. Harnessing the potential of biological recognition elements for water pollution monitoring. ACS Sensors 7 (3), 704–715. https://doi.org/10.1021/acssensors.1c02579.

Sanderson, H., Brown, R.S., Hania, P., McAllister, T.A., Majury, A., Liss, S.N., 2019. Antimicrobial resistant genes and organisms as environmental contaminants of emerging concern: addressing global public health risks. https://doi.org/10.1016/B978-0-12-813290-6.00007-X.

Suganthi, M., Ramesh, N., Sivakumar, C.T., Vidhya, K., 2019. Physiochemical analysis of ground water used for domestic needs in the area of perundurai in erode district. Int. Res. J. Multidiscip. Technovat. https://doi.org/10.34256/irjmtcon89.

Whaiduzzaman, M., Barros, A., Chanda, M., Barman, S., Sultana, T., Rahman, M.S., Roy, S., Fidge, C., 2022. A review of emerging technologies for IoT-based smart cities. Sensors 22 (23), 9271. https://doi.org/10.3390/s22239271.

World Health Organization, 2019. Safer water, better health. https://doi.org/10.2471/BLT.19.234906.

Wu, J., 2020. Challenges for safe and healthy drinking water in China. Curr. Environ. Health Reports 7, 292–302. https://doi.org/10.1007/s40572-020-00274-5.