E CONF SERIES

International Conference on Educational Discoveries and Humanities
Hosted online from Moscow, Russia
Website: econfseries.com 16th September, 2025

DEVICE AND SOFTWARE FOR ACCURATE MONITORING OF GROUNDWATER PARAMETERS USING KNN ALGORITHM

Jamolov Khudoyorkhon Muzaffar uglu,
Djumanov Jamoljon Khudoyqulovich,
Rakhmonov Shahzod Maruf uglu.
E-mail: xudoyorjamolov@gmail.com

Abstract

In this article, a novel device and software solution were developed for high-precision measurement and analysis of groundwater hydrogeochemical parameters, including pH, electrical conductivity, total dissolved solids (TDS), temperature, and other indicators. The proposed solution integrates the K-Nearest Neighbors (KNN) machine learning algorithm for data processing and enables real-time monitoring. The device operates using modern sensors, microcontrollers, and radio communication modules, while the software performs intelligent analysis of the collected data, providing visualization and predictive capabilities. This approach ensures high accuracy in measuring hydrogeochemical parameters and serves as an effective solution for sustainable management of groundwater resources. Groundwater quality can deteriorate due to pollution, making sustainable management and continuous monitoring essential. In this study, devices were developed and tested in wells to measure key parameters, and the collected samples were analyzed using machine learning algorithms.

Introduction

In the next step - we send (get) the data received from the receiver part of our radio module to our web application using the A9G GSM module. A9G is a multifunctional wireless communication module that includes GSM/GPRS and GPS functions, developed by Ai-Thinker. With this module, you can make calls, SMS messages, connect to the Internet, determine location via GPS and many other wireless communication operations. The A9G module is widely used in IoT (Internet of Things) projects, automated devices, GPS tracking systems and remote control systems. To power our device with an alternative energy source, we use a 300 V solar panel kit, which is a high-performance waterproof charging for the weather, strong output power, and a battery led 3.5V-5V. I used an Arduino Uno microcontroller as the central control part of the device. Arduino Uno is an open-source, easy-to-use microcontroller platform used in electronic devices, IoT, automation, and various other projects. It is based on the Atmel ATmega328P chip and is equipped with digital and analog input/output pins, a USB interface, power ports, and other functions. Arduino Uno is one of the most popular entry-level platforms for students, engineers, and programmers. In the Power Management section, we used the LM2596 regulator. Since our device integrates two wireless communication modules and these communication devices consume more than 2 amps of current when active, I paid special attention to choosing the right power source and used rechargeable lithium batteries. However, since the voltage coming from these lithium batteries is high, I used an LM2596 regulator to generate the necessary voltage for my Arduino microcontroller. The automated measuring device consists of a multi-channel measuring system for classifying and assessing the hydrogeochemical parameters of groundwater, a transmitting-receiving sensor, a GSM module for transmitting and receiving data, and a microprocessor for processing commands. In the process of creating its technological structure, an electronic schematic diagram was developed. The block diagram of the multi-channel measuring system for classifying and assessing the hydrogeochemical parameters of groundwater consists of the main module at the wellhead; a floating part located inside the well; a computer of the central control and monitoring system.

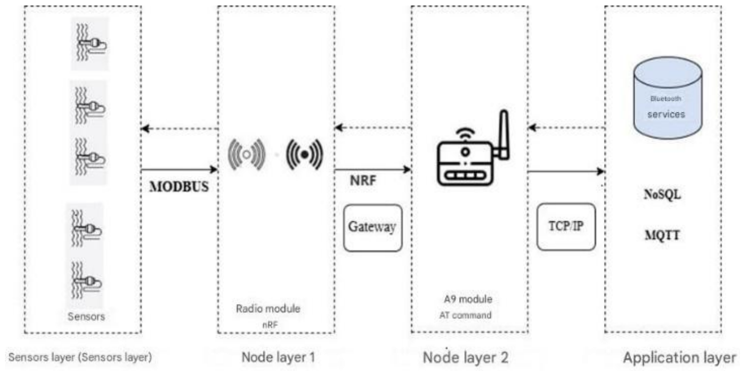

Fig.1. Architecture of an information and analytical system for groundwater monitoring.

Water is the key element of life, essential for the survival of all living organisms. For human health, at least 2 liters of clean water per day are recommended . Water quality affects not only people's well-being but also agriculture, industry, and the environment, making its regular monitoring and compliance with standards crucial. Conventional laboratory analyses are accurate but slow and resource-demanding, while recent advances in machine learning and real-time monitoring provide faster and more automated solutions. One promising method is the K-Nearest Neighbors (KNN) algorithm, which classifies new data by comparing it with existing datasets. Integrating KNN into IoT-based monitoring systems enables continuous and accurate classification of water quality, supporting timely decision-making and preventive measures .

This study explores the application of KNN for water quality classification using parameters such as pH, temperature, and TDS. The approach offers high efficiency, scalability, and cost-effectiveness, contributing to sustainable water resource management . Groundwater properties are defined through hydrological indicators (water level, slope, volume) and hydrochemical factors (salinity, salt content, organic, biogenic, and pollutant concentrations). Hydrogeological monitoring is therefore vital for managing groundwater quality and availability. Modern automated devices and software significantly enhance the accuracy and reliability of hydrogeochemical measurements, enabling continuous real-time data collection and advanced analysis of parameters such as pH, electrical conductivity, dissolved oxygen, and ion concentrations.

When collecting the dataset, we took it from our web application https://qudugsuvi.uz/home/1/.

Our web application is built on the Django framework, a high-level web framework for the Python programming language. Our web-based application enables the continuous observation and monitoring of well indicators throughout the measurement period. This functionality facilitates the collection of precise data accumulated over extended periods, allowing for comprehensive analysis and comparison of trends across different years. This increases the possibilities of groundwater monitoring.

Table 1. The Dataset Attributes

Data	Location	pH	TDS (mg/L)	Temperature (C°)	Water Level (cm)
2025- 01-01	Site A	7.2	450	11.1	2.8
2025 01-02	Site B	7.1	460	10.5	2.9
2025 01-01	Site B	6.9	400	10	3.2
2025 01-02	Site B	7.0	405	10	3.3

K-Nearest Neighbor (K-NN) The K-Nearest Neighbor (KNN) algorithm is a data classification based on the closest distance to an object or what is usually called a neighbor in training data and testing data . The following is the formula for K-Nearest Neighbor

$$j\%_{=1}(xi^{-}yi)^2 \qquad (1)$$

Where :
K = determines the data attribute
Xi = determines the training data
Yi = determines testing data or test data
K-Nearest Neighbor has calculation steps, namely:
Determine the set/parameter k = the closest number of data
Calculate the amount of new data with training data
Sort the closest data based on the minimum distance of the K value
Check the class data from the closest data
Determine the results from the closest data as the predicted value of the new data.
Calculating the distance between new data and old data can be calculated using several methods, one of which uses the Euclidean distance formula .
The formula for calculating Euclidean distance can be seen as follows:

$$(pH, - pH) + (TDS_t - TDSj)^2 + (Tempt - Tempj\ )^2 + (Leveli - Levelj)^2 \qquad (2)$$

pH_t, TDS_t, Тетр ^ Level $_t$ j- are the values for the data point
pHj, TDSj, Tempj, Levelj j- are the values for the data point d(i,j) i and j distance between points.
K- Nearest Neighbor Algorithm. Preparations before calculating are as below:
Cleaning data. Data cleaning is a step that needs to be carried out to deal with missing values and inconsistent data.
Data Selection. Data Selection is the stage ofreducing the amount of data used in the mining stage.

- still presenting the original data. In this data selection step, only data with important characteristics that are
- usually used in good water quality prediction classes are selected. In this study, all available attributes
- influenced the results. There are 21 attributes used including class/tag.
- Data Transformation. Data transformation is carried out to correct the shape and format of the data.

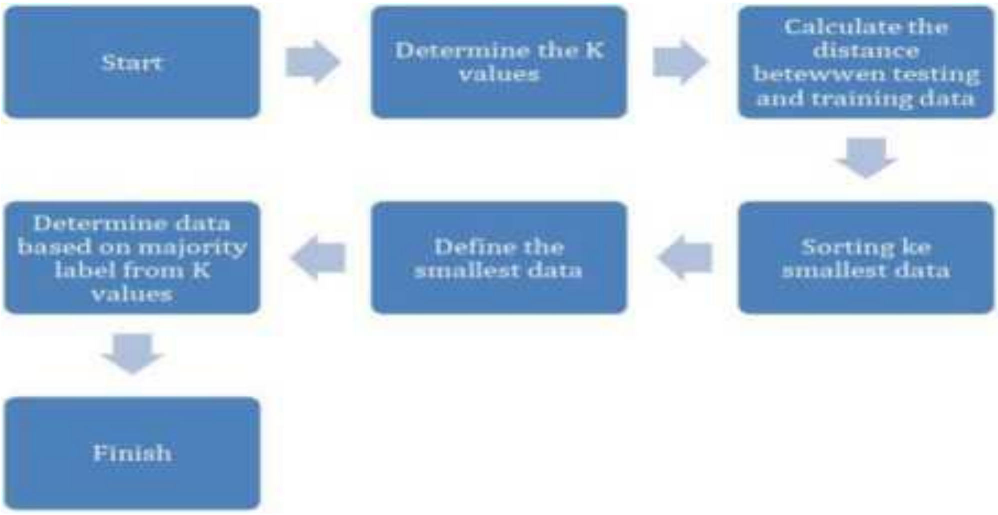

Fig.2. Classification of water quality based on KNN

The model evaluation phase determines the performance of the algorithm.
The integration of IoT sensors has enabled real-time data collection and analysis. The system transmitted data every 5 seconds, with an average processing and response time of 2.5 seconds. This performance meets the requirements of realtime

monitoring under dynamic water quality conditions. This approach enables more precise real-time prediction of groundwater quality parameters, including pH, TDS, temperature, and water level, through the classification capabilities of the KNN algorithm in water quality analysis.

Results

In the Python programming environment, we can import the libraries we need. Python libraries are used to import the necessary functions and tools into our program. Below are the libraries we use:

train_test_split: Divides the dataset into training and testing subsets. test_size=0.3 means 30% of the data is used for testing. random_state=42 ensures reproducibility of the split., KNeighborsClassifier: The KNN algorithm with n_neighbors defining how many neighbors are considered, accuracy_score: Calculates the accuracy of the predictions, classification_report: Provides precision, recall, and F1-score for each class, giving deeper insights into model performance.

Python libraries make it easy to perform specific tasks in our program. These libraries enable the implementation of a comprehensive pipeline for water quality detection, from sensor data acquisition to machine learning classification and visualization.

The Table 2 dataset for KNN classification in water quality analysis is structured as a table. Each row represents a water sample, while the columns include various features such as pH, temperature, dissolved solids, and conductivity.

Table 2. The Dataset KNN classification table

ID	pH	TDS (ppm)	Temperature(°C)	Water level(s m)	Quality label
1	7.2	150	25.5	2.5	Good
2	6.8	450	26.0	3.0	Moderate
3	5.5	701	28.0	1.8	Poor
4	6.9	155	24.0	2.2	Good
5	7.0	400	25.0	2.7	Good
6	6.5	700	28.5	3.5	Poor
7	7.1	520	25.8	3.0	Moderate
8	7.3	450	26.5	2.9	Good
9	6.7	620	27.5	3.2	Poor
10	7.4	390	25.2	2.8	Good

Conclusion

The KNN algorithm has proven to be a valuable tool in water quality monitoring and prediction. Its applications in classifying water samples based on parameters like pH, TDS, temperature, and water level demonstrate its potential in realtime decision-making systems, remote sensing, and environmental management. Despite its simplicity, KNN is effective in dealing with water quality data, especially when combined with appropriate preprocessing techniques and domain-specific knowledge. This paper focused on improving the scalability of KNN, processing large amounts of data from IoT sensors, and integrating it with more complex machine learning models to improve prediction accuracy and operational efficiency, and obtained effective results.

References

1. F. Rajabov, J. Djumanov, Kh. Jamolov. Kh. Urolboy, " Development of a TDS Measurement System Based on Frequency Impedance Spectroscopy for Water Composition Analysis Integrated with a Well Water Level Meter," 2025 IEEE 26th International Conference of Young Professionals in Electron Devices and Materials (EDM). Urgench DOI: 10.1109/EDM65517.2025.11096797, pp. 2400–2405, 08 August 2025.
2. R. Yakhshibaev, B. Turaev, Kh. Jamolov, N. Atadjanova, E. Kim and N. Sayfullaeva , "Development of a mathematical model for balancing the level and device for remote monitoring of groundwater parameters" 2021 International Conference on Information Science and Communications Technologies (ICISCT) DOI: 10.1109/ICISCT52966.2021.9670022, 17 January 2022.
3. F. Rajabov, J. Djumanov, K . Abdurashidova and Kh. Jamolov, "Creation of a Multi-Functional Device for Real-Time Monitoring of Water Quality Springer Proceedings in Mathematics & Statistics (ISMI 2024) 30 May 2025. pp 183–194.
4. F. Rajabov, J. Djumanov, Kh. Jamolov, "Creation of a multi-functional device for real-time monitoring in evaluation of the basic characteristics of water" DTAI - 2024 I international scientific - practical conference Tashkent 2024 year, September 11, 16-17p.
5. Djumanov.J, Anorboyev. E, Jamolov. Kh, "Analysis of the hydrogeological

conditions of the object in order to conduct automated monitoring of measurements of groundwater parameters” International Journal of Geology, Earth & Environmental Sciences ISSN: 2277-2081 An Open Access, Online International Journal Available at http://www.cibtech.org/jgee.htm, 2023 Vol. 13, pp. 161-167 (04.00.00; № 7)

6. Kh. Jamolov, J .Djumanov, F. Rajabov.“Water quality classification using KNN algorithm and real-time monitoring” Bulletin of TUIT: Management and Communication Technologies Journal https:// uzjurnal/tuitmct ISSN:2181-1083 DOI:0 № 3s (41)2024 (05.00.00. №31)

7. Kh.Jamolov, B.Turayev, R.Yakhshiboev, “Remote control of water distribution processes using modern technologies” Muhammad al- Khorezmiy avlodlari, 3(17) 2021, September 13, 2021. 100-104, (05.00.00; No. 10).

8. Rajabov, F.: Non-contact ultrasonic direct radiation metod for measuring the water level in wells and boreholes. ICISCT-2023. 28th–30th of September 2023, TUIT, Tashkent.