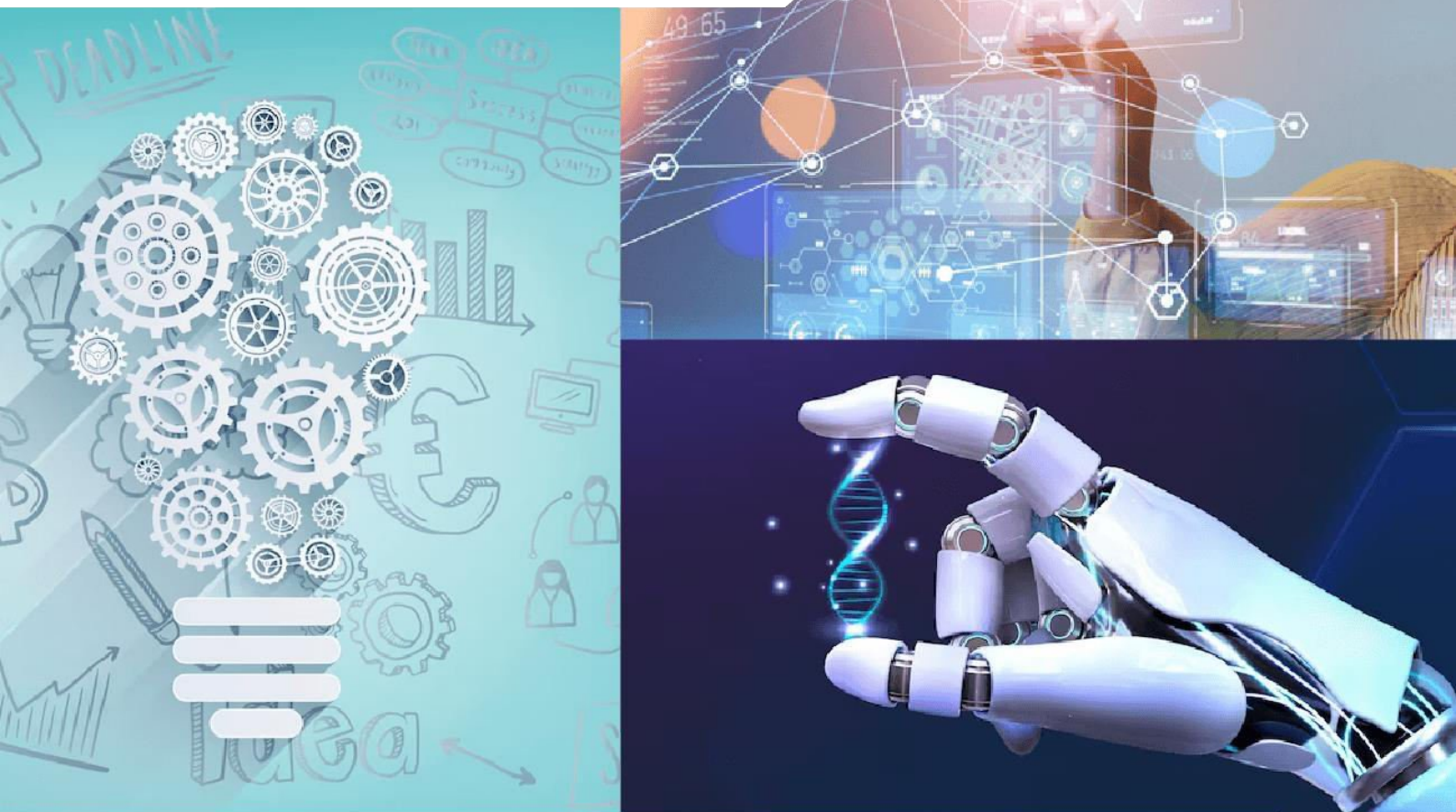




International Journal of Advanced Research in Education and Technology (IJARETY)

Volume 13, Issue 2, March - April 2026

Impact Factor: 8.152



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



Machine Learning-Based Groundwater Level Prediction Using Environmental Parameters

Jyothi

Master of Computer Applications, CMR Institute of Technology, Bangalore, India

ABSTRACT: Groundwater is a crucial source of freshwater that sustains agricultural, domestic, and industrial activities. Accurate forecasting of groundwater levels facilitates better water resource management and helps identify potential depletion risks at an early stage. This study is intended to predict groundwater level variations using environmental parameters such as rainfall, temperature, and humidity. There has been a lot of interest in machine learning techniques in recent years in environmental data analysis. They help explore groundwater distributions using atmospheric and environmental variables. Most of these methods focus on getting better prediction results. But these techniques remain inadequate for interpreting model behavior and translating findings into practical monitoring applications. This work proposes a machine learning-based model for groundwater level estimation. The system aims to address the existing limitations. The system uses a Random Forest-based regression model trained on a temporal dataset containing 1461 entries. The framework includes data cleaning and dataset visual analysis. It also performs groundwater risk evaluation. SHAP-based explainable AI techniques aid in the explanation of the predictive results. The study results show strong predictive performance, having an R^2 value of 0.972 and an RMSE of 0.139 meters. We think this system can be a useful tool for monitoring groundwater and ensuring responsible water use.

KEYWORDS: Groundwater forecasting, Random Forest, SHAP analysis, temperature and precipitation information, water level monitoring, and depletion risk.

I. INTRODUCTION

One of the planet's most valuable natural resources is groundwater. In many regions of the world, people use it to run industries, drink, and farm. People are forced to rely on subterranean water for daily needs in areas where rivers and lakes are unavailable or dry up during specific seasons of the year. However, groundwater reserves have been under a lot of strain in recent years due to factors like increased farming, fast population growth, and climate change. Water levels have been declining in many places as a result, and the water supply has become erratic. Serious issues like food shortages, environmental harm, and a lack of clean water for humans result from this. Thus, monitoring changes in groundwater levels over time is crucial. This facilitates better planning and ensures that there is always an adequate supply of water for water management authorities. For many years, groundwater has been studied by scientists using computer simulations and models based on physics. These instruments can provide some helpful details regarding subterranean water systems. However, in order for them to function effectively, they need a lot of specific information about the local geology. Acquiring this data is difficult, time-consuming, and expensive. Thus, these techniques are not very common in practice. This issue prompted researchers to consider data-driven strategies. One such method is machine learning. It makes connections between groundwater behavior and environmental conditions by learning from historical data. Ensemble techniques like neural networks, random forests, and support vector machines, found widespread application in hydrology-related research. These methods are especially respected for their capacity to handle nonlinear and high-dimensional patterns inherent in environmental datasets. A considerable body of research has drawn on these approaches to predict groundwater levels.

Wang [3] confirmed that Random Forest-based ensemble models accurately estimate underground water fluctuations from environmental inputs. Kanyama et al. [1] showed that climatic parameters like precipitation and temperature can be applied within machine learning pipelines to forecast aquifer-level variations. Integrated strategies fusing ML along with deep learning. Architectures have further strengthened forecasting accuracy in time-series groundwater studies [5], [10]. Research on drought-affected areas has shown that incorporating temperature, rainfall, and humidity data helps uncover groundwater trends and inform resource planning efforts [6], [9]. Investigations drawing on borehole observation records have similarly demonstrated that ML techniques can meaningfully boost the reliability and trustworthiness of groundwater prediction and resource evaluation [2]. Review studies draw attention to the reality of machine learning models broadly adopted for simulating subsurface water dynamics, owing to their capacity to handle high-dimensional environmental data while delivering improved forecasting outcomes [4], [7], [8].

In this study, we developed a machine learning model that uses temperature, humidity, and rainfall as inputs to forecast groundwater levels. We investigated the effects of these three factors on groundwater depth using Random Forest regression. A dataset of 1461 daily records gathered over four years was used to train the model. It picks up on the pattern of how environmental factors affect groundwater. Additionally, the system evaluates and categorizes the risk of groundwater depletion. To determine which input factors have the biggest effects on the predictions, we employed SHAP. In order to enable users to enter values and receive predictions instantly, the entire system was also implemented as a web application using Streamlit. This paper is arranged as follows. The system's design is covered in Section II. The experiments we conducted and their outcomes are discussed in Section III. Our conclusion and some suggestions for future groundwater prediction and water management research are provided in Section IV.

II. PROPOSED SYSTEM

Using temperature, humidity, and rainfall as inputs and a machine learning model, our system forecasts groundwater levels. These three variables and other weather-related factors have a direct impact on groundwater. We used a structured workflow to study these variables that consists of data preprocessing, feature scaling, model training, prediction, and visualization. The input features for a groundwater level prediction system can be seen in Fig. 1.

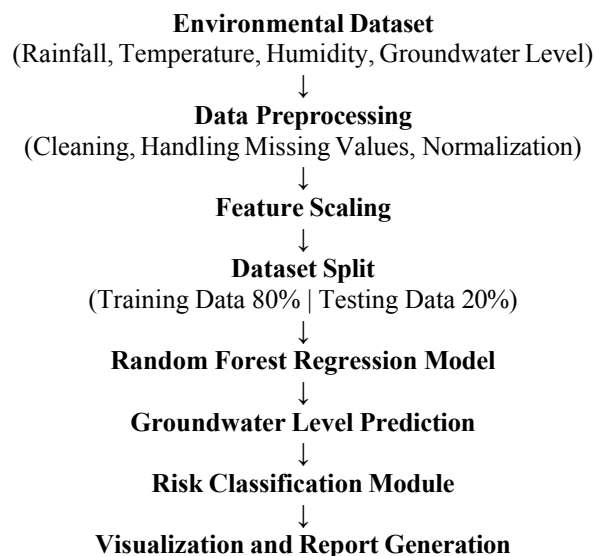


Fig. 1. Architecture of the proposed groundwater level prediction system

Figure 1 outlines the workflow of the proposed groundwater prediction framework. The process starts with an environmental dataset that includes rainfall, temperature, humidity, and groundwater level observations. These variables contribute greatly to groundwater recharge and depletion processes.

Initially, the dataset undergoes preprocessing where missing values are managed, and the data is normalized to maintain consistency. Feature scaling is then performed to standardize the input variables. After preprocessing, the dataset is partitioned into training and evaluation sets using an 80:20 ratio. Historical data are used in training the model. Test data assist with examine their accuracy. An RF regression model analyzes the link between environmental parameters and groundwater level variations. The model then estimates groundwater levels using environmental data inputs. The estimated values are classified to detect possible depletion conditions. Finally, the results are shown through a graph and documentation to assist groundwater monitoring.

Mathematical Analysis

The developed framework employs the Random Forest–based regression approach for predicting groundwater levels. During training, Random Forest creates a collection of decision trees and aggregates their predictions to produce the final prediction. This approach was the one we selected due to its capability to represent nonlinear patterns between environmental variables and groundwater levels.

The method derives a final prediction from many decision trees to improve accuracy and minimize overfitting. It also

performs effectively with environmental datasets where strong associations exist between climatic variables. These capabilities enable the RF model to be suitable for modeling groundwater variations influenced by hydro-meteorological conditions.

The prediction function is expressed as:

Random Forest Prediction Function:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N T_i(x)$$

where $\hat{y}$ represents the predicted groundwater level, N signifies the number of decision trees, $T_i(x)$ represents the prediction yielded by the i^{th} decision tree, and x represents the input feature vector consisting of rainfall, temperature, and humidity.

To evaluate model performance, the RMSE (Root Mean Square Error) metric is used:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Where y_i represents the actual groundwater level, and $\hat{y}_i$ denotes the predicted groundwater level.

Role-Based Access Control (RBAC)

To enhance system security and usability, the framework also integrates RBAC (Role-Based Access Control) through a secure login mechanism using username and password authentication. Different users are granted particular roles that define the functionalities they are permitted to access. This system permits access restricted to authorized individuals, prediction results, data analysis tools, and system management features, thereby preserving both accuracy and safety of groundwater prediction data.

III. RESULTS AND DISCUSSION

This section describes the implementation process and assesses the efficiency of the developed groundwater level prediction system. The system was implemented by applying machine learning techniques to analyze environmental parameters and produce groundwater level predictions. The obtained results indicate how effective the Random Forest-based regression model is in determining the connection between environmental variables and groundwater level variations.

A. Implementation Environment

Our system was implemented in Python because it offers reliable support for data analysis and machine learning-based applications. Pandas and NumPy handled data preparation and transformation. The RF model was built using Scikit-learn. Plotting libraries were used to visualize groundwater prediction outcomes and environmental trends. A Streamlit web interface provides interactive visualization. It also allows live predictions. Users can enter environmental parameters to examine groundwater level forecasts.

B. Dataset Description

Our dataset in this work captures key environmental variables known to drive changes in subsurface water depth. It records daily observations of relative humidity, precipitation, air temperature, and groundwater level, covering four years with 1461 entries in total. Precipitation captures rainfall input, while temperature and humidity reflect atmospheric factors that influence both evapotranspiration and aquifer recharge. Before training the model, the data underwent a cleaning process to remove inconsistencies, and the final dataset was verified to be suitable for time-series predictive modeling.

C. Training and Evaluation Process

There were two divisions of the collection, non-overlapping subsets following an 80:20 division to assess their quality, the trained model, which generalizes well to new inputs. Roughly four-fifths of the records were fed into the model of Random Forest regression during training, with the remaining fifth held back solely for performance evaluation. The training portion enabled the model to learn the relationships between environmental parameters and groundwater level variations. The dataset reserved for testing was used to measure how accurately the trained model could estimate groundwater levels for previously unseen environmental conditions.

D. Model Performance and Results

The total performance of the built groundwater forecasting model was assessed using common regression metrics, including R^2 and RMSE. The evaluation demonstrated the trained RF model yielded an R^2 value of 0.972, meaning that approximately 97.2% of the observed variability in groundwater depth is explained by the three environmental inputs used. An RMSE of just 0.139 meters was recorded, reflecting a very narrow gap between the model's outputs and the actual measured values. The integration of a strong R^2 score paired with a low RMSE indicates that the predicted water levels align closely with the actual observations, demonstrating strong predictive capability of the predictive framework.

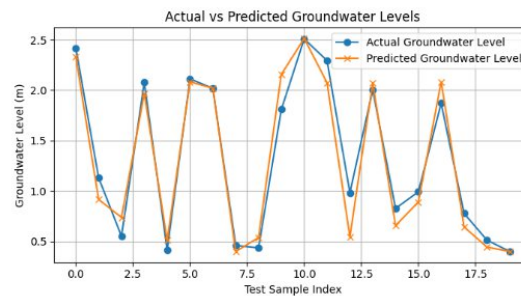


Fig. 2. Comparison of actual and predicted groundwater levels using the Random Forest– based regression model

Figure 2 illustrates the comparison between the actual groundwater levels derived from the dataset and the groundwater levels predicted by the Random Forest– based regression model. The x-axis shows the test sample index, while the y-axis represents groundwater level values. The blue line indicates the actual groundwater levels obtained from the dataset, and the orange line represents the values predicted by the machine learning model. The graph shows that the predicted levels closely follow the trend of the actual groundwater levels across the selected test samples. The small difference between predicted and observed values indicates that the relationship between temperature, humidity, and rainfall was effectively taught to the RF model, and groundwater level variations. This graphical comparison provides additional confirmation. The evaluation metrics confirm the reliability of the groundwater prediction framework. The system also provides visualizations of groundwater trends and prediction results through the Streamlit interface, helping users understand groundwater behavior and analyze prediction outputs more easily. However, some limitations remain. Environmental simulations are used in the study. Because of this, the model might function differently when applied to actual groundwater monitoring datasets. The dataset does not include hydrological and geological factors that affect groundwater levels. Predictive accuracy and reliability may be increased by including more conditions and actual monitoring data.

Table 1: Compares machine learning models' performance.

Model	R^2 Score	RMSE(m)
Linear Regression	0.781	0.452
Decision Tree	0.891	0.284
Random Forest (Proposed)	0.972	0.139

IV. CONCLUSION

A machine learning method for estimating groundwater levels is suggested in this analysis. temperature, humidity, and other inputs. Rainfall is utilized in the model. The relationship between climate conditions is examined using a Random Forest regression model. changes in groundwater levels. The outcomes indicate that the model successfully represents the associations between the variables. The model produces reliable groundwater level forecasts. We tested the model using regression metrics like the coefficient of determination and estimation error. Our model performed well in estimating groundwater-level variations. The platform includes plotting tools and an interactive interface. Users can

view groundwater variations through these tools. Users can also use these to better understand prediction results. This methodology can facilitate groundwater level monitoring. It also helps provide better insight into groundwater variations under different environmental conditions. Future research may use actual groundwater observation datasets from various regions. This can help further verify the developed model. It may also improve sustainable groundwater management.

REFERENCES

- [1] Y. Kanyama, R. Ajoodha, H. Seyler, N. Makondo, and H. Tutu, “**Application of machine learning techniques in forecasting groundwater levels in the Grootfontein aquifer,**” Proc. 2nd Int. Multidisciplinary Information Technology and Engineering Conf. (IMITEC), 2020.
- [2] N. Iqbal, A. N. Khan, A. Rizwan, R. Ahmad, B. W. Kim, K. Kim, and D. H. Kim, “**Groundwater level prediction model using correlation and difference mechanisms based on borehole data for sustainable hydraulic resource management,**” IEEE Access, vol. 9, 2021.
- [3] P. Wang, “**Prediction of groundwater levels based on a random forest regression algorithm,**” Proc. IEEE 4th Int. Conf. on Information Technology, Big Data and Artificial Intelligence (ICIBA), 2024.
- [4] H. Tao, M. M. Hameed, H. A. Marhoon, M. Zounemat-Kermani, S. Heddami, S. Kim, et al., “**Groundwater level prediction using machine learning models: A comprehensive review,**” Neurocomputing, 2022.
- [5] A. Tiwari and H. Himanshu, “**A hybrid machine learning and deep learning model for time-series groundwater level prediction,**” Proc. 2025 Smart Technologies, Communication & Robotics (STCR), 2025.
- [6] E. A. Hussein, C. Thron, M. Ghaziasgar, A. Bagula, and M. Vaccari, “**Groundwater prediction using machine-learning tools,**” Algorithms, vol. 13, no. 11, Art. no. 300, 2020.
- [7] S. Kim, J. Lee, and H. Park, “**Development of machine learning and deep learning models with SHAP strategy for predicting groundwater levels in South Korea,**” Scientific Reports, vol. 15, 2025.
- [8] X. Wang, Y. Zhang, and L. Chen, “**Enhancing groundwater predictions by incorporating response lag effects in machine learning models,**” Journal of Hydroinformatics, vol. 27, no. 2, pp. 338–356, 2025.
- [9] F. Feng, H. Ghorbani, and A.E. Radwan, “**Predicting groundwater level using traditional and deep machine learning algorithms,**” Frontiers in Environmental Science, vol. 12, Art. no. 1291327, 2024.
- [10] C. J. Okolie, T. Smit, and R. Guruprasad, “**Application of machine learning and deep learning for predicting groundwater levels in the West Coast Aquifer System, South Africa,**” Earth Science Informatics, vol. 17, pp. 1–18, 2024.

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975

Impact Factor: 8.152