**Literature Review: Machine Learning and Deep Learning for Groundwater Level Prediction**

The sustainable management of groundwater resources is increasingly critical in the face of climate variability, population growth, and intensifying agricultural demands. Accurate prediction of groundwater levels (GWL) is essential for informed decision-making in water resource management, particularly in regions vulnerable to drought or overexploitation. Recent advances in computational capacity and the availability of environmental data have positioned *machine learning* (ML) and *deep learning* (DL) as powerful tools for modeling complex hydrological systems. Unlike traditional hydrological models, which are often based on explicit physical equations, ML and DL methods are very effective at identifying non-linear patterns and relationships within large datasets, offering robust predictive capabilities (Aggarwal, 2018; Sit et al., 2020).

*Machine learning* encompasses a subset of artificial intelligence focused on developing algorithms that learn from data to make predictions or classifications. ML models are broadly categorized into three types:

*Table 1- Overview of the three main types of machine learning and examples of commonly used models (Alpaydin, 2020)*

| Type of Learning | Definition | Common Models |
|---|---|---|
| **Supervised Learning** | The model is trained on a labeled dataset, where both input features and expected outputs are known. It learns to map inputs to outputs for prediction. | Random Forest (RF), Support Vector Machines (SVM), Decision Trees (DT), Gradient Boosting (GBM) |
| **Unsupervised Learning** | The model is trained on data without labels. It identifies hidden patterns or structures, such as clusters or principal components, within the dataset. | K-means Clustering, Principal Component Analysis (PCA) |
| **Reinforcement Learning** | The model learns through interaction with an environment by receiving rewards or penalties for its actions. It optimizes decision-making over time. | Q-learning, Policy Gradient Methods, Deep Q-Networks (DQN) |

The implementation of ML models typically follows a structured pipeline composed of several critical steps. These include data splitting (into training and testing sets), preprocessing, feature engineering (encompassing feature extraction and selection), model training, and subsequent validation or testing, a process that has been widely used in agricultural and environmental studies to ensure robust model performance (Figure 1).

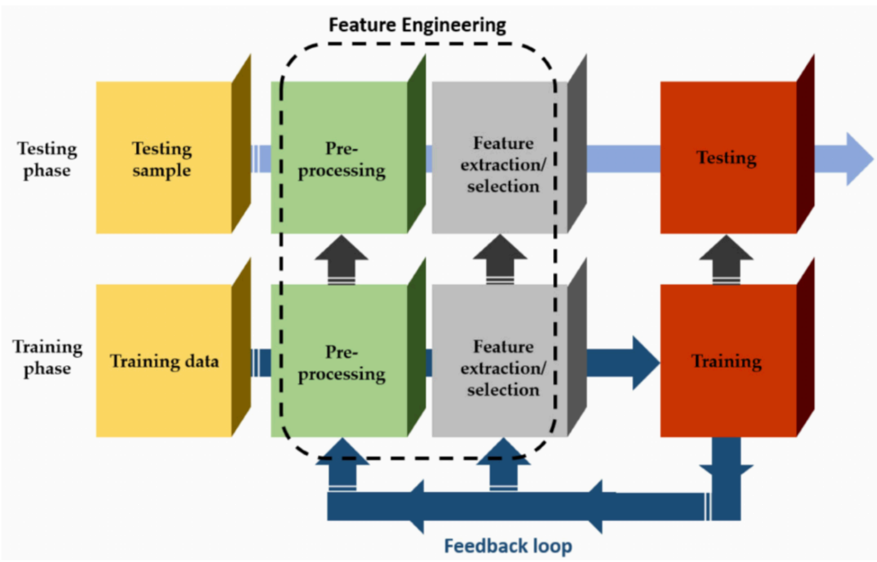


*Figure 1- General workflow for developing machine learning algorithms (Benos et al, 2021)*

Among machine learning approaches, deep learning has emerged as a powerful subfield focused on the use of artificial neural networks with multiple layers, capable of modeling intricate patterns in high-dimensional data (Aggarwal, 2018). While DL models can also be supervised, unsupervised, or based on reinforcement learning, they stand out for their ability to automatically learn complex features directly from raw data. This makes them especially useful for modeling spatial and temporal dependencies in environmental systems, as demonstrated in various hydrological applications (Sit et al., 2020).

Key DL architectures include:

- **Artificial Neural Networks (ANNs)**: Effective for capturing complex, non-linear relationships in data.
- **Convolutional Neural Networks (CNNs)**: Specialized in detecting spatial and temporal patterns, particularly in noisy datasets.
- **Long Short-Term Memory (LSTM) networks**: Designed for modeling long-term dependencies in time-series data.
- **Non-linear Autoregressive Networks with Exogenous Inputs (NARX)**: Dynamic networks that incorporate external variables, such as precipitation or land use, for enhanced predictions.

The choice of ML or DL model depends on factors such as data availability, temporal resolution, and the hydrological context. In groundwater studies, these methods have shown promise in capturing interactions between variables like precipitation, evapotranspiration, and hydraulic operations, enabling accurate GWL predictions (Pham et al., 2022; Wunsch et al., 2021).

**Applications of ML/DL in Groundwater Studies**

Groundwater level forecasting is critical for sustainable resource management, particularly in agricultural regions subject to climatic variability and intensive irrigation. ML and DL models have demonstrated significant potential in developing robust predictive tools, provided they are supported by high-quality, comprehensive datasets.

In South Korea, a study within the Four Major Rivers Project examined GWL prediction near the Baekje Weir using five ML algorithms: Random Forest (RF), Artificial Neural Network (ANN), Support Vector Regression (SVR), Gradient Boosting (GB), and Extreme Gradient Boosting (XGBoost). The XGBoost model outperformed others in both training and testing phases, highlighting the influence of variables such as distance to the weir, seasonality, and hydraulic management operations (e.g., full or partial gate openings). These findings underscore the applicability of ML in systems governed by hydraulic infrastructure (Lee et al., 2022).

In Iran, Feng et al. (2024) compared traditional and DL algorithms for GWL prediction in Izeh, using historical data from 2002 to 2022. The study evaluated Support Vector Machine (SVM), Decision Trees (DT), Random Forest (RF), Convolutional Neural Networks (CNN), and Generative Adversarial Networks (GAN). The CNN model achieved superior performance, with a coefficient of determination ($R^2$) of 0.9948 and a root mean square error (RMSE) of 0.0558. The integration of signal processing techniques, such as Wavelet Transform, enhanced the model's ability to capture complex hydrological patterns, demonstrating the strength of DL in intricate systems.

In Germany, Wunsch et al. (2021) compared three neural network architectures—LSTM, CNN, and NARX—for GWL forecasting based on time-series data. The results indicated that LSTM and NARX models excelled in short-term predictions (up to one month), while CNNs performed better for medium- to long-term forecasts. These findings emphasize the importance of aligning model selection with the temporal scope of the prediction, informing methodologies for diverse planning horizons.

In Bangladesh, Pham et al. (2022) tested seven ML algorithms to predict GWL in drought-prone regions, using meteorological and hydrological data from 1981 to 2017. The models included Random Forest (RF), Random Tree (RT), Support Vector Machine (SVM), M5P, REP Tree, Decision Stump, and Locally Weighted Learning Regression (LWLR). Ensemble methods, specifically Bagging-RT and Bagging-RF, delivered the best predictive accuracy, with notably low RMSE values. This study highlights the efficacy of ML in supporting water resource policies in vulnerable regions.

In Morocco, El Bilali et al. (2021) investigated groundwater quality forecasting in the Berrechid aquifer, focusing on its suitability for irrigation. The study applied Support Vector Regression (SVR), Random Forest (RF), Artificial Neural Network (ANN), and Adaptive Boosting (Adaboost) to predict indicators such as total dissolved solids (TDS), sodium percentage (Na%), sodium adsorption ratio (SAR), and potential salinity (PS). Both SVR and RF models achieved high correlation and low RMSE, proving reliable for mitigating salinization risks and supporting agricultural water management.

Additionally, Müller et al. (2021) explored the predictive performance of deep learning models—including Multilayer Perceptron (MLP), Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM)—for GWL forecasting in Butte County, California, using daily observations from 2010 to 2018. The study utilized meteorological and hydrological variables such as precipitation, streamflow, and temperature. Through surrogate-based hyperparameter optimization using Gaussian Processes (GP) and Radial Basis Functions (RBF), the MLP model consistently outperformed the more complex DL architectures in terms of accuracy and training efficiency. This finding underscores that simpler models, when properly optimized, may offer superior performance in data-limited or noisy hydrological environments.

*Table 2 - Summary of the literature review regarding the prediction of the groundwater level using machine learning models*

| Reference | Variables | Period | Method | Comments |
|---|---|---|---|---|
| Lee et al. (2022) | Distance to the weir, seasonality, hydraulic management operations | Not specified | RF, ANN, SVR, GB, XGBoost | XGBoost achieved the best performance; study influenced by hydraulic structures. |
| Feng et al. (2024) | Historical GWL data | 2002–2022 | SVM, DT, RF, CNN, GAN | CNN achieved the best performance ($R^2$ = 0.9948); used Wavelet Transform. |
| Wunsch et al. (2021) | GWL time series | Not specified | LSTM, CNN, NARX | LSTM and NARX performed better for short-term forecasts; CNN for medium- to long-term forecasts. |
| Pham et al. (2022) | Meteorological and hydrological data | 1981–2017 | RF, RT, SVM, M5P, REP Tree, Decision Stump, LWLR; Bagging-RT, Bagging-RF | Bagging models performed best; applied to drought-prone regions. |
| Müller et al. (2019) | Precipitation, streamflow, temperature | 2010–2018 | MLP, RNN, CNN, LSTM + optimization with RBF and GP | MLP outperformed other models after surrogate-based optimization; study in Butte County, California. |
| El Bilali et al. (2021) | TDS, Na%, SAR, PS | Not specified | SVR, RF, ANN, Adaboost | SVR and RF achieved high correlation and low RMSE; study in Berrechid aquifer. |

### Synthesis and Implications

The reviewed studies demonstrate that the effectiveness of ML and DL models for GWL and groundwater quality prediction depends on the hydrological context, data availability, and temporal resolution. Algorithms like XGBoost, CNN, and ensemble methods (e.g., Bagging-RF) consistently outperform others in specific scenarios, while the choice between short- and long-term forecasting influences model suitability. A common theme across all studies is the necessity of well-structured, high-quality datasets, a principle guiding the current research on the Tejo Vulnerable Zone, where a robust pipeline integrates historical piezometric, precipitation, and groundwater quality data.

---

### References


Aggarwal, C. C. (2018). *Neural networks and deep learning: A textbook*. Springer.

Alpaydin, E. (2020). *Introduction to machine learning* (4th ed.). MIT Press.

Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. *Sensors*, *21*(11), 3758. https://doi.org/10.3390/s21113758

El Bilali, A., Taleb, A., & Brouziyne, Y. (2021). Groundwater quality forecasting using machine learning algorithms for irrigation purposes. *Agricultural Water Management*, *245*, 106625. https://doi.org/10.1016/j.agwat.2020.106625

Lee, H., Kim, T., & Yoon, Y. (2022). Groundwater level forecasting using machine learning: A case study of the Baekje Weir in South Korea. *Water Resources Research*, *58*(8), e2022WR032779. https://doi.org/10.1029/2022WR032779

Müller, J., Park, J., Sahu, R., Varadharajan, C., Arora, B., Faybishenko, B., & Agarwal, D. (2021). Surrogate optimization of deep neural networks for groundwater predictions. *Journal of Global Optimization*, *81*, 203–231. https://doi.org/10.1007/s10898-020-00912-0

Pham, Q. B., Kumar, M., Di Nunno, F., Elbeltagi, A., Granata, F., Islam, A. R. M. T., Talukdar, S., Nguyen, X. C., Ahmed, A. N., & Tran Anh, D. (2022). Groundwater level prediction using machine learning algorithms in a drought-prone area. *Neural Computing and Applications*, *34*, 10751–10773. https://doi.org/10.1007/s00521-022-07009-7

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

Sit, M., Demiray, B. Z., Xiang, Z., Ewing, G. J., Sermet, Y., & Demir, I. (2020). A comprehensive review of deep learning applications in hydrology and water resources. *Journal of Hydrology*, *587*, 124983. https://doi.org/10.1016/j.jhydrol.2020.124983

Wunsch, A., Liesch, T., & Broda, S. (2021). Groundwater level forecasting with artificial neural networks: A comparison of long short-term memory (LSTM), convolutional neural networks (CNNs), and non-linear autoregressive networks with exogenous input (NARX). *Hydrology and Earth System Sciences*, *25*(3), 1671–1687. https://doi.org/10.5194/hess-25-1671-2021