ACADEMY JOURNAL OF SCIENCE AND ENGINEERING
Available online @
www.academyjsekad.edu.ng
AJSE 19 (2) 2025
Original Research

THE APPLICATION OF MACHINE LEARNING ALGORITHMS IN MODELLING GROUNDWATER LEVEL FLUCTUATIONS OF THE SOKOTO BASIN

Samson Alfa, Haruna Garba and Augustine Odeh

Department of Civil Engineering, Nigeria Defence Academy, Kaduna, Nigeria

---

Abstract

Groundwater level forecasting is essential for the sustainable management of water resources, especially water scarce regions such as the Sokoto Basin. This study utilises the application of machine learning algorithms, specifically Long Short-Term Memory (LSTM), eXtreme Gradient Boosting (XGBoost)and Random Forest (RF) algorithms to predict groundwater levels across six boreholes within the Sokoto Basin. Modelling was carried out for six boreholes located in the basin with groundwater level as target variable considering rainfall, soil moisture, temperature and humidity as feature variables. Model performance was evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and coefficient of Determination (R^2). Among the models, the XGBoost algorithm demonstrated the highest performance, producing predictions closely aligned with observed groundwater levels. Hyperparameter tuning via grid search further optimized its performance. The LSTM model also showed strong performance, particularly in capturing the peaks and valleys of the groundwater level time series. The RF model exhibited reliable performance across most locations. The study offers a practical framework for regions with limited funding for groundwater monitoring, enabling effective water resource management. This approach can support proactive measures such as water-use restrictions and drought alerts to prevent groundwater depletion, particularly during dry years, thereby contributing to the sustainability of water resources in the Sokoto Basin and similar regions.

Keywords: Groundwater, Modelling, Machine Learning, Forecasting, Sokoto

---

1.0 INTRODUCTION

Groundwater is a vital source of drinking water, irrigation, and industrial use. However, groundwater resources are under increasing pressure from overexploitation, land use, pollution, and climate change (Akintayo et al., 2022; Iqbal et al., 2021). Water usage and

Corresponding author's e-mail:samsonalfa@gmail.com

management are inextricably linked to various facets of human life, encompassing personal, domestic, recreational, industrial, agricultural, and even luxurious demands. However, water also poses significant risks to humanity, stemming from droughts, floods, and the effects of climate change. This highlights the imperative for a strong and harmonious relationship between human civilization and water resources, which can only be achieved through thoughtful development, planning, and management of this precious natural asset (Zaresefat & Derakhshani, 2023). Understanding how these different weather conditions influence the aquifer's recharge, the resulting groundwater levels and the knowledge about the groundwater variations can be used for quantifying its availability.

In hydrology, groundwater modelling is an essential tool for forecasting and managing the behaviour of subsurface water resources (Tao et al., 2022). However, field data collection or monitoring of variables such as water levels, hydraulic conductivity, and water quality is often limited to a few locations due to constraints like time, cost, and safety. To address the challenge of unmeasured or unsampled sites, interpolation techniques are employed. These methods estimate values at unmonitored locations by leveraging spatial patterns and relationships derived from data collected at nearby sites. By filling gaps in data coverage, interpolation enhances the accuracy and utility of groundwater models, as demonstrated in previous studies (Almahawis, 2018). Algorithms for machine learning can be applied in situations where modelled phenomena are obscure, difficult to observe, or poorly understood. This is because their core capability is the ability to learn from empirical data (Deka, 2019), where he delves into the use of machine learning in various civil engineering research. Civil Engineers and everybody else interested in water resources will find the sudden growth in the application of machine learning algorithms in hydrology to be an intriguing and encouraging development. The introduction of big data and artificial intelligence in hydrology has resulted in several new advances in the sustainable development of groundwater resources in several parts of the world (Gaffoor et al., 2020; Zaresefat & Derakhshani, 2023).

For groundwater level forecasting, numerical, statistical, or physical based models are traditionally the main tool; however, they have some practical limitations, including the need for large amount of data and input parameters (Chen et al., 2019). In many cases, data is limited on one hand, and obtaining accurate predictions is more important than understanding underlying mechanisms (Kanyama et al., 2020). Statistical models, however, do not take nonlinear interactions into account (Arabameri et al., 2021), thus as a result artificial-intelligence-based machine-learning (ML) models have been developed. ML approaches based on data mining have determined the conditions required to improve groundwater capacity.

Machine learning methods have gained popularity in groundwater studies due to

Corresponding author's e-mail:samsonalfa@gmail.com

their ability to handle large datasets and identify complex patterns. Decision trees, random forests, and neural networks are among the models frequently employed for groundwater level prediction. Neural networks, including deep learning models, have shown potential in capturing non-linear relationships and temporal dynamics in groundwater data . Djurovic et al., (2015), corroborated the aforementioned claim by demonstrating the predictive capability of Artificial Neural Network (ANN) and an Adaptive Neurofuzzy Inference System (ANFIS) in forecasting groundwater level changes using well data and other hydrological conditions as inputs. Mohanty et al. (2010) used three ANN algorithms (GDX, LM, BR) and historical data from 19 sites to predict groundwater levels in a tropical humid region. One appealing feature of ANNs is their ability to develop a relation between the outputs and inputs of a process without the physics being explicitly furnished to them.

XGBoost has been particularly effective due to its gradient boosting framework, which improves prediction accuracy by combining multiple weak learners (Chen & Guestrin, 2016). Random forests, another ensemble method, are valued for their robustness and ability to handle overfitting (Breiman, 2001). Studies by (Gaffoor et al., 2020; Vu et al., 2021; Chu et al., 2022) demonstrate the successful application of these ML methods in different geographical contexts, highlighting their adaptability and effectiveness. However, the performance of machine learning models is highly dependent on the quality and quantity of the input data, as well as the feature engineering process.

This study aims to address the limitations of existing methods by exploring a data-driven and adaptable approach for predicting groundwater levels and variations in the Sokoto Basin. To achieve this, the study investigates the relationships between groundwater levels and key environmental variables such as rainfall, soil moisture, temperature and humidity

1.1 Study Area

The Sokoto Basin in the northwest of Nigeria occupies approximately 65,000 square kilometres and is situated between latitudes 10° and 14° N and longitudes 3° and 7° E. The region as seen in figure 1 comprises of Sokoto and Kebbi States that is bordered on the north and west by Niger Republic and on the east by Zamfara state. The basin is divided into three main physiographic units: the coastal lowland of the Niger and lower Rima rivers, the Sokoto lowlands of the north/centre, and the uplands or high plains of the east and southeast averaging 700 meters in height. The primary drainage system in the basin is composed of the Rima and Sokoto rivers (Wali et al., 2023), which converge at Dundaye. Due to the influx of recharge during the dry season from several tributaries, base flow, springs, and the parched water body, the Sokoto River and its tributaries remain perennial in nature.

Corresponding author's e-mail:samsonalfa@gmail.com

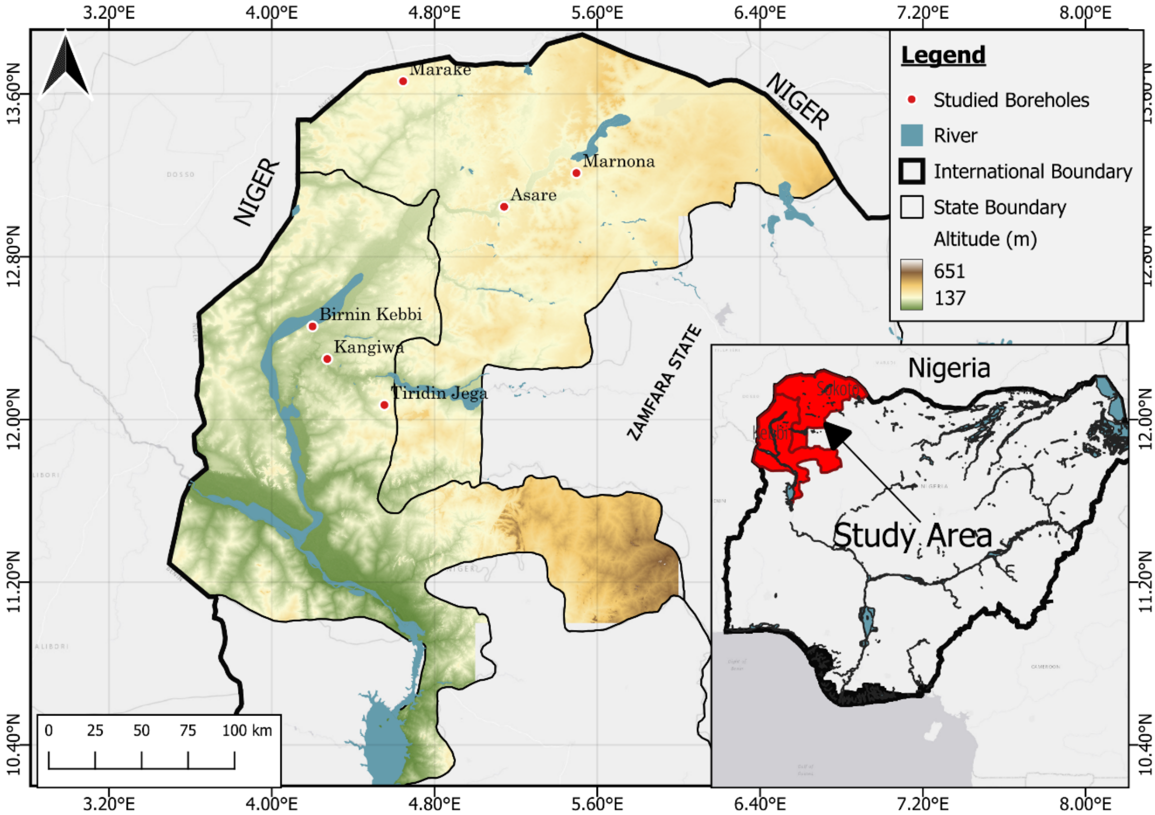

Figure 1: Location Map of Study Area

The study area is a part of northwest Nigeria's Sokoto-Rima hydrogeological region. It represents the Nigerian portion of the Illumeden sedimentary basin (Wali et al., 2023), which is centred in Niger underlain by interbedded semi consolidated gravel, sand, clay, and some limestone. According to Obaje et al. (2020) the sedimentary sequences are stratigraphically subdivided from bottom to top into several formations: the late Jurassic to early Cretaceous Illo and Gundumi Formations, the Maestrichtian Rima Group, the late Paleocene Sokoto Group and the Eocene-Miocene Gwandu Formation (Figure 2). The study area falls under moderate to limited aquifer productivity classes with common borehole yields in the range 0.1 - 10l/s (Heckmann et al., 2022). Particularly interesting is the Sokoto and Rima group falling under low to limited productive aquifers that can supply communities via motorised pumps and hand pumps (Figure 2). These are indicative of aquifers that are generally local and discontinuous, with low permeability and groundwater storage capacity.

Corresponding author's e-mail:samsonalfa@gmail.com

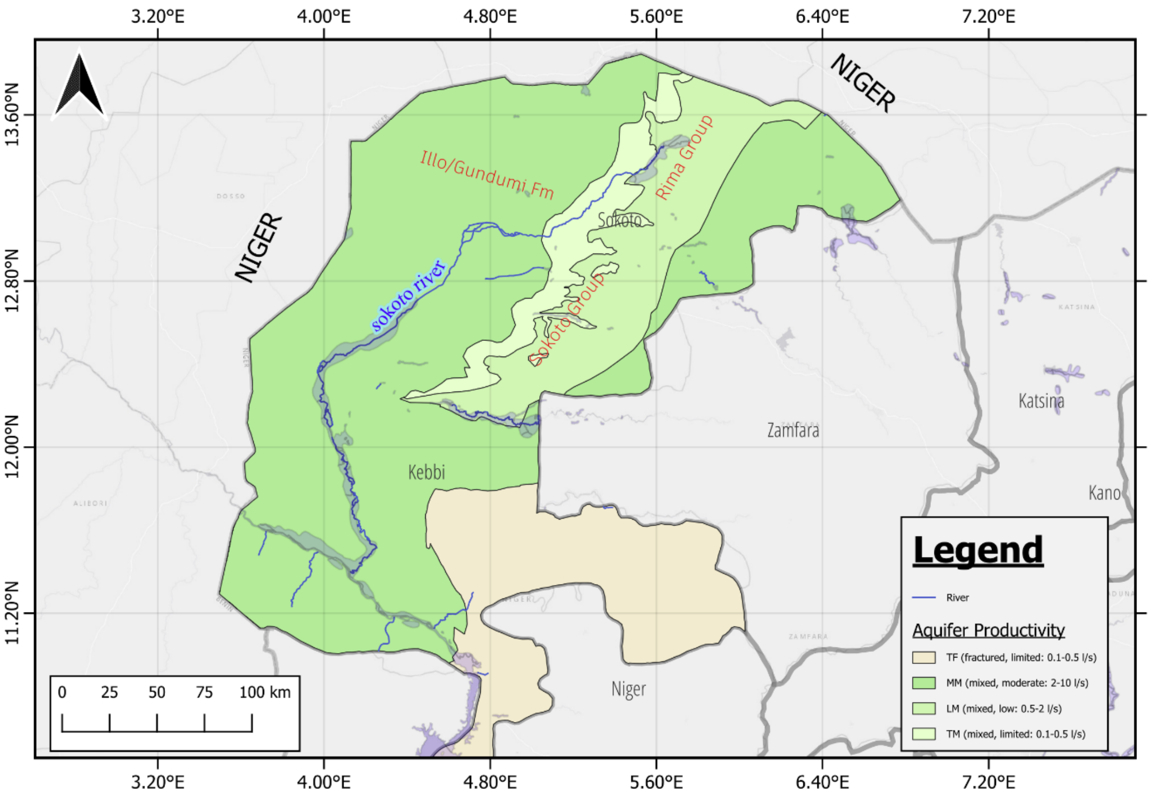

Figure 2: Hydrogeological Map showing aquifer productivity of the Sokoto Basin as depicted in "Groundwater resources in the ECOWAS region" Map *(The Federal Institute for Geosciences and Natural Resources [BGR] et al., 2022)*

2.0 MATERIALS AND METHODS

2.1 Data Acquisition

Daily groundwater level data were obtained from Nigeria Hydrological Services Agency (NIHSA) as the target variable. The groundwater level is measured in metres (m) above ground level. A total of 9 boreholes were found within the vicinity of the study area. However, not all boreholes were still active and the majority of them had significant missing data. It was therefore necessary to filter the datasets to select boreholes that would give us the best results.

At first, the research did not include boreholes with fewer than five years of data and more than 30% of missing data. This meant that six boreholes in the basin could be used for modelling. The time series of two selected boreholes is shown in Figure 3.

Corresponding author's e-mail:samsonalfa@gmail.com

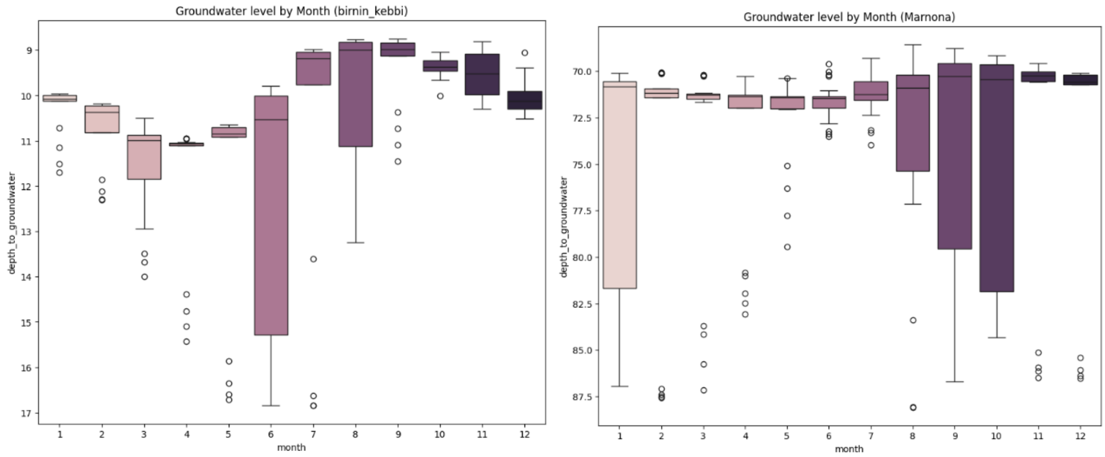

Figure 3: Groundwater fluctuation by month

Rainfall, surface temperature, air humidity, and soil moisture were among the environmental variables used in this research that were sourced from the National Aeronautics and Space Administration (NASA) Langley Research Centre (LaRC) Prediction of Worldwide Energy Resource (POWER) Project funded through the NASA Earth Science/Applied Science Program. The data was last accessed on the 12/04/2024. Rainfall pattern by month is illustrated in Figure 4 showing peak rainfall in August and least rainfall in the month of November to April.

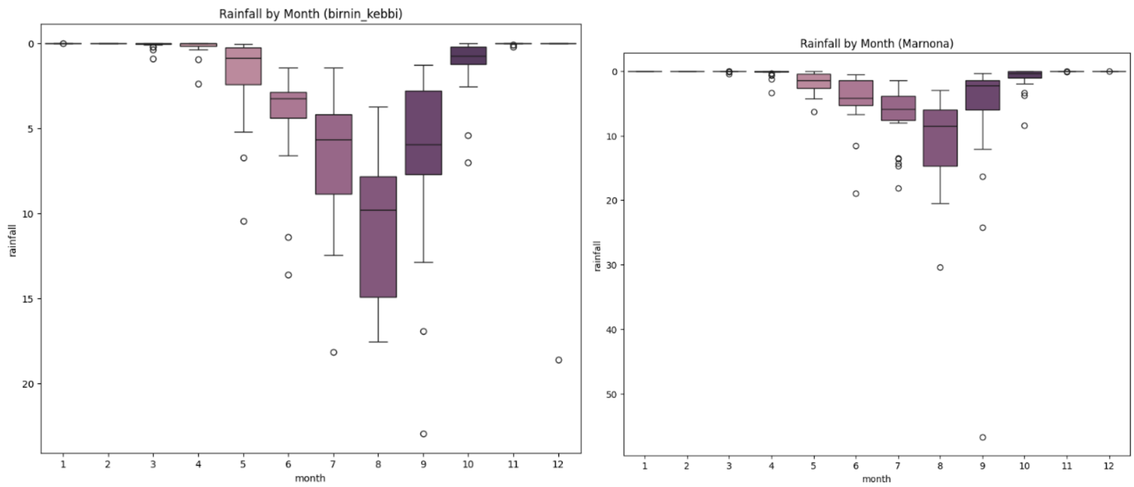

Figure 4: Rainfall by month

Corresponding author's e-mail:samsonalfa@gmail.com

2.2 Methods

For groundwater level prediction, we employed three different machine learning models namely LSTM, XGBOOST and RF models. The entire workflow for this study, including data processing, model development, and analysis, was implemented in Python using Google Colab. The learned models are used to forecast groundwater levels. evaluates the model's performance using three different indices. Until a drop in these indices is noticed after a number of consecutive iterations, the models stay in the training phase.

The tensorflow library in Python was used to model all the neural network architectures. Pandas, NumPy, Seaborn and Matplotlib were libraries imported for management, processing and visualisation of data. The workflow, which consists of four major phases with unique sub-steps for each, is shown in Figure 5.

The XGBRegressor function from the XGBoost library in Python was employed to model groundwater levels. The data was first scaled to standardize feature values, enhancing the model's performance. To fine-tune the model's hyperparameters, a parameter search was conducted using time series cross-validation (TSCV) and grid search (gsearch) algorithm, systematically exploring different hyperparameter values to select the optimal combination. With the best hyperparameters identified, the XGBRegressor model was retrained on the training dataset. The calibrated model was then used to make predictions on the test dataset, ensuring accurate groundwater level forecasts.

Employing the RandomForestRegressor module from Scikit-learn, the dataset was initially divided into training and testing datasets with an 80:20 split (Seidu et al., 2023). The variables were then scaled before the model was fitted. In order to generate a prediction, the model was first trained using training data and then applied to test data. Next, the features that were most significant to the RF model were determined by looking at the feature importance.

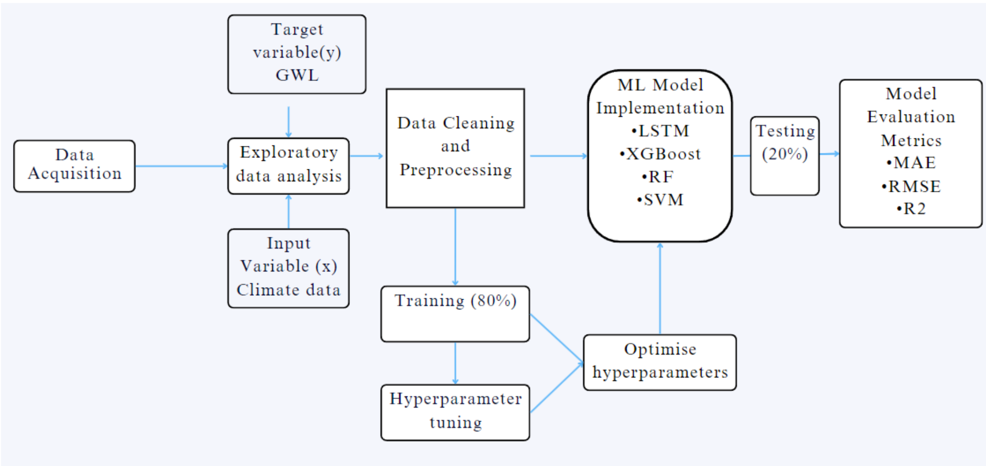

Figure 5: Study Flow chart for data selection and iteration of predictive model to be chosen.

2.2.1 Long Short-Term Memory Network (LSTM): A specialized ANN

Long Short-Term Memory – usually just called "*LSTM* are a special kind of Recurrent Neural Network (RNN) that is designed to address the issue of vanishing and exploding gradients in traditional RNNs, which can hinder their ability to learn long-term dependencies(Chu et al., 2022). The specific calculation method of a single memory unit at time *t* is as follows. The internal structure of an LSTM memory unit at time t and its correlation with the state of neighbouring time memory units. Ascertain what will happen to the information before time t. In this stage, the forget gate will be used to erase any superfluous information. The sigmoid activation function σ is used by the forget gate. The Equation 1 by (Chu et al., 2022) is used to compute the forget gate f_t;

$$f_t = \sigma\left(W_f \cdot [X_t h_{t-1}] + b_f\right) \qquad (1)$$

Where the weight matrix is represented by W_f. the input gate X_t, and the bias of the forget gate is represented by b_f

There are two steps involved in updating the unit status and deciding what to do with the new input data. First, the decision to update or disregard the new data is made by the sigmoid activation function. Second, the passed value's relevance level is determined by assigning weight to it via the tanh function. The input gate is obtained using Equation 2:

$$i_t = \sigma\left(W_f \cdot [X_{t,} h_{t-1}] + b_i\right) \qquad (2)$$

Where W_c stands for the weight matrix of the the calculation unit and bc stands for the offset term of the calculation unit. The cell state for the current input is then determined using h_{t-1} and $X_{t,}$, according to the Equation 3:

$$\widetilde{C_t} = tanh(W_c . [X_t h_{t-1}] + b_c) \qquad (3)$$

Subsequently, the cell state, C_t, is determined using C_{t-1} and f_t, as shown in Equation 4:

$$C_t = f_t . C_{t-1} + i_t . \widetilde{C_t} \qquad (4)$$

To determine the memory unit's hidden layer state at time *t*. First, as can be seen in equation 5, O_t in the forget gate selects which portion of the cell state is output by the activation function sigmoid.

$$O_t = \sigma(W_o \cdot [X_t h_{t-1}] + b_o) \qquad (5)$$

Ultimately, the formula indicates that the output of LSTM is determined by $tanh(C_t)$and the output of O_t as shown in Equation 6.

$$h_t = O_t .\ tanh(C_t) \qquad (6)$$

Corresponding author's e-mail:samsonalfa@gmail.com

2.2.2 eXtreme Gradient Boost (XGBoost)

One of the gradient boosting machine implementations is called XGBoost which is recognised as one of the finest performing algorithms applied for supervised learning (Osman et al., 2021). XGBoost is preferred by data scientists owing to fast execution speed out of core compute (Bedi et al., 2020; Nasir et al., 2022; Osman et al., 2021). To minimise training errors the individual trees are continually trained on the residual output of the preceding trees using this iterative procedure (Ali et al., 2022). The equation 7 is an example of the prediction's mathematical expression:

$$\hat{Y} = \emptyset(X) = \frac{1}{n}\sum_{k=1}^{n} f_k(X), \tag{7}$$

where $\hat{Y}$ is the forecasted value of the target, X denotes the input variable, K is a value that ranges from 1 to n, f_k is the function between input and output variables, and n is the number of trained functions by boosting trees. To train multiple functions f_k, in XGBoost, the loss function must be minimised, as outlined in equation 8.

$$\mathrm{L}(\emptyset) = \sum_i l(\hat{y}_i,\ y_i) + \sum_k \Omega(f_k), \tag{8}$$

$$\Omega(f_k) = \gamma T + \frac{1}{2} \cdot \lambda \|\omega\|^2 \tag{9}$$

Where, $\mathrm{L}(\emptyset)$ is the regularised function, i is the loss function measurement between $\hat{y}_i$ (prediction value) and y_i (actual value). where Ω (equation 9) is a regularisation factor that keeps the model from becoming less overfitting and error-prone by preventing the construction of more trees. ω is the score vector on the leaf, λ is the penalty parameter, T is the number of leaves, and γ is the complexity of the leaf.

2.2.3 Random Forest (RF)

Random Forest regression is an ensemble of decision trees DT (based on the results of multiple decision trees). Using a subsample of the dataset, each tree is trained and final value is given by averaging the whole ensemble. The final result of the RF algorithm is estimated from the result of each DT (Arabameri et al., 2021). This is achieved by using different predictive parameters in each generated tree and also by resampling data with replacement (W. Chen et al., 2019). Equation 10 may be used to determine the regression's concluding outcomes.

$$\hat{y}(x_i) = \frac{1}{B}\sum_{b=1}^{B} T_b(x_i)\ , \tag{10}$$

A RF model has two user-defined parameters: B, which denotes the number of trees in the forest, and T, which denotes the number of characteristics that are used to divide the nodes. Where, y is the target variables and x are the feature variables

2.3 Evaluation Criteria

The effectiveness of machine learning models in representing the data is assessed by evaluating their performance with various metrics (Lazzeri, 2021). The model is refined in response to feedback, iteratively, until the target accuracy is attained.

The study compared the performance of three machine learning models—LSTM

Corresponding author's e-mail:samsonalfa@gmail.com

(Long Short-Term Memory), XGBoost (Extreme Gradient Boosting), and Random Forest (RF)—for predicting groundwater levels in the Sokoto Basin. The models were evaluated based on Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the Coefficient of Determination (R^2) which are expressed in equations 11 – 13 (Shakya et al., 2022; Yi et al., 2024). The number of samples available is indicated by n, and the observed data and mean observed data are represented by O_i and $\bar{O}$ respectively. The predicted values and mean predicated value are represented by P_i and $\bar{P}$ (model output).

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|P_i - O_i| \tag{11}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=0}^{n}(P_i - O_i)^2} \tag{12}$$

$$R^2 = \frac{\left[\sum_{i=1}^{n}(O_i - \bar{O}).(P_i - \bar{P})\right]^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2 \sum_{i=0}^{n}(P_i - \bar{P})^2} \tag{13}$$

3.0 RESULTS AND DISCUSSION

3.1 Model Performance Comparison

The XGBoost algorithm outperformed the other models with the lowest MAE, RMSE and a mean R^2 score of up to 0.849 across all boreholes as shown in Table 1. It outperformed the LSTM algorithm at four of the borehole sites. Among the three algorithms tested, XGBoost took the longest to build due to the use of the grid search tool for hyperparameter optimization, and it also recorded the highest scores for four borehole sites.

The LSTM algorithm, while typically suited for time series data, performed less effectively in this context, possibly due to the complexity and noisiness of the groundwater data. The model recorded an average R^2 score of 0.765 across all boreholes. It also had the lowest MAE and RMSE for the Kangiwa and Marake sites, outperforming the XGBoost model in these areas. Among the three algorithms tested, the LSTM model required the least computational time to build.

The RF regression tree also performed well at three sites, achieving an R^2 score greater than 0.5 for the boreholes tested. However, it had the lowest performance among the three algorithms. The RF model does not rely on the importance assigned by a single decision tree but rather selects features randomly during training. Overall, the model's performance was reasonable.

Overall, XGBoost consistently outperforms other models in most locations, particularly in terms of MAE and RMSE, and achieves higher R^2 values, indicating better fit and predictive accuracy. LSTM also performs well but is generally outperformed by XGBoost.

Corresponding author's e-mail:samsonalfa@gmail.com

Table 1: Comparison of Evaluation Metrics of the Three ML Models employed in the Sokoto Basin. The top-performing model is marked in blue

LOCATION	METRIC	LSTM	XGBOOST	RF
KANGIWA	MAE	0.622	0.757	2.198
	RMSE	0.818	2.086	2.662
	R^2	0.972	0.774	0.633
MARNONA	MAE	2.311	0.747	4.277
	RMSE	3.463	2.858	5.647
	R^2	0.624	0.745	0.005
BKEBBI	MAE	0.270	0.090	0.456
	RMSE	0.309	0.188	0.690
	R^2	0.850	0.944	0.239
MARAKE	MAE	0.234	0.214	0.793
	RMSE	0.242	0.563	1.019
	R^2	0.495	-1.714	-7.895
ASARE	MAE	0.082	0.016	0.119
	RMSE	0.110	0.051	0.172
	R^2	0.811	0.961	0.547
TJEGA	MAE	0.425	0.075	0.673
	RMSE	0.586	0.254	0.898
	R^2	0.838	0.971	0.632

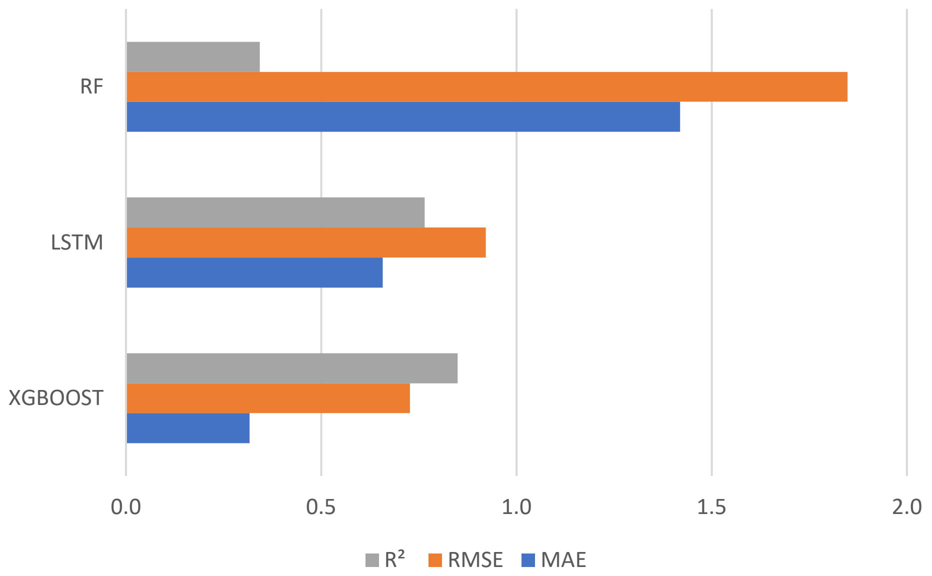

Figure6: Bar Graph Showing Model Averages

Corresponding author's e-mail:samsonalfa@gmail.com

3.2 Feature Importance

The best performing model, the XGBoost was particularly utilized to assess the importance of various input parameters. By analysing the feature importance scores generated by XGBoost, we constructed input combinations using the most influential parameters, as illustrated in Figure 7.

The XGBoost analysis revealed that soil moisture is the most critical factor in predicting groundwater levels, followed by temperature and humidity. Interestingly, rainfall was found to be the least significant parameter according to the importance scores assigned by XGBoost, suggesting that its influence on groundwater levels in the study area is minimal compared to the other factors.

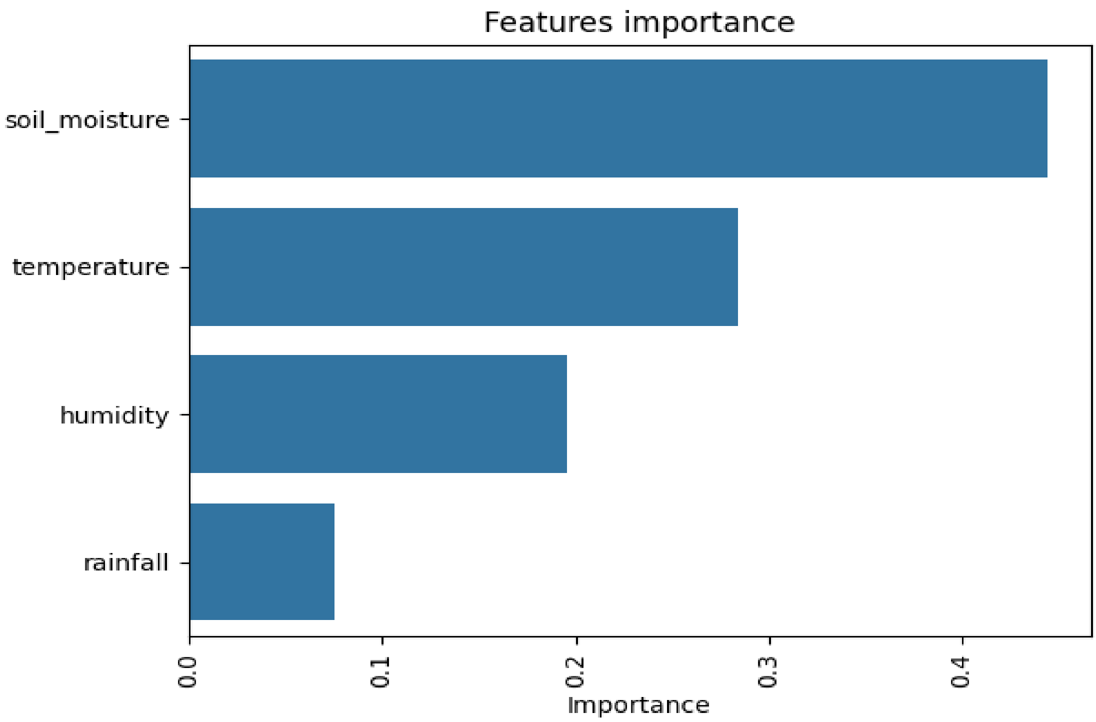

Figure 7: Analysis of Feature Importance using XGBoost Model

Finally, rainfall is the feature that recorded the lowest correlation score of 0.08. Despite the expectation that rainfall would have a higher correlation score due to its role in increasing water levels, the results did not align with this assumption. Similar study by Kanyama et al. (2020) experienced the same phenomena which prompted the use of decomposed signals of the time series which is common practice in machine learning. The significant noise in the dataset (see Figure 4) led to lower scores for the actual rainfall values, affecting the feature's ability to capture relationships effectively (Gibson, 2020)

Corresponding author's e-mail:samsonalfa@gmail.com

3.2.1 Implementation of LSTM

The LSTM model was trained using the training dataset and then applied to the test dataset for predictions. The hyperparameters for LSTM model including the learning rate, hidden unit and optimizer to achieve the best performance were set through a trial-and-error method for groundwater level forecasting across different wells.

Figure 8 expresses the comparison between the actual and predicted groundwater levels using the LSTM model for various boreholes, highlighting the model's performance. It means that LSTM model provides accurate predictions, effectively capturing the peaks and valleys of the observed groundwater levels and closely aligning with the actual data points. In the Kangiwa borehole, the LSTM model's predictions match the actual groundwater levels closely, as indicated by the MAE of 0.62 and RMSE of 0.81, resulting in an R^2 value of 0.97. This level of precision is also evident in the other boreholes asides from Marnona where it could not retrace after capturing the peaks.

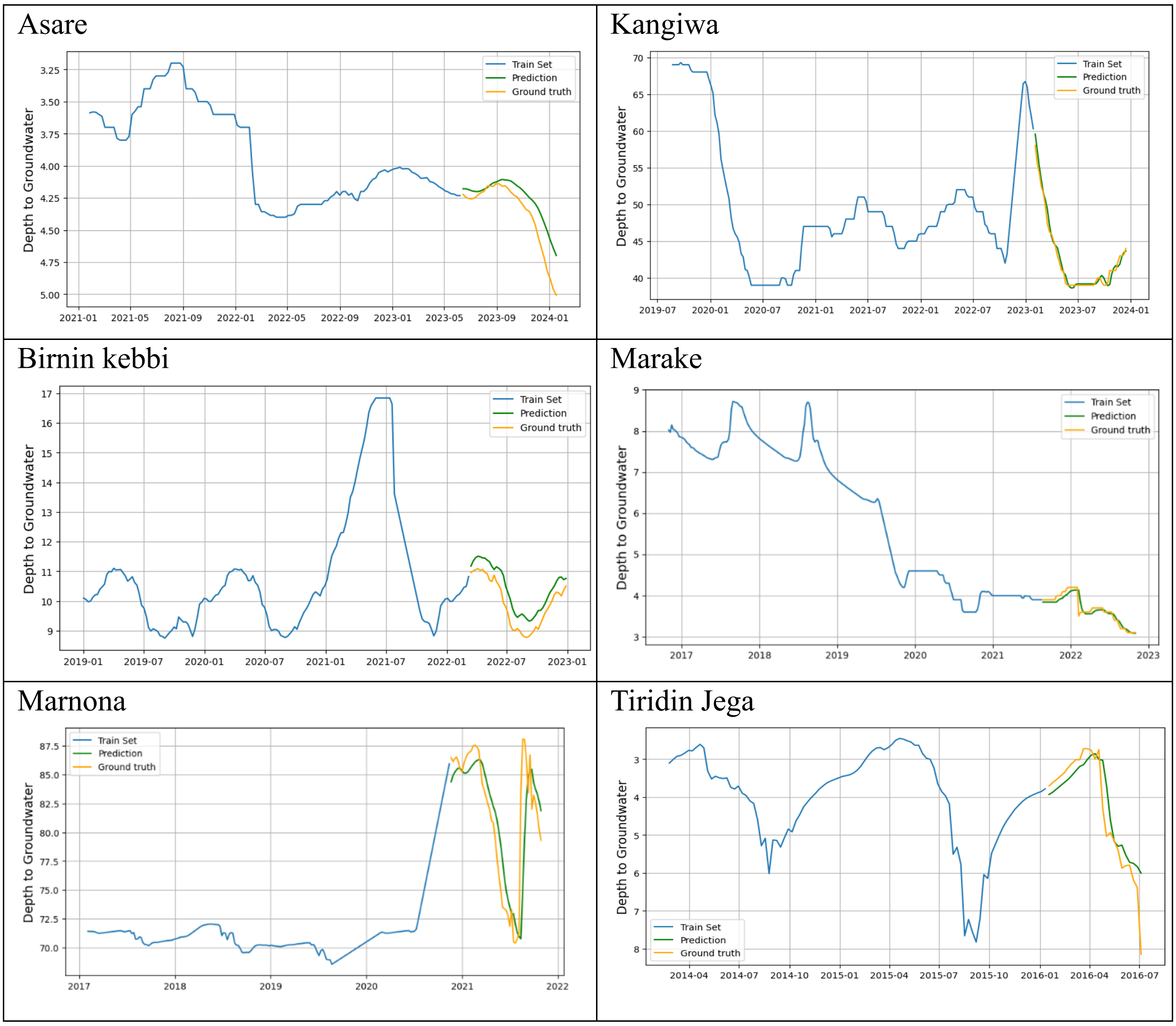

Figure 8**:** Time-series plot comparing the actual and predicted groundwater levels at selected boreholes using LSTM

Corresponding author's e-mail:samsonalfa@gmail.com

3.1.2 Implementation of XGBOOST

The XGBOOST prediction graphs exhibit random peaks due to the model's sensitivity to noise, unlike the smoother predictions produced by the LSTM model. This noise-induced variability is particularly evident in the Marnona and Marake sites (see Figure 9). Despite this, the models overall yielded strong performance, as reflected in the error indices, with MAE ranging from 0.016 to 0.757m and RMSE ranging from 0.051 to 2.859m. Figure 9 plots the ground truth values for groundwater level against the predicted values.

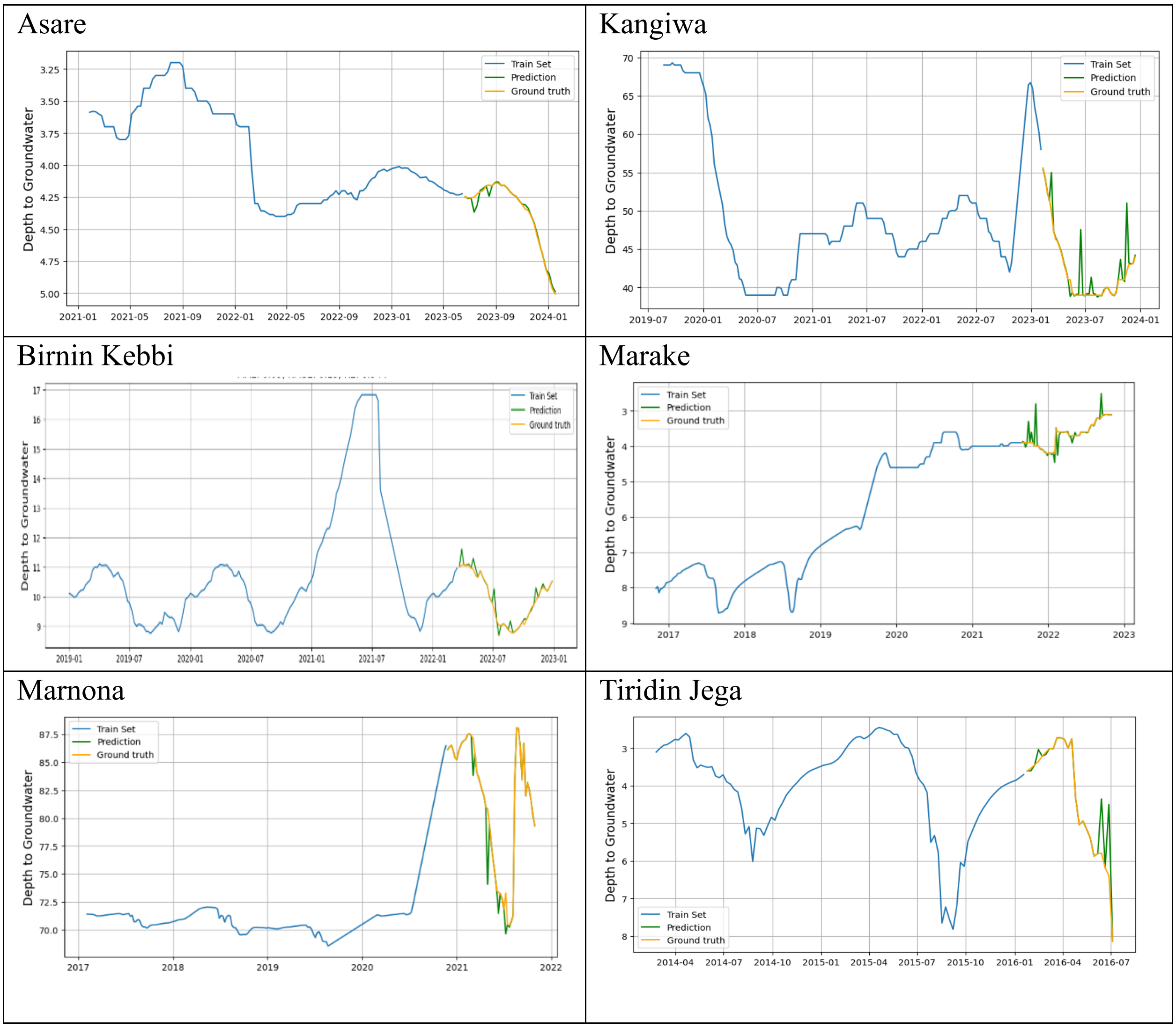

Figure 9**:** Time-series plot comparing the actual and predicted groundwater levels at selected boreholes using XGBoost

2.2.3 Implementation of Random Forest Model

The Random Forest is the last method that was implemented on our dataset and Figure 10 illustrates the performance of the Random Forest (RF) model across the six borehole locations, demonstrating its effectiveness in predicting groundwater levels. As seen in Figure 4.4, the model predictions exhibited noticeable

Corresponding author's e-mail:samsonalfa@gmail.com

variations in their trends, failing to capture the underlying gradual pattern in groundwater levels with the utmost precision which can be attributed to hyperparameter tuning. This was particularly evident in the case of the Marake borehole, where the R^2 value of -7.895 presented a significant outlier. Nevertheless, the overall performance of the models was deemed to be satisfactory. This assertion is corroborated by the error metrics, which indicate that MAE values ranged between 0.016 and 0.757m, while RMSE values fell between 0.051 and 2.859m. The models were able to estimate groundwater levels with an acceptable degree of accuracy, despite the inherent complexities of the system.

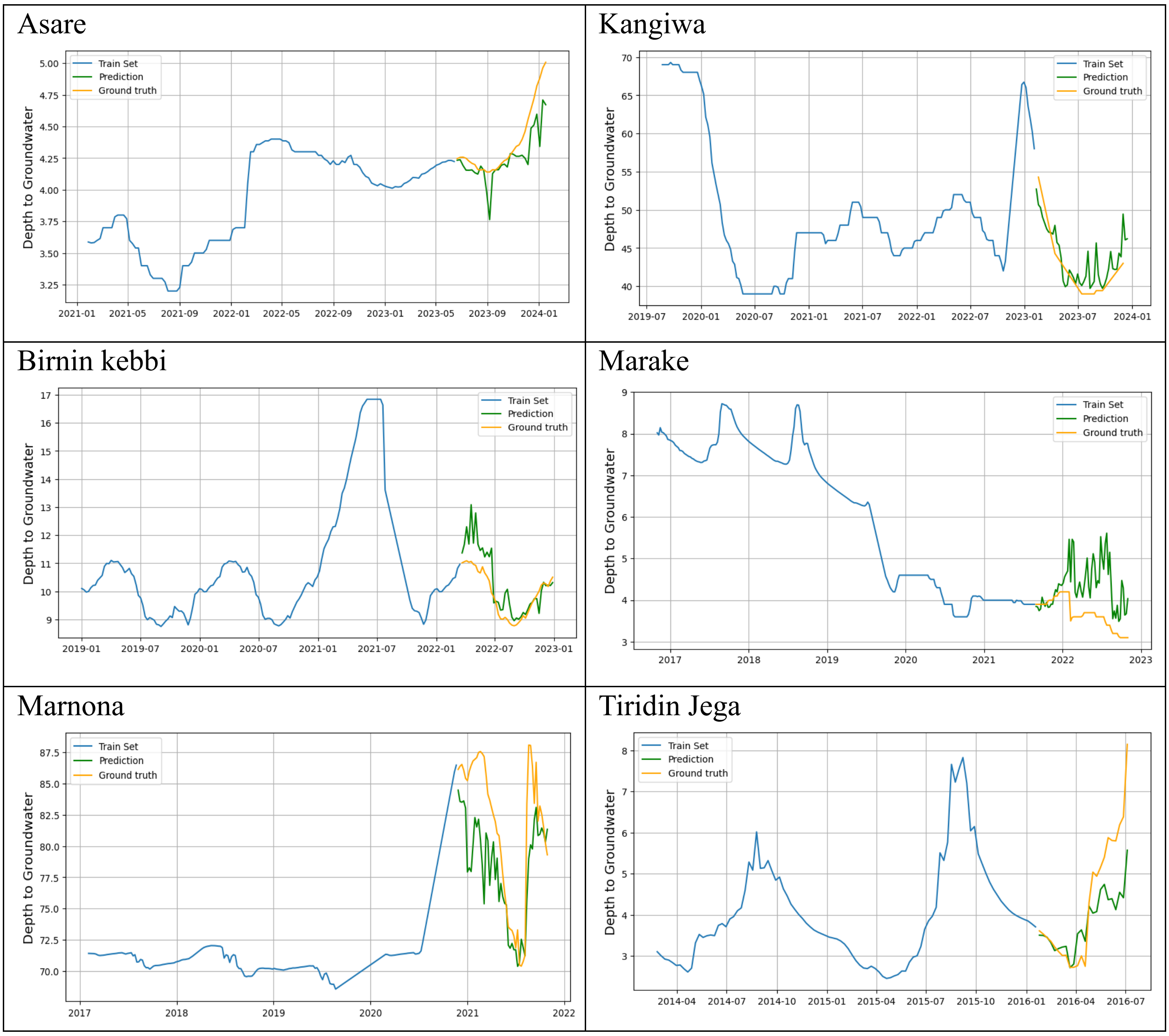

Figure 10**:** Time-series plot comparing the actual and predicted groundwater levels at selected boreholes using RF

Corresponding author's e-mail:samsonalfa@gmail.com

4.0 CONCLUSION

This study evaluated the application of machine learning methods for predicting groundwater levels in the Sokoto Basin. A comparative analysis of different machine learning models demonstrated that XGBoost effectively predicts groundwater levels using inputs such as precipitation, rainfall, temperature, soil moisture, and humidity. The LSTM model also showed strong performance, particularly in capturing the peaks and valleys of the groundwater level time series, with MAE and RMSE values ranging from 0.016 – 0.757m and 0.051 – 2.859m, respectively. The RF model exhibited reliable performance across most locations.

References

Akintayo, T., Bello, S. A., Ugwu, S. J., & Mwali, C. N. (2022). Determination Of Groundwater Potential Areas in Kaduna Metropolis, Kaduna, NIGERIA. *FUDMA Journal of Sciences*, *5*(4), 194–204. https://doi.org/10.33003/fjs-2021-0504-806

Ali, Z. H., Burhan, A. M., Kassim, M., & Al-Khafaji, Z. (2022). Developing an Integrative Data Intelligence Model for Construction Cost Estimation. *Complexity*, *2022*, 1–18. https://doi.org/10.1155/2022/4285328

Almahawis, M. (2018). *Assessing Groundwater Storage and Groundwater Level Fluctuations in The Area of Fort Collins, Colorado.*

Arabameri, A., Pal, S. C., Rezaie, F., Nalivan, O. A., Chowdhuri, I., Saha, A., Lee, S., & Moayedi, H. (2021). Modeling groundwater potential using novel GIS-based machine-learning ensemble techniques. *Journal of Hydrology: Regional Studies*, *36*, 100848. https://doi.org/10.1016/j.ejrh.2021.100848

Bedi, S., Samal, A., Ray, C., & Snow, D. (2020). Comparative evaluation of machine learning models for groundwater quality assessment. *Environmental Monitoring and Assessment*, *192*(12), 776. https://doi.org/10.1007/s10661-020-08695-3

BGR, BGS, EAGWM, & UNESCO. (2022). *Ressources en eaux souterraines dans la région de la CEDEAO – Groundwater resources in the ECOWAS region. 1: 5 000 000* [Map]. doi: 10.25928/GWR-ECOWAS.1

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 785–794.

Chen, W., Tsangaratos, P., Ilia, I., Duan, Z., & Chen, X. (2019). Groundwater spring potential mapping using population-based evolutionary algorithms and data mining methods. *Science of The Total Environment*, *684*, 31–49. https://doi.org/10.1016/j.scitotenv.2019.05.312

Chu, H., Bian, J., Lang, Q., Sun, X., & Wang, Z. (2022). Daily Groundwater Level Prediction and Uncertainty Using LSTM Coupled with PMI and Bootstrap Incorporating Teleconnection Patterns Information. *Sustainability*, *14*(18), 11598. https://doi.org/10.3390/su141811598

Deka, P. C. (2019). *A primer on machine learning applications in civil engineering*. CRC Press.

Djurovic, N., Domazet, M., Stricevic, R., Pocuca, V., Spalevic, V., Pivic, R., Gregoric, E., & Domazet, U. (2015). Comparison of Groundwater Level Models Based on Artificial Neural Networks and ANFIS. *The Scientific World Journal*, *2015*, 1–13. https://doi.org/10.1155/2015/742138

Gaffoor, Z., Pietersen, K., Jovanovic, N., Bagula, A., & Kanyerere, T. (2020). Big Data Analytics and Its Role to Support Groundwater Management in the Southern African Development Community. *Water*, *12*(10), 2796. https://doi.org/10.3390/w12102796

Gibson, K. (2020). *The application of machine learning for groundwater level prediction in the Steenkoppies compartment of the Gauteng and North West dolomite aquifer, South Africa* [PhD Thesis]. University of the Free State.

Heckmann, M., Brugeron, A., O Dochartaigh, B., & Lewis, M. (2022). *Groundwater resources in the ECOWAS region: Expected aquifer productivity*.

Huang, X., Gao, L., Crosbie, R. S., Zhang, N., Fu, G., & Doble, R. (2019). Groundwater Recharge Prediction Using Linear Regression, Multi-Layer Perception Network, and Deep Learning. *Water*, *11*(9), 1879. https://doi.org/10.3390/w11091879

Iqbal, N., Khan, A.-N., Rizwan, A., Ahmad, R., Kim, B. W., Kim, K., & Kim, D.-H. (2021). Groundwater Level Prediction Model Using Correlation and Difference Mechanisms Based on Boreholes Data for Sustainable Hydraulic Resource Management. *IEEE Access*, *9*, 96092–96113. https://doi.org/10.1109/ACCESS.2021.3094735

Kanyama, Y., Ajoodha, R., Seyler, H., Makondo, N., & Tutu, H. (2020). Application of Machine Learning Techniques In Forecasting Groundwater Levels in the Grootfontein Aquifer. *2020 2nd International Multidisciplinary Information Technology and Engineering Conference (IMITEC)*, 1–8. https://doi.org/10.1109/IMITEC50163.2020.9334142

Lazzeri, F. (2021). *Machine Learning for Time Series Forecasting with Python®*. John Wiley & Sons.

Mohanty, S., Jha, M. K., Kumar, A., & Sudheer, K. P. (2010). Artificial Neural Network Modeling for Groundwater Level Forecasting in a River Island of Eastern India. *Water Resources Management*, *24*(9), 1845–1865.

Corresponding author's e-mail:samsonalfa@gmail.com

https://doi.org/10.1007/s11269-009-9527-x

Nasir, N., Kansal, A., Alshaltone, O., Barneih, F., Sameer, M., Shanableh, A., & Al-Shamma'a, A. (2022). Water quality classification using machine learning algorithms. *Journal of Water Process Engineering*, *48*, 102920. https://doi.org/10.1016/j.jwpe.2022.102920

Obaje, N., Faruq, U., Bomai, A., Moses, S., Ali, M., Adamu, S., Essien, A., Lamorde, U., Umar, U., Ozoji, T., Okonkwo, T., Lukman, A., Idris Nda, A., Faruq, Z., & Umar, M. (2020). A Short Note on the Petroleum Potential of the Sokoto Basin in North-western Nigeria. *Journal of Petroleum Science and Engineering*, *4(1)*, 34–38. https://doi.org/10.11648/j.pse.20200401.14

Osman, A. I. 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. *Ain Shams Engineering Journal*, *12*(2), 1545–1556. https://doi.org/10.1016/j.asej.2020.11.011

Seidu, J., Ewusi, A., Kuma, J. S. Y., Ziggah, Y. Y., & Voigt, H.-J. (2023). Impact of data partitioning in groundwater level prediction using artificial neural network for multiple wells. *International Journal of River Basin Management*, *21*(4), 639–650. https://doi.org/10.1080/15715124.2022.2079653

Shakya, C. M., Bhattacharjya, R. K., & Dadhich, S. (2022). Groundwater level prediction with machine learning for the Vidisha district, a semi-arid region of Central India. *Groundwater for Sustainable Development*, *19*, 100825.

Tao, H., Hameed, M. M., Marhoon, H. A., Zounemat-Kermani, M., Heddam, S., Kim, S., Sulaiman, S. O., Tan, M. L., Sa'adi, Z., Mehr, A. D., Allawi, M. F., Abba, S. I., Zain, J. M., Falah, M. W., Jamei, M., Bokde, N. D., Bayatvarkeshi, M., Al-Mukhtar, M., Bhagat, S. K., … Yaseen, Z. M. (2022). Groundwater level prediction using machine learning models: A comprehensive review. *Neurocomputing*, *489*, 271–308. https://doi.org/10.1016/j.neucom.2022.03.014

Vu, M. T., Jardani, A., Massei, N., & Fournier, M. (2021). Reconstruction of missing groundwater level data by using Long Short-Term Memory (LSTM) deep neural network. *Journal of Hydrology*, *597*, 125776. https://doi.org/10.1016/j.jhydrol.2020.125776

Wali, S. U., Alias, N., Harun, S. B., Mohammed, I. U., Garba, M. L., Atiku, M., Gada, M. A., & Hamisu, I. (2023). Hydrogeological Analysis of Cretaceous and Tertiary Aquifers in Semiarid Sokoto Basin, Northwestern Nigeria: Implications for Sustainable Groundwater Development. *Global Journal of Earth Science and Engineering*, *10*, 27–55.

Corresponding author's e-mail:samsonalfa@gmail.com

https://doi.org/10.15377/2409-5710.2023.10.3

Yi, S., Kondolf, G. M., Sandoval Solis, S., & Dale, L. (2024). Groundwater Level Forecasting Using Machine Learning: A Case Study of the Baekje Weir in Four Major Rivers Project, South Korea. *Water Resources Research*, *60*(5), e2022WR032779. https://doi.org/10.1029/2022WR032779

Zaresefat, M., & Derakhshani, R. (2023). Revolutionizing Groundwater Management with Hybrid AI Models: A Practical Review. *Water*, *15*(9), 1750. https://doi.org/10.3390/w15091750

Corresponding author's e-mail:samsonalfa@gmail.com