گـۆڤـاری زانـکـۆی ڕاپــەڕیــن
Journal of University of Raparin.
مجلة جامعة رابرين
E-ISSN: 2522 – 7130 P-ISSN: 2410 – 1036

DOI: https://doi.org/10.26750/3867r447
Date Received: 10/08/2025
Date Accepted: 16/11/2025
Date Published: 20/06/2026

Data-Driven Analysis of Pond-Based Recharge in Daratoo and Tarin Pilot Sites in Erbil City Using Internet of Things and Machine Learning

Khalat Khalid Muhammed[1], Shuokr Qarani Aziz[2]

Khalatkhalid1@gmail.com shuokr.aziz@su.edu.krd

[1]Department of Civil Engineering, College of Engineering, Salahaddin University-Erbil, Erbil, Kurdistan Region, Iraq.

[2]Directorate of Irrigation Erbil, General Directorate of Water Resources, Ministry of Agriculture and Water Resources, Erbil, Kurdistan Region, Iraq.

Abstract

Groundwater depletion is a pressing issue in semi-arid regions. In Erbil's environs, over-extraction and reduced precipitation have driven local water-table declines. This study evaluates the effectiveness of pond-based managed aquifer recharge (MAR) at two pilot sites, Tarin and Daratoo, by integrating high-frequency Internet of Things (IoT) monitoring with machine learning (ML). Diver-HUB sensors recorded water pressure, temperature, electrical conductivity, and salinity every 15 minutes in recharge and monitoring wells. Pressure readings were barometrically corrected and used for time-series prediction of groundwater depth. Four modelling approaches (linear regression, neural networks, extreme Gradient Boosting (XGBoost), and gradient boosting (GBoost) were compared, and a random forest feature-importance analysis identified water pressure as the dominant predictor. During the wet season, observed water-table rises were 1.29 m at Tarin and 2.4 m at Daratoo, indicating successful short-term recharge. Linear regression produced the best predictive performance on test data (near-unity R^2 and very low RMSE), but these results are tempered by site specificity, the limited number of pilot sites, and the short observation window. The contribution of this work is a data-driven, deployable framework that combines IoT and ML to support real-time monitoring, inform MAR design and operation, and guide adaptive groundwater management. Broader validation across more sites and longer periods is recommended to establish transferability and long-term sustainability.

Keywords: Groundwater depth prediction, IoT, linear regression, MAR, ML

ئەنجامدانی شیکردنەوە بە پشتبەستن بە داتاو زانیارییەکان بۆ دوبارە تێرکردنەوەی چینەکان ئاوی ژێر زەوی لە ڕێگەی پۆندەکانەوە (گۆمەکان) لە شاری هەولێر بە بەکارهێنانی ئینترنێت و فێربونی ئامێری

خەڵات خالد محمد[1] ، شکر قرنی عزیز[2]

[1]بەشی ئەندازیاری شارستانی، کۆلێژی ئەندازیاری، زانکۆی سەلاحەدین-هەولێر، هەولێر، هەرێمی کوردستان، عێراق.

[2]بەڕێوەبەرایەتی ئاودێری هەولێر، بەڕێوەبەرایەتی گشتی سەرچاوەکانی ئاو، وەزارەتی کشتوکاڵ و سەرچاوەکانی ئاو، هەولێر، هەرێمی کوردستان، عێراق.

پۆختە

بەکارهێنانی ئاوی ژێر زەوی لە رادە بەدەر هۆکاری دڵەڕاوکێ ی بەهێزە لە ناوچە نیمچە وشکاییەکان، وەکو شاری هەولێر، لەبەر زۆر بەکارهێنان و قۆرخ کردنی ئاوی ژێر زەوی و کەم بارینی باران وایانکردوە ئاستی ئاوی ژێر زەوی دابەزینێکی زۆر بەخۆوە ببینێت. لەم توێژینەوەدا هەڵسەنگاندن کراوە بۆ دوبارە تێرکردنی ئاوی ژێر زەوی ئەویش بە تێرکردنی ئەو چینانەی کە ی ژێر زەوی لە خۆدەگرن بە بەکارهێنانی کە لە سەر پۆنددەکان (گۆمەکان) دامەزراون (MAR) بە بەکارهێنانی تەکنیکی چاودێری کردنی ئینتەرنێتی کەل و پەلەکان و(IOT) و فێربونی ئامێری (ML) لە دو شوێنی تاقیکردنەوەی تارین و دارەتو.

داتاو زانیاری دەربارەی پاڵەپەستۆی ئاو, و پلەی گەرمای، و گەیاندنی تەزوی کارەبایی (EC) , وسوێری بە رادەی بەرز کۆککرایەوە (بە جیاوازی کاتی بۆ ماوەی هەر 15 خولەکێك) بە بەکارهێنانی ئامێری هەستیاری (Diver-HUB) کەلە ناو ئەو بیری ئاوی چاودێری کراو بۆ دوبارە تێرکردنەوە جێگیرکراون. هەروەها چوار خوارزمی فێربونی ئامێری بەکارهێنران بۆ: (linear regression, neural network, Extreme Gradient Boosting (XG Boost), and gradient boosting),شیکردنەوەی زنجیرەی کاتەکان بۆ پێشبینی کردنی قوڵایی ئاوی ژێر زەوی ، وە لە کاتی راستکردنەوەی خوێندنەوەکانی بارۆمێتری پاڵەپەستۆ، ئەنجامەکان ئاماژە بەوە دەکەن کە پۆندەکان (گۆمەکان) ی کە هەڵدەستن بە دوبارە تێرکردنی ئاستی ژێر زەوی، ئاستی ئاوی ژێر زەوی وەرزی باران بارینیدا لە تارین بۆ نزیکەی (1,29) مەتر بەرز کردۆتەوە،وە ئاوی ژێر زەوی لە دارەتو بۆ نزیکەی (2،4) مەتر بەرزبۆتەوە، بەمەش بە کۆگاکردنی ئاو لە کۆگاکانی کۆکردنەوەی ئاوی ژێر زەوی سەرکەوتوبو. (Simple linear regression) باشتر بو لە کۆمەڵەکانی تری کە زۆر ئاڵۆز بون،کە توانای دەستنیشان کردنی نزیکەی یەك بو (R2~1.00). وهەڵەی رەگی دوجای مامناوەندی نزیك لە ئاستی روبەری زەوی واتە

(<0.001). هەروەها شیکردنەوەی گرنگی سیفەتەکانی دارستانە هەرەمەکیەکان راستی ئەوەی سەلماند کەوا پەڵەپەستۆی ئاو یەکێکە لە سیفەتە هەرە گرنگەگان، بەمەش توانای رونکردنەوەی نمونەکە زیاتر کرد, ئەم ئەنجامانە ئاماژە بەوە دەکەن کە خوارزمیە ساکارەکان تا رێژەیەك لە توانایاندا هەیە پێشبینی یەك پێگە بکەن لە کاتێکدا پەیوەندي نێوان پاڵەپەستۆ و قوڵایی نیمچە هێڵیە (ئاسۆییە). کۆکردنەوەی نێوان تۆڕەکانی هەستیاری ئینتەرنێتی کەل و پەلەکان (IOT) و فێربونی (ML) سیتەمێکی بەڵێندەر دابین دەکات کە توانای گەشەپێدان و پەرەسەندنی بەڕێوەبردنی ئاوی ژێر زەوی هەبێت، لە کاتی راستەقینەی کرداریدا، نەرم و نیان و توانای خۆ گونجاندنی هەیە، پشت بە داتاو زانیاریەکان دەبەستێت، هەروەها بە بەردەوامی کۆنترۆڵی لافاوەکان دەکات لە ناوچە ئاوەدانکراوەکان و نیمچە وشکایي یەکان.

کلیله وشەکان: پێشبینی کردنی قوڵایی ئاوی ژێر زەوی، ئینتەرنێتی کەل و پەلەکان، فێربونی ئامێری MAR، پێچی هێڵکاری، تێرکردنی ئاوی ژێر زەوی.

1. Introduction

Groundwater in the Erbil Basin and the Kurdistan Region of Iraq and surrounding areas is being rapidly depleted. Recent studies report dramatic water-table declines on the order of tens of meters in the past two decades driven by rising urban and agricultural demand, along with unregulated pumping. For example, Mustafa and Mawlood (2024) estimate declines of 32–55 m (average ~46 m) between 2004 and 2022 across Erbil's sub-basins. They attribute this to intensive abstraction, noting that continuous depletion of groundwater levels has been recorded due to uncontrolled exploitation from both legal and illegal wells. Population growth, extended dry periods, and poor management further compound the problem. Hydrologically and environmentally, the aquifer loss is severe. Lower recharge has driven down spring and base flows and expanded arid zones. Mustafa and Mawlood (2023) note that reduced recharge causes groundwater levels to drop, which in turn leads to producing reduced streamflow, land subsidence, and heightened drought vulnerability in Erbil city. At the same time, warming has concentrated salts, while rising temperatures have lowered groundwater levels and impacted EC in the aquifer, reflecting declining water quality.

Several regional investigations have begun to assess recharge and vulnerability. For instance, a GIS-based fuzzy AHP analysis found that only 30% of the Erbil Basin area has moderate to very high natural recharge potential (Hamad, 2022). Groundwater flow models quantify recharge and leakage. For example, one study found river aquifer leakage of ~33,432 m³/day but warned that under lower recharge (e.g., climate change) the aquifer’s dry area will expand (Mustafa and Mawlood, 2024). Using MCDA, Al-Kakey et al. (2022) even identified two specific dam and pond sites on Erbil’s eastern uplands as promising managed-recharge structures. These findings underline the need for targeted, data-driven managed aquifer recharge. In arid Iraq, artificial recharge via infiltration ponds or injection wells is widely advocated to arrest groundwater decline (Al-Kakey et al.,2022). Critically, experts urge site-specific monitoring. According to Mustafa and Mawlood (2024) deploying distributed sensor networks in Erbil’s wells to enable real-time water-level tracking. However, no published study has yet combined such IoT-based monitoring with machine learning analysis to optimize managed aquifer recharge design and evaluate system performance. This represents a clear research gap, as integrating continuous field data and AI-driven models could significantly improve MAR evaluation and ensure sustainable aquifer recovery in the Kurdistan Region, including Erbil.

Constructed ponds within natural recharge zones offer a nature-based approach to replenishing depleted aquifers. By capturing stormwater and runoff in engineered basins, infiltration ponds increase percolation into underlying aquifers, resulting in measurable increases in groundwater levels (Sufyan et al. 2024). Beyond local efforts, Managed Aquifer Recharge using ponds and basins has seen success internationally. In Colorado, irrigated stream infiltration basins generated seasonal water table mounding up to 2.5 m (Deng & Bailey, 2022). In Iraq’s Dibdibba aquifer, treated wastewater recharge ponds achieved sustained annual rises of 0.2 m (Hassan et al., 2023). In Spain, riverbed infiltration galleries and farm pond networks have offset up to 30% of extraction rates, while in Australia, infiltration basins integrated with aquifer storage and recovery (ASR) systems have demonstrated recovery efficiencies above 85% (Dillon et al., 2019).

Recent advances in Internet of Things (IoT) technology enable dense, real-time monitoring of groundwater and pond performance. Low-cost pressure transducers coupled with cellular or LoRa telemetry allow automated water-level and temperature data collection at 15-minute intervals (Calderwood et al., 2020;

Espinoza-Ortiz et al., 2023). Reviews highlight that IoT-based systems improve the detection of short-term recharge events and can drive adaptive management through remote dashboards and alerts (Aderemi et al., 2021). According to Aziz and Muhammed (2025), AI and machine learning (ML) with smart infrastructure like permeable pavements and green roofs can greatly enhance urban stormwater management. Their research demonstrates how AI may be used to forecast water quality, runoff, and sustainable planning.

Ameen et al. (2025) evaluate traditional and AI-powered methods for calculating IDF curves from rainfall records in Koya City, Iraq, from 2005 to 2022. They evaluate an LSTM-based RNN, linear regression, SVR, and the Gumbel distribution using data from 2005–2015 for calibration and 2016–2022 for testing. With RMSEs of 1.44 mm/hr, MAE of 0.81 mm/hr, and R^2 of 0.99 compared to RMSEs of 9.13, 10.76, and 6.19 mm/hr, the RNN-LSTM performs noticeably better than the others. Given Koya City's rugged terrain, projections for 2043 point to an increase in the frequency and intensity of short-duration storms, highlighting the necessity for adaptive stormwater infrastructure. Complementing these sensing innovations, ML and deep learning (DL) algorithms have been increasingly applied to hydrological time-series forecasting. Convolutional neural networks (CNNs) and long short-term memory (LSTM) networks have achieved RMSEs as low as 0.056 m in large aquifer systems (Feng et al., 2024), while random forests and XGBoost have reconstructed groundwater trends across spatially distributed wells with high accuracy (Tran et al., 2025; Zaresefat & Derakhshani, 2023). Nevertheless, systematic evaluations of ML/DL approaches on IoT-derived data in localized MAR settings remain rare. This study presents the first systematic evaluation of high-frequency IoT monitoring integrated with machine learning models in the Erbil Governorate to quantify and predict groundwater depth responses to pond-based managed aquifer recharge (MAR) at two pilot sites, Tarin and Daratoo. To compare prediction accuracy, time-series sensor data were analyzed using linear regression, neural network, XGBoost, and gradient boosting models. However, the study has some limitations, such as limited internet connectivity at the sites, which occasionally affected real-time data transmission and potentially led to gaps in the dataset. Additionally, the study was conducted at only two pilot sites, which may limit the generalizability of the findings to areas with different hydrogeological conditions.

2. Materials and Methods

2.1. Site Descriptions

The investigation was conducted at two pilot MAR projects located northeast of Erbil City, first pilot was Tarin Deep Recharge Well Project (UTM 38N: Easting 423,504 m, Northing 4,014,537 m; approximately 15 km northeast of Erbil; elevation 705 m a.s.l.) and the Daratoo Shallow Recharge Well Project (UTM 38N: Easting 418,813 m, Northing 3,995,536 m; approximately 12 km northeast of Erbil; elevation 490 m a.s.l.). The sites were selected for their contrasting hydrogeologic settings: Tarin is located at the foothills of Tarin Mountain, with concentrated hillslope runoff, whereas Daratoo lies on a shallow alluvial plain with broad infiltration potential. They were also selected for their logistical suitability for instrumentation and controlled ponding. At Tarin, the project comprises two deep recharge wells (each completed to 150 m) and an adjacent monitoring well (150 m). Well coordinates and completion data are summarized in Table 1.

Table 1: Details and locations of the monitoring and shallow recharge wells at the Tarin pilot project site (Directorate of Irrigation, 2024)

No	Name	X	Y	Depth (m)
1	Deep Recharge Well 1	423669.37	4014542.81	150
2	Deep Recharge Well 2	423578.4	4014611.1	150
3	monitoring well	423342.01	4014599.93	150

Borehole lithologic logs (Figure 1) record alternating sand, mixed sand and gravel, and clay and silt horizons. In particular, coarser sand and gravel units at intermediate depths form the principal permeable pathways for vertical percolation. The Tarin infiltration pond has a measured storage capacity of 131,546 m³ and is sited to intercept and retain hillslope runoff for controlled infiltration into the underlying fractured alluvial aquifer. In addition, site photographs and satellite imagery of the Tarin pond and well installations are presented in Figure 2.

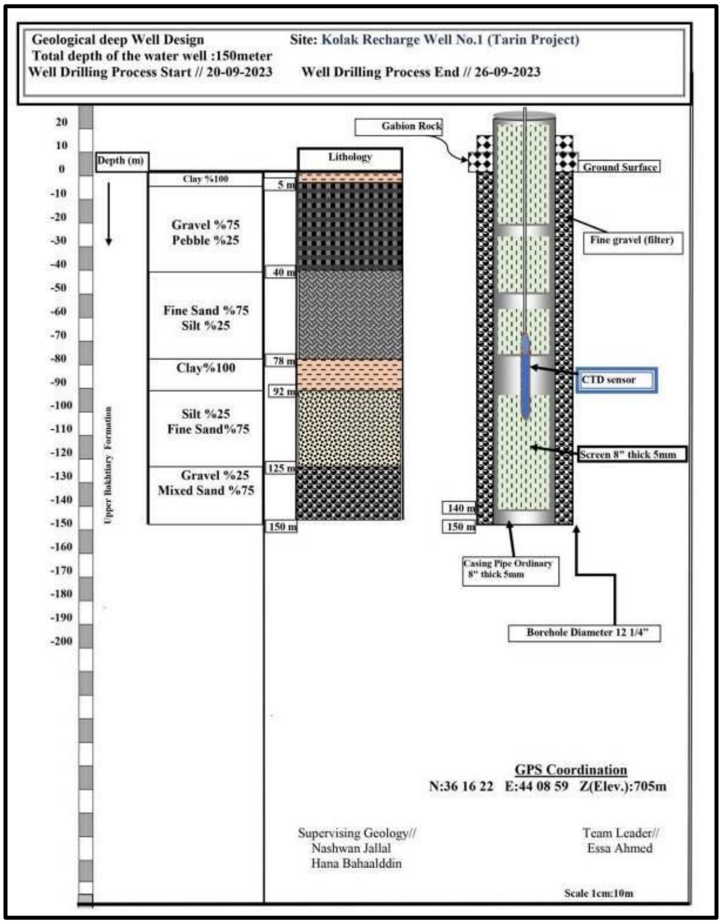


Figure 1: Borehole lithologic log and well-completion diagram for the Tarin pilot site (Directorate of Irrigation, 2024)

Figure 2: Two recharge wells and one monitoring well at the Tarin pilot site (Captured on Jun 22, 2025)

The Daratoo scheme consists of three shallow recharge wells, each with a depth of 12 m and a diameter of 1 m, positioned along the pond embankments, as well as one downstream monitoring well completed to 150 m (Table 2). Lithologic logs for Daratoo (Figure 3) indicate dominantly mixed sand and silt deposits with intercalated gravel horizons that favor rapid infiltration into shallow aquifer layers. However, local fine-grained lenses may moderate percolation. The Daratoo pond storage capacity at full supply level is 300,000 m³. In addition, site photographs and satellite imagery of the Daratoo pond and well installations are presented in Figure 4.

Table 2: Details and locations of the monitoring and shallow recharge wells at the Daratoo pilot project site (Directorate of Irrigation, 2024)

No	Name	X	Y	Depth (m)
1	Shallow Recharge Well 1	41873.49	3995539.04	12
2	Shallow Recharge Well 2	418835.7	3995480.1	12
3	Shallow Recharge Well 3	418813	3995448	12
4	Monitoring well	418427.67	3995730.1	12

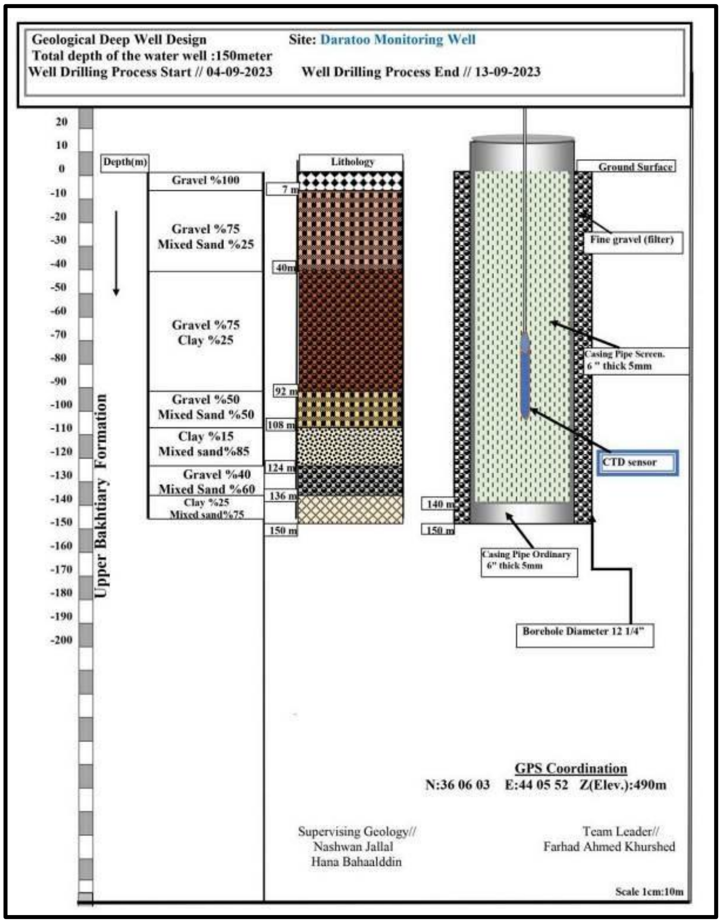


Figure 3: Lithologic log and well completion sketch for the Daratoo shallow recharge site, Directorate of Irrigation (2024)

Figure 4: Three recharge shallow wells and one monitoring well at the Daratoo pilot project site (Captured on April 17, 2025)

All wells were constructed using conventional hydrogeological practice, including steel casing, gravel pack, and screened intervals sized to intercept permeable lithologies. Each monitoring well houses a Diver HUB sensor string that records water pressure, which is barometrically corrected to derive groundwater depth, as well as water temperature, electrical conductivity, and salinity at 15-minute intervals. Sensor metadata, including model, serial number, calibration date, and installation depth, are provided in Supplementary Table S1. Data quality control included barometric correction, removal of spurious outliers, and periodic manual depth checks during site visits. At both sites, a Diver HUB IoT sensor string was installed directly at the well screen to capture hydrological and environmental parameters every fifteen minutes (Diver HUB, 2025). This temporal resolution was selected to balance data storage with accuracy while providing sufficient detail to assess groundwater recharge and support artificial intelligence and machine learning modeling. In addition, the system records water pressure, which is used to calculate groundwater depth after barometric correction, together with simultaneous measurements of water and air temperature. Electrical conductivity and salinity sensors provide continuous water quality data, while a built-in barometer logs ambient pressure for accurate water level adjustment.

2.2. Data Collection and Preparation

The primary dataset for this study was collected using the Diver HUB platform, an IoT based groundwater monitoring system provided by the Directorate of Irrigation in Erbil and installed at the Tarin and Daratoo pilot recharge sites. Monitoring covered the wet and early dry seasons. At Daratoo, data were collected from 20 February to 20 April 2024. At Tarin Well 1, monitoring extended from 11 February to 2 June 2024, while at Tarin Well 2 it continued until 29 July 2024. The platform recorded water pressure, which was converted to groundwater depth after barometric correction, as well as water and air temperature, electrical conductivity, salinity, and ambient air pressure at 15-minute intervals. This sampling interval provides sufficient temporal resolution to resolve storm driven hydrographs, including the rising limb, peak, and recession associated with ponding events. In addition, it captures short term dynamics and temporal derivatives such as rates of change and lagged responses that are useful for feature engineering. It also provides dense time series data required for robust training and validation of machine learning and deep learning models. Sensors were factory calibrated prior to deployment and were subjected to periodic field checks and cleaning to minimise fouling and sensor drift. Manufacturer performance specifications, including pressure accuracy of ±0.05% FS and temperature accuracy of ±0.1 °C, informed the error assessment. Data were quality controlled through barometric correction, removal of clear outliers, and cross validation of automated readings with manual depth measurements during routine site visits. As a result, these procedures produced continuous and high-quality records suitable for quantifying infiltration events, assessing aquifer response to ponding, and developing data driven predictive models.

2.3. Groundwater Depth Measurement at Tarin and Daratoo Recharge Wells

The study aimed to determine groundwater depth for three shallow recharge wells in Daratoo and two deep recharge wells in Tarin. The Directorate of Irrigation in Erbil provided manual groundwater readings taken at each location at different time intervals. The total pressure at the sensor depth, which is the sum of air pressure and water pressure, is recorded by a pressure sensor installed in the monitoring well. The following formula, in which all parameters are expressed in units of length in meters, was used to determine groundwater depth. Tables 3, 4, and 5 present the manual readings obtained for the Daratoo, Tarin 1, and Tarin 2 wells. In addition, Table 6 summarizes groundwater depth measurements at the Tarin 2 site from Diver HUB sensors and manual observations.

Table 3: Tarin 1 (Deep Recharge Well) (Directorate of Irrigation, 2024)

Date and time	Manual groundwater depth (m)
3/4/2024 11:00	48.90
3/18/2024 11:00	73.50
4/3/2024 11:00	40.70
4/24/2024 11:00	50.60
5/14/2024 11:30	39.12
6/6/2024 10:45	63.55
7/22/2024 10:45	71.40
8/14/2024 10:40	72.60
9/23/2024 11:00	59.14
10/17/2024 10:25	84.58
11/26/2024 12:45	90.68

Table 4: Tarin 2 (Deep Recharge Well) (Directorate of Irrigation, 2024)

Date and time	Manual groundwater depth (m)
3/4/2024 11:00	39.16
3/18/2024 11:00	40.30
4/3/2024 11:00	39.48
4/24/2024 11:00	40.45
5/14/2024 11:30	39.68
7/22/2024 10:45	57.56

Table 5: Daratoo (Shallow Recharge Well) (Directorate of Irrigation, 2024)

Date and time	Manual groundwater depth (m)
2/29/24 12:00 PM	98.60
3/18/24 1:10 PM	97.90
4/24/24 1:30 PM	95.55
5/14/24 1:40 PM	95.65
6/6/24 1:15 PM	96.18
7/22/24 12:15 PM	97.59

Table 6. Groundwater depth calculations from Diver-HUB monitoring at Tarin 2 site

Date and Time	Air Pressure (cm.H_2O)	Diver Pressure (cm.H_2O)	Water Pressure (m.H_2O)	Reference Sensor Depth (m)	Manual GW Depth (m)
2/20/2024 13:00	956.75	4130.217	31.735	70.895	39.16
3/18/2024 11:30	951.50	3988.575	30.371	70.671	40.30
4/3/2024 13:00	952.65	4033.042	30.804	70.644	39.84
4/24/2024 13:00	952.825	3977.650	30.248	70.698	40.45
5/14/2024 11:30	950.20	4048.950	30.988	70.668	39.68
7/22/2024 10:30	941.05	2256.100	13.150	70.710	57.56
Average	950.83	3739.089	27.883	70.714	42.83

At the Tarin 2 monitoring site (VEI_X6134), groundwater measurements were recorded on 2/11/2024 at 17:00 (UTC+03:00). Groundwater depth was determined using diver-measured water pressure and a reference sensor depth based on the following formula:

$$\text{GW Depth [m]} = Z_{\text{ref}} - H_{\text{water}} \quad (1)$$

where Z_{ref} is the reference sensor depth and H_{water} is the water pressure expressed in meters of water column.

For example, at this site, the reference sensor depth was 70.70 m, and the diver-measured water pressure was 3058.108 cm H_2O. Converting this water pressure to meters gives:

$$H_{\text{water}} = \frac{3058.108}{100} = 30.581 \text{ m } \mathrm{H_2O} \tag{2}$$

The data logger groundwater depth was then calculated as:

$$\text{GW Depth} = 70.70 - 30.581 = 40.119 \text{ m} \tag{3}$$

The data obtained through the Diver-HUB platform served as the core foundation of this study. Although the dataset was extensive and fully utilized during the analysis and modeling phases, its size made it impractical to present in full within the paper.

2.4. Predictive Modeling through Machine Learning and Deep Learning

To predict groundwater depth responses to pond-based MAR, this study applied a suite of machine learning and deep learning algorithms to high frequency sensor data collected from two pilot sites. The models included linear regression, feedforward neural networks, XGBoost, gradient boosting, and random forest. All algorithms were implemented in the Python programming language using the Google Colab environment. Input features included time stamped measurements of diver water pressure, air pressure, air temperature, water temperature, electrical conductivity, and salinity recorded at 15-minute intervals. Prior to modeling, barometric correction was applied to derive groundwater depth from diver pressure readings. In addition, standard data preprocessing techniques such as outlier removal, normalization, and timestamp alignment were used to ensure dataset integrity. The models were trained to identify patterns in environmental variables that influence groundwater depth. Feature importance was assessed using the random forest algorithm to determine which input variables contributed most to model predictions. Furthermore, correlation matrices were computed to examine relationships among variables and to support feature selection. Model performance was evaluated using the coefficient of determination R^2 and root mean square error RMSE. Comparative plots were generated to visualize agreement between predicted and observed groundwater depths. Among the tested models, linear regression yielded the highest predictive accuracy across sites, indicating a near linear relationship between input variables and groundwater response at the monitored wells. Overall, this modeling framework enabled accurate and high-resolution predictions of groundwater levels and supported the assessment of MAR system performance. However, the analysis was limited by short monitoring periods and site-specific data, which restrict broader generalization. Nevertheless, the integration of IoT sensor networks with machine learning and deep learning modeling represents a scalable approach for real time groundwater management in semi arid regions.

3. Results and Discussions

3.1. Groundwater Recharge Analysis for Tarin and Daratoo Ponds

Continuous IoT-based sensor monitoring revealed distinct seasonal groundwater responses at both pilot sites. At the Tarin pond (deep well), water levels rose significantly during spring, increasing from approximately 39.16 m in late February to about 40.45 m by late March 2024, representing a net gain of around 1.29 m. The most rapid rise occurred in late April and corresponded with peak seasonal rainfall. In contrast, the Daratoo pond shallow well initially declined from 98.60 m on 28 April to a minimum of 96.2 m by 6 June 2024 before recovering by nearly 2.4 m as infiltrating precipitation gradually reached the groundwater table. These observations demonstrate that small, well designed recharge ponds can produce substantial groundwater mounding in semi arid regions. They also highlight the important role of high frequency automated monitoring in capturing the full dynamics of seasonal water balance, which would be difficult to achieve using traditional manual measurements (Aderemi et al., 2021).

However, the 1.29 m increase at Tarin is significantly higher than the average annual rise of about 0.2 m reported for a treated wastewater recharge pond in Dibdibba Aquifer in Iraq (Hassan et al., 2023). However, it falls within the range of localized maxima up to 2.5 m documented in managed aquifer recharge systems linked to irrigated stream corridors in Colorado (Deng and Bailey, 2022). Similarly, the 2.4 m rebound at Daratoo surpasses the minor gains projected by long-term pond recharge simulations for the Dibdibba setting (Hassan et al., 2023). These findings suggest that the pond design and site conditions in this study provided favorable infiltration and recharge performance. Furthermore, even relatively simple surface impoundments, when strategically located and properly managed, can achieve groundwater level increases comparable to those produced by engineered direct injection schemes. It should be noted, however, that the stronger responses observed at the Tarin and Daratoo sites occurred only during short periods and under relatively low precipitation conditions, which may have amplified the apparent gains. This suggests that the observed increases reflect short-term, localized responses rather than sustained long-term trends, highlighting the influence of site-specific conditions and pond operation on recharge effectiveness. In summary, the results confirm that constructed recharge ponds are an effective and cost-efficient method to enhance aquifer storage in arid and semi-arid regions. The study also demonstrates that continuous IoT monitoring is essential for capturing temporal and seasonal fluctuations in groundwater levels, enabling more informed water management decisions. Together, the integration of strategically located ponds and real-time sensor data allows for improved assessment of recharge performance, supports the optimization of pond operations, and provides valuable insights for sustainable groundwater management and potential flood mitigation (Aderemi et al., 2021).

3.2. Comparative Analysis of Machine Learning Models for Depth Prediction in Diver Data

In this study, four predictive algorithms were applied including the ordinary least-squares regression, a simple feedforward neural network, XGBoost, and GBoost, to three Diver-HUB time series (Daratoo, Tarin-1, and Tarin-2) to forecast groundwater depth (Diver-HUB, 2025). Because the Daratoo sensor failed after April 20, 2024, its record count is smaller, whereas Tarin 1 and Tarin 2 provide richer histories, with Tarin 1 being the most extensive. A random forest feature-importance analysis (Figures 5, 8, and 11) showed that water pressure is the strongest predictor, with diver pressure (cmH_2O) contributing significantly. Seasonal and time-of-day variables (month and hour) and temperature readings played a much smaller role. Correlation matrices (Figures 6, 9, and 12) confirmed a tight coupling between water and diver pressures, underscoring that pressure alone captures most of the signal we need for depth estimation.

When plotted as time series and as predicted versus observed scatter plots (Figures 7, 10, and 13), the four algorithms show markedly different behaviour across the three sites. Quantitative performance (Table 7) indicates that ordinary linear regression delivered near-perfect fits on the Diver-HUB test data (Daratoo: $R^2 = 0.99995$, RMSE = 0.00285 m; Tarin 1: $R^2 = 1.000$, RMSE = 0.00029 m; Tarin 2: $R^2 = 1.000$, RMSE = 0.00029 m). The feedforward neural network tracked observed depths closely at the Tarin sites (Tarin 1: $R^2 = 0.99$, RMSE = 0.00841 m; Tarin 2: $R^2 = 0.99851$, RMSE = 0.1656 m) but performed poorly on the Daratoo record ($R^2 = -1.436$, RMSE = 0.61815 m). By contrast, the tree ensemble models (XGBoost and the GBoost variant) produced unstable results, with moderate skill at Tarin 1 (XGBoost $R^2 \approx 0.79$, RMSE ≈ 0.34 m) but strongly negative R^2 and very large RMSE at Daratoo and Tarin 2 (for example, XGBoost at Tarin 2: $R^2 = -0.648$, RMSE = 5.499 m), indicating predictions worse than the baseline mean.

These findings suggest that for high frequency, single site groundwater sensors, simpler models often outperform complex ensembles because the dominant predictor, barometrically corrected diver pressure, has a direct and near linear relationship with groundwater depth, and there are few additional informative features. Moreover, the analysis is based on single site, high resolution time series with a relatively limited effective sample size (Daratoo’s record was truncated by a sensor failure), conditions that favor parsimonious models and increase the risk of overfitting for flexible learners such as random forest, XGBoost, and deep neural networks.

This contrasts with studies that report benefits from ensemble or deep learning approaches in spatially distributed or large aquifer datasets, where nonlinear processes, site-to-site heterogeneity, and indirect predictors make sophisticated models advantageous (Tran et al., 2025; Feng et al., 2024). Consequently, in this focused setting, linear regression’s transparency, computational efficiency, and robustness make it the most practical and reliable choice, while more complex methods are likely to outperform linear models when applied to larger, multi-site, or more heterogeneous monitoring networks.

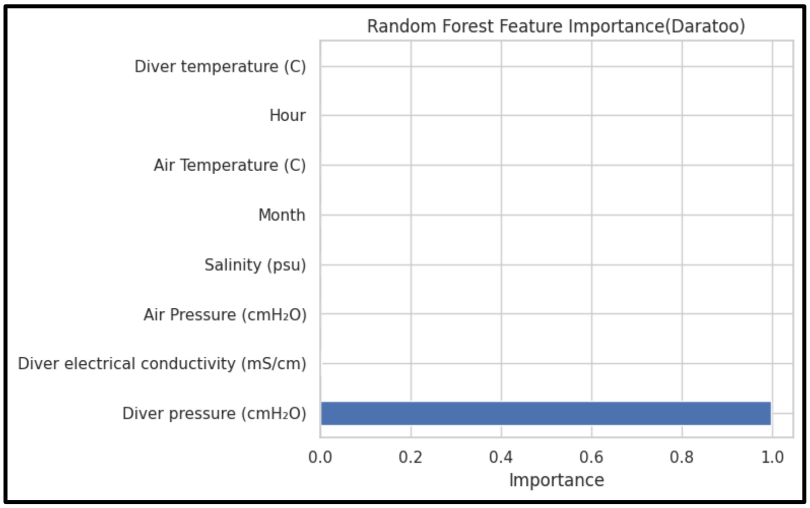
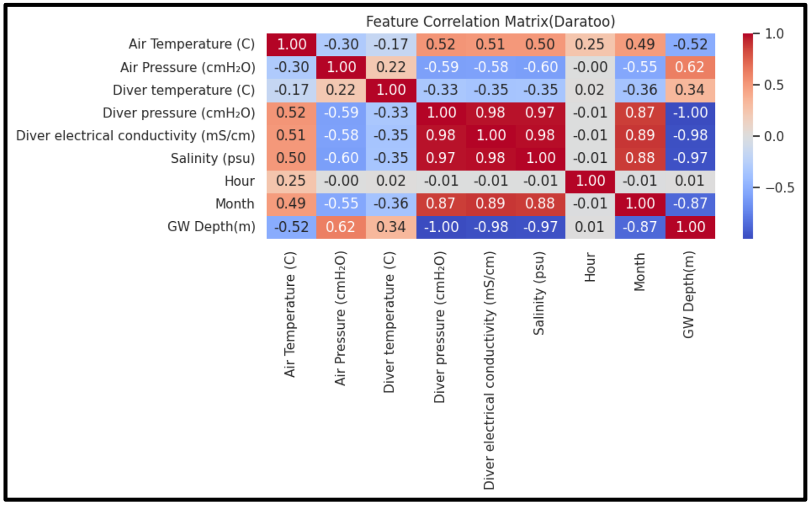


Figure 5: Random Forest feature importance for Daratoo dataset



	Air Temperature (C)	Air Pressure (cmH$_2$O)	Diver temperature (C)	Diver pressure (cmH$_2$O)	Diver electrical conductivity (mS/cm)	Salinity (psu)	Hour	Month	GW Depth(m)
Air Temperature (C)	1.00	-0.30	-0.17	0.52	0.51	0.50	0.25	0.49	-0.52
Air Pressure (cmH$_2$O)	-0.30	1.00	0.22	-0.59	-0.58	-0.60	-0.00	-0.55	0.62
Diver temperature (C)	-0.17	0.22	1.00	-0.33	-0.35	-0.35	0.02	-0.36	0.34
Diver pressure (cmH$_2$O)	0.52	-0.59	-0.33	1.00	0.98	0.97	-0.01	0.87	-1.00
Diver electrical conductivity (mS/cm)	0.51	-0.58	-0.35	0.98	1.00	0.98	-0.01	0.89	-0.98
Salinity (psu)	0.50	-0.60	-0.35	0.97	0.98	1.00	-0.01	0.88	-0.97
Hour	0.25	-0.00	0.02	-0.01	-0.01	-0.01	1.00	-0.01	0.01
Month	0.49	-0.55	-0.36	0.87	0.89	0.88	-0.01	1.00	-0.87
GW Depth(m)	-0.52	0.62	0.34	-1.00	-0.98	-0.97	0.01	-0.87	1.00

Figure 6: Feature correlation matrix for Daratoo dataset

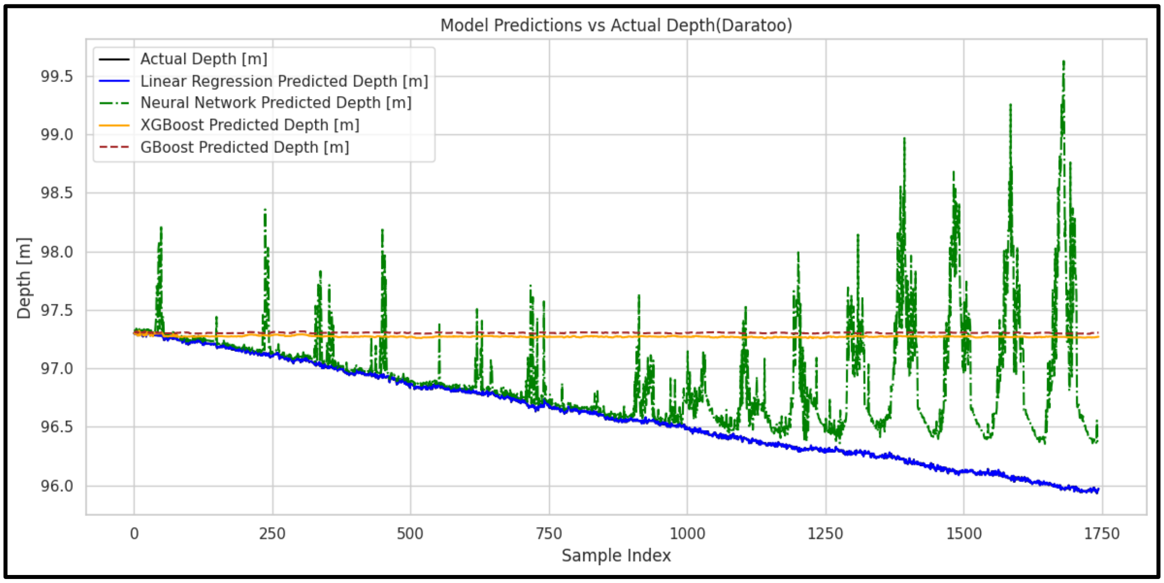


Figure 7: Model predictions versus actual depth for Daratoo dataset

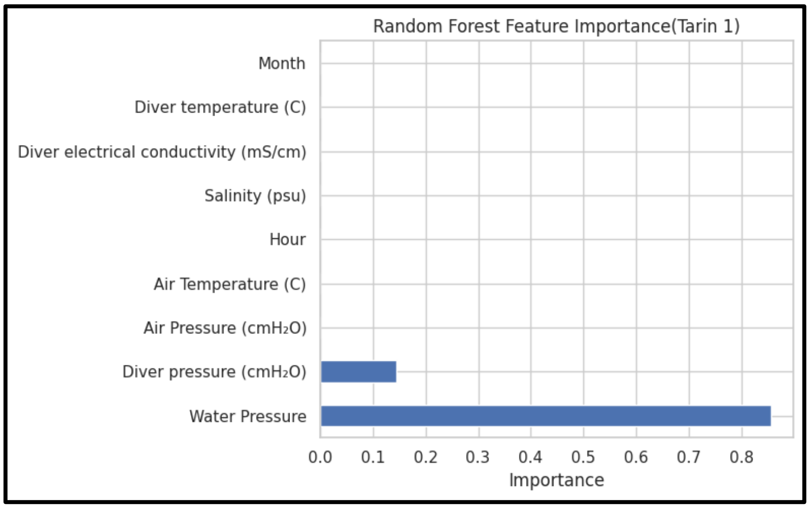
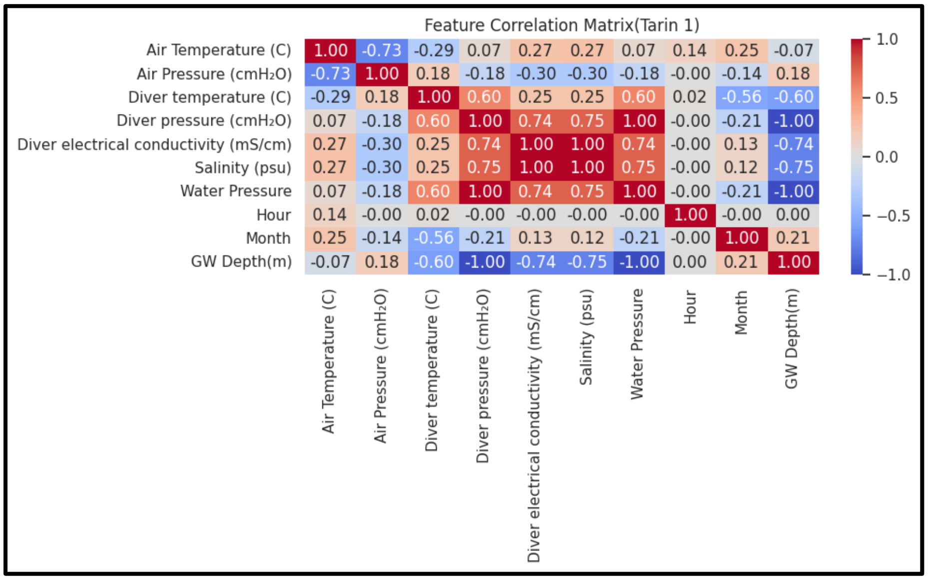


Figure 8: Random Forest feature importance for Tarin 1 dataset



	Air Temperature (C)	Air Pressure (cmH₂O)	Diver temperature (C)	Diver pressure (cmH₂O)	Diver electrical conductivity (mS/cm)	Salinity (psu)	Water Pressure	Hour	Month	GW Depth(m)
Air Temperature (C)	1.00	-0.73	-0.29	0.07	0.27	0.27	0.07	0.14	0.25	-0.07
Air Pressure (cmH₂O)	-0.73	1.00	0.18	-0.18	-0.30	-0.30	-0.18	-0.00	-0.14	0.18
Diver temperature (C)	-0.29	0.18	1.00	0.60	0.25	0.25	0.60	0.02	-0.56	-0.60
Diver pressure (cmH₂O)	0.07	-0.18	0.60	1.00	0.74	0.75	1.00	-0.00	-0.21	-1.00
Diver electrical conductivity (mS/cm)	0.27	-0.30	0.25	0.74	1.00	1.00	0.74	-0.00	0.13	-0.74
Salinity (psu)	0.27	-0.30	0.25	0.75	1.00	1.00	0.75	-0.00	0.12	-0.75
Water Pressure	0.07	-0.18	0.60	1.00	0.74	0.75	1.00	-0.00	-0.21	-1.00
Hour	0.14	-0.00	0.02	-0.00	-0.00	-0.00	-0.00	1.00	-0.00	0.00
Month	0.25	-0.14	-0.56	-0.21	0.13	0.12	-0.21	-0.00	1.00	0.21
GW Depth(m)	-0.07	0.18	-0.60	-1.00	-0.74	-0.75	-1.00	0.00	0.21	1.00

Figure 9: Random Forest feature importance for Tarin 1 dataset

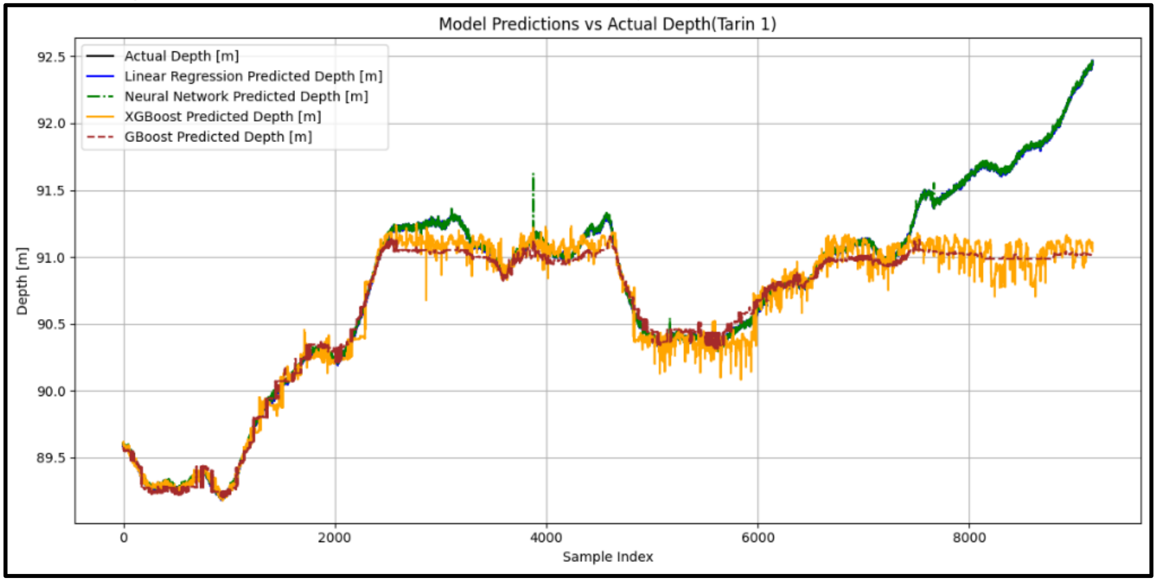


Figure 10: Model predictions versus actual depth for Tarin1 dataset

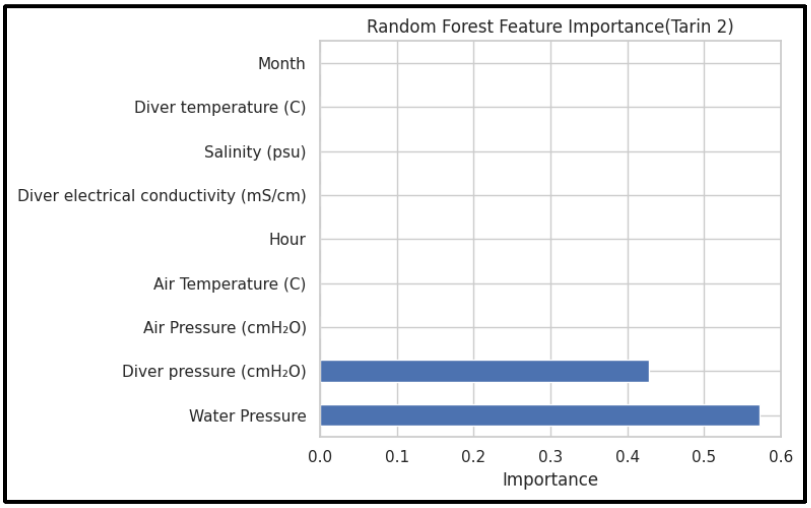
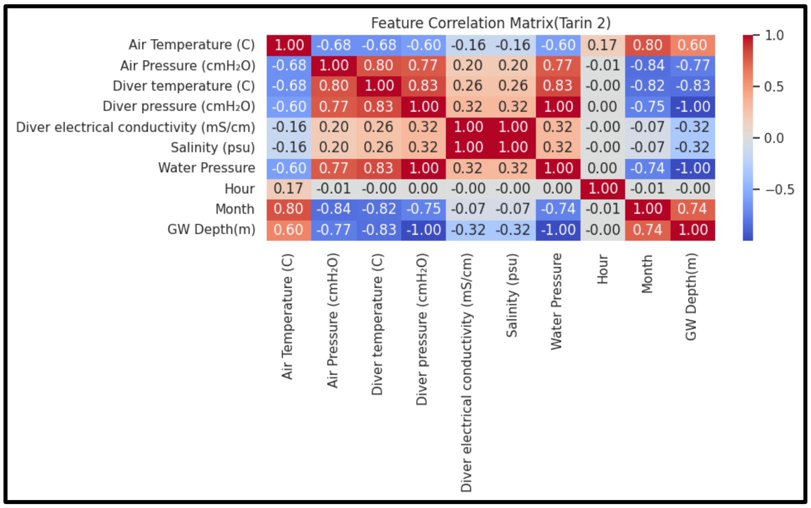


Figure 11: Random Forest feature importance for Tarin 2 dataset



	Air Temperature (C)	Air Pressure (cmH₂O)	Diver temperature (C)	Diver pressure (cmH₂O)	Diver electrical conductivity (mS/cm)	Salinity (psu)	Water Pressure	Hour	Month	GW Depth(m)
Air Temperature (C)	1.00	-0.68	-0.68	-0.60	-0.16	-0.16	-0.60	0.17	0.80	0.60
Air Pressure (cmH₂O)	-0.68	1.00	0.80	0.77	0.20	0.20	0.77	-0.01	-0.84	-0.77
Diver temperature (C)	-0.68	0.80	1.00	0.83	0.26	0.26	0.83	-0.00	-0.82	-0.83
Diver pressure (cmH₂O)	-0.60	0.77	0.83	1.00	0.32	0.32	1.00	0.00	-0.75	-1.00
Diver electrical conductivity (mS/cm)	-0.16	0.20	0.26	0.32	1.00	1.00	0.32	-0.00	-0.07	-0.32
Salinity (psu)	-0.16	0.20	0.26	0.32	1.00	1.00	0.32	-0.00	-0.07	-0.32
Water Pressure	-0.60	0.77	0.83	1.00	0.32	0.32	1.00	0.00	-0.74	-1.00
Hour	0.17	-0.01	-0.00	0.00	-0.00	-0.00	0.00	1.00	-0.01	-0.00
Month	0.80	-0.84	-0.82	-0.75	-0.07	-0.07	-0.74	-0.01	1.00	0.74
GW Depth(m)	0.60	-0.77	-0.83	-1.00	-0.32	-0.32	-1.00	-0.00	0.74	1.00

Figure 12: Random Forest feature importance for Tarin 2 dataset

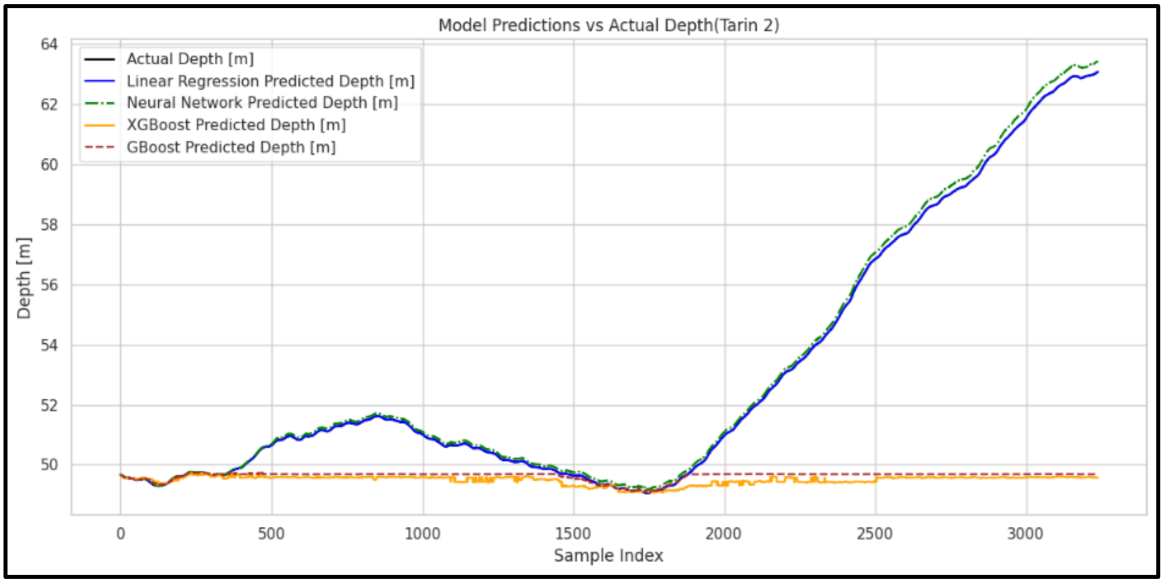


Figure 13: Model predictions versus actual depth for Tarin1 dataset

Table 7: Comparison of ML and DL models for water depth prediction across different datasets

No	Name	Model Comparison	R2 Score	RMSE
1	Daratoo	Linear Regression	0.99995	0.00285
		Neural Network	-1.43639	0.61815
		XG Boost	-2.85152	0.7769
		G Boost	-3.13216	0.80502
2	Tarin 1	Linear Regression	1	0.00029
		Neural Network	0.99	0.00841
		XG Boost	0.79406	0.34267
		G Boost	0.78327	0.35153
3	Tarin 2	Linear Regression	1	0.00029
		Neural Network	0.99851	0.1656
		XG Boost	-0.64792	5.49891
		G Boost	-0.59798	5.41493

4. Conclusions

The paper highlights the potential of integrating IoT sensor networks and advanced ML models to monitor and control pond-based MAR. Low-cost Diver-HUB pressure sensors were used to collect high-frequency water levels and temperature, allowing real-time observations of aquifer dynamics. This ongoing monitoring found that pilot sites experienced substantial groundwater recharge, with water levels increasing 1.29 m at Tarin and 2.4 m at Daratoo sites. Statistical learning models, particularly the linear regression model, provided highly accurate estimates of groundwater depths in this study, though some uncertainty remains due to data limitations and potential challenges in generalizing the results to other regions.

This integrated IoT and ML approach provides a cost-effective, scalable solution for real-time groundwater management, which increases the efficiency of the pond-based recharge systems and can lead to adaptive management. Results showed that recharge ponds can substantially improve groundwater replenishment when they are well-designed, continuously monitored by an IoT system, and analyzed through ML. Such a system contributes to sustainable water resources management in a growing world facing water scarcity challenges. Future work should focus on expanding sensor networks, developing adaptive management systems, integrating climate and land-use data, and testing this IoT–ML framework in diverse hydrogeological settings to confirm its broader applicability and robustness

5. Conflict of Interest

The author declares that there is no conflict of interest regarding the publication of this study.

6. Acknowledgments

The authors gratefully acknowledge the Civil Engineering Department, College of Engineering, Salahaddin University in Erbil, for their support in carrying out this research. We also extend our appreciation to the Directorate of Irrigation of Erbil for providing the IoT data essential to this study.

7. References

Aderemi, B.A., Olwal, T.O., Ndambuki, J.M. and Rwanga, S.S., 2021. A review of groundwater management models with a focus on IoT-based systems. *Sustainability*, *14*(1), pp.148. https://doi.org/10.3390/su14010148.

Al-Kakey, O., Othman, A. and Merkel, B., 2023. Identifying potential sites for artificial groundwater recharge using GIS and AHP techniques: a case study of Erbil basin, Iraq. *Kuwait Journal of Science*, *50*(1B). DOI: 10.48129/kjs.11917.

Ameen, S.M., Aziz, S.Q., Dawood, A.H., Sabir, A.T. and Hawez, D.M., 2025. Utilizing Machine Learning and Deep Learning for Precise Intensity-Duration-Frequency (IDF) Curve Predictions. Polytechnic Journal, 15(1), pp.3. https://doi.org/10.59341/2707-7799.1848.

Aziz, S.Q., Ahmed, B.M., Mahmod, A.A. and Othman, S., 2023c. Improvement of Erbil City Environment and Increasing Irrigated Areas through Simulated Hydraulic Design from Greater-Zab River Water. Academic Journal of Nawroz University (AJNU), 12(4). DOI: 10.25007/ajnu.v12n4a1656.

Aziz, S.Q. and Muhammed, K.K., MODERN METHODS AND ARTIFICIAL INTELLEGNCE IN STORMWATER HARVESTING AS SUSTAINABLE SOLUTIONS. The 3rd International Conference on Engineering and Innovative Technology (ICEIT 2024), October 30-31, 2024, Erbil, Kurdistan Region, Iraq. DOI: 10.31972/iceit2024.053.

Bates, B., Kundzewicz, Z. and Wu, S., 2008. Climate change and water. Intergovernmental panel on climate change secretariat. Available at: https://www.ipcc.ch/site/assets/uploads/2018/03/climate-change-water-en.pdf.

Calderwood, A.J., Pauloo, R.A., Yoder, A.M. and Fogg, G.E., 2020. Low-cost, open-source wireless sensor network for real-time, scalable groundwater monitoring. Water, 12(4), p.1066. https://doi.org/10.3390/w12041066.

Deng, C. and Bailey, R.T., 2022. Assessing the Impact of Artificial Recharge Ponds on Hydrological Fluxes in an Irrigated Stream–Aquifer System. Hydrology, 9(5), p.91. https://doi.org/10.3390/hydrology9050091.

Dillon, P., Stuyfzand, P., Grischek, T., Lluria, M., Pyne, R.D.G., Jain, R.C., Bear, J., Schwarz, J., Wang, W., Fernandez, E. and Stefan, C., 2019. Sixty years of global progress in managed aquifer recharge. Hydrogeology journal, 27(1), pp.1-30. https://doi.org/10.1007/s10040-018-1841-z.

Diver-HUB, 2024. *Diver-HUB IoT groundwater monitoring system*. [online] Available at: https://diver-hub.com/Bro/Home/Index [Accessed 11 Jun. 2025].

Directorate of Irrigation, 2024. Report on conducting pilot projects to implement three Managed Aquifer Recharge methods within Erbil Governorate, funded by UNICEF.

Espinoza Ortiz, M., Apún Molina, J.P., Belmonte Jiménez, S.I., Herrera Barrientos, J., Peinado Guevara, H.J. and Santamaria Miranda, A., 2023. Development of low-cost IoT system for monitoring piezometric level and temperature of groundwater. Sensors, 23(23), p.9364. https://doi.org/10.3390/s23239364.

Feng, F., Ghorbani, H. and Radwan, A.E., 2024. Predicting groundwater level using traditional and deep machine learning algorithms. Frontiers in Environmental Science, 12, p.1291327. https://doi.org/10.3389/fenvs.2024.1291327.

Hama, R.H., Hamad, R.T. and Aziz, F.H., 2014. Climate change in relation to rainfall and temperature in Erbil province, Kurdistan, Iraq. Tunisian Association of Digital Geographic Information, 8Th International Congers Geo Tunis, Tunis, 8. Available: https://www.researchgate.net/publication/286392629_Climate_change_in_relation_to_rainfall_and_temperature_in_Erbil_province_Kurdistan_Iraq.

Hamad, R 2022, "Erbil Basin Groundwater Recharge Potential Zone Determination using Fuzzy-Analytical Hierarchy Process (AHP) in North Iraq," *Tikrit Journal for Agricultural Sciences*, 22(3):175–190, https://doi.org/10.21203/rs.3.rs-1315621/v1.

Hassan, W.H., Ghanim, A.A., Mahdi, K., Adham, A., Mahdi, F.A., Nile, B.K., Riksen, M. and Ritsema, C., 2023. Effect of artificial (pond) recharge on the salinity and groundwater level in Al-Dibdibba Aquifer in Iraq using treated wastewater. Water, 15(4), p.695. https://doi.org/10.3390/w15040695.

Koncagül, E., Tran, M. and Connor, R., 2021. The United Nations world water development report 2021: valuing water; facts and figures. Available: https://unesdoc.unesco.org/ark:/48223/pf0000375751.

Margat, J. and Van der Gun, J., 2013. Groundwater around the world: a geographic synopsis. Crc Press. DOI: 10.1201/b13977.

Mustafa, J.S. and Mawlood, D.K., 2023. Estimating of groundwater recharge in North, Central, and South Basin of Erbil. Mathematical Modelling in Civil Engineering, 18(1), pp.14-24. DOI: 10.14738/tecs.122.16758.

Mustafa, JS & Mawlood, DK 2024, "Assessment of the Groundwater in Erbil Basin with Support of Visual MODFLOW," *Journal of Ecological Engineering*, 25(4):203–227, https://doi.org/10.12911/22998993/184184.

Sheik, A.G., Kumar, A., Sharanya, A.G., Amabati, S.R., Bux, F. and Kumari, S., 2024. Machine learning-based monitoring and design of managed aquifer rechargers for sustainable groundwater management: scope and challenges. Environmental Science and Pollution Research, pp.1-34. https://doi.org/10.1007/s11356-024-35529-3.

Sufyan, M., Martelli, G., Teatini, P., Cherubini, C. and Goi, D., 2024. Managed Aquifer Recharge for Sustainable Groundwater Management: New Developments, Challenges, and Future Prospects. Water, 16(22), pp.1-28. https://dx.doi.org/10.3390/w16223216.

Tran, T.V., Peche, A., Kringel, R., Broemme, K. and Altfelder, S., 2025. Machine Learning-Based Reconstruction and Prediction of Groundwater Time Series in the Allertal, Germany. Water (20734441), 17(3). DOI: 10.3390/w17030433.

Zaresefat, M. and Derakhshani, R., 2023. Revolutionizing groundwater management with hybrid AI models: A practical review. Water, 15(9), p.1750. https://doi.org/10.3390/w15091750.