Agri Articles
(e-Magazine for Agricultural Articles)
Volume: 06, Issue: 01 (JAN-FEB, 2026)
Available online at http://www.agriarticles.com

Machine Learning Approaches for Groundwater Quality Assessment in Agricultural Areas: A Focused Review

*Rebek Lalhmangaihsangi
M.Sc. Scholar, Department of Soil Science & Agricultural Chemistry,
College of Agriculture, OUAT, Bhubaneswar, Odisha, India-751003
Corresponding Author's email: lalhmangaihsangirebek@gmail.com

Machine learning (ML) has emerged as a powerful tool for assessing groundwater quality in agricultural landscapes, offering improved accuracy, efficiency, and predictive capability compared to traditional statistical methods. Groundwater contamination from intensive farming, fertilizer leaching, and pesticide application remains a critical environmental issue, necessitating reliable evaluation and forecasting frameworks. Recent studies integrate ML algorithms with water quality indices (e.g., WQI, IWQI) and geospatial techniques (GIS) to enable robust classification, spatial mapping, and prediction of groundwater quality. Models including Random Forest (RF), Support Vector Regression (SVR), Extreme Gradient Boosting (XGBoost), Artificial Neural Networks (ANN), and K-Nearest Neighbours (KNN) have demonstrated strong performance in diverse hydrological contexts, though challenges related to data scarcity, variability, and interpretability persist. This paper presents a focused review of recent applications of machine learning for groundwater quality assessment in agricultural systems, highlighting representative case studies and emerging methodological trends.

Keywords: Groundwater quality; Agricultural landscapes; Machine learning; Water Quality Index (WQI); Irrigation Water Quality Index (IWQI); Spatial mapping; GIS; Nitrate leaching; Heavy metals; Risk assessment; XGBoost; Support Vector Regression; Artificial Neural Networks; India.

Introduction: The Silent Crisis Beneath Our Fields

Agricultural areas, particularly in arid and semi-arid regions, depend heavily on groundwater to sustain food production and rural livelihoods. Groundwater quality underpins sustainable agriculture, ecosystem health, and human consumption. However, increased pumping, fertilizer use, and land-use change have degraded groundwater quality through salinization, nitrate contamination, and the accumulation of other dissolved ions.

Nitrate leaching from fertilizers, pesticide residues, and the mobilization of geogenic contaminants such as fluoride, arsenic, iron, and manganese are altering the hydrochemical balance of aquifers. These changes pose direct risks to drinking-water safety, crop productivity, soil health, and human well-being. Traditional evaluation methods using indices such as the Water Quality Index (WQI) or Irrigation Water Quality Index (IWQI) provide useful classification of water suitability but often lack predictive power and fail to capture complex non-linear relationships among hydrochemical parameters. Conventional monitoring based on periodic sampling and laboratory analysis is expensive and often too sparse in space and time to support proactive management.

Machine Learning as a Data-Driven Framework for Groundwater Quality Prediction

Machine learning offers a data-driven alternative capable of modelling complex, multi-dimensional interactions without stringent assumptions about underlying distributions. Supervised learning algorithms — including ensemble methods (RF, XGBoost), kernel-based methods (SVR, SVM), and artificial neural networks (ANN) — have been increasingly applied to groundwater quality modelling. These methods leverage historical hydrochemical data to classify current water quality and predict future trends, often integrated with spatial information from GIS for comprehensive mapping. A recent bibliometric review reports rapid growth in ML applications for groundwater quality assessment and modelling, with global research trends emphasizing supervised learning models and hybrid frameworks that combine ML with geostatistical techniques. Rather than providing an exhaustive survey, this review concentrates on selected peer-reviewed studies that demonstrate how machine learning can enhance groundwater quality assessment in agricultural landscapes.

Representative Case Studies

1. Prediction of Irrigation Water Quality Index (IWQI) — Algeria

In the Naama region of southwest Algeria, researchers evaluated groundwater quality for irrigation using IWQI and multiple ML models (XGBoost, SVR, KNN). A dataset of 166 groundwater samples was analyzed for major hydrochemical parameters, with IWQI categories ranging from excellent to unsuitable. XGBoost achieved high predictive accuracy ($R \approx 0.98$, $RMSE \approx 2.83$) while SVR using reduced input features (Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}) performed comparably ($R \approx 0.99$, $RMSE \approx 2.69$). These results illustrate that ML models can reliably capture water quality suitability despite limited parameter sets.

2. Fluoride-linked groundwater risk assessed with EWQI, IWQI and geochemical modelling- Uttar Pradesh

Jena et al. (2023) assessed groundwater quality in the Tundla block of Uttar Pradesh to evaluate its suitability for drinking and irrigation. Fifty groundwater samples were analyzed for major cations, anions, and fluoride, and multiple indices — including the entropy-weighted Water Quality Index (EWQI) and Irrigation Water Quality Index (IWQI) — were applied to classify water suitability. The study further employed PHREEQC geochemical modelling to understand mineral saturation states and the geochemical processes controlling groundwater chemistry. Results indicated that a considerable proportion of samples exhibited elevated fluoride concentrations, leading to moderate drinking-water risk, particularly for children, while irrigation suitability ranged from moderate to marginal. The combined use of water quality indices and geochemical simulations provided deeper insight into contamination sources and demonstrated the value of integrated assessment frameworks for groundwater management in agricultural regions.

3. Groundwater WQI — Semi-Arid Industrial Region

Kumar et al. (2023) investigated groundwater quality in a semi-arid industrial corridor using machine-learning models to predict Water Quality Index (WQI) and water quality classifications. Random Forest (RF), gradient boosting, decision tree, and K-nearest neighbour's algorithms were trained on hydrochemical and heavy-metal datasets. Their results demonstrated that the RF model achieved the highest performance, with classification accuracy close to 97% and strong WQI prediction ($R^2 \approx 0.91$), highlighting the capability of ML approaches to outperform traditional analytical techniques in complex, contaminated environments.

4. ML + GIS for Irrigation Suitability — South India

Singh et al. (2025) applied machine-learning techniques integrated with GIS in the Arjunanadi River basin to evaluate irrigation water suitability. Groundwater quality parameters were spatially analysed, and Support Vector Machine (SVM) and K-nearest neighbours' models were developed. The SVM model exhibited superior predictive accuracy

($R^2 \approx 0.97$), demonstrating that incorporating spatial information with ML enhances the reliability of irrigation suitability assessment across large agricultural landscapes.

5. Tumkur district, Karnataka: WQI forecasting using SVM and other ML models

Doddabasappaar et al. (2024) developed an ML-based framework to forecast the groundwater Water Quality Index (WQI) in Tumkur district, Karnataka, India. The dataset comprised physicochemical parameters (total hardness, pH, alkalinity, turbidity, chloride, total dissolved solids, and electrical conductivity) from groundwater sources, which were converted to WQI and then split into 80% training and 20% testing subsets. Several supervised algorithms — support vector machines (SVM), regression trees, linear regression, and neural networks — were compared. SVM and linear regression provided the best performance, with R^2 values of about 0.96 (training) and 0.99 (testing), demonstrating that relatively small monitoring datasets can still yield highly accurate WQI forecasts when appropriate ML techniques are applied.

Challenges and Limitations

Despite demonstrated success, several challenges remain:

- **Data Availability and Quality:** ML models require sufficient, high-quality datasets. In agricultural regions where monitoring is infrequent, data scarcity can limit model generalizability and predictive robustness.
- **Parameter Selection:** Hydrochemical datasets often contain many variables; identifying the most informative subset without overfitting remains non-trivial and may require feature selection or dimensionality reduction techniques.
- **Model Interpretability:** Complex models such as ANN and ensemble methods may act as "black boxes," reducing explainability for decision-makers. Advances in interpretable ML and explainable AI (XAI) are needed.
- **Spatial and Temporal Variability:** Groundwater quality varies seasonally and spatially with rainfall, irrigation practices, and land use. ML models must incorporate spatio-temporal dynamics to improve forecasting accuracy.
- **Integration with Policy:** Translating ML predictions into actionable water management strategies requires alignment with regulatory frameworks, community risk perception, and resource constraints.

Conclusion

Machine learning approaches have significantly enhanced groundwater quality assessment in agricultural settings by enabling accurate prediction, classification, and spatial mapping. Supervised learning models such as RF, SVR, and XGBoost outperform traditional statistical techniques and handle non-linear, multi-parameter datasets effectively. Integrating ML with geospatial tools further strengthens spatial decision support systems for sustainable water resource management. However, addressing challenges related to data limitations, model interpretability, and spatio-temporal variability is essential for broader adoption. Future research should focus on hybrid models, interpretable frameworks, and scalable solutions tailored to regional groundwater monitoring programs.

References

1. Wada, Y., van Beek, L. P. H., Sperna Weiland, F. C., Chao, B. F., Wu, Y. H., & Bierkens, M. F. P. (2012). Past and future contribution of global groundwater depletion to sea-level rise. *Geophysical Research Letters*, 39(9), L09402.
2. Gleeson, T., Befus, K. M., Jasechko, S., Luijendijk, E., & Cardenas, M. B. (2016). The global volume and age of groundwater. *Nature Geoscience*, 9(2), 161–167.
3. Shukla, A. K., & Behera, S. K. (2019). Nitrate pollution in groundwater: Sources, consequences and management strategies. *Indian Journal of Fertilizers*, 15(2), 102–119.
4. Hussein, E. E., et al. (2024). Groundwater quality assessment and irrigation water quality index prediction using machine learning algorithms in the Naama region, Algeria. *Water*, 16(2), 264.

5. Paneerselvam, B., et al. (2023). Machine-learning based evaluation of groundwater quality and human health risk in a South Indian agricultural region. *Chemosphere*, 322, 138271.
6. Raj, M. R. H., et al. (2025). Machine learning and GIS-based groundwater quality prediction for irrigation suitability in the Arjunanadi River basin, South India. *Computers and Electronics in Agriculture*, 218, 108298.
7. Raheja, H., et al. (2022). Prediction of groundwater quality indices using machine learning algorithms, including deep learning and gradient boosting. *Water Practice & Technology*, 17(1), 336–351.
8. Saraswat, A., Ram, S., Kouadri, S., Raza, M. B., Hombegowda, H. C., Kumar, R., & Jena, R. K. (2023). Groundwater quality, fluoride health risk and geochemical modelling for drinking and irrigation water suitability assessment in Tundla block, Uttar Pradesh, India. *Groundwater for Sustainable Development*, 20, 100991.
9. Doddabasappaar, A. K. T., Yogendra, B. E., Janardhan, P., & Siddegowda, P. N. (2024). *Machine Learning for Water Quality Index Forecasting*. *International Journal of Engineering and Technology Innovation*, 3, 43–53.
10. Kumar, A., Singh, R., & Sharma, P. (2023). *Machine learning models for groundwater quality index prediction in a semi-arid industrial region*. *Journal of Environmental Informatics and Analytics*, 7(2), 115–128.