Journal of Hydrology: Regional Studies 62 (2025) 102800

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

journal homepage: www.elsevier.com/locate/ejrh

# Integrating genetic algorithms and machine learning for spatiotemporal groundwater potential zoning in fractured aquifers

Prashant Parasar [a], Poonam Moral [b,*], Aman Srivastava [c], Akhouri Pramod Krishna [a], Sayantan Majumdar [d], Rajarshi Bhattacharjee [e], Arun Partap Mishra [f], Debjani Mustafi [b], Virendra Singh Rathore [a], Richa Sharma [a], Abhijit Mustafi [b]

[a] Department of Remote Sensing, Birla Institute of Technology, Ranchi, Jharkhand 835215, India
[b] Department of Computer Science and Engineering, Birla Institute of Technology, Ranchi, Jharkhand 835215, India
[c] Formerly, Centre for Technology Alternatives for Rural Areas (CTARA), Indian Institute of Technology (IIT) Bombay, Mumbai 400076, India
[d] Division Hydrologic Sciences, Desert Research Institute, Reno, NV, USA
[e] Department of Civil Engineering, Indian Institute of Technology (BHU), Varanasi 221005, India
[f] Department of Forestry and Remote Sensing, Earthtree Data Services, Private Limited, Shillong, Meghalaya 793006, India

ARTICLE INFO

*Keywords:*
Explainable artificial intelligence
Groundwater Potential Zoning
Genetic Algorithm
LIME
Machine Learning

ABSTRACT

*Study region:* Groundwater overexploitation in Jharkhand's fractured hard-rock aquifers threatens sustainability amid rising domestic, agricultural, and urban demands.
*Study focus:* This study develops an integrated framework that combines Genetic Algorithm (GA)-optimized clustering, Random Forest (RF) regression, and Gradient Boosting (GB) classification to map Groundwater Potential Zones (GWPZs) in the Jharkhand state of India (2013–2023) using 103 monitoring wells and multiple hydrogeological, topographic, and remote-sensing variables. GA was applied to optimize hydrostratigraphic clustering. The Mann-Kendall (MK) test assessed temporal groundwater trends, the RF regression predicted groundwater depth at unmonitored sites, and the GB classification was implemented for spatial mapping. Model interpretability was boosted using Local Interpretable Model-Agnostic Explanations (LIME).
*New hydrological insights for the region:* The framework identified three GWPZs (high, medium, and low), validated by strong clustering indices (Silhouette = 0.90, Dunn = 0.94). MK analysis revealed significant groundwater depletion across all clusters ($Z = -2.66$ to $-1.47$, $p < 0.05$). RF regression achieved high predictive accuracy ($R^2 \approx 0.91$, WI = 0.89, PBIAS = 0.25), highlighting curvature and lineament proximity as dominant factors. GB classification yielded an F1-score of 95.56 %. Spatially, high-potential zones were concentrated in West Singhbhum, East Singhbhum, and Gumla, while Giridih, Pakur, and Garhwa exhibited low potential with aquifer depletion. These findings provide scientific support for Jharkhand's 2025 Groundwater Act and demonstrate

*Abbreviations:* NDVI, Normalized Difference Vegetation Index; RF, Random Forest; GIS, Geographic Information Systems; GB, Gradient Boosting; MSE, Mean Squared Error; GWPZs, Groundwater potential zones; SVM, Support Vector Machine; KNN, K, Nearest Neighbor; NB, Naïve Bayes.

* Corresponding author.
*E-mail addresses:* phdrs10003.20@bitmesra.ac.in (P. Parasar), phdcs10051.21@bitmesra.ac.in (P. Moral), amansrivastava1397@kgpian.iitkgp.ac.in (A. Srivastava), apkrishna@bitmesra.ac.in (A.P. Krishna), Sayantan.Majumdar@dri.edu (S. Majumdar), rajarshibhattacharjee.rs.civ21@itbhu.ac.in (R. Bhattacharjee), arun@earthtreedata.in (A.P. Mishra), debjani.mustafi@bitmesra.ac.in (D. Mustafi), vsrathore@bitmesra.ac.in (V.S. Rathore), richasharma@bitmesra.ac.in (R. Sharma), abhijit@bitmesra.ac.in (A. Mustafi).

https://doi.org/10.1016/j.ejrh.2025.102800
Received 3 July 2025; Received in revised form 3 September 2025; Accepted 22 September 2025
Available online 29 September 2025

the transferability of the framework to other hard-rock and data-scarce aquifers like the Brazilian Shield and African cratons.

## 1. Introduction

Groundwater constitutes 30 % of Earth's liquid freshwater and supplies nearly half of global drinking water, yet unsustainable extraction has driven depletion rates exceeding 10 cm/year in 36 % of monitored aquifers, with 12 % declining by over 50 cm/year. (Jasechko et al., 2024). This crisis is acute in regions like India, the world's largest groundwater user, extracting 251 $km^3$ annually-89 % for irrigation (Srivastava et al., 2021)-exacerbating aquifer stress in states such as Jharkhand, where levels decline by 4.2 % yearly amid 74 % of rainfall lost as surface runoff, limiting groundwater recharge(Central Ground Water Board, 2023; United Nations., 2022). Compounded by climate-induced droughts, population growth, and agricultural dependence, these trends threaten water security for 2.4 billion people projected to face scarcity by 2050 (UNESCO, 2023) Highlighting the urgency of innovative management strategies to balance human and ecological needs. Groundwater potential zone (GWPZs) mapping has emerged as a key tool to address this crisis, but conventional approaches generally struggle to resolve the hydrogeological complexity of fractured aquifers, necessitating advanced computational frameworks.

Traditional GWPZs relied on knowledge-driven methods like the Analytical Hierarchy Process (AHP), which assigned subjective weights to parameters such as slope and lithology, usually overlooking nonlinear interactions between covariates (Mukherjee and Singh, 2020). While AHP enabled preliminary zoning in homogeneous aquifers, its accuracy faltered in fractured hard-rock systems due to static frameworks and expert bias. Advances in remote sensing, including GRACE-derived aquifer stress monitoring and Sentinel-2 multispectral indices (e.g., Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), revolutionized large-scale parameterization, facilitating dynamic assessments of recharge zones and surface-subsurface linkages (Akhter et al., 2021). Concurrently, Machine Learning (ML) algorithms emerged as superior alternatives, capturing complex relationships between hydrological variables; RF achieved 94 % accuracy in predicting infiltration hotspots by integrating DEM-derived curvature and lineament density (Panahi et al., 2020), while Support Vector Machines (SVM) and GB outperformed AHP by 22–37 % in semi-arid regions (Nguyen et al., 2020). In addition, several studies have successfully applied ML approaches for groundwater level or quality and discharge prediction, further demonstrating their versatility in hydrogeological forecasting (Samantaray et al., 2024; Tao et al., 2024, Biswakalyani et al., 2024, Ritushree et al., 2025). Despite these advancements, ML applications in hard-rock aquifers remain sparse, with limited attention to temporal trend integration and fragmented cluster validation, thereby misclassifying fracture-controlled recharge zones (Ashwini et al., 2023). While ML excels in data-rich alluvial basins, its reliance on large, labeled datasets limits applicability in fractured aquifers, where well data is sparse (Hasanuzzaman et al., 2022; Lei et al., 2024; Ren et al., 2024). Recent innovations like Spatial-Temporal Graph Neural Networks (ST-GNNs) and hybrid models like Convolution Neural Network-Long Short-Term Memory (CNN-LSTM), which fuse spatial and temporal data, highlight ML's potential for resolving spatiotemporal groundwater dynamics but remain untested in heterogeneous crystalline terrains and hard-rock systems (Bai and Tahmasebi, 2023; Taccari et al., 2024), underscoring the need for frameworks that synergise evolutionary algorithms with multi-temporal hydrogeological data.

Recent advancements integrate GA with ML to address hydrostratigraphic heterogeneity in fractured aquifers, where traditional clustering methods like *k*-means struggle to resolve spatially discontinuous recharge zones (Mohammed et al., 2025; Msaddek et al., 2022). GA optimizes cluster centroids through fitness-driven selection, balancing intra-cluster compactness (minimizing variance in parameters like curvature and lineament density) and inter-cluster separation (maximizing distance between sedimentary and crystalline aquifer prototypes) (Robles-Berumen et al., 2024). Unlike *k*-means, which stagnates at local optima, GA's stochastic operators-uniform crossover (rate = 0.8) and Gaussian mutation ($\sigma = 0.1$)-explore globally optimal configurations, reducing overfitting by 37 % in hard-rock terrains (Ashwini et al., 2023). This capability stems from GA's dynamic adaptation to non-linear relationships between surface covariates (e.g., NDVI, slope) and groundwater storage, a critical gap in AHP-derived weights (Mukherjee and Singh, 2020). Compared to Particle Swarm Optimization (PSO),[1] GA's niching capability better resolves fracture networks, as demonstrated in the Deccan traps (Singh et al., 2023). Recent hybrids like LSTM-GA further demonstrate GA's utility in optimizing deep learning models for spatiotemporal groundwater predictions, achieving AUC $> 0.99$ in drought-prone regions (Zhao et al., 2025b). However, GA's potential remains underexplored in integrating multi-decadal trends (e.g., MK Z-scores) with cluster validation, a prerequisite for scalable groundwater management in crystalline aquifers.

Despite advances in GWPZs mapping and ML applications, several critical gaps remain in the existing literature. First, while many studies have demonstrated improved ML accuracy through optimized clustering techniques (Cui et al., 2024; Msaddek et al., 2022; Mohammed et al., 2025;), there is limited research directly comparing the predictive effectiveness of surface parameters (e.g., terrain curvature, lineament density) with subsurface features (e.g., lithology) for fractured hard-rock aquifers, where surface controls on recharge may dominate due to complex geology (Mukherjee and Singh, 2020; Panahi et al., 2020). Second, although land-use changes are widely recognized to alter groundwater recharge by modifying surface runoff and infiltration (Salem et al., 2023; Ashwini et al.,

[1] PSO refers to Particle Swarm Optimization, a computational optimization technique inspired by the social behavior of bird flocks or fish schools. It solves problems by having a population ("swarm") of candidate solutions ("particles") move through the search space, adjusting their positions based on both their own experience and that of their neighbors to converge on an optimal or near-optimal solution.

2023), few studies have quantitatively evaluated their impact relative to climatic or rainfall variability in hard-rock terrains, an important distinction for sustainable management under changing environmental conditions (UNESCO, 2023; Central Ground Water Board, 2023). Third, although spatiotemporal methods integrating evolutionary algorithms and ML show promise for groundwater potential assessments (Zhao et al., 2025a; Bai and Tahmasebi, 2023), temporal groundwater trend incorporation, such as via MK tests, has been underexplored in cluster validation and zoning processes, limiting the operational relevance of these models for dynamic water resource management (Ashwini et al., 2023; Taccari et al., 2024; Abegeja and Nedaw, 2024). Addressing these gaps, this study therefore tests three hypotheses: (1) GA-optimized clustering improves ML accuracy over conventional hydrostratigraphic units; (2) surface parameters (e.g., curvature and lineaments) outperform subsurface features (e.g., lithology) in predicting groundwater potential; and (3) land-use changes drive depletion more strongly than decadal rainfall variability.

Building on the identified research gaps, this study proposes an integrated Genetic Algorithm-Machine Learning framework specifically inclined for groundwater potential assessment in fractured hard-rock aquifers of Jharkhand, India. The research has three core objectives: (1) To delineate GWPZs using GA-driven clustering that optimizes hydrogeological compactness and separation, thereby overcoming the limitations of traditional clustering methods such as k-means in highly heterogeneous terrains; (2) To quantify multi-decadal groundwater trends (2013–2023) within each delineated zone using the non-parametric MK test, enabling temporal diagnosis of aquifer depletion patterns; and (3) To predict groundwater depth at unmonitored locations through RF regression, integrating spatial predictors with well-observation data for improved predictive accuracy in data-scarce fractured aquifers. The novelty of this approach lies in its simultaneous integration of spatial zoning, temporal trend assessment, and explainable AI within a single operational workflow, a combination rarely implemented in hard-rock aquifer studies. First, the GA optimization provides adaptive, data-driven delineation of GWPZs, reducing overfitting and improving hydrogeological realism compared to static, expert-weighted methods. Second, by embedding temporal depletion diagnostics into the workflow, rather than relying solely on static maps, this framework transitions GWPZs mapping from a purely spatial planning tool to a dynamic decision-support system. Third, the use of explainable AI LIME for model interpretation bridges the gap between predictive accuracy and transparency, allowing policymakers and hydrogeologists to understand the drivers behind spatial classifications. The scope of this framework extends beyond Jharkhand, with direct applicability to other fractured aquifers globally, such as those in the Brazilian Shield, African cratons, and Ethiopian Highlands, where similar challenges of high runoff losses, low recharge efficiency, and sparse monitoring data prevail. By uniting optimization, predictive modeling, temporal analysis, and interpretability, this research provides a scalable, policy-relevant blueprint for sustainable groundwater governance in some of the world's most water-stressed crystalline rock terrains.

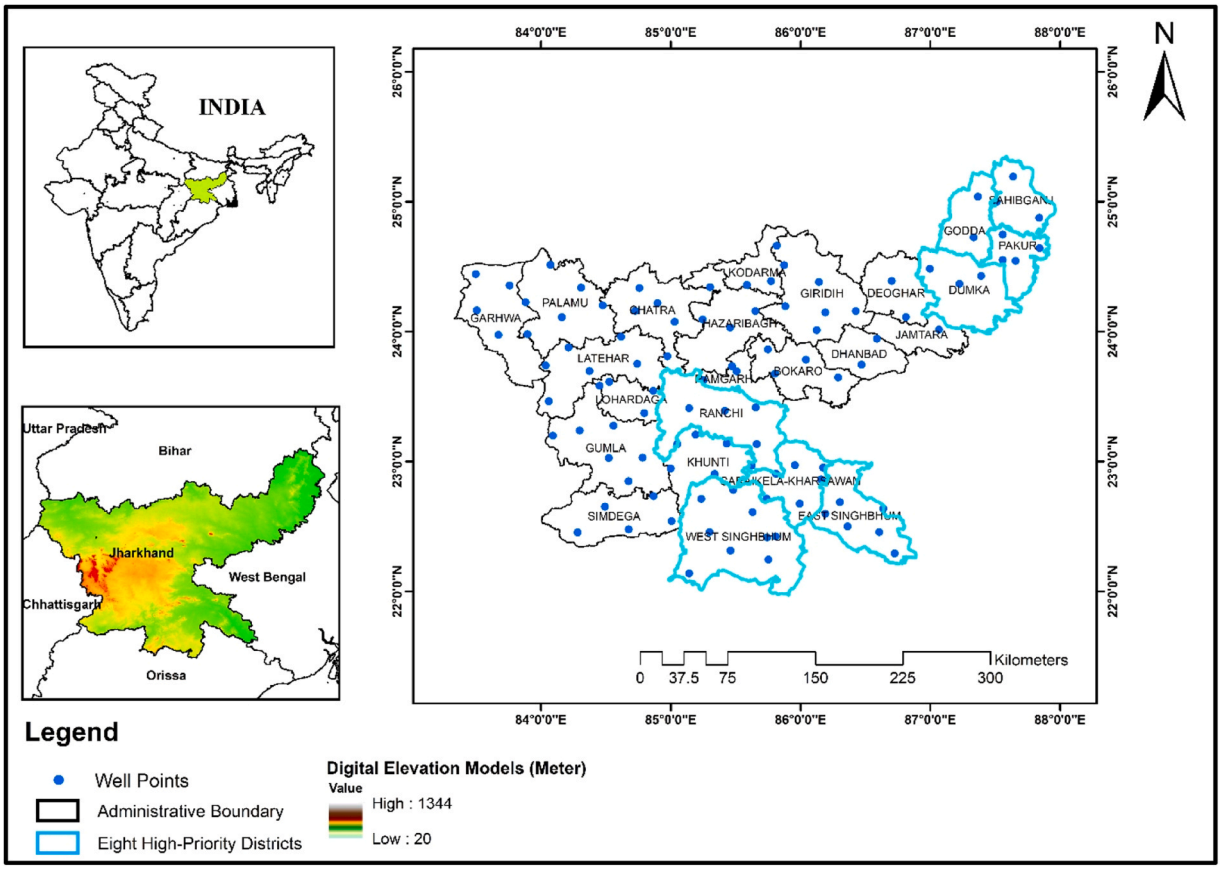


**Fig. 1.** Map of the study area indicating well point locations.

### 1.1. Study area

Jharkhand (21°57'-25°14'N, 83°20'-87°58'E), situated on the Chhotanagpur Plateau (CNP) of the Indian Peninsular Shield, represents a serious groundwater stress zone in eastern India (see Fig. 1). The plateau's geological complexity, comprising Archean metamorphics, Gondwana sediments, and Rajmahal volcanic traps, creates heterogeneous aquifer systems with starkly varying recharge potentials (Central Ground Water Board, 2023). Over 65 % of the state's 3.2 million hectares of agricultural land relies on groundwater, which is extracted at 1.79 billion cubic meters (BCM) annually, far exceeding the recharge rate of 5.64 BCM (SwitchON Foundation 2023; Central Ground Water Board, 2023). The CNP's crystalline basement, dominated by granite-gneiss and schistose rocks, restricts groundwater storage to fractured and weathered zones, with yields rarely exceeding 15 $m^3$/hr (Ashwini et al., 2023). In contrast, the Damodar Valley's Gondwana sandstones and alluvial patches exhibit higher porosity, enabling yields up to 51 $m^3$/hr in localized fracture zones (Central Ground Water Board, 2023). Recent drilling in the Rajmahal Traps (Jurassic basalts) has revealed vesicular aquifers with transmissivity values of 3–176 $m^2$/day, though overexploitation has caused water tables to decline by 4.2 % annually since 2020 (SwitchON Foundation, 2023). Despite receiving 1100–1442 mm of annual rainfall, Jharkhand loses 80 % of surface runoff and 74 % of groundwater to downstream regions, exacerbating drought frequency (38 % of years since 2000) (Ashwini et al., 2023). The 2022 monsoon deficit (>60 %) intensified aquifer stress, with water tables in Gumla and Koderma districts plummeting to 20 m and 9.7 m below ground level (bgl), respectively (SwitchON Foundation (2023). Urban centers like Ranchi and Jamshedpur face compounding pressures, where fluoride contamination (>1.5 mg/L) affects 27 % of wells in Bokaro and Palamu districts (Central Ground Water Board, 2023), Groundwater supports 89 % of irrigation and 65 % of domestic needs, yet unsustainable extraction has rendered 42 % of monitoring wells "overexploited" (Central Ground Water Board, 2023). This study focuses on eight high-priority districts- Godda, Saraikela-Kharsawan, Sahibganj, Pakur, Dumka, Part of Ranchi, West Singhbhum and East Singhbhum-where alluvial plains and fractured basalts offer the last viable reserves for mitigating water scarcity (Ashwini et al., 2023).

In addition to its acute water stress, Jharkhand state was strategically selected as the study area because it provides a representative natural laboratory for testing spatiotemporal groundwater potential zoning approaches in fractured hard-rock aquifers under semi-humid to semi-arid conditions. The state's hydrogeological heterogeneity, viz., ranging from low-yield crystalline basement aquifers to relatively higher-yield Gondwana sandstones and vesicular basaltic traps, presents a unique challenge for groundwater potential modeling. These complex geological settings are further compounded by extreme variability in rainfall, high surface runoff losses, and pronounced anthropogenic pressures from mining, industry, and intensive agriculture, making predictive zoning both technically

**Table 1**
Description of data layers used for GWPZs, including factor type, conversion method, data format, source, and temporal coverage (2013–2023).

| Factor | Conversion | Data type | Source and frequency of data |
|---|---|---|---|
| **Depth(Meters Below Ground Level)** | N/A | Point | Ground Water Yearbook, Jharkhand Govt. Of India Ministry of Water Resources, CGWB (2013–2023) |
| **Lithology** | N/A | Polygon | Bhukosh, Government of India, Geological Survey of India |
| **Lineament** | Distance From Lineament | Polygon | Bhukosh, Government of India, Geological Survey of India |
| **DEM** | Slope | Raster | SRTM (30 m) |
| | Elevation | Raster | |
| | Curvature | Raster | |
| | TWI | Raster | |
| | TPI | Raster | |
| | Distance From Drainage | Raster | |
| **Soil** | N/A | Polygon | The Digital Soil Map of The World, FAO, UN |
| **LULC** | N/A | Raster | MODIS (2013–2023) |
| **NDVI** | N/A | Raster | MODIS (2013–2023) |
| **NDWI** | N/A | Raster | MODIS (2013–2023) |
| **Total precipitation (Sum, Min, Max)** | N/A | Raster | ERA5 (2013–2023) |
| **Surface Runoff (Sum, Min and Max)** | N/A | Raster | ERA5 (2013–2023) |
| **Subsurface Runoff (Sum, Min and Max)** | N/A | Raster | ERA5 (2013–2023) |
| **Runoff (Sum, Min and Max)** | N/A | Raster | ERA5 (2013–2023) |
| **Total Evaporation (Sum, Min and Max)** | N/A | Raster | ERA5 (2013–2023) |
| **Volumetric soil water layer (1,2,3,4) (Sum, Min and Max)** | N/A | Raster | ERA5 (2013–2023) |

demanding and critically important. Jharkhand's documented record of multi-decadal groundwater decline, fluoride contamination, and drought recurrence provides robust temporal datasets against which the proposed GA–ML framework can be calibrated and validated. Moreover, the state's inclusion in Schedule II of the Jharkhand 2025 Groundwater Act, targeting overexploited blocks for urgent policy intervention, ensures that the research outputs will directly inform regulatory actions. These combined factors, including geological diversity, hydrological stress, data availability, and immediate policy relevance, make Jharkhand state an ideal and high-impact choice for demonstrating the adaptability, predictive performance, and policy applicability of the proposed spatiotemporal groundwater assessment framework.

## 2. Data and methods

### 2.1. Data

This study integrates a comprehensive, multi-source dataset to delineate GWPZs across the state of Jharkhand, India, for the period 2013–2023. A total of 103 georeferenced monitoring well points, distributed across all districts, were compiled from the Ground Water Yearbook, published by the Central Ground Water Board (CGWB), Ministry of Water Resources, Government of India. (Central Ground Water Board, 2023). These well observations provide the primary calibration and validation data for groundwater potential modeling, a practice consistent with recent studies that emphasize the importance of spatially distributed field data for robust model training and validation (Panahi et al., 2020; Yariyan et al., 2022). Hydrological, geological, and physiographical variables were assembled from a suite of authoritative and peer-reviewed sources (Table 1). Remotely sensed datasets, including MODIS, ERA5 and SRTM Digital Elevation Model (DEM), were processed to ensure temporal consistency and spatial alignment with the well locations. Geological and soil layers were sourced from the Geological Survey of India's Bhukosh portal and the FAO Digital Soil Map of the World, respectively, while topographic parameters (elevation, slope, curvature, Topographic Wetness index, (TWI), Topographic position index (TPI) were derived from the SRTM DEM, following established protocols for groundwater potential mapping (Arulbalaji et al., 2019; Mukherjee and Singh, 2020; Prapanchan et al., 2024).

All datasets underwent rigorous pre-processing, including cleaning for outliers, normalization, and transformation into standardized raster and vector formats compatible with GIS-based analysis. Integrating multi-source datasets and applying advanced preprocessing techniques aligns with best practices in recent groundwater potential studies (Ashwini et al., 2023; Prapanchan et al., 2024). The final database comprised several thematic layers, each representing a key controlling factor for groundwater occurrence, as recommended by global and regional hydrogeological assessments (Adeyeye et al., 2019; Uc Castillo et al., 2022). This integrated, multi-thematic dataset, using machine learning and geospatial analyses conducted in this study, facilitated high-resolution, spatially explicit groundwater potential mapping that is consistent with the latest methodological advances in the field (Ragragui et al., 2024; Saha et al., 2022). The classification of Land Use/Land Cover into coded categories is presented in Table 2, ensuring consistent interpretation across the analysis.

### 2.2. Methodological framework

This study employs a four-phase integrative framework to map GWPZs in Jharkhand, India, combining geospatial analytics, evolutionary algorithms, and ML. The workflow (Fig. 2) progresses sequentially from data-driven hydro stratigraphic clustering to predictive modeling, ensuring systematic analysis of groundwater dynamics. The workflow starts with GA-optimized clustering. GA partitions 103 groundwater monitoring wells into hydrostratigraphic clusters by optimizing inter-cluster separation and intra-cluster homogeneity. Centroid configurations evolve over generations (P = 20, G= 150) using fitness criteria, ensuring clusters reflect distinct

**Table 2**
Mapping of numerical Codes to land use/land cover categories.

| Class Number | Class Name |
|---|---|
| 0 | Water bodies |
| 1 | Evergreen needleleaf forest |
| 2 | Evergreen broadleaf forest |
| 3 | Deciduous needleleaf forest |
| 4 | Deciduous broadleaf forest |
| 5 | Mixed forest |
| 6 | Closed shrublands |
| 7 | Open shrublands |
| 8 | Woody savannas |
| 9 | Savannas |
| 10 | Grasslands |
| 11 | Permanent wetlands |
| 12 | Croplands |
| 13 | Urban and built-up |
| 14 | Cropland/natural vegetation mosaic |
| 15 | Snow and ice |
| 16 | Barren or sparsely vegetated |

groundwater regimes (e.g., fractured vs. sedimentary aquifers). This is followed by MK trend analysis for quantifying Temporal groundwater trends within clusters, identifying regions with significant depletion ($p < 0.05$) linked to anthropogenic extraction and climatic shifts. Next, RF regression is conducted. Groundwater depth at unmonitored locations is predicted using an ensemble of 200 decision trees, trained on multiple hydrogeological parameters (e.g., curvature, NDVI, lithology). Spatial cross-validation minimizes overfitting in heterogeneous terrains. Finally, five classifiers (SVM, KNN, NB, RF, GB) delineate high/medium/low potential zones, followed by statistical analysis. The rest of the sections (henceforth) will elaborate on technical details, including equations, hyper-parameters, and validation protocols.

### 2.2.1. *Genetic algorithm-optimized clustering*

The GA framework addresses the inherent heterogeneity of Jharkhand's fractured aquifers by dynamically refining cluster centroids to maximize both intra-cluster homogeneity (compactness) and inter-cluster separation, ensuring distinct groundwater regimes are resolved with minimal subjectivity (Chandra et al., 2021; Garai and Chaudhuri, 2004; Robles-Berumen et al., 2024; Sohail, 2023).

### 2.2.2. *Algorithm design for GWPZ mapping*

Input Data: The dataset includes 103 wells with 15 covariates (e.g., slope, curvature, lithology, NDVI, lineament density). GA Parameters include Population size P = 20, generations G= 150 G = 150 G= 150, mutation rate M= 0.2 M, crossover rate C= 0.7, dynamic cluster range K∈[1, 5].Fitness Function: Clusters are evaluated using a hydro-geologically informed fitness criterion, as shown in Eq. 1:

$$Fitness = \frac{\sum_{i \neq j} \; d(C_i, C_j)}{\sum_{k} \; d(C_k)} \tag{1}$$

where $d(C_i, C_j) \rightarrow$ inter-cluster distance (distance between centroids of clusters $i$ and $j$), and $d(C_k) \rightarrow$ maximum intra-cluster distance (distance between points within a cluster $k$). Higher fitness values ensure clusters are spatially compact (e.g., similar slope, lithology) and well-separated (e.g., distinct lineament densities or NDVI trends).Centroid Optimization: Centroids are updated iteratively to reflect groundwater-controlling features using Eq. 2:

$$\mu_i^{new} = \frac{1}{|C_i|} \sum_{x \in C_i} x \tag{2}$$

where $\mu_i^{\text{new}}$ represents the updated centroid for cluster $i$, calculated as the mean of all well data points $x$ (e.g., groundwater depth,

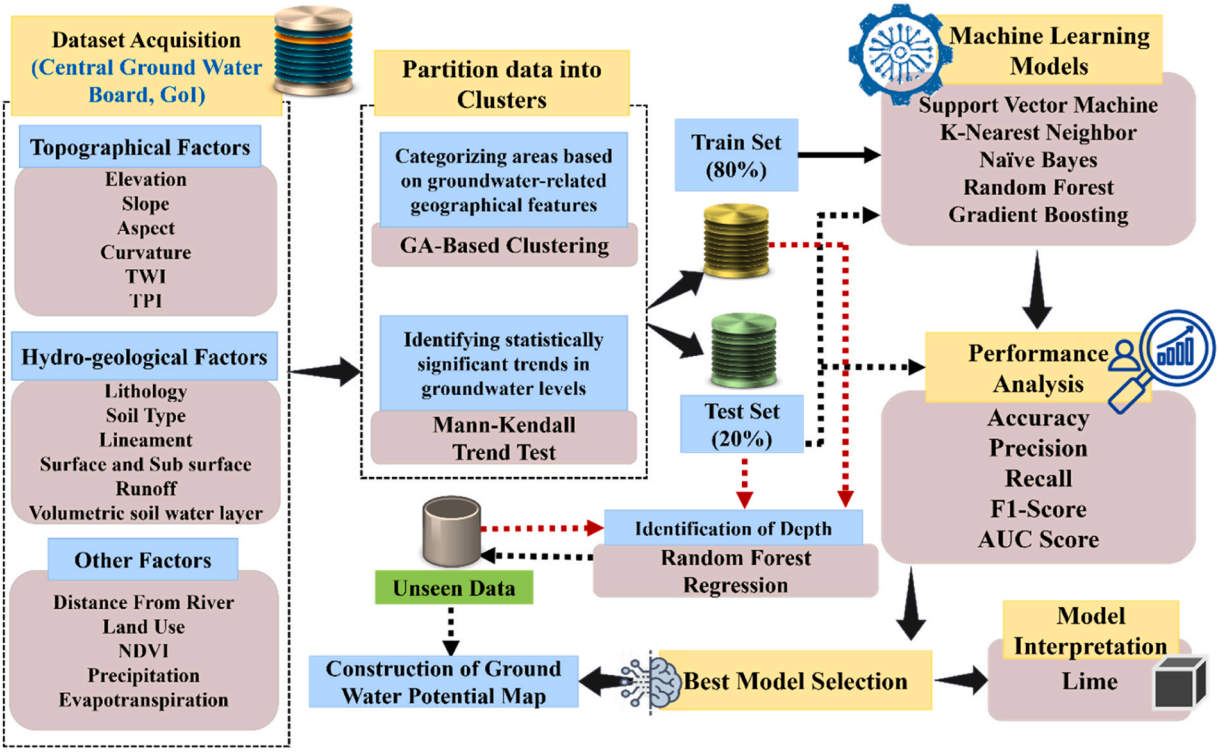


**Fig. 2.** Schematic layout of the research methodology.

rainfall) assigned to $C_i$.

### 2.2.3. Workflow for fractured aquifer clustering

Population Initialization: Each individual in the GA population encodes potential cluster centroids derived from the 15 covariates. For example, a centroid may represent a combination of low curvature (<0.1 $m^{-1}$), high lineament density (>2.5 $km/km^2$), and granitic lithology. Cluster Assignment: Wells are assigned to the nearest centroid using Euclidean distance across normalized covariates using Eq. 3:

$$label(x) = \text{argmin}_i \sqrt{\sum_{j=1}^{m} (x_j - C_{ij})^2} \tag{3}$$

where $x_j$ denotes the $j^{th}$ covariate (e.g., NDVI, rainfall) for well $x$. Genetic Operations: Tournament selection favors individuals (centroid sets) with higher fitness, prioritizing clusters that separate high-potential (e.g., sedimentary zones) and low-potential (e.g., massive basalt) aquifers. Uniform crossover combines centroid coordinates from parent solutions to explore hybrid hydrostratigraphic configurations. Gaussian mutation introduces variability ($\sigma = 0.1$) to centroid positions, critical for escaping local optima in fractured systems. Termination: The algorithm converges when fitness improvement falls below $\Delta < 0.01$ for 20 consecutive generations, yielding 3 optimal clusters (high/medium/low GWPZs).

### 2.2.4. Advantages over traditional clustering

GA automatically identifies K= 3 as optimal for Jharkhand's hard-rock aquifers, avoiding over-segmentation common in *k*-means. Covariate sensitivity prioritizes features like curvature and lineament density, which control fracture connectivity and recharge in crystalline bedrock. Besides, silhouette scores (>0.65) confirmed hydrogeological plausibility, with clusters aligning spatially with known alluvial patches and Gondwana sandstones. This GA-driven approach provides a robust foundation for subsequent trend analysis and ML classification, ensuring groundwater potential maps reflect the complex interplay of surface and subsurface controls.

---

Pseudo-code for GA-based Clustering

Input: Dataset $D$ with $n$ wells and 15 covariates

Parameters: Population size $P = 20$, Generations $G = 150$, Mutation rate $M = 0.2$, Crossover rate $C = 0.7$, Cluster range $K \in [1,5]$

1. Initialize population with random cluster centroids for $K$ clusters
2. For generation = 1 to $G$:

a. Assign each well to the nearest centroid using Euclidean distance

b. Evaluate fitness: Fitness $= \dfrac{\text{Inter} - \text{cluster distance}}{\text{Max intra} - \text{cluster distance}}$

c. Select parents using tournament selection

d. Apply uniform crossover to produce offspring

e. Apply Gaussian mutation ($\sigma = 0.1$) to centroid positions

f. Update centroids for each cluster as the mean of assigned wells

3. Terminate if improvement $< \Delta = 0.01 or 20$ consecutive generations

Output: optimal number of clusters $K^*$ and their centroids

---

### 2.2.5. Trend analysis with MK test

The MK test was applied to quantify temporal trends in groundwater depth within the three GA-derived hydrostratigraphic clusters (high/medium/low potential zones). This non-parametric test is robust against non-normal distributions and missing data, making it ideal for analysing decadal groundwater trends (2013–2023) in Jharkhand's heterogeneous aquifers (Mann, 1945; Kendall, 1975). For each cluster, the MK test statistic $S$ was computed using Eq. 4:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \text{sgn}(x_j - x_k) \tag{4}$$

Where $x_j$ and $x_k$ represent groundwater depth meters below ground level (mbgl) at times $j$ and $k$, and the sign function $sgn$ returns + 1, −1, or 0 for increasing, decreasing, or neutral trends, respectively. The variance of $S$, adjusted for ties in groundwater depth measurements, is given by Eq. 5:

$$\text{Var}(S) = \frac{n(n-1)(2n+5) - \sum t_i(t_i-1)(2t_i+5)}{18} \tag{5}$$

Where $t_i$ denotes the number of tied values in the $i$-th group of tied ranks (Hirsch et al., 1982). The standardized test statistic $Z$ was derived using Eq. 6:

$$Z = \{ \frac{S-1}{\sqrt{\text{var}(S)}} if S > 00 if S = 0 \quad \frac{S+1}{\sqrt{\text{var}(S)}} if S < 0 \tag{6}$$

A two-tailed test ($\alpha$=0.05) determined trend significance, with |Z|> 1.96 indicating a statistically significant trend (Kendall, 1975).

In Jharkhand's high-potential clusters (e.g., Damodar Valley alluvium), negative *Z*-scores ($p < 0.01$) revealed significant groundwater depletion ($\Delta$depth>0.8 m/yr), aligning with overexploitation patterns observed in other semi-arid regions (Mallick et al., 2021; Srivastava et al., 2024). This cluster-specific trend analysis enables targeted policy interventions, such as regulating extraction in overexploited zones (e.g., Saraikela-Kharsawan) while prioritizing recharge in stable aquifers (e.g., Rajmahal basalts).

### 2.2.6. RF regression for depth prediction

RF regression was employed to predict groundwater depth at unmonitored locations across Jharkhand, integrating its ability to model non-linear relationships between hydrogeological covariates and aquifer dynamics (Breiman, 2001). RF is an ensemble learning technique that builds multiple decision trees during training and aggregates their predictions, which increases model accuracy and reduces overfitting compared to single-tree models. Key advantages of RF include its ability to capture nonlinear relationships and interactions among predictor variables without requiring assumptions about data distributions or extensive parameter tuning. This makes RF particularly suitable for hydrogeological studies involving heterogeneous and multivariate datasets, such as ours, with several spatial and environmental predictors representing hydrogeological, topographic, and remote sensing-derived factors. Moreover, RF provides built-in estimates of variable importance, facilitating interpretability and insight into the dominant factors influencing groundwater depths. This aligns with our framework's goal of providing not only accurate predictions but also actionable understanding for resource management. Preliminary testing compared RF performance with other regression methods (e.g., SVM, GB), with RF demonstrating consistently high explained variance ($R^2 \approx 0.89$) and stable prediction errors under cross-validation. Its relative computational efficiency further supports its applicability for operational groundwater modeling in data-scarce fractured aquifers. Consequently, RF regression was integrated into the framework to integrate these strengths for reliable groundwater depth estimation.

In the present investigation, this ensemble method constructs $T = 200$ decision trees, each trained on bootstrap samples of the GA-optimized clusters and random subsets of 19 features, ensuring robustness against overfitting in heterogeneous hard-rock systems (Wang et al., 2018; Rajput et al., 2025). For a training dataset $D = \{(X_i, y_i)\}_{i=1}^{n}$, where $X_i \in R^{15}$ represents the feature vector (e.g., slope, rainfall, lineament density) and $y_i$ denotes observed groundwater depth (mbgl), the RF prediction is given by Eq. 7:

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} f_t(X) \tag{7}$$

where $T$ is the number of decision trees, and $f_t(X)$ represents the prediction of the $t^{th}$ tree. Permutation importance quantified the contribution of covariates, prioritizing parameters like curvature and lineament density that govern fracture connectivity in hard-rock aquifers (Pham et al., 2022).The hyperparameter ranges for RF were determined based on best practice guidelines in ensemble learning for environmental modeling, prior hydrological applications (e.g., Pham et al., 2022; Lee et al., 2020), and preliminary pilot experiments on our dataset to balance accuracy and computational cost. The search space included number of trees [100, 200, 300], maximum tree depth [10, 15, 20], and minimum samples per leaf [1, 3, 5]. These ranges were chosen to capture the trade-off between model complexity and generalization, ensuring robustness in data-scarce and heterogeneous terrains. Hyperparameters, including tree depth (max_depth = 15) and minimum leaf samples (min_samples_leaf = 3), were selected from this search with 5-fold spatial cross-validation to address spatial autocorrelation (Wang et al., 2018).

Pseudo-code for RF Regression
Input: Training dataset *T* with features *X* (15 covariates) and labels *y* (groundwater depth)
Parameters: Number of trees = 200, max_*depth* = 15, min _*samples_leaf* = 3
1. For each tree *t* in 1 to 200:

a. Draw a bootstrap sample from T
b. Select a random subset of features
c. Grow a decision tree:

o At each node, split on the feature that maximizes variance reduction
o Stop when max_*depth* or min_*samples_leaf* is reached

2. Aggregate predictions from all trees (average for regression)
3. Evaluate performance using $R^2$ and Mean Squared Error (MSE) on test data
4. Compute permutation importance for each feature
Output: Predicted groundwater depths, feature importance scores

### 2.2.7. GB classifiers for GWPZ mapping

The GB classifier was employed to delineate GWPZs across Jharkhand's geologically heterogeneous aquifers. GB builds an ensemble of decision trees in a sequential manner, where each new tree attempts to correct the residuals of the previous ensemble (Friedman, 2001). The additive model at iteration $m$ is given by Eq. 8:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \tag{8}$$

Where $h_m(x)$ is a weak learner, and $\gamma_m$ is the learning rate. The hyperparameter ranges and search space were determined based on a combination of best-practice guidelines for boosting algorithms in environmental modeling, prior hydrological machine learning

applications (e.g., Dhilsath and Samuel, 2021; Lee et al., 2020), and preliminary pilot experiments on our dataset. The number of estimators [50, 100, 150, 200] was chosen to capture the trade-off between underfitting (too few trees) and overfitting (too many trees). Learning rates [0.01, 0.05, 0.1] reflected commonly effective values for structured tabular data, balancing model stability and convergence speed. Maximum tree depths [2, 3, 4] were kept shallow to prevent overfitting given the high dimensionality of hydrogeological features. Subsample ratios [0.6, 0.8, 1.0] were included to introduce stochasticity for better generalization. A comprehensive hyperparameter tuning process was conducted using randomized grid search (Dhilsath and Samuel, 2021) combined to optimize model performance. The optimal configuration: 150 trees, learning rate of 0.1, max depth of 3, and subsample ratio of 0.8, is selected based on the lowest validation loss. Early stopping is applied to terminate training once validation loss plateaued, ensuring efficient convergence and preventing overfitting.

Incorporating feature subsampling (80 %) and stochastic gradient descent, the model achieved improved generalization, particularly within complex hard-rock terrains common to Jharkhand. This approach ensures robust and spatially consistent GWPZs mapping in data-scarce and geologically varied environments.

Pseudo-code for GB Classification
Input: Training dataset $T$ with features $X$ and labels $y$ (GWPZ class: High/Medium/Low)
Parameters: $n\_estimators = 150$, $learning\_rate = 0.1$, $max\_depth = 3$, $subsample = 0.8$
1. Initialize model with constant prediction (log-odds of $y$)
2. For $m = 1$ to 150

a. Compute residuals = Negative gradient of loss function
b. Fit regression tree $h_m(x)$ to residuals with $max\_depth = 3$
c. Compute optimal weight $\gamma_m$ for $h_m(x)$ by minimizing loss
d. Update model: $F_m(x) = F_{m-1}(x) + learning\_rate \times \gamma_m \times h_m(x)$

3. Apply early stopping if validation loss stops improving
Output: Final classifier for GWPZ mapping

### 2.2.8. Performance matrix for assessing ML classifier

M L classifiers were employed to categorize GWPZs based on hydrological and environmental features. The primary objective of this classification process was to construct GWPZs for new, unseen data using a dataset generated from the GA-based clustering approach. This structured dataset ensured the classification models captured key groundwater characteristics across different spatial regions. To achieve this, five different classification models were implemented and evaluated: SVM (Naghibi et al., 2017), KNN (Zaib et al., 2022), NB (Phong and Pham, 2023), RF (Miraki et al., 2019), and GB (Sachdeva and Kumar, 2021). The classification performance was evaluated using key metrics, including Accuracy, Precision, Recall, and F1-Score. These metrics are computed using True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) to assess the effectiveness of the classification models in predicting GWPZs (Table 3).

### 2.2.9. Cluster validation metrics

The quality of clusters generated through GA-based clustering was evaluated using four widely adopted validation indices: Silhouette Score, Davies–Bouldin Index, and Dunn Index (Rojas-Thomas and Santos, 2021; Rubinos et al., 2024). These metrics provide complementary perspectives on cluster compactness and separation, ensuring reliable delineation of GWPZs.

*2.2.9.1. Silhouette score (S).* The Silhouette Score assesses how similar a sample is to its own cluster compared to other clusters. It is defined as (Eq. 9):

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{9}$$

where $(a(i))$ is the mean intra-cluster distance for point $(i)$, and $(b(i))$ is the mean nearest-cluster distance. Scores range from –1 to + 1, with values closer to 1 indicating well-separated clusters.

**Table 3**
Performance metrics for machine learning classifiers.

| Metrics | Formula | Description |
|---|---|---|
| Accuracy | $\frac{TP + TN}{TP + TN + FP + FN}$ | Measures the overall correctness of the classifier in predicting GWPZs. |
| Precision | $\frac{TP}{TP + FP}$ | Correctly evaluating how many regions predicted as high, medium, or low GWPZs are correct. |
| Recall | $\frac{TP}{TP + FN}$ | Determines the ability of the model to identify GWPZs correctly. |
| F1-Score | $\frac{2 * Precision * Recall}{Precision + Recall}$ | Provides a balance between Precision and Recall. |

*2.2.9.2. Davies–Bouldin index (DBI).* The DBI measures the average similarity between each cluster and its most similar cluster (Eq. 10):

$$DBI = \frac{1}{k}\sum_{i=1}^{k} \max_{j \neq i}\left(\frac{\sigma_i + \sigma_j}{d(c_i, c_j)}\right) \tag{10}$$

where $(k)$ is the number of clusters, $(\sigma_i)$ is the average intra-cluster distance of cluster $(i)$, and $(d(c_i, c_j))$ is the distance between cluster centroids. Lower values indicate better clustering.

*2.2.9.3. Dunn index (DI).* The Dunn Index captures the ratio of the minimum inter-cluster distance to the maximum intra-cluster distance (Eq. 11):

$$DI = \frac{\min_{i \neq j} d(c_i, c_j)}{\max_{k} \delta(C_k)} \tag{11}$$

where $(d(c_i, c_j))$ is the distance between clusters $(i)$ and $(j)$, and $(\delta(C_k))$ is the intra-cluster distance of cluster $(k)$. Higher values represent compact and well-separated clusters.

*2.2.10. Explainable AI for groundwater potential classification using LIME for GB classifier*

To boost the interpretability of the classification model, this study employed the LIME technique (Moral et al., 2024; Yariyan et al., 2022; Yen et al., 2021). LIME provides understanding of how individual predictions are made by approximating the behavior of complex models with locally interpretable surrogate models. This method allows us to assess the contribution of each feature in classifying GWPZs and validate the consistency of the machine learning model with hydrological knowledge.

LIME operates by perturbing input data and training a simpler, interpretable model to approximate the local behaviors of the complex model. This helps in understanding which features significantly influence the classification for different GWPZs. From the GB interpretation, it is observed that Curvature reflecting terrain shape, Distance from lineament such as faults and fractures, LULC, etc., are among the most influential factors because they directly affect how water moves across and infiltrates the landscape, with lineaments often serving as pathways for groundwater flow, and terrain curvature influencing water accumulation and runoff (Khan et al., 2021; Lee et al., 2020; Zou et al., 2025; Maity et al., 2024). LULC impacts recharge rates by determining surface permeability and vegetation cover. LIME helps verify this importance by showing how these features contribute to the classification of individual instances.

## 3. Results and discussion

### 3.1. Clustering

This study applies GA-based clustering to categorize the region's geographical data into distinct groups based on similar characteristics. This unsupervised learning approach effectively analyzes critical factors such as soil type, land use, topography, and rainfall patterns to identify areas with similar groundwater potential (Kodihal and Akhtar, 2024; Zhao et al., 2025a). The number of clusters (K=3) was obtained as an optimal value directly from the GA optimization process, which simultaneously determined the best cluster count and centroid positions based on the defined fitness function. The GA clustering process resulted in three optimal clusters, ensuring a well-balanced data partitioning. The corresponding cluster-wise fitness values were 0.33, 0.36, and 0.34, reflecting the

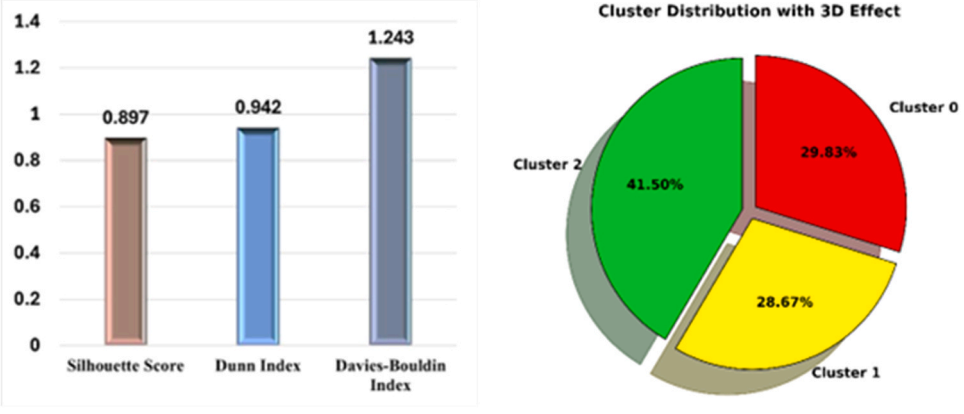


**Fig. 3.** (a) Performance evaluation of GA-based clustering (b) Cluster-wise distribution of GWPZs using GA-based clustering.

optimal distribution of spatial patterns within the dataset.

Multiple validation metrics have been utilized to evaluate the performance of the clustering process, including the Silhouette Score, Davies-Bouldin Index, Calinski-Harabasz Index, and Dunn Index (Rojas-Thomas and Santos, 2021; Rubinos et al., 2024). The obtained scores indicate a well-separated and compact clustering outcome, as illustrated in Fig. 3(a). A high Silhouette Score (0.897) and Dunn Index (0.942) suggest the clusters exhibit strong internal cohesion and good separation. Conversely, the Davies-Bouldin Index (1.243) further reinforces the effectiveness of the GA-based clustering in forming well-defined groups. In addition to performance evaluation, the distribution of the resulting clusters was analysed to understand the spatial grouping of GWPZs. Fig. 3(b) presents the proportion of data points assigned to each cluster.

Fig. 4 illustrates the variability of groundwater depths across GA-derived clusters classified as High, Medium, and Low potential zones. The High-potential cluster is associated with shallower depths, reflecting areas with better groundwater availability, while the Low-potential cluster exhibits greater median depths, indicating limited water availability. The presence of outliers in all clusters highlights spatial heterogeneity in groundwater conditions. This visualization validates that the GA-based clustering effectively distinguishes potential zones based on depth variability, as shown in Fig. 5

The complex dataset was transformed by leveraging GA for clustering into meaningful and interpretable clusters, providing a clearer visualization of GWPZs. This technique improves the spatial understanding of hydrological variations and supports groundwater resource management. The clustered outputs serve as a crucial foundation for further analysis in groundwater susceptibility mapping, as depicted in Fig. 4. This figure illustrates well points categorized through GA-based clustering, where each well represents groundwater data spanning from 2013 to 2023.

### 3.2. MK trend test

Groundwater level trends identified using the MK test have been widely applied to assess long-term water resource changes (Barman et al., 2023; Mishra et al., 2023). The results of the MK trend test for the clusters showed significant trends in groundwater levels. Cluster 1 demonstrated an even stronger negative trend, with an S value of −32,355.0, a Z value of −2.66, and a p-value of 0.0029, reflecting a very strong and highly significant downward trend in groundwater levels ($p < 0.01$). Similarly, Cluster 2 exhibited a highly significant negative trend, with an S value of −52,816.0, a Z value of −2.13, and a p-value of 0.0077, indicating a strong decline in groundwater levels over time ($p < 0.01$). Meanwhile, Cluster 3 showed a moderate but significant negative trend, with an S value of −7037.0, a Z value of −1.48, and a p-value of 0.03, indicating a clear decline in groundwater levels ($p < 0.05$).

Regions with the least negative trend or stable groundwater levels are categorized as high-potential zones, indicating better groundwater availability. Regions with moderate negative trends in groundwater levels are classified as medium potential zones. In contrast, those exhibiting strong to very strong negative trends are categorized as low potential zones due to significant groundwater depletion risks. Based on these classifications, Cluster 1 is labeled as high potential, Cluster 2 as medium potential, and Cluster 3 as low potential, as illustrated in Fig. 6.The MK test analysis revealed both upward and downward trends in groundwater levels across different observation sites. Positive Z-statistics, supported by Sen's slope, suggest zones experiencing long-term recharge or reduced extraction, while negative values highlight areas undergoing continuous depletion due to over-extraction and limited recharge. The statistical significance ($p < 0.05$) confirms that these observed trends are not random but are driven by persistent hydroclimatic and anthropogenic factors. This critical assessment underscores the necessity of integrating groundwater trend analysis into sustainable management planning.

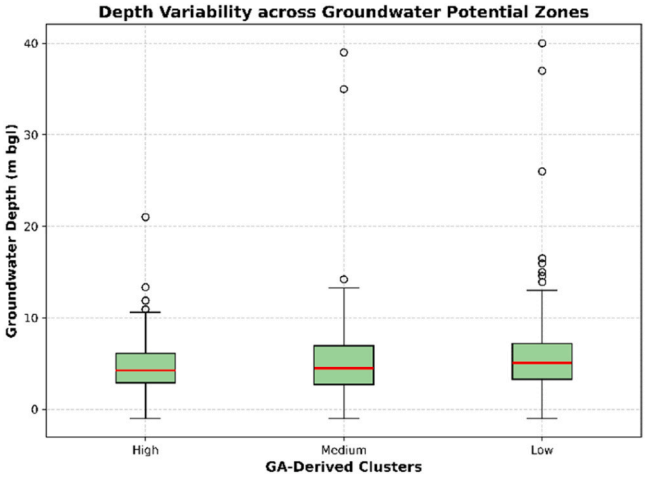


**Fig. 4.** Depth variability across GWPZs derived from GA clustering.

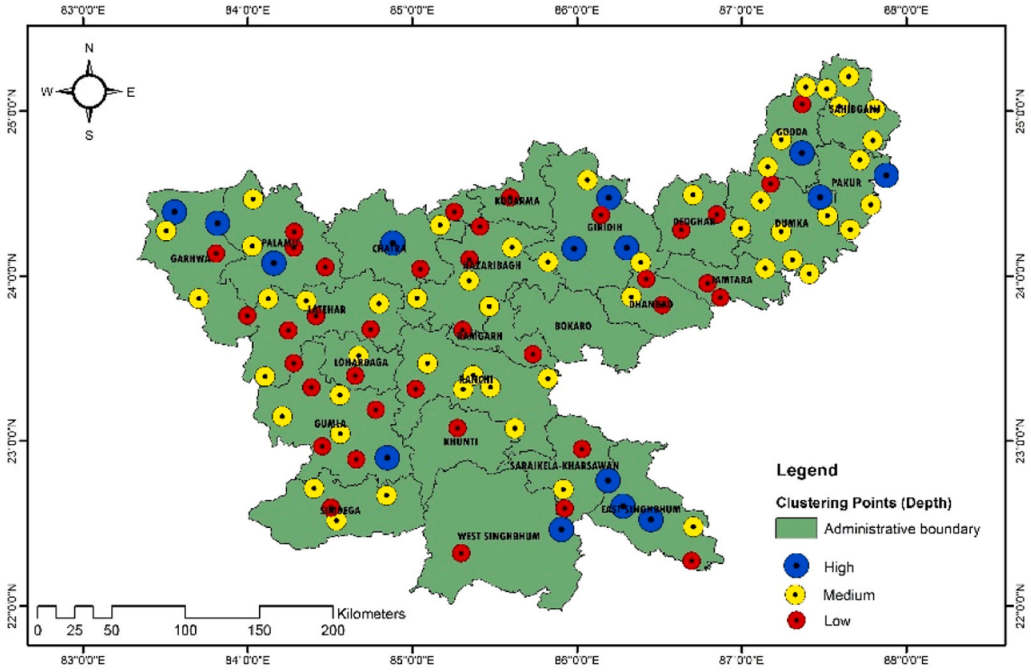


**Fig. 5.** Categorization of groundwater potential using GA-based clustering.

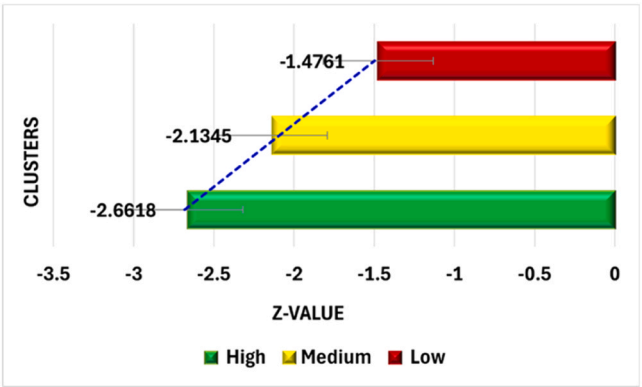


**Fig. 6.** Trend analysis of groundwater levels using MK test.

### 3.3. RF regression

In this study, RF Regression is employed to predict groundwater depth based on the clusters derived from GA-based clustering applied to the original dataset. The training (80 %) and testing (20 %) data were derived from the clustered dataset, ensuring that the model effectively captures variations in groundwater characteristics across different zones. This regression model's primary objective is to predict groundwater depth for new, unseen data based on key hydrological and environmental factors. By integrating an ensemble of decision trees, RF Regression minimizes overfitting while improving prediction accuracy. The model's performance was evaluated using Mean Squared Error (MSE) and the Coefficient of Determination ($R^2$).

The obtained MSE value of 2.56 indicates a low prediction error, while the $R^2$ (Mahakur et al., 2025) value of 0.91 demonstrates a highly accurate fit of the model to the data. Several studies have demonstrated the effectiveness of regression techniques for groundwater depth prediction due to its ability to handle feature interactions (Majumdar et al., 2022; Naghibi et al., 2017; Pradhan et al., 2024). In addition to $R^2$ and MSE, Willmott's Index of Agreement (WI) and Percent Bias (PBIAS) (Samantaray et al., 2025) is also employed to provide a more robust evaluation of model performance. The obtained WI of 0.89 indicates a high level of agreement between predicted and observed values, while the PBIAS of 0.24 suggests that the model predictions are nearly unbiased. Figs. 7(a) and (b) compares actual and predicted values for training and testing datasets using the RF Regressor.

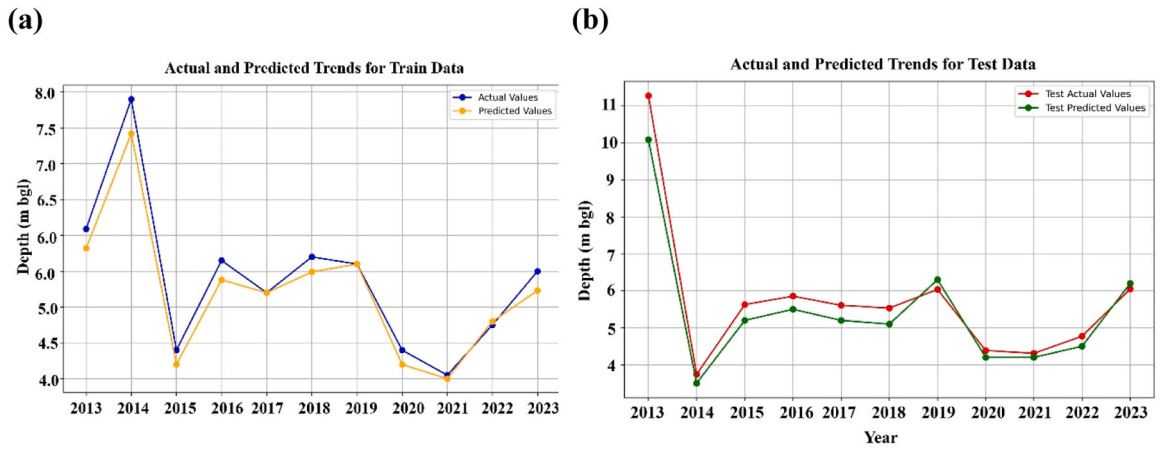


**Fig. 7.** Actual and Predicted Values on Training (a) and Testing (b) Data with RF Regressor.

### *3.4. Performance evaluation of GB classifier*

Table 4 summarizes the key parameters and their optimal values for all classifiers considered, ensuring transparency in the comparative analysis. GB consistently outperformed the other classifiers across all metrics (Table 5 and Fig. 8).

Fig. 9 further demonstrates GB's superior predictive ability, with the confusion matrix showing high class-wise accuracy and the ROC curve (Janitza et al., 2013) indicating strong discriminatory power. RF also performed competitively, while SVM (Naghibi et al., 2017) and KNN (Zaib et al., 2022) yielded slightly lower results. NB exhibited relatively lower accuracy, likely due to its feature independence assumption. Given this superior performance, GB was selected for further interpretability analysis using LIME and for constructing the final GWPZs map, as detailed in Sections 3.5 and 3.6.Fig. 10

### *3.5. LIME-based explanation of model outcomes*

Lime analysis showed that curvature, proximity to lineaments, LULC, and TWI consistently dominated classification across all GWPZs (Fig. 11.). because they directly affect how water moves across and infiltrates the landscape, with lineaments often serving as pathways for groundwater flow and terrain curvature influencing water accumulations and runoff (Khan et al., 2021; Lee et al., 2020; Zou et al., 2025). LULC impact recharge rates by determining surface permeability and vegetation cover. By applying LIME to different groundwater potential classes, the feature contributions for three representative samples are visualized in Fig. 11. In low-potential zones, convex terrain, large distances from lineaments, and dry surface conditions restricted recharge. By contrast, concave topography, proximity to lineaments, favorable soil moisture, and vegetated land cover supported higher recharge in moderate and high-potential zones.

These findings align with real-world hydrological conditions. For instance, regions with strong curvature, proximity to lineaments, and favorable soil and vegetation features tend to support higher groundwater recharge (class 2: High Potential Zone). In contrast, areas with steep or convex terrain, limited surface water presence, and large distances from hydrological features often correspond to lower groundwater potential (class 0: Low Potential Zone). By explicitly linking model predictions with hydrological variables and referencing categorical codes (Table 2), the LIME results are now both interpretable and scientifically grounded. This increases the reliability of classification results and supports data-driven decision-making in groundwater resource management.

Beyond classification accuracy, such interpretability provides practical value for groundwater management. By pinpointing the specific features that drive predictions in different zones, LIME boosts model transparency and supports targeted interventions. For example, prioritizing recharge structures in concave terrain near lineaments or avoiding development in convex, runoff-prone regions. Integrating explainable AI into groundwater studies thus ensures that predictions are not only reliable but also actionable for sustainable resource planning.

**Table 4**
Parameters and optimal values considered for the classifiers.

| Model | Parameters considered | Optimal values used in study |
|---|---|---|
| RF | Number of trees (n_estimators), Max depth, Min samples per leaf, Max features | 200, 15, 3, sqrt |
| GB | Number of estimators, Learning rate, Max depth, Subsample ratio, Feature subsampling | 150, 0.1, 3, 0.8, 80 % |
| SVM | Kernel type, Regularization parameter (C), Gamma | RBF, 10, scale |
| KNN | Number of neighbors (k), Distance metric, Weights | 5, Euclidean, uniform |
| NB | Distribution assumption | Gaussian |

**Table 5**
Comparative analysis of different models based on classification performance metrics.

| Models | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| SVM | 94.53 % | 95.01 % | 94.51 % | 94.31 % |
| KNN | 94.68 % | 95.13 % | 94.67 % | 94.46 % |
| NB | 93.84 % | 94.45 % | 93.84 % | 93.57 % |
| RF | 95.51 % | 95.82 % | 95.51 % | 95.37 % |
| GB | 95.67 % | 95.99 % | 95.66 % | 95.55 % |

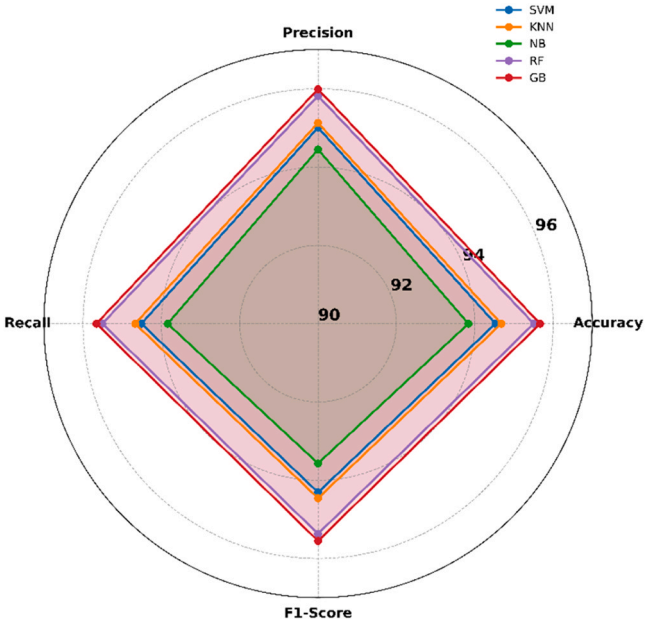


**Fig. 8.** Radar chart visualization of classifier performance metrics.

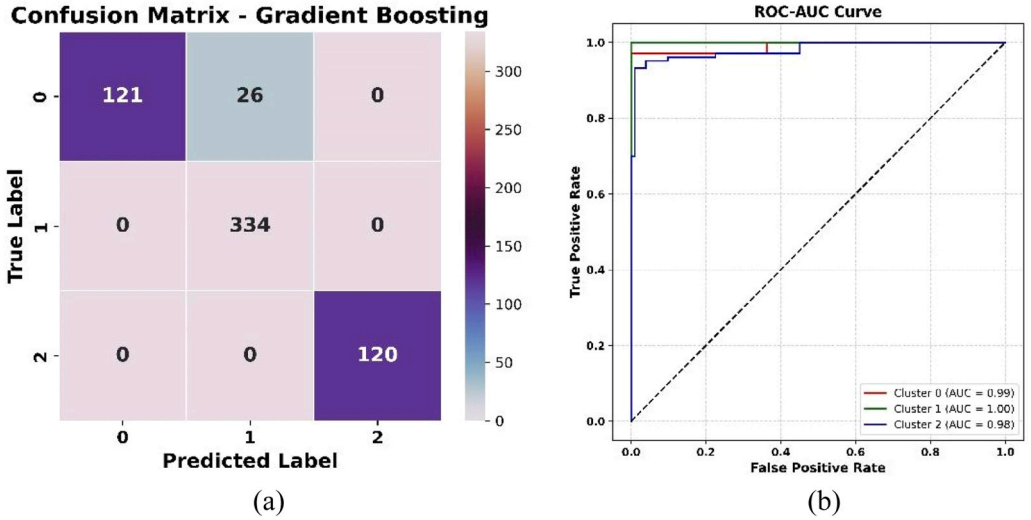


**Fig. 9.** Performance evaluation of the GB classifier for GWPZs classification (a) confusion matrix illustrating the distribution of correctly and incorrectly classified samples across groundwater potential classes (b) AUC–ROC curve showing the trade-off between true positive rate and false positive rate.

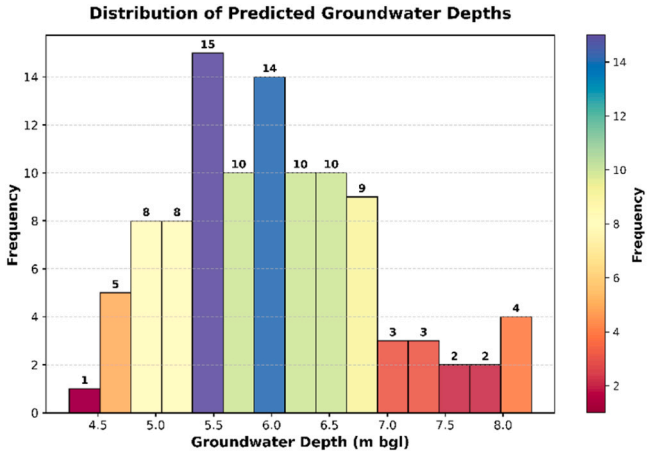


**Fig. 10.** Histogram of predicted groundwater depths showing the frequency distribution across different depth ranges (meter bgl).

### *3.6. Construction of groundwater potential map using GB*

The GB model classified Jharkhand into three GWPZs (Fig. 12). High-potential areas are concentrated in West Singhbhum, East Singhbhum, Saraikela-Kharsawan, Khunti, and Gumla; medium-potential regions include Ranchi, Bokaro, Dhanbad, and Hazaribagh; and low-potential zones occur in Giridih, Pakur, parts of Simdega, and Garhwa. These results highlight critical areas requiring targeted recharge and conservation measures.

The GWPZs map was developed using GB, an ensemble learning technique capable of capturing complex relationships between hydrogeological, topographical, and climatic factors. Unlike traditional deterministic or simple statistical models, GB iteratively refines weak learners to increase predictive accuracy, ensuring optimal classification of GWPZs. The primary goal is to classify regions into high, medium, and low GWPZs for new, unseen data, supporting sustainable groundwater management. Similar approaches integrating machine learning and GIS have been employed elsewhere; for example, a study in North Central Nigeria combined remote sensing with weighted overlay analysis to identify high-potential zones in sedimentary terrain and low-potential zones in rugged areas (Adeyeye et al., 2019).

The methodology involved data collection and preprocessing, incorporating hydrogeological and environmental attributes derived from GA-based clustering. The GB model was then trained and validated to classify groundwater availability. Model performance was evaluated using accuracy, precision, recall, and F1-score, and validated predictions were visualized in ArcGIS (10.4), producing the final GWPZ map.

Interpretation of the map revealed that areas with high groundwater potential generally have stable water levels, high recharge rates, favorable soil permeability, and dense vegetation, making them suitable for sustainable extraction. Regions with medium potential exhibit moderate availability influenced by seasonal variations, while low-potential areas show significant depletion trends and require urgent water conservation measures, such as rainwater harvesting, artificial recharge, and controlled extraction policies.

Groundwater management is particularly critical in Jharkhand, where over 60 % of irrigation, 85 % of rural water supply, and nearly half of urban water consumption depend on groundwater (Saha and Ray, 2019; Raj et al., 2024). Previous studies have emphasized the integration of hydrogeological, climatic, and topographical variables to improve groundwater prediction accuracy, including ensemble machine learning applications in Rajasthan (Machiwal et al., 2011). ML-GIS integration has consistently demonstrated improved prediction accuracy and improved spatial delineation of GWPZs (Lee et al., 2020; Kumar et al., 2023; Vafadar et al., 2023).

### *3.7. Method transferability*

While this study presents a robust framework for delineating GWPZs in Jharkhand, its utility extends beyond local application, provided key contextual adaptations are addressed. The modular integration of GA-based clustering, RF regression, and GB classification provides a transferable architecture for fractured crystalline aquifers globally. However, successful application to other hydrogeological settings depends on careful consideration of data representativeness, predictor relevance, and spatiotemporal variability.

#### *3.7.1. Transferability potential*

The methodology developed in this study benefits from the use of public-domain, globally accessible datasets such as ERA5 hydroclimatic reanalysis (precipitation, runoff, evaporation, soil moisture), MODIS vegetation indices (NDVI, NDWI), SRTM-derived digital elevation models, FAO global soil datasets, and geological information from national or regional survey agencies (Srivastava and Jain, 2025). The consistent formatting and broad availability of these datasets make the framework highly adaptable to other

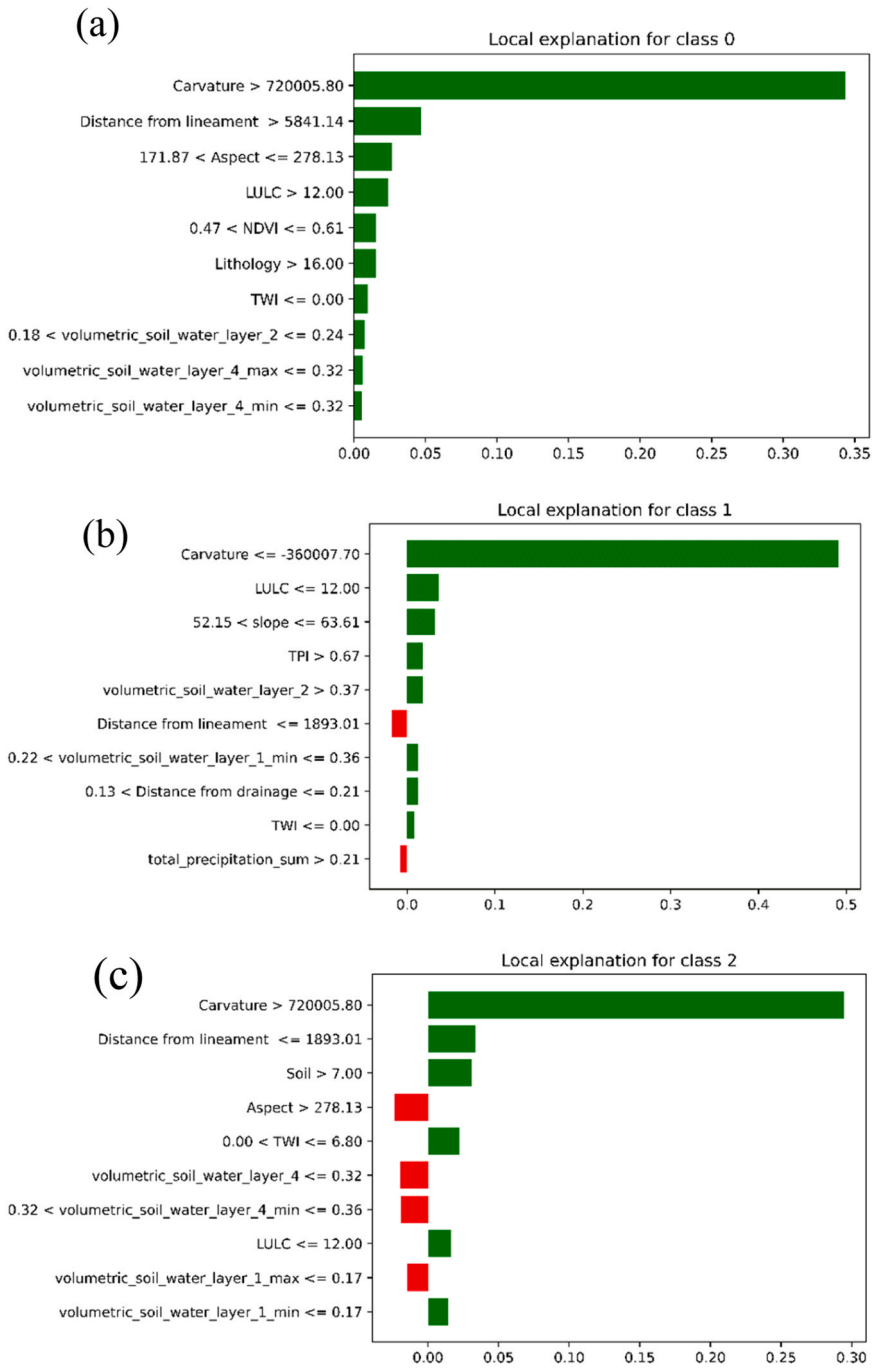


**Fig. 11.** LIME-based visual explanations for GWPZs classification: (a) class 0; (b) class 1; (c) class 2.

fractured aquifers in data-scarce regions, including the Brazilian Shield, Southern African cratons, and Ethiopian Highlands. These regions share key environmental and geological characteristics with Jharkhand's fractured hard-rock aquifers, including Precambrian crystalline basement rocks, sparse but critical fracture-controlled groundwater systems, and similar challenges of high surface runoff and limited natural recharge. These analogs exhibit comparable hydrogeological heterogeneity, climatic variability, and land-use pressures that affect groundwater availability and recharge dynamics. Furthermore, these regions commonly suffer from limited in-situ groundwater monitoring, making remote sensing and model-based approaches essential for resource assessment. The public-domain datasets used in our study are globally consistent and regularly updated, ensuring compatible spatial and temporal coverage for these territories. Although local calibration and validation would be required to account for site-specific factors such as fracture density, land cover types, and human activities, the adaptable GA–ML framework's reliance on universally available

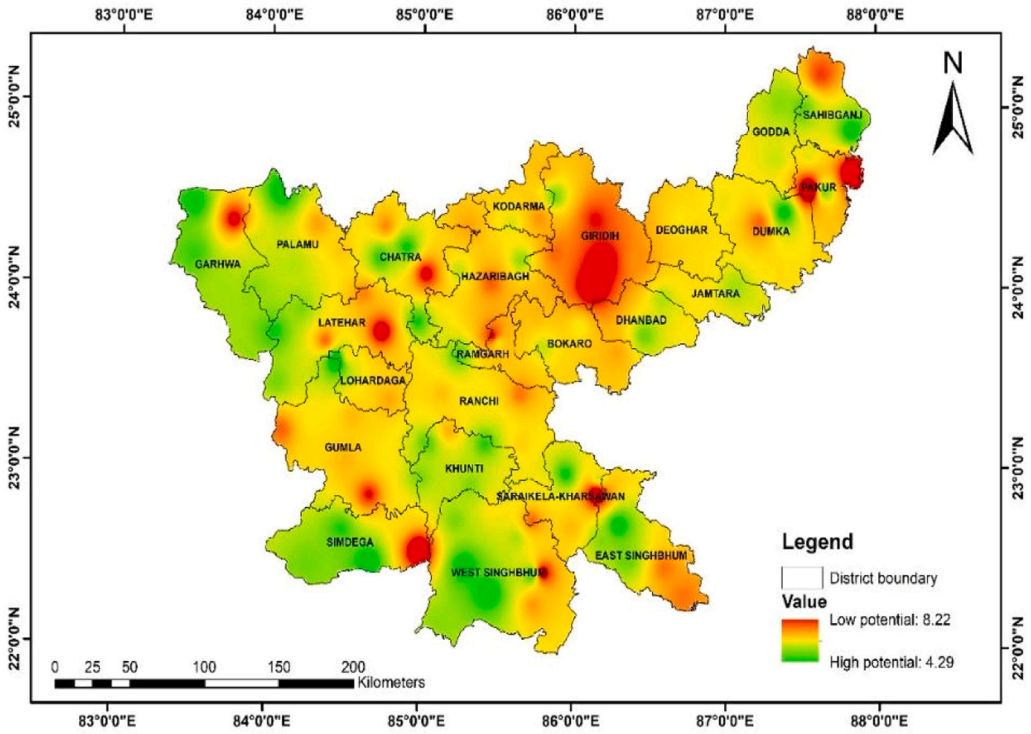


**Fig. 12.** Spatial distribution of GWPZs in Jharkhand State.

environmental variables gives it strong potential for transferability. Thus, the framework serves as a scalable, cost-effective blueprint for sustainable groundwater governance in various fractured aquifer settings worldwide where direct data remain scarce.

A key strength lies in the flexibility of the GA-based clustering technique, which does not depend on predefined geological boundaries but instead identifies recharge-relevant clusters dynamically, making it particularly suited for lithologically complex terrains. Additionally, the use of explainable artificial intelligence LIME improves the transparency and interpretability of model outputs, enabling domain experts in various regions to validate and trust the classification logic based on their local hydrogeological knowledge. This study provides a robust and scalable alternative to traditional empirical and geostatistical methods by integrating ML-based classification with GIS-based spatial visualization. This approach enables policymakers, hydrologists, and environmental managers to prioritize conservation efforts, develop sustainable groundwater extraction policies, and mitigate over-extraction risks, ultimately ensuring equitable and long-term groundwater availability in the region.

### *3.8. Contribution to sustainable development goals (SDGs)*

By allowing the identification of groundwater-rich and water-stressed regions, this study directly supports Sustainable Development Goal 6 (Clean Water and Sanitation). This allows for targeted interventions for equitable water access and conservation. Target 6.6, which focuses on preserving and rehabilitating ecosystems related to water, including aquifers, is in line with the machine learning-based spatial approach. Moreover, the approach supports SDG 13 (Climate Action) by emphasizing water resilience and incorporating hydroclimatic factors, particularly in relation to adjusting to water scarcity in the face of climate change (Jain et al., 2024). Similar integration of AI-driven technologies for sustainable groundwater planning has been shown in studies by Rao et al. (2025) in Visakhapatnam District, Andhra Pradesh, India, and by Masud et al. (2025) across Asia. However, the current study stands out for providing a scalable and transferable model that is especially appropriate for fractured aquifer systems. This approach increases decision-making transparency for policymakers by providing clear, data-driven understanding of groundwater availability and management needs. Overall, it serves as a practical and replicable framework for other regions facing similar hydrogeological and socio-environmental challenges, supporting the development and implementation of effective sustainability policies.

### *3.9. Advantages, limitations, and scope of the proposed model*

The proposed GA-ML framework for GWPZs in fractured aquifers provides several significant advantages. First, the integration of GA-driven clustering dynamically optimizes spatial hydrogeological heterogeneity, overcoming the limitations of traditional, static clustering techniques like k-means that usually struggle in complex fractured rock terrains. This adaptive clustering improves spatial compactness and separation, resulting in more accurate and realistic delineation of GWPZs. Second, the incorporation of multi-decadal

temporal trend analysis via the MK test allows the framework to go beyond static mapping, providing key understandings of the temporal dynamics of groundwater depletion within each zone. Third, the use of RF regression and GB classification integrates ensemble machine learning methods known for their robustness against overfitting, nonlinear relationship modeling, and strong predictive accuracy, particularly suited for data-scarce environments. Additionally, explainable AI through LIME yields transparency and interpretability by elucidating key drivers influencing groundwater potential, fostering trust and implementable visions for policymakers. The reliance on publicly available, globally consistent environmental and remote sensing datasets further ensures scalability and adaptability to other fractured aquifers worldwide, facilitating reproducibility and broader applicability.

Despite these strengths, the model also has limitations that pave the way for future research. While the GA-ML framework effectively integrates spatial and temporal groundwater assessments, the temporal component currently functions more as post hoc trend validation rather than being fully integrated into the zoning and classification algorithm. Deepening this temporal integration, such as through spatiotemporal graph neural networks or hybrid AI architectures, could further improve dynamic prediction capabilities. The model's reliance on available surface and hydrogeological parameters means local site-specific factors such as detailed fracture density, subsurface heterogeneities, and anthropogenic influences may be underrepresented, necessitating future incorporation of higher-resolution geophysical data and human activity indices. Furthermore, though the public datasets employed enable global transferability, the model requires local calibration and validation to accommodate regional environmental variability. Future directions include coupling this framework with real-time monitoring systems and expanding AI explainability tools to better assist decision-makers. By addressing these limitations, the framework can evolve into a more comprehensive, predictive, and operational tool for sustainable groundwater management across diverse fractured aquifer systems globally.

## 4. Conclusions

This study advances the science of GWPZs mapping in fractured hard-rock aquifers by presenting a novel, fully integrated GA-ML framework that unites spatial optimization, multi-decadal trend analysis, and explainable AI into a single operational workflow. Applied to Jharkhand, India, and drawing on several globally available hydrogeological, topographic, climatic, and remote-sensing factors, the framework delineated three distinct GWPZs with high spatial accuracy (Silhouette score 0.89; Dunn Index 0.94) and robust predictive performance (RF regression $R^2 \approx 0.91$; GB classification precision 95.99). Crucially, the MKl trend analysis revealed statistically significant groundwater depletion across all potential zones, demonstrating the added value of temporal diagnostics embedded alongside spatial zoning. Model interpretation via LIME identified terrain curvature, distance from lineaments, and land use/land cover as dominant controls, a finding that directly addresses our research hypotheses on the relative influence of surface parameters and anthropogenic change. By generating transparent, interpretable geospatial outputs, the framework provides policymakers a science-backed basis for implementing Jharkhand's 2025 Groundwater Act through targeted extraction limits and recharge interventions in depletion hotspots.

Beyond its strong local performance, the framework's transferability is a major contribution: by relying on globally consistent, public-domain datasets (ERA5, MODIS, SRTM, FAO soils), it is adaptable to other data-scarce fractured aquifers sharing similar hydrogeological characteristics, such as those in the Brazilian Shield, Southern African cratons, and Ethiopian Highlands. Nevertheless, full integration of temporal dynamics into the clustering process, incorporation of finer-scale fracture and geophysical data, and explicit modeling of climate change and human water-use impacts present promising future research directions. Extending the framework with advanced deep learning architectures, real-time monitoring inputs, and higher-resolution environmental covariates could boost predictive power and temporal responsiveness. Ultimately, by bridging optimization, prediction, interpretability, and policy alignment, this work sets a methodological benchmark for reproducible, dynamic groundwater resource assessments in some of the world's most hydro-geologically challenging and water-stressed environments.

## Funding

Not applicable.

## CRediT authorship contribution statement

**Mishra Arun Pratap:** Writing – review & editing, Writing – original draft, Visualization, Methodology. **Rajarshi Bhattacharjee:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Conceptualization. **Sayantan Majumdar:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Methodology, Funding acquisition, Conceptualization. **Akhouri Pramod Krishna:** Writing – review & editing, Supervision, Project administration, Investigation, Conceptualization. **Abhijit Mustafi:** Validation, Supervision, Project administration, Conceptualization. **Richa Sharma:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Virendra Singh Rathore:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Project administration, Conceptualization. **Debjani Mustafi:** Writing – review & editing, Writing – original draft, Validation, Supervision, Investigation, Data curation, Conceptualization. **Aman Srivastava:** Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis. **Poonam Moral:** Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Methodology, Formal analysis, Conceptualization. **Prashant Parasar:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

## Declaration of Generative AI and AI-assisted technologies in the writing process

While preparing this work, the authors improved their language using OpenAI Cloud Computing Platform Version 4, Grammarly Premium, and Quillbot. After using these services, the authors reviewed and edited the content as needed and took full responsibility for the publication's content.

## Declaration of Competing Interest

Author Arun Pratap Mishra was employed by the company Earthtree Data Services Private Limited The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Acknowledgment

This research is part of a project supported by Department of Science and Technology (DST) (Government of India) under the PURSE 2022 Grant (SR/PURSE/2022/130(G)). We also thank the Desert Research Institute, Raggio Parkway Reno, Nevada, United States of America, for covering the open-access publication costs.

*Institutional Review Board Statement*

Not applicable

## Data availability

Data will be made available on request.

## References

Abegeja, D., Nedaw, D., 2024. Identification of groundwater potential zones using geospatial technologies in meki catchment, Ethiopia. Geol. Ecol. Landsc. 1–16. https://doi.org/10.1080/24749508.2024.2392380.

Adeyeye, O.A., Ikpokonte, E.A., Arabi, S.A., 2019a. GIS-based groundwater potential mapping within dengi area, north central Nigeria. Egypt. J. Remote Sens. Space Sci. 22 (2), 175–181. https://doi.org/10.1016/j.ejrs.2018.04.003.

Akhter, G., Ge, Y., Iqbal, N., Shang, Y., Hasan, M., 2021. Appraisal of remote sensing technology for groundwater resource management perspective in Indus basin. Sustainability 13 (17), 9686. https://doi.org/10.3390/su13179686.

Arulbalaji, P., Padmalal, D., Sreelash, K., 2019. GIS and AHP techniques based delineation of groundwater potential zones: a case study from Southern Western Ghats, India. Sci. Rep. 9 (1), 2082. https://doi.org/10.1038/s41598-019-38567-x.

Ashwini, K., Verma, R.K., Sriharsha, S., Chourasiya, S., Singh, A., 2023. Delineation of groundwater potential zone for sustainable water resources management using remote sensing-GIS and analytic hierarchy approach in the state of jharkhand, India. Groundw. Sustain. Dev. 21, 100908. https://doi.org/10.1016/j.gsd.2023.100908.

Bai, T., Tahmasebi, P., 2023. Graph neural network for groundwater level forecasting. J. Hydrol. 616, 128792. https://doi.org/10.1016/j.jhydrol.2022.128792.

Barman, J., Biswas, B., Soren, D.D.L., 2023. Groundwater trend analysis and regional groundwater drought assessment of a semi-arid region of rajasthan, India. Environ. Dev. Sustain. 27 (2), 3415–3439. https://doi.org/10.1007/s10668-023-04022-1.

Biswakalyani, C., Samantaray, S., Satpathy, D.P., 2024. Assessment of spatio-temporal characteristics of groundwater quality at khordha district, India. Discov. Civ. Eng. 1, 28. https://doi.org/10.1007/s44290-024-00032-8.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324/METRICS.

Central Ground Water Board (CGWB). (2023). Dynamic groundwater resources of Jharkhand (2020–2021). In Ministry of Jal Shakti, Government of India.

Chandra, A., Sharath, M.N., Pani, A., Sahu, P.K., 2021. A multi-objective genetic algorithm approach to design optimal zoning systems for freight transportation planning. J. Transp. Geogr. 92, 103037. https://doi.org/10.1016/j.jtrangeo.2021.103037.

Cui, Z., Chen, Q., Luo, J., Ma, X., Liu, G., 2024. Characterizing subsurface structures from hard and soft data with multiple-condition fusion neural network. Water Resour. Res. 60 (11), e2024WR038170. https://doi.org/10.1029/2024WR038170.

Dhilsath, F.M., Samuel, S.J., 2021. Hyperparameter tuning of ensemble classifiers using grid search and random search for prediction of heart disease. Computational Intelligence and Healthcare Informatics. Wiley, pp. 139–158. https://doi.org/10.1002/9781119818717.ch8.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29 (5). https://doi.org/10.1214/aos/1013203451.

Garai, G., Chaudhuri, B.B., 2004. A novel genetic algorithm for automatic clustering. Pattern Recognit. Lett. 25 (2), 173–187. https://doi.org/10.1016/j.patrec.2003.09.012.

Hasanuzzaman, M., Mandal, M.H., Hasnine, M., Shit, P.K., 2022. Groundwater potential mapping using multi-criteria decision, bivariate statistic and machine learning algorithms: evidence from chota nagpur plateau, India. Appl. Water Sci. 12 (4), 58. https://doi.org/10.1007/s13201-022-01584-9.

Hirsch, R.M., Slack, J.R., Smith, R.A., 1982. Techniques of trend analysis for monthly water quality data. Water Resour. Res. 18 (1), 107–121. https://doi.org/10.1029/WR018I001P00107.

Jain, S., Srivastava, A., Khadke, L., Chatterjee, U., Elbeltagi, A., 2024. Global-scale water security and desertification management amidst climate change. Environ. Sci. Pollut. Res. 31 (49), 58720–58744. https://doi.org/10.1007/s11356-024-34916-0.

Janitza, S., Strobl, C., Boulesteix, A.-L., 2013. An AUC-based permutation variable importance measure for random forests. BMC Bioinforma. 14 (1), 119. https://doi.org/10.1186/1471-2105-14-119.

Jasechko, S., Seybold, H., Perrone, D., Fan, Y., Shamsudduha, M., Taylor, R.G., Fallatah, O., Kirchner, J.W., 2024. Rapid groundwater decline and some cases of recovery in aquifers globally. Nature 625 (7996), 715–721. https://doi.org/10.1038/s41586-023-06879-8.

Kendall, M.G. 1975. Rank Correlation Methods. 4th Edition, Charles Griffin, London. - References - Scientific Research Publishing. (n.d.). Retrieved September 5, 2023, from ⟨https://www.scirp.org/⟩(S(351jmbntvnsjt1aadkposzje))/reference/ReferencesPapers.aspx?ReferenceID=2099295.

Khan, U., Faheem, H., Jiang, Z., Wajid, M., Younas, M., Zhang, B., 2021. Integrating a GIS-Based Multi-Influence factors model with hydro-geophysical exploration for groundwater potential and hydrogeological assessment: a case study in the karak watershed, Northern Pakistan. Water 13 (9), 1255. https://doi.org/10.3390/w13091255.

Kodihal, S., Akhtar, M.P., 2024. Sustainable groundwater recharge potential zone identification: an AHP-OWA approach integrating future rainfall and land-use projections. Water Resour. Manag. 38 (3), 1079–1098. https://doi.org/10.1007/s11269-023-03710-x.
Kumar, M., Singh, P., Singh, P., 2023. Machine learning and GIS-RS-based algorithms for mapping the groundwater potentiality in the bundelkhand region, India. Ecol. Inform. 74, 101980. https://doi.org/10.1016/j.ecoinf.2023.101980.
Lee, S., Hyun, Y., Lee, S., Lee, M.-J., 2020a. Groundwater potential mapping using remote sensing and GIS-Based machine learning techniques. Remote Sens. 12 (7), 1200. https://doi.org/10.3390/rs12071200.
Lei, J., Fang, H., Zhu, Y., Chen, Z., Wang, X., Xue, B., Wang, N., 2024. GPR detection localization of underground structures based on deep learning and reverse time migration. NDT E Int. 143, 103043. https://doi.org/10.1016/j.ndteint.2024.103043.
Machiwal, D., Jha, M.K., Mal, B.C., 2011. Assessment of groundwater potential in a Semi-Arid region of India using remote sensing, GIS and MCDM techniques. Water Resour. Manag. 25 (5), 1359–1386. https://doi.org/10.1007/s11269-010-9749-y.
Mahakur, V., Mahakur, V.K., Samantaray, S., Ghose, D.K., 2025. Prediction of runoff at ungauged areas employing interpolation techniques and deep learning algorithm. HydroResearch 8, 265–275. https://doi.org/10.1016/j.hydres.2024.12.001.
Maity, R., Srivastava, A., Sarkar, S., Khan, M.I., 2024. Revolutionizing the future of hydrological science: Impact of machine learning and deep learning amidst emerging explainable AI and transfer learning. Appl. Comput. Geosci. 24, 100206. https://doi.org/10.1016/j.acags.2024.100206.
Majumdar, S., Smith, R., Conway, B.D., Lakshmi, V., 2022. Advancing remote sensing and machine learning-driven frameworks for groundwater withdrawal estimation in arizona: linking land subsidence to groundwater withdrawals. Hydrol. Process. 36 (11). https://doi.org/10.1002/hyp.14757.
Mallick, J., Talukdar, S., Alsubih, M., Salam, R., Ahmed, M., Kahla, N.Ben, Shamimuzzaman, Md, 2021. Analysing the trend of rainfall in asir region of Saudi Arabia using the family of Mann-Kendall tests, innovative trend analysis, and detrended fluctuation analysis. Theor. Appl. Climatol. 143 (1–2), 823–841. https://doi.org/10.1007/s00704-020-03448-1.
Mann, H.B., 1945. Nonparametric tests against trend. Econometrica 13 (3), 245. https://doi.org/10.2307/1907187.
Masud, M.M., Shamem, A.S.M., Saif, A.N.M., Bari, M.F., Mostafa, R., 2025. The role of artificial intelligence in sustainable water management in Asia: a systematic literature review with bibliographic network visualization. Int. J. Energy Water Resour. 9 (1), 247–265. https://doi.org/10.1007/s42108-024-00319-7.
Miraki, S., Zanganeh, S.H., Chapi, K., Singh, V.P., Shirzadi, A., Shahabi, H., Pham, B.T., 2019. Mapping groundwater potential using a novel hybrid intelligence approach. Water Resour. Manag. 33 (1), 281–302. https://doi.org/10.1007/s11269-018-2102-6.
Mishra, S., Chauhan, M., Sundaramurthy, S., 2023. Assessment of groundwater trends in bhopal, madhya pradesh: a statistical approach. Sustainability 15 (15), 11971. https://doi.org/10.3390/su151511971.
Mohammed, M.A.A., Szabó, N.P., Szűcs, P., 2025. High-resolution characterization of complex groundwater systems using wireline logs analyzed with machine learning classifiers and isometric mapping techniques. Model. Earth Syst. Environ. 11 (2), 85. https://doi.org/10.1007/s40808-024-02263-1.
Moral, P., Mustafi, D., Sahana, S.K., 2024. PODBoost: an explainable AI model for polycystic ovarian syndrome detection using grey wolf-based feature selection approach. Neural Comput. Appl. 36, 18627–18644. https://doi.org/10.1007/s00521-024-10171-9.
Msaddek, M.H., Moumni, Y., Ayari, A., El May, M., Chenini, I., 2022. Artificial intelligence modelling framework for mapping groundwater vulnerability of fractured aquifer. Geocarto Int. 37 (25), 10480–10510. https://doi.org/10.1080/10106049.2022.2037729.
Mukherjee, I., Singh, U.K., 2020a. Delineation of groundwater potential zones in a drought-prone semi-arid region of east India using GIS and analytical hierarchical process techniques. CATENA 194, 104681. https://doi.org/10.1016/j.catena.2020.104681.
Naghibi, S.A., Ahmadi, K., Daneshi, A., 2017. Application of support vector machine, random forest, and genetic algorithm optimized random forest models in groundwater potential mapping. Water Resour. Manag. 31 (9), 2761–2775. https://doi.org/10.1007/s11269-017-1660-3.
Nguyen, P.T., Ha, D.H., Avand, M., Jaafari, A., Nguyen, H.D., Al-Ansari, N., Van Phong, T., Sharma, R., Kumar, R., Le, H.Van, Ho, L.S., Prakash, I., Pham, B.T., 2020. Soft computing ensemble models based on logistic regression for groundwater potential mapping. Appl. Sci. 10 (7), 2469. https://doi.org/10.3390/app10072469.
Panahi, M., Sadhasivam, N., Pourghasemi, H.R., Rezaie, F., Lee, S., 2020a. Spatial prediction of groundwater potential mapping based on convolutional neural network (CNN) and support vector regression (SVR). J. Hydrol. 588, 125033. https://doi.org/10.1016/j.jhydrol.2020.125033.
Pham, Q.B., Tran, D.A., Ha, N.T., Islam, A.R.M.T., Salam, R., 2022. Random forest and nature-inspired algorithms for mapping groundwater nitrate concentration in a coastal multi-layer aquifer system. J. Clean. Prod. 343, 130900. https://doi.org/10.1016/j.jclepro.2022.130900.
Phong, T.Van, Pham, B.T., 2023. Performance of Naïve Bayes tree with ensemble learner techniques for groundwater potential mapping. Phys. Chem. Earth Parts A/B/C. 132, 103503. https://doi.org/10.1016/j.pce.2023.103503.
Pradhan, A., Adams, K.H., Chandrasekaran, V., Liu, Z., Reager, J.T., Stuart, A.M., Turmon, M.J., 2024. Modeling groundwater levels in California's central valley by hierarchical Gaussian process and neural network regression. J. Geophys. Res. Mach. Learn. Comput. 1 (4). https://doi.org/10.1029/2024JH000322.
Prapanchan, V.N., Subramani, T., Karunanidhi, D., 2024. GIS and fuzzy analytical hierarchy process to delineate groundwater potential zones in Southern parts of India. Groundw. Sustain. Dev. 25, 101110. https://doi.org/10.1016/j.gsd.2024.101110.
Ragragui, H., Aouragh, M.H., El-Hmaidi, A., Ouali, L., Saouita, J., Iallamen, Z., Ousmana, H., Jaddi, H., El Ouali, A., 2024. Mapping and modeling groundwater potential using machine learning, deep learning and ensemble learning models in the saiss basin (Fez-Meknes region, Morocco). Groundw. Sustain. Dev. 26, 101281. https://doi.org/10.1016/j.gsd.2024.101281.
Raj, S., Rawat, K.S., Kumar, S., Almuflih, A.S., Almakayeel, N., Qureshi, M.R.N., 2024. Mapping and validation of groundwater potential zone from alluvial plain: a case study of muzaffarpur, bihar, India. Geol. Ecol. Landsc. 1–16. https://doi.org/10.1080/24749508.2024.2429841.
Rajput, J., Kushwaha, N.L., Srivastava, A., Vishwakarma, D.K., Mishra, A.K., Sahoo, P.K., Elbeltagi, A., 2025. Developing Machine Learning Models for Predicting Daily Relative Humidity and Solar Radiation Using Lagged Time Series Data Inputs in A Semi-Arid Climate. J. Atmos. Sol.-Terr. Phys. 106619. https://doi.org/10.1016/j.jastp.2025.106619.
Rao, M.S., Rao, G.R., Krishna, G.M., Ratnam, K.N., 2025. Revolutionizing groundwater suitability with AI-Driven spatial decision Support-A remote sensing and GIS approach for visakhapatnam district, andhra pradesh, India. J. Geogr. Inf. Syst. 17 (1), 23–44. https://doi.org/10.4236/jgis.2025.171002.
Ren, Q., Li, L., Wang, J., Jiang, R., Li, M., Feng, J., 2024. Dynamic evolution mechanism of the fracturing fracture system—enlightenments from hydraulic fracturing physical experiments and finite element numerical simulation. Pet. Sci. 21 (6), 3839–3866. https://doi.org/10.1016/j.petsci.2024.09.004.
Ritushree, B., Panda, S., Sahoo, A., Samantaray, S., Satapathy, D.P., 2025. Prediction of groundwater level and potential zone identification in keonjhar, odisha based on machine learning and GIS techniques. Frankl. Open 11, 100250. https://doi.org/10.1016/j.fraope.2025.100250.
Robles-Berumen, H., Zafra, A., Ventura, S., 2024. A survey of genetic algorithms for clustering: taxonomy and empirical analysis. Swarm Evolut. Comput. 91, 101720. https://doi.org/10.1016/j.swevo.2024.101720.
Rojas-Thomas, J.C., Santos, M., 2021. New internal clustering validation measure for contiguous arbitrary-shape clusters. Int. J. Intell. Syst. 36 (10), 5506–5529. https://doi.org/10.1002/int.22521.
Rubiños, M., Díaz-Longueira, A., Timiraos, M., Michelena, Á., García-Ordás, M.T., Alaiz-Moretón, H., 2024. A comparative analysis of algorithms and metrics to perform clustering (pp. 63–72). https://doi.org/10.1007/978-3-031-73910-1_7.
Sachdeva, S., Kumar, B., 2021. Comparison of gradient boosted decision trees and random forest for groundwater potential mapping in dholpur (Rajasthan), India. Stoch. Environ. Res. Risk Assess. 35 (2), 287–306. https://doi.org/10.1007/s00477-020-01891-0.
Saha, R., Baranval, N.K., Das, I.C., Kumaranchat, V.K., Reddy, K.S., 2022a. Application of machine learning and geospatial techniques for groundwater potential mapping. J. Indian Soc. Remote Sens. 50 (10), 1995–2010. https://doi.org/10.1007/s12524-022-01582-z.
Saha, D., Ray, R.K., 2019. Groundwater resources of India: potential, challenges and management. Groundwater Development and Management. Springer International Publishing, pp. 19–42. https://doi.org/10.1007/978-3-319-75115-3_2.
Salem, A., Abduljaleel, Y., Dezso, J., Dénes, L., 2023. Integrated assessment of the impact of land use changes on groundwater recharge and groundwater level in the Drava floodplain. Hungary. Sci. Rep. 13. https://doi.org/10.1038/s41598-022-21259-4.
Samantaray, S., Sahoo, A., 2024. Groundwater level prediction using an improved ELM model integrated with hybrid particle swarm optimisation and grey wolf optimisation. Groundw. Sustain. Dev. 26, 101178. https://doi.org/10.1016/j.gsd.2024.101178.

Samantaray, S., Sahoo, A., Yaseen, Z.M., Al-Suwaiyan, M.S., 2025. River discharge prediction based multivariate climatological variables using hybridized long short-term memory with nature inspired algorithm. J. Hydrol. 649, 132453. https://doi.org/10.1016/j.jhydrol.2024.132453.

Singh, D., Vardhan, M., Sahu, R., Chatterjee, D., Chauhan, P., Liu, S., 2023. Machine-learning- and deep-learning-based streamflow prediction in a hilly catchment for future scenarios using CMIP6 GCM data. Hydrol. Earth Syst. Sci. 27 (5), 1047–1075. https://doi.org/10.5194/hess-27-1047-2023.

Sohail, A., 2023. Genetic algorithms in the fields of artificial intelligence and data sciences. Ann. Data Sci. 10 (4), 1007–1018. https://doi.org/10.1007/s40745-021-00354-9.

Srivastava, A., Jain, S., 2025. Remote Sensing for Environment Assessment: Multispectral, Hyperspectral, and Thermal Imaging Applications. In: Saritha, V., Pande, C. B., Singh, R., Shahid, M. (Eds.), Remote Sensing for Environmental Monitoring. Springer, Singapore. https://doi.org/10.1007/978-981-96-5546-5_1.

Srivastava, A., Khadke, L., Chinnasamy, P., 2021. Web Application Tool for Assessing Groundwater Sustainability—A Case Study in Rural-Maharashtra, India. In: Vaseashta, A., Maftei, C. (Eds.), Water Safety, Security and Sustainability. Advanced Sciences and Technologies for Security Applications. Springer, Cham. https://doi.org/10.1007/978-3-030-76008-3_28.

Srivastava, A., Maity, R., Desai, V.R., 2024. Investigating Spatio-Temporal Trends and Anomalies in Long-Term Meteorological Variables to Determine If Maharashtra is an Emerging Warming State in India. In: Sreekeshava, K.S., Kolathayar, S., Vinod Chandra Menon, N. (Eds.), Civil Engineering for Multi-Hazard Risk Reduction. IACESD 2023. Lecture Notes in Civil Engineering, 457. Springer, Singapore. https://doi.org/10.1007/978-981-99-9610-0_25.

2023 SwitchON Foundation. (2023). *Groundwater depletion in Jharkhand: Status, challenges, and mitigation strategies*. Retrieved from https://www.switchon.org.in/wp-content/uploads/2024/03/Groundwater-Depletion-on-rise-in-Jharkhand-.pdf.

Taccari, M.L., Wang, H., Nuttall, J., Chen, X., Jimack, P.K., 2024. Spatial-temporal graph neural networks for groundwater data. Sci. Rep. 14 (1), 24564. https://doi.org/10.1038/s41598-024-75385-2.

Tao, H., Abba, S.I., Al-Areeq, A.M., Tangang, F., Samantaray, S., Sahoo, A., Siqueira, H.V., Maroufpoor, S., Demir, V., Bokde, N.D., Goliatt, L., 2024. Hybridized artificial intelligence models with nature-inspired algorithms for river flow modeling: a comprehensive review, assessment, and possible future research directions. Eng. Appl. Artif. Intell. 129, 107559. https://doi.org/10.1016/j.engappai.2023.107559.

Uc Castillo, J.L., Martínez Cruz, D.A., Ramos Leal, J.A., Tuxpan Vargas, J., Rodríguez Tapia, S.A., Marín Celestino, A.E., 2022. Delineation of groundwater potential zones (GWPZs) in a Semi-Arid basin through remote sensing, GIS, and AHP approaches. Water 14 (13), 2138. https://doi.org/10.3390/w14132138.

UNESCO, 2023. Imminent risk of a global water crisis, warns the UN World Water Development Report 2023 | Retrieved June 6, 2025, from ⟨https://www.unesco.org/en/articles/imminent-risk-global-water-crisis-warns-un-world-water-development-report-2023⟩.

United Nations (UN). 2022. The United Nations world water development report 2022: groundwater: making the invisible visible.

Vafadar, S., Rahimzadegan, M., Asadi, R., 2023. Evaluating the performance of machine learning methods and geographic information system (GIS) in identifying groundwater potential zones in Tehran-Karaj plain, Iran. J. Hydrol. 624, 129952. https://doi.org/10.1016/j.jhydrol.2023.129952.

Wang, X., Liu, T., Zheng, X., Peng, H., Xin, J., Zhang, B., 2018. Short-term prediction of groundwater level using improved random forest regression with a combination of random features. Appl. Water Sci. 8 (5), 125. https://doi.org/10.1007/s13201-018-0742-6.

Yariyan, P., Avand, M., Omidvar, E., Pham, Q.B., Linh, N.T.T., Tiefenbacher, J.P., 2022a. Optimization of statistical and machine learning hybrid models for groundwater potential mapping. Geocarto Int. 37 (13), 3877–3911. https://doi.org/10.1080/10106049.2020.1870164.

Yen, H.P.H., Pham, B.T., Phong, T.Van, Ha, D.H., Costache, R., Le, H.Van, Nguyen, H.D., Amiri, M., Tao, N., Van, Prakash, I., 2021. Locally weighted learning based hybrid intelligence models for groundwater potential mapping and modeling: a case study at gia lai province, Vietnam. Geosci. Front. 12 (5), 101154. https://doi.org/10.1016/j.gsf.2021.101154.

Zaib, S., Najib, M.S., Mohd Daud, S., Zahari, M.F., Mohamad, M., 2022. The study of groundwater source by using KNN classification (pp. 563–574). https://doi.org/10.1007/978-981-16-8690-0_51.

Zhao, Y., Ma, J., Li, Y., Cheng, K., Zhang, M., Liu, Z., Yang, F., 2025. Carbon emission based predictions of anthropogenic impacts on groundwater storage at typical basins in 2050. Research 8, 0680. https://doi.org/10.34133/research.0680.

Zhao, Y., Yang, L., Pan, H., Li, Y., Shao, Y., Li, J., Xie, X., 2025. Spatio-temporal prediction of groundwater vulnerability based on CNN-LSTM model with self-attention mechanism: a case study in hetao plain, Northern China. J. Environ. Sci. 153, 128–142. https://doi.org/10.1016/j.jes.2024.03.052.

Zou, L., Su, Y., Xu, X., Li, W., Zhao, H., Zhang, M., Jiao, S., 2025. Study on the propagation law of hydraulic fractures in heterogeneous permeability reservoirs. PLOS One 20 (7), e328689. https://doi.org/10.1371/journal.pone.0328689.