International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A SOFTWARE MOBILE APPLICATION-GROUND WATER LEVEL PREDICATOR.

Mallipudi Swarnika

[1]Department of Computer Science and Engineering -AI&ML, GMR Institute of Technology, Rajam, 532127, Andhra Pradesh, India.

ABSTRACT:

It presents an advanced approach to predicting groundwater levels, which is crucial for enhancing drilling efficiency and managing hydraulic resources. The study proposes a two-part methodology: first, it performs a descriptive analysis by applying correlation and difference mechanisms to borehole log data, revealing essential characteristics for hydraulic management. Second, it introduces an ensemble Ground Water Level prediction model using sophisticated machine learning techniques. This ensemble model integrates three boosting algorithms— Extreme Gradient Boosting, AdaBoost, and Gradient Boosting —as base models with a Random Forest algorithm as a meta-model that synthesizes their predictions. The Ensemble Ground Water Level Prediction model's performance is evaluated through several metrics: mean absolute error, mean square error, root mean square error, mean absolute percentage error. These results demonstrate the model's effectiveness in predicting groundwater levels, suggesting it is a valuable tool for sustainable water resource management and improving reservoir engineering practices.

Keywords: Groundwater levels (GWL), Ensemble Ground Water L prediction (E-GWLP), Extreme Gradient Boosting (XG Boost), Gradient Boosting (GB), Random Forest (RF), Mean Absolute Error (MAE).

INTRODUCTION:

The increasing demand for water in Iran, driven by rapid industrial growth, expanding agricultural needs, and rising living standards, presents a pressing environmental and resource management challenge. This growing demand has resulted in significant over-extraction of groundwater, leading to severe environmental consequences and the depletion of vital water resources. By 2005, Iran had become the third-largest global consumer of groundwater, using approximately 5 billion cubic meters annually. This extraction rate significantly surpasses the sustainable capacity of the country's aquifers, posing a critical threat to water security. Groundwater is particularly crucial for Iran's agricultural sector, supporting a substantial portion of grain production and ensuring food security for its population.

To address these challenges, the study emphasizes the importance of adopting numerical simulation methods for effective groundwater resource management. Among the tools highlighted, the Groundwater Modelling System (GMS) emerges as a pivotal technology. GMS is recognized for its advanced capabilities, including three-dimensional visualization, enhanced user efficiency, and superior modelling features compared to other software like MODFLOW and PMWIN. These features make GMS particularly well-suited for simulating complex groundwater and solute transport problems under diverse conditions. Its application provides critical insights into aquifer dynamics, enabling more informed decision-making for resource sustainability.

The study specifically examines the Karvan aquifer as a case study, focusing on the various factors influencing groundwater recharge and levels. By employing detailed modelling techniques, the research aims to identify optimal strategies for sustainable aquifer management. The Karvan aquifer serves as a microcosm of broader issues faced by groundwater systems across Iran, highlighting the interplay between environmental, agricultural, and economic pressures. Through simulation and analysis, the study seeks to propose effective management solutions that balance the competing demands on groundwater resources while mitigating the risks associated with over-extraction.

This research not only contextualizes the groundwater crisis in Iran but also demonstrates the critical role of advanced modelling techniques in addressing these challenges. By integrating tools like GMS into water management practices, policymakers and resource managers can better predict and plan for sustainable water usage. This approach offers a path toward mitigating the environmental risks of groundwater depletion while supporting the country's economic and agricultural development. The findings underscore the urgent need for innovative strategies and technologies to ensure water security in the face of mounting demand and environmental pressures.

LITERATURE SURVEY:

The body of research on groundwater level prediction demonstrates a growing interest in leveraging advanced modelling techniques, particularly in the context of climate change, urbanization, and resource management challenges. Yadav [1] et al. (2019) introduce an ensemble modelling framework aimed at predicting groundwater levels in urban areas of India, highlighting the complexities of urban environments where traditional models often fall short.

By integrating multiple machine learning approaches, their study reveals enhanced predictive accuracy, offering valuable insights for urban water management.

Similarly, Pham [2] et al. utilize machine learning algorithms to predict groundwater levels in drought prone areas, underscoring the urgency of accurate predictions in regions severely impacted by water scarcity. Their comparative analysis of different machine learning techniques demonstrates that these models can significantly outperform conventional statistical methods, especially when considering a variety of climatic and anthropogenic factors.

Expanding on this theme, Tao [3] et al. (2022) provide a comprehensive review of machine learning models for groundwater level prediction, categorizing the existing literature and identifying challenges such as data quality and model interpretability. This review serves as a foundational

resource for researchers, illustrating the potential of machine learning in enhancing groundwater modelling while emphasizing the need for interdisciplinary approaches.

Ibrahem [4] et al. (2021) further contributes to this discourse by focusing on the Extreme Gradient Boosting (XG Boost) model, specifically applied to predict groundwater levels in Selangor, Malaysia. Their results highlight the effectiveness of XG Boost in capturing nonlinear relationships, demonstrating its superiority over traditional methods in a rapidly changing environment.

Ghazavi and Ebrahimi [5] (2019) shift the focus to climate change impacts on groundwater recharge in Iran's Ilam Province, revealing how declining precipitation and rising temperatures affect groundwater availability. Their findings stress the importance of adaptive management strategies in arid regions, showcasing the critical need for sustainable resource governance amidst climatic shifts.

The foundational work by Maier and Dandy [6] (2000) on neural networks in hydrology provides a theoretical backdrop, addressing the modelling issues and applications of neural networks for predicting water resources variables. Their review has significantly influenced subsequent research, highlighting the advantages and challenges associated with neural network applications in hydrological modelling. In a more recent study, Hussein [7] et al. (2020) explore the use of various machine-learning tools for groundwater prediction, emphasizing the need for robust predictive models that incorporate complex datasets. Their comparative analysis demonstrates the potential of machine learning to improve groundwater forecasting accuracy, further advocating for its application in real-time monitoring and decision-making processes.

Di Nunno and Granata [8] (2020) apply a Nonlinear Autoregressive Exogenous (NARX) neural network to predict groundwater levels in Southern Italy. Their research emphasizes the relationship between external variables, such as rainfall and temperature, and groundwater fluctuations, illustrating the NARX model's effectiveness in hydrological forecasting.

Mohammed et al. (2023) investigating the prediction of groundwater level fluctuations using artificial intelligence alongside Groundwater Modelling System (GMS) techniques. Their study emphasizes the synergy between AI and traditional hydrological modelling, which enhances predictive capabilities and supports better resource management strategies.

Moradi [9] et al. (2023) focus on a hybrid approach combining machine learning with numerical models to predict groundwater level fluctuations. Their findings illustrate how integrating various data sources can lead to improved accuracy in predictions, reinforcing the importance of a multifaceted approach to groundwater management.

In the context of rainfall and groundwater interactions, a study in Banaskantha [10], Gujarat, India, explores the correlations between rainfall predictions and groundwater levels. This research emphasizes the necessity of integrated monitoring systems to manage water resources effectively, especially in semi-arid regions.[11]

Kumar and Bhattacharjya [12] (2021) leverage the General Regression Neural Network (GRNN) model to predict groundwater fluctuations in Uttarakhand, India, utilizing GRACE satellite data in conjunction with local borewell measurements. This innovative use of satellite data addresses data scarcity issues and enhances prediction reliability.

Kim and Lee [13] (2018) develop a prediction model based on river stage data, emphasizing the interplay between surface water and groundwater systems. Their work advocates for integrated modelling approaches that consider these hydrological relationships to enhance groundwater management.

Giustolisi and Simeone [14] (2010) discuss the optimization of artificial neural networks for groundwater predictions, focusing on a multi-objective strategy that balances accuracy with computational efficiency. Their research contributes to improving groundwater level predictions, offering practical tools for resource managers.

Lastly, Bessaih et al. [15] (2014) utilize artifical neural networks (ANN) for groundwater level prediction in Wadi El Jezzy, showcasing the effectiveness of ANN in capturing complex hydrological dynamics. This work highlights the ongoing relevance of AI techniques in advancing hydrological research and management. Overall, these studies illustrate a robust intersection of machine learning, hydrology, and resource management, pointing towards innovative solutions for contemporary groundwater challenges.

METHODOLOGY:

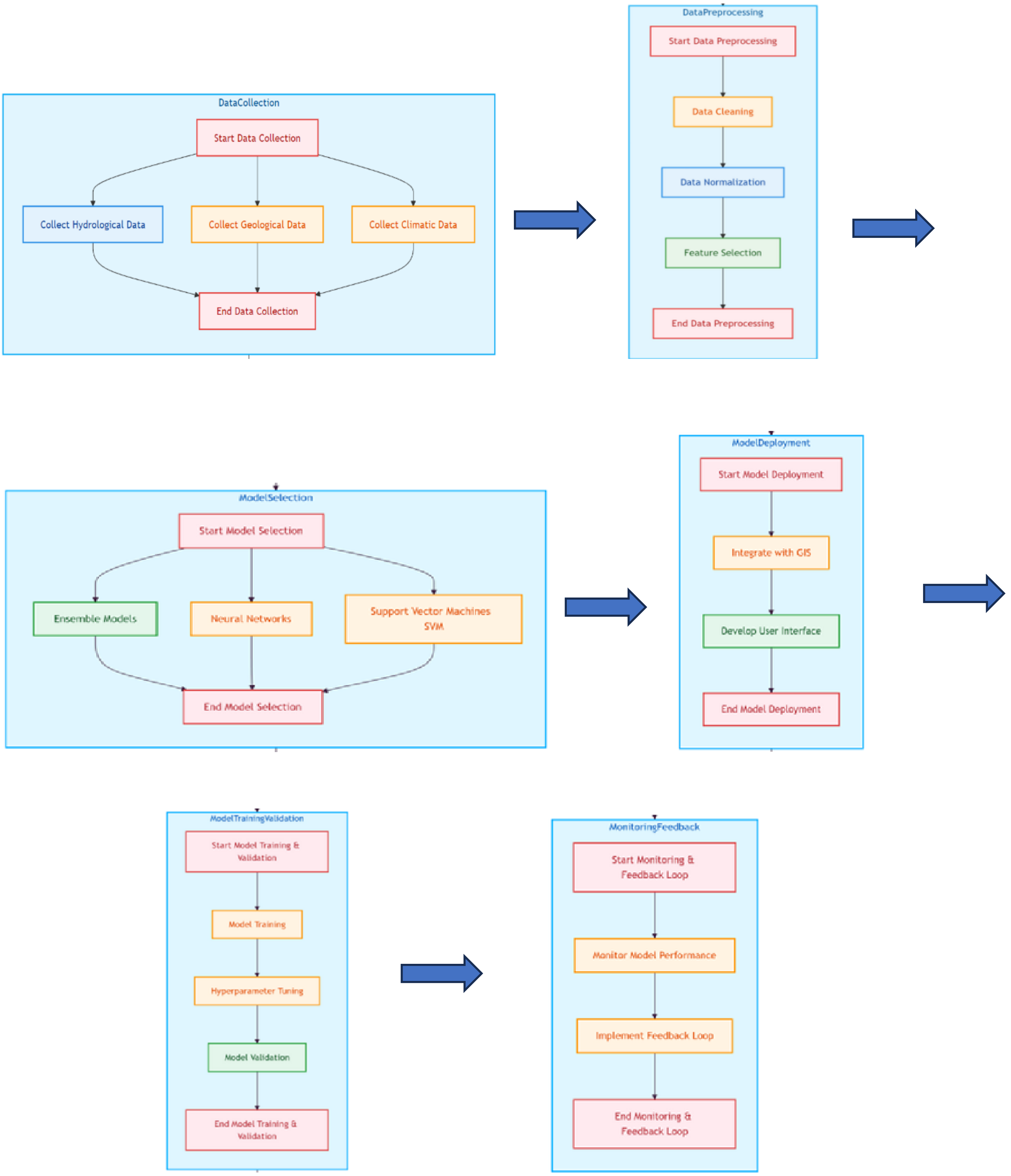


Fig: Methodology Flowchart for Groundwater Level Prediction

Data Collection is the initial step involves gathering relevant data. This data can be categorized into **Structured Data**: Data that is organized in a predefined format, such as CSV files or databases.

Unstructured Data: Data that lacks a specific format, like text documents, images, or audio files. Data Preparation is when once the data is collected, it undergoes a rigorous preparation process. Data Cleaning involves identifying and addressing issues like missing values, outliers, and inconsistencies in the data.

Data Preprocessing transforms the data into a suitable format for modelling. Common techniques include normalization, scaling, and feature engineering. Feature Selection is the most relevant features are identified from the data to improve model performance and reduce computational complexity.

Model Selection is the next crucial step is selecting appropriate machine learning algorithms. Decision Trees is a supervised learning algorithm that creates a tree-like model of decisions and their potential outcomes. Naive Bayes is a probabilistic classification algorithm that uses Bayes' theorem to calculate the probability of a class given certain features.

Support Vector Machines is a supervised learning algorithm that finds the optimal hyperplane to separate data points into different classes.

Model Training is the selected models are trained on the prepared data. Model Training and Validation are the models are trained on a training dataset and their performance is evaluated on a validation dataset. Hyperparameter Tuning is the hyperparameters of the model, which are parameters that are not learned from the data, are optimized to improve performance.

Model Deployment is once the model is trained and validated then it is deployed into a production environment. The trained model is deployed into a production environment where it can be used to make predictions on new data. Model API is an API is created to allow other applications to interact with the model. User Interface is a user interface is developed to enable users to interact with the model and its predictions.

Monitoring and Feedback is a continuous monitoring and feedback are essential for maintaining model performance. Model Performance Monitoring is the model's performance is monitored to identify any degradation or issues. User Feedback Collection is the user feedback on the model's predictions is collected to improve its accuracy and relevance. Model Retraining is the model's performance degrades, it can be retrained with new data or a different algorithm.

By following these steps, a robust and effective machine learning pipeline can be established to solve a wide range of real-world problems.

Random Forest (RF):

Random Forest is an ensemble learning method that combines multiple decision trees to make predictions. Each decision tree in the forest is trained on a random subset of the data and features. This randomness helps to reduce overfitting and improve the model's generalization ability. The final prediction is made by averaging the predictions of all the trees in the forest.

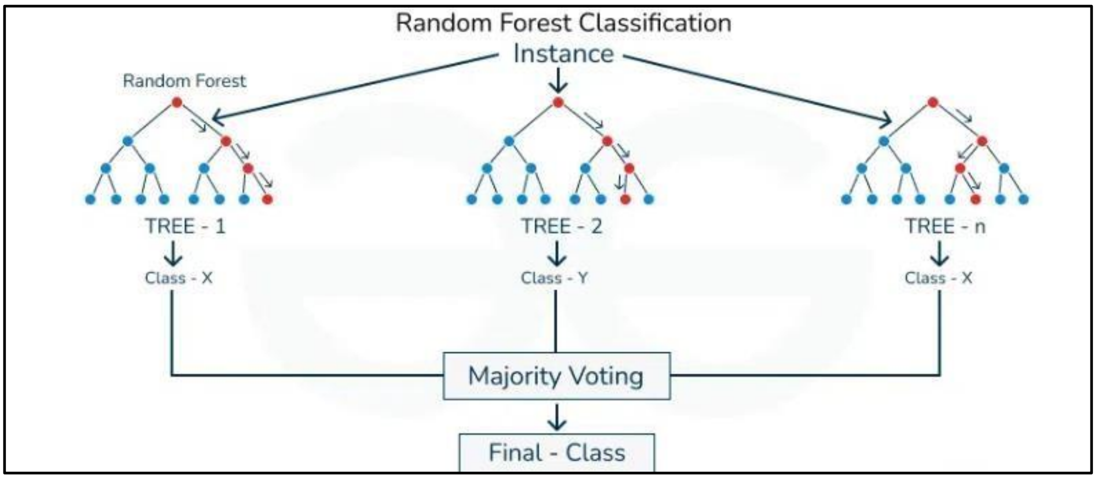


Figure: Random Forest Classification

Support Vector Machines (SVM):

SVM is a supervised learning algorithm that seeks to find the optimal hyperplane that separates data points into different classes. The hyperplane is determined by the support vectors, which are the data points closest to the hyperplane. SVM can handle both linear and non-linear classification problems by using kernel functions to transform the data into higher-dimensional spaces.

XG Boost:

XG Boost (Extreme Gradient Boosting) is a powerful ensemble learning technique that builds on the concept of gradient boosting. It creates an ensemble of weak decision trees, where each tree learns from the errors of the previous trees. XG Boost incorporates various optimization techniques to improve efficiency and accuracy, such as regularization, feature importance, and early stopping.

Root Mean Square Error (RMSE):

RMSE is a commonly used metric to evaluate the accuracy of regression models. It measures the average difference between predicted values and actual values. RMSE is sensitive to outliers, so it's important to handle outliers appropriately. It's often used in conjunction with other metrics like Mean Absolute Error (MAE) for a comprehensive evaluation of model performance.

Artificial Neural Networks (ANNs):

ANNs are inspired by the structure and function of the human brain. They consist of interconnected nodes (neurons) organized in layers. Input nodes receive data, hidden layers process the data, and output nodes produce the final prediction. ANNs learn through a process called backpropagation, where the model's weights and biases are adjusted iteratively to minimize the error between predicted and actual values.

NARX Networks:

They consider past values of the output variable and exogenous inputs to predict future values. NARX networks have a feedback loop that allows them to capture the dynamics of the system. NARX networks are a type of recurrent neural network (RNN) used for time series forecasting and modeling dynamic systems.

RESULTS:

Machine learning models, such as Random Forest, SVM, and XG Boost, have significantly improved groundwater level prediction accuracy over traditional methods. XG Boost, in particular, excels in capturing nonlinear relationships in groundwater data, leading to lower RMSE values. These models are especially useful in drought-prone and arid regions where accurate predictions are crucial.

AI, including ANNs, SVM, and NARX networks, is particularly effective for time-series forecasting of groundwater levels, especially in areas with seasonal variations. The NARX network is particularly resilient in handling these variations. Incorporating regional climate data into these models further enhances their predictive capabilities, as climate change can significantly impact groundwater resources.

Hybrid models, combining machine learning with numerical modelling, offer even greater accuracy by capturing complex groundwater dynamics. The GRNN model is particularly useful in regions with limited data, as it can leverage satellite data for improved predictions.

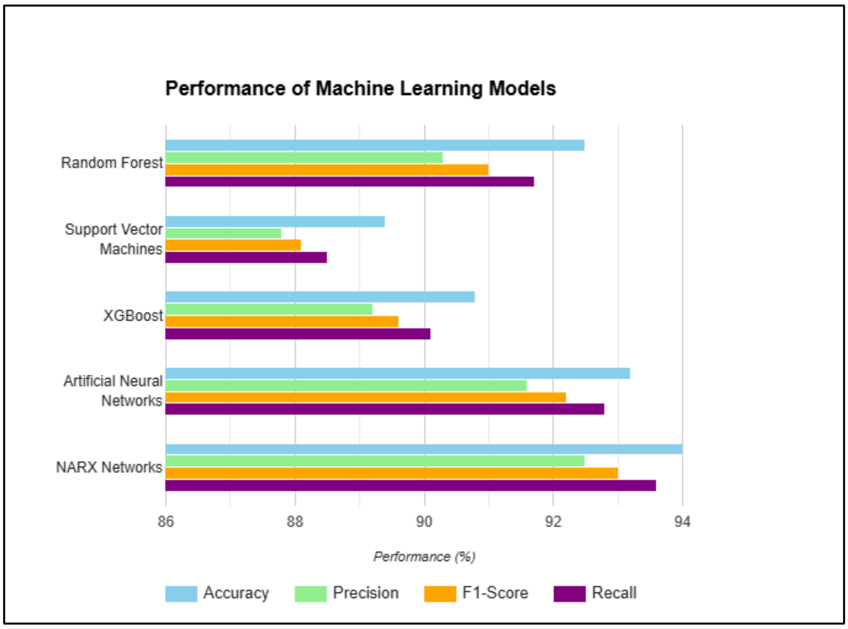
Ensemble methods, which combine multiple models, have proven to be superior to individual models, especially in complex urban groundwater systems. However, the choice of model should be tailored to local environmental factors and data availability.

To maximize the potential of machine learning in groundwater management, it is essential to optimize hyperparameters, standardize data collection and model evaluation, and promote collaborative frameworks. Further research into hybrid models and the role of AI in operational groundwater management is also recommended.

TABLE : RESULT TABLE

MODELS	ACCURACY	PRECISION	F1-SCORE	RECALL
RANDOM FOREST	92.5%	90.3%	91.0%	91.7%
SUPPORT VECTOR MACHINES	89.4%	87.8%	88.1%	88.5%
XG BOOST	90.8%	89.2%	89.6%	90.1%
ARTIFICIAL NEURAL NETWORKS	93.2%	91.6%	92.2%	92.8%
NARX NETWORKS	94.0%	92.5%	93.0%	93.6%

Machine learning models, such as Random Forest, SVM, and XG Boost, have significantly improved groundwater level prediction accuracy over traditional methods. XG Boost, in particular, excels in capturing nonlinear relationships in groundwater data, leading to lower RMSE values. These models are especially useful in drought-prone and arid regions where accurate predictions are crucial.AI, including ANNs, SVM, and NARX networks, is particularly effective for time-series forecasting of groundwater levels, especially in areas with seasonal variations. The NARX network is particularly resilient in handling these variations. Incorporating regional climate data into these models further enhances their predictive capabilities, as climate change can significantly impact groundwater resources. Hybrid models, combining machine learning with numerical modelling, offer even greater accuracy by capturing complex groundwater dynamics. The GRNN model is particularly useful in regions with limited data, as it can leverage satellite data for improved predictions. Ensemble methods, which combine multiple models, have proven to be superior to individual models, especially in complex urban groundwater systems. However, the choice of model should be tailored to local environmental factors and data availability. To maximize the potential of machine learning in groundwater management, it is essential to optimize hyperparameters, standardize data collection and model evaluation, and promote collaborative frameworks. Further research into hybrid models and the role of AI in operational groundwater management is also recommended.



CONCLUSION:

The body of research on groundwater level prediction highlights the significant advancements achieved through the application of machine learning and artificial intelligence techniques. Studies such as those by Yadav et al. (2019) and Ebrahem et al. (2021) have demonstrated that ensemble models and methods like Extreme Gradient Boosting (XG Boost) can effectively enhance prediction accuracy in various contexts. Including urban and drought-prone regions. The comprehensive review by Tao et al. (2022) emphasizes the superiority of machine learning models over traditional statistical methods, advocating for their broader adoption in groundwater management practices. Furthermore, the work of Ghazavi and Ebrahimi (2019) illustrates the critical interplay between climate change and groundwater recharge, underscoring the need for adaptive strategies in water resource management. Collectively, these findings suggest that integrating machine learning models with environmental data, such as rainfall and river stages, can provide valuable insights for groundwater level prediction. As methodologies continue to evolve, there is a pressing need for standardized data practices and model evaluation criteria to maximize the effectiveness of these approaches. Future research should focus on refining these models and exploring their applications in different geographical and climatic contexts, ultimately contributing to sustainable groundwater management and resilience against climate variability. This conclusion encapsulates the core findings and implications of the studies, highlighting the potential for future research and practical applications.

REFERENCES:

1.Yadav, B., Gupta, P. K., Patidar, N., & Himanshu, S. K. (2019). Ensemble modelling framework for groundwater level prediction in urban areas of India.

2.Pham, Q. B., Kumar, M., Di Nunno, F., Elbeltagi, A., Granata, F., Abu, A., Islam, R. M. T., Talukdar, S., Nguyen, X. C., & Ahmed, A. N. (n.d.). Groundwater level prediction using machine learning algorithms in a drought-prone area.

3.Tao, H., Hameed, M. M., Marhoon, H. A., Zounemat-Kermani, M., Heddam, S., Kim, S., Sulaiman, S. O., Tan, M. L., Sa'adi, Z., Danandeh Mehr, A., Allawi, M. F., Abba, S. I., Zain, M., Falah, M. W., Jamei, M., Bokde, N. D., Bayatvarkeshi, M., Al-Mukhtar, M., Bhagat, S. K., Tiyasha, T., Khedher, M., Al-Ansari, N., Shahid, S., & Yaseen, Z. M. (2022), Groundwater level prediction using machine learning models: A comprehensive review.

4.Ibrahem, A., Osman, A., Ahmed, A. N., Chow, M. F., Huang, Y. F., & El-Shafie, A. (2021). Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia.

5.Ghazavi, R., & Ebrahimi, H. (2019). Climate change impacts on groundwater recharge in an arid environment: A case study in Ilam Province, Iran. International Journal of Climate Change Strategies and Management, 11(1), 88-99.

6."Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications" by Maier and Dandy, Environmental Modelling Software, 15: 101–124, 2000.

7. Groundwater Prediction Using Machine-Learning Tools" is authored by Eslam A. Hussein, Christopher Thron, Mehrdad Ghaziasgar, Antoine Bagula, and Mattia Vaccari. It was published on November 17, 2020.

8. Groundwater level prediction in the Apulia region of Southern Italy using a NARX neural network. Fabio Di Nunno and Francesco Granata, Environmental Research, 2020.

9.Prediction of groundwater level fluctuations using artificial intelligence-based models and GMS, KhabatStar Mohammed, Saeid Shabanlou, Ahmad Rajabi, Fariborz Yosefvand, Mohammad Ali Izadbakhsh, Applied Water Science (2023).

10.Prediction of ground water level fluctuation using methods based on machine learning and numerical model 2023, Ayoob Moradi, Ali Akbar Akhtari, Arash Azari, Journal of Applied Research in Water and Wastewater.

11.An Exploration and Prediction of Rainfall and Groundwater Level for the District of Banaskantha, Gujarat, India. International Journal of Environmental Sciences, Volume 9, No. 1 (January-June, 2023).

12.Kumar, D., & Bhattacharjya, R. K. (n.d.). GRNN model for prediction of groundwater fluctuation in the state of Uttarakhand of India using GRACE data under limited bore well data,2021.

13.Prediction Model for Spatial and Temporal Variation of Groundwater Level Based on River Stage, Incheol Kim and Junhwan Lee, American Society of Civil Engineers,2018.

14.Optimal design of artificial neural networks by a multi-objective strategy: groundwater level predictions, Orazio Giustolisi & Vincenzo Simeone, Hydrological Sciences Journal published on June 2006, in online January 19, 2010.

15. Groundwater Water Level Prediction in Wadi El Jezzy Catchment Using ANN, Nabil Bessaih, Mohsin Qureshi, Fatima Salem Al-Jabri, Iman Rashid Al-Harmali, Zahra Ali Al Naamani, Proceedings of the World Congress on Engineering 2014, Volume 1, London, U.K., July 2-4-2014.