Predicting Groundwater Quality Using Machine Learning Models: A Case Study in Semnan Province, Iran

Setareh Zaman[1], Amir Derakhshanpour[2], Kimia Peyvandi*[3]

[1] Bachelor's Student, Department of Computer Science, Semnan University, Semnan, Iran; setareh.zaman82@gmail.com

[2] Master's Student, Department of Computer Science, Semnan University, Semnan, Iran; amirderakhshanpour@semnan.ac.ir

[*3]Assistant Professor, Department of Computer Science, Semnan University, Semnan, Iran; kpeyvandi@semnan.ac.ir

ABSTRACT

Groundwater in arid and semi-arid regions faces increasing pressures from overexploitation, climate variability, and declining quality, making accurate forecasting essential for sustainable management. This study examines two decades of hydrological and chemical data from Semnan Province, Iran (2003–2023), including precipitation, piezometric levels, and key water quality parameters. Eight nonlinear machine learning models, covering tree-based, ensemble, kernel-based, instance-based, and neural networks, were tested. Temporal patterns were captured using a walk-forward validation combined with exponentially weighted averaging. Ensemble methods, especially Random Forest and XGBoost, delivered the best performance, with an average R^2 of 0.90, RMSE of 820.0, and RMAE of 21. These results show that combining different environmental data with advanced nonlinear models provides a reliable and flexible way to forecast groundwater quality in water-scarce areas.

Keywords
Machine Learning,
Groundwater,
Water Quality,
Walk-forward Validation

1. INTRODUCTION

Scarcity of water is now one of the most urgent environmental and socio-economic concerns of the 21st century. Population, urbanization, and industrialization have placed record demands on freshwater resources, while climate change continues to exacerbate trends of drought and alter hydrological cycles [1]. Among the different sources of freshwater, groundwater is of primary significance in arid and semi-arid nations such as Iran, where surface water is limited and seasonal; In such nations, groundwater is the primary source of drinking water, supplies water for agriculture, and is used for industry purposes pivotal to domestic economies [1]. But all this growing dependence on groundwater has resulted in over pumping, decreasing water tables, and declining water quality, none of which are positive signs for future sustainability.

With deepening water scarcity, conventional and non-conventional water sources have been distinguished. Conventional sources such as surface water and shallow groundwater are naturally replenished and have been the focal point of water supply infrastructure for centuries [2]. However, in arid and semi-arid regions, such as large parts of Iran, these sources are no longer capable of meeting increasing demands driven by population growth, agricultural expansion, and urbanization; Therefore, attention has turned to non-conventional sources of groundwater, i.e., deeper or saline aquifers with reduced recharging and more irregular quality patterns [3]. Such non-conventional reservoirs will become crucial in the coming years because of the dearth of water and therefore their sustainable production demands careful monitoring and accurate forecasting of their behavior [1]. This is due to their complex hydrogeochemical processes and vulnerability to climatic change and human intervention.

In the context of Semnan Province, such unconventional groundwater resources are characterized by notable spatial and temporal variability in both quality and level. Key hydro chemical indicators such as sodium (Na^+) and chloride (Cl^-) concentrations reflect salinization trends resulting from excessive abstraction, mineral dissolution and evapotranspiration [3]. Meanwhile, piezometric water-level data represent the quantitative dimension of groundwater depletion, while rainfall data capture the climatic input driving recharge processes. The integration of these variables enables a comprehensive understanding of how climatic factors and geochemical processes jointly influence groundwater anomalies [1]. Their inclusion in the modeling framework ensures that both the quantity (water-level) and quality (Na^+, Cl^-) dimensions of groundwater dynamics are represented, providing a robust foundation for predictive analysis.

This has necessitated the employment of machine learning (ML) techniques, which can analyze large, heterogeneous data and detect latent patterns that control unconventional groundwater quantity and quality variations [4]. The present study uses and compares different ML algorithms to forecast the anomalies of Semnan Province's unconventional groundwater resources and offers insight that can be applied to more data-driven and sustainable management practices. Semnan Province in central Iran provides a representative case in point. Its low rainfall and high evaporation climatic regime make groundwater the dominant and sometimes the only stable water supply. Semnan has witnessed drastic fluctuations of groundwater levels and increasing water quality anomalies during the last decades, mainly caused by over-abstraction, land use change, and irregular recharging regimes. It is necessary to understand and predict such deviations not just to avoid further depletion but to make management decisions data-oriented and proactive [2].

Traditional statistical and hydrologic models, while useful, have the drawback of being unable to capture the nonlinear and multivariate nature of groundwater systems involving complex interactions between climatic, geologic, and anthropogenic processes [5]. To this end, machine learning (ML) methods have become increasingly popular as effective alternatives [6]. ML techniques can identify sophisticated relationships in large data to predict hydrological variables more precisely and more flexibly [6]. These techniques have been applied with success in different environmental applications like rainfall prediction, modeling of water quality, and prediction of drought [7]. Their effectiveness, however, is based on the selection of suitable algorithms, type and quality of input data, and fitness of the model to the specifics of the area of study [2, 4].

The present research attempts to explore and estimate the ability of some ML models in predicting groundwater anomalies in Semnan Province, Iran. By integrating long-term databases of precipitation data records, piezometric level data records, and primary water quality parameters, the present research provides a comprehensive discussion of the performance of a broad range of machine learning approaches over real-world environmental processes. Comparative analysis not only seeks to identify the most effective predictive model but also to construct understanding of determinants of groundwater dynamics that lie under. Ultimately, the product of this study is expected to assist in creating more effective groundwater management and increase understanding of how data-driven models can assist in enhancing sustainable water resource planning in regions at risk for scarcity and degradation.

The rest of this paper is organized to support our central objective: Section 2 reviews related works in groundwater quality, with attention to temporal and spatial forecasting; Section 3 outlines the methodology, first introducing the dataset and detailing feature engineering and preprocessing (3.1), then describing the experimental design and evaluation metrics (3.2), and finally the models used in this research (3.3); Section 4 reports experimental results, beginning with the main outcomes, followed by a comparison with all used methods; Section 5 interprets these findings and discusses limitations; and Section 6 concludes with recommendations and future works .

2. Related Works

Recent years have witnessed growing interest in applying machine-learning (ML) techniques to groundwater and water-quality prediction, driven by the increasing availability of hydrogeological and environmental data.

Karimidastenaei et al. [1] synthesized global research on Unconventional Water Resources (UWRs), classifying twelve UWR types and mapping their climatic and geographic suitability. Their findings indicate that UWR adoption is highly constrained by local climatic, geographic, economic, and political factors, leading to limited large-scale applicability. Although publications on UWRs have increased rapidly in the past decade, the authors highlight persistent knowledge gaps and call for more quantitative assessments, socioeconomic analyses, and pilot projects to evaluate real-world scalability and sustainability.

Zhu et al. [2] provided a comprehensive review of machine-learning applications in various kinds of aquatic environments, including surface water, groundwater, drinking water, sewage, and seawater, and demonstrated that data-driven models effectively capture the nonlinear relationships that traditional process-based methods often fail to represent. Their review covered applications ranging from construction and monitoring to pollution control and water-treatment optimization. They emphasized the scalability and pattern-recognition capabilities of ML algorithms but also noted recurring challenges such as data quality, interpretability, and transferability. To strengthen the role of ML in water management, they recommended hybrid physics, ML frameworks, improved data pipelines, and uncertainty-aware modeling.

In the context of groundwater-level (GWL) prediction, Mousavimehr et al. [3] evaluated twelve machine-learning models using satellite-derived inputs (GRACE, GLDAS, ERA5) for the Tehran region. Their results showed that ensemble methods such as Random Forest and AdaBoost consistently outperformed single learners, achieving higher determination coefficients (R^2) and lower RMSE and RMAE values. Incorporating multiple satellite-based variables—terrestrial water storage, soil moisture, and precipitation—significantly improved prediction accuracy. The study also revealed spatial variability in model errors, particularly larger deviations in northern regions, reflecting local hydroclimatic and topographic differences. These findings support the use of ensemble and multi-source approaches for groundwater monitoring in data-scarce regions.

Feng et al. [4] compared traditional machine-learning models (SVM, Decision Tree, Random Forest) with deep learning architectures (RNN, CNN, GAN) for GWL prediction, concluding that data-driven methods effectively capture nonlinear and temporal dynamics in hydrogeological systems. Their work and related studies suggest that while tree-based models provide robust baselines for small datasets, recurrent networks such as LSTM and GRU excel at modeling sequences and multi-step forecasts. Furthermore, recent research has shown that CNNs can outperform RNNs in cases dominated by local temporal or spatial dependencies, due to their noise robustness and computational efficiency. GANs, meanwhile, are primarily utilized for data augmentation and generating synthetic time series. Across studies, the importance of careful feature selection, sensitivity analysis, and rigorous validation using metrics such as RMSE and R^2 has been repeatedly emphasized.

Yadav et al. [5] focused on groundwater-quality prediction using ensemble and deep learning methods. Their comparative study of Random Forest, XGBoost, and deep neural networks demonstrated that simple, rapidly measurable physicochemical parameters, such as pH, total hardness, and total dissolved solids, can accurately predict composite indices like the Environmental Water Quality Index (EWQI) at local scales. They found that tree-based ensemble models offer strong baseline performance and robustness to noisy, heterogeneous data, while deep networks can capture complex nonlinear relationships when sufficient data are available. Furthermore, models based on a reduced set of handheld-measurable parameters enable near-real-time field monitoring, which is critical for small-area water management. Their study also stressed the importance of external validation and uncertainty assessment to ensure model transferability across different hydrogeological contexts.

Yan et al. [6] conducted a meta-analysis of more than 170 studies applying machine learning to water-quality prediction, categorizing them into indicator-level and index-level modeling approaches. They reviewed preprocessing techniques, data acquisition methods, and model architecture including ensemble trees, SVMs, CNNs, RNNs, XGBoost, and DNNs. The authors identified emerging trends such as hybrid modeling, data augmentation, and the integration of hydrodynamic knowledge into ML frameworks. Common evaluation criteria—R^2, RMSE, and RMAE—were consistently reported, while recurring challenges included data scarcity, heterogeneity, and cross-site generalization. The study advocated for incorporating physical process knowledge and developing uncertainty-aware frameworks to improve robustness and interpretability.

In a complementary domain, Goorabi et al. [7] employed Persistent Scatterer Interferometric Synthetic Aperture Radar (PS-InSAR) with Sentinel-1 time series data to map urban subsidence in Isfahan between 2014 and 2019. Their analysis revealed deformation rates ranging from −5 to −100 mm/year and linked these patterns to geological and geomorphological units. The integration of ascending and descending SAR passes improved deformation mapping accuracy. Consistent with prior remote-sensing studies, their work underscored the importance of combining multi-temporal InSAR data with geomorphologic and hydrogeological indicators to attribute subsidence to aquifer depletion, lithology, and land-surface characteristics. They also recommended multi-sensor and ground-based validation (GNSS, leveling, boreholes) to strengthen causal interpretation.

Finally, Zhi et al. [8] argued that deep learning presents a promising direction for inland water-quality prediction, as it can extract high-dimensional relationships that traditional statistical or process-based models often overlook. Their review demonstrated that deep architecture not only enhances predictive accuracy but also aids in identifying influential drivers and generating new hypotheses. They emphasized, however, that challenges related to data demands, interpretability, and transferability remain significant. The authors advocated for integrating deep learning with physical modeling and developing uncertainty-quantified frameworks to achieve more reliable and interpretable outcomes.

3. Data and Methodology

This section describes the data sources, study area, and analytical approach used to develop and evaluate machine-learning models for predicting groundwater quality in Semnan Province, Iran. We use historical groundwater-monitoring records supplemented with meteorological and land-use predictors collected for two decades. After data cleaning, imputation, and feature engineering, several candidate models were trained and tuned using spatially and temporally aware cross-validation to avoid leakage. Model performance is assessed with RMSE and R^2 for continuous targets.

3.1. Dataset

To enrich the groundwater observation dataset, rainfall and piezometric records were integrated through a three-step spatial and temporal linkage process that preserves geographic accuracy and hydrological causality.

1. Linking rainfall to groundwater by station codes A reference table containing each station's unique identifier and coordinates in both geographic (lat/long) and UTM (X, Y) systems was used to join rain-gauge records (which were provided in lat/long) with groundwater records (provided in UTM). Precipitation measurements were matched to groundwater stations via the shared station codes from this reference table, ensuring direct, non-interpolated assignment of rainfall values to the corresponding groundwater monitoring locations and preserving the dataset's spatial integrity. We understand the basic parts of the dataset are the more optional way of the ground water temporal ways.
2. Semi-annual aggregation and temporal synchronization of rainfall Monthly precipitation totals were aggregated into two semi-annual sums per station: March–August (first half) and September–February (second half). To account for the typical lag between rainfall and groundwater recharge, a six-month temporal offset was applied: rainfall occurring in the second half of year t was paired with groundwater measurements taken in the first half of year t+1, while rainfall in the first half of year t was paired with groundwater measurements from the second half of the same year. These semi-annual, lagged rainfall variables were appended to the groundwater records by station code, ensuring the predictor set reflects the delayed hydrologic response.
3. Incorporation of piezometric measurements and derivation of a water-level index Independent piezometric records were incorporated to provide direct information on the aquifer head where the original groundwater dataset lacked water-table measures. For each groundwater observation, the nearest piezometric observation (by

geographic coordinates) was matched and used to compute a water-level variable as the difference between ground surface elevation and piezometric head. For temporal consistency, June piezometric readings were used to represent the first half-year and December readings for the second half; if those months were missing, the nearest available month within the same half-year was substituted (ties resolved by rotation to avoid bias). The resulting water-level index was merged with the groundwater dataset and used as an explanatory variable to improve representation of subsurface conditions and prediction stability.

After these three preprocessing steps, the resulting dataset consists of approximately 5,600 semi-annual groundwater observations collected from monitoring wells across Semnan Province between 1382 and 1402 (2003–2023).

During exploratory data analysis (EDA), both numerical and visual analyses were performed to better understand the relationships between predictors and the target variable (Electrical Conductivity, EC). Correlation analysis revealed that Na^+ and Cl^- had correlation coefficients close to 1 with EC, indicating almost complete overlap and redundancy as shown in Figure 1.

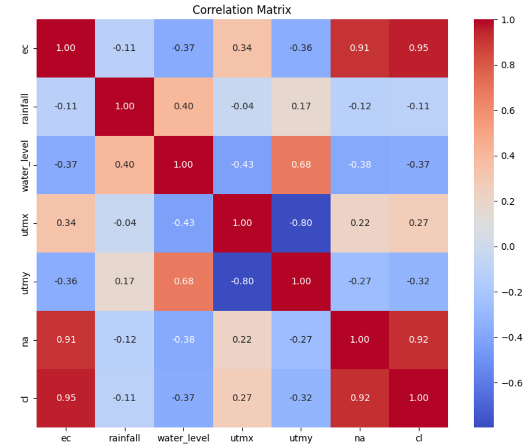

Figure 1. Correlation matrix of the hydrological and chemical features

To further examine the nature of these dependencies, scatter plots were drawn for rainfall, water-level, Na^+, and Cl^- against EC as seen in Figure 2. The plots confirmed that Na^+, Cl^-, and TH displayed near-linear relationships with EC, while rainfall and water-level showed highly nonlinear patterns. Based on these observations, Na^+, Cl^-, and TH were excluded from modeling to avoid multicollinearity and overfitting, while the remaining features were retained for nonlinear regression modeling.

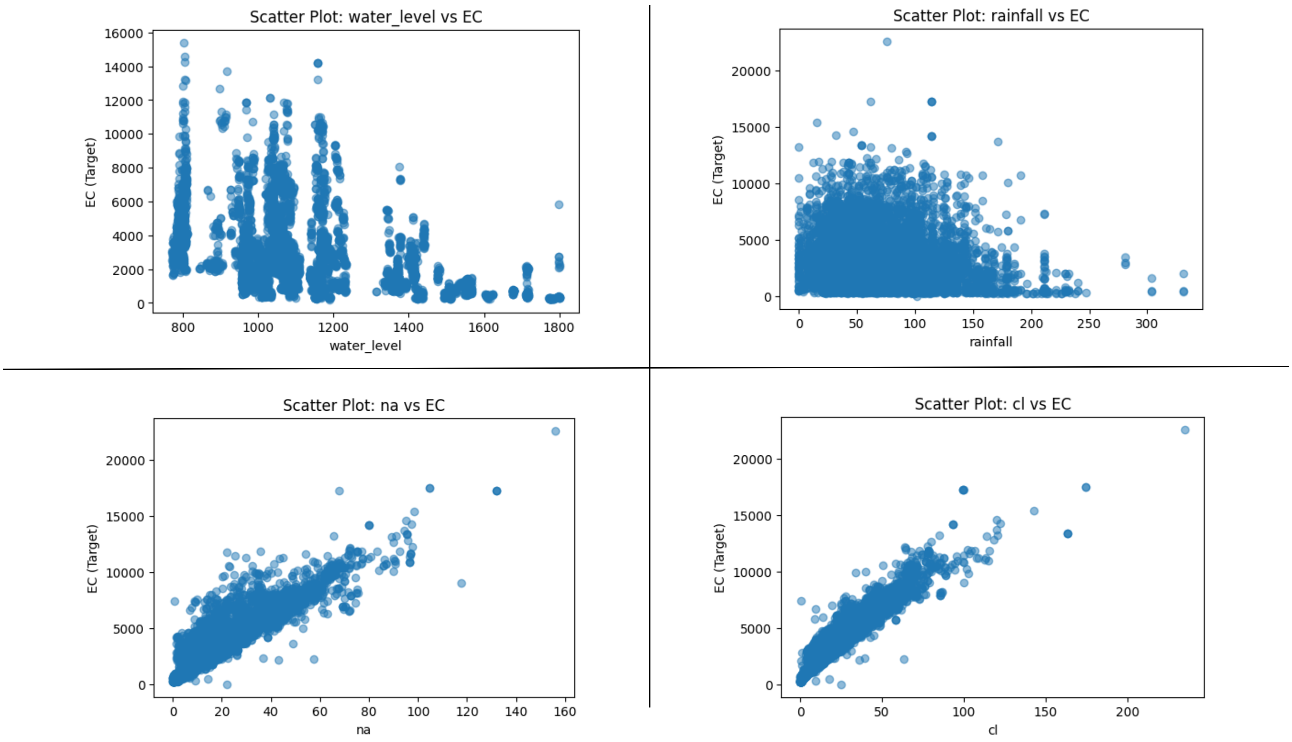

Figure 2. Scatter plots of rainfall, water-level, Na⁺, and Cl⁻ versus EC

3.2. Design And Metrics

In conventional machine learning procedures, data is divided into training and test subsets through a random Train/Test split. While this strategy is fine for time-independent and space-independent data, it is not suitable for the present research since data possess strong temporal as well as regional continuity. Random partitioning can create overlapping data within subsets, i.e., instances where data of the same year are present in both training and test sets. Overlapping causes data leakage, which amplifies and warps model performance estimates. To address this problem, a Walk-Forward Validation procedure was applied. In the sequential approach, the model is initially trained on first-year data and tested on the subsequent year. In the last of each iteration, the year to be predicted is appended to the training set and the model is retrained to predict the subsequent year. Iterations are repeated until the final year in the available dataset. Preservation of temporal order and hence simulating real forecasting situations where models predict based on past data, is the best advantage of this approach. Moreover, Walk-Forward Validation also makes it possible to assess performance stability over time, which is essential for temporally structured hydrological data. Walk-Forward Validation was applied in the present study for the period 2003–2024 (1382–1403). The model was first trained on 2003–2011 (1382–1390) data and tested for 2012 (1391). After each test, the most recently tested year was added to training data, and the procedure was repeated iteratively through 2023 (1402).

To put together model performance over decades, an exponentially weighted mean was calculated, giving higher weight to more recent years (e.g., 2020–2023). This is because more recent hydrological trends are more important when it comes to forecasting groundwater. The weighting has been done as follows:

$$W_i = \frac{e^{\propto i}}{\sum_{j=1}^{n} e^{\propto j}} \tag{1}$$

where W_i is the weight for year i, α is the smoothing coefficient that controls the rate of exponential growth, and n is the total number of years considered.

This configuration ensures that more up-to-date data have higher influence on the aggregated performance metric, aligning model assessment with the ever-changing nature of hydrological systems. In general, Walk-Forward Validation and weighted averaging usage precluded data leakage, preserved temporal realism, and created a strong and legitimate assessment of predictive performance.

3.3. Models

As discussed in Section 3.1, exploratory data analysis indicated strong linear dependencies between EC and several chemical indicators (Na^+, Cl^-, TH), leading to their exclusion. The remaining features, semi-annual rainfall, water-level index, and spatial coordinates (UTM X, Y), exhibited complex nonlinear relationships with EC as seen in Figure 2. Therefore, the modeling phase focused on nonlinear regression methods capable of capturing these intricate dependencies. Pure linear methods were therefore deemed inappropriate. Instead, the study employed a set of nonlinear machine learning regressors that are well-suited to identify sophisticated, multi-dimensional patterns.

The final models employed were:

- **Decision Tree for Regression (DT)**: A tree-based method that partitions the input space recursively to fit simple prediction models in each region, capturing nonlinear relationships efficiently [۹].
- **Random Forest for Regression (RF)**: An ensemble of decision trees that reduces overfitting by averaging predictions across multiple trees, enhancing robustness and generalization [۱۰].
- **Gradient Boosting Regression (GBR)**: Sequentially builds an ensemble of weak learners (trees), where each new tree corrects errors from the previous ones, providing high predictive accuracy [۱۱].
- **Support Vector Regression (SVR)**: A kernel-based method that maps inputs into high-dimensional spaces to fit a regression function while controlling model complexity [۱2].
- **K-Nearest Neighbors for Regression (KNN)**: An instance-based approach that predicts outputs based on the average of the nearest neighbors in the feature space, capturing local patterns [۱3].
- **XGBoost for Regression (XGB)**: A scalable and regularized gradient boosting framework, optimized for speed and performance [14].
- **LightGBM for Regression (LGBM)**: A highly efficient gradient boosting decision tree algorithm using histogram-based techniques for faster computation [15].
- **Multilayer Perceptron (MLP)**: A feedforward neural network capable of approximating complex nonlinear mappings between inputs and outputs [16].

These models represent a broad set of nonlinear learning processes, spanning from instance-based (KNN) and kernel-based (SVR) to ensemble and gradient boosting solutions (RF, GBR, XGB, LGBM), and finally to neural architectures (MLP). Such diversity allowed the study to entirely examine performance stability and generalization across varying model complexities.

Hyperparameters were tuned through grid search for every step in the Walk-Forward Validation scheme outlined above. Model performance was tracked using three complementary regression metrics:

- Root Mean Squared Error (RMSE)
- Root Mean Absolute Error (RMAE)
- The Coefficient of Determination (R^2)

RMSE emphasizes larger errors by squaring them, RMAE provides a scale-independent measure of the average absolute deviation relative to the observed mean, and R^2 quantifies the proportion of variance in the target variable explained by the model. To make the long-term comparison realistic, yearly derived values were merged through an exponentially weighted average, assigning greater weight to more recent years for greater relevance and consistency with the evolving hydrologic behavior of the region. In this manner, the final assessment included both historic reliability and modern forecasting relevance.

4. Experimental Results

The predictive performance of the evaluated machine learning models was examined under a walk-forward validation framework to capture year-to-year variations. Figures 3–5 illustrate these results, showing R^2, RMAE, and RMSE, respectively, across the study period. These visualizations provide a clear sense of both the models' predictive accuracy and their temporal stability.

4.1. Random Forest and XGBoost

Random Forest (RF) and XGBoost (XGB) clearly outperformed all other models. RF achieved a weighted mean R^2 of 0.846 with mean RMSE and RMAE of 821 and 21, respectively, while XGB had a weighted mean R^2 of 0.831, RMSE of 874, and RMAE of 22. Both models consistently captured the dominant trends in groundwater parameters across most years, often exceeding R^2 values of 0.9 as seen in Figure 3. Their success stems from their ability to model nonlinear relationships among precipitation, piezometric levels, and chemical variables, as well as to handle multicollinearity. RF's ensemble of decision trees provides interpretable feature importance rankings, whereas XGB's gradient-based optimization with regularization supports strong generalization. The low and stable RMSE and RMAE values as seen in Figures 4 and 5, indicate that their predictions were closely aligned with observed measurements.

4.2. Gradient Boosting and LightGBM

Gradient Boosting (GB) and LightGBM (LGBM) also delivered reliable predictions, although with slightly lower accuracy. GB exhibited a weighted mean R^2 of 0.688, while LGBM achieved 0.818. Both methods remained relatively stable under the walk-forward scheme, demonstrating robustness to data drift and temporal shifts in groundwater regimes. LGBM's leaf-wise tree growth and histogram-based learning allowed faster convergence without substantial loss of accuracy, making it computationally efficient for large-scale hydro-environmental datasets.

4.3. Decision Tree and K-Nearest Neighbors (KNN)

Decision Tree (DT) demonstrated moderate predictive performance, with a weighted mean R^2 of 0.808, RMSE of 932, and RMAE of ۲1. While DT successfully captured local patterns and nonlinear interactions among precipitation, piezometric levels, and groundwater chemistry, its performance varied noticeably across years as seen in Figures 3–5 . In some years, R^2 exceeded 0.95, indicating that the model could accurately follow dominant trends, but in other years, particularly 1397–1399, R^2 dropped below 0.7, reflecting the model's sensitivity to temporal variability and the limited size of training data per year. This fluctuation suggests that single-tree models can overfit specific patterns while struggling with broader temporal shifts. KNN showed moderate performance, with a weighted mean R^2 of 0.812. While it could detect local patterns effectively, its accuracy fluctuated across years due to sensitivity to the number of neighbors and feature scaling as seen in Figures 3–5. This variability indicates that KNN may struggle with long-term trends and nonlinear interactions when data are sparse or unevenly distributed.

4.4. Multi-Layer Perceptron (MLP) and Support Vector Regression (SVR)

The Multi-Layer Perceptron (MLP) neural network demonstrated limited predictive power (weighted mean $R^2 = 0.107$, RMSE = 2100, RMAE = 39). Despite the theoretical capacity of neural networks to capture nonlinear functions, the relatively small dataset and limited features restricted the model's ability to generalize. Support Vector Regression (SVR) performed poorly, with negative weighted mean R^2 (-0.064), RMSE of 2293, and RMAE of 40. The negative R^2 indicates that SVR's predictions were worse than simply using the mean of observed values. This poor performance likely resulted from insufficient training data per year, sensitivity to kernel choice, and an inability to capture temporal variations effectively.

4.5. Overall Observations

Overall, the results emphasize the superiority of ensemble-based methods, particularly RF and XGB, for predicting unconventional groundwater quality using heterogeneous, multi-source datasets. Their consistent accuracy and stability under walk-forward validation indicate that these models not only fit historical data well but also retain predictive power when applied to unseen years, a critical requirement for sustainable water management. Figures 3–5 collectively demonstrate these findings, highlighting both the absolute performance of each model and their year-to-year stability.

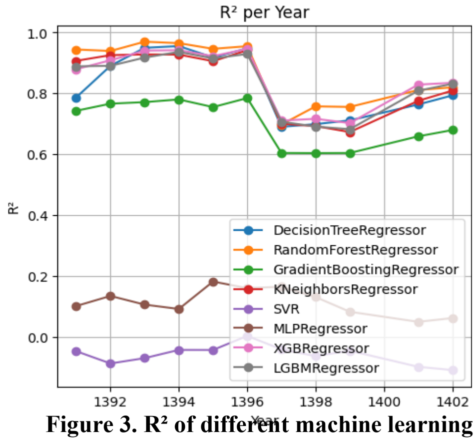

Figure 3. R² of different machine learning models across years

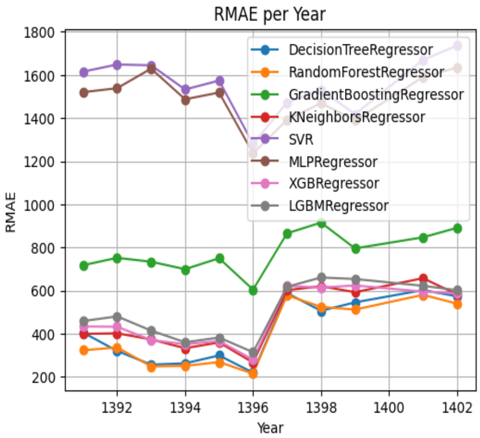

Figure 4. RMAE of different machine learning models across years

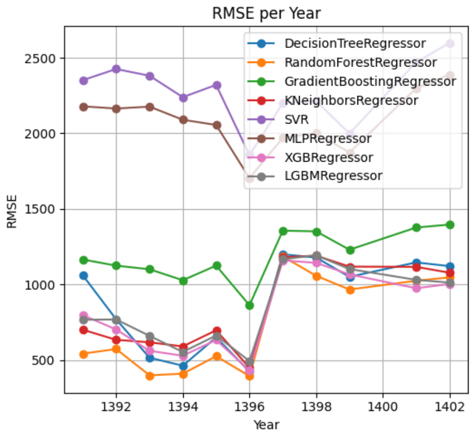

Figure 5. RMSE of different machine learning models across years

5. DISCUSSION

These findings reinforce the idea that ensemble-based machine learning methods are particularly well suited for groundwater anomaly prediction in complex hydrological systems. The outstanding performance of Random Forest and XGBoost demonstrates their robustness in handling multivariate, noisy data while effectively capturing nonlinear relationships among meteorological and hydro chemical variables. Random Forest's ability to reduce variance through averaging and XGBoost's gradient-based optimization with regularization both contribute to their superior predictive power and generalization across different years. These results highlight that ensemble diversity and feature interaction modeling are more influential than simply increasing model complexity.

In contrast, the relatively weak performance of SVR and MLP suggests that higher model sophistication does not necessarily translate to better accuracy when the dataset is limited or temporally imbalanced. SVR's sensitivity to kernel choice and MLP's dependence on extensive training data likely constrained their capacity to model inter-annual variations. This aligns with earlier findings in hydro-environmental modeling, where kernel-based or deep methods often underperform in small or noisy datasets due to overfitting and poor extrapolation capabilities. The moderate results of KNN and Gradient Boosting further emphasize that predictive success in groundwater studies depends strongly on how well the algorithm structure matches the statistical properties of the data.

What makes this work distinct is its focus on realistic forecasting. By using walk-forward validation and integrating different sources of hydrological information, the framework captures how models behave over time rather than just in static historical splits. This makes it a practical and adaptable approach for predicting groundwater anomalies in other arid and semi-arid regions, where conditions change from year to year and reliable forecasting is essential for sustainable management.

6. CONCLUSION

This study presented a comparative analysis of eight machine learning models, Decision Tree, Random Forest, Gradient Boosting, XGBoost, LightGBM, K-Nearest Neighbors (KNN), Support Vector Regression (SVR), and Multi-Layer Perceptron (MLP), for predicting groundwater quality anomalies in Semnan Province, Iran. By integrating precipitation, piezometric, and chemical indicators under a walk-forward validation framework, the analysis enabled a consistent evaluation of model performance over successive years, reflecting their temporal generalization capacity.
Among all models, Random Forest (R^2 = 0.846, RMSE = 821.06, RMAE =21) and XGBoost (R^2 = 0.831, RMSE = 873.73, RMAE = 22) achieved the highest performance, showing excellent accuracy and consistency across years. Their ensemble structures enabled effective handling of nonlinear relationships and multicollinearity among hydrological and chemical predictors.
In contrast, MLP (R^2 = 0.107) and SVR (R^2 = –0.064) exhibited the weakest predictive performance. Even after testing different kernels and hyperparameters, their limited dataset size and few input features hindered their ability to capture temporal patterns effectively.
Overall, the results underscore the strong advantage of ensemble-based algorithms, particularly Random Forest and XGBoost, in forecasting groundwater quality from complex and heterogeneous datasets. The proposed framework shows that integrating diverse hydrological indicators with a temporally consistent validation approach yields accurate and interpretable predictions, providing a practical foundation for sustainable groundwater management in arid and semi-arid regions.

7. Acknowledgement

We would like to express our gratitude to the Research Department of the Semnan Regional Water Company for providing us with dataset on unconventional water, which served as the foundation of this study.

8. References

[1] Karimidastenaei, Z.; Avellán, T.; Sadegh, M.; Kløve, B.; Torabi Haghighi, A., "Unconventional water resources: Global opportunities and challenges", *Science of the Total Environment*, Vol. 827, Article 154429, 2022.

[2] Zhu, M.Y.; Wang, J.W.; Yang, X.; Zhang, Y.; Zhang, L.; Ren, H.; Wu, B.; Ye, L., "A review of the application of machine learning in water quality evaluation", *Eco-Environment & Health*, Vol. 1, pp. 107–116, 2022.

[3] Mousavimehr, S. M.; Kavianpour, M. R., "Estimating groundwater levels in Tehran Province using ensemble learning algorithms", *Contributions of Science and Technology for Engineering*, Vol. 2, No. 1, pp. 51-63, 2025.

[4] Feng, L.; Ghorbani, H.; Radwan, A. E., "Predicting groundwater level using traditional and deep machine learning algorithms", *Frontiers in Environmental Science*, Vol. 12, Article 1291327, 2024.

[5] Yadav, A.; Raj, A.; Yadav, B., "Enhancing local-scale groundwater quality predictions using advanced machine learning approaches", *Journal of Environmental Management*, Vol. 370, Article 122903, 2024.

[6] Yan, X.; Zhang, T.; Du, W.; Meng, Q.; Xu, X.; Zhao, X., "A Comprehensive Review of Machine Learning for Water Quality Prediction over the Past Five Years", *Journal of Marine Science and Engineering*, Vol. 12, Article

[۷] Goorabi, A.; Karimi, M.; Yamani, M.; Perissin, D., "Land subsidence in Isfahan metropolitan and its relationship with geological and geomorphological settings revealed by Sentinel-1A InSAR observations", *Journal of Arid Environments*, Vol. 181, Article 104238, 2020.

[۸] Zhi, W.; Appling, A. P.; Golden, H. E.; Podgorski, J.; Li, L., "Deep learning for water quality", *Nature Water*, Vol. 2, pp. 228–241, 2024.

[۹] Quinlan, J. R., "Induction of decision trees", *Machine Learning*, Vol. 1, No. 1, pp. 81–106, 1986.

[۱۰] Breiman, L., "Random forests", *Machine Learning*, Vol. 45, No. 1, pp. 5–32, 2001.

[۱۱] Friedman, J. H., "Greedy function approximation: A gradient boosting machine", *Annals of Statistics*, Vol. 29, No. 5, pp. 1189–1232, 2001.

[1۲] Drucker, H.; Burges, C. J. C.; Kaufman, L.; Smola, A.; Vapnik, V., "Support vector regression machines", *Advances in Neural Information Processing Systems*, Vol. 9, pp. 155–161, 1996.

[1۳] Altman, N. S., "An introduction to kernel and nearest-neighbor nonparametric regression", *The American Statistician*, Vol. 46, No. 3, pp. 175–185, 1992.

[1٤] Chen, T.; Guestrin, C., "XGBoost: A scalable tree boosting system", *Journal of Machine Learning Research*, Vol. 17, pp. 1–5, 2016.

[1٥] Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Xu, Y., "LightGBM: A highly efficient gradient boosting decision tree", *Advances in Neural Information Processing Systems*, Vol. 30, 2017.

[1٦] Hornik, K.; Stinchcombe, M.; White, H., "Multilayer feedforward networks are universal approximators", *Neural Networks*, Vol. 2, No. 5, pp. 359–366, 1989.