scientific reports

Check for updates

OPEN

Forecasting land-use and land-cover change for groundwater sustainability in the Muvattupuzha basin using CA-Markov (2033–2050)

Alagulakshmi K[1], Sneha Gautam[1,2,3]✉, G. Prince Arulraj[1], Suneel Kumar Joshi[4] & Chang-Hoi Ho[3]✉

Rapid urbanization and land use and land cover (LULC) change have affected groundwater dynamics and its quality in many river basins. The present study uses an integrated framework combining multi-temporal Landsat imagery, geospatial analysis, multivariate statistics, and Machine Learning (ML) approaches to understand LULC changes and groundwater dynamics and its quality degradation. The supervised classification was used in the present study, which shows that built-up land increased significantly from 12.3% (329.13 km^2) in 2003 to 44.4% (1,187.11 km^2) in 2023, mainly due to the conversion of agricultural and forested land. Furthermore, future LULC dynamics by the CA-Markov model indicate continuous landscape transformation, with net conversions into built-up and forested areas during the periods 2023–2033 and 2033–2043, respectively, while there is a decline in water bodies and agricultural land use, and their rates of change stabilize over the periods approaching 2043–2050. Multivariate statistical analyses, such as correlation analysis, Principal Component Analysis (PCA), and Cluster Analysis, identify both geogenic processes and human activities as dominant determinants of groundwater hydrochemistry. To investigate the relationships between physicochemical parameters and nitrate variability, 3 ML models were employed: Random Forest (RF), Support Vector Regression (SVR), and XGBoost. Model interpretation using SHapley Additive exPlanations (SHAP) showed that Mg^{2+}, Ca^{2+}, and alkalinity are the significant factors influencing nitrate distribution, reflecting buffering reactions and redox-controlled processes. An integrated framework combining LULC, hydrogeochemical, and ML techniques provides a strong foundation for assessing groundwater. It offers insights into sustainable land-use planning and groundwater management in rapidly urbanizing tropical basins.

Keywords Muvattupuzha basin, Land use change, CA–Markov modelling, Groundwater quality, Machine learning, SHAP model, Nitrate contamination

India, the world's second most populous country, exhibits remarkable diversity in its topography, climate, environment, ecology, hydrology, geology, land use, and socioeconomic conditions[1]. This diversity strongly influences the regional and local development pattern, resource availability, and environmental sustainability across the nation[2]. Such characteristics make India a representative case for examining broader global challenges related to land and food systems under rapid demographic and environmental change. At the global scale, Food security is a crucial requirement for ensuring sustainable agricultural production to support the continuously growing global population[3,4]. In the present context, achieving higher agricultural productivity while improving land management and conservation practices remains a significant challenge. Considering this, Land Use and Land Cover (LULC) change has emerged as an important global issue, primarily driven by human activities that alter land surfaces and disrupt environmental ecosystems at regional, local, and international scales[5]. Land use refers to the way humans modify land for habitation and economic purposes.

[1]Division of Civil Engineering, Karunya Institute of Technology and Sciences Deemed University, Karunya Nagar, Coimbatore, Tamil Nadu 641114, India. [2]Water Institute, A Centre of Excellence Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu 641 114, India. [3]Department of Climate and Energy Systems Engineering, Ewha Womans University, Seoul, Republic of Korea. [4]Geo Climate Risk Solutions Pvt. Ltd, Visakhapatnam, Andhra Pradesh 530048, India. ✉email: snehagautam@karunya.edu; hoch@ewha.ac.kr

In contrast, land cover refers to the physical characteristics of the Earth's surface, encompassing vegetation, soil, water, geology, and geomorphological features[6,7]. In recent years, LULC changes have become increasingly valuable for formulating effective policies addressing environmental and natural resource challenges. Furthermore, detailed LULC mapping plays a crucial role in implementing and monitoring[8]. In this case, this piece emphasizes the need to regularize land management practices amid urbanization and changes in LULC. The application of Remote Sensing (RS) & Geographical Information System (GIS) in monitoring LULC change using traditional machine-learning methods for analyzing satellite images is considered. While Artificial Neural Networks (ANNs) are also used to forecast the impacts of LULC change, the Cellular Automata (CA) method is reliable. Integration of demographic, economic, and environmental information into future modelling efforts will be achieved using machine-learning classifiers such as Random Forest (RF) and support vector machines, so that the more conventional Maximum Likelihood Classification (MLC) can be outperformed. Predictive models like CA-Markov are also used to model spatial patterns relevant to land management and conservation, despite issues with the non-stationarity of transition probabilities. LULC changes significantly influence the hydrological balance and groundwater sustainability in humid tropical basins, where high rainfall, intense agricultural activity, deforestation, and rapid urban expansion coexist[9]. These basins, such as those in Kerala, are highly sensitive to anthropogenic modifications, as variations in land cover alter infiltration, recharge, and contamination pathways in groundwater systems[10]. Moreover[11,12], studied significant LULC changes in the Western Ghats, including forest degradation and conversion to agricultural or plantation uses. Recent studies highlight the growing use of ML models for predicting groundwater quality, demonstrating their ability to capture complex nonlinear interrelationships among hydrochemical parameters. ML models like RF and XGBoost have shown better predictive performance for contaminants in groundwater, including nitrate levels and comprehensive water quality indices, than conventional methods[13,14]. RF, XGBoost, and neural networks were used to predict groundwater quality, and ML methods were reported to achieve high accuracy across a wide array of physicochemical parameters. Modern efforts also integrated explainable AI approaches, notably SHAP[15,16], to interpret ML predictions and quantify the contribution of important predictors to groundwater quality outcomes, pinpointing TDS and hydrogeochemical indicators as the leading variables. These interpretable machine learning frameworks provide transparency and allow easy integration with hydrogeochemical process knowledge, thereby eliminating the so-called "black-box" issue in traditional machine learning applications in hydrogeology[17]. Built-up expansion reduces percolation and increases surface runoff, while the decline of vegetation and wetlands diminishes the basin's natural recharge capacity[18]. Understanding these spatial-temporal transitions is thus crucial for maintaining groundwater quality and ensuring its availability under changing land-use dynamics. Despite extensive research on LULC mapping and hydrological modelling, a notable gap remains in linking CA-Markov-based LULC projections with groundwater risk assessment[19]. Analyzing global and basin-scale studies conducted in Kerala provides the necessary regional context to interpret LULC change dynamics and their environmental consequences. Studies of the major river basins-Notably the Periyar[20], Bharathapuzha[21], and Chaliyar[22] and other basins such as Thoothupuzha[23], Achankovil, Kallada, Meenachil, Valapattanam, Kuppam, Neyyar, and Vamanapuram regularly report rapid urban growth, plantation dominance, and the accompanying loss of agricultural land and wetlands, along with increased hydrological stress and localized deterioration of groundwater quality[27,31]. Most existing studies emphasize either land transformation or groundwater trends separately, rarely integrating them to predict future groundwater vulnerability under different LULC scenarios. This disconnect limits our ability to anticipate how future land transitions, especially urbanization and agricultural intensification, may affect groundwater quality and sustainability. Despite this corpus, most studies accentuate either shallow LULC or hydrological effects with inadequate integration of groundwater quality data, especially at sub-basin scales. The Muvattupuzha basin was therefore selected based on its hydrological sensitivity, rapid land-use change, increasing dependence on groundwater, and the availability of consistent multi-temporal datasets, which together enable an integrated LULC and groundwater assessment using spatial and ML-based methodologies[32,33]. The present study addresses this gap through 3 hypotheses: H1: LULC transitions from 2003 to 2023 have significantly modified groundwater recharge and contamination potential; H2: CA-Markov-based LULC forecasts for 2033–2050 can reliably predict zones of groundwater stress when combined with quality indicators; and H3: explainable modelling frameworks can reduce uncertainty in linking LULC change trajectories with groundwater quality dynamics. By integrating multi-temporal satellite data, CA-Markov simulations, and groundwater quality indices, this research produces spatially explicit forecasts that are both scientifically rigorous and interpretable. The contribution lies in advancing explainable forecasting of LULC and groundwater interactions, incorporating uncertainty-aware validation metrics, and establishing a unified framework that quantitatively links future land transitions to groundwater quality degradation or improvement. This integrative approach supports data-driven decision-making for sustainable groundwater management in rapidly transforming humid tropical landscapes.

Study area description

The Muvattupuzha River Basin is situated in the central part of Kerala, India, spanning latitudes 9°30′45″ to 10°07′46″ N and longitudes 76°14′14″ to 76°56′24″ E (Fig. 1). Given the basin's total catchment area of approximately 2,670 km², the present groundwater quality investigation was confined to the effective drainage area of about 1,550 km², as explicitly stated in the Study Area description. It spans portions of the Ernakulam, Kottayam, and Idukki districts, covering an approximate drainage area of 1,554 km². The river originates from the confluence of the Thodupuzha, Kaliyar, and Kothamangalam tributaries, which drain the high-altitude regions of the Western Ghats and flow westward before merging with the Vembanad Lake near Vaikom. The basin shows a distinct east-to-west elevation gradient, with the highlands rising above 1,800 m and the lowlands reaching nearly sea level. The topography can be grouped into 3 broad zones: highlands characterized by steep slopes, dense forests, and lateritic soils; midlands with gently undulating terrain supporting plantations and

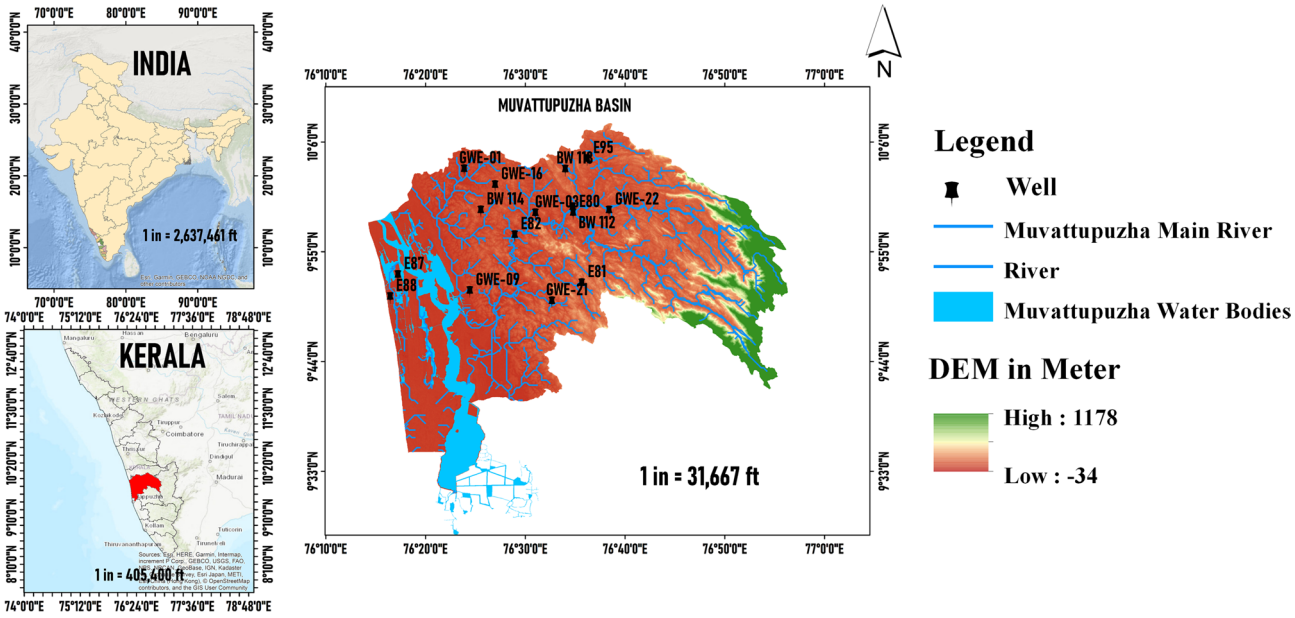


Fig. 1. Spatial extent of the Muvattupuzha River Basin showing major tributaries, elevation gradients, and sub-watershed boundaries used for hydrological analysis. The study area map analysis was generated using ArcGIS Pro software.

settlements; and lowlands comprising alluvial plains dominated by paddy fields, backwaters, and urban centers. This topographic diversity governs the drainage pattern, infiltration behavior, and groundwater recharge potential across the basin. Hydrologically, the Muvattupuzha River is perennial, receiving flows from several tributaries. The basin is strongly influenced by monsoonal rainfall, with nearly 80% of the annual runoff occurring during the Southwest (SW) monsoon from June to September. The total annual average flow is approximately 4,000 million m^3[31]. Reservoirs such as Malankara and Bhoothathankettu significantly alter natural discharge patterns, whereas smaller water-retaining structures, including check dams and ponds, play a crucial role in enhancing local groundwater recharge and sustaining baseflow continuity during dry periods.

The basin has a humid tropical monsoon climate, with a mean annual rainfall of 2,800–3,000 mm. Temperatures range between 22 °C and 32 °C, and the Southwest monsoon brings the primary rain, while the Northeast (NE) monsoon provides supplementary precipitation between October and November. The basin's socio-ecological environment is influenced by the >1.2 million inhabitants, whose livelihoods are primarily derived from agriculture and plantation crops, such as rubber, coconuts, and spices, as well as from new urban employment opportunities. Increasing urbanization, industrialization, and land clearing have heightened environmental pressure, resulting in the depletion and contamination of groundwater and the destruction of natural habitats. Social and ecological interactions thus determine key long-term sustainability among the basin's freshwater resources. Groundwater is primarily found in phreatic and semi-confined settings, stored within weathered and fractured crystalline rocks, as well as lateritic and alluvial deposits. Coastal alluvial aquifers tend to yield high volumes of freshwater but remain highly susceptible to pollution and saltwater invasion. Lateritic aquifers yield moderately and also have high replenishment potential. Crystalline fractured aquifers, commonly described as chalk formations, tend to lack the capacity to retain stored material and yield freshwater only at localized fracture zones. Water levels range from 2 to 15 m below ground level (bgl). It is controlled by topography and sub-surface lithology, the distribution and character of the countryside, and the seasonally uneven distribution of rainfall.

Approach and methodology

Data and pre-processing

The present study integrated multi-temporal Landsat imagery (2003, 2013, 2023) with ancillary geospatial and field datasets to evaluate LULC changes and their impacts on groundwater in the Muvattupuzha River Basin. Landsat 7 ETM + and Landsat 8/9 OLI/TIRS data (30 m spatial resolution) were used and classified using the Maximum Likelihood Algorithm (MLA) into five classes: built-up, agriculture, forest, barren, and water bodies. Accuracy assessments achieved a reliability of over 85%. Complementary datasets, including SRTM-derived DEMs, soil characteristics, and proximity to rivers and roads, were standardized to a common projection (UTM 43 N) to support CA-Markov modelling. Groundwater-quality data (EC, TDS, NO_3^-, SO_4^{2-}, Ca^{2+}, Mg^{2+}, Na^+) from the Kerala State Groundwater Department were used to compute the Groundwater Quality Index (GWQI) via the Inverse Distance Weighting (IDW) interpolation approach. Together, these datasets established a robust spatio-temporal foundation for assessing surface-subsurface interactions and groundwater sustainability in the river basin.

Methodology

Figure 2 illustrates the methodological framework for the current study, comprising 5 key components: obtaining Landsat imagery from the USGS, pre-processing, generating training data for LULC classification, identifying LULC classes and change detection, and forecasting future LULC changes. The GWQ evaluation, based on the EWQI, GWQ mapping, ML-based geographical prediction, indicator analysis using SHAP values, and secondary GWQ data gathered, is included in the study that follows. It is essential to note that the research regions were defined based on LULC predictions and GWQ evaluations. The detailed methodology is mentioned in Annexure I.

Results

Spatio-temporal pattern of LULC

Figure 3; Table 1 depict the changes in LULC for the year 2003–2023 in the Muvattupuzha river basin. The LULC has been classified into 5 classes: Built-up, Water Bodies, Barren Land, Agricultural Land, and Forest. The spatial distribution of the LULC map for 2003 indicates that it is covered by forest land, occupying 1297.06 km^2 (48.49%), followed by barren land at 451.04 km^2 (16.86%), water bodies at 274.37 km^2 (10.26%), agricultural land at 323.36 km^2 (12.09%), and built-up area at 329.13 km^2 (12.30%). However, built-up area had expanded by about 541.91 km^2 (20.26%), indicating urban growth during 2013. Water bodies had a minor increase to 305.19 km^2 (11.41%), while barren land nearly doubled to 911.47 km^2 (34.07%) during 2013. Agricultural land also increased to 678.90 km^2 (25.38%), indicating increased cultivation activities during 2013. However, forest cover has drastically reduced to 237.49 km^2 (8.88%) in 2013 compared to 2003, indicating significant deforestation within the basin. Further, the built-up area reached 1187.11 km^2 (44.41%) during 2023 (Kappa coefficient of 0.85), indicating considerable urban growth compared to previous decades. Water bodies showed a minor decrease to 272.04 km^2 (10.18%). Barren land declined sharply to 216.41 km^2 (8.10%), while agricultural land fell to 180.65 km^2 (6.76%), indicating a conversion of cropland to other uses. The forest area has partially recovered to 816.98 km^2 (30.56%), suggesting restoration or regeneration of vegetative cover over the last decade of the total region, thus indicating notable vegetation recovery or afforestation efforts by 2023.

Furthermore, the transition analysis has been carried out to quantify spatial and temporal changes in LULC. The change detection matrices were tabulated to determine the persistence, gain, and loss between the various land[32]. The step helped quantify the magnitude and direction of transitions, for example, from agricultural land to built-up areas or from forest to agricultural land[33]. Transition intensity analysis was also conducted to determine the prevailing conversions and swap changes, providing clues to the rate of urbanization and landscape alteration. Spatial change patterns were also represented through[34]. The analysis highlighted that built-up expansion mainly occurred around the central region [35]. Figures 3 and 4 display the LULC maps for the 3 years under study, along

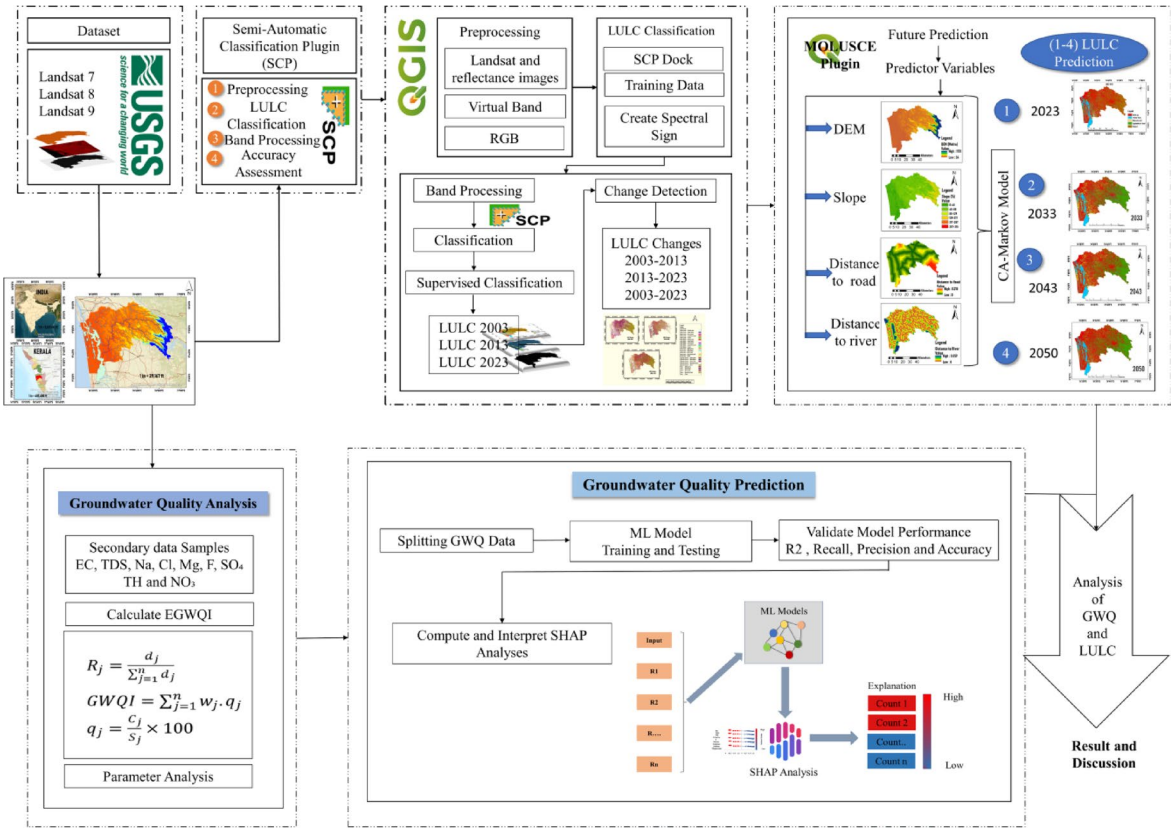


Fig. 2. Methodological flowchart of the current study.

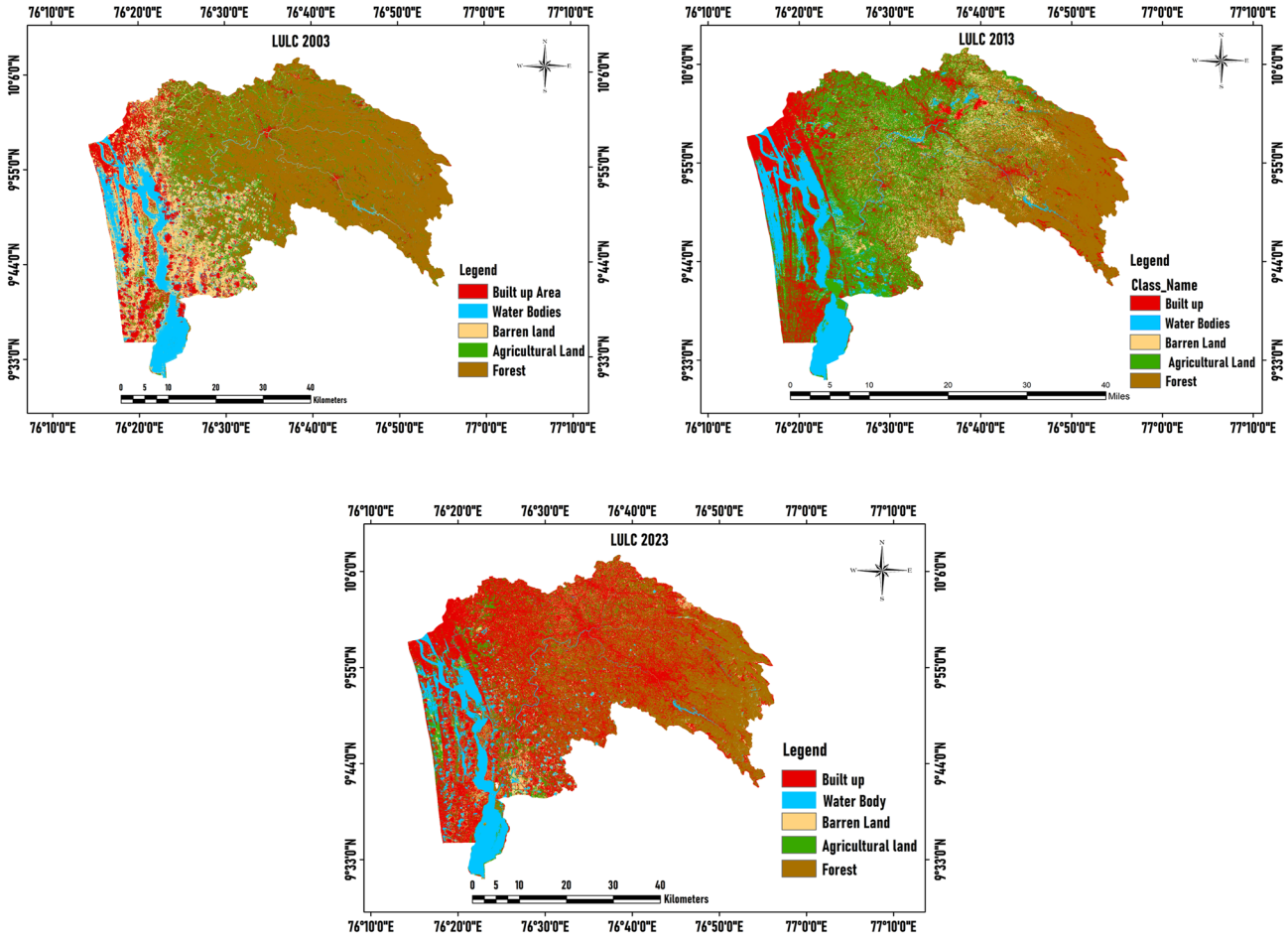


Fig. 3. Spatio-temporal distribution of LULC change for 2003, 2013, and 2023. Spatio-temporal three-year LULC mapping used the QGIS 3.42.1 version software.

Year	2003			2013			2023		
LULC categories	Area (km^2)	Area (%)	Kappa	Area (km^2)	Area (%)	Kappa	Area (km^2)	Area (%)	Kappa
Built-up area	329.13	12.30		541.91	20.26		1187.11	44.41	
Water bodies	274.37	10.26		305.19	11.41		272.04	10.18	
Barren land	451.04	16.86	0.75	911.47	34.07	0.8	216.41	8.10	0.85
Agricultural land	323.36	12.09		678.9	25.38		180.65	6.76	
Forest	1297.06	48.49		237.49	8.88		816.98	30.56	

Table 1. Accuracy metrics.

with the spatial hotspots of change. The transition analysis accuracy metrics, tabulated in Table 2, support the plausibility of the changes detected relative to observed land development. The present transition analysis served as the empirical foundation for formulating future scenarios through CA-Markov modeling.

LULC change 2003–2023

Figure 4 illustrates that from 2003 to 2013, the LULC underwent significant changes, primarily due to urban growth and agricultural intensification. The built-up area increased from 329.13 to 541.91 km^2, representing a rise of 212.78 km^2 (≈ 65.1%). This suggests rapid urbanization, possibly driven by infrastructure development and population growth. Water Bodies demonstrated moderate growth of 30.82 km^2 (≈ 1.15%), indicating enhanced water storage or the creation of artificial reservoirs.

Barren Land increased significantly from 451.04 to 911.47 km^2, an increase of 460.43 km^2 (≈ 17.21%), presumably due to deforestation and the encroachment of vegetated land into non-productive areas. Agricultural Land almost doubled, rising by 355.54 km^2 (≈ 13.29%), indicating agricultural encroachment onto forests. Forest Cover declined sharply from 1297.06 to 237.49 km^2, representing a loss of 1059.57 km^2 (≈ – 39.61%), indicating a significant loss of forest cover during this decade. Overall, from 2003 to 2013, the region was dominated by

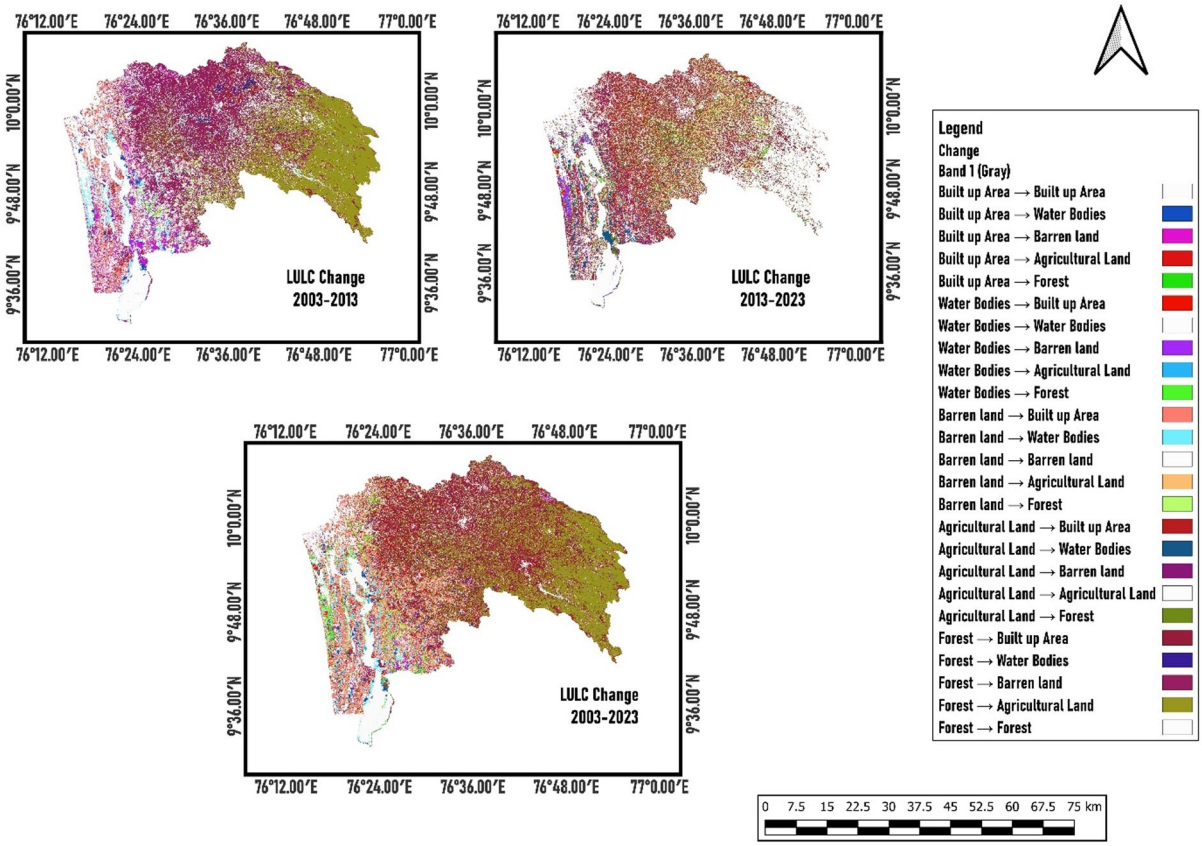


Fig. 4. Hotspot change map (2003–2023). LULC change detection mapping was prepared using QGIS 3.42.1 version software.

	LULC	Built-up area (%)	Water bodies (%)	Barren land (%)	Agricultural land (%)	Forest (%)
2003–2013						
2003	Built-up area	55.85	5.98	32.46	3.25	2.46
	Water bodies	6.05	73.56	15.83	2.55	2.01
	Barren land	28.21	12.21	49.77	3.70	6.10
	Agricultural land	28.06	1.60	50.62	11.75	7.97
	Forest	9.52	1.80	28.76	46.76	13.15
2013–2023						
2013	Built-up area	77.60	4.09	4.40	6.68	7.23
	Water bodies	11.47	64.21	11.40	5.34	7.58
	Barren land	51.10	4.84	13.23	9.70	21.14
	Agricultural land	25.93	0.88	2.28	3.33	67.57
	Forest	38.08	1.84	9.22	7.27	43.59
2003–2023						
2023	Built-up area	67.29	7.69	8.39	8.75	7.88
	Water bodies	15.60	63.81	4.65	7.54	8.40
	Barren land	49.33	11.64	10.08	13.39	15.56
	Agricultural land	60.78	2.46	7.29	18.29	11.18
	Forest	38.91	0.90	5.50	49.97	4.73

Table 2. Transition matrix (2003–2023).

high deforestation and conversion into barren, agricultural, and built-up land. The period from 2013 to 2023 marks a continuation of urban expansion, accompanied by some recuperation of vegetation and a reduction in barren land. The built-up area also increased further, from 541.58 to 1187.11 km^2, adding 645.53 km^2 (≈ 24.15%), which confirms the ongoing urban sprawl. Water bodies decreased by 32.38 km^2 (≈ − 1.21%), possibly due to

surface water loss from reclamation or sedimentation. Agricultural Land decreased from 910.85 to 180.65 km², representing a loss of 730.20 km² (≈ − 27.32%), possibly due to conversion to built-up areas. The Forest Area experienced a notable recovery, increasing from 678.89 to 816.98 km², a gain of 138.09 km² (≈ 5.17%), possibly due to agricultural activities. Barren Land decreased by 21.04 km² (≈ − 0.79%), indicating a decrease in exposed soil as land was brought under other uses. Therefore, from 2013 to 2023, urbanization increased, agriculture declined, and forests regrew. Over the 20 years from 2003 to 2023, the basin underwent significant landscape changes.

The built-up area increased from 328.38 to 1187.11 km², adding 858.73 km² (≈ + 32.12%), the most significant increase of any class. Water Bodies barely fluctuated with a minor decrease of 1.83 km² (≈ − 0.07%), indicating minimal hydrological change. Barren Land decreased by 269.96 km2 (≈ − 10.10%), suggesting land reclamation or re-vegetation. Agricultural land increased by 493.68 km² (≈ approximately + 18.47%), suggesting a long-term expansion in the cultivable area, despite a reduction in the recent decade. Forest Cover lost the most, decreasing significantly from 1297.02 to 216.41 km², by 1080.61 km² (≈ − 83.97%), highlighting ongoing deforestation pressure over the study period. Between 2003 and 2023, the Muvattupuzha Basin underwent a significant transformation, evolving from a forested landscape to a built-up, agricultural-dominated landscape, transitioning from a natural to a human-altered environment.

Transition dynamics

Table 2 summarizes the LULC transition matrices derived from maps classified for the first period (2003–2013), the second period (2013–2023), and the entire period (2003–2023). The diagonal values in each matrix indicate the persistence of every LULC class, while the off-diagonal values represent the extent of the transition between two classes.

The total LULC change matrix for the decade 2003–2023 indicates that nearly 66–68% of the landscape remained stable, while around 32–34% exhibited visible changes between land use categories. Built-up areas showed the greatest persistence, with 67.29% remaining in their original areas, and they also showed the greatest expansion, covering 8.39% of bare land and 8.75% of agricultural land. Water bodies were moderately stable, with 63.81% of their coverage remaining unchanged. However, 15.60% of their coverage changed to built-up land, indicating encroachment and a decrease in surface water coverage. Barren land was the most dynamic, unstable class, exhibiting only 10.08% persistence and large-scale conversions: 49.33% transitioned into built-up land and 13.39% into agricultural land, reflecting intensified reclamation activities and developmental pressure. Agricultural land exhibited 18.29% persistence, but the most significant transition was to built-up land (60.78%), indicating continuous transformation of agricultural land into settlements. Forest cover exhibited the least persistence, with only 4.73% remaining unchanged over time. Meanwhile, 49.97% of the forest cover was converted to agricultural land, and 38.91% to built-up areas.

Table 2 also shows the transition proportions in a row-wise manner. It was calculated with respect to the total forest area in 2003 (1297 km²), rather than the net loss in forest. Therefore, the values 38.91% (transitioned to built-up) and 49.97% (transitioned to agriculture) correspond to gross transitions rather than net change. That is, 38.91% corresponds to 504.9 km² of the forest area in 2003, which was built up, and 49.97% corresponds to 648.3 km², which was converted to agriculture. However, a substantial amount of area also transitions from other classes, such as agriculture, into forest, thereby offsetting the gross loss. Therefore, the net reduction in forest cover between 2003 and 2023 is 480 km² (from 1297 to 817 km²). On a global scale, the findings demonstrate that urban agglomeration land has steadily increased at the expense of agricultural land, barren land, and forests in the basin. This trend of expansion points to rapid urbanization and human-induced land change over the past twenty years. The predominance of urban gain and the attendant reduction in natural and semi-natural areas indicate considerable landscape transformation and ecological stress in the Muvattupuzha Basin between 2003 and 2023.

Actual and predicted LULC

Validation of the CA-Markov-predicted LULC map for 2023 was performed by comparing the simulated land-use distribution with the classified actual LULC. Category-specific area statistics and their respective percentage shares are summarized in Table 3. Figure 5 illustrates the close agreement between the observed and predicted land-use patterns across the basin. The dominant land-use class is the built-up area, which occupies 1,187.11 km² (44.41%) in the actual LULC map and 1,169.96 km² (43.77%) in the predicted map. The built-up class achieved a classification accuracy of 88% (Kappa coefficient = 0.87), indicating strong concordance between the observed and simulated datasets and confirming that the model effectively captures the dynamics of urban expansion.

Year	LULC actual – 2023		LULC prediction-2023		Kappa	
LULC categories	Area (km²)	Area (%)	Area (km²)	Area (%)	Accuracy	Validation
Built-up area	1187.11	44.41	1169.96	43.77	88.71%	0.833
Water bodies	272.04	10.18	309.52	11.58		
Barren land	216.41	8.10	210.81	7.89		
Agricultural land	180.65	6.76	187.37	7.01		
Forest	816.98	30.56	795.53	29.76		

Table 3. Actual and predicted LULC for 2023.

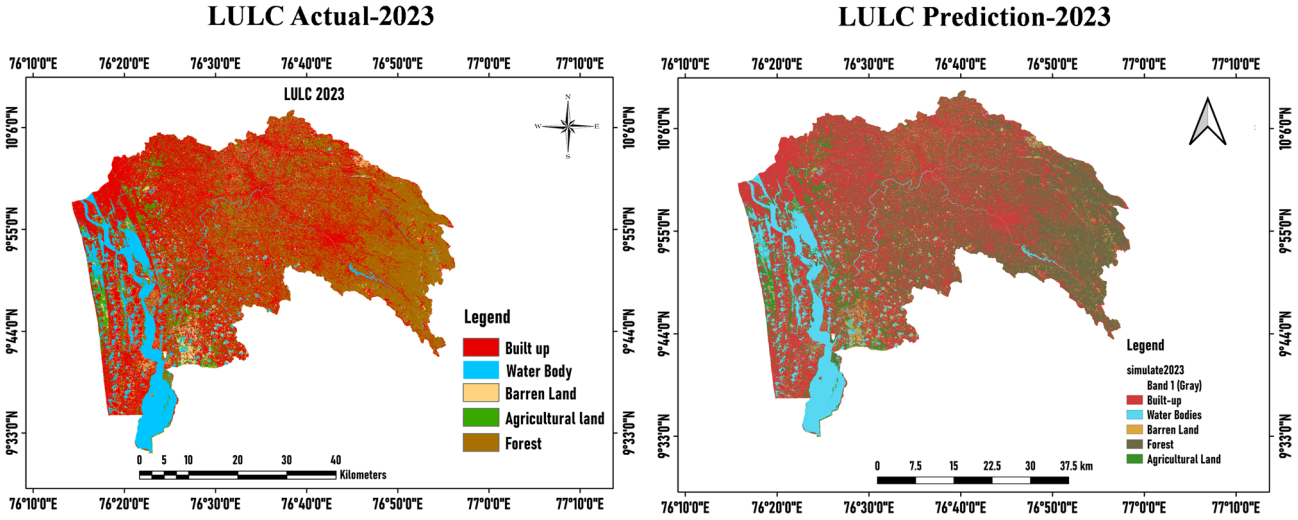


Fig. 5. Analysis of actual and predicted LULC in 2023. Spatio-temporal three years LULC mapping used the QGIS 3.42.1 version software.

Water bodies cover 272.04 km^2 (10.18%) in the actual LULC, while the predicted map shows a greater extent of 309.52 km^2 (11.58%). Barren land occupies 216.41 km^2 (8.10%) on the actual map and 210.81 km2 (7.89%) on the predicted map, indicating a marginal deviation between the observed and simulated areas. Agricultural land accounts for 180.65 km^2 (6.76%) in the actual LULC and 187.37 km^2 (7.01%) in the predicted LULC, showing small variations in the modeled agricultural extent. Forest land is 816.98 km^2 (30.56%) in the actual map and 795.53 km^2 (29.76%) in the predicted map, indicating a slight underestimation of forest cover in the simulation. Overall, the CA-Markov predicted LULC distribution closely matched. The 2023 LULC, with the least deviation across all land-use categories, thereby demonstrating satisfactory model performance for basin-scale LULC simulation. Moreover, Model validation was performed using Kappa statistics derived from a pixel-based confusion matrix comparing observed and simulated LULC maps in ArcGIS Pro. Spatial neighborhood influence was incorporated using a 5 × 5 focal statistics window, which allowed land-use transitions to be influenced by surrounding cells and ensured realistic spatial contiguity in the simulated maps.

Kappa statistics

The confusion matrix Table A6 illustrates high concordance between the observed and the predicted classes of LULC, as reflected in the dominance of entries on the main diagonals. In general, the predicted map showed an overall accuracy of 88.71%, suggesting a very reliable classification. This indicates near-perfect agreement between the two datasets, reinforcing the strength of the prediction model, with a Kappa coefficient value of 0.833. Built-up areas have the highest classification accuracy and the least confusion among other land use classes. This is because urban regions are spatially persistent and distinct spectrally. Water bodies show limited misclassification, mainly with barren land and forest margins, which could be attributed to seasonal variation and mixed pixels along shoreline boundaries.

Bare land is moderately confused with forest and water bodies, probably because of the spectral similarity in dry conditions and transitional characteristics in land cover. Some mutual misclassifications between forest and agricultural classes occur, particularly in fragmented areas and along edges, reflecting heterogeneity in land use and a gradual transition in land cover across the basin. The high overall accuracy and Kappa statistic indicate that the used modeling framework successfully captured the spatial land-use dynamics. The validation results confirmed that the predicted land use/land cover map is suitable for further analyses related to environmental assessment and groundwater sustainability.

Suitability analysis

Suitability mapping based on physical and accessibility factors

The land-use suitability map was developed through a weighted overlay methodology that integrated slope, elevation, soil type, distance to roads, and distance to rivers. All thematic layers were resampled to 30 m to ensure consistency with the DEM and satellite datasets. The resulting suitability index was then categorized into 5 classes: low, moderate, suitable, strong, and highly strong suitability (Table A5). Suitability classes are spatially distributed as shown in Fig. 6, where the strongest classes are found in low-lying areas with gentle slopes and high road network accessibility. Low-suitability zones are associated with steep terrain, higher-order elevations, and proximity to river buffer zones, where environmental and developmental constraints are very high.

Table A5. Area statistics by quantitative suitability classes show that strong suitability accounts for the most significant proportion of the study area, followed by robust and suitable classes. This indicates that a substantial area of the basin has relatively better physical and infrastructural conditions for future controlled land-use expansion. Smaller area proportions are represented by low- and moderately suitable classes, indicating

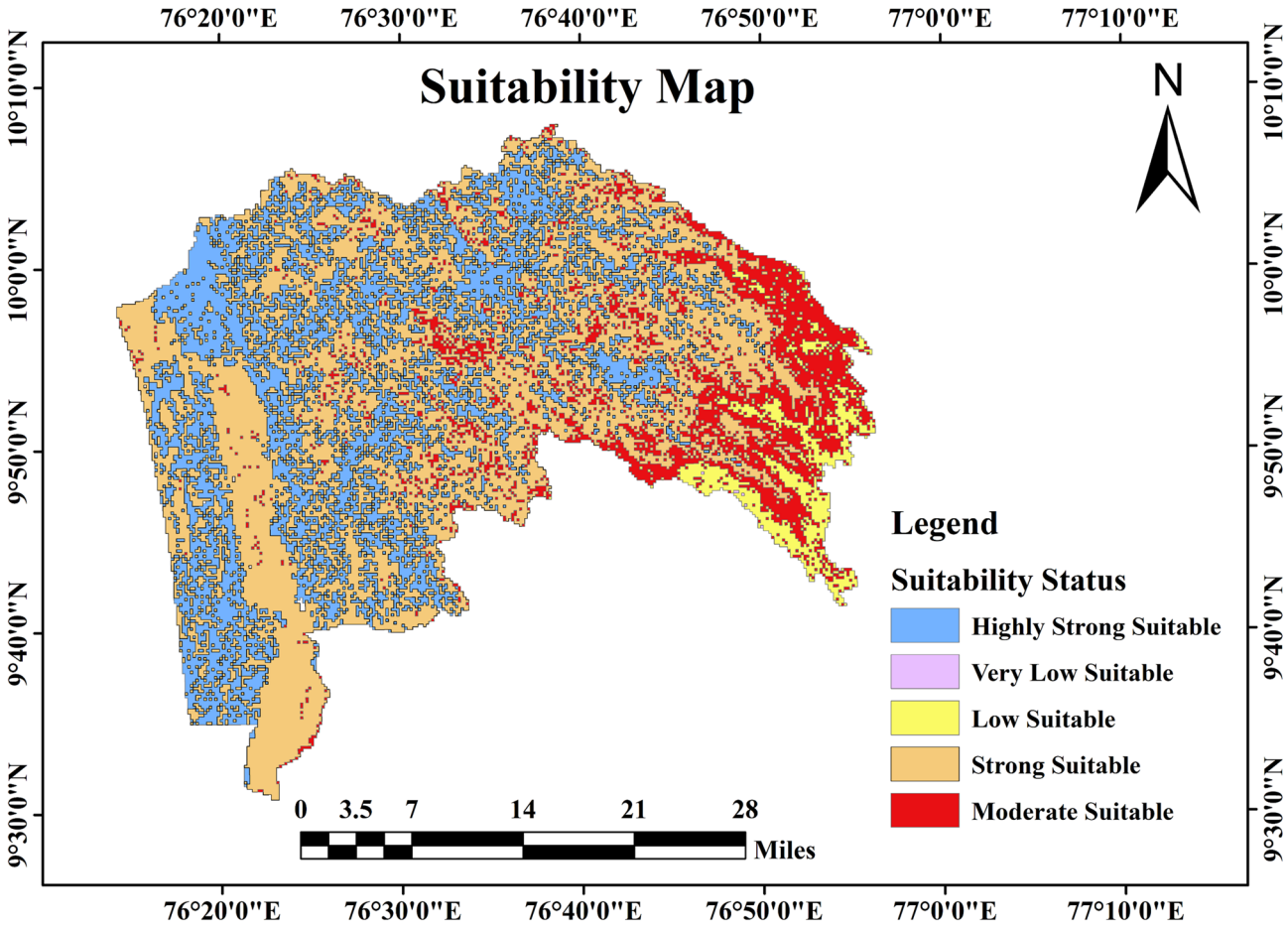


Fig. 6. Suitability analysis map. Suitability mapping was prepared using QGIS 3.42.1 version software.

environmentally sensitive or physically constrained areas. The suitability classification spatially resolves the hierarchy of the potential land use. It presents a critical input for land-use transition modeling using CA-Markov.

Future LULC scenarios

Figure 7; Table 4 show the years 2033, 2043, and 2050, during which the landscape exhibited moderate changes among the primary LULC classes. Between 2023 and 2033, the built-up area grew by 3.31 km^2 (0.12%), marking the beginning of urbanization and infrastructure development.

Conversely, the area of water bodies decreased by 9.88 km^2 (0.37%), possibly due to land reclamation and reduced surface water storage. The barren land category increased by 13.82 km^2 (0.52%), indicating a conversion from agricultural or vegetated land. Agricultural land, on the other hand, suffered a notable loss of 40.95 km^2 (1.53%), which may be due to encroachment by built-up areas or reduced cultivation. Forest cover increased by 33.65 km^2 (1.26%), indicating localized regeneration or afforestation processes. More conspicuous spatial changes were evidenced within the period 2033 to 2043. The urbanized area expanded by 80.85 km^2 (3.03%), indicating rapid urbanization and population-driven changes in land use. Water bodies continued to contract by 69.94 km^2 (2.62%), as did a slight decline in barren land of 28.74 km^2 (1.08%). Agricultural land decreased by 17.17 km^2 (0.64%), as expected due to ongoing urban sprawl. The forest cover showed a minimal increase of 0.65 km^2 (0.02%), reflecting relative stability in vegetated areas despite pressure from surrounding land uses. Between 2043 and 2050, the spatial pattern seemed to stabilize, with slight variation in class areas. The built-up area experienced a moderate decline of 21.95 km^2 (0.821%), suggesting potential saturation of expansion. Water bodies registered a minimal fall of 9.82 km^2 (0.367%), and barren land saw a decrease of 14.47 km^2 (0.541%). Agricultural land increased by 19.01 km^2 (0.711%), and forest cover was practically unchanged, with an insignificant fall of 16.17 km^2 (0.623%). Generally speaking, these trends demonstrate a distinct trend of built-up consolidation from 2023 to 2043, followed by spatial stabilization up to 2050, and corresponding decreases in agricultural and water-covered lands.

Table 5 presents the Kappa accuracy values for the forecasted years (2033, 2043, and 2050), indicating the consistency of the CA-Markov model in predicting future LULC trends. The Built-up Area shows a steady reduction in Kappa values from 0.86 (2033) to 0.83 (2043) and 0.79 (2050), suggesting that as urbanization accelerates, the model's predictive consistency slightly diminishes with increasing spatial heterogeneity and the intricacy of built-up area growth processes. Although Kappa values for other classes were not provided, the trend

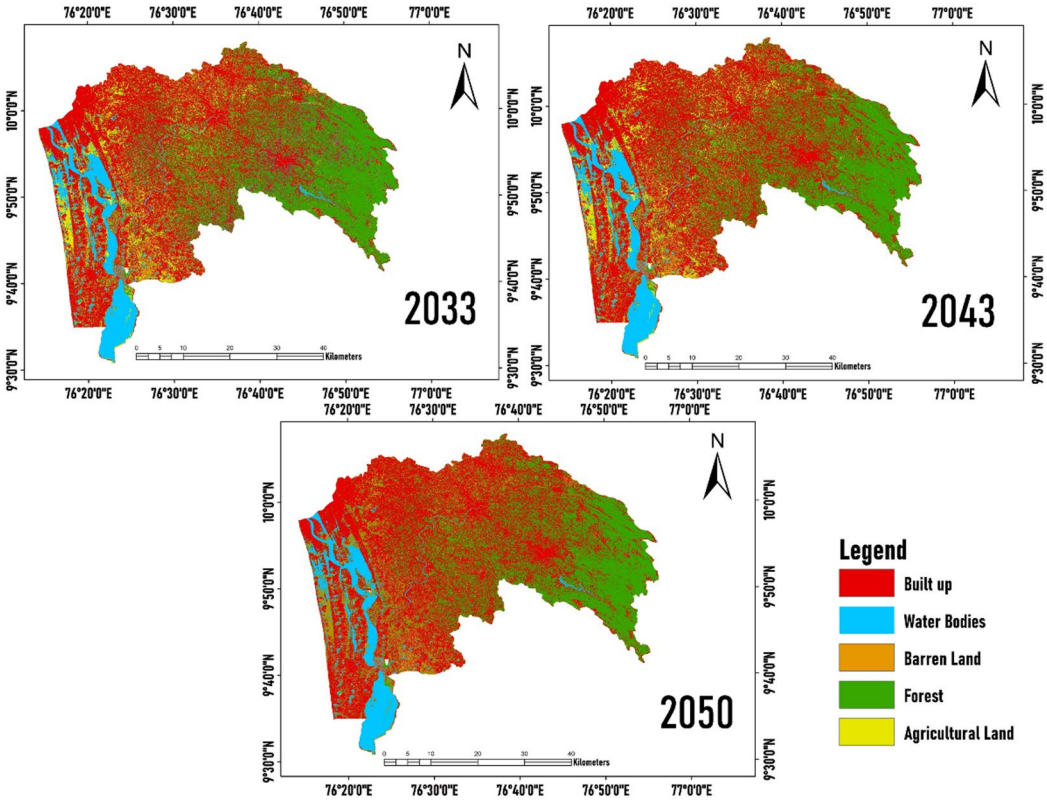


Fig. 7. Future scenario maps (2033, 2043, 2050). Spatio-temporal future prediction LULC mapping was prepared using the QGIS 3.42.1 version software.

Year	2023–2033		2033–2043		2043–2050	
LULC categories	Area (km^2)	Area (%)	Area (km^2)	Area (%)	Area (km^2)	Area (%)
Built-up area	3.31	0.12	80.85	3.025	21.95	0.821
Water bodies	−9.88	−0.37	−69.94	−2.616	−9.82	−0.367
Barren land	13.82	0.52	−28.74	−1.075	−14.47	−0.541
Agricultural land	40.95	−1.53	17.17	0.642	19.01	0.711
Forest	33.65	1.26	0.65	0.024	−16.67	−0.623

Table 4. Projected class areas.

Year	LULC 2033			LULC 2043			LULC 2050		
LULC categories	Area (km^2)	Area (%)	Kappa	Area (km^2)	Area (%)	Kappa	Area (km^2)	Area (%)	Kappa
Built-up area	1173.27	43.94		1254.1	46.97		1276.08	47.79	
Water Bodies	299.64	11.22		299.7	8.6		219.88	8.23	
Barren land	146.42	5.48	0.86	117.68	4.41	0.83	103.21	3.87	0.79
Agricultural land	829.23	31.06		846.4	31.7		865.41	32.41	
Forest	224.63	8.41		224.27	8.43		208.61	7.81	

Table 5. Future prediction area and Kappa.

suggests that higher Kappa values (>0.8) in the initial projections indicate a high level of agreement between the predicted and reference maps, supporting the model's ability to capture short-term transitions reliably. The decrease in Kappa by 2050 is slight, indicating moderate accuracy (range 0.75–0.80), consistent with long-term projections, which tend to increase uncertainty. Overall, the model demonstrates high classification reliability and robust predictive ability through 2043, with a slight dip by 2050 due to the synergistic effects of land-use complexity and the growing built-up area in the Muvattupuzha basin.

Figure 8 illustrates the multi-resolution budgeting analysis for 2033, 2043, and 2050, which assesses the accuracy of LULC predictions under varying spatial and quantitative data conditions. Every one of the three subplots is associated with a projected year's top (2033), middle (2043), and bottom (2050). In 2033, the multiple-resolution budgeting analysis depicts continuous enhancement in agreement values with increasing resolution. The situations with limited spatial and quantitative data have moderate accuracy, whereas those with medium to perfect information have high consistency, indicating consistent short-term model performance.

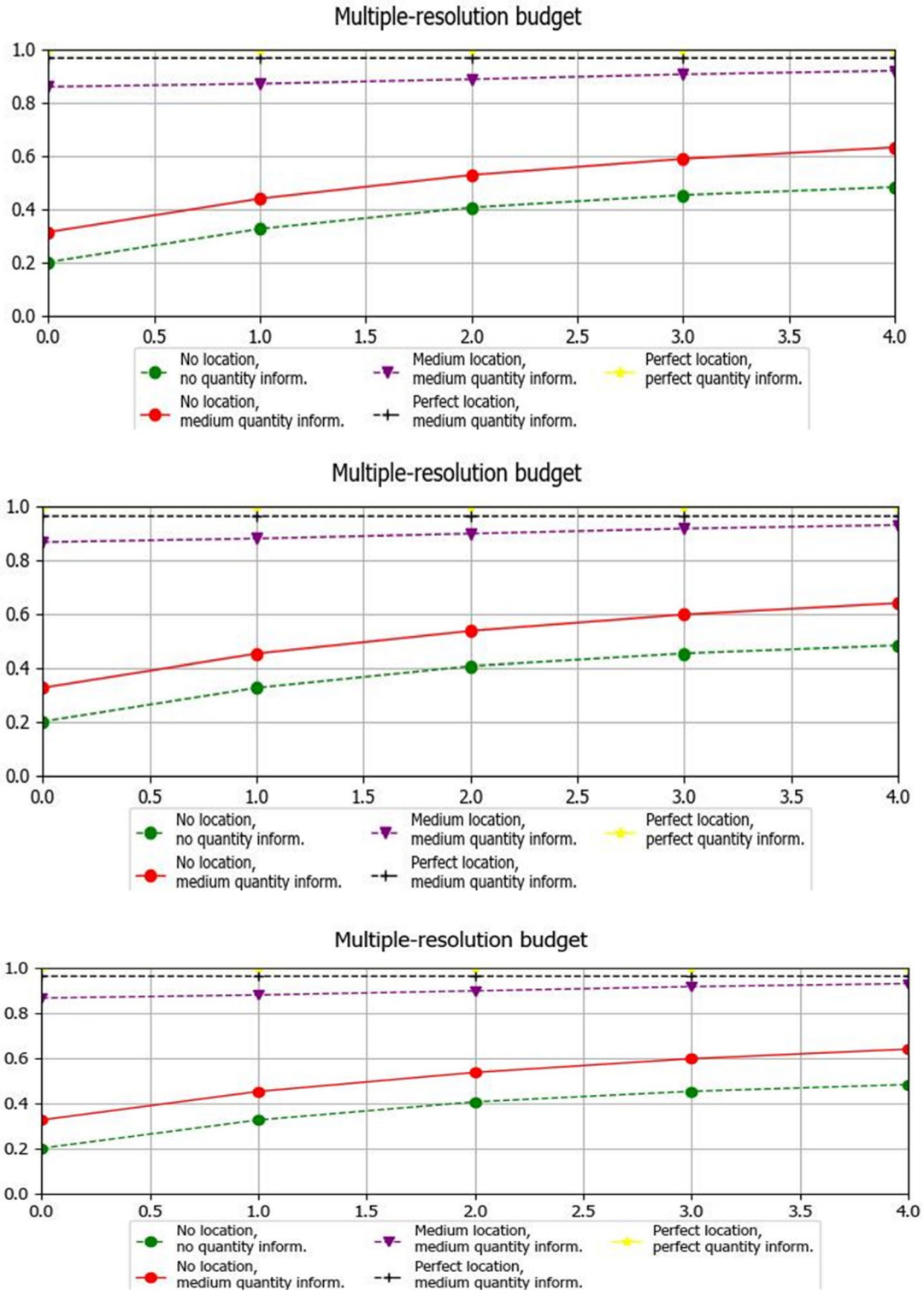


Fig. 8. Multiple-resolution budget analysis for the years 2033, 2043, and 2050. Multiple-resolution budget was prepared the QGIS 3.42.1 version software.

In 2043, the pattern of agreement is not dissimilar, although the trend for improvement weakens marginally, implying a marginal decline in predictability precision from that seen in 2033. The model continues to perform well under conditions with satisfactory spatial and quantitative information, demonstrating high reliability for mid-term predictions. In 2050, the consensus remains optimistic but shows a slightly lower improvement in low-information situations, due to compounded uncertainty in long-term projections. Yet, in both cases, the model still exhibits high precision and consistent prediction results, despite the precise definition of spatial and quantitative information.

Groundwater quality analysis

Physiochemical

Figure 9 shows that the physicochemical properties of groundwater, as measured in 2003, 2013, and 2023, exhibit evident spatial and temporal fluctuations, reflecting both natural and anthropogenic factors within the basin. In 2003, the EC and TDS levels were comparatively moderate in the majority of wells, with a maximum EC of 5300 µS/cm and TDS of 3180 mg/L at well E88, suggesting potential mineralization or contamination from proximal anthropogenic activities. Conversely, wells such as GWE-22 and GWE-21 had low EC and TDS levels, showing comparatively freshwater conditions. Sulfate concentrations were typically below the critical level, ranging from 3 to 12 mg/L in the majority of wells. In contrast, sodium, calcium, and magnesium had moderate concentrations, indicating an evenly balanced ionic composition. Alkalinity concentrations ranged from 5 to 265 mg/L, with higher values in urbanized or farming areas, possibly due to bicarbonate enrichment. Nitrate concentrations were low, indicating minimal leaching of fertilizer or sewage during that period. In 2013, minor fluctuations in EC and TDS were observed in various wells, including BW112 and GWE-03, indicating an improvement in groundwater quality in some aquifers. However, high sulfate and sodium levels in wells E80 and E87 indicated ongoing chemical weathering or anthropogenic sources. Calcium and magnesium levels showed moderate variation, with alkalinity values increasing in wells near agricultural or built-up areas, possibly due to greater infiltration of domestic wastewater or leaching of fertilizers. Fluoride levels were within allowable limits in the majority of wells; however, nitrate concentrations varied slightly, suggesting localized contamination sources.

In 2023, EC and TDS values increased again in certain wells, such as BW113, E80, and E88, underscoring the impact of ongoing land-use changes and urban development. The remarkably high EC and TDS values of E88 (937.33 µS/cm and 562.40 mg/l) reaffirmed its ongoing susceptibility to contamination. Sodium, sulfate, and calcium concentrations also increased in some wells, likely due to ion exchange and leaching of soil or anthropogenic contamination. Alkalinity concentrations ranged from 5 to 163 mg/L, indicating moderate buffering capacity, while fluoride concentrations were generally low. Total hardness showed moderate variation, with higher levels in wells E80 and E88, suggesting higher concentrations of calcium and magnesium ions. Nitrate levels increased in some wells compared to 2013, particularly in BW113 and E95, suggesting potential agricultural runoff or leakage from domestic sewage. Generally, the physicochemical analysis suggests that groundwater quality at the study site has undergone moderate to significant changes over the past two decades. While some wells are stable and of acceptable quality, others exhibit clear mineralization, increased salinity, and nutrient enrichment, which are likely associated with land-use changes, agricultural activities, and urbanization within the basin.

Correlation

Figure 10 presents the correlation heatmaps for 2003, 2013, and 2023, showing the relationships among the major physicochemical parameters of groundwater in the Muvattupuzha basin. The majority of the ionic constituents showed high positive correlations in 2003. EC, TDS, Na, Ca, and Mg exhibit nearly perfect relationships ($r > 0.9$),

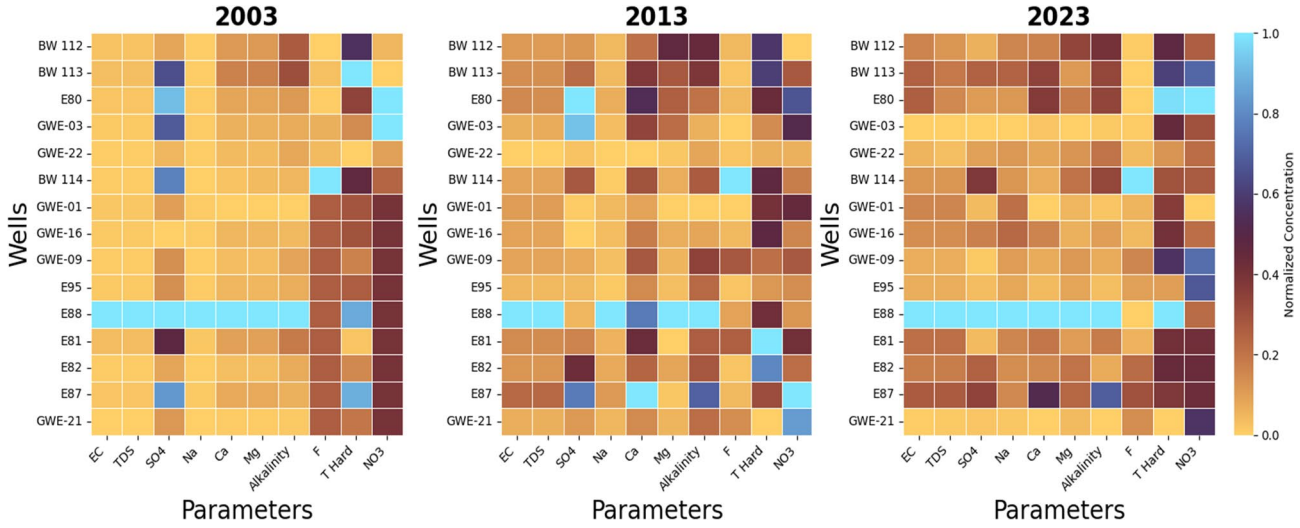


Fig. 9. Physicochemical groundwater parameter vs. wells in three years. Parameters vs. wells data frame the heatmap used MATLAB R2025 version (**a**).

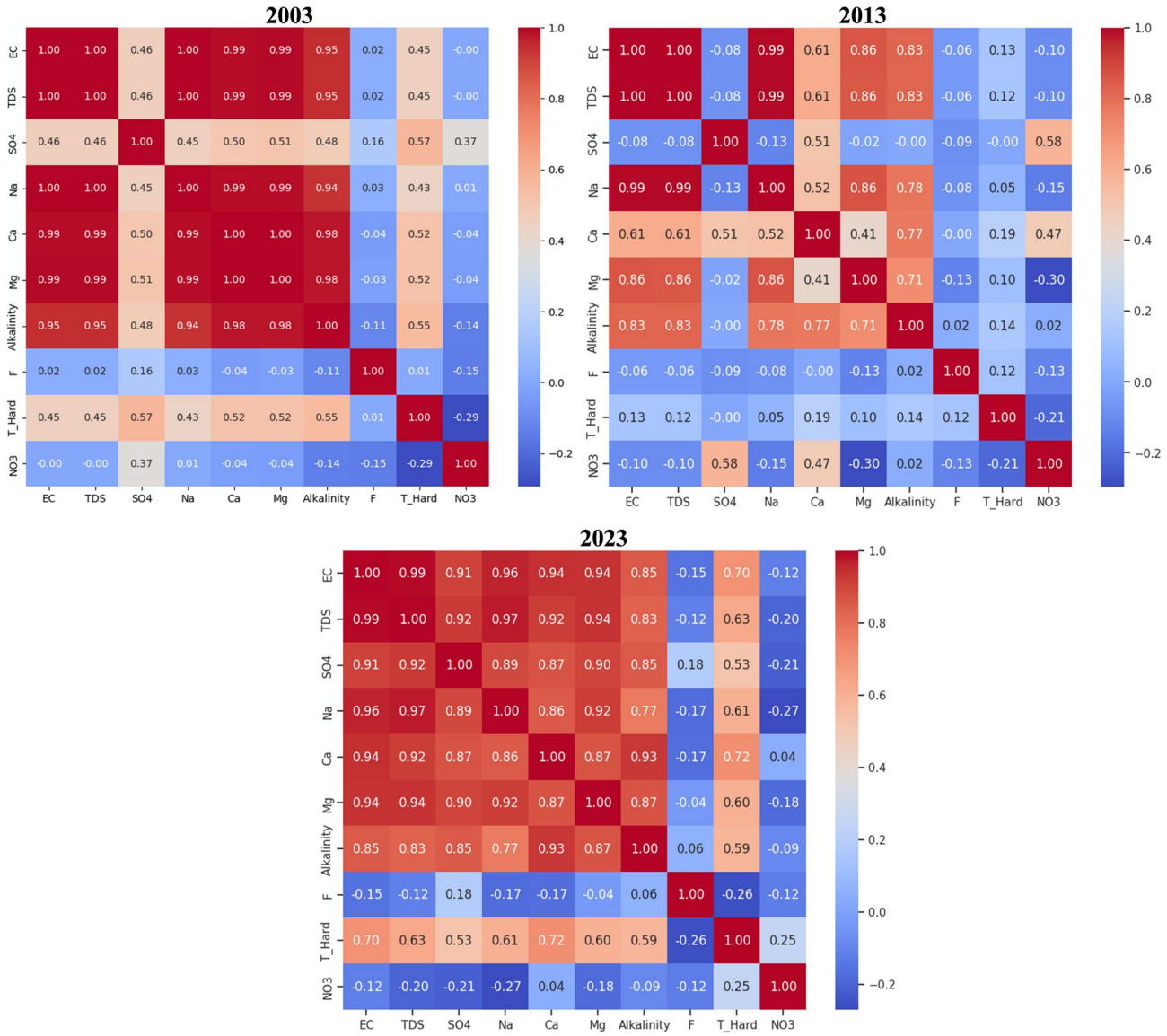


Fig. 10. Relationship between major physicochemical parameters of groundwater in the study area MATLAB 2025(**a**).

which signify that the ionic strength of groundwater was dominated primarily by dissolved salts. Total hardness and alkalinity also show moderate to strong correlations with TDS and EC, reflecting natural geochemical processes such as rock–water interactions. F and NO_3^- have weak or inverse correlations with all parameters, suggesting limited anthropogenic influence at this time. During 2013, correlations between EC, TDS, Na, Ca, and Mg are strong, but reductions in some pairs are slight, reflecting the emergence of variability in groundwater chemistry. The moderate correlations among alkalinity, SO_4, and the major cations suggest that agricultural and domestic contributions may have begun to influence groundwater chemistry. A negative or weak correlation between nitrate and other substances suggests contamination rather than basin-wide contamination from a localized source. By 2023, correlations among most parameters have further intensified, particularly between EC, TDS, Na, Ca, Mg, and SO_4 ($r > 0.9$), reflecting an additional ionic burden and a rise in salinity and hardness due to increased anthropogenic inputs and LULC modifications. The continued negative or weak correlations between fluoride and nitrate, as well as other ions, suggest that their sources are more local, presumably due to agricultural runoff and point contamination. Generally, the period between 2003 and 2023 reflects an increase in ionic interdependence and groundwater mineralization, as would be expected with the expansion of built-up and agricultural land in the basin.

Figure 11 shows that a comparative analysis of groundwater quality parameters for 2003, 2013, and 2023 reveals a distinct temporal trend towards increased mineralization and nitrate enrichment. The average EC rose from around 210 µS/cm in 2003 to 278 µS/cm in 2013, and then to 320 µS/cm in 2023, demonstrating increasing ionic accumulation in the aquifer system. Similarly, TDS increased from 130 mg/L (2003) to 165 mg/L (2013) and 190 mg/L (2023), representing improved leaching and human-induced inputs. Sulfate (SO_4^{2+}) levels fluctuated

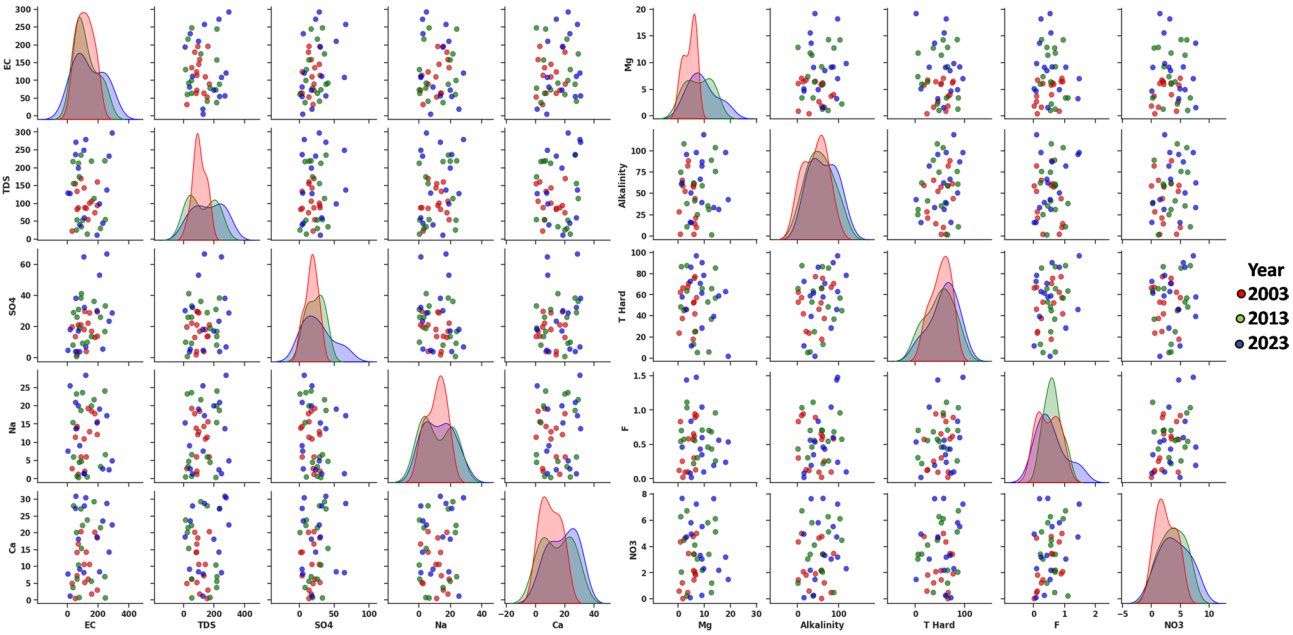


Fig. 11. Pairwise scatter plot in GWQ parameters MATLAB 2025(**a**).

moderately, increasing from 4.1 mg/L in 2003 to 5.6 mg/L in 2023. In contrast, those of sodium (Na^+) and calcium (Ca^{2+}) rose from 7.8 to 18.5 mg/L to 11.6 and 24.3 mg/L, respectively, during the same period.

Mg^{2+} and alkalinity levels also increased, with Mg increasing from 6.9 mg/L in 2003 to 9.8 mg/L in 2023, and alkalinity from 72 mg/L to 96 mg/L, indicating rising carbonate weathering and recharge from anthropogenic origins. Total hardness increased from 78 mg/L (2003) to 104 mg/L (2023), remaining within tolerance limits but indicating a trend toward more complex water types. There was a significant increase in NO_3^- concentration from 1.8 mg/L in 2003 to 3.9 mg/L in 2013 and 6.2 mg/L in 2023, indicating increasing infiltration of agricultural runoff and domestic sewage. There were strong positive correlations between EC, TDS, Na, Ca, and Mg, validating their shared geochemical control and dissolution mechanism; however, pH showed a poor correlation with the other parameters. The pattern of nitrate data points reveals local contamination areas but not even distribution. Generally, the findings indicate a tendency towards a decrease in groundwater quality from 2003 to 2023, primarily driven by ion enrichment and nitrate build-up, attributed to natural hydrochemical development and the anthropogenic effects of land-use modification and agricultural intensification in the region.

Descriptive statistical analysis

Building upon the previous physicochemical evaluation, the descriptive statistical analysis of groundwater parameters provided a more in-depth understanding of the distributional patterns and variability of EC, TDS, SO_4, Na, Ca, Mg, Alkalinity, F^-, TH, and NO_3^- concentrations in the Muvattupuzha Basin. This type of analysis is crucial for situating the physicochemical dataset within context before applying machine learning or multivariate techniques, and for understanding the variability present within the data that will impact predictive modeling. Figure 12 shows the distribution of major groundwater quality parameters, including EC, TDS, SO_4, Na, Ca, Mg, Alkalinity, F^-, TH, and NO_3, across three time periods: 2003, 2013, and 2023. In 2003, the majority of parameters had broad, asymmetric distributions, with higher peaks at lower concentrations. EC and TDS exhibited right-skewed patterns, indicating higher variability and the presence of outliers with extremely high values. SO_4, Na, and Alkalinity distributions were also shifted, showing potential localized contamination or natural mineralization. In 2013, density curves were denser with a marginal left shift in EC and TDS, suggesting a decrease in mean concentration. Ca, Mg, and F^- exhibited moderate peak distributions, whereas NO_3 maintained a wide tail, indicating its oscillating concentration across sampling locations. In 2023, distributions for most parameters were smoother and thinner. EC, TDS, and large cations (Na, Ca, Mg) exhibited decreased peaks and lower variability, indicating an overall improvement in water quality uniformity. SO_4 and NO_3 retained slight skewness, implying some lingering anthropogenic effect, whereas alkalinity and total hardness looked more stabilized near the central tendency.

Figure 13 Hierarchical dendrogram showing clustering of the Muvattupuzha groundwater wells into groups with similar hydrochemistry by using Ward's linkage method. A leaf of the dendrogram (e.g., E88, BW113, GWE-01) corresponds to a groundwater well, whereas the "Distance" refers to the Euclidean dissimilarity between wells or clusters in feature space derived from PCA. Lower distances correspond to wells with similar chemical compositions, while higher distances indicate greater hydrochemical variation. Ward's connection reduces within-cluster variance, grouping wells with similar groundwater chemistry based on EC, TDS, and major ions. Three dominant hydrochemical clusters were observed: Cluster 1 have high EC, TDS, Na, and SO_4, indicating mineralized or anthropogenically affected water. Cluster 2 showing moderately mineralized transitional water.

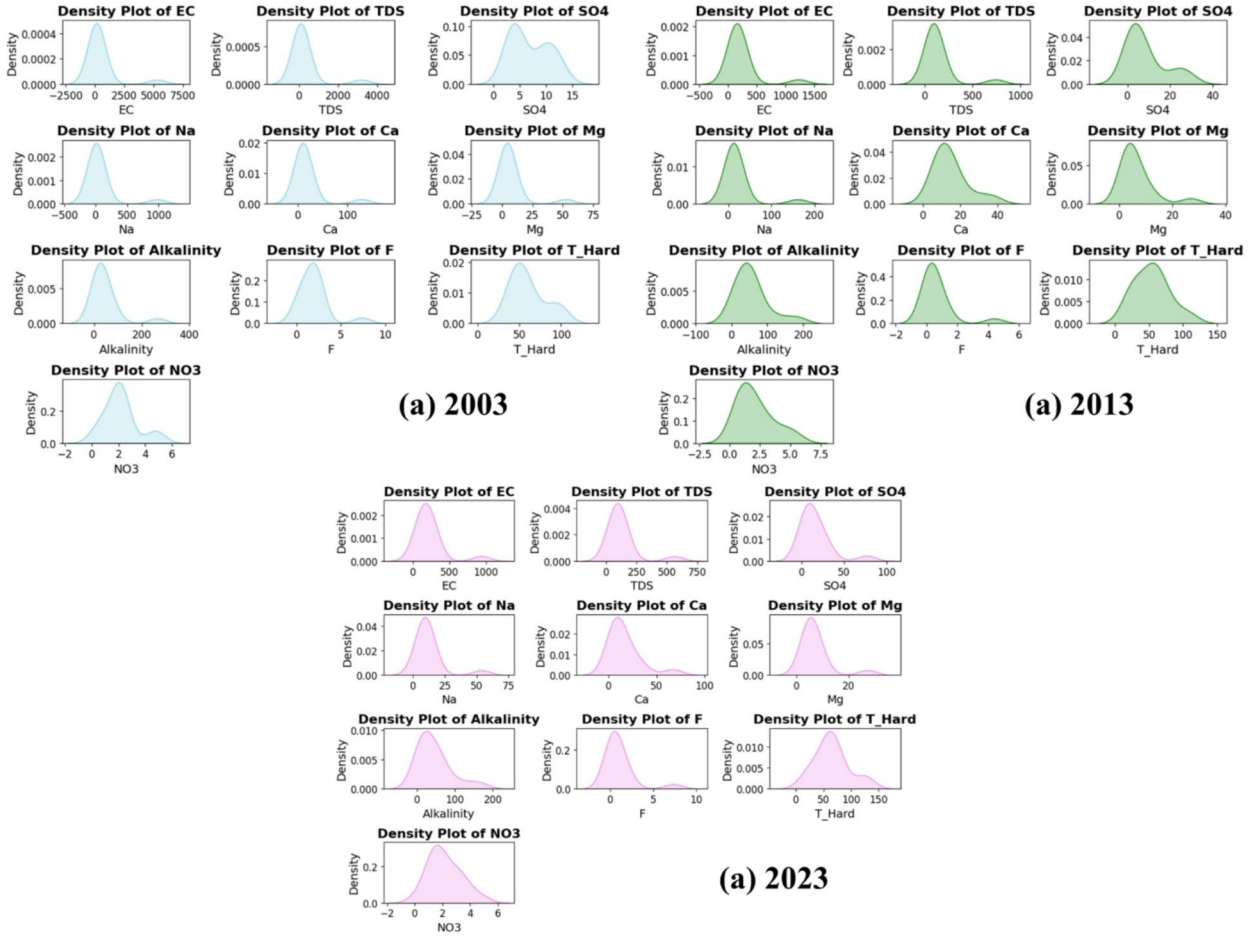


Fig. 12. Identified the distributional behavior and variability of GWQ parameters MATLAB 2025 (**a**).

Cluster 3 characterized by low EC and TDS, which is indicative of younger recharge zones or shallow aquifers in the study area.

Principal component analysis

After conducting a correlation analysis, Principal Component Analysis (PCA) was used to dimensionally reduce and identify the leading gradients influencing variability in groundwater quality parameters within the basin. This is an essential step in groundwater research, as it helps identify dominant processes, simplify intricate data, and make it more interpretable for further clustering or predictive modeling. The PCA and Cluster Analysis of groundwater data collected in 2003, 2013, and 2023 provided detailed insights into the spatio-temporal variability of hydrochemical parameters in the Muvattupuzha basin. The scree plot revealed that PC1 and PC2 accounted for nearly 72% of the total variance, with PC1 contributing approximately 58.5% and PC2 approximately 13.5%. This trend illustrated that both of these components explained most of the large-scale variations in groundwater quality. In 2003, wells E88, E87, and BW113 exhibited extremely high EC and TDS levels, which evidently affected PC1, reflecting mineralization and salinity gradients within the aquifer. These wells were located in areas of extensive land use and potential interaction with weathered or fractured crystalline rocks, where ionic strength was elevated due to rock–water interactions. The PC1 loadings for EC and TDS were high and positive, indicating they were major contributors to total variability. The other parameters, like Ca^{2+}, Mg^{2+}, and Na^{+}, also took positive loading, which implies that the initial component explained the geogenic dissolution of silicates and carbonates responsible for salinity in groundwater. Conversely, NO_3 showed a clear correlation with PC2, distinct from EC and TDS, suggesting that its fluctuations were driven by human-made inputs rather than natural geochemical processes. Wells E80, BW114, and GWE-03 showed comparatively higher nitrate levels, attributed to agricultural effluent runoff and domestic effluent seepage. The orthogonality of nitrate loading in PC2 ensured that salinity and nitrate pollution originated from different sources.

The PCA pattern remained the same in 2013; however, the proportion of variance explained by PC1 decreased slightly, reflecting balanced contributions from both anthropogenic and geogenic processes. Wells such as E88 and E87 continued to control the salinity-related axis (PC1), characterized by high EC (up to 1230 µS/cm) and TDS (up to 738 mg/L). In contrast, wells GWE-09 and GWE-21 exhibited moderate mineralization, accompanied by rising nitrate levels. This suggested local nutrient enrichment, perhaps due to increased agricultural land use at the time. As of 2023, PC1 still described mineralization, with E88 (EC = 937 µmhos/cm, TDS = 562 mg/L) and

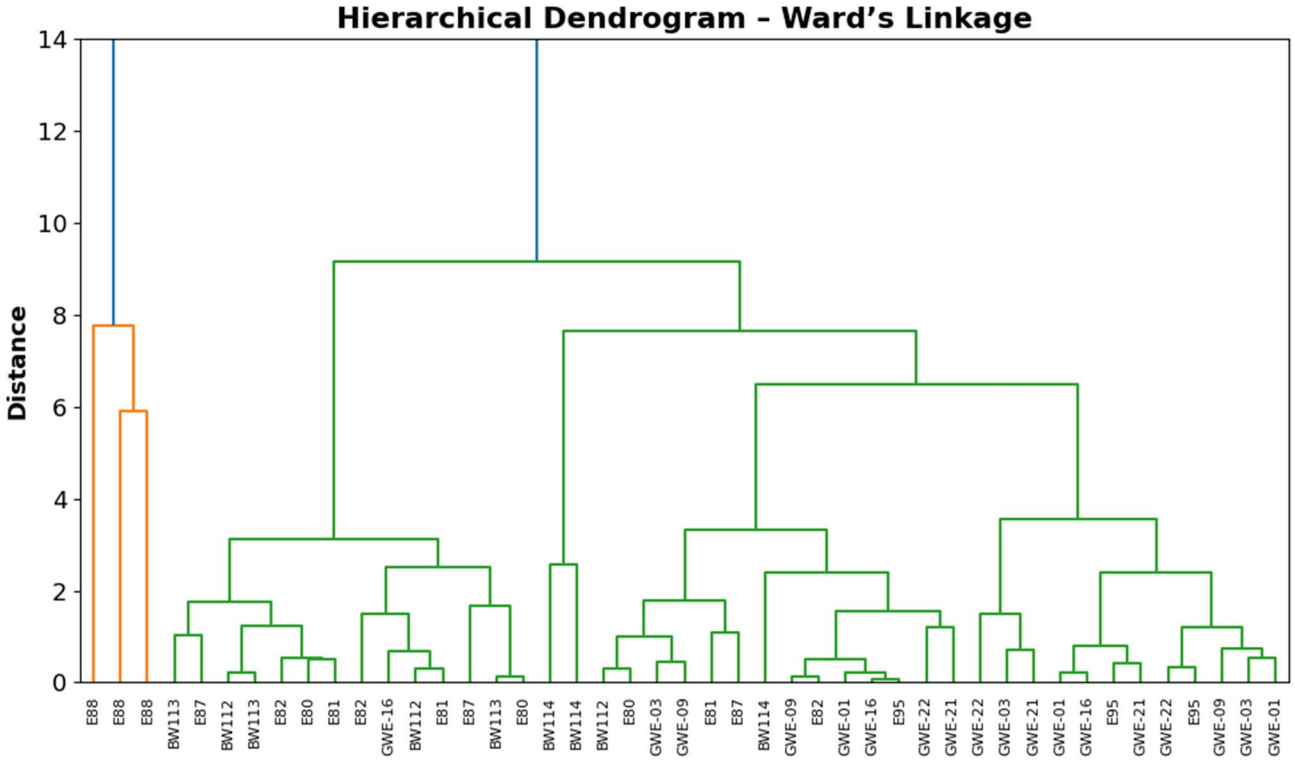


Fig. 13. Well-based groundwater quality using Ward's linkage method MATLAB 2025 (**a**).

E87 (EC = 287 µmhos/cm, TDS = 172 mg/L) still at the highest end of the scale. But a steady drop in ionic content at wells such as BW112, GWE-03, and GWE-21 indicated partial dilution or enhanced recharge conditions. Nitrate levels, on the other hand, rose substantially in wells E80 (4.72 mg/L) and BW113 (3.43 mg/L), indicating that PC2's anthropogenic impact persisted, possibly due to ongoing fertilizer application and urban growth.

The biplot clearly demonstrated that EC, TDS, Ca, and Mg vectors clustered along PC1, while nitrate (NO_3) stretched along PC2, forming an orthogonal relationship that highlighted the dual hydrochemical control. The Cluster Analysis also complemented these PCA results. Hierarchical clustering classified the wells into three predominant clusters: Cluster I consisted of wells with low EC, TDS, and NO_3^- levels, such as GWE-22, GWE-21, and GWE-03, indicating good-quality, recharge-controlled areas. Cluster II: Wells of moderate ionic strength, such as E81, E82, and BW112, are representative of mixed impact by both natural weathering and minimal anthropogenic influences. Cluster III: Wells E88, E87, and BW113, which exhibit high salinity and fluctuating nitrate levels, are indicative of pollution-prone areas associated with intensive land use and shallow aquifers.

Cluster analysis

Throughout the three study periods, the spatio-temporal trend of these clusters showed a progressive migration of a few wells (e.g., BW114, E80) from Cluster II to Cluster III, suggesting an increase in anthropogenic stress in specific localities. Figure 14 indicates that, when combined, PCA and cluster analysis showed that EC and TDS were the primary controlling factors of the groundwater hydrochemical composition, defining the salinity-related axis (PC1). At the same time, nitrate was a secondary, human-caused effect (PC2). The repeated isolation of these constituents throughout 2003–2023 highlighted that both geogenic and anthropogenic processes jointly contributed to groundwater quality, with salinity accumulation induced by rock–water interactions and nitrate enrichment from surface contamination sources. These results underscore the necessity for a comprehensive management approach that integrates both natural and human-induced factors to safeguard groundwater in the Muvattupuzha basin.

Figure 15 illustrates the Elbow, Silhouette, and K-means clustering analysis as a whole, showing the spatial and temporal clustering of Muvattupuzha basin groundwater samples based on their hydrochemical attributes in 2003, 2013, and 2023. Panels (a) and (b) successfully identify the ideal number of clusters (k). In Elbow Method (a), a steep drop in the Within-Cluster Sum of Squares (WCSS) is noted up to k = 2, after which the curve levels off, which means two clusters adequately explain the dataset's variance. This is corroborated by the Silhouette Score (b), which reaches a maximum value (~0.7) at k = 2, which again confirms the existence of two well-separated and compact clusters. Panel (c), showing the PCA-based representation (PC1: 58.5%, PC2: 13.5%), distinguishes the groups distinctly. Cluster 1 (blue) accounts for wells with higher EC, TDS, Na, SO_4, Ca, and Mg, characteristic of mineralized or anthropogenically affected groundwater. In contrast, Cluster 2 (orange) includes wells with lower EC and TDS, reflecting fresher, less mineralized recharge areas. Temporal analysis exhibits clear hydrochemical evolution. In 2003, wells such as E88 and BW113 reported high salinity and TDS, indicating intense mineralization and potential contamination from built-up areas or agricultural sources. In

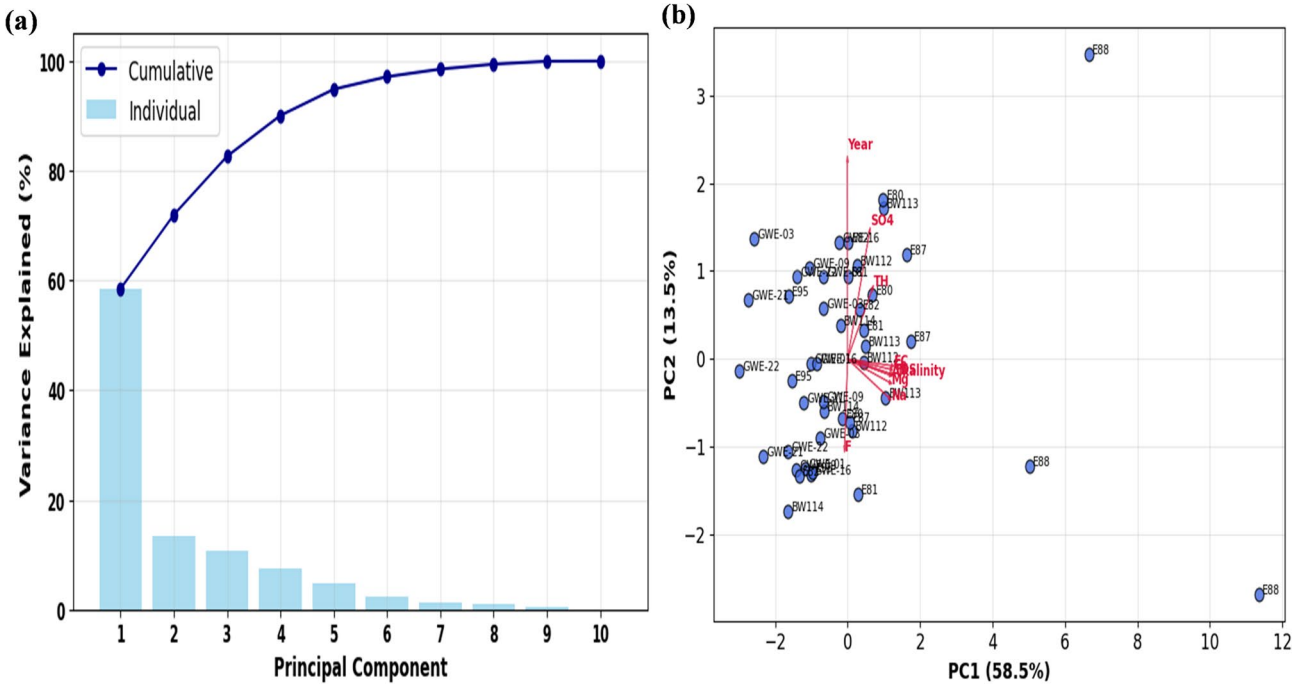


Fig. 14. PCA analysis of water quality parameters (**a**) Screen plot and (**b**) PCA biplot loading of parameters and wells Origin 2018.

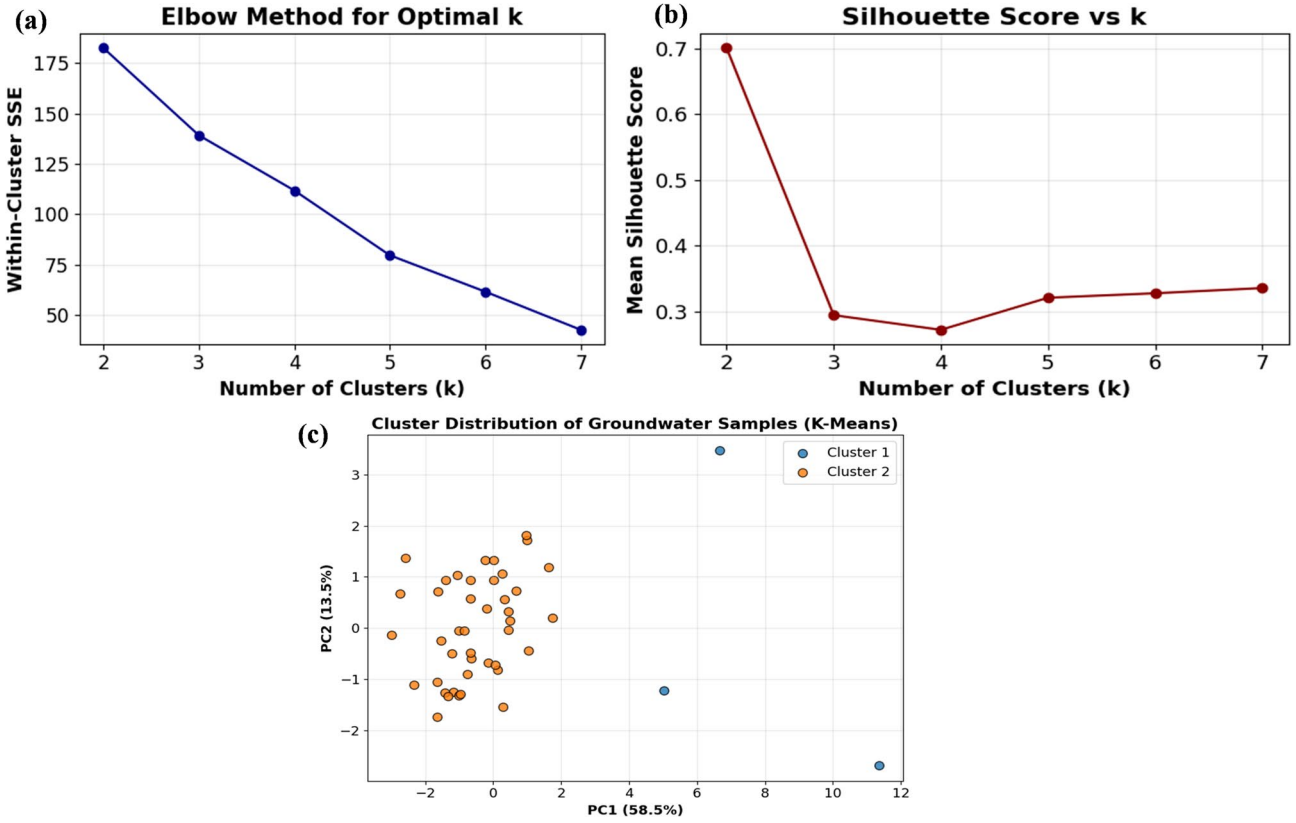


Fig. 15. (**a**-**c**) Validation of cluster and groundwater samples visualization origin 2018.

2013, moderate improvement was observed in some wells, possibly due to dilution from recharge or a decrease in anthropogenic input. By 2023, the increasing concentrations of SO4 and Na in wells such as BW113 and E80 indicate renewed mineralization and a declining dilution capacity. Wells such as GWE-03, GWE-16, and GWE-21 consistently had low EC and TDS levels, indicating steady, clean aquifer areas. Generally, clustering results suggest that the Maiduguri aquifer exhibits spatio-temporal differentiation, regulated by LULC modification, return flows from agriculture, and natural geochemical processes, resulting in several zones of low-, moderate-, and high-mineralization groundwaters.

Predictive machine learning analysis of GWQ

After identifying major groundwater quality trends using PCA and cluster analysis, predictive modelling was conducted to determine whether nitrate levels could be accurately modelled from the major physicochemical parameters: EC, TDS, SO_4, Na, Ca, Mg, Alkalinity, F^-, TH, and NO_3.

This was prompted by initial findings that nitrate showed relatively little correlation with these broad water quality parameters, as indicated by low correlations and separate loadings on the PCA axes. The performance of several ML models (LR, RF, SVR, and XGBoost) was compared using conventional performance metrics: Coefficient of determination (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). The predictive performance of the models was evaluated using the coefficient of determination (R^2), as shown in Fig. 16a–d. All models yielded positive R^2 values, indicating better performance than a trivial mean-based predictor of nitrate concentration. However, there was significant variability in predictive accuracy among the models.

Furthermore, LR achieved the lowest RMSE (0.76) and MAE (0.71), indicating moderately improved generalization, though it still failed to reproduce significant patterns (Fig. 16a). SVR and XGBoost models

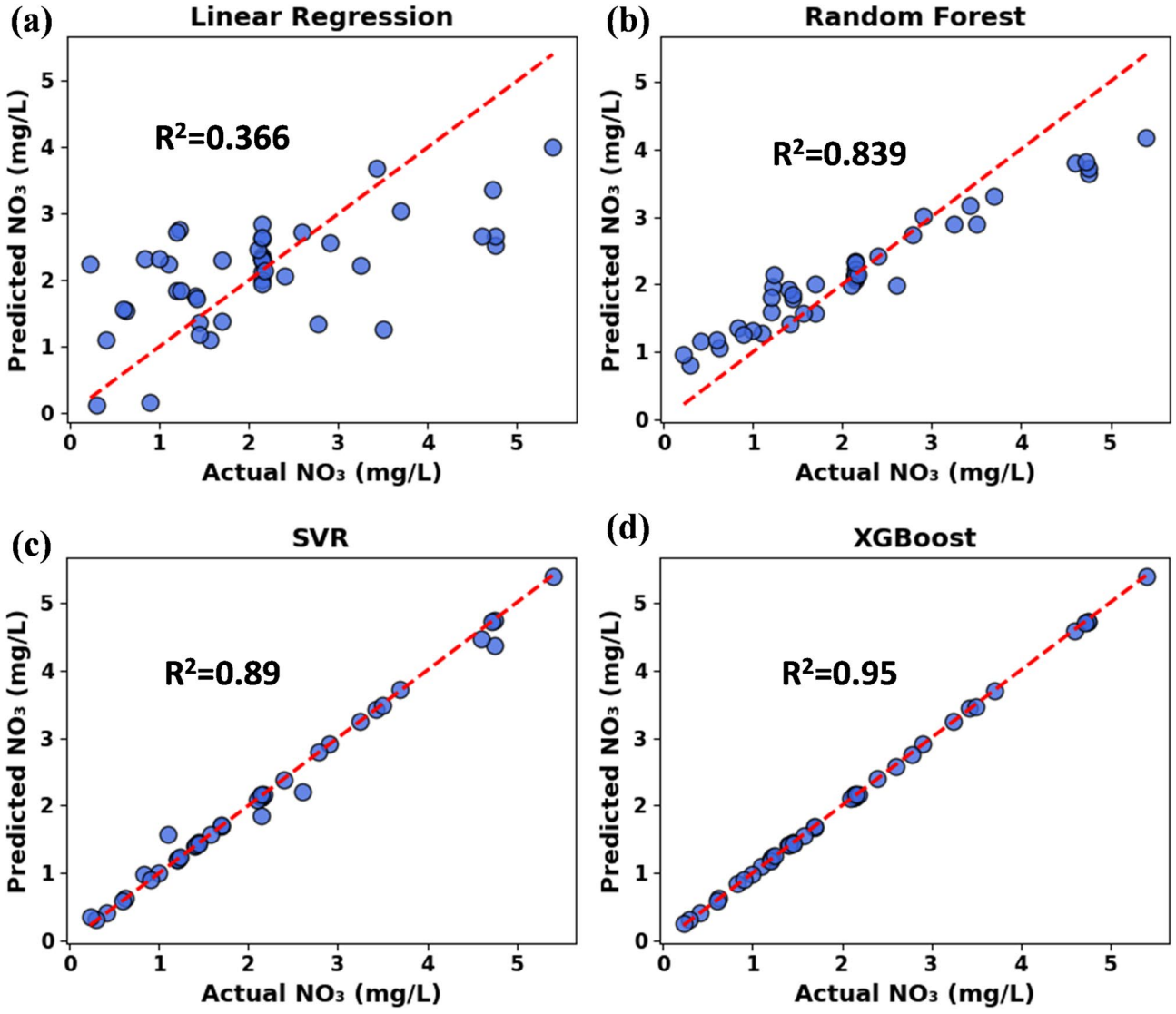


Fig. 16. Performance evaluation of ML models in groundwater quality predicting nitrate concentration using parameters MATLAB 2025 (**a**).

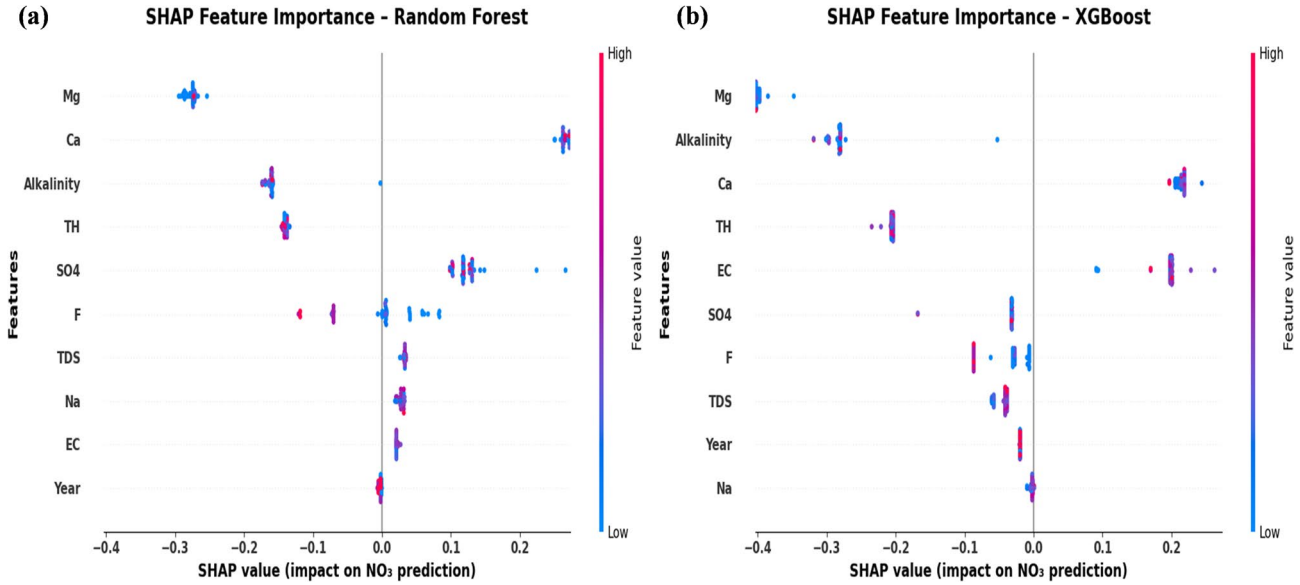


Fig. 17. Illustrates the SHAP feature importance for predicting nitrate levels MATLAB 2025 (**a**).

were on par with each other, though equally disappointing compared to a baseline, with RF having the most significant values for error (Figure. 16b). These results complement the previous multivariate results, where NO_3 was found to vary on a principal component (PC2) orthogonal mainly to the EC and TDS gradients, and to group separately from overall regimes of mineralization determined in the data. The statistical independence here indicates that nitrate levels in Maiduguri groundwater are controlled by mechanisms not represented by pH, EC, or TDS alone.

These factors are very likely to include site-specific anthropogenic inputs, such as proximity to agricultural areas, septic contamination, or local soil permeability, which have been extensively reported to control nitrate distribution in shallow aquifers[36–38]. The significance of these model predictions lies in their importance for regional groundwater safety management. They indicate the insufficiency of using general water quality parameters as surrogates for nitrate contamination. Direct monitoring of nitrates remains crucial for early detection of health hazards, particularly in semi-arid built-up areas where rapid infiltration from improperly managed sanitation systems can create localized hotspots of nitrate buildup[39,40]. Additionally, the relatively strong cluster and PCA findings for EC and TDS reaffirm that these parameters are useful for discriminating between generalized groundwater facies, lending support to the targeted monitoring analysis. The weak predictability of nitrate, however, underscores a pressing need for future studies to include broader environmental predictors, e.g., land-use patterns, depth to the water table, or seasonal recharge dynamics, to improve modelling performance[41–44]. This strategy could ultimately facilitate advanced risk mapping and guide sustainable management strategies adapted to Muvattupuzha's hydrogeological and socio-economic contexts.

Importance of SHAP analysis

Figure 17 illustrates the relative significance and effect of input features on the prediction of nitrate (NO_3) via Random Forest and XGBoost models.

The most significant features in the Random Forest model that control NO_3 prediction are Ca, Mg, TH, SO_4, and F. Both Ca and Mg exhibit both positive and negative SHAP values, indicating that changes in their concentrations significantly influence NO_3 concentrations in either direction. Total hardness and SO_4 contribute moderately, whereas measures like TDS, Na, and EC contribute less. The 2003–2023 feature has less effect, indicating temporal constancy in the nitrate trend. For XGBoost, the critical predictors are Mg, Alkalinity, Ca, TH, and EC. In this case, Mg and Alkalinity have a more pronounced negative SHAP effect, which means that increases in these features are associated with decreases in NO_3 prediction. On the other hand, EC and Ca show positive SHAP values, indicating a positive correlation with NO_3. The characteristics SO_4, F, and TDS contribute slightly, while Na and 2003–2023 are still the least contributive.

Discussion

LULC

The LULC dynamics of the Muvattupuzha basin between 2003 and 2023 are characterized by a significant landscape transformation, primarily driven by rapid urbanization, shifts in agricultural patterns, and fluctuations in forest cover. During the twenty years under consideration, the region has undergone a transformation from a predominantly forested area to one shaped more by human-made land uses[66]. Accordingly, the built-up class exhibited the most significant growth: it expanded from 329.13 sq.km (12.30%) in 2003 to 1187.11 sq.km (44.41%) in 2023, thereby clearly indicating heightened urban expansion driven by population growth, infrastructure development, and industrialization[45]. Throughout this report, an increase in built-up is defined

by remote sensing in a general sense to include compact urban cores, low-density suburban sprawl, peri-urban settlements, mixed residential and commercial zones, industrial zones, transportation networks, paved surfaces, and scattered linear development along roads[5,46,47]. Such urban sprawl primarily occurred at the expense of agricultural and forest areas, thereby indicating increased anthropogenic pressure on natural ecosystems. While the basin area of forest land was 48.49% in 2003, further degradation by 2013 had reduced it to just about 8.88%. These fluctuations in forest cover between 2003, 2013, and 2023 should not be viewed as examples of widespread deforestation followed by rapid natural regeneration. Instead, the changes primarily reflect classification sensitivity to the treatment of vegetation and dense tree cover in medium-resolution Landsat imagery[48,49]. In Kerala, rubber and mixed trees have spectral properties similar to those of natural forests, particularly in mature growth phases, which introduce variability into class assignment based on the classification scheme used, the composition of training samples, and the definition of thresholds[50,51]. In The 2013 classification, agricultural land was primarily designated as such, resulting in an ecologically unrealistic 82% loss of forest cover. In contrast, for 2003 and 2023, plantation- and tree-dominated land covers were included within the broader forest class. This apparent increase in forest cover between 2013 and 2023 does not reflect rapid natural forest regeneration over a single decade. Still, it reflects changes in the classification of mature vegetation that develop a dense canopy structure within 10–15 years and possess a forest-like spectral signature in Landsat data[52]. Similar classification-induced forest cover oscillations have been reported more widely in LULC studies from Kerala and the Western Ghats, and only reiterate that changes largely reflect methodological and vegetation-management effects rather than unrealistically rapid forest loss and recovery[53–56]. This is explicitly included in the revised manuscript to avoid misinterpretation of forest dynamics. This reduction indicates deforestation, encroachment, and conversion for agriculture and built-up purposes. Evaluated that deforestation in the Western Ghats has measurable implications for watershed functioning, including altered infiltration rates, reduced recharge potential, increased surface runoff, and disruption of subsurface flow regimes[57]. Such landscape-level modifications directly influence groundwater availability and quality, particularly in basins like Muvattupuzha, where steep topography and intensive land conversion amplify hydrological sensitivity. However, a 30.56% recovery in forest cover in 2023 highlights emerging trends in reforestation, natural regeneration, and policy-driven afforestation efforts over[58,59]. This suggests that environmental awareness, combined with sustainable land management strategies, may have helped mitigate some of the previous degradation.

Agricultural land represented a highly variable trend during the period under consideration[60]. Between 2003 and 2013, the area under crop cultivation expanded significantly, likely due to increasing food demand and the conversion of previously forested and barren lands into cropland. This expansion represents a well-documented phase of agricultural intensification in central Kerala, supported by conducive monsoonal conditions during 2000–2020, increased surface-water availability, and the expansion of home gardens and mixed perennial cropping systems. These land-use types are classified as agricultural land in satellite-based LULC mapping. The subsequent contraction from 2013 to 2023 reflects a structural transition in land use, driven by rapid urbanization, infrastructure development, and the conversion of farmland to built-up areas, rather than by classification discrepancies. LULC transition analysis indicates that former agricultural areas were predominantly transformed into built-up land, driven by residential expansion, transportation infrastructure, and peri-urban development, particularly along major corridors within the basin[61]. A substantial proportion was also converted into agricultural land, especially rubber and mixed tree plantations, which represent a common land-use transition pathway in central Kerala[62,63]. Minor portions transitioned into barren or mixed-use categories associated with construction and land abandonment. These agricultural-to-urban and agricultural-to-built-up transitions are well documented across Kerala's river basins and align with regional land-use transition models. This trend began to reverse in the following decade, with the agricultural area shrinking sharply by 2023, reflecting the large-scale conversion of farmland into built-up areas [64]. Barren land also showed dynamic changes, increasing significantly up to 2013 due to vegetation loss and deforestation, but then decreasing markedly by 2023, indicative of conversions to built-up or vegetated land. This decrease in barren land during the latter period is indicative of reclamation and development activities, as well as increased vegetation cover in previously degraded zones[65,66]. The water body category generally fluctuated minimally; however, localized areas showed decreases, apparently due to land reclamation, sedimentation, and urban encroachment along riverbanks and wetlands [67]. The magnitude of land transformation is further emphasized by the LULC transition matrix, which indicates that approximately one-third of the total area underwent noticeable class transitions between 2003 and 2023. Among the changes, built-up land shows the highest persistence and the most significant expansion into other classes. The Muvattupuzha Basin falls within the rapidly expanding Kochi urban corridor and exhibits widespread ribbon development, suburban housing, institutional campuses, and road-centric growth. These features are large enough to appear built-up in medium-resolution Landsat data, so the growth largely reflects urban sprawl and land surface sealing rather than a population-driven increase in density within the Muvattupuzha municipality itself. Such large-scale basin-level built-up growth has also been observed in several central Kerala -river basins when areas of peri-urban and linear development are included. It is also important to note that the basin boundary is a hydrological unit and, as such, neither administrative nor municipal; comparisons to statutory urban area percentages at the state level are therefore not appropriate. Observed growth largely reflects the conversion of agricultural land, fallow fields, and agricultural fringes into dispersed, built-up, and infrastructure-heavy areas, consistent with reported land-use trends in Kerala over the past two decades. A significant portion of agricultural and forest classes has been converted into built-up areas, while barren and vegetated classes[66,68]. The limited persistence of natural land classes indicates ecological stress and imbalance resulting from intensive land conversion.

Temporal Kappa accuracy assessments confirm that classification is reliable and consistently improving, from 0.75 in 2003 to 0.85 in 2023.

Table 3 evaluates the CA-Markov model's performance in predicting urban growth dynamics, achieving ~ 88% accuracy and a Kappa value of 0.87 for the basin. The results indicate that built-up growth reflects stable spatial patterns, constrained by existing infrastructure and settlement clustering. A slight underestimation of built-up areas suggests that rapidly changing peri-urban zones cannot be captured by historical transition probabilities. Water body classes are overestimated due to seasonal variations and spectral confusion, as the framework lacks hydrological processes and seasonal fluctuation factors. Minor variations in agricultural land indicate alteration in the cropping pattern and land conversion. Transitional fringes between agricultural and built-up areas are comparatively difficult to simulate, as they rely more on short-term socio-economic factors. The match between observed and predicted barren land class suggests a high temporal stability for the class. However, classification is uncertain due to its spectral similarity with built-up and fallow lands. Slightly underestimated forest classes indicate localized forest conversion or fragmentation. The difficulty increases further, as some plantation types also exhibit similar spectral values. The validation results support the suitability of the CA-Markov framework for basin-scale LULC prediction, especially for major anthropogenic classes. However, certain uncertainties remain regarding the hydrologically sensitive classes; the model results agree reasonably well with the observed distribution, thus supporting the model's further application for LULC projection through 2050. The integration of a seasonal normalization technique, a hydrologic variable, and a socio-economic driver in subsequent studies may improve predictive accuracy. The suitability pattern shows a strong influence from slope and distance to roads, which were weighted higher because both factors are directly relevant to accessibility and construction feasibility. Slopes below gentle thresholds and proximity to major roads consistently yield higher suitability values. Distance to rivers is used as a constraint, resulting in lower suitability values near river corridors to reflect flood risk and environmental protection considerations. Soil properties and elevation represent second-order but significant control on suitability, particularly distinguishing between moderately suitable and highly suitable zones. The integrated approach ensures that no single factor dominates the suitability outcome, thereby enhancing the robustness of the analysis.

The suitability map plays a crucial role in the CA-Markov model by guiding spatial allocations in land-use change. High and very high suitability classes represent areas where future built-up or agricultural expansion is most likely, while low suitability zones limit the likelihood of land conversion. With suitability constraints, CA-Markov avoids unrealistic expansion into environmentally sensitive or physically unsuitable areas, hence enhancing the spatial realism of future LULC scenarios and bringing model outputs closer to ground conditions. This indicates growing confidence in the spatial interpretation of classification precision. The noted gradual increase in accuracy also speaks to the effectiveness of the methods used[69]. The CA-Markov projections depict continued urban expansion through 2033 and 2043, followed by stabilization toward 2050. The built-up area would retain its leading position in land cover, with further gradual losses in agriculture and water body areas. The modest near-term recovery in the forest category suggests that, beyond 2043, there may be some ecological balance or constraints on further reforestation. Predictive uncertainty increases for longer-term projections, a common phenomenon in spatial simulation models due to the compounding of variability in land-use behavior and exogenous socio-economic drivers. This is evident in the decrease in Kappa values from 0.86 in 2033 to 0.79 in 2050. In sum, the changes over time denote the transformation of the Muvattupuzha basin from a predominantly natural landscape to a human-modified urban environment. The consistent growth in built-up areas, coupled with the contraction of agricultural and forest lands, reflects the strong impact of urbanization, infrastructure development, and population growth[70,71]. Although recent signs of forest recovery and a reduction in barren land are encouraging, the loss of agricultural space and alteration of hydrological surfaces pose challenges for the sustainable management of land and water resources. To ensure balanced development in the future, strategic land-use planning, strict environmental regulation, and effective conservation programs must be adopted to maintain ecological stability while accommodating inevitable urban growth[72,73].

Groundwater quality analysis

The overall spatio-temporal pattern of groundwater quality in the Muvattupuzha Basin demonstrates a clear hydrochemical evolution influenced by both natural geogenic processes and intensifying anthropogenic pressures. The physicochemical analysis (Fig. 9; SI Table 7 A) reveals that EC and TDS increased steadily from 2003 to 2023, particularly in wells E88, BW114, and E80, indicating progressive mineralization associated with urban expansion and reduced dilution capacity. Such trends are consistent with observations in semi-urban basins of southern India and other tropical regions, where rapid land-use transformation alters the recharge-discharge balance and enhances solute loading[74,75]. The concurrent rise in Na, Ca, and Mg supports enhanced silicate and carbonate weathering under variable redox conditions, a process similarly reported in the Mekong Delta[75] and parts of the North China Plain[76]. The correlation results in Fig. 10 highlight strong positive associations among EC, TDS, Na, Ca, and Mg, confirming that mineralization is primarily governed by ion exchange and dissolution of aquifer materials. The consistent independence of nitrate (NO_3) from these major ions indicates that its occurrence is largely anthropogenic, originating from agricultural leaching and domestic effluent infiltration rather than natural hydrogeochemical controls. The observed threefold increase in mean NO_3 concentration from 2003 to 2023 aligns with the rising nitrate contamination reported globally in shallow aquifers, such as those in the Indo-Gangetic plains, the Nile Delta[84], and the Gharb Plain, Morocco[77].

Descriptive statistical plots (Fig. 12) further emphasize the non-normal, right-skewed distributions of EC, TDS, and Na^+, reflecting localized hotspots of salinity and contamination. By 2023, slightly narrower density curves suggest improved water uniformity in some recharge areas, possibly due to dilution from monsoonal infiltration or reduced fertilizer intensity in plantations. However, persistent skewness in the distributions of SO_4 and NO_3 confirms that anthropogenic stress remains spatially uneven. Similar distributional asymmetry in groundwater solutes has been reported in the Cauvery Delta[79] and Cuddalore district[80], indicating that such skewed patterns are typical in mixed-land-use basins with variable recharge. Figure 14 from PCA provided

further insight into the dominant hydrochemical processes. The first principal component (PC1) accounted for approximately 62% of the total variance and was primarily controlled by EC, TDS, Ca, Mg, and Na, indicating geogenic mineralization and salinity gradients. The second component (PC2), influenced mainly by NO_3, represented anthropogenic nutrient loading. The orthogonality of NO_3 along PC2 reaffirms that nitrate contamination is decoupled from salinity processes. Similar dual-component patterns, separating natural and anthropogenic influences, have been documented in aquifers of the Tral Plateau of Iran[81] and the Mediterranean coast[82]. The clustering analysis in Fig. 15 complements these results by delineating three hydrochemical facies:

- Cluster I: low-mineralized recharge zones (wells GWE-21, GWE-22, GWE-03).
- Cluster II: transitional zones of mixed influence (E81, E82, BW112).
- Cluster III: high-salinity and nitrate-impacted wells (E88, E87, BW113).

The progressive migration of wells such as E80 and BW114 from Cluster II to III between 2003 and 2023 reflects intensifying land-use pressure and reduced aquifer resilience. Comparable temporal facies shifts were observed in the groundwater of the Western Ghats[83] and northern Thailand[84], where rapid agricultural conversion triggered similar hydrochemical transitions. Machine-learning-based models resulted in positive R^2 values for nitrate (Fig. 16), indicating that regression models using conventional physicochemical inputs performed better than a trivial mean-based predictor of nitrate concentration. This outcome reinforces the PCA results, indicating that unmeasured environmental variables, such as land use, aquifer depth, and soil permeability, influence nitrate behavior. Although SVR produced the lowest RMSE (0.76), its predictive capacity remained limited, echoing findings by[85] in Bangladesh and[13] in Iran, where nitrate prediction required spatial or climatic covariates. The SVR and XGBoost models, despite strong generalization in other hydrochemical contexts, underperformed for nitrate, underscoring the need to integrate spatial LULC and hydrological predictors in future modelling frameworks.

The SHAP values in Fig. 17 show that the models consistently assign higher importance to alkaline earth elements (Ca^{2+} and Mg^{2+}), total hardness, sulfate (SO_4^{2-}), and alkalinity. At the same time, parameters such as EC, TDS, and Na^{+} contribute relatively little. Coexistent positive and negative SHAP values for Ca and Mg point to nonlinear effects and context-dependent relationships in nitrate estimates, rather than purely monotonic effects. This follows the general pattern observed in groundwater systems, where nitrate mobility is controlled by site-specific geochemical buffering, ion exchange, and aquifer heterogeneity. This is reinforced by the relatively low contributions of both EC and TDS to the multivariate conclusions, which are in line with our starting hypothesis that nitrate is not strongly controlled by general mineralization gradients[86,87].

In contrast, SHAP patterns indicated localized, process-driven controls, likely related to anthropogenic inputs such as agricultural activities, sanitation practices, and variations in soil permeability. The interpretations are consistent with previous studies, which conclude that nitrate contamination in shallow aquifers is often decoupled from the bulk hydrochemical facies and is controlled by land-use-specific processes. The SHAP analysis should therefore be considered not as a demonstration of robust predictive power, but rather as a hypothesis-generating framework that highlights possible controlling variables and nonlinear interactions that are often absent in more traditional statistical approaches[88,89]. In the RF model, Ca and Mg had the highest SHAP values, suggesting an indirect influence on nitrate behavior through hardness and aquifer lithology. In XGBoost, Mg and Alkalinity emerged as dominant negative drivers, indicating that more buffered, mineral-rich waters tend to suppress nitrate accumulation, possibly through enhanced denitrification in reducing environments. Such feature-importance trends have been observed globally, including in machine-learning studies of nitrate dynamics in California's Central Valley[90] and the Pravara River[90]. Together, the driver importance plots underline the multivariate and site-specific nature of nitrate contamination, where no single physicochemical factor alone dictates risk, but rather their interaction within changing land-use and hydrological contexts. Integrating all results, the study demonstrates that groundwater quality in the Muvattupuzha Basin has undergone a measurable transformation over the past two decades, shifting toward higher mineralization and localized nitrate enrichment. The combined use of correlation analysis, PCA, clustering, and SHAP-based interpretation[92]. These findings are consistent with research on rapidly urbanizing tropical watersheds, underscoring that sustainable groundwater management must integrate continuous quality monitoring with LULC regulation and nutrient-input control to prevent long-term aquifer degradation.

Spatio-temporal analysis of LULC changes and their impact on groundwater quality

The Muvattupuzha basin exhibits pronounced spatio-temporal dynamics that couple rapid demographic change, intense monsoonal precipitation, shifting land uses, and heterogeneous subsurface conditions, all of which shape groundwater quantity and quality. Between 2003 and 2023, the study area experienced accelerated urban expansion and the conversion of vegetated/agricultural land to built surfaces and plantations, resulting in precise LULC trajectories characterized by increases in impermeable cover, fragmentation of natural land, and localized intensification of quarrying and built-up areas. These landscape transitions alter hydrological partitioning (reduced infiltration, higher surface runoff) and create new pathways for contaminant mobilization from surface sources into shallow aquifers[93,94]. Recent regional LULC assessments for the Muvattupuzha catchment, therefore, provide the essential spatial context for linking observed groundwater trends to land-use drivers[95]. The basin's hydroclimatic setting accentuates these anthropogenic pressures. The Muvattupuzha catchment lies within a humid tropical regime, with average annual rainfall exceeding 3000 mm, concentrated during the southwest monsoon months. Such high, seasonally pulsed precipitation produces large surface flows and episodic recharge events; when combined with expanded impervious cover, however, the effective recharge to shallow aquifers becomes both spatially heterogeneous and temporally erratic[96]. Intense monsoon runoff also increases the potential for mobilizing surface-derived pollutants (sediment, nutrients, agrochemicals, and urban

effluents) into stream networks and connected alluvial aquifers during high-flow periods[97,98]. The spatial pattern of urban expansion was overlaid with the locations of groundwater monitoring wells to explicitly link land-use modelling results with changes in groundwater chemistry. Wells located within identified urbanization hotspots consistently recorded higher levels of contamination. For instance, well (E88) is located in a Coastal zone, and industries projected by the CA–Markov model to have rapid built-up expansion from 2013 to 2023, and recorded elevated NO_3, EC, and TDS concentrations, reflecting increasing anthropogenic input. Likewise, wells BW113 and BW114, located in Muvattupuzha town and positioned close to a high-density built-up growth corridor, showed persistent degraded groundwater quality and, hence, seemed to be influenced by urban runoff, septic effluents, and reduced infiltration capacity. In contrast, wells located in relatively stable forested or agricultural zones had lower solute concentrations, underscoring the spatial sensitivity of groundwater chemistry to localized land-use transitions.

Demographic and socio-economic trends amplify exposure pathways. The principal urban centers in the basin, including Muvattupuzha municipality, have experienced steady population growth over recent decades, resulting in increased domestic water demand and generating larger volumes of wastewater and septic effluent. Census figures indicate that town populations, household density, and peri-urban settlement footprints contribute to intensified groundwater abstraction and localized contamination risks where sanitation or wastewater treatment infrastructure is inadequate. The combined effect of growing population and land conversion is therefore a dual stress: increased pollutant loading per unit area and greater extraction from shallow wells[99].

Soil and geologic heterogeneity exert first-order controls on both contaminant attenuation and recharge patterns. Lateritic and mixed alluvial soils dominate much of the basin; lateritic profiles are typically well-drained but low in organic matter and nutrient retention capacity, whereas alluvial deposits in valley bottoms provide higher porosity and more direct hydraulic connectivity with streams[100,101]. These contrasts mean that pollutant fate varies sharply over short distances: nutrients and salts applied on upland lateritic slopes may be rapidly mobilized during monsoon runoff. At the same time, low-lying alluvial zones may receive concentrated loads that more readily infiltrate the water table. Thus, spatial variations in soil type and vadose-zone thickness must be considered when interpreting groundwater quality heterogeneity[102]. Empirical assessments of groundwater chemistry from the basin and adjacent study areas reflect these combined influences. Field surveys and hydrochemical analyses report generally acceptable ranges for many physicochemical parameters in rural parts of the basin. Still, also document localized exceedances of elevated electrical conductivity, increased TDS, and episodic spikes in nitrate and specific ions, particularly in peri-urban and agricultural hotspots. These signatures are consistent with fertilizer leaching, sewage/septic effluent intrusion, and salinization processes in lowland areas subject to irrigation return flows or reduced flushing[39,103–107]. The literature therefore supports a mechanistic link: LULC change, along with high monsoonal fluxes and heterogeneous soils, produces spatially variable groundwater degradation, concentrated near pollutant sources and in hydraulically connected alluvial corridors. Taken together, the evidence recommends a targeted monitoring and land-management approach: spatially stratified sampling across soil and LULC classes, control of point and diffuse pollutant loads in rapidly urbanizing zones, preservation, or restoration of recharge-friendly land covers in critical subbasins, and regulation of quarrying and effluent discharges to reduce acute impacts[108]. Only by linking high-resolution LULC change maps with hydrogeological and socio-demographic data can effective mitigation strategies be designed to sustain groundwater resources in the Muvattupuzha basin.

Conclusion

The integrated assessment of LULC dynamics and their influence on groundwater quality between 2003 and 2023 reveals a profound transformation in the Muvattupuzha River Basin, leading to a significant disruption in its hydro-environmental balance. Over the two decades, the basin has transitioned from a predominantly forested landscape to one increasingly dominated by built-up and human-modified environments. The built-up area expanded dramatically from 12.30% (329.13 km^2) in 2003 to 44.41% (1,187.11 km^2) in 2023, mainly at the expense of agricultural and forest land, while water bodies and vegetation cover declined notably. Hydrochemical investigations confirmed that the rising concentrations of EC, TDS, Ca, Mg, and Na reflect both enhanced rock–water interaction and anthropogenic inputs. NO_3 exhibited strong spatial variation, indicating that fertilizer leaching, septic effluents, and urban runoff were the primary sources of contamination. The GWQI classification revealed deterioration in water quality over time, with most wells affected in 2003–2023, particularly the coastal zone (E88) well. These findings signify that uncontrolled urban sprawl and agricultural intensification have exerted cumulative pressure on groundwater quality and ecosystem sustainability.

Statistical and ML analyses provided deeper insight into the mechanisms controlling groundwater chemistry. Correlation and PCA revealed that both geogenic and anthropogenic factors govern groundwater composition, while cluster analysis identified three major hydrochemical facies, showing a temporal shift toward more contaminated groups. Advanced predictive modeling using RF, SVR, and XGBoost demonstrated strong potential for forecasting groundwater quality trends. However, nitrate dynamics showed nonlinear relationships, indicating that physicochemical parameters alone cannot fully explain its variability. SHAP-based interpretation identified Mg, Ca, and Alkalinity as key regulators of nitrate concentration, reflecting denitrification processes and the aquifers' buffering capacity. The findings underscore the need to integrate hydrochemical monitoring, geospatial analysis, and machine learning-based forecasting into sustainable water resource management frameworks. Future planning must focus on protecting recharge zones, regulating urban expansion, and implementing adaptive strategies to mitigate pollution hotspots, ensuring the long-term ecological and hydro-environmental stability of the basin.

Despite these robust findings, several limitations should be acknowledged. Due to data availability, the LULC analysis was confined to a 20-year window; similarly, observations of groundwater quality were available only for discrete years (2003, 2013, and 2023), limiting the ability to evaluate continuous temporal trends. Monthly

records of groundwater quality were not continuously available at many monitoring stations, further restricting the ability to capture seasonal variability, particularly in the pre- and post-monsoon seasons. Detailed aquifer parameters were not available for many locations-for example, porosity, hydraulic conductivity, and fracture connectivity-which prevented the direct inclusion of hydrogeological controls imposed by the subsurface; therefore, the controlling role of lithology and structure has been interpolated indirectly through hydrochemical indicators and statistical relations. However, the holistic LULC and machine learning framework used in this study provides a reliable basin-scale evaluation of groundwater vulnerability. The outputs underscore the imperative need to safeguard recharge zones, control unplanned urban growth, and mitigate nitrate pollution hotspots to ensure sustainable groundwater management and long-term hydro environmental resilience of the Muvattupuzha Basin.

Data availability

All data generated or analysed during this study are included in this published article [and its supplementary information files].

Received: 18 November 2025; Accepted: 2 February 2026
Published online: 13 February 2026

References

1. Dong, S., Guo, H., Chen, Z., Pan, Y. & Gao, B. Spatial Stratification Method for the Sampling Design of LULC Classification Accuracy Assessment: A Case Study in Beijing, China. *Remote Sens. (Basel)* https://doi.org/10.3390/rs14040865 (2022).
2. Li, X. et al. Hydrochemical characteristics and nitrate health risk assessment in a shallow aquifer: insights from a typical Low-Mountainous region. *Water* **17** (24), 3516. https://doi.org/10.3390/w17243516 (2025).
3. Siddiqui, K. Impact of population changes and economic growth in China and India. *World* (2024).
4. UN-DESA, *World Population Prospects* https://www.un.org/development/desa/pd/2022
5. Murmu, J., Radhadevi, L., Pande, C., Bandaru, M. & Kumar, M. Indicators of sustained agriculture, impacts of LULC and weather parameters on ET: Case study in Chota Nagpur Plateau. *Environ. Sustain. Indic.* **27**, 100836. https://doi.org/10.1016/j.indic.2025.100836 (2025).
6. Tassi, A. & Vizzari, M. Object-oriented lulc classification in Google Earth engine combining snic, glcm, and machine learning algorithms. *Remote Sens. (Basel)*. **12** (22), 3776 (2020).
7. Shaji, J., Sajith, S. L., Joseph, J. & Ramachandran, K. K. LULC change along Central Kerala coast and perception on implementation of CRZ notification. *National Conf. Geosp. Technol.* (2017).
8. Pandey, S. & Kumari, N. Prediction and monitoring of LULC shift using cellular automata-artificial neural network in Jumar watershed of Ranchi District, Jharkhand. *Environ. Monit. Assess.* https://doi.org/10.1007/s10661-022-10623-6 (2023).
9. Nair, S. B. & CJ, P. Urbanization in Kerala—What Does the Census Data Reveal? *J. Human Dev.* vol. (2017).
10. Gopinath, G. Chemistry of groundwater in lateritic terrains of the Muvattupuzha river. (2017).
11. Giriraj, A., Irfan-Ullah, M., Murthy, M. S. R. & Beierkuhnlein, C. Modelling spatial and temporal forest cover change patterns (1973–2020): A case study from South Western Ghats (India). *Sensors* **8**(10), 6132–6153. https://doi.org/10.3390/s8106132 (2008).
12. Reddy, C. S. Assessment and monitoring of long-term forest cover changes (1920–2013) in Western Ghats biodiversity hotspot. http://www.natureasia.com
13. Zegaar, A., Ounoki, S. & Telli, A. Machine learning for groundwater quality classification: A step towards economic and sustainable groundwater quality assessment process. *Water Resour. Manage.* **38** (2), 621–637 (2024).
14. Che Nordin, N. F. et al. Groundwater quality forecasting modelling using artificial intelligence: A review. *Groundw. Sustain. Dev.* https://doi.org/10.1016/j.gsd.2021.100643 (2021).
15. Yang, S. et al. Spatial mapping and prediction of groundwater quality using ensemble learning models and SHapley additive explanations with Spatial uncertainty analysis. *Water (Switzerland)*. **16** (17), Sep–2024. https://doi.org/10.3390/w16172375 (2024).
16. Xiong, H. et al. Critical role of vegetation and human activity indicators in the prediction of shallow groundwater quality distribution in Jianghan plain with LightGBM algorithm and SHAP analysis. *Chemosphere* **376**, 144278. https://doi.org/10.1016/j.chemosphere.2025.144278 (2025).
17. Pandya, H., Jaiswal, K. & Shah, M. A comprehensive review of machine learning algorithms and its application in groundwater quality prediction. *Arch. Comput. Methods Eng.* **31** (8), 4633–4654 (2024).
18. Maya, K., Santhosh, V., Padmalal, D. & Kumar, S. R. A. Impact of mining and quarrying in Muvattupuzha river basin, Kerala-An overview on its environmental effects. *Bonfring Int. J. Industrial Eng. Manage. Sci.* **2** (1), 36–40 (2012).
19. Anand, B., Rekha, R. S., Radhakrishnan, N. & Ramaswamy, K. Analysis of LULC change dynamics and its impact assessment using CA-ANN model in part of Coimbatore region, India. *GeoJournal* https://doi.org/10.1007/s10708-023-10944-0 (2023).
20. Kumar, A. A., Dipu, S. & Sobha, V. Seasonal variation of heavy metals in Cochin estuary and adjoining periyar and Muvattupuzha rivers, Kerala, India. *Global J. Environ. Res.* **5** (1), 15–20 (2011).
21. Ali, H. Y., Priju, C. P. & Prasad, N. B. N. Delineation of Groundwater Potential Zones in Deep Midland Aquifers Along Bharathapuzha River Basin. *Kerala Using Geophysical Methods, Aquat Procedia.* **4**, 1039–1046. https://doi.org/10.1016/j.aqpro.2015.02.1 (2015).
22. Shahul Hameed, A. et al. Isotopic characterization and mass balance reveals groundwater recharge pattern in Chaliyar river basin Kerala, India.. *J Hydrol Reg Stud.* **4**, 48–58. https://doi.org/10.1016/j.ejrh.2015.01.003 (2015).
23. Ribinu, S. K., Prakash, P., Khan, A. F., Bhaskar, N. P. & Arunkumar, K. S. Hydrogeochemical characteristics of groundwater in Thoothapuzha River Basin, Kerala, South India. *Total Environ. Res. Themes* https://doi.org/10.1016/j.totert.2022.100021 (2022).
24. Alappuzha, K. Government of Kerala groundwater department, vol. no. May, 1–19. (2020).
25. Parthasarathy, K. S. S. & Kundapura, S. Spatio-Temporal Analysis on the Optical Properties of Vembanad Lake, Kerala, India–A Remote Sensing Approach, (2023).
26. Devi, A. B., Deka, D., Aneesh, T. D., Srinivas, R. & Nair, A. M. Predictive modelling of land use land cover dynamics for a tropical coastal urban City in Kerala, India. *Arab. J. Geosci.* **15** (5), 399 (2022).
27. Prasad, G. & Ramesh, M. V. Spatio-Temporal Analysis of Land Use/Land Cover Changes in an Ecologically Fragile Area—Alappuzha District, Southern Kerala, India. *Nat. Resour. Res.* **28**, 31–42. https://doi.org/10.1007/s11053-018-9419-y (2019).
28. Selvan, S. C., Kankara, R. S., Prabhu, K. & Rajan, B. Shoreline change along Kerala, south-west Coast of India, using geo-spatial techniques and field measurement. *Nat. Hazards*. **100** (1), 17–38 (2020).
29. Balchand, A. N. & Nambisan, P. N. K. Effect of Puip-Paper effluents on the water quality of Muvattupuzha river emptying into Cochin backwaters, (1986).

30. Beegam, S. N. & ArulRaj, P. G. Journal of critical reviews effect of population growth on land use and runoff of Muvattupuzha Sub-basin.
31. Reconaissance Survey Report Sand Auditing Of Muvattupuzha River. Ernakulam District, 2019. [Online]. Available: www.ties.org.in
32. Hasan, M., Haque, R. & Rahman, M. Case Studies in Chemical and Environmental Engineering Identifying the land use land cover (LULC) changes using remote sensing and GIS approach: A case study at Bhaluka in Mymensingh, Bangladesh. 7, (2022).
33. Njoku, E. A. & Tenenbaum, D. E. Remote Sensing Applications: Society and Environment Quantitative assessment of the relationship between land use / land cover (LULC), topographic elevation and land surface temperature (LST) in Ilorin, Nigeria. 27, (2022).
34. Townshend, J. R., Gayler, J. R., Hardy, J. R., Jackson, M. J. & Baker, J. R. Remote sensing letters: preliminary analysis of LANDSAT-4 thematic mapper products. *Int. J. Remote Sens.* **4** (4), 817–828. https://doi.org/10.1080/01431168308948606 (1983).
35. Handavu, F., Chirwa, P. W. C. & Syampungani, S. Socio-economic factors influencing land-use and land-cover changes in the Miombo woodlands of the copperbelt Province in Zambia. *Policy Econ.* **100**, 75–94. https://doi.org/10.1016/j.forpol.2018.10.010 (2019).
36. Krishnaraj, A. & Honnasiddaiah, R. Multi-spatial-scale land/use land cover influences on seasonally dominant water quality along middle Ganga basin. *Environ. Monit. Assess.* **195** (12), 1434. https://doi.org/10.1007/s10661-023-12059-y (2023).
37. Fan, F., Weng, Q. & Wang, Y. Land use and land cover change in Guangzhou, China, from 1998 to 2003, based on landsat TM / ETM+ imagery. 1323–1342, (2007).
38. Thein, A. M. & Htwe, A. N. Based on Principal Component Analysis of Land Use Land Cover Change Detection Using Landsat Satellite Images (Case study Mandalay City). *IEEE Conf. Comput. Appl. (ICCA)* https://doi.org/10.1109/ICCA51723.2023.10181968 (2023).
39. Khan, R. & Jhariya, D. C. Assessment of land-use and land-cover change and its impact on groundwater quality using remote sensing and GIS techniques in Raipur City, Chhattisgarh, India. *J. Geol. Soc. India.* **92**, 59–66 (2018).
40. Zhang, Z. et al. Impact of Land Use/Land Cover and Landscape Pattern on Water Quality in Dianchi Lake Basin, Southwest of China. *Sustain.* https://doi.org/10.3390/su15043145 (2023).
41. Islam, M. Y., Nasher, N. M. R., Karim, K. H. R. & Rashid, K. J. Quantifying forest land-use changes using remote-sensing and CA-ANN model of Madhupur Sal Forests, Bangladesh. *Heliyon* **9** (5), e15617. https://doi.org/10.1016/j.heliyon.2023.e15617 (2023).
42. Dash, P., Sanders, S. L., Parajuli, P. & Ouyang, Y. Improving the accuracy of land use and land cover classification of landsat data in an agricultural watershed. *Remote Sens. (Basel).* **15** (16). https://doi.org/10.3390/rs15164020 (Aug. 2023).
43. Potapov, P. The Global 2000–2020 Land Cover and Land Use Change Dataset Derived From the Landsat Archive: First Results. *Front. Remote Sens.* https://doi.org/10.3389/frsen.2022.856903 (2022).
44. Brontowiyono, W., Asmara, A. A., Jana, R., Yulianto, A. & Rahmawati, S. Land-Use Impact on Water Quality of the Opak Sub-Watershed, Yogyakarta, Indonesia. *Sustain.* https://doi.org/10.3390/su14074346 (2022).
45. Oliphant, A. J. et al. Mapping cropland extent of Southeast and Northeast Asia using multi-year time-series Landsat 30-m data using a random forest classifier on the Google Earth Engine Cloud. *Int. J. Appl. Earth Observation Geoinform.* **81**, 110–124 (2019).
46. Patel, S., Indraganti, M. & Jawarneh, R. N. A comprehensive systematic review: impact of land Use/ land cover (LULC) on land surface temperatures (LST) and outdoor thermal comfort. *Build. Environ.* **249**, 111130. https://doi.org/10.1016/j.buildenv.2023.111130 (2024).
47. BIS. Indian Standard Drinking Water Specification (Second Revision), *Bureau of Indian Standards*, vol. IS 10500, no. May, pp. 1–11, [Online]. (2012). Available: http://cgwb.gov.in/Documents/WQ-standards.pdf
48. Guidelines for Drinking-water Quality FIRST ADDENDUM TO THIRD EDITION. Volume 1 Recommendations WHO Library Cataloguing-in-Publication Data, (2006).
49. Srivastava, M., Srivastava, P. K., Kumar, D. & Kumar, A. A comprehensive assessment of uranium in groundwater using IDW and EWQI in the Sahibganj district of Jharkhand, India. (2024).
50. Dewana, B. R., Prasetyo, S. Y. J. & Hartomo, K. D. Comparison of IDW and Kriging Interpolation Methods Using Geoelectric Data to Determine the Depth of the Aquifer in Semarang, Indonesia. *Jurnal Ilmiah Teknik Elektro Komputer dan. Informatika* **8**(2), 215. https://doi.org/10.26555/jiteki.v8i2.23260 (2022).
51. Mondal, K. C., Rathod, K. G., Joshi, H. M. & Mandal, H. S. Impact of land-use and land-cover change on groundwater quality and quantity in the Raipur, Chhattisgarh, India: A remote sensing and GIS approach. *IOP Conf. Ser. Earth Environ. Sci.* https://doi.org/10.1088/1755-1315/597/1/012011 (2021).
52. Benaissa, C., Bouhmadi, B. & Rossi, A. An assessment of the physicochemical, bacteriological quality of groundwater and the water quality index (WQI) used GIS in Ghis Nekor, Northern Morocco. *Sci. Afr.* https://doi.org/10.1016/j.sciaf.2023.e01623 (2023).
53. Pareta, K. et al. Groundwater quality assessment for drinking and irrigation purposes in the Ayad river basin, Udaipur (India). *Groundw. Sustain. Dev.* **27**, 101351. https://doi.org/10.1016/j.gsd.2024.101351 (2024).
54. yedem Fentie, A., Mengistu, D. & Molla, G. Assessment of groundwater quality for drinking purpose using GIS based WQI methods, in Koga irrigation. *Water Sci.* **38** (1), 618–631 (2024).
55. Deshmukh, K. K. & Aher, S. P. Assessment of the impact of municipal solid waste on groundwater quality near the Sangamner City using GIS approach. *Water Resour. Manage.* **30**, 2425–2443 (2016).
56. Dhaduti, M. S., Hunashyal, A. M., Dhaduti, S. C., Jalagar, S. R. & Mathad, S. N. Assessment of groundwater quality of hubballi City, Karnataka, India by using Canadian Council of ministers of the environment water quality index, weighted arithmetic water quality index and Geospatial techniques. *J. Institution Eng. (India): Ser. A.* **105** (3), 581–587 (2024).
57. Kushe, V. P., Mishra, S. S. & Charhate, S. Assessment of ground water quality parameters during post monsoon season in three taluka of Sindhudurg district of Maharashtra using water quality index. *Sādhanā* **49** (2), 171 (2024).
58. Harman, B. I., Koseoglu, H. & Yigit, C. O. Performance evaluation of IDW, Kriging and multiquadric interpolation methods in producing noise mapping: A case study at the city of Isparta, Turkey. *Appl. Acoust.* **112**, 147–157. https://doi.org/10.1016/j.apacoust.2016.05.024 (2016).
59. Kawo, N. S. & Karuppannan, S. Groundwater quality assessment using water quality index and GIS technique in Modjo river Basin, central Ethiopia. *J. Afr. Earth Sc.* **147**, 300–311 (2018).
60. Bashir, E., Wasay, M. A., Naseem, S., Kaleem, M. & Shahab, B. Evaluation of groundwater quality from Shah Faisal Town, Karachi employing SPSS and GIS-IDW techniques. *J. Himal. Earth Sci.* **57** (2), 32–53 (2024).
61. Workneh, H. T., Chen, X., Ma, Y., Bayable, E. & Dash, A. Comparison of IDW, Kriging and orographic based linear interpolations of rainfall in six rainfall regimes of Ethiopia. *J. Hydrol. Reg. Stud.* https://doi.org/10.1016/j.ejrh.2024.101696 (2024).
62. Singh, R., Singh, A., Majumder, C. B. & Vidyarthi, A. K. Impact of pH, TDS, Chloride, and nitrate on the groundwater quality using Entropy-Weighted water quality index and statistical analysis: A case study in the districts of North India. *Water Conserv. Sci. Eng.* **9** (2), 86 (2024).
63. Kefi, M., Aden, M. M. & Ben Ali, B. Water quality monitoring for irrigation by the integration of water quality index in a geographic information system environment in Chiba watershed, Nabeul, Tunisia. *Water Conserv. Sci. Eng.* **9** (1), 20 (2024).
64. Geris, J. et al. Predicting land use and land cover changes for sustainable land management using CA-Markov modelling and GIS techniques surface water-groundwater interactions and local land use control water quality impacts of extreme rainfall and flooding in a vulnerable semi-arid region of Sub-Saharan Africa. *J. Hydrol. (Amst)* **609**, 127834. https://doi.org/10.1016/j.jhydrol.2022.127834 (2022).

65. Jothi, S. V. M. P. Detecting outliers in data streams using clustering algorithms. *Int. J. Innovative Res. Comput. Communication Eng.* **1**, 8 (2013).
66. Duraj, A. & Szczepaniak, P. S. Outlier detection in data streams—A comparative study of selected methods. *Procedia Comput. Sci.* **192**, 2769–2778 (2021).
67. Angiulli, F. & Fassetti, F. Distance-based outlier queries in data streams: the novel task and algorithms. *Data Min. Knowl. Discov.* **20** (2), 290–324 (2010).
68. Development of. A PCA-based land use/land cover classification utilizing Sentinel-2 time series. *Middle East. J. Agric. Res.* https://doi.org/10.36632/mejar/2022.11.2.42 (2022).
69. Pandey, H. K., Singh, V. K., Srivastava, S. K. & Singh, R. P. Groundwater quality assessment using PCA and water quality index (WQI) in a drought-prone area. *Sustain. Water Resour. Manag.* **9** (6), 197 (2023).
70. Torres-Martínez, J. A., Mahlknecht, J., Kumar, M., Loge, F. J. & Kaown, D. Advancing groundwater quality predictions: machine learning challenges and solutions. *Sci. Total Environ.* **949**, 174973. https://doi.org/10.1016/j.scitotenv.2024.174973 (2024).
71. El-Rawy, M. et al. An Integrated GIS and Machine-Learning Technique for Groundwater Quality Assessment and Prediction in Southern Saudi Arabia. *Water* https://doi.org/10.3390/w15132448 (2023).
72. Sahour, S. et al. Evaluation of machine learning algorithms for groundwater quality modeling. *Environ. Sci. Pollut. Res.* **30** (16), 46004–46021 (2023).
73. Siqi, W., Qiang, L., Xifeng, G., En, Z. & Jianping, Y. Fast and unsupervised outlier removal by recurrent adaptive reconstruction extreme learning machine. *Int. J. Mach. Learn. Cybernet.* **10** (12), 3539–3556 (2019).
74. Haggerty, R., Sun, J., Yu, H. & Li, Y. Application of machine learning in groundwater quality modeling-A comprehensive review. *Water Res.* **233**, 119745 (2023).
75. Yu, T. K. et al. Predicting potential soil and groundwater contamination risks from gas stations using three machine learning models (XGBoost, LightGBM, and),*Process Saf. Environ. Prot.*, **199**, 107249, doi: https://doi.org/10.1016/j.psep.2025.107249. (2025).
76. Weith, T. et al. Human-Environment Interactions 8 Sustainable Land Management in a European Context.http://www.springer.com/series/8599
77. Connor, R. The United Nations world water development report 2015: water for a sustainable world. in World Water assessment Programme. *U.N. Educ. Sci. Cultural Organization*. https://books.google.co.in/books?id=zQV1CQAAQBAJ 2015. https://books.google.co.in/books?id=zQV1CQAAQBAJ
78. Hua, Y., Yan, D. & Liu, X. Environmental and Sustainability Indicators Assessing synergies and trade-offs between ecosystem services in highly urbanized area under different scenarios of future land use change. *Environmental and Sustainability Indicators.* **22**, 100350. https://doi.org/10.1016/j.indic.2024.100350 (2024).
79. Tahir, Z. et al. Predicting land use and land cover changes for sustainable land management using CA-Markov modelling and GIS techniques. *Sci. Rep.* https://doi.org/10.1038/s41598-025-87796-w (2025).
80. Ariti, A. T., van Vliet, J. & Verburg, P. H. Land-use and land-cover changes in the central rift Valley of ethiopia: assessment of perception and adaptation of stakeholders. *Appl. Geogr.* **65**, 28–37. https://doi.org/10.1016/j.apgeog.2015.10.002 (2015).
81. Kumar, V. & Agrawal, S. Urban modelling and forecasting of landuse using SLEUTH model. *Int. J. Environ. Sci. Technol.* https://doi.org/10.1007/s13762-022-04331-4 (2023).
82. Nery, T. et al. Comparing supervised algorithms in Land Use and Land Cover classification of a Landsat time-series. *IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, IEEE 5165–5168. (2016).
83. Priyadarisini, D. & Umadevi, G. A System Dynamics Model for Assessing Land-Use Transport Interaction Scenarios in Chennai, India. *Sustain.* https://doi.org/10.3390/su15076297 (2023).
84. Miao, Z., Brusseau, M. L., Carroll, K. C., Carreón-Diazconti, C. & Johnson, B. Sulfate reduction in groundwater: characterization and applications for remediation. *Environ. Geochem. Health.* **34** (4), 539–550. https://doi.org/10.1007/s10653-011-9423-1 (2012).
85. Das, B. & Pal, S. C. Assessment of groundwater recharge and its potential zone identification in groundwater-stressed Goghat-I block of Hugli District, West Bengal, India. *Environ. Dev. Sustain.* **22** (6), 5905–5923. https://doi.org/10.1007/s10668-019-00457-7 (2020).
86. Lei, X. et al. Coupling coordination analysis of urbanization and ecological environment in Chengdu-Chongqing urban agglomeration. *Ecol. Ind.* **161** (December 2023), 111969. https://doi.org/10.1016/j.ecolind.2024.111969 (2024).
87. Bao, C. & He, D. Scenario modeling of urbanization development and water scarcity based on system dynamics: A case study of Beijing–Tianjin–Hebei urban agglomeration, China. *Int. J. Environ. Res. Public. Health* https://doi.org/10.3390/ijerph16203834 (2019).
88. Ozhukayil, J., Sebastian, L. & Chandramohanakumar, N. Comparative study on dissolved trace metal concentrations of iron and manganese in Muvattupuzha river. *Int. J. Math. Trends Technol-IJMTT*, **38**, (2016).
89. Shaina Beegam, N., ArulRaj, G. P. & Brema, J. An examination of land use and land cover changes in muvattupuzha river basin using GIS. *Int. J. Recent Technol. Eng.* **8**, 7957–7960 (2019).
90. 44_Kerala MPR June 2020.
91. *muvattupuzha area ssurrounding industries.*
92. Wang, H., He, Q., Liu, X., Zhuang, Y. & Hong, S. Landscape and urban planning global urbanization research from 1991 to 2009: A systematic research review. *Landsc. Urban Plan.* **104**, 3–4. https://doi.org/10.1016/j.landurbplan.2011.11.006 (2012).
93. Keerthi Naidu, B. N. & Chundeli, F. A. Assessing LULC changes and LST through NDVI and NDBI Spatial indicators: a case of Bengaluru, India. *GeoJournal* **88** (4), 4335–4350. https://doi.org/10.1007/s10708-023-10862-1 (2023).
94. Abbas, Z., Yang, G., Zhong, Y. & Zhao, Y. Spatiotemporal change analysis and future scenario of lulc using the CA-ANN approach: A case study of the greater bay area, China. *Land. (Basel)* https://doi.org/10.3390/land10060584 (2021).
95. Patel, A. et al. Results in Engineering Novel approach for the LULC change detection using GIS & Google Earth Engine through spatiotemporal analysis to evaluate the urbanization growth of Ahmedabad city. 21 (2023).
96. Vijay, A. & Varija, K. Spatio-temporal classification of land use and land cover and its changes in Kerala using remote sensing and machine learning approach. *Environ. Monit. Assess.* **196** (5), 459 (2024).
97. Akhila, R. & Pramada, S. K. Land use land cover change detection and prediction using land change modeler–A case study of kerala: R Akhila and SK Pramada. *J. Earth Syst. Sci.* **134** (3), 138 (2025).
98. Prakasam, C. & R, A. Impact of changing urban landscapes on forest degradation: A study on a part of Western Ghats, India. *Environ. Monit. Assess.* **196** (3), 256 (2024).
99. Chinnasamy, P. & Honap, V. U. Spatiotemporal variations in soil loss across the biodiversity hotspots of Western Ghats Region, India. *J. Earth Syst. Sci.* **132** (2), 90 (2023).
100. Drissia, T. K., Sreya, P., Eldho, T. I. & Dinesan, V. P. Impact of land use/land cover and climate change on streamflow variations in heterogeneous river basins of south-western India. *Int. J. River Basin Manag.* 1–18, (2024).
101. Nawab, N. P. S., Nimitha, M., Muthukumar, A. & Muthuchamy, M. Remote sensing and GIS for monitoring and assessing forest susceptibility to climate change: A spatio-temporal study on protected area of Western ghats, India. *J. Sci. Res.* **66** (4), 7–14 (2022).
102. Arumugam, T., Kinattinkara, S., Velusamy, S., Shanmugamoorthy, M. & Murugan, S. GIS based landslide susceptibility mapping and assessment using weighted overlay method in wayanad: A part of Western Ghats, Kerala. *Urban Clim.* **49**, 101508 (2023).
103. Natarajan, S. Flood Inundation Mapping by Multi-criteria Decision Analysis—A Study on Recent Floods—Idukki District, Kerala-India. *Int. Conf. Adv. Mater. Modeling Analysis Sustain. Resilient Infrastruct.* 191–201. (2025).

104. Khan, H. H., Khan, A., Ahmed, S. & Perrin, J. GIS-based impact assessment of land-use changes on groundwater quality: study from a rapidly urbanizing region of South India. *Environ. Earth Sci.* **63**, 1289–1302 (2011).
105. Acharya, S., Hori, T. & Karki, S. Assessing the spatio-temporal impact of landuse landcover change on water yield dynamics of rapidly urbanizing Kathmandu Valley watershed of Nepal. *J. Hydrol. Reg. Stud.* **50**, 101562. https://doi.org/10.1016/j.ejrh.2023.101562 (2023).
106. de Gomes, K. M., Saad, S. I. & Mota da Silva, J. Hydrological implications of agricultural expansion on natural and degraded lands in Northeastern Brazil. *J. South. Am. Earth Sci.* **167**, 105785. https://doi.org/10.1016/j.jsames.2025.105785 (2025).
107. Nambiar, S. R., Satheendran, S. S. & Dhanya, R. Land Use/Land Cover Changes of Kannur District, Kerala From 1969 to 2024: A Geospatial Investigation. *Int. Conf. Adv. Mater. Modeling Analysis Sustain. Resilient Infrastruct*. 133–143. (2025).
108. Varunprasath, K., Islam, M. N. & Amritha, P. S. Land use and land cover analysis in the Alappuzha District, South Kerala, India, in India III: Climate Change and Landscape Issues in India: A Cross-Disciplinary Framework, Springer, 291–307. (2025).
109. Li, L., Huang, X. & Yang, H. Scenario-based urban growth simulation by incorporating ecological-agricultural-urban suitability into a Future Land Use Simulation model. *Cities* https://doi.org/10.1016/j.cities.2023.104334 (2023).
110. Prayag, A. G., Zhou, Y., Srinivasan, V., Stigter, T. & Verzijl, A. Assessing the impact of groundwater abstractions on aquifer depletion in the cauvery Delta, India. *Agric. Water Manag.* **279**, 108191. https://doi.org/10.1016/j.agwat.2023.108191 (2023).
111. Anjaly, C. S., Sathian, K. K., Anu, V. & Jinu, A. Assessment and mapping of water quality of a shallow aquifer near an industrial belt using Hydro-chemical parameters and irrigation water quality index. *J. Agricultural Eng. (India).* **60** (1), 60–73. https://doi.org/10.52151/jae2023601.1797 (2023).
112. Aju, C. D., Achu, A. L., Prakash, P., Raicy, M. C. & Reghunath, R. An integrated statistical-geospatial approach for the delineation of flood-vulnerable sub-basins and identification of suitable areas for flood shelters in a tropical river basin, Kerala. *Geosyst. Geoenvironment.* **3** (2), 100251. https://doi.org/10.1016/j.geogeo.2024.100251 (2024).
113. Satterthwaite, D., Mcgranahan, G. & Tacoli, C. Urbanization and its implications for food and farming,. 2809–2820. https://doi.org/10.1098/rstb.2010.0136 2010. https://doi.org/10.1098/rstb.2010.0136
114. Prasood, S. P., Mukesh, M. V., Rani, V. R., Sajinkumar, K. S. & Thrivikramji, K. P. Urbanization and its effects on water resources: Scenario of a tropical river basin in South India. *Remote Sens Appl* https://doi.org/10.1016/j.rsase.2021.100556 (2021).
115. Shimod, K. P., Vineethkumar, V., Prasad, T. K. & Jayapal, G. Effect of urbanization on heavy metal contamination: a study on major townships of Kannur District in Kerala, India. *Bull. Natl. Res. Cent.* https://doi.org/10.1186/s42269-021-00691-y (2022).
116. Devi, A. & Nair, A. Effects of urbanization in a shallow coastal aquifer: an integrated GIS-based case study in Cochin, India. *Groundw. Sustain. Dev.* **15**, 100656. https://doi.org/10.1016/j.gsd.2021.100656 (Aug. 2021).
117. Salim, M. Z. et al. A comprehensive review of navigating urbanization induced climate change complexities for sustainable groundwater resources management in the Indian Subcontinent. *Groundw. Sustainable Dev.* **25**, 101115. https://doi.org/10.1016/j.gsd.2024.101115 (2024).
118. Fan, C. & Wang, Z. Spatiotemporal characterization of land cover impacts on urban warming: A Spatial autocorrelation approach. *Remote Sens. (Basel).* **12** (10), 1–17. https://doi.org/10.3390/rs12101631 (2020).
119. Digra, M., Dhir, R. & Sharma, N. Land use land cover classification of remote sensing images based on the deep learning approaches: a statistical analysis and review. *Arab. J. Geosci.* https://doi.org/10.1007/s12517-022-10246-8 (2022).
120. Kumar, N. V., Mathew, S. & Swaminathan, G. Analysis of groundwater for potability from Tiruchirappalli City using backpropagation ANN model and GIS. *J. Environ. Prot. (Irvine Calif).* **01** (02), 136–142. https://doi.org/10.4236/jep.2010.12018 (2010).
121. Tran, D. D. et al. Environmental pressures on livelihood transformation in the Vietnamese Mekong delta: implications and adaptive pathways. *J. Environ. Manage.* **377**, 124597. https://doi.org/10.1016/j.jenvman.2025.124597 (2025).
122. Tian, J. et al. Spatiotemporal monitoring of water storage in the North China plain from 2002 to 2022 based on an improved GRACE downscaling method. *J. Hydrology: Reg. Stud.* **59**, 102370. https://doi.org/10.1016/j.ejrh.2025.102370 (2025).
123. Sanad, H. et al. Ecological and Health Risk Assessment of Heavy Metals in Groundwater within an Agricultural Ecosystem Using GIS and Multivariate Statistical Analysis (MSA): A Case Study of the Mnasra Region, Gharb Plain, Morocco. *Water.* https://doi.org/10.3390/w16172417 (2024).
124. Saranya, T. & Saravanan, S. A comparative analysis on groundwater vulnerability models—fuzzy DRASTIC and fuzzy DRASTIC-L. *Environ. Sci. Pollut. Res.* **29**, 86005–86019. https://doi.org/10.1007/s11356-021-16195-1 (2022).
125. Subbarayan, S., Thiyagarajan, S., Karuppanan, S. & Panneerselvam, B. Enhancing groundwater vulnerability assessment: comparative study of three machine learning models and five classification schemes for Cuddalore district. *Environ. Res.* https://doi.org/10.1016/j.envres.2023.117769 (2023).
126. Fallahzadeh, R. A. et al. Spatial distribution and health risk assessment of nitrate in drinking water: A case study in the central plateau of Iran. J. Environ. Health Sustain. Dev. (2024).
127. Ayadi, Y. et al. Groundwater potential recharge assessment in Southern mediterranean basin using GIS and remote sensing tools: case of Khalled-Teboursouk basin, karst aquifer. *Appl. Geomatics.* **16** (3), 677–693 (2024).
128. Sivakumar, V. et al. An integrated approach for an impact assessment of the tank water and groundwater quality in Coimbatore region of South India: implication from anthropogenic activities. *Environ. Monit. Assess.* **195**, 88 (2023).
129. Ngernkerd, P., Choowong, M., Choowong, N. & Surakiatchai, P. Late pleistocene climate variation on the Khorat Plateau, Northeastern Thailand inferred from the remnants of sand dunes, (2021).
130. Sheikh Sayed, R. et al. (n.d.). Technological and Sustainable Approaches to Groundwater Resource Assessment and Management in the Context of Urban Expansion: A Case Study in Keraniganj, Bangladesh.
131. Jahanshahi, A., Booij, M. J., Patil, S. D. & Gupta, H. Impact of land use land cover change on catchment hydrological response in 576 Iranian catchments. *J. Arid Environ.* **231**, 105463. https://doi.org/10.1016/j.jaridenv.2025.105463 (2025).
132. Li, X. et al. Identifying the Spatial pattern and driving factors of nitrate in groundwater using a novel framework of interpretable stacking ensemble learning. *Environ. Geochem. Health.* https://doi.org/10.1007/s10653-024-02201-1 (2024).
133. Ransom, K. M., Nolan, B. T., Stackelberg, P. E., Belitz, K. & Fram, M. S. Machine learning predictions of nitrate in groundwater used for drinking supply in the conterminous United States. *Sci. Total Environ.* https://doi.org/10.1016/j.scitotenv.2021.151065 (2022).
134. Li, Z. Extracting Spatial effects from machine learning model using local interpretation method: an example of SHAP and XGBoost. *Comput. Environ. Urban Syst.* **96**, 101845 (2022).
135. Wang, H., Liang, Q., Hancock, J. T. & Khoshgoftaar, T. M. Feature selection strategies: a comparative analysis of SHAP-value and importance-based methods. *J. Big Data.* **11** (1), 44 (2024).
136. Rohde, M. M. et al. A machine learning approach to predict groundwater levels in California reveals ecosystems at risk. *Front. Earth Sci. (Lausanne).* **9**, 784499 (2021).
137. Aher, S., Deshmukh, K., Gawali, P., Zolekar, R. & Deshmukh, P. Hydrogeochemical characteristics and groundwater quality investigation along the basinal cross-section of Pravara River, Maharashtra, India. *J. Asian Earth Sciences: X.* **7**, 100082. https://doi.org/10.1016/j.jaesx.2022.100082 (2022).
138. Yang, S. et al. Spatial mapping and prediction of groundwater quality using ensemble learning models and Shapley additive explanations with Spatial uncertainty analysis. *Water* **16** (17), 2375. https://doi.org/10.3390/w16172375 (2024).
139. Bewket, W. & Sterk, G. Farmers' participation in soil and water conservation activities in the Chemoga Watershed, Blue Nile Basin. *Ethiopia. Land Degrad. Dev.* https://doi.org/10.1002/ldr.492 (2002).

140. Prajapati, G. S., Rai, P. K., Mishra, V. N., Singh, P. & Shahi, A. P. Remote sensing-based assessment of waterlogging and soil salinity: A case study from Kerala, India. *Results Geophys. Sci.* **7**, 100024. https://doi.org/10.1016/j.ringps.2021.100024 (2021). February.
141. Ayalew, S. E., Niguse, T. A. & Aragaw, H. M. Hydrological responses to historical and predicted land use/land cover changes in the Welmel watershed, Genale Dawa Basin, ethiopia: implications for water resource management. *J. Hydrol. Reg. Stud.* **52**, 101709. https://doi.org/10.1016/j.ejrh.2024.101709 (2024).
142. Chen, D., Elhadj, A., Xu, H., Xu, X. & Qiao, Z. A study on the relationship between land use change and water quality of the Mitidja watershed in Algeria based on GIS and RS. *Sustain.* https://doi.org/10.3390/SU12093510 (2020).
143. Halder, S. et al. Understanding hydrological responses through LULC analysis and predictive modelling (MLPNN-MC Model): A study of Bandu Sub-watershed (India) over three decades. *Artif. Intell. Geosci.* **6** (2), 100152. https://doi.org/10.1016/j.aiig.2025.100152 (2025).
144. Census of India. Census of India 2011 District Census Handbook Ernakulam, p. 516, (2011).
145. Dash, S. S. & Maity, R. Effect of climate change on soil erosion indicates a dominance of rainfall over LULC changes. *J. Hydrol. Reg. Stud.* **47**, 101373. https://doi.org/10.1016/j.ejrh.2023.101373 (2023).
146. Liu, S., Wang, L. & Guo, C. Heavy metal pollution and ecological risk assessment in brownfield soil from Xi'an, China: An integrated analysis of man land interrelations. *PLoS One.* **15**, e024139 (2020).
147. Hao, Z. et al. 2025. Dynamic soil erosion in response to LULC changes in mountainous areas of southwest China over the last 40 years: A case study of the Erhai Basin in Yunnan province, Environmental and Sustainability Indicators. 27: 100755. https://doi.org/10.1016/j.indic.2025.100755 2025 https://doi.org/10.1016/j.indic.2025.100755
148. Mechal, A., Fekadu, D. & Abadi, B. Multivariate and water quality index approaches for Spatial water quality assessment in lake Ziway, Ethiopian rift. *Water Air Soil. Pollut.* **235** (1), 78. https://doi.org/10.1007/s11270-023-06882-9 (2024).
149. Santos, R. S. S. et al. Groundwater contamination in a rural municipality of Northeastern brazil: application of Geostatistics, Geoprocessing, and geochemistry techniques. *Water Air Soil. Pollut.* **235** (3), 179 (2024).
150. The, S., Journal, G. & Mar, N. Interdependent urbanization in an urban world: an historical overview, **164**: 85–95, (2015).
151. Lekshmi, A. & Lancelet, P. T. Trend of urbanisation in Ernakulam with respect to Kerala. *J. Global Resour.* **5** (02), 41–48 (2019).
152. Liu, W., Zhang, L., Hu, X., Meng, Q. & Qian, J. International Journal of Applied Earth Observation and Geoinformation Nonlinear effects of urban multidimensional characteristics on daytime and nighttime land surface temperature in highly urbanized regions: A case study in Beijing China.. *Int. J. Appl. Earth Observation Inf.* **132**, 1–12 (2024).

Acknowledgements

The authors would like to acknowledge Karunya Institute of Technology and Sciences for providing the required facilities and logistical support during this research. We are very grateful to the anonymous reviewers for their comments and time on our paper.

Author contributions

A.K. designed the study, carried out the data collection, performed LULC classification, groundwater quality analysis, and prepared the initial draft of the manuscript.S.G. supervised the research, contributed to the conceptualization, methodology development, geospatial and hydrochemical interpretation, and thoroughly revised and edited the manuscript.G.P.A. assisted in data interpretation, CA–Markov modelling, and contributed to the refinement of the results and discussion.S.K.J. supported the statistical analyses, machine-learning modelling, and validation procedures.C.H.H. contributed to the methodological framework, interpretation of findings, and critical revision of the manuscript for intellectual content.All authors reviewed the manuscript and approved the final version.

Funding

The work of C-HH was supported by the National Research Foundation of Korea (NRF) funded by the Korean Government (MSIT; RS-2025-00555756) and the Ministry of Education (RS-2018-NR031078).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-026-38961-2.

Correspondence and requests for materials should be addressed to S.G. or C.-H.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2026