Journal of
Water Resource Engineering and Management

ISSN: 2349-4336
Volume 12, Issue 1, 2025
January–April
DOI (Journal): 10.37591/JoWREM

https://journals.stmjournals.com/jowrem

Research JoWREM

Harnessing Hydrogeological Parameters: Prediction of Water Probability and Levels for Water Well Construction Using AI-Enabled Models

Pranavi Kannepogu[1,*], T. Jhansi Rani[2], V.S.C. Prasanthi[3], Deepthi Marada[4]

Abstract

The AI-Based Decision Support System for Water Well Construction utilizes data from the National Aquifer Mapping and Management System (NAQUIM) and employs advanced AI techniques like regression analysis, decision trees, and neural networks. This system predicts crucial parameters for water well construction, including location suitability, water-bearing zone depths, and groundwater quality. By integrating large datasets such as lithology, geophysical logs, and aquifer maps provided by the Central Ground Water Board (CGWB), the system enhances the precision of predictions. Its user-friendly graphical interface allows users to select specific locations and obtain detailed insights effortlessly. The AI-driven approach offers a scalable and accurate alternative to traditional manual methods, improving the efficiency of decision-making in water well planning. By providing reliable and actionable insights, the system supports sustainable groundwater management, promoting more effective resource utilization and reducing environmental impact. This innovative solution significantly advances groundwater decision support practices.

Keywords: Water well construction, groundwater management, artificial intelligence (AI), NAQUIM data, decision support system

INTRODUCTION

Access to clean and reliable water sources is crucial for sustaining life and supporting various socioeconomic activities. In many regions, particularly those with limited access to centralized water supply systems, groundwater extraction through wells plays a vital role in meeting water demand. However, the success of water well construction depends heavily on understanding subsurface hydrogeological conditions, which can vary significantly from one location to another. To aid in the decision-making process for water well construction, the Central Ground Water Board (CGWB) in India developed the National Aquifer Mapping and Management System (NAQUIM). NAQUIM provides a wealth of geospatial data and information related to the relevant parameters. While these datasets hold immense potential for guiding well construction activities, their effective utilization requires advanced analytical techniques and decision support systems. In recent years, artificial intelligence (AI) has emerged as a powerful tool for analyzing complex datasets, identifying patterns, and making predictions. By leveraging AI techniques, it is possible to extract actionable insights from NAQUIM data and assist

***Author for Correspondence**
Pranavi Kannepogu
E-mail: ranavikannepogu@gmail.com

[1-3]Student, Department of Information Technology, Gayatri Vidya Parishad College of Engineering for Women, Visakhapatnam, Andhra Pradesh, India
[4]Assistant Professor, Department of Information Technology, Gayatri Vidya Parishad College of Engineering for Women, Visakhapatnam, Andhra Pradesh, India

Received Date: October 28, 2024
Accepted Date: January 16, 2025
Published Date: January 20, 2025

Citation: Pranavi Kannepogu, T. Jhansi Rani, V.S.C. Prasanthi, Deepthi Marada. Harnessing Hydrogeological Parameters: Prediction of Water Probability and Levels for Water Well Construction Using AI-Enabled Models. Journal of Water Resource Engineering and Management. 2025; 12(1): 16–28p.

users in making informed decisions regarding water well construction. However, despite the growing interest in AI-based approaches to groundwater management, there remains a need for tailored solutions that address the specific challenges associated with well construction.

LITERATURE REVIEW

M.A. Gossell presented a comprehensive study on the utilization of advanced techniques to optimize the design of groundwater production wells, addressing a common challenge in groundwater management. It has been shown that production wells may be drilled deeper than necessary, leading to increased installation costs and potential issues with water quality. Furthermore, the study identified instances of poor-quality water within the well, emphasizing the importance of considering water quality variations in well design [1].

Thakur addressed the critical issue of arsenic contamination in groundwater, particularly focusing on Bihar, India. It presents an overview of existing research on arsenic contamination in drinking water, emphasizing studies conducted globally and in Bangladesh, where arsenic contamination is widespread. Overall, this underscores the urgent need for sustainable mitigation by elucidating the depth-related distribution of as in drinking water sources [2].

The study by Liu et al. (2015) presented a novel approach for estimating discharge and water depth in ungauged rivers by integrating hydrologic and hydraulic modeling with satellite data. It addresses the challenge of predicting river dynamics in areas lacking ground-based measurements, which is crucial for managing water resources, especially in light of future freshwater scarcity. This study contributes to advancing our ability to understand and manage river systems, especially in regions where conventional monitoring infrastructure is lacking [3].

Narendra Singh Chandel and Chakraborty addressed crop water stress through an AI-enabled mobile device, showing a fusion of computer vision, deep learning, and edge computing technologies. This highlights the significance of real-time stress detection in agriculture, emphasizing the limitations of traditional methods and the potential of AI-driven solutions. This review discusses the evolution of AI and image processing techniques in crop stress monitoring, underscoring the shift towards cost-effective, portable, and user-friendly devices for on-field applications [4].

Barkman and Davidson (1972) addressed the crucial issue of water quality in waterflood or water disposal projects, emphasizing the potential impact of suspended solids on injection wells. It is proposed that methods to interpret water quality data are obtained through membrane filters or cores and to predict well impairment caused by suspended solids [5].

Balraj Singh, Parveen Sihag, et al. (2021) addressed the critical issue of water quality assessment through the development and comparison of soft computing techniques for predicting the Water Quality Index (WQI) in three sub-watersheds in Iran. This study underscores the significance of employing advanced computational techniques to predict water quality indices, particularly in regions facing water quality degradation, and highlights the potential of ANN as a viable tool for such predictions through artificial neural networks (ANN), generalized regression neural network (GRNN), and adaptive neuro-fuzzy interference systems (ANFIS) [6].

Md. Saikat Islam Khan and Nazrul Islam presented a comprehensive approach to water quality prediction and classification, addressing a critical challenge in environmental management. This provides novel insights by integrating Principal Components Analysis (PCA) with regression algorithms for WQI prediction and employing a classification approach, thereby offering a dynamic and effective solution for water quality assessment. Additionally, a Gradient Boosting Classifier was utilized for water quality status classification, achieving notable accuracy [7].

Journal of Water Resource Engineering and Management
Volume 12, Issue 1
ISSN: 2349-4336

Banerjee et al. (2022) addressed the pressing issue of water contamination prediction using location-based parameters, such as latitude, longitude, and elevation, coupled with machine learning techniques. Highlighting the global challenge of maintaining water quality during industrial growth and pollution [8].

METHODOLOGY

System Architecture and Implementation

The system architecture integrates a machine learning model with a user-friendly web-based interface for predicting suitable water well locations. It comprises a backend powered by AI algorithms and a frontend designed for intuitive user interaction. The implementation ensures seamless data flow, efficient processing, and clear visualization of results.

Development of the Graphical User Interface

Developing a Graphical User Interface (GUI) for a web-based AI-enabled water well predictor can greatly enhance user experience and facilitate interaction with predictive models. Some of the common choices for choosing a programming language and GUI framework that suits the requirements include Python with libraries like Tkinter, PyQt, or wxPython, or web-based frameworks like Flask or Django for building browser-based interfaces (Figure 1) [9].

To design the user interface, the layout, and functionality of the GUI must be sketched out. The key features that users need, such as input forms and data visualization tools, need to be considered for the result to be displayed. The interface is intuitive and user-friendly. Data visualization tools can be integrated to present model outputs clearly and understandably.

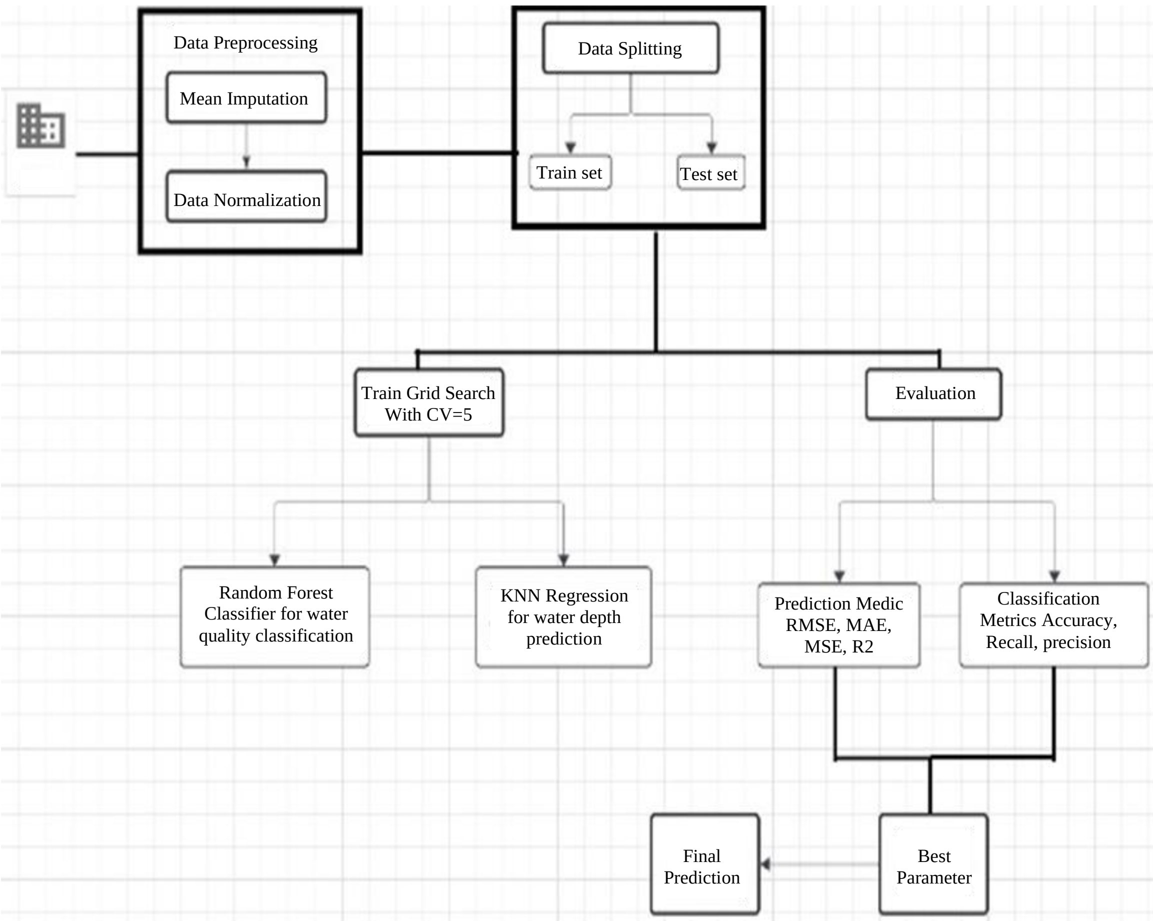

Figure 1. The architecture of the AI model.

Connections between the GUI and predictive models can be established to enable seamless interaction. In addition, the utilization of appropriate APIs or libraries to invoke model functionalities based on user inputs and displaying the results within the GUI is required. Developing a comprehensive and user-friendly GUI for an AI-enabled water well predictor involves careful planning and integration of various components. By focusing on user experience and interactive visualizations, the system can effectively assist users in making informed decisions regarding water well construction [10].

Development of AI Model and User Interface

GitHub Repository Setup

A GitHub repository was created to host all codes and files related to the development of the AI model and user interface. The repository serves as a centralized version of the control system, allowing collaborative development and tracking of changes over time. Detailed documentation and README files are included in the repository to provide instructions for setting up the environment, running the code, and contributing to the project.

Model Development

Machine learning models for predicting groundwater-related parameters were developed using the Python programming language and relevant libraries, such as scikit-learn, TensorFlow, or PyTorch. The codes for data preprocessing, feature engineering, model training, and evaluation were implemented and organized within the GitHub repository. Hyperparameter tuning techniques, such as grid search or Bayesian optimization, have been applied to optimize model performance and generalization.

Streamlit Integration

Streamlit, a Python library for building interactive web applications, was utilized to create a user-friendly interface for accessing predictive models. The Streamlit code was developed and stored within the GitHub repository, allowing seamless integration with the backend AI models. User interface components, such as input forms, data visualizations, and result displays, were implemented using Streamlit's intuitive syntax and built-in widgets.

Frontend Development

HTML, CSS, and JavaScript codes for customizing the appearance and layout of the user interface were included in the GitHub repository. Responsive design principles were applied to ensure compatibility with various devices and screen sizes, thus enhancing user experience and accessibility. Interactive elements, animations, and transitions were implemented to improve usability and engagement within the user interface.

Deployment and Testing

The developed AI model and user interface were deployed and tested in a controlled environment to validate their functionality and performance.

Continuous integration and deployment (CI/CD) pipelines were set up to automate the deployment process and ensure smooth updates and versioning of the application.

Extensive testing, including unit, integration, and user acceptance tests, was conducted to identify and address any bugs or issues prior to public release.

AI-Based Predictive Models for Water Well Construction

Ground Water Quality Assessment

The predictive model for groundwater quality assessment is an integral component of the decision support system proposed in this study, aimed at evaluating the expected quality of groundwater or water probability at specific locations. By leveraging a comprehensive analysis of hydrogeological, environmental, and anthropogenic parameters derived from data provided by the CGWB, this model facilitates informed decision-making regarding water resource management and environmental protection (Figure 2).

In this study, a predictive modeling approach for assessing water potability was performed using a random forest classifier. The dataset used in this study contains features related to water quality, including pH levels, hardness, and solid concentration. The dataset was preprocessed to handle missing values and class imbalance (Figure 3).

Class imbalance can be handled by separating the data into two classes (potable and non-potable) and upsampling the minority class (potable) to balance the dataset. The upsampled data with the majority class (not potable) are then combined, and the data can be shuffled to remove any inherent order.

The random forest classifier was trained on the preprocessed data after splitting them into training and testing sets. Hyperparameter tuning was performed using GridSearchCV, a technique that systematically searches through a specified parameter grid and cross-validates the model to determine the optimal hyperparameters. The tuned hyperparameters included the number of estimators and the minimum number of samples per leaf.

The performance of the model was evaluated using classification metrics including precision, recall, F1-score, and support for each class. In addition, the accuracy of the model was assessed using a test set. The final model achieved a high level of accuracy, thereby demonstrating its efficacy in predicting water potability. This study contributes to the field of water quality assessment by providing a robust predictive modeling framework that can aid in identifying potable water sources, thereby ensuring public health and safety.

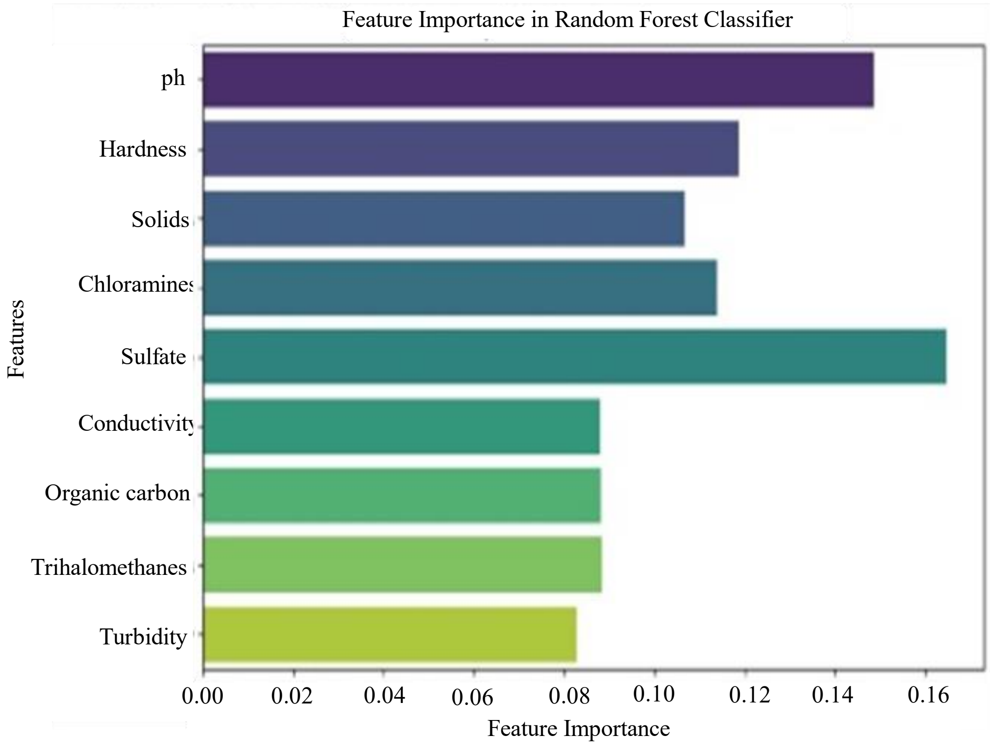

Figure 2. Feature importance in random forest classifier.

```
notpotable  = data[data['Potability']==0]
potable = data[data['Potability']==1]
df_minority_upsampled = resample(potable, replace = True, n_samples = 1200)
data = pd.concat([notpotable, df_minority_upsampled])
data = shuffle(data)
```

Figure 3. Handling class imbalance.

Underground Water Levels Prediction for Water-Bearing Zones

The model aims to estimate the depth at which water-bearing zones can be expected to be encountered in a given location, utilizing a combination of geological and hydrogeological data derived from NAQUIM provided by the CGWB. Information related to topography, land use, and proximity to surface water bodies may also be considered as input features. Feature engineering may involve data transformation, scaling, or the creation of composite features to capture complex relationships within a dataset (Figure 4).

This study presents a predictive modeling approach for estimating water depth based on latitude and longitude coordinates using K-nearest neighbor (KNN) regression. The dataset utilized in this study contained geographical coordinates (latitude and longitude) and the corresponding water depth measurements. The workflow of the study included data preprocessing where outliers were detected and removed from the dataset using the Interquartile Range (IQR) method to ensure data quality and reliability (Figure 5).

The latitude and longitude coordinates were extracted as features (X), and the water depth was selected as the target variable (y) for the prediction. Then, the dataset was split into training and testing sets (80% training, 20% testing) to facilitate model training and evaluation. Features were standardized to have a mean of 0 and a standard deviation of 1 to ensure uniformity and improve model performance (Figure 6).

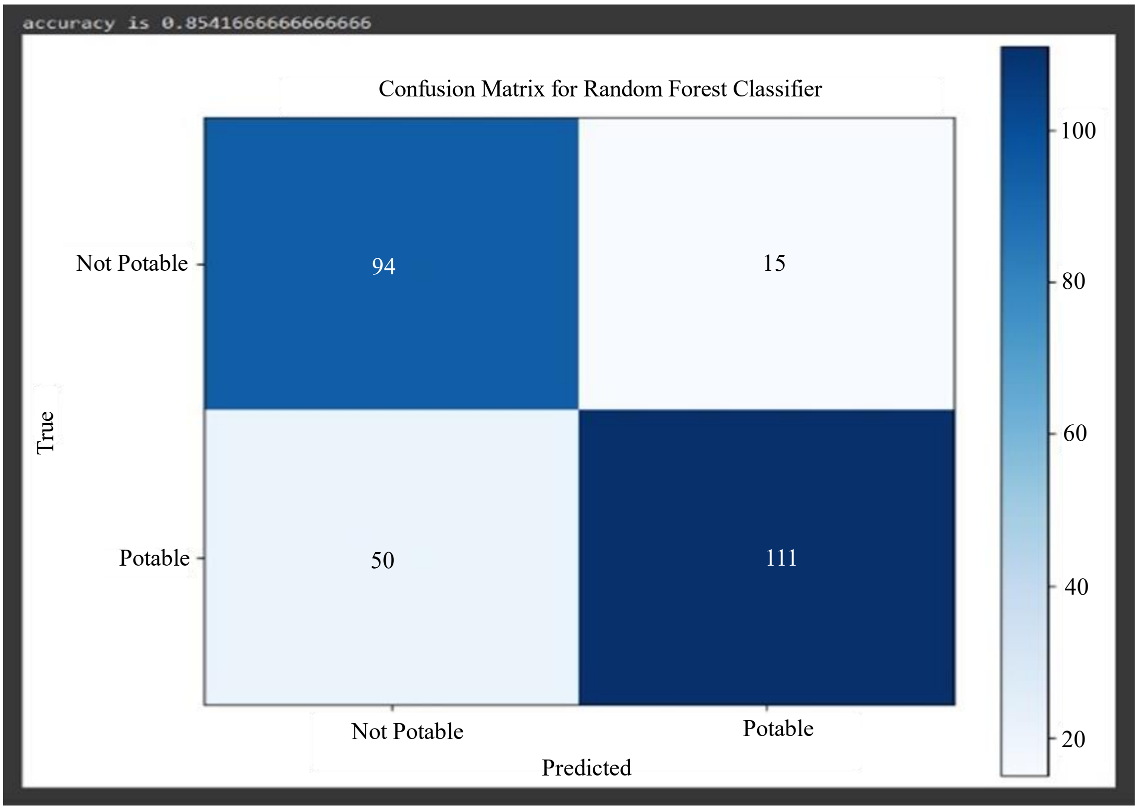

Figure 4. Accuracy of the water probability obtained on training NAQUIM data of CGWB.

```
# Outlier detection and removal
Q1 = data.quantile(0.25)
Q3 = data.quantile(0.75)
IQR = Q3 - Q1
data = data[~((data < (Q1 - 1.5 * IQR)) | (data > (Q3 + 1.5 * IQR))).any(axis=1)]
```

Figure 5. Interquartile Range (IQR) method to detect and remove outliers from data.

A KNN regression model was trained on the training data with a specified value of k (number of neighbors). Cross-validation was employed to assess the model's performance, using the negative mean squared error (MSE) as the evaluation metric. This ensures the robustness and generalization of the model. The trained model was evaluated on the test set using the MSE to quantify the accuracy of water depth predictions (Figure 7).

The results of this study demonstrate the efficacy of the KNN regression approach in predicting water depth based on geographical coordinates. The methodology presented in this study can be applied to various environmental monitoring and management applications, contributing to the advancement of hydrological studies and water resource management. The performance of the depth prediction model was evaluated using metrics such as mean absolute error, root mean square error, and coefficient of determination (R-squared).

RESULTS AND EVALUATION

Performance of AI Model

Several metrics and techniques can be employed to assess the accuracy, reliability, and effectiveness of predictive models. Some of the techniques that are used to predict include accuracy metrics where the mean absolute error (MAE) or the MSE, measures the average magnitude of errors between the predicted and actual values. By taking the inputs of the water probability, latitude, and longitude from the user, the area suitable for the construction of a water well can be predicted (Figures 8–11).

```
# Extract features (latitude and longitude) and target variable (water depth)
X = data[['LAT', 'LON']].values
y = data['depth'].values

# Split data into train and test sets (80% train, 20% test)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Figure 6. Feature extraction and data splitting.

```
Mean Cross-Validation RMSE: 2.8187229716204065
Mean Squared Error: 2.7791767686723294
Predicted Water Depth: 4.74
```

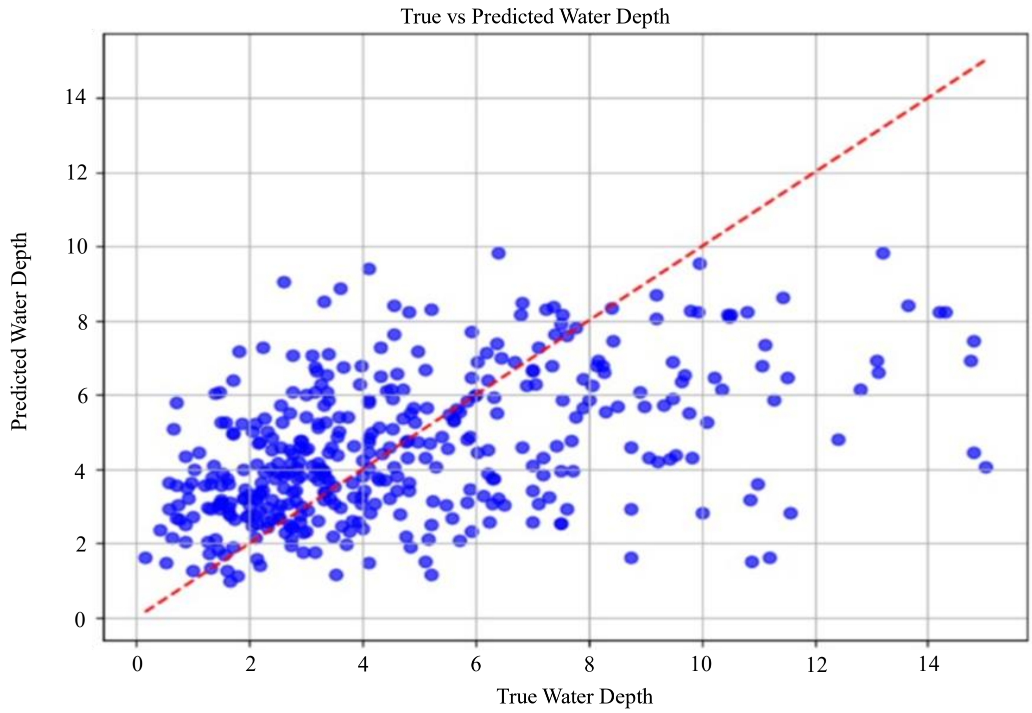

Figure 7. MSE of the predicted depth obtained on training NAQUIM data of CGWB.

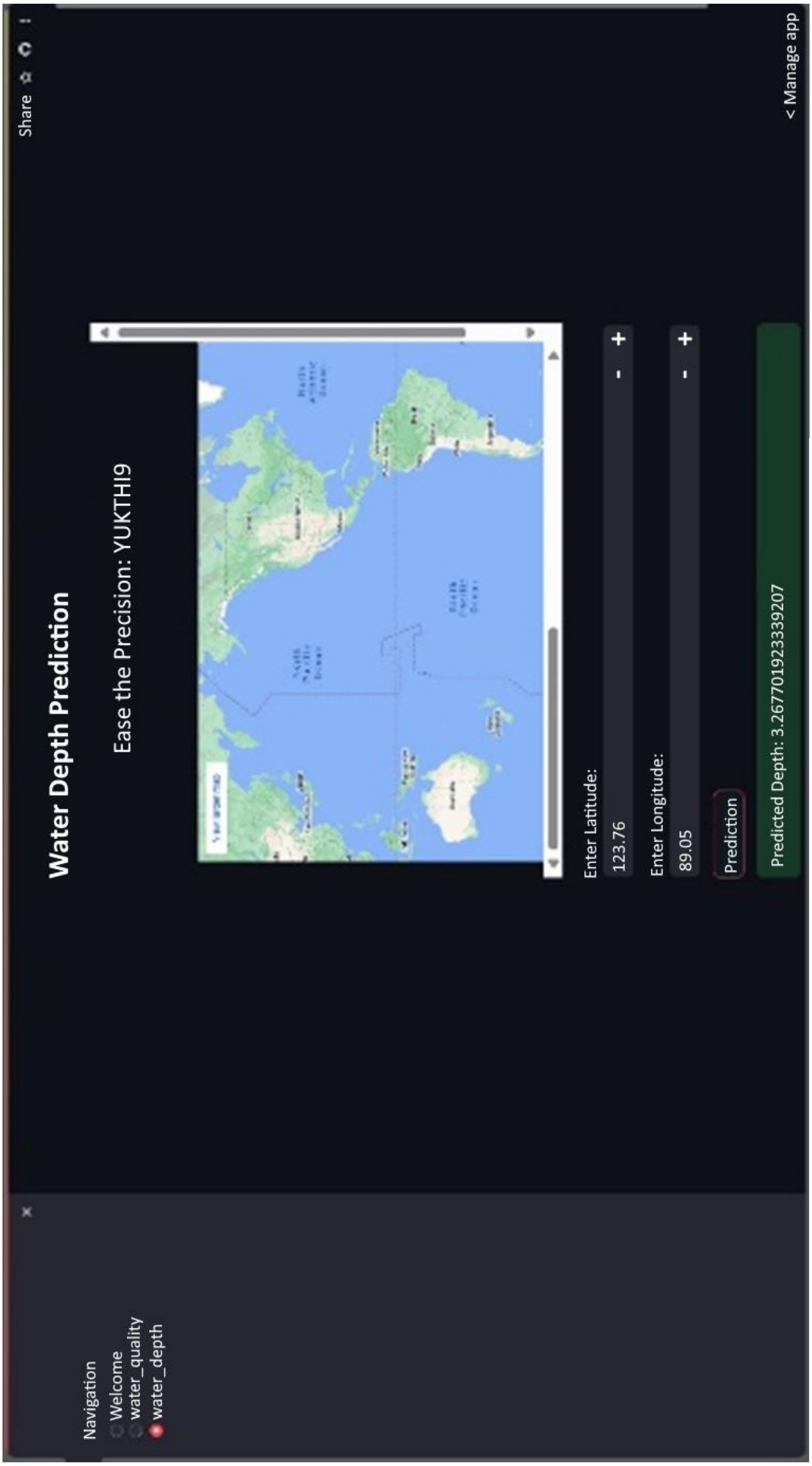

Figure 8. Selection of the input coordinates through map integration.

Figure 9. Prediction of water probability or quality.

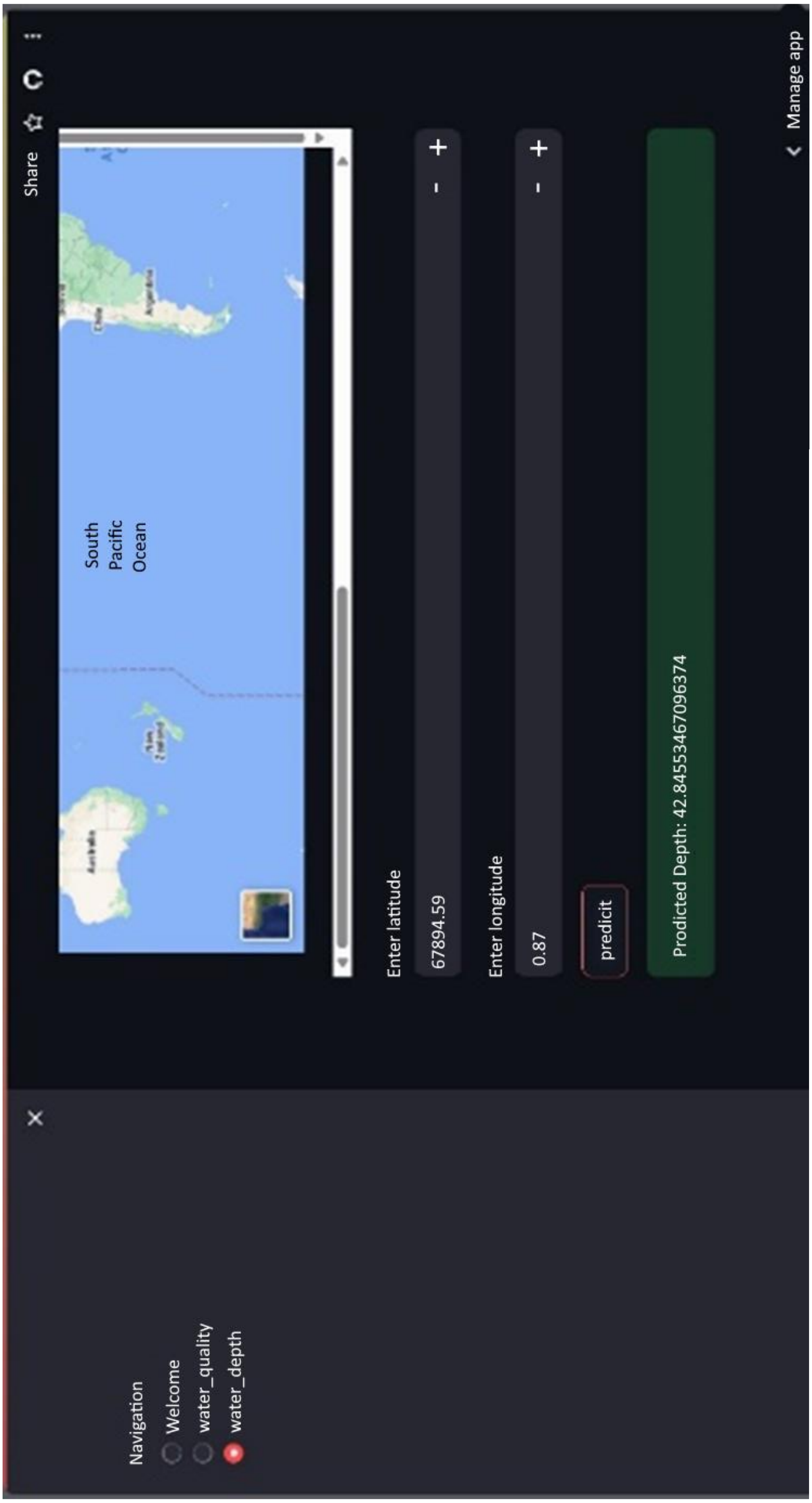

Figure 10. Prediction of underground water levels.

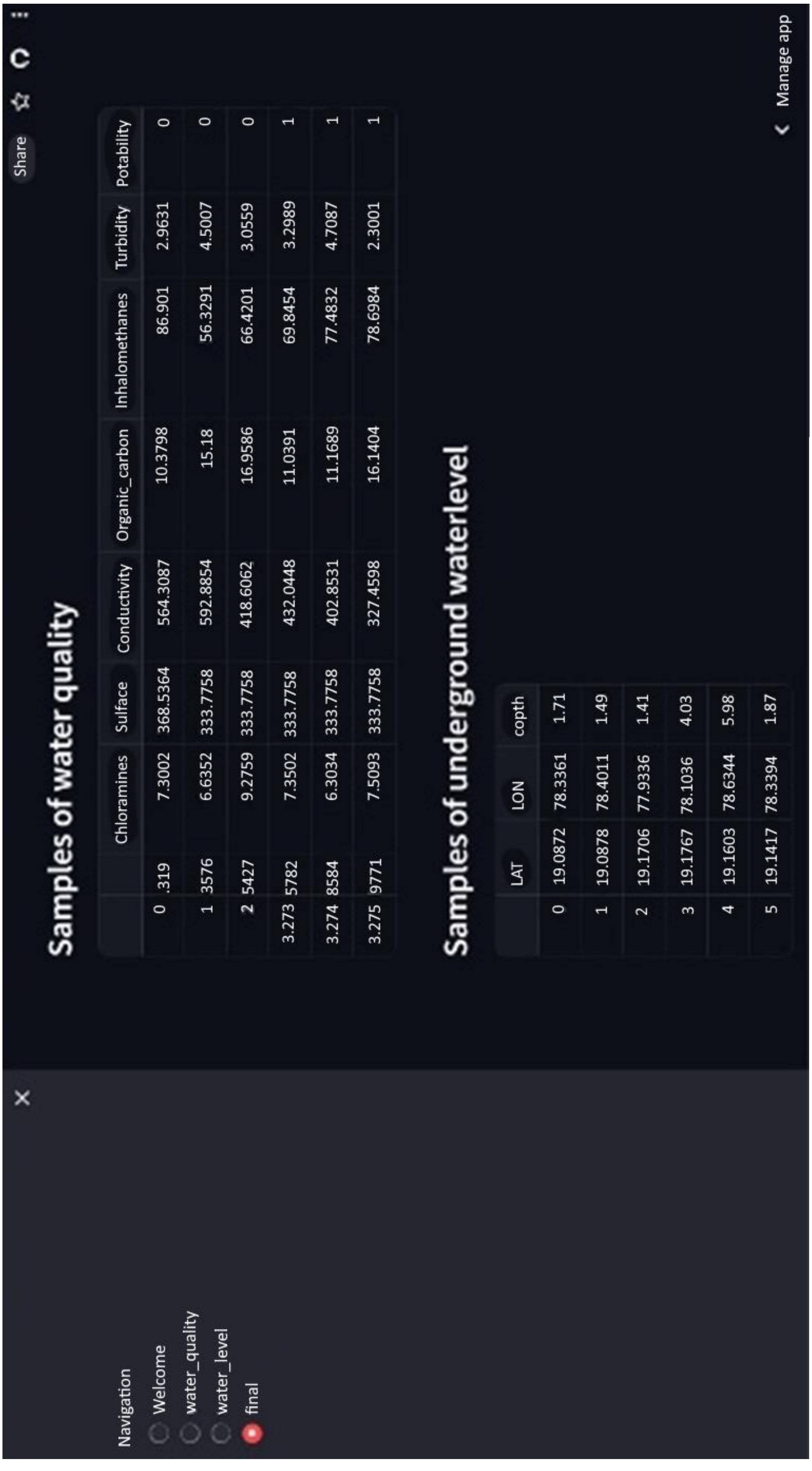

Figure 11. Selected locations with water quality and water levels.

FUTURE DIRECTIONS

While the research has achieved significant milestones in enhancing decision support for water resource management, there are several avenues for future research and development: Enhanced Data Integration is required to explore opportunities to integrate additional datasets and sources of information, including real-time monitoring data, remote sensing imagery, and crowd-sourced data, to improve the accuracy and timeliness of predictive models.

Spatial and temporal analyses are essential for incorporating spatial and Temporal Analysis techniques to account for the dynamic nature of groundwater systems and to assess long-term trends, seasonal variations, and climate change impacts on groundwater quality and availability.

The development of uncertainty quantification methodologies to quantify and propagate uncertainties in predictive modeling outputs enables users to make robust decisions under conditions of uncertainty and variability.

CONCLUSION

This research introduced a comprehensive approach to support decision-making in water well construction through the development of predictive models and a user-friendly graphical interface. Leveraging NAQUIM data provided by the CGWB, our predictive models have demonstrated promising performance in assessing suitability, depth prediction, discharge estimation, and drilling technique selection for water wells. The user-friendly GUI facilitated seamless interaction with the models, enabling them to make informed decisions regarding groundwater management and infrastructure development. Through rigorous performance evaluation, the accuracy, reliability, and robustness of the predictive models were validated, ensuring their practical utility in diverse geographical and hydrogeological contexts.

Limitations and Potential Challenges

Although the proposed research paper presents a promising approach to aid decision-making in water well construction through predictive modeling and a user-friendly interface, several potential limitations and challenges should be considered.

- Validating predictive models requires robust testing against independent datasets or real-world observations.
- Compliance with regulatory requirements and ethical considerations such as data privacy and confidentiality are paramount when handling sensitive information related to groundwater resources and environmental data.
- Integrating the system with existing water management platforms, databases, and workflows may present technical challenges and interoperational issues. Seamless integration ensures system adoption and usability within established frameworks.

REFERENCES

1. Gossell MA, Nishikawa T, Hanson RT, Izbicki JA, Tabidian MA, Bertine K. Application of flowmeter and depth-dependent water quality data for improved production well construction. Ground Water. 1999;37(5):729–35. doi:10.1111/j.1745-6584.1999.tb01165.x. PMID: 19125926.
2. Thakur BK, Gupta V, Bhattacharya P, Jakariya M, Tahmidul Islam M. Arsenic in drinking water sources in the Middle Gangetic Plains in Bihar: An assessment of the depth of wells to ensure safe water supply. Groundw Sustain Dev. 2021;12:100504. doi:10.1016/j.gsd.2020.100504.
3. Liu G, Schwartz FW, Tseng KH, Shum CK. Discharge and water-depth estimates for ungauged rivers: Combining hydrologic, hydraulic, and inverse modeling with stage and water-area measurements from satellites. Water Resour Res. 2015;51(8):6017–35. doi:10.1002/2015WR 016971.
4. Chandel NS, Chakraborty SK, Chandel AK, Dubey K, Subeesh A, Jat D, et al. State-of-the-art AI-enabled mobile device for real-time water stress detection of field crops. Eng Appl Artif Intell. 2024 May;131:107863. doi:10.1016/j.engappai.2024.107863.

5. Barkman JH, Davidson DH. Measuring water quality and predicting well impairment. J Pet Technol. 1972;24(7):865–73. doi:10.2118/3543-PA.
6. Singh B, Sihag P, Singh VP, Sepahvand A, Singh K. Soft computing technique-based prediction of water quality index. Water Supply. 2021;21(8):4015–29. doi:10.2166/ws.2021.157.
7. Khan MSI, Islam N, Uddin J, Islam S, Nasir MK. Water quality prediction and classification based on principal component regression and gradient boosting classifier approach. J King Saud Univ Comput Inf Sci. 2022;34(8 Pt A):4773–81. doi:10.1016/j.jksuci.2021.06.003.
8. Banerjee K, Bali V, Nawaz N, Bali S, Mathur S, Mishra RK, Rani S. A machine-learning approach for prediction of water contamination using latitude, longitude, and elevation. Water. 2022;14(5):728. doi:10.3390/w14050728.
9. Central Ground Water Board (CGWB), Ministry of Jal Shakti, Department of Water Resources, River Development and Ganga Rejuvenation, Government of India. (2024). Retrieved from http://cgwb.gov.in.
10. India Water Resources Information System (IWRIS), Ministry of Jal Shakti, Department of Water Resources, River Development and Ganga Rejuvenation, Government of India. (2024). Retrieved from http://indiawris.gov.in.

