"An Integrated Remote Sensing–GIS–Artificial Intelligence Framework for Groundwater Quality Forecasting in India"

Dr. Devendra Kumar Singh, PhD (Environmental Science)

Chief Technical Officer, Dept. of Cardiology, SGPGIMS Lucknow (U.P.), India.

Corresponding Author Email: dksingh.pgi@gmail.com | ORCID: https://orcid.org/xxxx-xxxx-xxxx-xxxx

How to Cite this Article:

Singh, D. K. (2026). An Integrated Remote Sensing–GIS–Artificial Intelligence Framework for Groundwater Quality Forecasting in India. International Journal of Creative and Open Research in Engineering and Management, <i>02</i>(8), 1-9. https://doi.org/10.55041/ijcope.v2i8.109

https://doi.org/10.55041/ijcope.v2i8.109

Abstract—

Background: Groundwater is essential for drinking water, agriculture, industry, and ecological sustainability in India; however, its quality is increasingly affected by geogenic processes, agricultural activities, urbanization, industrialization, and climate variability. Conventional monitoring is spatially limited and resource-intensive.

Objective: This study aimed to develop an integrated remote sensing-geographic information system-artificial intelligence framework for groundwater quality assessment and forecasting in India.

Methods: Groundwater-quality observations were integrated with satellite-derived variables, including land-use/land-cover, NDVI, NDWI and land-surface temperature, together with geological, climatic, terrain, hydrogeological, and anthropogenic datasets. Random Forest, XGBoost, Support Vector Regression and Artificial Neural Network models were proposed and evaluated using R^2, RMSE, MAE, and MAPE.

Results: The integrated framework enabled spatial prediction, identification of groundwater quality hotspots, assessment of environmental drivers, and comparison of Artificial Intelligence predictions with conventional interpolation methods.

Conclusion: The RS–GIS-AI stands for **Remote Sensing**, **Geographic Information System**, and **Artificial Intelligence**. This combined framework integrates satellite or aerial data capture, spatial mapping and analysis, and smart computer learning algorithms to study the Earth. framework provides a scalable decision-support approach for predictive groundwater monitoring, targeted sampling, contamination risk assessment, and sustainable groundwater management in India.

Keywords: Groundwater Quality; Remote Sensing; Geographic Information System (GIS), artificial intelligence, machine learning, spatial Modelling, groundwater forecasting.

Introduction

a. Background and Rationale

Importance of Groundwater Quality in India

Groundwater is a vital freshwater resource in India that; supports drinking-water supply, agriculture, industries, and rural livelihoods. Water quality directly affects public health, food production, and sustainability. Increasing contamination from fluoride, arsenic, nitrate, salinity, industrial effluents, agricultural chemicals, and untreated wastewater makes systematic groundwater quality assessments essential for sustainable water resource management in the region.

Challenges in Monitoring and Forecasting Groundwater Quality

Monitoring groundwater quality across India is challenging because aquifers are spatially heterogeneous, and groundwater conditions vary with geology, climate, land use, recharge, and anthropogenic activity. Conventional monitoring depends on field sampling and laboratory analyses, which can be costly, time-consuming, and spatially limited. Consequently, forecasting quality in unsampled areas is challenging.

Role of Remote Sensing, , and Artificial Intelligence in Environmental Science

Remote sensing provides spatial information on land use, vegetation, surface temperature, moisture, and environmental changes, whereas Geographical Information System integrates these datasets for spatial analysis and mapping. Artificial intelligence can identify complex nonlinear relationships between environmental factors and groundwater quality. Therefore, their integration offers an advanced framework for prediction, hotspot identification, monitoring, and sustainable groundwater management.

b. Objectives of the Study

Primary Aim of the Research

The primary aim of this study is to develop an integrated Remote Sensing–GIS–Artificial Intelligence framework for assessing and forecasting groundwater quality in India. The framework seeks to combine groundwater observations with spatial, environmental, climatic, hydrogeological, and land use information to generate reliable groundwater quality predictions and support evidence-based planning, monitoring; and management.

Specific Objectives of the Integrated Framework

The specific objectives are to integrate remote sensing and GIS datasets with groundwater quality observations, identify the major environmental and anthropogenic factors influencing groundwater quality, develop and compare AI-based predictive models, generate spatial groundwater quality and contamination risk maps, validate model performance using statistical indicators, and identify groundwater quality hotspots and vulnerable areas requiring enhanced monitoring and management.

Significance in the Context of Environmental Science

This study is significant because it demonstrates how geospatial technologies and artificial intelligence can strengthen the conventional methods of environmental monitoring. By integrating diverse datasets and predictive modelling, the framework can improve the understanding of groundwater quality dynamics, support the early identification of contamination risks, optimize monitoring resources, and provide spatial decision-support information. This can also contribute to climate-resilient water management and sustainable development in India.

a. Groundwater Quality in India

Review of Existing Studies on Groundwater Quality Issues in India

Groundwater-quality studies in India have identified fluoride, arsenic, nitrate, salinity, iron and other chemical contaminants as important regional concerns. The 2025 assessment analyzed **14,978 groundwater samples and; founding** that **71.7% complied with BIS permissible limits**, while **28.3% exceeded limits for one or more parameters**. Nitrate showed the highest exceedance (**20.7%**), followed by fluoride (**8.05%**) and electrical conductivity (7.23%), respectively.

Factors Affecting Groundwater Quality

Groundwater quality is controlled by the interactions among **geogenic, hydrogeological, climatic, and anthropogenic factors**. Geological weathering, mineral dissolution, aquifer characteristics, groundwater residence time, and geochemical reactions influence the natural concentrations of fluoride, As, and Fe. Agricultural fertilizers, sewage, industrial discharge, mining, urbanization, and excessive abstraction can further deteriorate the water quality. Recent reviews have confirmed that groundwater contamination varies regionally, according to geological and anthropogenic pressures. ,

Gaps in Current Research

Despite extensive monitoring, groundwater quality research in India remains constrained by spatially uneven observations, heterogeneous aquifer conditions; and limited continuous forecasting. Although the national CGWB dataset provides a strong monitoring foundation; conventional sampling remains point-based and cannot fully characterize unsampled locations. Recent studies have increasingly explored machine learning; however, there remains a need for **integrated frameworks** that combine groundwater chemistry, satellite-derived variables, hydrogeology, and climate and land-use information for scalable prediction and early warning applications.

b. Remote Sensing and GIS in Environmental Monitoring

Use of Remote Sensing and GIS in Environmental Science

Remote sensing provides repeated large-area observations of land use/land cover, vegetation, surface temperature, moisture, and surface water conditions, whereas GIS integrates these data with geological, climatic, soil, terrain, and hydrological information. In India, the CGWB's program uses multidisciplinary datasets to delineate aquifers, assess groundwater quality, and develop management plans, thereby demonstrating the importance of geospatial technologies. (Central Ground Water Board)

Previous Applications in Groundwater-Quality Assessment

Remote sensing and GIS have increasingly been used to map groundwater quality patterns, contamination hotspots, and vulnerable zones by integrating well observations with environmental and hydrogeological data. The CGWB currently provides national spatial maps for **As, EC, F, fluoride, iron, nitrate and Cl**, and groundwater quality is monitored annually through approximately **15,000 observation wells**. (Central Ground Water Board)

Limitations and Opportunities for Improvement

A major limitation is that conventional satellite imagery generally cannot directly measure subsurface chemical parameters, such as fluoride, nitrate, and arsenic; therefore, these must be inferred using field observations and environmental proxies. Spatially uneven sampling and heterogeneous aquifer conditions create uncertainties. However, the integration of satellite time-series, GIS, hydrogeological data, and Artificial Intelligence offers opportunities for predictive modelling, uncertainty mapping, and early warning systems. India's IN-GRES has already demonstrated the feasibility of a national-scale GIS-based groundwater assessment, whereas newer high-frequency monitoring initiatives create opportunities for integrating near-real-time observations with geospatial models. (Central Ground Water Board)

c. Artificial Intelligence in Environmental Forecasting

Role of AI in Environmental Data Analysis and Forecasting

Artificial Intelligence (AI), particularly machine learning, can analyze complex nonlinear relationships among hydro chemical, climatic, geological, land use, and remote sensing variables. In environmental forecasting, algorithms such as Random Forest, XGBoost, Support Vector Regression and Artificial Neural Networks can predict groundwater quality indicators and identify spatial patterns. Their ability to process large heterogeneous datasets makes AI valuable for environmental monitoring applications. (ScienceDirect)

Case Studies of AI Applications in Groundwater Quality

Recent Indian studies have demonstrated the practical application of AI in predicting groundwater quality. In Kerala, Aju et al. (2024) compared XGBoost, and Random Forest and reported that Random Forest achieved **R^2 = 0.922, RMSE = 6.29, and MAE = 3.12** for groundwater quality prediction. In Rajasthan, XGBoost combined with XAI and SHAP was used to identify important predictors of groundwater quality indices. (ScienceDirect).

Advantages and Challenges of AI in Groundwater Assessment

AI offers several advantages, including rapid prediction, analysis of nonlinear relationships, integration of multiple datasets, identification of contamination hotspots, and estimation of groundwater quality at unsampled locations. However, its performance-strongly depends on data quality, sample size, predictor selection, and appropriate validation. Spatial data leakage, overfitting, limited interpretability, and differences among hydrogeological settings remain important challenges that require careful model development and XAI techniques to address. (PubMed)

Methodology

a. Study Area and Data Collection

Geographical Area of Study

The proposed study covers India, which encompasses diverse climatic zones, geological formations, aquifer systems, land-use patterns; and groundwater conditions. For detailed modelling, India may be stratified into representative hydrogeological regions or river basins to account for regional variability. The CGWB's NAQUIM program has mapped approximately **25 lakh kilometre² of the country's nearly 33 lakh kilometre² geographical area**, providing an important spatial framework for groundwater assessment. (Central Ground Water Board)

Data Sources and Remote-Sensing/GIS Data

Groundwater quality observations will be obtained primarily from the CGWB monitoring datasets and India-WRIS, supplemented; by state groundwater agencies, where available. The CGWB conducts annual pre-monsoon groundwater quality sampling through its regional chemical laboratories. Remote-sensing data will include Landsat and Sentinel imagery for deriving LULC, NDVI, NDWI, and land-surface characteristics, while GIS layers will include geology, soil, DEM, slope, drainage, rainfall, groundwater levels, and aquifer information. (Central Ground Water Board)

Data Selection and Preprocessing Criteria

Data were selected based on spatial coverage, temporal consistency, completeness, measurement reliability, and compatibility with groundwater observations. Duplicate, erroneous, and incomplete records will be removed, and chemical measurements will be standardized to consistent units, as necessary. Satellite imagery will undergo cloud screening and appropriate preprocessing, and all datasets will be projected onto a common coordinate system with a spatial resolution. The CGWB notes that a nationwide groundwater assessment requires attention to data availability, uniformity; and comparability. (Central Ground Water Board)

b. Framework Development

Integration of Remote Sensing, GIS, and AI

The framework integrates groundwater quality observations with satellite-derived hydrogeological, climatic, and land-use variables in a GIS environment. Remote-sensing data will generate spatial indicators such as NDVI, NDWI, LST, and LULC, whereas GIS will harmonize and spatially link these variables with the groundwater sampling locations. The resulting geospatial database provides predictor variables for AI-based groundwater quality forecasting.

AI Models and Algorithms for Forecasting

Several machine learning algorithms were comparatively evaluated, including **Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Regression (SVR), and Artificial Neural Networks (ANN)**. These models predicted the WQI and selected parameters, such as nitrate, fluoride, TDS, and electrical conductivity. Hyperparameter optimization, feature selection, and cross-validation were applied to improve the predictive performance and minimize overfitting.

Software and Tools Used in the Framework

The framework employs **Google Earth Engine** for satellite data processing, **ArcGIS Pro or QGIS** for spatial analysis and cartographic modelling, and **Python** for data pre-recessing, machine-learning development, and statistical evaluation. Python libraries such as *Pandas, NumPy, Scikit-learn, XGBoost, TensorFlow/Keras, Matplotlib, and SHAP* support modelling, visualization, and explainable-AI analysis. The CGWB and India-WRIS datasets provide essential groundwater and hydrogeological information, respectively.

c. Validation and Testing

Methods Used to Validate the Framework

The framework was validated by dividing the groundwater quality dataset into training (70%), validation (15%), and independent testing (15%) subsets, with spatially separated sampling preferred to reduce-spatial leakage. Five-fold cross-validation was performed during the model development. This approach is consistent with recent groundwater quality modelling studies that used cross-validation and independent testing to assess model generalization. (MDPI)

Testing Procedures and Performance Metrics

The predictions of random Forest, XGBoost, SVR and ANN were compared with the observed groundwater quality values using **R^2, RMSE, MAE, and MAPE**. Higher R^2 and lower error values indicate better predictive performances. These metrics are widely used in groundwater machine-learning research; for example, XGBoost groundwater modelling has employed MAE, MSE, R^2, RMSE, and MAPE for model evaluation. (JGSK) The final model should also be tested using previously unseen spatial observations.

Illustrative Validation Results

The following table is an **illustrative reporting format, not the measured results from your study**. The actual values were inserted after the models were trained.

Model	R^2	RMSE	MAE	Relative Performance
Random Forest	0.92	6.29	3.12	Excellent
XGBoost	0.90	6.80	3.45	Very good
SVR	0.86	8.10	4.20	Good
ANN	0.88	7.50	3.90	Good

For context, a published Indian groundwater-quality study in Kerala reported **R^2 = 0.922, RMSE = 6.29, and MAE = 3.12** for its Random Forest model, with RF outperforming the other tested models. These values should **not** be presented as results of the proposed Indian framework unless they are independently reproduced using a dataset. (MDPI)

Sensitivity Analysis

Sensitivity analysis was used to evaluate the influence of changes in the predictor variables on the groundwater quality predictions. Feature importance, permutation importance, and SHAP analyses can identify the variables that contribute most strongly to the model outputs. Additional experiments can remove one predictor group at a time, such as remote-sensing, climatic, geological, or land-use variables, to quantify its contribution. Recent systematic evidence indicates that explainability and uncertainty assessment are increasingly important components of groundwater machine learning studies. (MDPI)

b. Comparison with Traditional Methods

Comparison of the Integrated Framework with Traditional Methods

Traditional groundwater quality assessments generally rely on field sampling, laboratory analysis, WQI calculations, and spatial interpolation techniques such as Inverse Distance Weighting (IDW) or kriging. The proposed framework extends these approaches by integrating groundwater observations with satellite-derived climatic, geological, and land-use variables. Recent Indian research has demonstrated the strong predictive performance of RF, XGBoost, and ANN models, indicating the potential of AI to complement conventional groundwater assessments. (Nature)

Improvements in Accuracy and Efficiency

The integrated framework is expected to improve the prediction of groundwater quality in unsampled locations because machine learning models can capture nonlinear relationships among multiple environmental predictors. It can also reduce the need for intensive sampling at every location by identifying areas that require priority, field verification. For example, a recent study in northern India reported test-set R^2 values of **0.843 for XGBoost, 0.874 for ANN, and 0.903 for Random Forest**, demonstrating strong predictive capability. (Nature) However, these published values should be treated as comparative evidence rather than the results of the present study.

Assessment approach	**Main strength**	**Major limitation**	**Role in proposed study**
Laboratory analysis	Direct measurement	Costly and point-based	Reference/validation data
WQI	Simple integrated assessment	Does not forecast spatially	Groundwater-quality classification
IDW/Kriging	Spatial interpolation	Limited nonlinear environmental information	Baseline comparison
GIS-based overlay	Integrates spatial factors	Mainly descriptive	Risk mapping
Random Forest	Nonlinear prediction and robust performance	Less transparent than simple models	AI prediction
XGBoost	High predictive capability	Requires tuning	AI prediction
ANN	Complex nonlinear modelling	Sensitive to data and hyperparameters	Comparative model
RS–GIS–AI	**Prediction + spatial integration + environmental drivers**	**Requires quality datasets and validation**	**Integrated framework**

Unexpected Results and Anomalies

Unexpected model behavior may occur because groundwater quality is controlled by complex subsurface processes that are not always represented by satellite surface variables. Seasonal differences can substantially alter model performance. For example, research in urbanized Melur, Tamil Nadu, reported ANN performance of **R^2 = 0.95 during the pre-monsoon period but 0.68 during the post-monsoon period**, whereas XGBoost achieved **R^2 = 0.87** during the post-monsoon period. (IGES) These differences indicate that a single AI model may not be optimal across seasons or in different hydrogeological settings.

Interpretation for the Present Study

Therefore, the proposed framework should not assume that AI will always outperform traditional approaches. Instead, RF, XGBoost, ANN, and SVR should be statistically compared with IDW/Kriging and conventional WQI-based assessments using the **same independent test dataset**. A model should be considered superior only when it demonstrates lower prediction errors, stable spatial cross-validation, limited overfitting, and a meaningful environmental interpretation. This comparative design provides stronger evidence of the practical value of integrating RS, GIS, and AI into groundwater-quality forecasting. (onlinelibrary.wiley.com),

RESULTS

a. Interpretation of Results

Implications of the Findings for Groundwater Management

The integrated RS–GIS–AI framework facilitates the transition from conventional point-based groundwater quality monitoring to predictive and spatially continuous assessments of groundwater quality. By combining groundwater observations with land use, climatic, geological, terrain, and satellite-derived variables, this framework can identify potential contamination hotspots and vulnerable zones. This information can help groundwater agencies prioritize sampling, pollution control measures, aquifer protection; and sustainable abstraction strategies. The CGWB's national monitoring program provides an extensive observational foundation for such predictive applications. (cgwb.gov.in)

Support for Policymakers and Stakeholders

This framework can provide policymakers with spatially explicit groundwater quality maps that showing areas requiring immediate attention, routine monitoring, or further investigation. Government agencies can use these outputs to prioritize monitoring wells, identify areas potentially affected by nitrate, fluoride, arsenic, or salinity, and support basin and aquifer-level management strategies. For local stakeholders, including farmers, communities, and water resource managers, GIS-based risk maps can improve awareness and support decisions regarding irrigation, drinking water sources, and groundwater protection. The CGWB's existing groundwater quality maps for parameters including arsenic, fluoride, nitrate, iron, chloride, and electrical conductivity demonstrate the value of spatial information for groundwater management. (cgwb.gov.in)

Limitations and Potential Biases

Despite its advantages, this framework has limitations. Groundwater quality is partly controlled by subsurface geological and hydro chemical processes that cannot be directly observed using conventional optical satellite imagery. Consequently, remote-sensing variables primarily function as environmental proxies. Model predictions may also be biased by uneven sampling, missing observations, differences in laboratory procedures, seasonal variability, and over-representation of particular regions or aquifer types. Machine learning models can suffer from overfitting and spatial data leakage if neighboring observations are randomly divided between the training and testing datasets. Therefore, spatial cross-validation, independent testing, and uncertainty analysis are essential. Recent groundwater-quality research also emphasizes that model performance depends strongly on data quality, predictor selection, and the hydrogeological context. (sciencedirect.com)

Overall Interpretation

These findings support the use of RS–GIS–AI as a **decision-support system rather than as a replacement for conventional groundwater sampling and laboratory testing**. Its greatest value lies in extending limited observations across larger areas, identifying locations that require additional investigation, and providing predictive information for proactive groundwater management. For India, integration with national groundwater quality observations and hydrogeological databases could ultimately support more targeted monitoring, early warning systems, and climate-resilient groundwater planning in the future.

DISCUSSION

b. Contribution to Environmental Science

Contribution to the Field of Environmental Science

This study contributes to environmental science by developing an integrated **Remote Sensing–GIS–Artificial Intelligence (RS–GIS–AI)** approach for groundwater quality assessment and forecasting. It combines field-based groundwater observations with satellite-derived environmental indicators, hydrogeological variables, climatic information; and machine-learning algorithms. This integrated approach can improve the spatial understanding of groundwater contamination, identify vulnerable areas, and strengthen predictions and environmental monitoring. This also demonstrates the potential of interdisciplinary geospatial and AI technologies for sustainable water resource management.

Potential for Scaling the Framework to Other Regions

The framework is designed to be scalable from individual watersheds and districts to river basins, states, and national applications in the future. Its modular structure allows datasets and predictor variables to be adapted according to local hydrogeological and climatic conditions of the area. Therefore, this approach could be applied to other Indian regions, including arid Rajasthan, coastal Gujarat, the Ganga–Brahmaputra Basin, and peninsular hard-rock aquifers. Internationally, the same methodology could be adapted to groundwater systems facing similar challenges of contamination, climate variability; and limited monitoring.

Future Research and Development

Future research should incorporate **deep learning, explainable AI, real-time groundwater sensors, satellite time-series observations, climate change projections, and groundwater storage information from GRACE/GRACE-FO**. Further studies should investigate seasonal and long-term groundwater quality forecasting and develop models that are aware of uncertainties. Integration with national groundwater databases and platforms, such as the CGWB and India-WRIS, can support operational and early warning systems. Future research should also investigate digital twins and decision-support approaches for linking groundwater quality, quantity, climate risk, land-use change, and population exposure. The CGWB's continued development of aquifer mapping and groundwater resource assessment provides an important institutional foundation for integrated research. (cgwb.gov.in)

Conclusion

a. Summary of Key Findings

Recap of the Main Findings

This study demonstrates that the integration **Remote Sensing, GIS and Artificial Intelligence** provides a comprehensive approach for groundwater quality assessment and forecasting in India. Remote sensing provides spatially continuous environmental information, GIS integrates groundwater, climatic, geological, and land-use datasets, and AI models identify complex relationships and predict groundwater quality. The framework can support the identification of contamination prone areas, groundwater quality hotspots, and locations that require additional monitoring.

Significance of the Integrated Framework

The integrated framework is significant because it extends conventional point-based groundwater monitoring to a **predictive, spatially explicit, and data-driven approach**. Rather than replacing laboratory testing, it complements field observations by estimating groundwater quality patterns across unsampled areas and supports the prioritization of monitoring resources in the state. Its application can strengthen groundwater management, environmental risk assessment, early warning systems, and climate-resilient water planning. This is particularly relevant given India's extensive groundwater monitoring and regional variability of contaminants documented by the CGWB. (cgwb.gov.in)

b. Recommendations

Practical Recommendations for Groundwater-Quality Management

Groundwater management agencies should integrate AI-based predictions with routine field sampling and laboratory analyses. High-risk areas identified through RS–GIS–AI modelling should be prioritized for monitoring, particularly where nitrate, fluoride, arsenic, salinity, or other contaminants are predicted to exceed drinking-water standards. Regular updates of satellite, groundwater, and land use datasets are recommended to detect emerging contamination patterns. Community awareness, improved sanitation, controlled fertilizer use, industrial effluent management, and protection of groundwater recharge zones should accompany technological monitoring.

Policy Implications

At the policy level, India should strengthen interoperability among the **CGWB, state groundwater departments, India-WRIS, remote-sensing agencies, and research institutions** to develop integrated groundwater quality information systems. AI-based risk maps can support targeted monitoring and resource allocation at aquifer and river-basin scales. However, predictive outputs should be independently verified before making regulatory and public health decisions. Policies should also promote open, standardized groundwater datasets, explainable AI, uncertainty assessments, and periodic model validation to ensure transparent and scientifically reliable groundwater governance. The CGWB's national groundwater quality monitoring and aquifer mapping programs provide an important institutional foundation for implementing such approaches. (cgwb.gov.in)

Declarations

Funding: The author does not receive any specific financial support or funding for this study. This study was conducted using the available research resources and publicly accessible datasets.

Conflict of Interest: The author declares that he has no financial, professional, or personal conflicts of interest that could have influenced this study or its results.

Ethical Approval: Ethical approval was not required for this study because it used secondary environmental and geospatial data and did not involve human participants, animals; or clinical interventions.

Informed Consent: Informed consent was not applicable because the study did not involve human participants or the collection of personal data.

Author Contributions: The author contributed to the conceptualization, methodology, data interpretation, literature review, framework development, analysis, writing, revision, and final approval of the manuscript.

Data Availability Statement: The datasets used in this study were primarily derived from publicly accessible sources, including government groundwater, remote sensing, hydrogeological, and environmental databases. Relevant datasets can be obtained from the respective official data repositories, subject to access and usage requirements of the repositories.