

HYDRO-GEOTECHNICAL MAPPING OF MANAGED AQUIFER RECHARGE SITES: INTEGRATING GIS AND MACHINE LEARNING FOR ENHANCED GROUNDWATER REMEDIATION STRATEGIES

Usama Sher¹, Hassan Ali Sadiq², Md Abu Syed³, Nauman Gul^{*4}, Uzair Ali⁵,
Muhammad Awais Khan⁶, Sibga Tullah Tahir⁷

^{1,*4}University of Engineering and Technology (UET), Peshawar, Pakistan

²Civil Engineer, Pini Arabia for Engineering Consultancy, Riyadh, Saudi Arabia

³MSc Scholar, Department of Civil Engineering, Chongqing University, Chongqing, China

⁵Local Government and Rural Development Department, Khyber Pakhtunkhwa, Pakistan

⁶Comsats University Islamabad, Abbottabad Campus, Pakistan

⁷The University of Law, Manchester, United Kingdom (UK)

¹usamasher5400@gmail.com, ²hassanalisadiq@gmail.com, ³12300072@stu.cqu.edu.cn,

⁴naumangul7@gmail.com, ⁵engruzair91@gmail.com, ⁶enr.m.awaishkhan@gmail.com,

⁷sibgattahir@gmail.com

DOI: <https://doi.org/10.5281/zenodo.22909428> 

Keywords

MAR, GIS, Machine Learning, Groundwater Remediation, Hydrogeotechnical Mapping Site Suitability.

Article History

Received: 14 June 2026

Accepted: 05 September 2026

Published: 23 September 2026

Copyright @Author

Corresponding Author: *
Nauman Gul

Abstract

Recharging depleted aquifers, reducing degradation, and increasing storage are all possible outcomes of MAR. There are a number of factors that determine the efficiency of MAR, including site selection, hydrogeological conditions, soil and subsurface features, source water supply, contaminant risk, land use, and the hydraulic response of the receiving aquifer. It is possible that regional site selection approaches that make use of a single hydrogeological parameter or multi-criteria decision analysis based on expert opinion could be effective; however, these methods are unable to take into account the nonlinear interactions that occur between environmental and hydro-geotechnical elements. An integrated hydro-geotechnical mapping framework combines geographic information systems (GIS), remote sensing, hydrogeological characterisation, geotechnical research, machine learning (ML), and groundwater-quality assessment in order to identify, rank, and validate MAR sites. It is proposed that lithology, soil texture, hydraulic conductivity, infiltration capacity, aquifer thickness, depth to groundwater, recharge potential, slope, drainage density, lineament density, land use/cover, electrical conductivity, ground water quality, contamination vulnerability, source water quality, and infrastructure proximity all be taken into consideration. GIS has the ability to provide spatial information structure, harmonise data, spatially analyse data, map constraints, and visualise data. On the other hand, machine learning models such as Random Forest, Support Vector Machine, Gradient Boosting, Extreme Gradient Boosting, and Artificial Neural Networks have the capability to model non-linear relationships between environmental predictors and groundwater-response

indicators. A recent study has demonstrated that machine learning has the potential to enhance groundwater mapping prediction provided it is taught and evaluated independently. In addition to the accuracy of the models, the technique that has been suggested takes into account the quality of the data, cross-validation, uncertainty analysis, explainable machine learning, and field verification in order to take into account the hydrogeological plausibility and the uncertainty of the models. Following a transparent and open preliminary screening procedure, a decision framework that is based on GPS makes use of machine learning in order to significantly enhance spatial predictions. Additionally, the model takes into account the attenuation of groundwater contaminants, the compatibility of source water, the duration of transit, the storage of aquifers, the risk of clogging, and the efficiency of recovery. In order to select MAR sites, develop experimental recharge methods, and devise an adaptive groundwater management strategy, water resource planners can make use of hydro-geotechnical suitability maps. Those metropolitan areas that are experiencing groundwater contamination and depletion can benefit from the technique described above.

1. Introduction

Groundwater is one of the key elements of freshwater systems globally and is an essential water source for drinking and irrigation, industry and for supporting ecosystems in many areas. Its significance expands further in arid and semi-arid areas where surface-water supplies are significantly more seasonal than they are in other areas, and rates of demand for surface-water are outpacing decreases in precipitation (Glanville et al., 2023). Population development, urbanisation, agricultural and industrial activities, and climate variability have boosted groundwater withdrawal from many aquifer systems. Due to excessive groundwater extraction above aquifer recharge rates, well efficiency drops, pumping costs rise, and surface water systems related to the groundwater system are affected. Overextraction of coastal aquifers causes saltwater intrusion. MAR can recharge groundwater using a technique. MAR involves purposefully depositing water in the ground to retrieve it later, store it for use, improve the environment, or manage groundwater resources. Man-made and natural recharge facilities include infiltration basins, ponds, wells, trenches, galleries, riverbank filtration, check dams, etc. With proper water quality controls, the system can be used with surface water, stormwater, treated wastewater, and

seasonal flows. Thus, MAR is necessary for surface-water, groundwater, wastewater, flood, and climate adaptation coordination (McCurry & Pyne, 2022). Although it may have the potential to be implemented, MAR cannot be successfully implemented at every location (Zurqoni et al., 2020). While the surface characteristics of a site may seem to be favorable, the subsurface may be unfavorable, the hydraulic conductivity may be low, the aquifer thickness may be small, the water table may be shallow, the source water may be poor, the site may be highly vulnerable to contamination, or the storage capacity may be limited. Likewise, a seemingly permeable site may experience clogging, excessive mounding, inadequate recovery efficiency or undesirable groundwater-quality reactions (Stefan & Ansems, 2018). Therefore, the multiple conditions of the surface, subsurface, hydrogeological, geotechnical, environmental, and socioeconomic must be understood in order to select a site for the MAR. Multi-criteria decision analysis (MCDA) is a widely used method for mapping MAR suitability that is performed using a GIS. Alishahi Tabriz et al. (2019) identified a wide range of factors which were used in 63 GIS-MCDA studies and found that there was a great variability in the ways in which weights were assigned, thus reflecting the usefulness and methodological inconsistencies of

traditional suitability mapping. GIS can be especially useful because it allows for the standardization, integration, analysis and creation of suitability maps from heterogeneous spatial datasets. In the past, other MAR studies have mapped the potential sites based on the slope, soil, drainage density, infiltration rate, lithology, groundwater depth, land use, and electrical conductivity. (Stefan & Ansems, 2018).

But, in most cases, GIS-MCDA is based on pre-defined weights and relationships. These methods can present complex interactions in a simplified manner. For instance, the impact of soil texture on recharge could be influenced by slope, land cover, antecedent moisture, aquifer depth, rainfall intensity and sub surface stratigraphy. Groundwater recharge also can be influenced by lineament density, dependent on lithology and structural connectivity. These nonlinear interactions provide an opportunity for machine learning (Elghazouly et al., 2024) .

Machine learning has a data-driven approach that can learn relationships between several predictors and the groundwater or recharge response, as shown by Alishahi Tabriz et al. (2019) . Several algorithms have increasingly been utilized for groundwater potential, groundwater recharge-potential, and groundwater-quality mapping, such as: Random Forest, Support Vector Machine, Artificial Neural Networks, Gradient Boosting and others. For instance, recharge-potential mapping studies have shown that models using machine learning algorithms can capture complex patterns and relationships among environmental variables, even if they are nonlinear (Moumane et al., 2025). Directly integrating ensemble ML into suitability assessment of MAR has been a recent development. A 2026 study in the upstream Awash River basin found that an ensemble model (combination of Random Forest, Support Vector Machine and XGBoost) outperforms significantly an AHP-based Groundwater Potential Index (GWPI), highlighting the potential of integrating expert-based GIS with data-driven learning (EL-Omairi & El Garouani, 2023).

The overall goal of this article is to create a comprehensive conceptual framework for hydro-

geotechnical mapping, based on GIS and machine learning, of MAR sites. The framework aims to not only generate a suitability map, but to combine physical suitability, recharge performance, groundwater quality, contamination risk, and remediation potential. We highlight the significance of hydro-geotechnical elements, validate our models, analyses uncertainty, and verify our findings in the field.

2. Managed Aquifer Recharge and Groundwater Remediation

2.1 Concept of Managed Aquifer Recharge

MAR is a different type of natural recharge since it requires intentional manipulation of the hydrological system (El Moneam, 2023). The basic goal is to add water to an aquifer such that beneficial storage or improvement of the groundwater resource is accomplished. Depending on the project objective, MAR can: increase groundwater storage, provide drought reserves, reduce pumping stress, improve groundwater quality, prevent seawater intrusion, reduce floodwater impacts etc. However, from the global synthesis of MAR initiatives by UNESCO, it is clear that managed recharge can contribute to higher groundwater storage, better water security, improved environmental flows, avoidance of seawater intrusion and natural attenuation of some contaminants (Chekirbane et al., 2025). Nevertheless, MAR cannot be seen as just a way to place water in the subsurface. Therefore, the two aspects of quantity and quality need to be taken into account in a successful MAR program. The U.S. Environmental Protection Agency recognizes that the improved recharging of an aquifer needs to be assessed for aquifer storage potential, transmissivity, infiltration characteristics, recoverability, water quality, water availability, and future demand. These factors are directly applicable to the hydro-geotechnical mapping (Hartog & Stuyfzand, 2017) .

2.2 MAR as a Groundwater Remediation Strategy

The technologies traditionally used for groundwater remediation include pump-and-treat

systems, permeable reactive barriers, monitored natural attenuation, in-situ chemical treatment and bioremediation (Mystrioti & Papassiopi, 2024). When well designed, MAR can supplement these strategies. Groundwater recharge can affect groundwater remediation by:

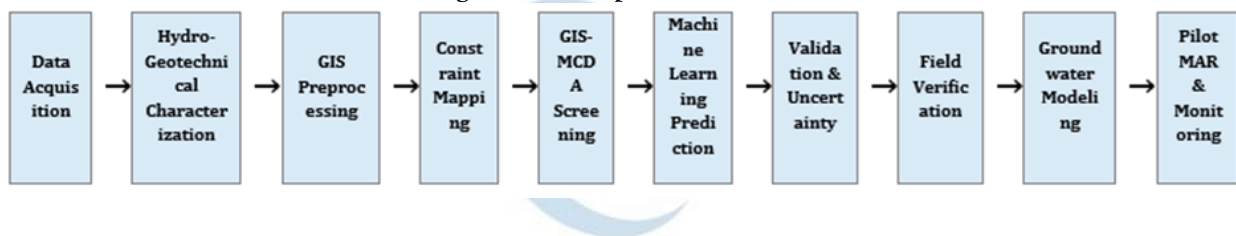
- Reducing the concentration of some dissolved pollutants,
- improved drainage of polluted areas; and
- variations in flow patterns; and
- natural remediation; and
- participation in biogeochemical reactions; and
- hydraulic containment;

Thus, recharge suitability should incorporate a source-water quality component in addition to the physical recharge potential. EPA is particularly explicit that recharge-water quality and contaminant effects should be taken into account during assessing enhanced aquifer recharge (E.A.R.) (Meles et al., 2024).

2. 3 Hydro-Geotechnical Controls on MAR Suitability

The proposed framework includes both hydrogeological and geotechnical parameters since the engineering properties of the near surface environment are determined to control recharge performance.

Figure 1 Conceptual Framework



3.1 Lithology and Geological Structure

Groundwater movement, storage capacity, permeability and porosity are controlled by lithology. Clayey formations may have limited vertical infiltration and sand and gravel provide good hydraulic conductivity. Fractured rock aquifers may offer good recharge potential when fractures are interconnected (Wendt et al., 2021). Geological features like faults, fractures and lineaments may significantly affect groundwater flow. Lineament density obtained from remote sensing can thus be used as a good spatial predictor. Note, however, that lineament density is not necessarily a positive recharge factor as some faults may be barriers to recharge. As the number of structural controls that are used increases in regional MAR mapping, lithology, distance to faults, and fault density were among the main suitability criteria for the Hari River Basin (Halim et al., 2025).

The textural class is related to hydraulic conductivity. Hydraulic conductivity is related to the textural class of the soil.

Infiltration is highly influenced by soil texture. Coarse sandy soils can typically allow more infiltration than clayey soils, however high permeable surfaces may provide limited opportunities for contaminant attenuation (Pauwels et al., 2023). Clay layers can impede infiltration, but can also offer some natural protection from the potential for rapid contaminant migration. Also, hydraulic conductivity is very significant because it indicates the speed at which water passes through the porous medium. Insufficient conductivity can cause surface ponding, high mounding or low recharge rates on a MAR site as described by Guo et al. (2022). Pumping tests, slug tests, laboratory permeability tests, grain-size relationships, geophysical interpretation of data, borehole information; and groundwater models can be used to obtain hydraulic conductivity. Therefore, the hydraulic conductivity may differ by several orders of magnitude, and should ideally be expressed at a more localized scale than the regional scale of the aquifer.

3.2 Infiltration Capacity

The infiltration capacity can be used to describe the rate of recharge-water entry into the vadose zone. Field infiltration tests can be direct measurements and can be interpolated in a GIS.

The infiltration rate is used in conjunction with: soil texture, initial soil moisture, soil compaction, land cover, sediment concentration, groundwater depth, vadose-zone thickness, and recharge-water quality. Furthermore, high infiltration capacity does not necessarily mean high MAR suitability. However, contaminants may reach the aquifer before they are sufficiently reduced before reaching the groundwater (Yumna et al., 2024) if infiltration is too fast and the zone of interest is thin.

3.3 Depth to Groundwater

Another important parameter is the depth to groundwater. Additional natural attenuation and storage can be provided by a sufficiently thick unsaturated zone. The shallow groundwater table could cause limited groundwater storage and groundwater mounding (Liu & Pan, 2024). Groundwater-depth observations from monitoring wells, borehole data, geostatistical interpolation, and groundwater models can be used to generate the spatial distribution of groundwater depth. The available unsaturated

storage should be accounted for as a function of season as it varies with the year.

The thickness of the aquifer and the amount of storage available in the aquifer. Another important factor to consider when evaluating a site's potential for groundwater storage is the aquifer's thickness. A highly permeable, extremely thin aquifer might not be the best strategic storage aquifer because it could store less groundwater than a thicker, moderately permeable aquifer. Aquifer extent, groundwater fluctuation, transmissivity, specific yield, storage coefficient, and saturation thickness are among the essential indicators. To simplify the expression, potential groundwater storage can be written as $V_s = A \times b \times S_y$, where A is the recharge area, b is the saturated thickness of the aquifer, and S_y is the specific yield. Reasons why the practical or recoverable storage might not live up to the theoretical potential include factors like operational constraints, groundwater-flow conditions, pumping, recharge efficiency, hydraulic connection, and regional hydrogeological heterogeneity (Raji & Abdulkadir, 2022).

4. GIS-Based Hydro-Geotechnical Mapping

GIS can give the spatial context in which various data sets can be combined. At least the following layers should be part of a conceptual GIS database for MAR site assessment:

Table 1 MAR (Layers)

Category	Variables
Topography	Elevation, slope, curvature
Soil	permeability, hydraulic conductivity
Geology	Lithology, geological formations, faults, fractures
Hydrology	Drainage density, stream proximity, catchment characteristics
Hydrogeology	Groundwater depth, aquifer thickness, transmissivity, storage

First each layer needs to be converted into a single spatial reference system and spatial resolution. This is because the input variables are in different units and ranges (Zhang et al., 2019). Slope, for instance, can be given in degrees; hydraulic conductivity in m/day; electrical conductivity in $\mu\text{S}/\text{cm}$ and groundwater depth in m. The

following parameters must be scaled to the same order of magnitude for a suitability model.

Each criterion can be placed on the same scale from 0 to 1 and then a conventional standardized suitability score can be calculated. If the factors are correlated and the higher the score the more suitable, then the standardized score will be equal

to $S_i = X_i - X_{\min} / X_{\max} - X_{\min}$. The standardized scores are then combined together using a Weighted Linear Combination (WLC) approach to calculate the overall MAR Suitability Index (MARS_I), $MARS_I = \sum w_i S_i$, where MARS_I is the overall suitability index, w_i is the weight assigned to criterion i , S_i is the standardized suitability score of criteria i , and n is the number of criteria. The weight should be assigned in such a way that their sum is one ($\sum w = 1$), meaning that the final MARS_I will be on the same scale of suitability (Ferrer-Ferrando et al., 2025). The GIS-MCDA approach has been employed extensively in the selection of MAR's sites, with variations in criteria and weighting techniques found in the literature (Imig et al., 2022).

5. Restrictions of traditional GIS-MCDA

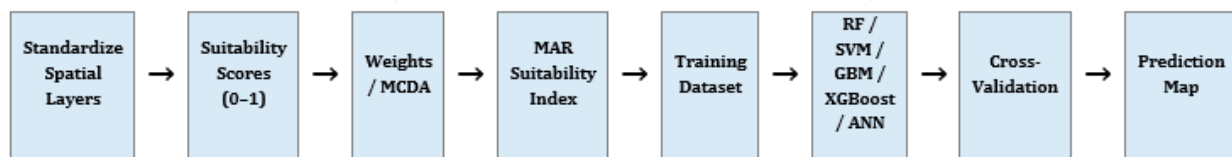
Although GIS-MCDA is relatively simple to implement, transparent, and accessible, there are some drawbacks. First, weights are often determined by experts. While it is important to have expert knowledge, experts may give different importance to the same parameter. Second, the assumption of a relationship between a parameter and recharge suitability is often linear or monotonic. This is a simplified form of hydrogeological system that is rarely seen in nature. Third, there can be double counting of variables associated with correlations. Slope and

drainage density, for instance, may have some correlation, while drainage and runoff may be partially correlated with land cover as well. The underlying data may be uncertain, and this uncertainty may not be represented in the final suitability map. Fifth, a suitability map cannot be used to determine actual recharge performance. A site could be very suitable but not work well due to unrecognized subsurface heterogeneity. These restrictions further promote the use of machine learning with GIS.

5.1 MAR Site Assessment Machine Learning

Machine learning can learn complex predictor-outcome correlations. Measured infiltration rate, recharge rate, groundwater-level reaction, groundwater potential, aquifer transmissivity, groundwater-quality response, successful vs unsuccessful infiltration sites, or groundwater mounding can be used for MAR evaluation. Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting Machine (GBM), XGBoost, ANN, and Logistic Regression are popular techniques. Recently, groundwater research employing machine-learning has grown rapidly. A recent comprehensive evaluation of over 200 groundwater-quality studies shown that ML is frequently utilised for parameter prediction in complicated groundwater systems, making data-driven approach acceptable (Haggerty et al., 2023).

Figure 2 GIS-MCDA to Machine Learning



5.2 Random Forest

Random Forest (RF) is a model that is very appealing for MAR mapping because it can model nonlinear relationships, accommodate a large number of predictors, model interactions, account for different variable scales, estimate variable importance, and can be used for heterogeneous datasets (Salman et al., 2024). A Random Forest model builds several decision trees from a

bootstrapped sample and a random subset of the features. Generally, the final estimate is made based on the collective trees. In addition, variables related to groundwater and its environs, which can also be viewed as spatial variables, represented as $X = \{\text{Slope, Soil, Lithology, K, Dgw, Rainfall, LULC, Lineament, EC...}\}$ can be used for Aquifer Recharge (MAR) application and predictive models. Moreover, spatial variables can be used to

model Aquifer Recharge (MAR), such as Slope, Soil, Lithology, K, Dgw, Rainfall, LULC, Lineament, EC... where KK denotes hydraulic conductivity and Dgw denotes groundwater depth. The combination of these variables define the physical, hydrological, geological and water-quality conditions that affect the performance of MAR. The model can then be used to estimate the probability of MAR suitability, P (MAR suitability) or in order to foretell the ongoing result, such as the rate of recharge. Compared to older, more simplistic approaches that relied on a single criteria to determine appropriateness, this new method makes it possible to more efficiently link complicated relationships among multiple factors.

5.3 SVM For Support

Support vector machines (SVMs) shine when the connection between the predictors and the target variables is not monotonic. By mapping the feature space, kernel functions make it possible to use complex decision boundaries. Using support vector machines (SVMs) to conduct groundwater recharge-potential mapping (GWRPM) has the potential to be especially useful when dealing with small data sets that contain high-dimensional environmental predictors (Wagle & Harikrishnan, 2021).

5.4 XGBoost with Gradient Boosting

Decision trees were used by gradient boosting models to correct errors made by earlier trees. Thanks to XGBoost, gradient boosting is a breeze. Overfitting can ruin these models' predictive abilities, so they'll need some tweaks. Spatial cross validation and separate test datasets can be used to assess them. Evidence from a recent study demonstrating the value of an ensemble of groundwater potential models (including AHP, RF, SVM, and XGBoost) is significant approach in MAR assessment. The ensemble MAR probability model was reported to achieve a higher ROC-AUC value of 0.913 and PR-AUC value of 0.946 for the Awash River study than the lower performance of their AHP-derived groundwater potential index (Wulan Sari & Prabowo, 2025).

5.5 Proposed Integrated GIS-ML Framework

Integrated GIS-ML Framework consists of eight stages:

- (1) data acquisition
- (2) hydro-geotechnical characterization with borehole logging
- (3) GIS preprocessing,
- (4) constraint mapping
- (5) GIS-MCDA screening
- (6) machine-learning prediction
- (7) validation of the model;
- (8) final suitability mapping and field checking

Potential workflow:

Input layers → Feature engineering → Training → Cross-validation → Model tuning → Independent testing → Prediction map

5.6 Groundwater Quality and Remediation Potential Integration

Groundwater quality and contamination risks are directly incorporated into the suitability assessment for Managed Aquifer Recharge (MAR) is a significant improvement over traditional groundwater recharge mapping (Zhang et al., 2019). The MAR suitability model should take into account the quality of the recharge source water and receiving aquifer, potential for site contamination, and site environmental and engineering constraints, in addition to the quantity of water that can be recharged. Overall MAR suitability can be stated as a function of $MAR_{overall} = f(Q_s, Q_a, R, S, C, E)$, that is, the overall suitability of a MAR depends on source-water quality (Q_s), aquifer receiving water quality (Q_a), recharge capacity (R), available storage capacity (S), contamination risk (C), and environmental and engineering constraints (E). Parameters such as electrical conductivity, total dissolved solids (TDS), pH, nitrate, ammonium, iron, manganese, arsenic, fluoride, chloride, sulfate, microbial indicators, and selected organic contaminants can be important parameters for assessing groundwater quality (Haggerty et al., 2023) The use of these parameters enables the evaluation to reveal, in addition to the areas of high physical recharge potential, those areas in

which recharge can have a positive effect on improving the condition of groundwater, or those areas in which recharge can cause a mobilization of contaminants and deterioration of aquifer conditions. The integrated approach is in line with recent U.S. Environmental Protection Agency (EPA) research which calls for both volume of groundwater replenishment and potential effects on receiving aquifers (Imig et al., 2022).

5.7 Contamination Vulnerability Mapping

Prior to managed aquifer recharge (MAR), groundwater vulnerability to contamination should be evaluated, and a high level of recharge vulnerability may also indicate a high vulnerability to groundwater contamination (Elzain et al. 2021). Aquifer depth to groundwater, aquifer media, recharge rate, soil properties, topography, vadose-zone properties, hydraulic conductivity, land use, and the location of potential contaminant sources are some of the factors that can be incorporated into an appropriate vulnerability assessment. The conventional DRASTIC model is a good first approximation to groundwater vulnerability and machine-learning (ML) can increase spatial prediction where there is adequate groundwater-quality data available. For instance, nitrate concentrations can be measured from monitoring wells, and used as a dependent variable, whereas land use, intensity of fertilizer application, groundwater depth, soil texture, rainfall and hydraulic conductivity can be used as predictive variables. A trained ML model can then produce spatial nitrate-risk surface which can be used to mark areas as different risks of nitrate contamination. This information can be integrated into the MAR suitability framework to distinguish high recharge potential with low contamination vulnerability from high recharge potential/high contamination vulnerability. High recharge potential and low contamination vulnerability can then be considered as favorable locations for implementing MAR and areas identified as high risk can be avoided or treated with extra precautions before undertaking recharge activities (Satyam & Patra, 2024).

5.8 Water-Source Compatibility

The suitability of a MAR site is not compatible without having access to a reliable source of recharge water. These recharge water sources include seasonal rivers, stormwater, treated municipal wastewater, properly treated industrial effluent, irrigation return flow, reservoir releases and excess flood water, and harvested rainfall. The spatial and temporal availability of these sources needs to be assessed, as water availability may fluctuate significantly from one season to the next, and from one year to the next. Example: There can be very favorable hydrogeological conditions for recharge, but no reliable source of water and long-term operation of MAR is not possible. On the other hand, a site near a plentiful seasonal river might have adequate recharge water, but the aquifer properties might be inappropriate, or infiltration might be restricted or other physical limitations. Thus, the concept of MAR feasibility should be considered as a product of site suitability, water availability and water quality and can be expressed as $\text{MAR feasibility} = \text{Site suitability} \times \text{Water availability} \times \text{Water quality}$. A more relevant insight into the theoretical potential for recharge, practical feasibility, and sustainable operational feasibility of MAR is provided by combining the three (Ebrahim et al., 2024). Additionally, this combination lowers the inclination to assume that physically favorable conditions are operational feasibility.

5.9 Model training and validation

If the observations are unreliable, then there is no way to rely on machine learning to make predictions. Ideally training data should be representative of appropriate and inappropriate conditions. A typical dataset may contain: 70% training data, 15% validation data; and 15% independent testing data. Random splitting, however, may not be appropriate for spatial data since observations that are near each other can be highly correlated. Therefore, if adequate data are available, it is better to use spatial cross-validation. The following can be used to assess the performance of the classification:

A thorough set of prediction-specific measures is needed to evaluate model performance. Performance in classification is measured using Accuracy $(TP+TN)/(TP+TN+FP+FN)$, Precision $(TP/TP+FP)$, Recall $(TP/TP+FN)$, and F1-score $(2(Precision \times Recall)/(Precision+Recall))$. The Precision-Recall AUC is better for imbalanced MAR datasets than the ROC-AUC, which assesses a classifier's ability to discriminate positive and negative cases across thresholds. To forecast continuous recharging, typical regression metrics include the Root Mean Square Error $(RMSE = \sqrt{(1/n) \sum (y_i - \hat{p}_i)^2})$ and Coefficient of Determination $(R^2 = 1 - [\sum (y_i - \hat{p}_i)^2 / \sum (y_i - \bar{y})^2])$. Because no single metric can determine the best model, the final decision must also guarantee that the method is spatially stable, hydrogeologically sound, and reliable in real-world field settings. Complex algorithms can be "black boxes" for ML environmental mapping. Decision makers must understand why a site is suitable. Explainable AI can fix this (Zhao et al., 2023). Feature significance analysis helps identify key features in random forest or XGBoost. SHAP can also be used to determine if values of a predictor are driving predicted suitability higher or lower. For instance, a model may show that hydraulic conductivity is most significant, groundwater depth is second most significant, soil permeability is third most significant, lineament density is fourth most significant and slope is fifth most significant. As a result, the results are comparable with knowledge of hydrogeology. The more extreme the relationship that the model implies, such as a strong association between a very low hydraulic conductivity and a very high recharge, the more the result should be examined, and not simply taken at face value. Maps of suitability will always have a level of uncertainty inherent in their development because they are derived from environmental, hydrogeological, and socio-economic data which may be incomplete or inaccurately known, or vary between locations. There are a number of uncertainty issues that could affect the results presented in this report, including the following: measurement errors, limited number of borehole observations,

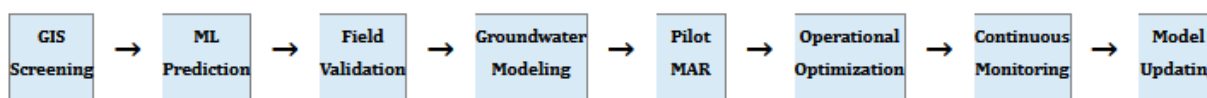
interpolation error, uncertainty of the hydraulic conductivities, uncertainty of recharge rates, temporal variation of groundwater levels, source water availability during the seasons, ML model uncertainty, classification errors, and uncertainty of the weights assigned to the individual suitability criteria. To assess the robustness of the final suitability map, each parameter and criterion weight can be varied in a specified range and a sensitivity analysis performed. For instance, in an input parameter, it can be written as $S_i^* = S_i(1 \pm \delta)$, where the unknown range is assumed to be $\pm \delta$. The areas which consistently fall into one of the highly suitable categories can be categorized in stable high priority MAR zones, while areas that fall into a very different category should be classified as uncertain and should be prioritized for further field investigation. Recent studies on groundwater recharge mapping have been focused on the need for robust spatial stability when input uncertainty is present, highlighting the added value of delineating robust ground water recharge suitability zones from uncertain or transitional zones (Pratavia et al., 2022).

GIS and machine-learning (ML) approaches need to work together for groundwater-flow modeling to be a tool for groundwater-flow planning for MAR, not a replacement. Once priority sites have been determined, groundwater-level changes, recharge mounding, flow directions, contaminant transport, well interference, recovery efficiency, boundary effects, and long-term aquifer storage can be simulated with a numerical model like MODFLOW. Hence, the integrated GIS-ML-groundwater-model workflow can implement a multi-layer approach to site selection and operational planning of MAR, establishing a more reliable operational planning that takes into account spatial screening, data-driven prediction, and physically-based simulation.

GIS screening → ML prediction → field validation → groundwater modeling → pilot MAR → operational optimization.

Spatial multi-criteria analysis and mathematical groundwater modelling have been shown to rank MAR sites (Abbas et al., 2024).

Figure 3: Decision-Support Chain



5.10 Proposed MAR Suitability Classification

A useful classification system divides sites into five categories:

Class	Interpretation	Recommended Action
Very High	hydro-geotechnical & water-quality conditions	Immediate field verification
High	limited constraints	Detailed feasibility study
Moderate	Some important uncertainties	Additional investigation
Low	environmental & technical constraints	Generally,
Unsuitable	environmental, Major physical, regulatory constraints	Exclude

5.11 Water-Stressed Urban and Semi-Arid Regions applications

The framework is especially applicable to urbanizing areas where multiple declines are occurring: groundwater depletion and land-use change. An example of the region is Kabul, Afghanistan. Groundwater decline and suitability of MAR were identified in the past and assessed by GIS and MCDA. Slope, drainage density, infiltration rate, unsaturated-zone thickness, soil type and electrical conductivity were included in the Kabul study. The ANP assessment results were used to classify about 10.3% of the study area as very suitable, 33.1% as suitable, and seven sites were prioritised on land use (Bierkens & Wada, 2019). This case illustrates the usefulness of GIS based MAR planning, but it is possible that other data, such as: groundwater-level time series, borehole yield, geophysical data, aquifer transmissivity, urban imperviousness, precipitation variability, contaminant sources, seasonal river flow, groundwater-quality trends and predicted climate conditions. Furthermore, the outcome would be a dynamic groundwater decision support system, rather than a static suitability map. The Integrated GIS-ML Approach has the following advantages: The integrated approach proposed herein has a number of advantages for MAR planning, including the ability of ML to approximate nonlinear relationships, the ability of GIS-based

mapping to approximate spatial relationships, and the ability to integrate heterogeneous datasets like satellite observations, borehole records, geophysical surveys, soil properties, water chemistry, and hydrological data. It also provides for better planning of groundwater remediation by adding contamination risk and water quality in addition to recharge potential. Moreover, MCDA offers a clear and easily-understood screening system, and explainable ML can pinpoint the significance of the major predictors. The framework is also flexible enough to enable ML models to be updated when new monitoring data becomes available, and can be expanded to facilitate the implementation of water management planning for climate change by the use of future scenarios related to rainfall variability, drought, land-use change and water demand.

5.12 Limitations and Challenges

Though there are benefits, the integrated approach has drawbacks. Performance of ML is dependent on training data quality. Water-level predictions from models may be poor in poorly monitored areas if the observations are sparse or are only from easily accessed areas. Secondly, groundwater systems are in three dimensions while many GIS suitability models are two dimensional. Thus, important vertical heterogeneity may be overlooked. Third,

correlations do not necessarily represent causation. An ML model can learn a statistical relationship with little or no physical basis. Fourth, historical training data may not represent future conditions due to changing climate and land use. Fifth, satellites do not provide a subsurface investigation. Sixth, in addition to being technically suitable, there are regulatory, social, economic and institutional issues to consider when implementing MAR. Last, careful monitoring is needed for groundwater remediation using MAR. The high suitability score is not reason enough to initiate recharge.

5.13 Future Research Directions

Research for future mapping of MAR should be directed towards developing integrated digital platforms for groundwater that incorporate sophisticated numerical approaches and ongoing hydrogeological monitoring. An important direction is physics-informed machine learning (PIML), which uses the principles of hydrogeology and groundwater-flow constraints to enhance the physical reliability of predictions made by ML models. A benefit of adding Spatiotemporal ML is the ability to consider temporal variations in groundwater levels, precipitation, land use, and water quality; and a benefit of adding explainable AI (XAI) is increased transparency and confidence for hydrogeologists and regulators.

5.14 Conclusion

Hydrogeotechnical mapping identifies and prioritises MAR possibilities based on geology, infiltration, groundwater depth to bedrock, aquifer storage, terrain, land use, water supply, and groundwater quality. GIS-MCDA framework enables transparency during preliminary screening, while ML techniques like Random Forest, SVM, Gradient Boosting, XGBoost, and neural networks identify complex nonlinear variable relationships and increase predictive capability. GIS-MCDA, ML, hydrogeological modelling, and field validation work best together. Groundwater quality, contamination vulnerability, aquifer geochemistry, and natural

attenuation are also incorporated, further strengthening the MAR planning. Finally, site selection for MAR should be a process of continuous improvement with the input of new monitoring and field data. Data acquisition, Hydro-geotechnical characterization, GIS screening, ML prediction, Explainability, Uncertainty analysis, Field validation, Groundwater modelling, Pilot MAR, Monitoring, Groundwater modelling updating. This holistic solution can make MAR suitability mapping a proactive groundwater-management system. In water-stressed areas, including rapidly urbanizing and semi-arid areas, it can help restore groundwater resources, enhance drought resilience, ease stresses on over-exploited groundwater resources, and help achieve water security over the long term.

REFERENCES

- Abbas, S. A., Bailey, R. T., White, J. T., Arnold, J. G., White, M. J., Åkerkasova, N., & Gao, J. (2024). A framework for parameter estimation, sensitivity analysis, and uncertainty analysis for holistic hydrologic modeling using SWAT+. *Hydrology and Earth System Sciences*, 28(1). <https://doi.org/10.5194/hess-28-21-2024>
- Alishahi Tabriz, A., Birken, S. A., Shea, C. M., Fried, B. J., & Viccellio, P. (2019). What is full capacity protocol, and how is it implemented successfully? *Implementation Science*: IS, 14(1). <https://doi.org/10.1186/s13012-019-0925-z>
- Bierkens, M. F. P., & Wada, Y. (2019). Non-renewable groundwater use and groundwater depletion: A review. In *Environmental Research Letters* (Vol. 14, Number 6). <https://doi.org/10.1088/1748-9326/ab1a5f>

- Chekirbane, A., Khemiri, K., Martins, T. N., Stefan, C., & Panagiotou, C. F. (2025). A Participatory GIS-based Multicriteria Decision Analysis Approach to Map the Geospatial Feasibility of Managed Aquifer Recharge in a Tunisian Coastal Watershed. *Environmental Processes*, 12(2). <https://doi.org/10.1007/s40710-025-00769-z>
- Ebrahim, G. Y., Stefan, C., Sallwey, J., & Lautze, J. (2024). Mapping the potential of managed aquifer recharge in Africa: GIS-based multi-criteria decision analysis approach. *Groundwater for Sustainable Development*, 27. <https://doi.org/10.1016/j.gsd.2024.101374>
- El Moneam, M. A. (2023). Review of artificial recharge prospects for augmentation of groundwater in Egypt: A case study of El Bustan extension area. In *Ain Shams Engineering Journal* (Vol. 14, Number 7). <https://doi.org/10.1016/j.asej.2022.101995>
- Elghazouly, H. G., Elnaggar, A. M., Ayaad, S. M., & Nassar, E. T. (2024). Framework for integrating multi-criteria decision analysis and geographic information system (MCDA-GIS) for improving slums interventions policies in Cairo, Egypt. *Alexandria Engineering Journal*, 86. <https://doi.org/10.1016/j.aej.2023.11.059>
- EL-Omairi, M. A., & El Garouani, A. (2023). A review on advancements in lithological mapping utilizing machine learning algorithms and remote sensing data. In *Heliyon* (Vol. 9, Number 9). <https://doi.org/10.1016/j.heliyon.2023.e20168>
- Elzain, H. E., Chung, S. Y., Park, K. H., Senapathi, V., Sekar, S., Sabarathinam, C., & Hassan, M. (2021). ANFIS-MOA models for the assessment of groundwater contamination vulnerability in a nitrate contaminated area. *Journal of Environmental Management*, 286. <https://doi.org/10.1016/j.jenvman.2021.112162>
- Glanville, K., Sheldon, F., Butler, D., & Capon, S. (2023). Effects and significance of groundwater for vegetation: A systematic review. In *Science of the Total Environment* (Vol. 875). <https://doi.org/10.1016/j.scitotenv.2023.162577>
- Guo, Z. Q., Li, P., Yang, X. M., Wang, Z. H., Lu, B. B., Chen, W. J., Wu, Y., Li, G. W., Zhao, Z. W., Liu, G. Bin, Ritsema, C., Geissen, V., & Xue, S. (2022). Soil texture is an important factor determining how microplastics affect soil hydraulic characteristics. *Environment International*, 165. <https://doi.org/10.1016/j.envint.2022.107293>
- Haggerty, R., Sun, J., Yu, H., & Li, Y. (2023). Application of machine learning in groundwater quality modeling - A comprehensive review. In *Water Research* (Vol. 233). <https://doi.org/10.1016/j.watres.2023.119745>
- Halim, P. R., Yusuf, A. A., Sumner, A., & Purnagunawan, R. M. (2025). Structural transformation and regional inequality differential in Indonesia. *Regional Studies, Regional Science*, 12(1). <https://doi.org/10.1080/21681376.2025.2565318>

- Hartmann, A., Gleeson, T., Wada, Y., & Wagener, T. (2017). Enhanced groundwater recharge rates and altered recharge sensitivity to climate variability through subsurface heterogeneity. *Proceedings of the National Academy of Sciences of the United States of America*, 114(11).
<https://doi.org/10.1073/pnas.1614941114>
- Hartog, N., & Stuyfzand, P. J. (2017). Water quality considerations on the rise as the use of managed aquifer recharge systems widens. *Water (Switzerland)*, 9(10).
<https://doi.org/10.3390/w9100808>
- Imig, A., Szabó, Z., Halytsia, O., Vracholi, M., Kleinert, V., & Rein, A. (2022). A review on risk assessment in managed aquifer recharge. In *Integrated Environmental Assessment and Management* (Vol. 18, Number 6).
<https://doi.org/10.1002/ieam.4584>
- Islam, M. R., Aziz, M. T., Alauddin, M., Kader, Z., & Islam, M. R. (2024). Site suitability assessment for solar power plants in Bangladesh: A GIS-based analytical hierarchy process (AHP) and multi-criteria decision analysis (MCDA) approach. *Renewable Energy*, 220.
<https://doi.org/10.1016/j.renene.2023.119595>
- Liu, C., & Pan, C. (2024). Quantifying the Infiltration Capacity of High-Turbidity Rivers Under the Conditions of Fine Particle Clogging and Resuspension. *Water Resources Management*, 38(4).
<https://doi.org/10.1007/s11269-023-03729-0>
- McCurry, G., & Pyne, D. (2022). Planning for Managed Aquifer Recharge Projects. *Groundwater*, 60(5).
<https://doi.org/10.1111/gwat.13226>
- Meles, M. B., Bradford, S., Casillas-Trasvina, A., Chen, L., Osterman, G., Hatch, T., Ajami, H., Crompton, O., Levers, L., & Kisekka, I. (2024). Uncovering the gaps in managed aquifer recharge for sustainable groundwater management: A focus on hillslopes and mountains. In *Journal of Hydrology* (Vol. 639).
<https://doi.org/10.1016/j.jhydrol.2024.131615>
- Moumane, A., Elmotawakkil, A., Hasan, M. M., Kranjčić, N., Batchi, M., Karkouri, J. Al, Đurin, B., Gomaa, E., El-Nagdy, K. A., & M. Youssef, Y. (2025). Integrating GIS, Remote Sensing, and Machine Learning to Optimize Sustainable Groundwater Recharge in Arid Mediterranean Landscapes: A Case Study from the Middle Draa Valley, Morocco. *Water (Switzerland)*, 17(15).
<https://doi.org/10.3390/w17152336>
- Mystrioti, C., & Papassiopi, N. (2024). A Comprehensive Review of Remediation Strategies for Soil and Groundwater Contaminated with Explosives. In *Sustainability (Switzerland)* (Vol. 16, Number 3).
<https://doi.org/10.3390/su16030961>
- Pauwels, R., Graefe, J., & Bitterlich, M. (2023). An arbuscular mycorrhizal fungus alters soil water retention and hydraulic conductivity in a soil texture specific way. *Mycorrhiza*, 33(3).
<https://doi.org/10.1007/s00572-023-01106-8>
- Prataviera, E., Vivian, J., Lombardo, G., & Zarrella, A. (2022). Evaluation of the impact of input uncertainty on urban building energy simulations using uncertainty and sensitivity analysis. *Applied Energy*, 311.
<https://doi.org/10.1016/j.apenergy.2022.118691>

- Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random Forest Algorithm Overview. In *Babylonian Journal of Machine Learning* (Vol. 2024). <https://doi.org/10.58496/BJML/2024/007>
- Satyam, S., & Patra, S. (2024). Innovations and challenges in adsorption-based wastewater remediation: A comprehensive review. *Heliyon*, 10(9). <https://doi.org/10.1016/j.heliyon.2024.e29573>
- Stefan, C., & Ansems, N. (2018). Web-based global inventory of managed aquifer recharge applications. *Sustainable Water Resources Management*, 4(2). <https://doi.org/10.1007/s40899-017-0212-6>
- Wagle, S. A., & Harikrishnan, R. (2021). Comparison of plant leaf classification using modified alexnet and support vector machine. *Traitement Du Signal*, 38(1). <https://doi.org/10.18280/TS.380108>
- Wendt, D. E., van Loon, A. F., Scanlon, B. R., & Hannah, D. M. (2021). Managed aquifer recharge as a drought mitigation strategy in heavily-stressed aquifers. *Environmental Research Letters*, 16(1). <https://doi.org/10.1088/1748-9326/abcfe1>
- Wu, D., Li, B., Li, Y., Li, Q., Sheng, C., Liu, J., Wang, M., Min, Y., Feng, J., Zhang, Y., & Yu, J. (2024). Characterization of Groundwater Hydrochemistry and Temporal Dynamics of Water Quality in the Northern Baiquan Spring Basin. *Water* (Switzerland), 16(17). <https://doi.org/10.3390/w16172519>
- Wulan Sari, B., & Prabowo, D. (2025). Analisis Perbandingan Prediksi Harga Rumah Dengan Random Forest, Gradient Boosting, dan XGBoost. *Intellect : Indonesian Journal of Learning and Technological Innovation*, 4(1). <https://doi.org/10.57255/intellect.v4i1.1385>
- Yumna, Witno, Najib, N. N., & Faqi, M. (2024). The Effectiveness of Public Green Open Space Capabilities in Reducing Flooding. *Indonesian Journal of Applied Research (IJAR)*, 5(2). <https://doi.org/10.30997/ijar.v5i2.424>
- Zhang, H., Xu, Y., & Kanyerere, T. (2019). Site assessment for MAR through GIS and modeling in West Coast, South Africa. *Water* (Switzerland), 11(8). <https://doi.org/10.3390/w11081646>
- Zhao, X., Zhao, Z., Huang, F., Huang, J., Yang, Z., Chen, Q., Zhou, D., Fang, L., Ye, X., & Chao, J. (2023). Application of environmental variables in statistically-based landslide susceptibility mapping: A review. In *Frontiers in Earth Science* (Vol. 11). <https://doi.org/10.3389/feart.2023.1147427>
- Zurqoni, Retnawati, H., Rahmatullah, S., Djidu, H., & Apino, E. (2020). Has arabic language learning been successfully implemented? *International Journal of Instruction*, 13(4). <https://doi.org/10.29333/iji.2020.13444a>